From Existing Practices to Best Practices: Improving the Quality and Consistency of Participant Assessment Methods in Cartographic User Studies

By

Travis Maclean White

Submitted to the graduate degree program in Geography and Atmospheric Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Co-Chair: Terry A. Slocum

____

Co-Chair: Stephen L. Egbert

____

Daniel R. Hirmas

____

William C. Johnson

____

Andrea Herstowski

Date Defended: 4 December 2017

The dissertation committee for Travis White certifies that this is the approved version of the following dissertation:

From Existing Practices to Best Practices: Improving the Quality and Consistency of Participant Assessment Methods in Cartographic User Studies

Co-Chair: Terry A. Slocum

Co-Chair: Stephen L. Egbert

Date Approved: 4 December 2017

Abstract

User studies are a cornerstone of cartographic research, delivering valuable insights into the design and use of maps, as well as the attitudes and behaviors of map users. The ever-expanding body of cartographic user studies maintains cartography as a forward-facing discipline: emerging technologies and methods are actively pursued and adopted, rather than dismissed. Alongside this growing, diversifying body of research and methods are two persistent, logistical challenges: the cartographic community's lack of an active database of past and present user studies, and the lack of formal guidelines for the selection of appropriate research methods. Over three articles, this dissertation explores how these challenges affect the critical assessment of user differences, and offers a practical solution to improve the overall quality of future cartographic user studies. The first article reviews participant assessment practices from ten years of refereed user studies: how participants were recruited, which information about the participants was assessed by the study authors, how that information was assessed, and the quality and consistency of the reported study details. The review reveals considerable inconsistency among these practices, either as a result of authors not doing an adequate job in designing their studies or, more likely, reporting on their studies. The second article examines recent developments in open-access data sharing, and uses those principles to guide the development of an online, collaborative, cartographic user study database that I hope cartographers will contribute to, and use, as a tool to guide their own user study designs. The third article examines the concept of best practices, which refers to methods or procedures that have been shown to produce superior results, and introduces a model to allow for their systematic identification. When combined, the database and the best practices identification model should assist cartographers in selecting the most appropriate methods to design a user study. The establishment of a cartographic user study database (CartoBase) and a systematic method for

identifying and implementing best practices have the potential to improve the overall quality of cartographic research.

Acknowledgments

For my loved ones and their extraordinary patience and encouragement and patience again.

Climb Mt. Fuji,

O snail,

But slowly, slowly!

– *Kobayashi Issa*

Table of Contents

List of Figures

List of Tables

CHAPTER 1. INTRODUCTION

1.1 Improving Cartographic User Studies

A persistent challenge facing cartographers is the tension between *old* and *new*. Emerging technologies such as interactive maps challenge established map design frameworks and theories of map usage. Online and mobile platforms challenge the requisite skills that constitute a mapmaker—or a map user. Traditional print maps are being replaced by dynamic, interactive maps. In this state of flux, mapping technology typically emerges first, habits and patterns of use follow, and the appropriate methods to evaluate these uses arise long afterwards.

But suppose that a cartographer wishes to evaluate a new mobile map interface. How should the cartographer decide on the best research questions to ask? How should he or she select the best research methods to answer those questions? The cartographer will likely examine the literature for guidance, but what if the technology under study is so new it is not yet covered? Or what if the research method(s) to be used are so novel that there is no consensus regarding their appropriate use?

This challenge is not limited to technological innovations. Within empirically-driven human subject research, established methods of investigation are giving way to novel approaches and analytical techniques (Gabbard et al. 1999; Bowman et al. 2002; Slocum et al. 2003; Robinson et al. 2005; Kramers 2008; Çöltekin 2015; Griffin 2015). Conventional quantitative methods (e.g., controlled lab experimentation) are being complemented—and even supplanted—by newly-adopted qualitative methods (e.g., User Experience Design approaches to usability testing) (Tobon 2002; Slocum et al. 2004; Robinson et al. 2005; Tobon 2005; Tsou and Curran 2008; Roth et al. 2015). New research findings and new ways of doing research challenge existing cartographic theories in two diametric ways. First, beneficially, by continuing epistemological dialogues that explore, confirm, and contradict the existing body of cartographic knowledge. Second, detrimentally, the rate and variety of new knowledge generation outpaces the community's ability to absorb it, resulting in a lag that affects a variety of stakeholders, including developers, users, and other cartographers developing their own research studies. In 2015, the International Cartographic Association identified this lag as a key research challenge (Griffin 2015; Ooms and Lapon 2015; Roth 2015; van Elzakker 2015; White 2015), and recently published four research agendas that strategize effective solutions (Çöltekin 2017; Griffin et al. 2017; Robinson et al. 2017; Roth et al. 2017).

One goal which shaped those research agendas has been to establish "best practices" guidelines that clarify how to select methods appropriate for cartographic research, including cartographic user studies (Cartwright et al. 2001; MacEachren and Kraak 2001; Slocum et al. 2001; Virrantaus et al. 2009; Moellering 2012; Roth et al. 2017). Conventional cartographic research centers on the *user*: the focus is on understanding how map users read, interpret, interact with, and derive meaning from maps. Accordingly, a wide assortment of qualitative and quantitative practices have been developed to evaluate the user. But without formal guidance, researchers may be unable to distinguish between sub-optimal or ad hoc practices and optimal or "best" practices. Thus, a formalized set of cartographic best practices guidelines is needed to facilitate the goal of pairing research questions with the best available user study methods. In my dissertation, I work toward this goal by focusing on a single component of user studies: the recruitment and assessment of human subjects prior to or after their participation in a user study experiment (*i.e.* "*participant assessment"*). Participant assessments are a useful, small-scale proxy for broader cartographic user studies for two reasons. First, participant assessments are important to nearly every user study, but there is little consistency regarding which methods should be used or how they should be built into user studies. Second, there is currently no consensus as to what constitutes "best practices" for participant assessment.

1.2 Research Challenges and Objectives

My research objectives were designed around participant assessments and address three challenges that stand in the way of establishing best practices guidelines for their selection and implementation in cartographic user studies. Each chapter in this dissertation addresses a separate challenge, building toward a broader goal of identifying and implementing best practices for all components of cartographic user studies, not just participant assessments.

Challenge #1: There has been no attempt to examine existing cartographic user studies in order to determine if they are being appropriately designed, executed, and reported on.

This challenge points to a methodological deficiency: we cannot guide the selection and application of particular study methods when we have not surveyed the methods actually being used, or whether those methods are being used effectively. No cartographer has yet published a review of user studies for the purpose of exploring noteworthy trends or issues affecting existing study methods. Chapter Two provides a systematic review of 104 cartographic user studies published in five refereed journals between 2005 and 2014. The review focuses on participant assessment practices: how participants were recruited, which user difference information was gathered by the study authors, how that information was gathered, and the quality and consistency of the reported study details. My objective was to compile these existing practices, search for notable patterns or issues, and use this information to identify best assessment practices and make recommendations for their appropriate use and reporting.

Challenge #2: Cartographers do not have access to a communal online database to archive user studies, share study details, recommend specific study methods, and store study data.

In order to establish a best practices resource, a necessary step is to first create and maintain a comprehensive online database that inventories the reported components of individual published studies, or "existing practices" (e.g., methods, protocols, participants, materials, procedures, analyses, outcome expectations, and limitations). Chapter Three reports on the development of CartoBase, an open-access, collaborative cartographic user study database. Built on my findings from Chapter Two, CartoBase has three purposes: (1) to aggregate and organize published cartographic user studies into a searchable archive; (2) to serve as an open-source research and communication tool to help cartographers explore issues related to the design, execution, and reporting of user study methods; and (3) to assist in the identification and implementation of best practices for participant assessment in cartographic user studies.

Challenge #3: There is no procedure to help cartographers make individual user study design decisions, such as how to compare and choose between two participant assessment methods.

Best practices guidelines will allow researchers to better define or anticipate which study methods are appropriate for different study contexts. But just as cartographers lack a database to keep track of user study methods and outcomes, so too do they lack a reliable method for identifying best practices. I contend that cartographers already possess a sufficient collection of user study best practices, but lack the necessary resources or procedures to distinguish them from all other existing practices. In Chapter Four, I examine the concepts of best practices and benchmarking to establish a systematic process for identifying best practices for user studies. The process, Cartographic User Study Solution-Driven Benchmarking (CartoSDB), is a novel, six-step benchmarking process that systematically searches and compares published user studies to identify the best available study methods or practices for any study context. CartoSDB defines a specific user study design problem or issue, establishes the criteria sufficient to compare potential "best practice" solutions, then searches CartoBase to identify the solution or practice that meets those criteria. Based on this information, cartographic researchers can make more informed and consistent decisions regarding the design, execution, and reporting of their own user studies.

These three challenges, and my efforts to address each in turn, concern the past, present, and future of cartographic user studies. Whichever guidelines we establish to improve research designs *now* must be founded on evidence from *past* research, and be adaptive to changing or emerging methodologies in the *future*. While the cartographic community has discussed—and continues to discuss—the shape and merits of a best practices resource (e.g., Cartwright et al. 2001; Çöltekin 2015; Roth et al. 2017), my dissertation research represents the first effort to make it a reality and lays the necessary groundwork to develop future best practices guidelines.

1.3 Definition of Common Terms

Six terms are frequently used in all chapters of this dissertation: user, participant, user study, user difference, participant assessment, and best practices. While each term has a variety of cartographic and non-cartographic definitions, I will limit their use to mean the following:

User refers to the intended end-user, audience, or population of a particular map product or service.

Participant refers to a representative user who has been recruited to participate in cartographic research. This term is often used interchangeably with *human subject* and *subject* in the cartographic literature.

User Study refers to any cartographic research utilizing quantitative and qualitative methods to measure and observe the attitudes, behaviors, abilities, and characteristics of study participants. My research focuses on experiment-based user studies, although the term encompasses a variety of study designs and research methods.

User Difference refers to any measurable or observable difference between individuals participating in a user study. This term can apply to any innate or learned characteristic, skill, or ability, such as sex, domain expertise, and sense of direction.

Participant assessment refers to any method used to gather information about user differences outside of the formal study experiment. For example, a pre-experiment questionnaire can record a participant's sex and measure their level of expertise and sense of direction.

Best Practice refers to a method, procedure, or set of guidelines that is considered superior to any comparable alternative practice intended to achieve the same purpose or outcome. What may be considered "best" in one study context may be worse in another study context.

1.4 Organization of the Dissertation

Three different articles found in Chapters Two, Three, and Four, build upon one another as I explore and offer solutions to each of the challenges listed above. Chapter Two reports the findings of a systematic review designed to examine the relationships between user differences, participant assessments, and contemporary cartographic user studies. Chapter Three describes the development of a collaborative, open-access cartographic user study database. Chapter Four describes a best practices identification procedure intended to improve the process of designing cartographic user studies. Each chapter is formatted as a journal article, with conventional Abstract, Introduction, Background, and Conclusion sections. Each chapter contains its own relevant literature review and statements of research problems, questions, and objectives.

In Chapter Five, I summarize how each Chapter contributes to an overall improved understanding of the challenges facing cartographic researchers, and the actions they can take to improve the overall quality of their user studies. I conclude this dissertation with recommendations for future research.

CHAPTER 2. UNDER-DESIGNED OR UNDER-REPORTED? PARTICIPANT ASSESSMENT METHODS IN CARTOGRAPHIC USER STUDIES

Abstract

User studies are a critical component of cartographic research, helping to evaluate the viability of mapping tools, techniques, and phenomena based on the performance of participants recruited for these studies. User differences, which are used to analyze and interpret study outcomes, are typically evaluated using a limited number of performance-based (e.g., completion time and correctness-of-response), skill-based (e.g., expertise and education), or demographic-based (e.g., sex and age) factors. Presented here is a systematic review of 104 cartographic user studies published in five refereed journals over a ten-year span. The review examines methods of assessing participants prior to or after their involvement in research studies and focuses on the extent to which user differences are under-reported and user studies are under-designed. The review reveals that researchers are not consistent regarding which user differences are accounted for, how they are assessed, and how they fit into the overall design of particular studies. Inconsistency is not necessarily indicative of poor study design; nevertheless, inconsistency may cause difficulties when researchers attempt to compare the outcomes of two separate studies. This inconsistency is a result of authors not doing an adequate job of either designing their studies or, perhaps more likely, reporting on their studies. Whether under-designed or under-reported, what this review makes clear is that cartographers need to improve the quality and consistency of user studies.

2.1 Introduction

User differences factor into the design of maps and user studies in similar ways. Within map design, effective design solutions are predicated on an understanding of the similarities and differences that define particular map audiences and their specific needs and preferences that must be accommodated. Likewise, valid and reliable user study designs are predicated on the expectations that the participant sample is representative of the intended users, and that the study uses appropriate methods to assess the relevant individual user differences. But which user differences are most relevant to the interpretation of maps or cartographic user study outcomes? And which methods should cartographers utilize to assess these differences?

User differences, and their influence on user study outcomes, have been a frequent focus of research (Slocum et al. 2001; Swienty 2005; Lobben 2007; Wilkening and Fabrikant 2013). Examples of well-researched user differences include domain expertise or experience (McGuinness 1994; Roth 2009, Baker 2012, Ishikawa 2013; Deeb et al. 2014), sex and gender (Prestopnik and Roskos-Ewoldsen 2000; Sholl et al. 2000; Deeb et al. 2012; Deeb et al. 2014), education (Taylor and Plewe 2006; Golledge et al. 2008; Roth 2009), spatial ability (Hegarty et al. 2002, Dillemuth 2009, Lee and Bednarz 2012; Pingel 2014), and color deficiency (Buckingham and Harrower 2007; Korpi et al. 2013). User differences such as these are important because they provide the basis for comparison of study results (for instance, the performance of experts compared to novices). A variety of methods have been developed to assess user differences, as have procedures to factor such differences into user studies. Unfortunately, as this research reveals, many such methods and procedures are underutilized by researchers.

The question of how user differences "fit" into user studies--which user differences to account for, how to measure user differences, and how to interpret user differences--remain unresolved challenges. To date, cartographers have not established standardized methods for assessing user differences or for determining how those differences impact study results. A lack of standardization is problematic for three reasons. First, it diminishes our ability to compare results among user studies. Second, it leads researchers to utilize ad hoc or inefficient assessment methods alongside reliable and validated methods. Third, it brings into question the quality of the research design and the reporting of published studies.

Assessing user differences enables between-subjects and within-subjects comparisons of participant performance, behavior, and attitudes. All study participants possess a set of innate and learned characteristics, skills, and abilities which affect study outcomes, either individually or in some combination. Time limitations and practicality will influence the number of differences that can be assessed in a study session—researchers do not have time to assess every user difference, and it would be wasteful to do so. But how do we determine which user differences are most relevant to a given context and which are relevant across *all* contexts? And how do researchers currently assess participants as part of their studies? To answer these questions, I conducted an empirical review of 104 published user studies from 96 refereed journals. While there *are* recent reviews of cartographic research methodologies (e.g., Montello 2002; Haklay and Tobon, 2003; Roth and Harrower, 2008; Balciunas 2011; Kinkeldey et al. 2014), no one has published an empirical comparison focusing on participant assessments.

In this article, *assessment* refers to the processes used to gather or measure user difference information not acquired as part of the formal study or experiment, such as demographic data collected in a questionnaire; most typically, assessments occur at the start or end of each participant study session. *Test* and *evaluation*, on the other hand, refer to any user difference information gathered or measured during the user study proper, including preparatory activities such as eye-tracking calibration or practice sessions used to familiarize participants with study task procedures. The "user" in "user study" refers to the intended end-user, audience, or population. *Subjects* or *participants* are representative users who have been recruited for participation in the study.

In the following sections I examine issues related to user differences and their assessment, contemporary user studies design, and the relationship between them. This information is essential for interpreting the participant assessments review that is the subject of this study. My review is limited to quantifiable assessments of users' differences, including qualitative data that were converted into metrics in contrast to, say, information gained from focus groups or expert interviews. Examples from this review include descriptive statistics of user characteristics such as median participant age or male-to-female ratio, discrete measurements of individual abilities such as spatial working memory, and Likert scale responses to a variety of mapping topics, tools, and software.

2.2 User Differences

User differences refer to any measurable or observable difference between two or more study participants. For example, researchers may gather information about participants' age and level of expertise to establish the groups in a between-subjects or between-groups experiment design. Demographic differences, such as sex and gender, are frequently gathered. Domain proficiency or competency in a specified topic or area—for instance, geographic literacy, Python programming language fluency, or familiarity with a study area—often described in terms of *expertise*, *experience*, and *education*, is also common. Many user studies make frequent use of participants' attitudinal feedback, such as Likert scale ratings of opinions and preferences. Less common are measurements of specific perceptual or cognitive abilities, such as sense of direction and spatial awareness.

A variety of methods have been devised to assess user differences. Most common are self-reported surveys and questionnaires; for instance, participants may complete a demographic and domain proficiency questionnaire at the start of a study, and then another to collect feedback at the conclusion of the study. Specially-designed assessments may be deployed to gather specific information, such as the Spatial Thinking Ability Test (Lee and Bednarz, 2012), although these often require a special protocol, software, and/or hardware to execute. Time also influences the selection of assessment methods. Experimenters have a finite amount of time to work with, and overlong studies may adversely affect participant engagement. Self-reported surveys and questionnaires are so prevalent because they require relatively little time to implement; complicated, or unfamiliar assessment methods (such as a mental rotation test) often require greater time commitments, which results in less available time for the study proper.

2.3 Overview of Contemporary User Study Design

Early map research was primarily behaviorist and focused on identifying which map elements and variables users found most functional (Montello 2002; McMaster and McMaster 2002); in contrast, current research largely takes a cognitive-based approach to examine why users find specific elements and variables functional (van Elzakkar et al. 2006; Slocum et al. 2009, 212; Fabrikant et al. 2010; Lobben et al. 2014). In both behaviorist and cognitive approaches, the map user has a prominent role, as the desired outcome of most studies is to improve the quality of cartographic communication between the map and its user. As a result, a common theme in the literature and textbooks is the importance of designing *for* the user (e.g., Robinson et al. 2005; van Elzakkar et al. 2007; Dent et al. 2008; Tsou and Curran 2008).

Contemporary map design and user study design both typically follow *User-Centered Design* (UCD) frameworks, wherein end-users have active roles throughout research processes and their behavior and attitudes influence the final design (Abras 2004, Soegaard and Dam 2012). In cartography, UCD decisions are expected to maximize the qualities of *effectiveness*, *efficiency*, *satisfaction*, and *appropriateness* in a map. Drawing upon standardizations in user-centered and human factors design (e.g., ISO 9241-11:1998; usability.gov), these four qualities may be considered as the most relevant to the map user, as they suggest every map should be clear, concise, enjoyable, and suitable to its context of use. Contemporary cartographic research is based upon this philosophy, and established design guidelines (Tsou 2011; Field and Demaj 2012) utilize this philosophy: maps must accommodate the user's needs and meet his or her expectations; thus, the user should be at or near the center of all research designs.

Recent literature (e.g., Plaisant 2004, Robinson et al. 2005, Demšar 2007; Lobben 2008, Roth and Harrower 2008; Balciunas 2011; Kinkeldey et al. 2014; Coltekin 2015) typically classifies cartographic user study methods according to a few prominent dichotomies: the mode of inquiry (behavioral vs. attitudinal), the general mode of evaluation (controlled experiment vs. usability testing), and the context of use under study (scripted vs. natural) (Figure 1). Researchers must identify which of these dichotomies apply to their particular research agendas, and use these to select from a wide range of available research methods. Potential methods include conventional surveys and questionnaires, card sorting, focus groups, verbal protocol analysis, eye-tracking, usability lab studies, and longitudinal case studies (see, for example, Demsar 2007; Robinson et al. 2005; Roth et al. 2011; Opach et al. 2014). Certain methods are better suited for specific study conditions; for example, eye-tracking can pair well with scripted, controlled studies designed to collect behavioral data (Fabrikant et al. 2008).

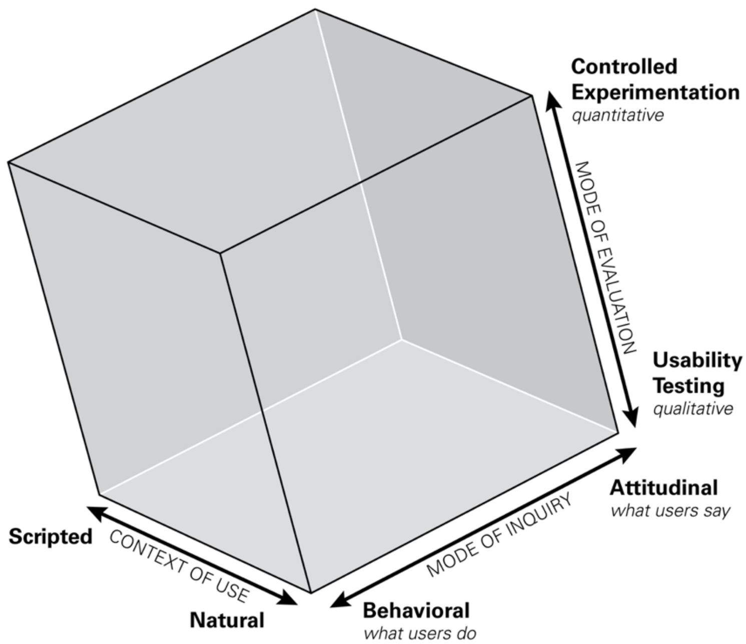


Figure 1. User study methods charted across three dimensions. Mode of Inquiry (behavioral vs. attitudinal), Mode of Evaluation (controlled experimentation vs. usability testing), and Context of Use (scripted vs. natural). Behavioral methods focus on participants' actions, whereas attitudinal methods focus on participants' stated beliefs. Controlled Experimentation methods are typically indirect, collect quantitative data, use large sample sizes, and produce generalizable and repeatable results, whereas Usability Testing methods are typically direct, collect qualitative data, use small sample sizes, and produce specific results. Scripted methods are devised by the researcher to focus on one specific tool, technique, or aspect of use, whereas natural methods try to approximate real-world scenarios. Adapted from Roth & Harrower 2008 and Rohrer 2014.

Once the appropriate method or combination of methods is selected, the remaining components of the study design can be organized. These components are typically organized into four categories (Coltekin 2015, Roth 2015): 1) the *stimuli* or *materials*, which constitute the map and software used in the study, 2) the *participants* or users recruited for the study, 3) the *procedure*, which encompasses the sequence of tasks comprising the study experiment, and 4) *analysis*, or the methods used to analyze and interpret study outcomes. The characteristics of each component also help to determine if the selected method(s) is appropriate for the particular research agenda. This article focuses on issues associated with the *participants* category, specifically how participant differences are assessed as part of a study. Which study method(s) a researcher selects will influence what kind of participant data can be collected, as well as the quantity and level of detail of that data (for instance, an anonymous email survey and a focus group will produce substantially different data).

Through their utilization of UCD frameworks, cartographers are presumed to thoroughly understand the needs, abilities, and limitations of intended users. Yet, this "user first, user last" design principle is more obvious in map design than in user study design. Cartographic texts (e.g., MacEachren 1995; Keates 1996; Slocum et al. 2009) are remarkably consistent in how they model map use situations and detail the role of map users (and their differences) in shaping effective map design solutions. In comparison, the literature does not provide a clear indication of which assessments of user differences should be used in typical user study designs. Experiments are typically built around factors and conditions that allow for obvious between- or within-groups comparisons—for example, comparing female to male performance, or undergraduate geography majors to non-majors—but the degree and extent to which user differences are examined vary considerably. User study design variability is not problematic per se—it is in fact necessary, given the broad scope of cartographic research—but it is troublesome if the researcher relies on inconsistent or ad hoc decision-making when creating a study design.

2.4 Developing a Review of Contemporary Participant Assessments

In this section I discuss the process for developing my review of participant assessments in cartographic user studies. First, I describe the selection of journals from which I obtained my sample of user studies. Second, I discuss the selection and organization of my user study sample, including patterns related to time and journal of publication. Third, I review my approach for organizing both the major components of each user study and the specific information within each study related to participant assessment.

2.4.1 Journal Selection Process

Journal selection was based on two criteria: their visibility among cartographic researchers and relevance to cartographic user studies. I surveyed five prominent refereed journals over a ten-year period, 2005-2014: *Annals of the Association of American Geographers*, *Cartographic Journal*, *Cartographic Perspectives*, *Cartographica*, and *Cartography and Geographic Information Science* (*CaGIS*)[1]. The ten-year time frame allowed me to build a sufficiently large sample of user studies, and helped ensure that these studies would adequately represent the various avenues of current research. While these five journals do reflect the current state of cartographic user studies, cartographers also frequently publish in a variety of related domain journals ranging from GIScience to human-computer interaction and cognitive psychology. It would also be interesting to examine user studies in those journals, but time did not permit me to do so.

2.4.2 Study Selection Process

I identified 153 refereed user study articles published between 2005-2014. For my analysis I retained experiment-based studies that necessitated some type of participant assessment by the

[1] This review originally included the *Professional Geographer*, which was dropped after I could only identify two cartographic user studies published between 2005-2014 which fit my study selection criteria.

authors. Using this guideline, 57 articles were removed from the sample pool: 37 articles did not contain a quantitative assessment component (e.g., case studies and small-sample usability testing), and an additional 20 articles were omitted because the authors did not report any user difference or participant assessment information. This produced a pool of 96 refereed journal articles reporting on 104 individual empirical user studies. A bibliography for this review is available at https://osf.io/ur6xn/. Figure 2 shows a breakdown of studies by journal, by year. In Figure 2A, we see that over ten years, the number of user studies published per year rose from five in 2005 to nineteen in 2014; although the trend is not uniform, there does seem to be a tendency for more user studies in recent years, suggesting that user studies are becoming more important. In Figure 2B, we see that nearly two-thirds of the studies were acquired from *The Cartographic Journal* and *CaGIS*. In comparison, *Cartographica* contains fewer experiment-based studies, the *Annals* has a much broader disciplinary focus, and *Cartographic Perspectives* produces fewer articles than the other journals.

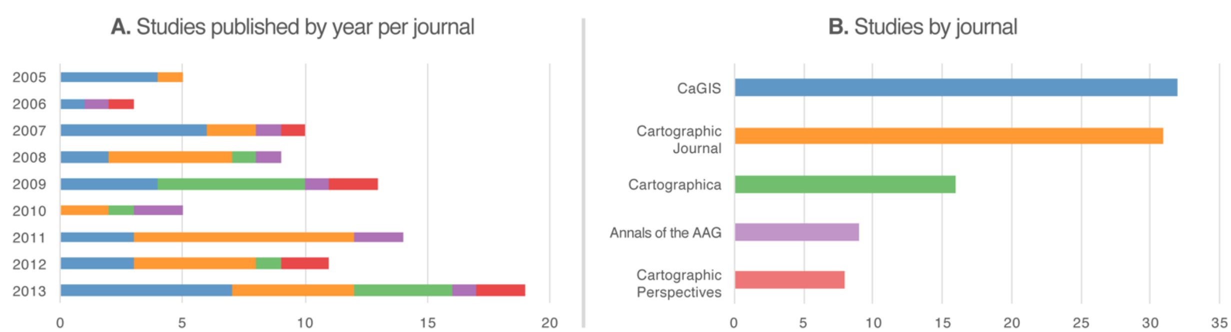


Figure 2. Breakdown of studies published per year by journal (A), and by the total number of studies by journal (B). The number of user study articles published per year rose from five in 2005 to nineteen in 2014. In most years, the majority of studies were acquired from *The Cartographic Journal* and *CaGIS*.

2.4.3 Approach to Organizing User Studies

Once the pool of user studies was finalized, I separated the published details for each study into two major groups: Study Information and Participant Information (Table 1). The components of the two columns under Study Information, *Study ID* and *Study Parameters*, contain descriptive information about both the publication (e.g., author, journal, year of publication) and the study itself (e.g., study design, study setting, session length). The three columns under Participant Information—*Sample Information*, *User Differences*, and *Assessment Methods*—describe the participant information gathered by each study. While many articles followed a "linear" section arrangement—Introduction, Materials, Methods, Results, Discussion—the location of participant information was highly variable. For this reason, I catalogued each study in two steps. First, I inventoried Study Information data, which were typically easy to locate and record. Second, I isolated all components related to issues of user differences and participant assessment and performed a detailed inventory of these. Depending on how an article was organized—for instance, Did the author include a separate subsection for participants?—the second step required substantially more time to complete.

The participant information under review was separated into three basic categories (Table 1). *Sample Information* contains descriptive information about the sample itself: the sample size, how participants were recruited and from where, and if they were remunerated for their participation. *User Differences* contains information about any differences gathered or measured during the study, such as a participant's sex, age, education level, or color blindness. *Assessment Methods* describes the methods used to gather or measure—that is, assess—each user difference.

Table 1. Organization of participant assessment details.

Study Information		**Participant Information**		
Study ID	*Study Parameters*	*Sample Information*	*User Differences*	*Assessment Methods*
Author	Topic of study	Sample size	Sex	Nature of assessment (controlled, self-reported)
Year	Study design	Recruitment method	Age	Timing of assessment
Journal	Inquiry method	Source of participants	Misc./Other info	Assessment medium
Volume	Factorial info	Remuneration method	Domain expertise	Special tools/hardware
Issue	Number of experiments		Education level	Source of assessment
Title	Session length		Definition of Expertise	
Abstract	Study setting		Major/Field	
	Study IVs		Cognitive ability	
	Study DVs		Physiological ability	

Some participant information, such as sample size or age, was reported at higher frequencies than others, such as map-reading expertise or color blindness. It is important to note that many of the recorded variables were context-dependent, and so their absence from a given study is not necessarily indicative of a flaw in that study's design. For example, if an author is not investigating issues related to color deficiency, it may be acceptable to omit assessments of color deficiency from the study. Many articles used a separate *Participants* subsection to report this information. Nearly every published study provided at least some information about their participant samples, and only two studies withheld participant assessment data entirely. The specific findings are reported on in the *Results* section below.

It is important to repeat that this review was performed using *published* information only. As already discussed, authors exhibited considerable variation in the level of detail they provided about participant assessments. While this review contains as much detail as provided by the authors, some authors were more descriptive or forthcoming than others. For instance, Study A provided a descriptive statistical breakdown of its sample by sex, gender identity, age, and grade level, whereas Study B simply reported its sample as consisting of "20 males and 30 females, ages 18 through 24". In both examples the reader is able to ascertain useful information about each participant sample, although the reportage of Study A is more descriptive. However, the absence of reported information does not necessarily mean an author neglected to do something; continuing this example, it is impossible to ascertain if the author of Study B did not collect more descriptive participant information, or if they did but simply chose not to report on it. Nevertheless, my interpretation of the review findings are based on the presumption that each author reported on their particular study fully, honestly, and accurately.

2.5 Results of the User Study Review

In this section I present and discuss the results of my review. First I summarize how participant assessments fit into a typical study, then I explain how I categorized this information. Next I provide context by discussing the settings and lengths of the reviewed studies, how subjects were recruited and remunerated for their participation, and the range of their sample sizes. Following these are separate discussions for the most prevalent user differences gathered across the study sample, including an explanation of their value to user studies in general and cartography specifically, the method of collection or assessment, and any relevant issues raised either by the author(s) or which surfaced during this review.

A typical user study is designed to progress through three major stages: Introduction (including practice or training sessions), Test Session, and Debriefing. Participant assessments may occur at any time during the study, but are typically deployed during the introduction or debriefing stages. Ideally, a researcher can explicate not only why a particular user difference may be critical to interpreting a study's outcomes, but where the assessment should occur in the study. For instance, comments and suggestions made during debriefing may bias subjective questionnaire responses (Nielsen 1993), making it necessary to administer them at the beginning of the debriefing stage.

Most reports communicated the salient user differences of the participant sample in one or a few paragraphs, often in a *Methods* subsection. For example, the following excerpt from Fabrikant et al. (2010, 22) is a prototypical participant profile paragraph:

> "Thirty students (thirteen male and seventeen female) from an undergraduate introductory human geography class, with a mean age of 20.1 years, took part in the experiment. They received a small amount of course credit in return for their participation. The test sample was judged to be a good sample of the desired novice user population, as the vast majority of the participants (72 percent) had not taken more than one geography class (including the first one they were just taking). On average, participants reported that they had used maps only occasionally and had never had training in meteorology, cartography, GIS, computer graphics, or graphic design."

Note the details reported here: sample size, sex ratio, where the recruited participants came from, education level, mean age, remuneration, a definition of a "novice" user, and five self-reported map-related skills. Additional details regarding how and when this information was assessed were provided elsewhere in the paper.

Here is another example, from Deeb (2012, 178):

> "A group of 80 map users participated in the study. They were divided according to the criteria experience and gender. The non-expert group included 50 participants in the beginning of their geographical education, with no previous training in cartography. The other group of participants consisted of 30 experts who work with maps on a daily basis and have at least a master's degree in geography. Of the 80 participants, 35 were female and 45 male divided in balance across experts (15 females and 15 males) whereas the non-expert group consisted of 20 females and 30 males. The average age of the users was 23.3 years, with the experts' average on 27.3 years and the non-experts on 20.4 years. By taking into account experience and gender, it was possible to detect what influence the users' backgrounds had on their preferences regarding labels on maps."

Again, note the reported details: sample size, grouping criteria, definition of a "novice" user, education level, training level, definition of an "expert" user, sex ratios, average ages, and the specific relationships to be examined (the effect of experience and gender on map label preference). These reporting strategies are helpful because they aggregate key user difference information *and* indicate to the reader how the authors will use this assessment data to analyze their study results. The convenience of logical content organization cannot be overstated; and yet only 59% of the studies (n = 61) included a standalone participant section; the remainder scattered this information throughout the paper.

2.5.1 Categorizing Reported Study Information

To systematize my review process, I used four basic categories to indicate how every study reported on each of the variables presented in Table 1:

- **Reported as "not gathered"**: Author explicitly stated they did *not* gather information on or perform some action related to the variable as part of the study. This did not occur very often. For instance, if the author stated they did not assess map expertise, then this was coded "Not Gathered".
- **Reported, details provided**: The author reported on the variable, and provided some level of detail about that variable. For example, with the statement "the median age of participants was 21 years," the author reported on the variable (*age*), and provided one key detail (median age). Note that authors occasionally indicated even more details were gathered than reported on.
- **Reported, no details provided**: The author reported on the variable, but did not provide any details. For example, authors frequently stated that their subjects received an incentive for their participation without indicating what form the incentives took, such as monetary payment or course credit.
- **Not reported**: The author did not report on the variable.

As the results will show, authors exhibited considerable variety in what—and to which level of detail—user difference and participant assessment information was reported. Across the majority of the reviewed studies, authors tended to report information but withheld useful details ("Reported, no details provided"), such as how participant assessment data fit into the study design or were intended to enhance study outcomes. Those authors who did provide sufficient detail ("Reported, details provided"), such as Taylor and Plewe 2006, Dillemuth 2005, and Ooms et al. 2012, explained how specific user differences can significantly impact user performance in experiments, or be used to improve the interpretation of the study results. For example, Dillemuth (2005, 289) stated, "individual differences have a significant impact on performance," while Ooms et al (2012, 240) understood the benefit of participant assessment data as a means to interpret user study results and "eliminate biases and explain…possible distortions in the data." But more often than not, authors failed to discuss the relevance of individual differences to the topic under study, nor did they typically explain why particular differences were assessed.

2.5.2 Study Setting & Session Length

89% of the studies (n = 93) reported information about the setting or location of the study (Table 2). Half of the studies (n = 53) occurred in a lab, 7% (n = 7) were conducted in the classroom, 13% (n = 13) were conducted online, 9% (n = 9) were field studies, and 11% (n = 11) were conducted in multiple locations (e.g., in the lab and field or online and in the classroom). The remaining 11 studies did not provide enough information to determine their specific setting(s). In each reported case, the location was logical given the nature of the study (e.g., controlled experiments in laboratories or navigational exercises in the field), with many of the reporting authors providing the conditions of each setting (e.g., an empty computer lab containing 30 workstations had its window blinds closed and room lights dimmed to minimize environmental distractions).

Table 2. Study settings and session length.

	% reporting study setting	*% reporting length of session*	*Session Length (median, in minutes)*
Study Setting (overall)	89%	34%	30
Lab	50%	17%	30
Classroom	7%	2%	30
Online	13%	5%	20
Field	9%	4%	90 [a]
Multi-site	9%	6%	45

[a] Excludes two field studies, one of which took place over one full day, the other over two days.

34% of the studies (n = 35) reported the approximate or exact lengths of their respective study sessions. The median length of all reported studies was 30 minutes, with online studies taking less time (median = 20 minutes) and field studies taking the most time (median = 90 minutes). Compared to study setting, session lengths were underreported, and these numbers should be taken with a grain of salt. For instance, the review may indicate that the median length of a classroom study is 30 minutes, but only two of the seven classroom studies reported this information. Considering how carefully study authors must manage their allotted experiment times, whether due to limited time availability or the limited attention spans of their participants, this under-reporting is unexpected.

2.5.3 Recruitment, Incentives, and Sample Size

2.5.3.1 Recruitment

Recruitment refers to the method used to obtain subjects for participation in a study, and *incentives* refers to whether subjects were offered some form of remuneration or compensation for their participation. The choice of recruitment method is a key study design issue, as it affects the representativeness of the sample to the intended audience, as well as the credibility of the study results (Nielsen 1993; Nielsen and Hinderer 2003; Panter and Sterba 2011; Mason and Suri 2012; Greiner 2015). Web-based methods such as Listserv emails, Internet message board posts, and Amazon's Mechanical Turk, and direct methods such as in-class announcements, all reach different subpopulations who may or may not represent the intended audience or user. For instance, an in-class announcement may attract a convenience sample of relatively homogeneous subjects such as undergraduates recruited from an intro-level geography class, which introduces selection bias by not accurately represent an entire population of map users. Moreover, differences in study performance occur based on whether participation is voluntary or mandated, and even between subjects who volunteer early versus those who volunteer late (MacDonald 1972; Nielsen 1993).

Table 3. Recruitment, Incentives, and Sample Size Reporting.

	Reported	*Not Reported*
Recruitment	37%	63%
Incentives	39%	61%
Sample Size	98%	2%

Table 3 shows the proportions of reporting and non-reporting for both issues. These articles varied in their levels of published details; some authors referenced the recruitment process in the article text without providing further details, whereas other described exactly how participants were contacted and compensated. Only 37% of user studies (n=38) provided information regarding participant recruitment. The most common methods reported were announcements made in classrooms or distributed through listserv emails. Only two studies used Amazon's Mechanical

Turk. Another performed a large-sample, international, anonymous web-based survey, but did not provide details as to how participants were recruited.

2.5.3.2 Incentives

Incentives are often necessary to encourage participation in a user study. Researchers must take care in selecting appropriate incentives, as they may motivate subjects to perform or behave in unpredictable or unwanted ways (Nielsen and Sova, 2003; Sharp et al. 2006). For example, participants may be overly-positive or overly-critical depending on the incentive type and amount, or may feel obligated to behave in a certain way so as to receive the incentive.

In a fashion similar to recruitment, only 39% of user studies (n = 40) provided incentive information (Table 3); of those, fifteen offered monetary remuneration, fourteen offered course credit, two offered both, and one study entered participants into a prize raffle. Interestingly, eight studies explicitly stated that no incentives were offered; these were perhaps the only instance wherein authors felt it necessary to acknowledge something that *wasn't* included in the study. While recruitment must have occurred in some form regardless of whether the topic was raised in the article, the same cannot be said for incentives. Of the studies that did not report on incentives, it is impossible to say whether incentives were offered or not. Interestingly, only 16 studies reported both recruitment and incentive information.

2.5.3.3 Sample Size

Sample size refers to the number of subjects participating in a user study, and is critical for making inferences about the population the sample represents (Nielsen 1993; Osborne 2008). 98% of the reviewed studies reported the size of their respective samples (Table 3). The reported number of participants per study varied depending on approach and setting; for instance, one usability evaluation (Demsar 2007) used just 4 participants, whereas Midtho and Nordvik (2007) recruited 1,421 participants for their web-based experiment. Grouping by study setting revealed a number of notable patterns (Table 4). Lab studies sample sizes ranged from 4 to 169 participants, with a median of 33. Studies conducted in the classroom—including large lecture halls—ranged from 34 to 739 participants, with a median of 125. Online studies ranged from 61 to 1,421 participants, with a median of 142. Field-based studies ranged from 2 to 1,140 participants (the large sample is from Lorenz et al. (2013), who conducted a multi-year study), with a median of 28. The number of subjects participating in multi-site studies ranged from 16 to 145, with a median of 35.

Table 4. Reported sample sizes, overall and by study setting.

Study Setting	*% reporting sample sizes*	*Sample Size (median)*	*Range*
Overall	98%	43	4 to 1421
Lab	100%	33	4 to 169
Classroom	100%	125	34 to 739
Online	100%	142	31 to 1421
Field	89%	48	8 to 1140
Multi-site	100%	35	18 to 145

All values rounded; 12 studies reported sample sizes, but not study setting, and one study reported neither.

The nature and scope of a study will help determine what sample size is required (other influencing factors include budget, time, and the intended audience population). Bernard (2006) discusses four factors that influence sample size selection: heterogeneity of the population, the number of subgroups in the analysis, the size of the subgroups, and precision of the measurements. Generally speaking, sample size should be large enough to minimize sampling errors among the collected data and produce reliable outcomes. A small sample size can easily obscure differences among the participants comprising a sample. Conversely, methods such as usability studies often require relatively few participants, as additional participants offer only diminished returns (five is common, although others advocate for more; see Nielsen 2000 and Tullis and Albert 2013 for in-depth discussions).

2.5.4 User Difference and Participant Assessment Findings

Of the 104 user studies in this sample, 99% (n = 103) reported information pertaining to at least one user difference (one paper summarized its study in a single paragraph, which contained no participant information). Only 39% (n = 40) of the reviewed studies reported using one or more user difference as an independent variable; the remaining studies used various map or test conditions (such as monitor configuration or symbol density) as independent variables.

Table 5 contains the rates at which authors reported specific user differences, and how often those differences were used as independent variables. The results were as varied as the experiments themselves. The most commonly reported participant information was either demographic in nature ("Sex and Gender," "Age") or related to expertise ("Expertise or Experience," "Education," "Academic Major or Professional Field"). Most information was self-reported by the participant (e.g., responding to "Rate your familiarity with geobrowsers"), although in some studies participants completed validated self-assessments (e.g., the Santa Barbara Sense of Direction Scale, Hegarty et al. 2002), and in fewer still, the researchers performed a controlled assessment directly on the participant (e.g., the Vandenberg Mental Rotations Test, Vandenberg and Kuse 1978).

Table 5. Reported rates of each user difference and their use as independent variables.

	Not gathered	*Reported, details provided*	*Reported, no details provided*	*Not reported*	*Used as IV*
Demographic Information					
Sex and Gender	1%	63%	1%	36%	10%
Age	1%	61%	1%	38%	5%
Misc. Personal Info	0%	9%	3%	88%	8%
Expertise					
Expertise or Experience	4%	54%	10%	32%	5%
Education	0%	38%	39%	23%	1%
Academic Major / Professional Field	2%	36%	28%	35%	11%
Cognitive/Physiological					
Cognitive/Perceptual	0%	20%	0%	80%	11%
Physiological	0%	12%	0%	88%	1%

All values rounded to nearest whole number.

2.5.4.1 Sex & Gender

Sex differences, or comparisons of performance between males and females, are often used to differentiate participant performance in spatial and geographic tasks. A great deal of geographic and other scientific literature is devoted to these topics (e.g., Gilmartin and Patton 1984; Chang and Antex 1987; Boardman 1990; Self and Golledge 1994; Montello et al. 1999; Abramov 2012). Among the studies I analyzed, the more recent studies distinguished between the terms sex (biologically or physiologically defined) and gender (socially constructed), whereas authors of the older studies tended to use the terms interchangeably. 63% of the studies (n = 65) reported the proportion of male to female participants. One study each were classified as "Not Gathered" and "Reported, No Details". 10% of the studies (n = 10) reported using sex or gender as independent variables.

2.5.4.2 Age

A participant's age can greatly influence their ability to perform particular map-related tasks, given the development gap that exists between young children and adults, or the familiarity of digital natives with modern technology to older, potentially less-technically savvy individuals. 61% of the studies (n = 63) reported the ages of their participant samples. Age information was reported in two principal ways: the author either provided descriptive statistics about the participant sample—typically median or average age—or simply stated the range of ages in the sample. As with sex, one study each was classified as "Not Gathered" and "Reported, No Details". Only five user studies reported using age as an independent variable in the analysis.

2.5.4.3 Miscellaneous Personal Information

Compared to sex, gender, and age, there were relatively few reports of other user differences having been assessed. Due to their relative rarity, these differences were grouped into a single catch-all category, *Miscellaneous Personal Information*. Just twelve studies reported having collected additional personal information such as nationality, place of residence, and political affiliation, although only nine of those reported any details. In eight of these studies, the miscellaneous personal information was used as an independent variable (for instance, Patterson and Jenny (2014) used participant nationality to compare differences in the perception of cross-blended hypsometric tints).

2.5.4.4 Expertise, Experience, and Education

Expertise, *experience*, and *education* each relate to some innate or learned skill, knowledge, or ability, and are known to be good predictors of performance (see, for example, Evans 1997; Kobus et al. 2001; Hope and Hunter 2007; and Roth 2009). While many study authors used these terms interchangeably, I used fixed definitions in my review. *Expertise* refers to proficiency in a particular domain, which may be broad (cartography in general) or narrow (mobile map use); *level of expertise* is often described as a dichotomy between novice and expert. Frequently, the boundaries between novice and expert categories are easily blurred, as subjects may be novices in certain areas and experts in others. For example, someone may be an expert at reading topographic maps but a novice at manipulating an interactive map display. Other common categories are *naive* (having no experience) and *intermediate* (somewhere between novice and expert). *Experience* refers to the firsthand acquisition of skills or knowledge within a particular domain; time and frequency are common measures of experience (e.g., daily use of mobile maps or years of work experience making mobile maps). *Education* usually refers to grade level (e.g., third-graders or undergraduates), although it can also refer to formal instruction or training within a particular domain. Generally speaking, *expertise* is a broader measure of domain proficiency, whereas *education* and *experience* can infer where and how proficiency was acquired.

As with sex and gender, the study authors frequently conflated the terms *expertise* and *experience*; for example, “map expertise” and “map experience” were often used to describe the same concept. In total, 54% of the studies (n = 56) reported details about participant expertise or experience, and another 10% (n = 10) discussed expertise or experience without providing substantive details. Despite this relatively high reporting rate, only 15% of the studies (n = 16) used either as an independent variable. Adding complication to this conflation of terms is how novices and experts were defined. 58% of studies (n = 60) provided direct or indirect definitions of *expertise*, or what separated *experts* from *novices*: 15% of these definitions were based on some assessment of expertise, 20% on education, 25% on experience, and the remaining 40% on some combination of the three. The majority of studies relied on inconsistent, ad hoc categorizations of *experts* and/or *novices* using various combinations of expertise, education, and experience. In these studies, a map expert was variously defined as a student who had completed graduate-level cartography coursework; someone who worked with maps on a daily basis; and someone with five or more years of thematic map data analysis experience.

77% of the studies (n = 80) reported the education (grade) levels of participants. This reporting frequency was high for two apparent reasons. First, the majority of these studies obtained convenience samples from university classes (e.g., Sui (2008, 9): "...university students can serve as a surrogate for the general, educated public..."), and their education status was reported along with the sample size ("Forty undergraduates volunteered to participate…"). Second, education was used more frequently than expertise or experience as the basis for making between-groups comparisons of *novices* (e.g., lower-level coursework in cartography) and *experts* (e.g., upper-level coursework in cartography). Only one study used the actual grade level of participants (undergraduates vs. graduates) as an independent variable to group and compare study results.

2.5.4.5 Academic Major or Professional Field

64% of the studies (n = 67) reported participants' academic majors or professional fields. Similar to education above, this high frequency was due to many researchers using convenience samples from undergraduate courses ("Twenty-five geography majors and twenty-five non-majors completed a…”). 30 participant samples were composed in part or entirely of geography majors (including cartography, GIS, and geomatics majors); six of those studies used a subject’s academic major to separate experts (geography majors) from novices (non-geography majors). Ignoring whether random or non-random student samples are representative of larger populations, using a subject’s professed major to establish expertise—particularly when undergraduate majors are not regarded as experts within their academic field—oversimplifies the individual differences that distinguish experts from non-experts.

2.5.4.6 Cognitive and Physiological Differences

I use the terms *cognitive* and *physiological* to refer to a variety of mental or bodily functions, such as memory recall or touch sensitivity. Cognitive abilities receive substantial attention in the literature (e.g., Committee on Support for Thinking Spatially 2006; Hegarty et al. 2002; Lloyd and Bunch 2003; Battersby, Golledge, and Marsh 2006; Lee and Bednarz 2009; Lee and Bednarz 2012: Toker et al. 2012), and authors of such studies typically produce highly descriptive, detailed reports of their participants’ user differences. The authors within my study sample gathered cognitive information only when said information was the topic of the study; for instance, Lee and Bednarz (2012) deployed multiple measures of spatial ability specifically to develop their Spatial Thinking Ability Test, and Buckingham and Harrower (2007) assessed their participants for color vision deficiency in their examination of color saturation in maps.

Among the reviewed studies, 32% (n = 33) included one or more cognitive or physiological assessment of user differences. 20% of the studies (n = 21) assessed one or more cognitive ability, and of those only 9% (n = 9) used one or more of these cognitive differences as an independent variable. Various aspects of spatial ability, such as spatial working memory or mental rotation ability, were the most frequently measured cognitive abilities. Cognitive assessment methods represent perhaps the most reliable and valid methods available to cartographers. Examples of methods used include the Vandenberg Mental Rotation Test (Vandenberg and Kuse 1978, used by Wilkening and Fabrikant 2013), the Santa Barbara Sense of Direction Scale (Hegarty et al. 2002, used by Dillemuth 2009), and the Spatial Thinking Ability Test (developed and used by Lee and Bednarz 2012). Also of note are the affiliated institutions of these studies' authors; fifteen of the reviewed studies were authored or co-authored by researchers from three institutions: University of Zurich, University of California, Santa Barbara, and University of South Carolina.

12% of the studies (n = 12) gathered information on one or more physiological function, none of which were used as an independent variable. The majority of assessments were vision-related—whether participants wore glasses or had a vision impairment, suffered color vision deficiency, or could see stereoscopic three-dimensional imagery (each of these were assessed in four studies). One additional study (Pingel 2014) assessed hearing ability. Interestingly, only one of the twelve eye-tracking studies reported having assessed participants' vision; this is surprising given that vision problems can affect equipment calibration or accuracy.

2.5.5 Review of Findings

While performing this review, four general patterns emerged regarding how researchers assessed and reported on user differences data:

1. They identified the specific participant assessments built into their studies, the data gathered, and how those data benefit the studies (e.g., Hegarty et al. 2002; Dillemuth 2005; Lee and Bednarz 2012; Pingel and Schinazi 2014).

2. They identified specific user differences and how they fit into the study design, but did not explain which assessment methods were used (e.g., Taylor and Plewe 2006; Ishikawa et al. 2005).
3. They identified specific user differences, but did not explain how those differences fit into the overall study design (e.g., Li and Qin 2014).
4. They did not perform participant assessments, or perhaps they did perform assessments but failed to report on them (e.g., Ghosh et al. 2012).

These patterns conform to, and helped shape, the four categories I used to organize the reported user difference and participant assessment data, discussed in *Categorizing Reported Study Information* above (Reported as "not gathered"; Reported, details provided; Reported, no details provided; Not reported). The first and second outcomes are optimal, because they communicate to the reader a high degree of detail and make the study itself more transparent. In the third and fourth outcomes, however, the quality of communication declines.

This points to what is perhaps the most noteworthy finding of the review: cartographic researchers are either under-designing their user studies, or they are under-reporting their user studies, and in most cases it is impossible for the reader to tell the difference. To illustrate this issue, Table 6 displays the "Reported, No Details Provided" and "Not Reported" rates of each study variable discussed in this review, ordered from the highest to lowest "Not Reported" rate. Omitting or not reporting on study details—for reasons ranging from word count limitations to the potentially extraneous nature of the information—can force the reader to infer that information for themselves, which can lead to miscommunication or misinterpretation. To be sure, many "not reported" user difference and participant assessment variables were likely irrelevant to their study contexts (e.g., academic major is irrelevant if the participants are elementary school students). Assuming one could extrapolate these rates to all other aspects of user studies not covered in this review, a lot of information is potentially being left out of user study reports. This raises yet another point—word count limitations may have a direct effect on reporting rates. If true, then researchers would benefit from an additional communication outlet for their study data, such as an online research database and repository.

Table 6. Frequency of "Reported, No Details" and "Not Reported".

Study Detail	*Reported, no details provided*	*Not reported*
Misc. Personal Info	3%	88%
Physiological Info	0%	88%
Cognitive/Perceptual	0%	80%
Recruitment	4%	63%
Incentives	0%	61%
Age	1%	38%
Sex and Gender	1%	36%
Academic Major / Professional Field	28%	35%
Expertise or Experience	10%	32%
Education	39%	23%
Study Setting	0%	11%
Sample Size	0%	2%

Effective communication is an essential part of scholarship, and improves the overall integrity and reproducibility of research (Nosek and Bar-Anan 2012; Nosek, Spies, and Motyl

2012). Omitting content or details from a published report reduces the transparency of that study, diminishes its overall quality, and even calls into question other study decisions made by the researchers. This is not to suggest that the studies reviewed here exemplify poor research, and I am in no way accusing cartographers of designing poor user studies. The issue is that poor communication between authors and their audiences is an unnecessary barrier which forces the reader to make inferences for information that should have been stated explicitly by the author. And, this review indicates that under-reporting is common enough to merit corrective action.

2.5.6 Limitations and Future Directions

There are a few key limitations of the present review. First and foremost, cartographers communicate their research in a variety of formats and settings beyond the five journals selected for this review. The sample of 104 user studies represents only a fraction of the cartographic research that has been conducted over the decades. Second, I excluded dozens of user studies because they lacked a quantitative component, but user differences are important no matter the quantitative or qualitative nature of the study in question. Third, the narrow scope of this review—user differences and participant assessments—revealed similarly narrow findings; a more comprehensive review which accounts for all user study components is required if cartographers wish to uncover all of the study design and reporting practices that benefit, and inhibit, modern cartographic research.

Taking into account the limitations described above, and the relative patterns of under-design and under-reporting that emerged from this review, my next step will be to create a publicly-available cartographic user study database. The expectation is that an expanded database will help cartographers to both catalogue existing studies, and identify and implement best participant assessment practices. In follow-up publications I will discuss the process of database construction and management, and demonstrate how the database can be used as a best practices resource. Cartographers need to exhibit better assessment practices—and better reporting practices—and a continuation of this research will help them meet both ends.

2.6 Conclusion

In this exploratory review, I systematically analyzed 104 cartographic user studies from 96 refereed journal publications. My goal was to identify any observable patterns of user difference assessment that was undertaken prior to or after a subjects' participation in a quantitative user study. I first provided an overview of contemporary user study design and the primacy of the *user* and *user-centered design* principles to cartographic research, then examined how researchers have fit assessments of their users into this research.

Several notable assessment and reporting patterns emerged. First the review provides evidence of which assessment practices are more preferred than others—which is not to be confused with which practices are demonstrably *best*. By way of example, recall that 63% of the reviewed studies assessed participants' domain expertise or experience. Unfortunately, few of those studies utilized valid and reliable assessments to measure and establish group differences; more often, researchers used seemingly ad hoc questions or non-validated means to determine expertise. For instance, Roth et al. (2009) detailed their process and source material used to define and assess expertise, whereas Resch (2013) simply determined expertise by having participants self-identify with one of four groups (No Experience, Some Experience, Experienced, or Expert). From a study design standpoint, the valid, reliable, and documented methods are the better options, but many researchers do not select (and possibly are not even aware of) these methods.

Second, the review findings also indicate that most (although certainly not all) user studies gather only those data immediately relevant to the variable(s) or concept(s) under study, and do not typically use participant assessment data to make broader connections with other published research. This suggests that cartographers can make improvements to existing user study methodologies by developing guidelines to help determine which user differences should be assessed in a given study context, and select appropriate assessment methods for those study contexts. For instance, while nearly every reporting study gathered participant information via a self-reported paper or digital questionnaire, few authors addressed issues such as their justification for selecting a particular survey question or how they avoided question bias. It is impractical to suggest cartographers adopt an exhaustive set of user differences to assess in every user study. Given the diverse nature of cartographic inquiry, the relevance of specific user differences will vary from study to study. A challenge to consider, then, is which user differences should be incorporated into cartographic user studies as a matter of standard practice, using which methods of assessment, and to what degree of specificity, given the often stringent time restraints.

The relevance of participant assessments is not a point of dispute; what is contentious is *which* user differences are relevant to cartographic inquiries, *how* they should be assessed, and whether cartographers are willing to implement a standardized approach for assessing each. Based on these review findings, I encourage cartographers to demonstrate more consistency and care when making and reporting on participant assessments.

CHAPTER 3. CONSTRUCTING A CARTOGRAPHIC USER STUDY DATABASE TO IDENTIFY BEST PRACTICES

Abstract

Over the past few decades, a wide assortment of qualitative and quantitative methods has been developed to conduct cartographic user study research. However, cartographers do not possess an up-to-date online user study database to keep track of existing or emerging user studies, nor the methods used to perform those studies. Introduced here is CartoBase, a prototype cartographic user study database. The paper describes the rationale behind the database and the decisions that shaped its design. CartoBase has three primary purposes: (1) it aggregates and organizes published cartographic user studies into a searchable archive; (2) it provides an open-source research and communication tool to help cartographers explore issues related to the design, execution, and reporting of user study methods; and (3) it can help researchers identify practices and methods to incorporate into their own user studies. Currently, the database contains details of 110 user studies published in five refereed journals between 2005 and 2014. CartoBase is designed to be open-access and collaborative so that the entire cartographic research community can contribute to and continually improve upon the database contents. The long-term goal for CartoBase is to provide the research community with a comprehensive bibliographic and encyclopedic research tool that details the experimental conditions under which every published cartographic user study was performed and reported on. With these tools made easily accessible, CartoBase will help improve the overall quality of future cartographic user studies.

3.1 Introduction

Over the past few decades, cartographic research has expanded to encompass a diversity of questions, methodologies, and technologies. Within experiment-based cartographic research, established methods of investigation are giving way to novel approaches and analytical techniques (Gabbard et al. 1999; Bowman et al. 2002; Slocum et al. 2003; Robinson et al. 2005; Kramers 2008; Çöltekin 2015; Griffin 2015; Roth et al. 2017). Conventional quantitative methods such as controlled lab experimentation are being complemented—and even supplanted—by newly-adopted qualitative methods such as the User Experience Design (UXD) approaches to usability testing (Tobon 2002; Slocum et al. 2004; Robinson et al. 2005; Tobon 2005; Tsou and Curran 2008; Roth et al. 2015; Roth et al. 2017). New research findings and new ways of doing research challenge existing cartographic theories in two diametric ways. First, they do so beneficially by continuing epistemological dialogues that explore, confirm, and contradict the existing body of cartographic knowledge. Second, detrimentally, the rate and variety of new knowledge generation outpaces the community's ability to absorb it, resulting in a lag that affects a variety of stakeholders, including product developers, map users, and cartographers developing their own research studies. The International Cartographic Association has identified this lag as a key research challenge (Griffin 2015; Ooms and Lapon 2015; Roth 2015; van Elzakker 2015), and recently published research agendas to address it (Griffin et al. 2017; Robinson et al 2017; Roth et al. 2017).

A long-standing goal of cartographic researchers has been to establish a "best practices" framework that clarifies how to select human subject research methods appropriate for cartography (Cartwright et al. 2001; MacEachren and Kraak 2001; Slocum et al. 2001; Virrantaus et al. 2009; Moellering 2012; Roth et al. 2017). Conventional cartographic research centers on the user: the focus is on understanding how map users read, interpret, interact with, and derive meaning from maps. Accordingly, a wide assortment of qualitative and quantitative methods have been developed to evaluate the user's attitudes, behavior, and performance. A best practices framework is intended to facilitate the challenge of pairing research questions to appropriate user study methods and to help researchers overcome the knowledge generation-absorption lag. One obstacle to this goal, however, is the absence of a reliable system for tracking the methods, study designs, and outcomes of published user studies.

Cartographic research is composed of decades of user studies published in a diverse number of journals and texts, many of which remain relevant today. And yet, to this day, studies are not easily discoverable; aside from occasional bibliographies such as Finn and Thunen (2014), there is no central resource that consistently keeps track of existing or emerging cartographic research. Moreover, experimental data and metadata are rarely made openly available *and* accessible outside of published reports—this is unsurprising for older studies that predate modern digital storage methods, but it is arguably indefensible for practicing researchers. Researchers who are designing studies or looking to compare results and make inferences using related studies are impeded and unable to move research forward. In lacking these tools, cartographers are unduly prevented from engaging in the best (and what should be standard) possible practices.

To establish a comprehensive best practices resource, a necessary first step is to create and maintain a comprehensive online database that inventories the reported components of individual published studies, or "existing practices" (e.g., methods, protocols, participants, materials, procedures, analyses, outcome expectations, and limitations). This central database should not be a simple bibliography. It must be a tool that guides the design of cartographic user studies, and allows researchers to discover for themselves the strengths and advantages—or weaknesses and disadvantages—of existing user study methods. Ideally, this tool would also provide a practical way to identify and integrate new approaches with traditional methods.

Problem Statement

While examining broader problems associated with proper data storage and sharing, and how to incentivize a user base to contribute enough data and metadata to make a research database and repository widely useful, Open Research advocate Cameron Neylon (2011) wrote:

> "“Making it easy to put your data on the web” and “helping others to find your data” solve problems that most researchers don’t think they have… A successful data repository system will start by solving a different problem, a problem that all researchers recognize they have, and will then nudge the users into doing the additional work [to] make that data useful to other researchers."

What problem do all cartographic researchers share? If our long-term objective is to provide and promote best practices, I assert our fundamental problem is one of organizing and aggregating our human subjects research into a structured, archival database. Cartographers do not possess an up-to-date online user study database, nor does the literature suggest that one is under development. Lacking such a tool has a cascading effect on all facets of cartographic user studies, especially in how they are designed and reported on. An “existing practices” database would provide three immediate benefits as a resource and a tool. First, it would act as a foundation for the “best practices for user study design” knowledge-base desired by the cartographic research community (by detailing the findings, advantages, and successes by which best practices are determined). Second, it would facilitate decision-making during the planning phases of new studies, particularly for researchers investigating similar phenomena or employing similar methods. Third, assuming the database allows collaborative editing, it would expedite researchers' abilities to review, compare, and keep up with existing research.

In the following sections I detail the rationale, methods, and materials used to create a prototype online collaborative cartographic user study research database ("CartoBase" for short). CartoBase serves as a bibliographic, encyclopedic, and research tool detailing the experimental conditions under which individual cartographic user studies were performed and reported on. For purposes of timeliness, manageability and focus, the database is currently limited to the recruitment and assessment of human subjects prior to their participation in their respective studies. Although narrow in scope, this information is sufficient to demonstrate the feasibility and value of a comprehensive user study database. The long-term goal for the database is to expand it to include other study elements, such as materials used in a study, experiment task sequencing, or the statistical methods used to analyze outcomes. While the process for adding this information is relatively straightforward, it will also require considerable time and resource allocations.

More specifically, the database has three purposes. The first purpose is to aggregate and organize published cartographic user studies into a searchable archive. As the body of cartographic research continues to grow—and expand into emerging fields such as visual analytics—so does the need for a dedicated, up-to-date resource to help keep track of who researches which topics, when, and what the associated findings were. The second purpose is to serve as an open-source research and communication tool to help cartographers explore issues related to the design, execution, and reporting of user study methods. Over the past two decades, data accessibility and sharing have become key issues for scientists, research organizations, and funding agencies;

keeping in line with these open access philosophies would improve the transparency of cartographic research and help cartographers evaluate the quality or viability of specific methods or findings. The third purpose is to help identify and implement cartographic user study best practices, and specific to the context of my research, best *participant assessment* practices. Improving access to published research data and findings will not lead to better cartographic research without a framework in place to help determine which methods or techniques lead to the most desirable—or best—outcomes. Thus, CartoBase is designed to be a tool that can facilitate best practices identification and implementation.

3.2 Open Access: Background and Initiatives

Scientific research has recently entered an era of "data-intensive scientific discovery,'' where ''all of the science literature is online, all of the science data is online, and they interoperate with each other.'' (Hey, Tansley, & Tolle 2009, xxx). In the early 2000s, three separate international coalitions of researchers issued public statements on the nature and relevance of Open Access (OA) to scientific research and scientific communication: the Budapest Open Access Initiative (2002), the Bethesda Statement on Open Access Publishing (2003), and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). The Budapest statement (2002) defined the benefits of OA to literature as:

> "...its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet

itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

The basic idea underlying each of these sorts of public statements is that research data and literature (especially public-funded research) should be made available online without price barriers and, where possible, without permission barriers (Suber 2012, 4). The Public Library of Science (PLOS; plos.org) and the Open Access Scholarly Publishers Association (OASPA; oaspa.org) are two prominent leaders of this movement. According to them and other OA practitioners (e.g., Wicherts et al. 2006; Reichman 2011; Suber 2012; Nosek et al. 2015), achieving OA is dependent on access to the Internet, the consent of researchers or applicable copyright-holders, and widespread changes to research and publishing cultures.

OA initiatives are not limited to accessibility, as they provide and promote best practices for research data management, privacy, copyrights, publication, and the time and scope of data sharing (Wicherts et al. 2006; Yarkoni 2010; Reichman 2011; Nosek et al. 2015). Ideally, adopting OA initiatives will enable greater transparency, openness, and reproducibility—what Nosek et al. (2015, 1422) refer to as "core values of science"— as routine components of research strategy. Ware and Mabe (2015) identify rapid technological developments, increasing data volume, growing numbers of publishing scientists, and expanding publication venues as key factors forcing researchers to adopt new data standardization and accessibility strategies. Data management and sharing in particular have received a lot of attention. Changing views of the nature of data ownership—for instance, do publicly-funded data belong in the public domain, and if so when should they be made available, and in which format(s)?—has led to the development of practical tools and procedures that promote data (and research) transparency and access.

3.2.1 Research Data Repositories

Traditional methods of data discovery and access are fast becoming obsolete due to various unnecessary impediments. Too often, the only publically accessible information about an experiment's methods, outcomes, etc. appear in peer-reviewed publications (Chavan and Penev, 2011) which, due to space constraints, make little or no mention of the raw or derived data from which those results are produced, nor the experimental metadata (Paton 2008, 33). Further, the unstructured, textual format of academic journals (Cruz-Toledo et al. 2012, 1) is time-consuming and difficult to search and query, and often exists behind a paywall. For these reasons, there is a growing need for more structured, streamlined methods to preserve and access research data, as well as effective strategies to ensure researchers contribute to and maintain these data.

The emphasis on OA, and the view that OA is both a best practice and an ethical obligation, has led to the development of multiple publication, data management, and data sharing strategies. Pampel et al. (2013) collectively termed these publication strategies "Research Data Repositories" (RDR), and identified four types of RDR: institutional, disciplinary, multi-disciplinary, and project-specific. The primary purpose of an RDR is twofold: first, it aggregates and organizes existing research into a structured, searchable archive; second, it stores digital research data in a variety of file formats for access and reuse by other researchers (Pampel et al. 2013). Currently, the international Registry of Research Data Repositories (re3data.org) has identified and indexed over 1,500 research data repositories from different academic disciplines.

Two noteworthy examples of multi-disciplinary, comprehensive OA and RDRs are the nonprofit Center for Open Science (COS) and Zenodo, both of which launched in 2013. COS envisions "a future scholarly community in which the process, content, and outcomes of research are openly accessible by default" (cos.io/about/mission/ COS 2017). COS's OA strategy is rooted in the belief that OA is itself a best practice, and that structured implementation of OA principles will lead to the discovery of even better practices throughout the entire research lifecycle (Nasek 2012a; 2012b; 2015). In comparison, the general purpose repository Zenodo (http://about.zenodo.org/) was created by the European Organization for Nuclear Research (CERN) and the OpenAIRE Consortium in accordance with the Wilkinson et al. (2016) FAIR principles for scientific data management and stewardship based on open best practices (data should be Findable, Accessible, Interoperable, and Reusable). Both entities recognize that establishing OA practices requires substantial changes to existing incentives and infrastructure that support research, and the business models that dominate academic publication and communication.

Despite the diversity among existing OA and RDR strategies, each offers a number of overlapping benefits (Paton 2008; Tenopir 2011; Pampel et al. 2013; Nosek et al. 2015). These include improvements to:

- **Data Discovery and Sharing**. RDRs improve both the deposition and preservation of research data, and the discovery of relevant results from those data. By aggregating all research in a particular field or under a specific topic of interest, related studies and methods are easier to discover and review, and the need to perform unstructured research queries is minimized. By sharing datasets, data models, algorithms, programs, etc., other researchers can use this information to assist in the design and analysis of their own experiments, or import/export it for their own uses. According to a study by Publishing Research Consortium (2010), data sharing is one of the most important issues among scientists.

- **Validation and Replication**. Open access to experiment data (including raw, derived, and top-level results) enables other researchers to perform re-analysis studies to validate results. Not only is this a key part of the scientific process, but replication studies are valuable training tools. These issues are especially relevant if science is indeed undergoing a "replicability crisis," which refers to both the difficulty of replicating published findings as well as the reluctance of journals to publish replication studies in favor of, say, innovative research (recent examples dealing with this topic include Open Science Collaboration 2015; Baker 2016; Murphy 2016; and Martin and Clarke 2017).

- **Data Integrity and Transparency**. RDRs safeguard against misconduct such as data fabrication or falsification by preserving all experiment data in their original formats. The transparency offered by public repositories, which contain details regarding the design and execution of the research, helps to preserve the integrity of both the data and the research that produced those data. Importantly, researchers can compare published reports against the actual data, thus minimizing the likelihood of reporting bias or miscommunication. There is also a movement to add a Digital Object Identifier (DOI) to data and data models, which can reinforce data integrity by increasing the visibility of the researchers who build the models and collect the data.

- **Best Practices Sharing**. RDRs can help in the identification and adoption of best practices in every facet of research design, execution, and analysis. For example, RDRs can help optimize workflows by minimizing efforts to (re-)collect data if they are already publicly available, and can help researchers identify optimal practices and draft effective implementation strategies in their own research.

In many cases, OA initiatives and RDRs are a practical, effective solution to changing policies of funding agencies. For instance, National Science Foundation grant proposals require data management plans indicating how data will be stored and accessed (NSF 2005; 2011). As part of their recommendations, the German Science Foundation (Schönbrodt et al. 2016) defines standards (and restrictions) for data storage and sharing, how standards should be implemented, and clarifies the roles and responsibilities of researchers that produce and share data, and researchers that use those data. Other international agencies such as the National Institutes of Health (NIH 2006), the Organization for Economic Co-operation and Development (OECD 2004), and the European Research Advisory Board (ERAB 2006) have also established policies that encourage or require the open handling and sharing of digital research data (Pampel et al. 2013).

3.2.2 Moving Toward OA Cartography

The technical, financial, and cultural challenges of adopting OA in cartography are solvable. The big question is not whether the move toward OA cartography will occur but whether cartographers take a proactive or reactive stance in how they implement OA. For instance, the refereed journal *Cartographic Perspectives* has already moved from a subscription-based to an OA format. Clearly, it is in our best interest to devise and implement OA strategies that allow cartographers to adapt to

OA gradually rather than trying to force changes all at once; as Reichman et al. (2011, 704) state, "overcoming the cultural and sociological barriers to increased data access requires changing human behavior." Thus, improvements must occur in measured, deliberate steps that build upon one another, ensuring that newer, better practices introduced at each step are given ample time to become routine.

I argue that CartoBase, my database of cartographic user studies, represents a logical early step toward OA by improving the related processes of data discovery and data accessibility. Searching for existing or emerging research (i.e. data discovery) requires combing through various subscription-based services like ScienceDirect or search engines such as Google Scholar, which are susceptible to publication lags and inconsistent search results. And, even if data *are* discovered, there is no guarantee the researcher will have access to the majority of cartographic research, which is often published in closed-access journals. By archiving cartographic research in an open environment, CartoBase helps raise the visibility of new or obscure analytic approaches, reporting styles, and theoretical or practical interpretations of published research. The next section details the methods and materials used to create CartoBase.

3.3 CartoBase Overview

3.3.1 Database Requirements and Expectations

CartoBase was developed especially for academic and non-academic cartographers engaged in human subjects research. The database was designed to meet five common requirements and expectations of OA databases sourced from a variety of OA advocates (Hey, Tansley, & Tolle 2009; Tenopir et al. 2011; Suber 2012; Nosek et al. 2015). They are:

- **Open access**. The database should be free, open, public, and web-accessible. This includes permissions to download, share, use, and build upon its contents.
- **Comprehensive**. Ideally, the database should remain up-to-date and contain the bulk of published cartographic user studies, as well as the publications referenced by each of those studies.
- **Collaborative**. Database maintenance should occur collaboratively among the cartographic research community.
- **Searchable**. The database contents should be accessible through a variety of search and filtering interfaces.
- **Identifies best practices**. The database should serve as a best practices resource, and assist in the processes of best practice identification and implementation.

These criteria were drawn from multiple sources, including papers and workshop presentations at the 2015 and 2016 annual meetings of the American Association of Geographers, the 2015 and 2017 International Cartographic Conferences, informal discussions with members of the International Cartographic Association Commissions on Use, User, and Usability Issues and Cognitive Issues in Geographic Information Visualization (CogVis), as well as findings from my review of contemporary assessment methods (White 2017a).

As one might expect, the desired features, functions, and content of a truly comprehensive database are so numerous as to be impractical without substantial time and resource investments. For instance, the webpages of both the Use, User, and Usability Issues (http://use.icaci.org/) and CogVis (http://cogvis.icaci.org/) commissions list a number of desired criteria, including links to references across all user studies, information about user study methods and techniques, a detailed

FAQ of why/when/how to implement those techniques, and tips for data analysis. Multiple stakeholders (who are themselves members of the ICA commissions) stated that a cartographic research database should be capable of directing researchers to best practices at every stage of their research, ranging from the best possible match between study questions and methods of inquiry, to the formatting and organization of published articles. In its ideal, overly-ambitious form—one beyond the scope of my research—the desired database will be both encyclopedic and context-aware, in that it will contain information about a broad range of existing user studies, potential research agendas, and possible methods of inquiry, and be capable of adapting to the needs of any researcher using the database to assist in the design of the participant assessment component of their particular study.

In its current form, CartoBase only partially fulfills the six listed criteria. Its limited functionality is purposeful, as my current objective is to demonstrate the database as a proof-of-concept. In the future, as CartoBase grows, so will its functionality, and the stakeholder needs and expectations the database accommodates. I feel that it is critical that the future growth of CartoBase, as well as its maintenance, be collaborative and actively supported by the stakeholders it is designed to serve. So, as this project progresses, it will eventually require the backing of professional organizations and the larger cartographic community in the form of time, resources, and participation.

3.3.2 Methods and Materials

The development of CartoBase occurred over four major phases:

1. Study Inclusion: involved the identification and collection of user studies for inclusion in the database.

2. Study Classification: involved the classification and cataloguing of each user study, as well as the preparation of study details for inclusion in the database.
3. Database Construction: covered the processes of database construction and inputting study classification data.
4. Database Maintenance & Refinement: The fourth phase, which is ongoing, represents the work necessary to maintain the database and make its information accessible for view and use in a variety of software and file formats.

Phase 1: Study Inclusion

The refereed user studies comprising CartoBase were obtained from five prominent cartographic journals published between 2005-2014: *Annals of the American Association of Geographers* (AAG; formerly the *Annals of the Association of American Geographers*), *The Cartographic Journal*, *Cartographic Perspectives*, *Cartographica*, and *Cartography and Geographic Information Science* (CaGIS). These journals were selected due to their visibility among academic cartographers and the frequency with which published user studies appear in these journals. Each journal was accessed via the University of Kansas Libraries database (lib.ku.edu), with the exception of *Cartographic Perspectives*, which is currently the only open access English-language journal devoted to cartographic research (cartographicperspectives.org). In the future, CartoBase will expand to include cartographic user studies published in any refereed journal.

95 user study articles reporting on 110 separate experiments were published between 2005-2014 (Table 7). Each of the 110 studies contained some quantitative measure of individual, within-groups, or between-groups performance, likely explained by specific user differences and thus necessitating some type of participant assessment. 75 additional user study articles published in

these same five journals between the years of 2000-2003 and 2015-2016 have been gathered and await classification. An up-to-date bibliography of CartoBase entries is available at https://osf.io/ur6xn/.

Table 7. User study characteristics by refereed journal.

Journal	Studies	Experiments	Participants [a]
Annals of the AAG	9	12	839
Cartographic Journal	30	33	3161
Cartographic Perspectives	8	8	802
Cartographica	16	20	2440
CaGIS	32	37	3182

[a] Two participant sample sizes were not reported.

The remainder of this article employs the following terminology. The terms “study” and “user study” refer to a single published research event in which human subjects participated. An individual research event is an “experiment,” regardless of whether non-experimental methods were used (such as focus groups or expert interviews). “Experimenter” refers to the individual conducting the research; a single publication can report on multiple experiments. “Participants” refers to the human subjects recruited to participate in the experiment and provide the experimenter with data. “User differences” refers to the measured or observed differences among participants, which may occur before, after, or as part of an experiment. “Participant assessment” refers to the gathering of any user difference information that was not acquired as part of the formal experiment, such as background questionnaires that are completed before an experiment begins. Typically, participant assessments occur prior to or at the conclusion of an experiment session.

Phase 2: Study Classification

Once the user studies were compiled, the published details of each study were classified and catalogued to record every occurrence of user difference and participant assessment data. Specifically, in White (2017a) I recorded every instance where an author collected user difference data or utilized a participant assessment method, and categorized those data using 30 variables. CartoBase uses a similar approach, adding more stringent, descriptive categorizations to the types and levels of detail of the reported information. Study components were classified and catalogued across five dimensions: Study Identification, Study Parameters, Participant Sample Data, User Differences, and Participant Assessment Methods.

Study Identification covers the essential identifying information of individual studies. Categories include bibliographic information such as study title, DOI, author names and affiliations, abstract, and reference information. *Study Parameters* contains basic information about the design and execution of the study, such as method(s) of investigation, study setting, and study session lengths. *Participant Sample Data* contains general information about the participants of each experiment, including participant sample size and recruitment and remuneration methods. *User Differences* manages the specific user difference information gathered over the course of each study, such as a participant's age or level of education, and how those data fit into the reported experiment. Typically, user difference information was used to analyze study outcomes, although authors frequently reported user difference information without explaining its relevance to the study itself. The *Participant Assessment Methods* dimension describes the method(s) used to gather user difference information, including who administered it, and how and when an assessment occurred.

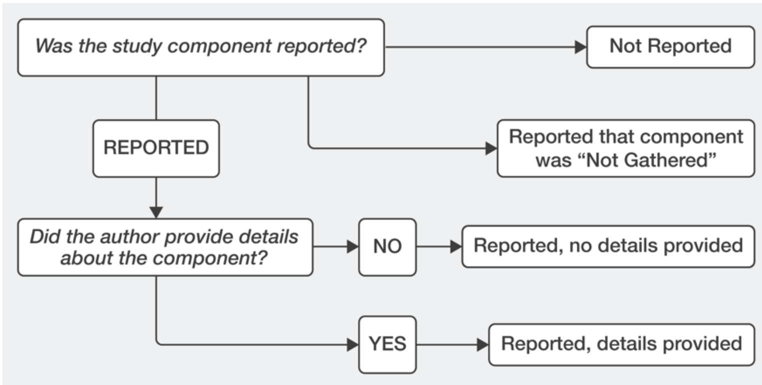


Figure 3. Variability in the reporting level of specificity. Consider this example from Fish et al. (2011, 352): "Seventy-eight students, mostly undergraduates, from 46 different majors at Michigan State University participated in the study. Each participant was compensated $10 for about 30 minutes of his or her time." The following study components were coded *Reported, details provided*: sample size (78), remuneration ($10), sample source (university students), session length (30 minutes). These components were coded *Reported, no details provided*: education level ("mostly undergraduates"), academic major (46 unidentified different majors). The study component "recruitment method" was coded as *Not reported* because the authors did not report this information here (or elsewhere in the article).

While bibliographic information such as the article title or DOI were simply recorded as they appeared in print, the majority of study components, required two levels of classification: first to indicate whether the variable was reported, and second to indicate the level of specificity of the reporting (Figure 3). This schema produced four possible outcomes:

- **Reported as "not gathered":** Author explicitly stated they did *not* collect a specific piece of information or perform some action. For instance, if the author stated they did not assess map expertise, then this was coded "Not Gathered".
- **Reported, details provided**: The information was collected and reported on in the published article. Examples include the range and median age of all participants, or stating that no participants reported suffering color blindness.
- **Reported, no details provided**: Author reported (or made a passing reference to) having collected some piece of information or performed some action, but provided no additional

information. For instance, if an author wrote “females preferred X” without reporting the number or ratio of female participants, the presumption is the author *did* gather this information even if it wasn’t reported.

- **Not reported**: Author did not report any information on the topic.

The contents of CartoBase are limited to the scope of my research—the identification and implementation of user differences and participant assessment best practices. As a result, many critical study design issues such as task ordering or selected methods of statistical analysis were not catalogued for inclusion in the current iteration. In the future, CartoBase will expand to catalogue these missing components. Also, note that CartoBase relies on *published* information; thus, there is no way to distinguish between *study design* omissions (e.g., something did not occur in the experiment) and *reporting* omissions (e.g., something occurred, but was omitted from the published article). The relevance of particular study variables is context-dependent, and their absence from a given study is not necessarily indicative of a flaw in that study’s design. For instance, if color deficiency is not relevant to the topic under study, we wouldn’t expect the study author to assess it as part of the study or address the issue in the published article.

Phase 3: Database Creation

Contemporary cartographic user study design models typically follow User-Centered Design (UCD) frameworks, wherein end-users have active roles throughout research processes and their behavior influences experiment outcomes and the final design of products (Abras 2004, Soegaard and Dam 2012). I created CartoBase according to a UCD framework developed by Robinson et al. (2005), which follows a six-stage iterative process (Figure 4): 1) *work domain analysis*, wherein end-users and developers communicate ideas and requirements of the final product; 2) *conceptual development*, the outline of desired features; 3) *prototyping*, the creation of working models; 4) *interaction and usability studies*, the informal and/or formal assessment of functionality and quality to identify design problems and offer revisions; 5) *implementation*, whereby revisions are made to the interface concept; and 6) *debugging*, whereby the developers address minor stability and usability issues prior to full release. Feedback loops are built into stages two, four, and five as a means to further improve the model.

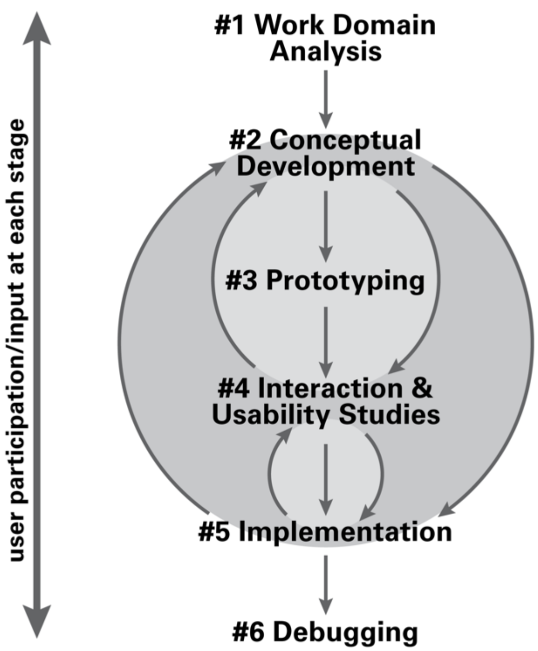


Figure 4. The Robinson et al. (2005) UCD model is an iterative, six-stage process devised for interactive maps, but is flexible enough to be adapted for most cartographic applications. Graphic from Roth et al. (2015).

The advantages of the Robinson et al. (2005) UCD model are twofold. First, there are multiple cases of its successful application within the research community (Fuhrmann et al. 2005; Kramers 2008; Tsou and Curran 2008; Marsh and Haklay 2010; Resch et al. 2013), particularly among cartographers who are also vocal advocates of a best practices resource and database. Second, the model helps organize which activities need to occur, and in which order, throughout the process. As described in this article, CartoBase is a stage 3 prototype, although it will undergo multiple iterations of stages 2 through 5 before entering stage 6.

For *Work Domain Analysis* (stage one) I drew upon three sources of end-user input: a ten-year review of cartographic user studies (White 2017a), presentations from conference paper sessions and workshops (April 2015, August 2015, April 2016), and informal conversations and interviews with colleagues and domain experts (February, April, October of 2016, and February 2017). The desired features of CartoBase, determined during *Conceptual Development* (stage two), are discussed in the previous section on stakeholder needs. The remainder of this section will explain the how the CartoBase prototype was created.

CartoBase was constructed using LibreOffice Base (libreoffice.org/discover/base), a front-end database management system that utilizes a relational database of structured tables to organize and store information, similar to Microsoft Access or Filemaker Pro. LibreOffice is free, open-source software distributed under the Lesser General Public License approved by the Open Source Initiative (available at: libreoffice.org/download/license). Its adherence to open access principles is the primary reason I selected this software over other options. The database and its contents are managed through a personal Open Science Framework project (https://osf.io/ur6xn/).

CartoBase contains five linked tables, one for each of the dimensions discussed in Phase 2 above: Study Identification, Study Parameters, Participant Sample Data, User Differences, and Participant Assessment. Figure 5 illustrates a single database entry. Each table is linked by the article DOI and a unique study ID. This data structure enables a variety of customizable queries to search for specific data or data combinations. Once the table variables were defined, the user study data were added to CartoBase by first reading each study, determining the relevant variables or information, and then entering the variables or information in the respective tables. Due to the variety of article structures used by the reporting authors, there was no consistent method for locating and extracting the relevant variables or information (as previously mentioned, some authors consolidated participant data into standalone subsections, whereas other authors scattered the same information throughout their articles).

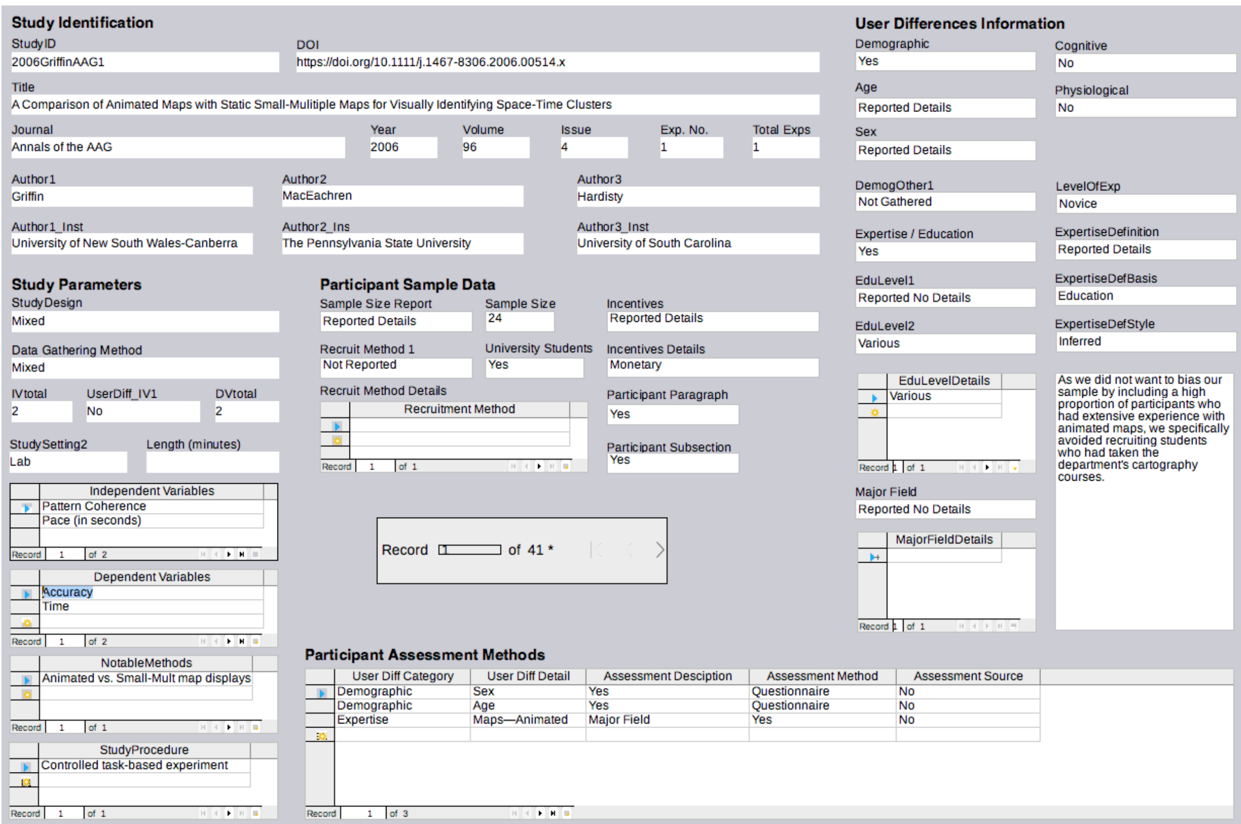


Figure 5. A sample CartoBase entry.

Phase 4: Database Maintenance & Refinement

The adoption and long-term success of CartoBase will depend on continual maintenance and refinement. Maintaining the database requires keeping pace with emerging user studies and adding those studies excluded from my limited sample. Eventually, CartoBase will expand to include those missing studies as well as study details that were purposely omitted from my research; ideally, future updates will occur in a collaborative environment where the researchers responsible for the research under question make direct contributions. In keeping with the tenets of the OA movement, all CartoBase documents are stored in a public COS Open Science Framework (OSF) repository (osf.io/ur6xn/). OSF is an RDR developed by COS to support research project management and collaboration, data archiving, data sharing, and version control, similar to GitHub (github.com). Interested parties can access research materials, view change logs, and if granted the necessary permissions, contribute to or make their own modifications to the database. Refinements will help ensure CartoBase itself becomes (and remains) a user-friendly resource and tool.

Currently, the database is limited by the scope of my research as well as the database software itself. LibreOffice Base is designed primarily for single users, which is sufficient for this iteration of the database, but problematic if attempting to accommodate multiple users. As a consequence, collaboration with other cartographic researchers, which is vital to the growth and longevity of CartoBase, cannot be practically achieved. In order to accommodate these stakeholders, the database will eventually have to migrate to a custom-built, publicly accessible website. As CartoBase grows in size in content and number of users and collaborators, it will be necessary to export this database and transfer it to a more robust, open source web-based database management system such as MySQL.

3.4 CartoBase Demonstration

As a resource and tool, CartoBase addresses several longstanding needs of cartographic researchers, including aggregating published user studies in a central repository, facilitating study design and study reporting decision-making, and providing a tool to identify and implement best participant assessment practices (and, eventually, best practices for all facets of user study research). CartoBase abbreviates the often tedious process of exploring or locating existing studies using multi-disciplinary databases and retrieval systems such as Web of Science, Elsevier Scopus, or Google Scholar, most of which operate behind a paywall; the convenience afforded by CartoBase can save valuable time, and allow the interested user to more quickly locate those studies most relevant to their own. Additionally, when viewing all user studies in aggregate, it is easier to establish groupings of relevant versus irrelevant research using simple search and query functions.

In this section, I present 1) a general data summary of the CartoBase contents, 2) explain the process for accessing and using the database, 3) demonstrate the sort of information the database is capable of providing, and 4) discuss the value of this information to cartographic researchers who are either planning or reporting the participant assessment components of their user studies.

3.4.1 A Data Summary of CartoBase

At the time of this writing, CartoBase contains information pertaining to 110 user studies published between 2005 and 2014 (Table 7). Due to various copyright permission issues, the database contains DOIs and key study details of each article, but not the published articles themselves. Until study authors start providing open access to their research articles and data, and contributing this information directly to CartoBase, CartoBase users will have to access each article independent of the database, such as through a university library system.

CartoBase uses a relational model structure to store the details of 72 selected study attributes into 16 data tables. Figure 6 shows the overall data structure of CartoBase; refer to

Appendix A for additional information on the database structure as well as a complete description of each data attribute. As indicated previously, the study details constitute a large set of information gathered from five broad study categories (*Study Identification*, *Study Parameters*, *Participant Sample Data*, *User Difference Information*, and *Participant Assessment Methods*), with an emphasis on user difference and participant assessment data. The database contains two levels for most participant information: whether the information was reported on (e.g., "Reported, details provided" or "Not reported,"), and the specific published details.

Figure 6. The organization of CartoBase by study attributes and data tables

Table 8 shows the aggregate frequency with which selected data categories were reported. For each dimension there are two or three variables that are more frequently reported than others.

For instance, nearly all authors reported the sample sizes (n = 108, 98%) and settings (n = 99, 90%) of their user studies, but only about one-third reported how participants were recruited (n = 39, 35%) or remunerated for their participation (n = 40, 36%). The reporting rates of some variables, such as education level (n = 86, 78%) or academic major (n = 71, 65%), can be attributed to the prevalence of university-derived convenience samples—the participant samples of 74 studies (67%) are fully or partially composed of graduate and undergraduate students. See White (2017a) for a thorough discussion of trends and issues related to these reporting rates.

Table 8. Reporting rates of selected user study data categories.

Participant Sample Data	*%*	*User Differences*	*%*	*Participant Assessments*	*%*
Sample Size	98	Sex	64	Assessment Method	68
Study Setting	90	Age	63	Assessment Format	42
Recruitment Method	35	Demographics	35	Self-Assessment	50
Incentive Method	36	Education level	78	Controlled Assessment	21
		Major / Field	65	Assessment Source	19
		Level of Expertise	84		

3.4.2 Accessing and Using CartoBase

A fully accessible version of CartoBase in LibreOffice Base format is available at its OSF project page (https://osf.io/ur6xn/), as are version-controlled tab-delimited files of all database entries and a bibliography of current entries. CartoBase functions like most relational databases. All user study records are indexed into related tables to facilitate various browsing, search, query, analysis, and export capabilities. The majority of table relationships are either one-to-many or many-to-many. For instance, one study experiment may be associated with many participant assessment methods, and many authors have published more than one study in CartoBase. For more details on how to access CartoBase, refer to the tutorial in Appendix B.

Each study experiment has a unique primary reference key as well as an easy-to-understand identifier label ("StudyID") containing the first author's last name, year of publication, publishing journal, and an alphanumeric character if multiple experiments took place. To illustrate, the two study experiments reported by Klippel et al. (2011) in the *Annals of the Association of American Geographers* were assigned the StudyIDs "Klippel_2011_AAG_1" and "Klippel_2011_AAG_2", respectively. Users can browse all studies using a standard search interface, filter data using specific criteria, or use queries to find specific studies or compare study data from different categories. Data are viewable in tables, spreadsheets, and interactive forms, and the user can generate summative reports and data visualizations of selected content. A strength of this structure is that new study information can be added by linking new data tables to existing study keys rather than rebuilding entire sections of the database.

3.4.3 Applications of CartoBase

To illustrate how the query and analysis capabilities of CartoBase can be incorporated into a user study design, I will demonstrate three related applications of the database. For purposes of this demonstration, imagine a search for published studies to help design a lab-based, between-groups experiment comparing the performances of expert and novice university students across a series of tasks. This hypothetical scenario is useful because each of the three study parameters—(1) lab-based, (2) between-groups, (3) expert vs. novice—feature prominently among the user studies within CartoBase. The first application demonstrates how to filter CartoBase to retrieve a subset of contextually-relevant studies using the three listed parameters. The second application emphasizes how to filter CartoBase to make meaningful comparisons and evaluations between key study details. The third application emphasizes how CartoBase can be filtered or searched to identify potential methods for assessing user expertise.

Application #1: Retrieve a subset of cartographic user studies that match the parameters of the hypothetical study context.

The first application of CartoBase demonstrates one of its fundamental functions: searching and sorting user study details to help identify a relevant subset of user studies. To accomplish this, I will apply three filters to select records in CartoBase. Access to existing research is critical when designing a new study, as published findings often provide the key insights and inputs that shape emerging studies. Unassisted data discovery typically requires a substantial time commitment, even when using an online database or retrieval system such as Web of Science or Google Scholar; to illustrate, consider how long it would take to locate relevant studies among the 3,000-plus results of a Google Scholar search for "cartography lab between groups expert novice" between the years 2005–2014[2]. As we shall see, CartoBase, on the other hand, optimizes the search and retrieval process because its contents are already limited to cartography, and allows the user to filter studies into a manageable subset using a variety of contextual information about each study.

A key feature of CartoBase is the ability to quickly search or filter the content using one or multiple study parameters. As the goal of this demonstration is to identify a subset of related lab-based, between-groups user studies based on assessments of expertise, we specify three selection

---

[2] Last accessed on October 16, 2017. The direct link is: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=cartography+lab+between+groups+expert+novice&btnG=

parameters: *Study Setting* ("Laboratory"), *Study Design* ("Between-Groups"), and *Level of Expertise* ("novice vs. expert"). Table 9 lists each parameter in order of their reporting rates across all 110 CartoBase studies. The *Study Setting* parameter filters out studies that occurred outside a laboratory and thus practices unsuited (or at least less suited) to a laboratory setting; laboratories are by far the most common setting (48%, n = 53). The *Study Design* parameter filters out studies that did not use a between-groups experiment design, which separates participants into two or more sample groups; only 20% of the user studies in CartoBase (n = 22) conducted a between-groups experiment.

Table 9. Reporting rates of selected study parameters.

Study Setting [a]	*n*	%	*Study Design*	*n*	%	*Level of Expertise*	*n*	%
Laboratory	53	48%	Within-Subjects	48	44%	All levels/Various	41	37%
Online	15	14%	Mixed	40	37%	Novice vs. Expert	18	16%
Field	12	11%	Between-Groups	22	20%	Not Specified	18	16%
Multi-site	10	9%				Novice only	17	16%
Classroom	9	8%				Expert only	16	15%

[a] 99 studies reporting

The third parameter, *Level of Expertise*, indicates which level(s) of expertise was assessed in a study, regardless of domain. User studies are often designed to assess differences within a specific user group (such as within-subjects or repeated-measures designs) or compare two or more groups (between-groups design). *Level of Expertise* is a common grouping mechanism of cartographers. Among the database studies, nearly equal proportions focused solely on N*ovice* participants (16%), *Expert* participants (15%), or *Novice vs. Expert* comparisons (16%) (Table 3). All remaining participant groups were either categorized as *All levels/Various* (37%) or *Not*

Specified (16%). *All levels/Various* means the study authors did not focus on one particular level of expertise or make comparisons between levels of expertise. *Not Specified* refers to studies that did not report on participants' levels of expertise. For purposes of this demonstration, we're interested in comparisons of novice and expert participants, rather than focusing on one specific level of expertise.

Table 10. Subset of studies matching all three filtering parameters.

Filter	*n*	*%*	*Study subset*
All studies	110	100%	
Lab-based	53	48%	
Lab-based + Between-groups	13	12%	
Lab-based + Between-Groups + Novice vs. Experts	3	3%	Lloyd 2005; Bearman 2010; Ooms 2014

Table 10 lists every study that fits all three selected parameters, if the three parameters are applied sequentially. Note how the pool of potential candidate studies reduces from 110 to 3. As one would expect, for every added parameter the number of retained studies decreases. The delimited group of user studies, while certainly not composed of *every* published study that may be relevant to the demonstration context, can provide valuable insights to help guide our hypothetical study design. Additional filtering parameters may be applied to further refine the subset. Unlike traditional literature searches—which often require tedious keyword, abstract, or text searches to find relevant study details—CartoBase allows the user to more quickly identify potentially relevant and useful user studies, and from there make quicker, more confident estimations of the relevance and usefulness of those studies.

Application #2: Identify an existing, high-quality definition of expertise suitable for the hypothetical study context.

The second application of CartoBase demonstrates how it can use contextual information about a user study to assist in the identification and comparison of specific study details. The focus here is to ascertain the quality of the definitions of expertise provided by the study authors. Why is this relevant? Many researchers use expertise as the basis of their between-groups or mixed experiment designs. However, unlike grouping characteristics with fixed dichotomies such as sex (male vs. female), a participant's perceived expertise will vary according to the definition in use. The quality of these definitions are determined both by the criteria used to establish expertise and the amount of detail reported in the published article. As this demonstration will show, some authors are more precise with their definitions of expertise, and thus there are stronger (and clearer) distinctions between groups in certain studies.

As with Application #1, the subset of comparable definitions of expertise were obtained by applying a three-parameter search filter. For this search, however, *Basis of Definition* ("Education") and *Definition Reporting Style* ("Explicit") were combined with *Level of Expertise* ("Novice vs. Expert"), which reduces the pool of studies from 110 to 12 (Table 11). *Basis of Definition* is self-referential; study authors based their definitions of expertise on a limited number of user characteristics, including education, domain experience or expertise, cognitive abilities, or some combination thereof. Because our hypothetical study involves university students, I have retained all studies that use a definition that is wholly or partially based on the participant's education.

Table 11. Subset of expertise definition matching all three filtering parameters.

Filter	*n*	*%*	*Study subset*
All studies	110	100%	
Basis of Definition: Education	41	37%	
Education + Explicit reporting style	29	26%	
Education + Explicit reporting style + Novice vs. Experts	12	11%	Lloyd & Bunch 2005; Cromley & Ye 2006; Çöltekin et al. 2009; Hegarty et al. 2009; Stewart & Kennelly 2010; Klippel, Hardisty, & Li 2011; Baker et al. 2012; Deeb, Ooms, & De Maeyer 2012; Popelka & Brychtova 2013; Deeb et al. 2014; Ooms, De Maeyer, & Fack 2014

Definition Reporting Style refers to the quality of reported definitions, which were either *Explicit* (clearly and explicitly defined in the article text) or *Inferred* (the definition was inferred from clues in the text). When an author does not provide an explicit definition nor adequately explain what makes for a "novice" or an "expert" in their particular study context, it is left to the reader to infer their meaning of expertise using clues from the text. Here is an example of an explicit definition from Stewart and Kennelly (2010, 524), which appears in Tables 11 and 12:

> "We considered as experienced those individuals with some graduate experience in computer science, graphic rendering with shading, and those individuals with some graduate experience in geography with cartographic hill shading."

Table 12 lists the twelve studies whose definitions of expertise fit the hypothetical context of this demonstration along with excerpts of the definitions themselves. While these twelve definitions are varied, each reveals to the reader the measureable criteria used to separate novices from experts. Generally speaking, it is easier to measure or assess expertise (or any user difference) if the criteria used to define it are clearly presented. From the reader's perspective, clear and concise information is easier to process, evaluate, and, if determined to be relevant, adopt for implementation. Of the twelve definitions, two (Cromley and Ye 2006; Çöltekin et al. 2009) are perhaps weaker in the sense that the articles do not wholly explain how their defining criteria were applied to establish two comparable groups. Certainly, other definitions are so specific as to have limited applicability outside their respective study contexts (Hegarty et al. 2009; Stewart and Kennelly 2010), but they still offer guidance as to how they may be adapted for use in other study contexts.

Table 12. Explicit definitions of expertise used to make novice-expert comparisons.

Study	*Definition excerpt*
Lloyd & Bunch 2005	Major vs. minor commitment to geography
Cromley & Ye 2006	Have they taken statistics course or how well do they understand frequency diagrams
Çöltekin et al. 2009	Self-assessed proficiency levels using graphics and spatial data
Hegarty et al. 2009	Experts (meteorology students) had a mean of 43 months' training in meteorology … None of the undergraduate students had studied meteorology at the university level
Stewart & Kennelly 2010	Experienced: graduate experience in computer science graphic rendering with shading; graduate experience in geography with cartographic hill shading
Klippel, Hardity, & Li 2011	Experts enrolled in 400-level course; non-experts enrolled in an introductory information science & technology class
Baker et al. 2012	Experience-based discrimination of expertise: Domain experience questionnaire, geoscience concept inventory; minimum requirement: courses in physical/historical geology, geologic mapping
Deeb, Ooms, & De Maeyer 2012	Experts: work with maps on a daily basis, MA+; Non-experts: students at beginning of geographic education, no previous carto experience
Popelka & Brychtova 2013	Differences between group of cartographers and non-cartographers were tested with use of Wilcoxon rank sum test.
Deeb et al. 2014	Expert: work with maps on a daily basis, MA+ in geography; novice: secondary students, no map training
Opach and Rød 2014	Self-assessed proficiency levels in cartography, GIS, and web technologies
Ooms, De Maeyer, & Fack 2014	Expert: MA + in geography or geomatics with carto training; novice: NO carto training, did not work with maps at professional level

As a whole, the fluid nature of expertise prevents researchers from using common designations that are comparable across all user study contexts. Participants who fit the criteria of experts in study A may be considered novices in study B. Whether one particular type of expertise is more relevant or meaningful than another is highly dependent on the study context—What is being studied? In which circumstances? Using which methods and techniques? And yet, as CartoBase reveals, many cartographers exhibit inconsistency in how they conceive of, define, or assess expertise—even under similar study contexts. CartoBase can facilitate the evaluation of these definitions, but it is up to the researcher to determine which are appropriate to their study topic.

Application #3: Identify suitably robust methods for assessing participant expertise.

The third application of CartoBase demonstrates how it can help identify suitable methods for assessing user expertise. Having identified a subset of studies (Application #1) and definitions of expertise (Application #2) that may be relevant to our hypothetical scenario, we can now examine the various methods used to assess participant expertise. As discussed in Application #2, the criteria used to define expertise should provide clues as to how it can be assessed. For example, an education-based definition of expertise may consider a participant's grade level, academic focus, course enrollment, class performance, or some other education-based measurement.

I argue that an ideal assessment method should possess four characteristics: validity, reliability, timeliness, and practicality. Validity means the method assesses what it purports to assess. Reliability means the assessments are consistent and comparable across studies. Proof of validity and reliability require empirical evidence, which can range from information gathered in a pilot study to refereed articles. I combined these two criteria in CartoBase under the variable

Assessment Source, which indicates whether the study authors provided source data or empirical evidence in support of a particular assessment method. Across all CartoBase studies, just 20% (n = 22) reported their assessment sources. The last two characteristics, timeliness and practicality, speak to the ease of using an assessment method during an experiment. Both are equally important to a researcher with a limited amount of time to conduct all procedures in a user study. A timely assessment method can be deployed in the shortest possible amount of time, and a practical method should be as uncomplicated as possible. Surprisingly, the timeliness and practicality of assessment methods were not discussed in any of the studies currently in CartoBase. Some study authors did discuss both in regards to the overall studies themselves, but not specifically for assessment methods. Thus, in the remainder of the discussion I will focus just on validity and reliability.

Table 13. Subset of expertise assessment methods filtered by level of expertise and source.

Filter	*n*	*%*	*Study subset*
Expertise (all levels)	71	100%	
Novice vs. Expert	18	30%	
Expertise (all levels) + Expertise Assessment Source	2	3%	Roth 2009; Baker et al. (2012)
Novice vs. Expert + Expertise Assessment Source	1	2%	Baker et al. (2012)

To identify which valid and reliable methods were used to assess novice and expert participants, I applied a pair of filters to CartoBase: *Level of Expertise* (“Novice vs. Expert) and *Expertise Assessment Source* (“Reported”). As Table 13 shows, only one study (Baker et al. 2012) used assessments that fit both parameters. Baker et al. (2012) measured expertise using assessments of domain-specific knowledge (a nineteen-question Geoscience Concept Inventory) and experience (the Domain Experience Questionnaire-Geology). Removing the *Level of Expertise* revealed just one other user study (Roth 2009) that reported *Expertise Assessment Source* for *any* level of expertise. Roth (2009) measured participant expertise across three realms (education/training, work experience, and self-reported expertise) and two domains (floodplain mapping, map use and interpretation) using just six questions.

In both Baker et al. (2012) and Roth (2009), the authors were able to generate responses that allowed for precise differentiation between types and levels of user expertise. The primary differences between each method, aside from the specificity of their respective domain topics, relate to their timeliness: whereas the Roth approach utilized just six questions that could conceivably be answered in a short timespan, the comprehensive approach taken by Baker required substantially more time to complete.

As Table 13 shows, 97% of study authors did not report having used empirically validated and/or reliable methods for assessing their participants' expertise. Rather, they appeared to have used ad hoc assessments which were described in varying detail. Some authors were fairly descriptive in how they determined expertise, utilizing an approach similar to Hope and Hunter (2007, 202), one of the eighteen studies comprising the *Novice vs. Expert* subset in Table 13:

> "The participants' experience in dealing with spatial data varied from novice undergraduates to experienced mature-aged postgraduate students employed in professional positions where they were using spatial information to assist in decision-making on a daily basis. The students subjectively classified themselves into one of three experience levels, which were described as: (1) novice (less than six months' experience

using GIS); (2) some experience (between six months and two years of experience of using GIS); and (3) experienced (more than two years' experience of using GIS)."

As CartoBase reveals, some authors were less forthcoming, only mentioning what was assessed without providing descriptive information about the assessment itself. For example, Ooms et al. (2012, 240), another study in the *Novice vs. Expert* subset, reported "…a questionnaire was presented that gathered background information, structured in several themes [including] familiarity with web maps…" but did not state how the questionnaire assessed web map familiarity. Other authors simply approximated participant expertise using nominal assignments, such as geography majors vs. non-majors. Unfortunately, as this application of CartoBase also reveals, too many study authors omit relevant information about the participant assessment methods used in their respective studies. This last finding reinforces the potential value of an OA database, which could store details such as those which were omitted from their final publications.

3.4.4 Demonstration Conclusion

CartoBase is capable of supplementing user study research decisions by effectively and efficiently representing a variety of contextually-relevant user study details. Whereas many of these details would otherwise remain obscured or difficult and time-consuming to obtain, CartoBase strives to make them as visible and organized as possible. The act of archiving and managing minute study details in a public forum makes existing research more accessible and transparent. From an author's perspective, organizing all of the details of their study and making them comparable with all other user studies is a form of quality assurance. From a reader's perspective, this enables researchers to more effectively use published studies to guide and strengthen their own user studies, and the approach may even lead to new insights about the topics they study, how they study them, and how they communicate their findings.

3.5 Conclusion

The purpose of CartoBase is to demonstrate how a sizable body of user study research can be aggregated and organized in an open access environment as a means to improve the quality and consistency of emerging cartographic research. Current processes for accessing and examining the details of published user studies are unnecessarily restrictive and at odds with the modern scientific move toward greater data accessibility, management, and sharing. The fact that cartographers lack an online cartographic research database—even a generic search-and-query bibliography—is unacceptable. Contradictorily, while such a database has long had widespread support, there is no evidence of any strategic follow-through; this is likely an effect of the time and resources necessary to implement such a database.

Establishing a truly comprehensive database of existing cartographic user studies is a long-term, resource-intensive undertaking beyond the practical scope of my research. The database is currently limited to a subset of cartographic research, from a small selection of peer-reviewed journals and a limited time frame, and has been tailored to investigate my specific research questions. There are some actions CartoBase cannot, and should not, perform. For example, while CartoBase provides information to help researchers make study design decisions, it does not make those decisions itself (nor is it capable of doing so). Nor does CartoBase render judgment on the perceived quality of a published study or the methods used by a researcher. It is up to the due diligence of the researcher to gather and weigh such information, and choose the best possible courses of actions. Because CartoBase only provides aggregated, categorized overviews of published user studies, it is up to the researcher to take the extra step of examining the text of the published study in detail.

CartoBase is intended to supplement current study design and reporting processes. However, there are a number of challenges to address if it is to evolve and gain acceptance in the cartographic community. The long-term challenges range from gathering more than just a limited study sample, to releasing a fully collaborative, certified RDR and database for the international cartographic community, and responding to these challenges will require considerable time and resource commitments. Perhaps even more onerous are the challenges of changing our present research and publication cultures. While raising awareness of OA may be a relatively simple matter, convincing researchers to incorporate OA as a matter of standard research practice is not. However, encouraging cartographers to share their data in an OA environment such as CartoBase is a logical—and valuable—step toward addressing these challenges, and improving the overall quality and consistency of cartographic research.

Bearing such limitations and challenges in mind, the current iteration of CartoBase serves as both an operational template and a proof-of-concept as my research progresses and the database evolves. The long-term goal of CartoBase, as both a bibliometric, encyclopedic database and cartographic research tool, is to enable the identification and implementation of best practices at all stages of cartographic user study design, execution, analysis, and reporting. The specific examples of user differences and participant assessments—often under-reported or overlooked entirely in the cartographic literature—effectively illustrate the value this database brings to the study design and reporting workflow.

The OA characteristics must also play a central role in CartoBase's continual development. Conventional cartographic research is becoming more open access and collaborative in nature. As long as we lack the tools and resources that encourage OA behaviors—improving data discovery and sharing, encouraging data integrity and transparency, and facilitating the identification and adoption of best practices—we put the quality of our own research at an unnecessary disadvantage.

CHAPTER 4. A PROCESS TO IDENTIFY AND IMPLEMENT BEST PRACTICES IN CARTOGRAPHIC USER STUDIES

Abstract

A key issue confronting the cartographic community is the notion of "best practices" and its lack of implementation within cartographic user studies. A review of the current literature indicates that cartographers embrace a diverse body of optimal, sub-optimal, and ad hoc qualitative and quantitative methods to conduct and report on their user studies. Building on the question of how cartographers can improve quality and consistency in user studies (White 2017a) and the development of a cartographic user study database (White 2017b), this paper examines the concept of best practices to establish a systematic process for their identification and implementation in user studies. The proposed process, Cartographic User Study Solution-Driven Benchmarking (CartoSDB), is a novel, user-friendly benchmarking process that systematically searches and compares the contextual information of published user studies to identify the best available study methods for any study context. Utilizing a simple six-step procedure, CartoSDB defines a specific user study design problem or issue and establishes the criteria sufficient to compare potential "best practice" solutions and the study contexts for which those practices were originally used. Based on this information, cartographic researchers can make more informed and consistent decisions regarding the design, execution, and reporting of their user studies.

4.1 Introduction

A key issue confronting the cartographic community is the notion of "best practices" and its lack of implementation within cartographic research (Çöltekin et al. 2017; Roth et al. 2017). Best practices refer to methods, procedures, or sets of guidelines that are considered superior to any comparable alternative practices intended to achieve the same outcome. Cartographic user studies refers to experimental research utilizing quantitative and qualitative methods to measure and observe the attitudes and behaviors of map users (Çöltekin 2015). By using and adhering to best practices, cartographers stand to gain reliable and valid results from user studies, including their design, implementation, and publication of the study's outcomes.

Unfortunately, cartographers have never formally established cartographic best practices, nor the conditions necessary to identify them. Instead, cartographers embrace a diverse body of optimal, sub-optimal, and ad hoc qualitative and quantitative methods to conduct and report on user studies. As established methods of investigation give way to novel approaches and analytical techniques, it is critical that cartographers effectively pair particular research questions with the best methods available (Montello 2002; Çöltekin 2015). The challenge is recognizing which practices can produce a desired outcome, and then being able to select the best available practice for a given study context.

Building on my question of how cartographers can improve quality and consistency in user studies (White 2017a) and my development of a cartographic user study database (White 2017b), this paper examines the concept of best practices in user studies to establish a system for their identification and implementation. Based in part on the findings of the aforementioned articles, I propose a novel, user-friendly best practices identification process suitable for decisions affecting the recruitment and assessment of participants prior to their involvement in a cartographic user study. The identification process, and my recommendations, are based on a large body of *best practices* and *benchmarking* literature that emerged in the 1980s (Camp 1989), matured in the 1990s (Bogan and English 1994; Camp 1995; Keehley 1997), and which continues to shape professional (Anand and Kodali 2008), scientific (National Research Council 2005; Cresswell et al. 2011; Oakley and Daudert 2016), and academic (Alstete 1995; Bifulco and Bretschneider 2001) landscapes. My ultimate intention is that the process will enable researchers to efficiently identify and implement the best available practices throughout a user study lifecycle.

The paper proceeds in seven sections. I begin with a review of my related research, with special emphasis on user differences and participant assessments. Next, I preview user study contexts, which have an important role in my best practice identification method. Following this, I critique the concepts and functions of best practices, clarify the need for best practices, and examine the conditions necessary to correctly identify them. Next, I provide critiques of existing benchmarking models used to identify best practices and their applicability to cartography. Then I propose and demonstrate a novel benchmarking model, Cartographic User Study Solution-Driven Benchmarking (CartoSDB), to identify best practices when designing user studies. Throughout the article, I address the short- and long-term challenges associated with identifying, implementing, and maintaining best practices within cartographic research. The intended outcome of this paper is *not* to produce a list stating which research practices are better or best, or instruct researchers in their correct implementation. Rather, my expectation is that the reader will gain a stronger appreciation for the necessity of pursuing best practices, and adopt my benchmarking process to make their own effective, efficient best practices decisions.

4.2 Review of Related Research

The purpose of this section is to review the research related to and underlying my search for participant assessment best practices. The first section introduces two important components of user studies, user differences and participant assessments, and the issues that compelled my search for best practices. The second section outlines the findings from a review of cartographic user studies (White 2017a), and how that led to the creation of a cartographic user study database known as CartoBase. The third section describes four purposes of CartoBase, a one-of-a-kind, open-access, collaborative user study database developed for the cartographic research community.

4.2.1 The Importance of Assessing User Differences

User differences and participant assessments are the application areas underlying my search for best practices. The term *user difference* refers to any measurable or observable difference between individuals participating in a user study, and can apply to any innate or learned characteristic, skill, or ability. Age, sex, domain expertise, and cognitive abilities such as spatial awareness are common examples. *Participant assessments*, the term given to any process outside the formal study used to gather or measure user difference data, are an essential component of user studies because they frequently form the basis for comparing study outcomes. Examples of participant assessments range from self-reported background questionnaires to specialty tests such as the Vandenberg Mental Rotations Test (Vandenberg and Kuse, 1978). The assessment of user differences is a critical component of nearly every user study, because 1) user differences influence user study outcomes, and 2) assessing those differences enables direct comparisons of study participant performance, behavior, and attitudes.

Whereas cartographers recognize the importance of user differences, they do not subscribe to standardized ways of assessing them. Three factors are responsible for the uneven variety of assessment methods currently in use:

1. The user difference being assessed, and the conditions under which they are assessed. Substantially different methods are required for assessing user differences such as age and spatial working memory, in different experiment environments, and when using different hardware and software.
2. The lack of validated assessment guidelines. In lieu of formal assessment design and selection guidelines, and based on my previous research (White 2017a), researchers rely on a combination of non-cartographic sources, their own best judgment, or other published user studies.
3. Uneven quality of published studies. As discussed in Section 2.2 below, study authors frequently omit useful assessment details from their published articles, which prevents other researchers from adopting those methods for their own use.

Taken together, these factors prevent cartographers from maintaining some measure of consistency from user study to user study. While the first factor is necessary and unavoidable, cartographers can take steps to curtail the other two. Establishing validated assessment guidelines and encouraging authors to include more details about participant assessments in their published article will help improve the overall quality and consistency of all participant assessment methods. If we produce community-authored assessment guidelines *specifically for cartographic research*, either as an edited volume or as a collaborative online resource, then cartographers can base a greater proportion of their study decisions on expert-derived knowledge. If we produce similar guidelines to standardize how user studies are reported, then we can set a "baseline" for the expected quality and quantity of future study articles.

4.2.2 Review of Cartographic User Studies

The literature provides few clues regarding how cartographers fit assessments of user differences into their user studies. To investigate how cartographers handle this in practice, I conducted a systematic review of 104 user studies published in five refereed journals over a ten-year span (2005-2014) (White 2017a). The goal was to identify any observable patterns regarding the selection of user differences, their methods of assessment and, if reported, the authors' justifications for those decisions. The study revealed three broad patterns. First, relatively few user differences are gathered in a typical study; age, sex, education level, and some measure of domain expertise are most common. Second, aside from three user characteristics (age, sex, education level), there is little consistency regarding which other differences are assessed, or how they are assessed. For instance, some authors use broad definitions of expertise (e.g., map expertise) and assess it in general terms (e.g., asking "indicate how often you use maps"), whereas others use narrow definitions (e.g., meteorological expertise) and more precise assessments (e.g., years of training or professional work experience in meteorology). Third, there are substantial ranges of quality and detail among the reported information; simply put, some authors do a measurably better job of reporting their research than others. Few authors report having used valid or reliable assessment methods to measure or establish group differences, suggesting that some researchers rely on ad hoc or non-validated methods even if better options are available.

Although my review did reveal which assessment practices are more frequent (and perhaps more preferred), there were few outright statements or clear evidence of which were demonstrably better or best. Given the diverse nature of study topics, the appropriateness of any particular user difference or assessment method is context-dependent, and will vary from study to study. While the importance of participant assessments is not in dispute, they merit more allotted time and attention to detail at all stages of a user study than they currently receive. An ongoing challenge, then, is providing cartographers with empirically-based best practices to guide their selection, implementation, and reporting of user differences and participant assessments.

4.2.3 CartoBase

The next phase of my research was to organize the published details of 110 user studies into a searchable database (White 2017b). I developed a prototype database, CartoBase (short for cartographic user study database), for four overlapping purposes (osf.io/ur6xn). First, CartoBase aggregates and organizes user difference and participant assessment details into a searchable archive; at present, there is no such publicly available archive. Second, CartoBase is both a communication and exploration tool that expedites the processes of 1) locating studies or topics of interest, 2) comparing the practices of two or more studies, and 3) grouping authors or studies based on who did what research. Third, CartoBase demonstrates, and is intended to promote, the value of open access and open research initiatives to the cartographic research community. Scientific research and publication culture have evolved over the past two decades to encourage (and often demand) the removal of barriers preventing data transparency, sharing, and accessibility; the future growth and success of CartoBase are predicated on cartographers collaborating to ensure our research—data *and* literature—is digitally open and accessible to all.

Fourth, and leading directly to this article, CartoBase enables a variety of measurable comparisons between its collection of existing studies, specifically those methods used to recruit, vet, and assess human subjects prior to their involvement in their respective studies. The purpose of such comparisons is to help in the identification of best participant assessment practices.

4.3 The Relevance of User Study Contexts

Most definitions of context (e.g., Abowd et al. 1999; Greenberg 2001; Reichenbacher 2003; Schlichtmann, H. 2009; Griffin et al. 2017) describe it as the circumstances or environment in which a situation or activity occurs. Contextual information gives us a frame of reference to make better decisions in a given situation, and we understand that context changes depending on the situation. For example, subway maps and in-car GPS devices are both navigational aids, but their intended contexts of use are completely different. A mapmaker must determine which contextual information is relevant for designing the best possible map for either situation. Subway commuters of all ages and literacies need to be able to determine their location and the route to their destination as quickly and correctly as possible; for this reason, subway maps typically adhere to a near-universal, geometrical, color-coded design that is recognizable. In-car GPS devices are intended for use in any location, in all traffic and weather conditions, and at any time of day, meaning the map and the information it displays—including the position of the car, the distance to the next turn, traffic congestion—have to be as precise, up-to-date, and, perhaps most importantly, as non-distracting as possible.

Context is so influential to cartographic research and design that it was the focus of a recent research agenda of the International Cartographic Association Commissions on "Use, User, and Usability Issues" and "Cognitive Issues in Geographic Visualization" (Griffin et al. 2017).

Cartographers rely on contextual information to design better maps and perform better research. In their research agenda, Griffin et al. proposed a model that could be used to systematically evaluate the relevant contextual information of two distinct map use situations, and determine if a particular map design is transferable across those contexts. I have integrated their concepts of *transferability* and *relevance* into my proposed method for identifying best practices for user studies.

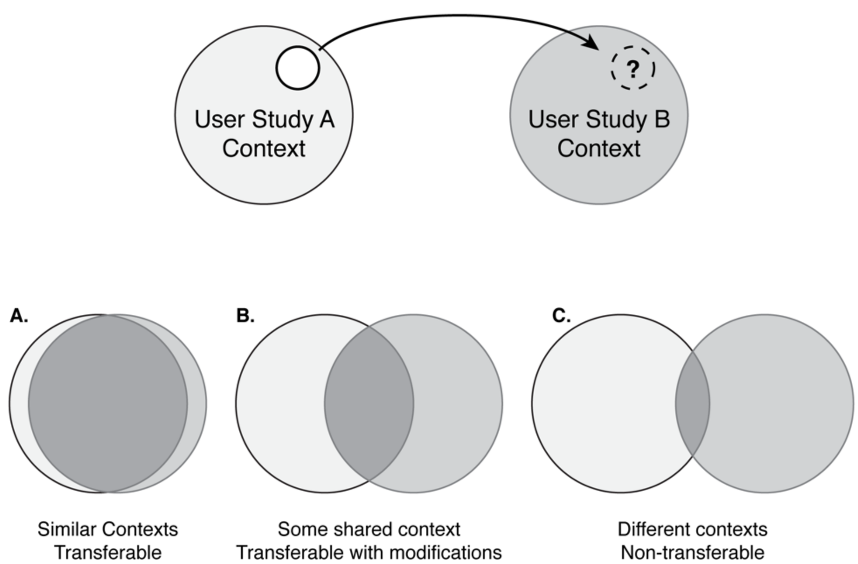


Figure 7. Illustrating the concept of transferability. The potential for a participant assessment practice to be transferable from its original study to a new study depends on the degree of similarity between the study contexts. For example, an assessment of expertise designed exclusively for an online experiment using participants sourced from Amazon Mechanical Turk (User Study A) may not be transferable to a full-day field mapping exercise involving graduate and postgraduate participants (User Study B). However, a comparison of their shared contextual information will indicate if the two studies are similar enough that the assessment of expertise will work in the field study. If the two contexts are similar (A), then the assessment is transferable. If the studies share some context (B), then the assessment may be transferable with minor or major modifications. If the contexts are different (C), then the assessment is non-transferable.

Originating from qualitative research, *transferability* refers to the extent to which research results can be transferred situationally to other contexts rather than generalized to all contexts (Lincoln and Guba, 1985). To determine if a practice is transferable from one situation to another, you should first identify the contextual factors that define each situation. Next, you should establish which of those contextual factors are the most relevant to each situation. Then you should determine how many relevant contextual factors they have in common. Based on the degree of similarity between those contexts, a practice is either deemed transferable (the situations are alike), transferable with modifications (the situations are similar but not identical), or non-transferable (the situations are not alike). Figure 7 illustrates the general principle.

I argue that what constitutes a "best" practice is distinctly context-dependent. Identifying and implementing any user study best practice necessitates an understanding of the context under which a particular study is being conducted. Thus, research methods are not one-size-fits-all, and neither are rules and guidelines cartographers have devised to guide the design of user studies; it is unlikely that cartographers will ever create a single research design that is suitable to all user study contexts. However, the contextual information of a user study will provide direct clues as to which study techniques or assessment methods are appropriate for that particular study. For these reasons, context and transferability are key components of my proposed method to identify best practices for any given user study context.

4.4 Best Practices

4.4.1 Best Practices Overview

The concept of "best practices" is common across many research domains, such as medicine (Cresswell et al. 2011), education (Fram and Camp 1995), journalism (Society of Professional Journalists 2014), social science research (Osborne 2008), business management (Camp 1998), and software development (Wilson et al. 2014). But what, specifically, defines a best practice?

There are nearly as many definitions of "best practice" as there are domains in which it is applied. Presented here are four examples:

- "Any practice, knowledge, know-how, or experience that has proven to be valuable or effective within one organization that may have applicability to other organizations." (O'Dell & Grayson 1998, 167)
- "Those practices that have been shown to produce superior results; selected by a systematic process; and judged as exemplary, good, or successfully demonstrated." (American Productivity and Quality Center, 1999)
- "A "best practice" is a results-oriented process or practice that enables an organization to accomplish performance results that cause it to lead its industry in achievement of goals. It operates in a frame of reference, and should be considered primarily in relationship to the environment in which it is found." (Saunders 1999, 20)
- "Best when compared to any alternative course of action and that it is a practice designed to achieve some deliberative end." (Bretschneider et al. 2005, 307)

These definitions—and many others (e.g., Osborne 2008; Klein et al. 2016)—emphasize a few key concepts. First, a practice is any action that produces a result. Second, best practices will enable superior performance or outcomes. Third, best practices must be identified using some formalized comparative or evidence-based process. Fourth, "best" is a moving target, and changes over time or according to context; what may be considered a best practice in a given scenario may not transfer to another scenario, or may be superseded by a better emerging practice.

4.4.2 The Need for Best Practices

So why pursue best practices, or develop a step-by-step procedure for identifying them? I suggest the following three related issues underlie why the search for and implementation of best practices in cartography is necessary, and would be beneficial to our research community.

- **The variety of cartographic research methods in use**

  Cartographers are employing a variety of methods to investigate many different questions, map use behaviors, and mapping technologies, many of which have been adapted from other domains. The diversity of methods reflects the nature of the issues being researched, ranging from geospatial big data (Robinson et al. 2017), interactivity (Roth et al. 2017), and geovisualization (Çöltekin et al. 2017), to the role of user differences (Elzakker and Wealands 2007; Dykes, Lloyd, and Radburn 2007). As the cartographer's methodological toolbox expands, so too, must our documentation that these methods are valid and reliable to the questions under study.

- **Evidence of inconsistent research methods & need for guidelines**

  As revealed in White (2017a), existing user study practices are inconsistent in the manner in which they are designed, implemented, and reported on. Cartographers employ a widening variety of methods to measure and report on user differences, not all of which are valid or reliable. At present, there is a general lack of guidelines to assist researchers as they design and implement user studies (Roth et al. 2017), including participant assessments. Without formal guidance, how can researchers, particularly those new to or unfamiliar with cartography, expect to distinguish between optimal and sub-optimal practices? Establishing formalized best practice guidelines will help cartographers answer this question and select appropriate study methods.

- **Best practices will improve user studies**

  Promoting and adhering to best practices will improve the overall quality and consistency of cartographic research. As such, the cartographic community should prioritize the development of a practical, systematic method to identify best practices. Fortunately, several things work to our advantage in this respect:

1. The size of the cartographic community. Cartography is a relatively small discipline with a limited academic and professional footprint (compared to, say, GIS), which fosters greater communication and collaboration among colleagues, and eases our ability to stay atop emerging trends and findings.
2. The amount of published cartographic research. While the body of cartographic research is substantial, it is also relatively manageable compared to, say, the broader domain of geographic information science. Any effort to account for the bulk of published cartographic user studies (and the bulk of existing methods) is a realistic, attainable goal.
3. Current online tools such as GitHub (github.com) and the Open Science Framework (osf.io) are available to help us organize existing research and existing practices, which can be used to ease the comparative process underlying best practice identification.

4.4.3 Criteria for a Best Practice

The literature points out that the term *best practice* is something of a misnomer, as it is unlikely there is a singular *best* way to measure or capture a particular map use phenomenon or experimental variable (Klein et al. 2016). The simplicity of the concept belies the complexities of identifying and implementing an undisputed *best* practice for a given scenario, and obscures the likelihood that more than one appropriate course of action may exist. For this reason, it is important to organize practices according to their perceived or measurable worth. Most practitioners use some variation of a good-better-best scale (Figure 8).

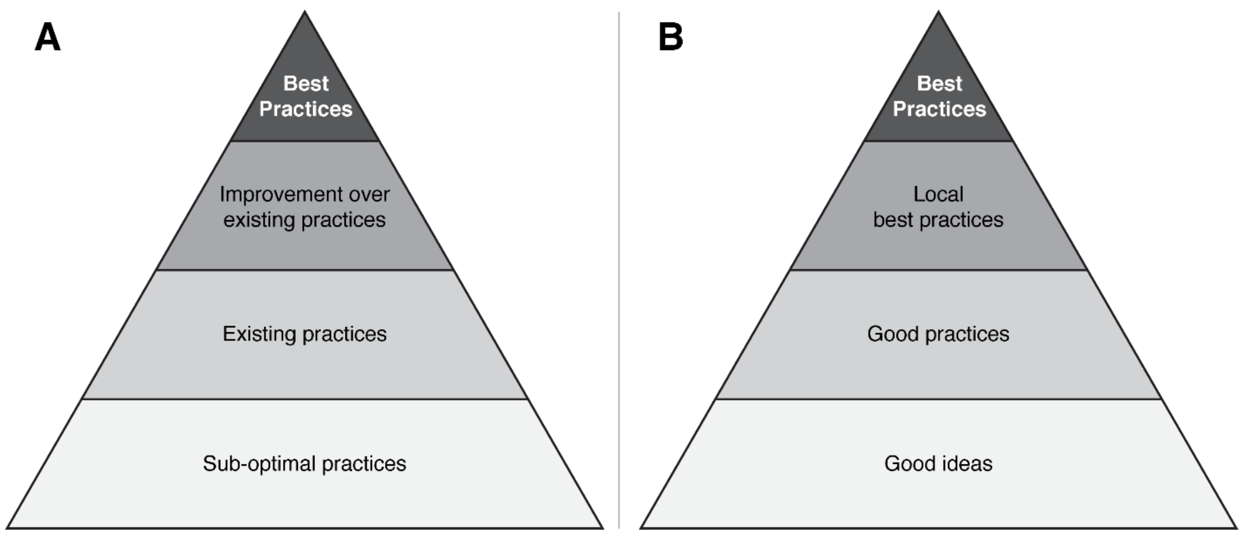


Figure 8. Two common best practices rankings. A) organizes all available practices into four tiers, placing *sub-optimal practices* at the bottom, followed by *existing practices*, *improvements over existing practices*, and *best practices*. B) organizes best practices into four levels of increasing quality and applicability: once proven, a *good idea* becomes a *good practice*, which can become a *local best practice* if it is the best approach for a specific set of circumstances; at the top are *best practices* that are applicable to all circumstances.

What actually *constitutes* a best practice depends on its context of use. For example, what a software developer considers a best practice may not translate to cartography; and what a critical cartographer considers a best practice may not be embraced by those taking a conventional, positivist approach. In addition to being moving targets, best practices also evolve: emerging techniques, technologies, and methodologies constantly reshape and influence which practices are optimal in a given scenario. Nevertheless, the literature suggests an overlapping set of generalized, essential criteria that apply to every best practice (see, for example, Camp 1995; Keehley 1997; Klein et al. 2016): that it shows measurable success, has a proven track record, is replicable with modifications, is transferable, and is recognized as innovative and important to the organization studying the practice. Taken together, a best practice is measurably better than the alternatives, is adaptable and transferable, and its potential users will recognize its value and want to adopt it in their existing work or research processes.

Confounding the establishment of a single set of best practice criteria for all user studies is the variety of contexts that shape each user study. While those generalized criteria listed above are helpful guideposts, they are not particularly informative when a researcher is looking for a single best practice to be used in a specific user study context. In these situations, identifying a suitable best practice is contingent on knowing the relevant contextual factors shaping a user study *and* the outcome the practice is expected to produce. For example, imagine a researcher is designing a repeated-measures navigational exercise to study the map reading abilities of elementary-aged children, and the exercise will take place on their school playground. If the researcher plans to assess their map literacy prior to the exercise, the study context will influence the choice of assessment: the study setting (outside, school playground) precludes a computer test, the age of the participants (children) means the test must be written to their age and development level, and the type of exercise (navigational) suggests that the test should focus on wayfinding rather than thematic map literacy. These study details provide enough contextual information—about the information the researcher expects to assess as well as the conditions under which the assessment will be performed—to establish a set of criteria that can determine if a given practice is potentially the "best" practice for that study. Contextual information is key to best practices in user studies. Therefore, the more specific the context is, the more likely a desired outcome can be achieved.

4.4.4 Conditions for Identifying Best Practices

While the criteria discussed in the previous section can indicate what *constitutes* a best practice in a given context, the criteria do not tell us explicitly *how* to identify one. Researchers cannot use ad hoc methods or rely on human judgment alone to identify best practices. Bretschneider et al. (2005, 307) suggest that two "necessary and sufficient" conditions must be met in order to appropriately identify the best practices for a particular context: *comparability of cases* and *completeness of cases*.

Comparability of cases (or the "comparability condition") refers to the ability to compare specific cases or practices based on their shared characteristics and the outcomes they produce. The more alike two practices are, the more possible it is to make a sufficient comparison between the two and determine which one is better or produces the more desirable result. For example, if a researcher wants to improve the language of their demographic questionnaire, they should compare it to other demographic questionnaires and not, say, spatial ability tests.

Completeness of cases (or the "completeness condition") refers to having a sufficient set of cases, studies, or practices to compare. Bretschneider et al. (2005) stipulate that sufficient comparability requires the inclusion of *all* comparable practices, arguing that the smaller the group under comparison becomes, the greater the likelihood that a better practice falls outside that group. For example, if a researcher wants to find the best demographic questionnaire available, they should compare as many questionnaires as possible and not limit their comparison to just two or three examples. The authors do acknowledge the impracticality of achieving true "completeness," and leave it up to the researcher to determine when a group of practices is sufficiently large enough to enable sufficient comparability.

According to Bretschneider et al. (2005), a limitation of many identification efforts is that the selected processes or practices are not sufficiently comparable, and/or the number of cases are insufficient to identify a "true" best practice. The comparability and completeness conditions address this by setting appropriately rigorous thresholds by which to make comparisons. In the next section I survey the procedure of "benchmarking for best practices" and propose a variation that utilizes CartoBase as a way to satisfy both the comparability and completeness conditions and help researchers identify best practices.

4.5 Benchmarking for Best Practices

4.5.1 Defining Benchmarking and Early Developments in Benchmarking

An early definition of benchmarking described it as a formalized, evidence-based, comparative process designed to evaluate "the products, services, and work processes of organizations that are recognized as representing best practices for the purpose of … improvement" (Spendolini 1992, 9). Regardless of who is undertaking the benchmarking—an organization, academic department, research group, or individual—the process involves searching for best practices, adapting them for implementation in one's own organization, process, or workflow, then monitoring the implementation of that practice over time to ensure its success (Keehley and Abercrombie, 2008). Benchmarking comparisons are made by setting a benchmark—a high standard or point of reference against which others are compared—and using a combination of expert opinion and quantitative or qualitative measurement to determine which specific practice can exceed that benchmark and produce a superior outcome (Camp, 1995, 24).

Much of the current benchmarking literature is based on the work of Robert Camp (1989, 1995), who popularized the process to help corporations identify and implement best practices within their own industries in order to gain competitive advantages. According to Camp (1995, 14-15), a benchmarking study serves four main purposes: (1) document strengths and weaknesses within one's own performance or work processes; (2) identify and document performance of industry leaders (referred to as "the competition"); (3) incorporate best practices of industry leaders; and (4) adapt and innovate those practices to become new industry leaders (i.e., the new benchmark). Benchmarking can look *internally* or *externally* to identify best practices. *Internal benchmarking* compares similar operations within one's own organization or domain to identify best practices (e.g., comparing different participant assessment methods of two cartographic researchers). *External benchmarking* looks outside one's domain to identify comparative partners and transferable best practices; for example, cartographers often look to psychology and cognitive science for effective experimental methods when conducting novel research (Andrienko et al. 2007; Çöltekin et al. 2016). Internal studies are typically easier to conduct than external studies, but both use similar processes to achieve the same goals.

Traditional benchmarking is a multi-step, team-based process requiring substantial financial resources, time (often six to twelve months), and organizational support. Most studies progress through four major phases:

1. **Planning the Study**. The objective of this phase is to determine the what, who, and, in some models, the how of the benchmarking study—what is to be benchmarked, who is to

be benchmarked against, and how the data will be collected. Best practice criteria—or the criteria for the desired solution—and key measurement variables are selected during this phase.

2. **Data Collection**. Performance data are collected during this phase.
3. **Data Analysis**. Best practices are identified during this phase by comparing performance data and determining which actions produce superior results.
4. **Adaptation for Improvement**. There are two objectives of this phase: 1) implement the identified best practice, and 2) monitor its implementation over time to ensure the adopted practice is improving performance as expected.

Regardless of their differences, three early steps in every model—1) determining what to benchmark, 2) what to benchmark against, and 3) gathering the relevant data—are critical to the overall success of the procedure (Camp 1995; Longbottom 2000; Anand and Kodali 2008). Expert input is critical across these steps to ensure the study collects and measures the appropriate data, and identifies a truly "best" practice. To remain effective over time, an adopted practice must be continually monitored to ensure it actually produces better results, and the benchmarkers must recalibrate or start anew if it does not. A benchmarking organization, department, or research group cannot become complacent, or assume that a "best" practice they have identified will always remain so.

4.5.2 The Migration of Benchmarking to other Domains and Fields

Over the past few decades, a number of benchmarking experts have adapted Camp's ideas across many academic and professional fields (for example, see Keehley 1997; Longbottom 2000;

Bretschneider et al. 2005; and Anand and Kodali 2008). A key advantage of academic and certain professional environments (as opposed to the private sector) is that best practices are more openly shared since the outcome is no longer tied solely to seeking some advantage over the competition. Because my proposed benchmarking model emerged from this non-competitive, academic environment, I will briefly review some applications of benchmarking for best practices that produced results similar to what I hope to achieve with cartographic user study design.

Early on, benchmarking experts (e.g., Spendolini 1992; Bogan and English 1994) recognized how the benchmarking process could complement existing methods of academic development, which often entail learning from academic journals and collaborating within a professional or peer network in an area of specialization. For example, by the end of the 1990s a number of higher education institutions had popularized benchmarking or similarly-styled studies as the preferred means for identifying best practices in areas such as admissions and student recruitment, student achievement, and administrative efficiency (Alstete 1995; Camp 1998; Saunders 1999). During the 2000s, and hastened by the predominance of the Internet, the adoption of best practices similarly pervaded a wide variety of professional and academic fields. As part of this broad adoption trend, the criteria that define a best practice have also become similarly varied. For example, the Society of Professional Journalists Code of Ethics (2014) prioritizes the role of ethics in journalistic best practices. According to the organization's website, these practices promote "the highest professional standards for journalists of all disciplines."

Benchmarking for best practices has been used extensively in fields related to cartography, such as geographic information science (Donert 2009; Schulze et al. 2013), data visualization (Ward & Theroux 1997; Ward et al. 2015; Bhowmick et al. 2017) and information and graphic design (Plaisant et al. 2008; Meirelles 2013). However, benchmarking efforts are notably absent from the cartographic literature, which suggests that the benchmarking process has not yet migrated to cartography, or perhaps that a method suitable to the needs of cartographers has not yet been developed. Also notable is how the concept of "best practices" is rarely discussed or defined in the cartographic literature, despite the fact most of our publications are concerned with the best or most appropriate ways of doing things; that is, cartographers actively pursue and promote best practices without using the term itself.

4.5.3 Solution-Driven Benchmarking

The widespread adoption of benchmarking for best practices has coincided with the expanding, centralizing role of the Internet. Data creation, storage, sharing, and searches now almost exclusively occur online. Any potential benchmarker can now execute a search for best practices if they know where to look, and how to appropriately identify them. Generally speaking, the Internet helps to circumvent many of the restrictive requirements of traditional benchmarking. Where a traditional search demands deliberately slow, narrow, resource-intensive data gathering and analysis from a large team, Internet searches are faster, broader, cheaper, and can potentially be conducted by an individual. One promising solution, known as *Solution-Driven Benchmarking* (SDB) (Keehley and Abercrombie 2008), has recently emerged to bypass traditional strictures and exploit the search potential of the Internet.

The SDB methodology was developed by those lacking the staff and financial resources required to perform a traditional benchmarking study, such as nonprofits and nongovernmental organizations. Although the solution-driven approach (Table 14) progresses through the same major phases as the traditional approach discussed previously, the result is a more streamlined and deductive process dependent upon just one or two people. The solution-driven methodology is purposely limited in scope, focusing on identifying solutions to specific problems rather than long or complex processes. Therefore, the search for best practices shifts from improving a large, complex process, such as the entire user study design process, and instead focuses on a specific component within that process that needs to be adjusted or improved.

Table 14. Solution-Driven Benchmarking (Keehley & Abercrombie 2008)

Phase	*Steps*	*Description*
Planning the study	1. Discover the problem	A problem must be recognized and defined by the researcher who initiates the search for promising practices.
	2. Establish criteria for solutions	Measures or other criteria are used to establish the goal or desired impact of the promising practice.
Collecting & Analyze Data	3. Search for promising practices	Search for promising practices by accessing professional networks and the Internet.
Adapting for improvement	4. Implement promising practices	Implement the best or promising practice that you have identified and are importing.
	5. Monitor progress	Continually monitor the practice that has been implemented to ensure it has a positive impact.

What makes SDB so promising for cartographic researchers is also what sets it apart from other benchmarking models. Unlike traditional benchmarking, SDB for best practices typically occurs in a non-competitive, non-corporate, online environment where information (problems, practices, and solutions) is more openly accessible and shared. Open access to information is the key component. As Keehley and Abercrombie (2008, 86) explain, a successful search for a best practice is dependent on the Internet *and* a "network of professionals and organizations that the benchmarker can tap for promising practices." Combining these resources with the narrow focus of the SDB process expedites the best practices search, especially if the benchmarker has access to a comprehensive database of domain-specific information such as CartoBase. In the following section I present and review a modified SDB process to help cartographic researchers identify and implement best practices as they design, execute, and report on cartographic user studies. For the remainder of this article, I will refer to this model as CartoSDB (Cartographic Solution-Driven Benchmarking).

4.6 A Benchmarking Model for Cartographic User Study Design

A key purpose of this article is to ascertain if a benchmarking model can be developed that could improve the overall quality and consistency of cartographic user studies. CartoSDB is intended to assist researchers in selecting appropriate practices in the context of their specific study, thereby reducing the perpetuation of ad hoc and/or unreliable design decisions. Whereas traditional benchmarking relies on measurements of performance, CartoSDB makes comparisons based on the qualitative characteristics of the desired practice and the contextual information of the study in question. The model format is a variation of the solution-driven method presented above, and synthesizes two critical components presented in Section Two: CartoBase and contextual information about a study. These two components help satisfy the completeness and comparability conditions necessary to identify best practices, as I will demonstrate below. This section proceeds with a description of the overall CartoSDB model, followed by a detailed description of each step in the model, including the action the benchmarker must perform and the anticipated outcome. I have integrated a case study into each step to demonstrate how the model functions.

4.6.1 Description of the model

The CartoSDB model is divided into six sequential steps (Table 15), and the entire process closely follows the five-step SDB model. Three key differences are reflected in the second, third, and fourth steps, which integrate *context* and *transferability* to establish the parameters for the search and comparison of best practices. As with the SDB method, CartoSDB is purposely streamlined to ensure that its user can successfully benchmark a best practice in a short amount of time, and with limited knowledge of benchmarking in general. When using CartoSDB, the desired outcome is itself a benchmark—that is, a point of reference against which each potential best practice is compared. The benchmarker needs to find evidence that links a promising practice with an improvement in the study design, then ensure that the evidence matches the criteria used to establish a successful resolution.

Table 15. Cartographic User Study Solution-Driven Benchmarking.

Steps	**Description**
1. Define the study problem	Define the component or problem in a user study for which a best practice is sought.
2. Establish criteria for the Best Practice	Qualitative or quantitative criteria are used to establish desired qualities and impact of the best practice.
3. Define the study context	Define the major contextual information of the user study and conditions under which the problem exists.
4. Search CartoBase for promising practices	Search CartoBase using best practice criteria and contextual information to identify one or more potential best practices.
5. Implement the promising practice	Implement the best or most promising practice that you have identified in CartoBase.
6. Report outcome of the selected practice	Report on the positive or negative impact of the practice so other researchers know to use or avoid it in future user studies

To illustrate the six-step CartoSDB process, I present a case study (referred to as Case Study below) using the first of two user studies reported in a 2016 article published in the *Annals* *of the American Association of Geographers* by Toru Ishikawa (2016)[3]. To provide some background, Ishikawa provided this description of the user study (p. 78):

> "The first study examines the conceptions of geospatial concepts by experts and novice students in GIScience. In this study, authors of GIScience textbooks and students in an introductory GIScience class are considered experts and novices, respectively, and the structures in which geospatial terms are explained in textbooks and conceptualized in students' minds are compared...Considering the strong correlation among spatial tests in the dimension of spatial visualization and possible effects of spatial visualization ability on geospatial thinking suggested in the literature, this research uses a major spatial test in the dimension, the Card Rotations Test, to assess students' spatial visualization ability."

For demonstration purposes, imagine that we are replicating Ishikawa's study, only this time we want to include an assessment of GIS proficiency for the student participants. Our intent is to identify the "best" assessment method; by establishing the criteria of the desired assessment and the relevant contextual information of the study, we can search CartoBase for one or more potential best practices, and add them to the study.

---

[3] Ishikawa (2016) was selected for the simulated case study simply because it has not yet been added to CartoBase; its use here is in no way suggestive that the study is flawed or problematic.

Step 1. Define the user study problem

To initiate a search for best practices, a researcher must first define exactly which component of their user study needs to be improved, or what problem needs to be solved. It is essential from the outset to concentrate on a singular issue. Defining a problem too broadly (e.g., I need a better way to assess *all* user differences) may impede successive steps or result in an ineffective search. The problem of the Case Study is straightforward: we need to identify an assessment of GIS proficiency and add it to the study procedure. Step 1 does not require access to any external resources, and the benchmarker should be able to complete it on their own. Upon completion of Step 1, the benchmarker should be able to state exactly how the desired best practice fits into the user study.

Step 2. Establish criteria for the Best Practice

After the researcher has defined the study problem, he or she will establish criteria which will be used in Step 4 to help determine the suitability of an identified promising practice. These criteria may include what the practice is expected to assess, information about the design or implementation of the practice, and possibly quantitative and/or qualitative characteristics of the practice. The benchmarker will judge a potential best practice based on its ability to meet these criteria.

For our Case Study, we will use one criterion for the desired assessment method: the method must contain some assessment of GIS proficiency, but need not be limited to GIS only (map and cartographic literacy tests could feasibly contain a GIS component). This criterion was selected based on my reading of the article itself, and is sufficient to demonstrate the capabilities of the CartoSDB model as well as what a researcher might do in a real-world application, although more criteria could easily be added. Note that currently there are no established guidelines, standards, or measures to rely on when establishing the criteria for a best practice. Upon completion of Step 2, the benchmarker will have identified the criteria that will act as search filters in Step 4.

Step 3. Define the user study context

Defining the relevant contextual factors that shape a user study or influence the conditions under which a problem exists is necessary to ensure that all potential best practices are suitably comparable (see the *comparability condition* introduced in Section 4.4) *and* potentially transferable (Step 4 below). Context plays key roles in the design and reporting of user studies as well as the concept of best practices. All best practices are relative—what may be best in one research context may be worse in another. The challenge for the benchmarker is determining which contextual information is most relevant to the study at hand. Note this step does not appear in the original SDB model; unlike most benchmarking models, CartoSDB prioritizes *context* as a key comparability measure.

In addition to the Case Study details provided in the model description, the following contextual information was gathered elsewhere in Ishikawa's article: Forty-four graduate students participated; the study was designed as a between-groups experiment; and the study occurred in the classroom, was paper-based, and took approximately thirty minutes to complete. Which contextual information indicates if an existing assessment method is transferable to the Case Study? Absent any formal guidance, and using our best judgment, we will presume there are three "relevant" contextual factors: (1) the participants are college-aged, (2) the entire study took approximately thirty minutes (short enough to occur in a single class meeting), and (3) the study is comparing novices to experts. As with the best practice criteria in Step 2, this contextual information is sufficient to demonstrate the capabilities of CartoBase, although other factors could be deemed more relevant. The benchmarker can use this contextual information to filter out inappropriate practices from their search. Step 3 does not require access to any external resources, and the benchmarker should be able to complete it on their own. Upon completion of Step 3, the benchmarker will have identified the key contextual factors that will act as search filters in Step 4.

Step 4. Search CartoBase for potential best practices

This step in the SDB model is what distinguishes it from traditional benchmarking methods; according to Keehley and Abercrombie (2008, 88), the key to success is online access to information shared by a domain-specific "network of professionals and organizations that the benchmarker can tap for promising practices." What sets CartoSDB apart from the SDB model are the central roles of CartoBase and the *transferability* process. Rather than having to look at all published research to find suitable user study methods, CartoSDB expedites the search by utilizing CartoBase. Through CartoBase, the benchmarker is able to satisfy the two conditions necessary to identify best practices. First, CartoBase is comprehensive[4], which satisfies the *completeness condition* stipulating that sufficient comparability requires the inclusion of all comparable studies. Because CartoBase aggregates published user study information in one online location, there is a strong likelihood that a best practice exists within the database. Second, CartoBase helps satisfy

[4] The current iteration of CartoBase is not truly comprehensive, meaning the completeness condition is not sufficiently met in the strictest sense. However, the database is sufficiently large to demonstrate the utility and viability of both CartoBase and the CartoSDB method.

the *comparability condition* by facilitating swift and focused searches for best practices using the criteria and relevant contextual information established in Steps 2 and 3.

The transferability process is unique to benchmarking because it ignores traditional quantitative targets that are not directly applicable to the user study environment like "the best practice must improve productivity by 15%." Instead the transferability process compares the contextual information of one user study which contains a potential best practice to the contextual information of the user study the best practice is intended for. This occurs when the benchmarker assigns the filters in CartoBase, since the database will remove any studies that do not "match" the relevant context.

Step 4 of has four basic stages:

1. Identify all potential best practices, which is achieved by filtering CartoBase using the established criteria from Step 2.
2. Identify all comparable user studies, which is achieved by filtering CartoBase using the relevant contextual information of the Case Study from Step 3.
3. Assuming one or more potential best practices remain after applying both sets of filters (criteria and context), examine each in detail to identify which one has the "best" practice. If no studies remain, you may need to reduce or adjust the filters.
4. Assuming a transferable practice has been identified, determine which modifications must be made to the practice in order to implement it into your own study context, such as adjusting the language or making it digital.

Applying these four stages to our Case Study, Table 16 lists every study that meets each of the individual and combined Case Study filters. The table displays four rows: 1) all studies in CartoBase, 2) all studies that match the best practice criteria, 3) all studies that match the relevant contextual information, and 4) all of the studies/practices that match both the criteria and context. Other criteria and contextual information could be added (or removed) to produce a smaller (or larger) group of potential best practices.

Table 16. Subset of CartoBase studies matching the Case Study filters.

Filter	*n*	*%*	*Study subset*
All CartoBase studies	110	100%	
Best Practice Criteria (assesses GIS proficiency)	16	15%	
Relevant Contextual Information (college-aged; approx. half hour length; novice vs. experts)	13	12%	
Best Practice Criteria + Relevant Contextual Information	2	2%	Hope and Hunter 2007; Opach and Rød 2014

As Table 16 shows, CartoBase reveals that just two assessments of GIS proficiency (from Hope and Hunter 2007 and Opach et al. 2014) meet both the best practice criteria *and* contextual filters. Because CartoBase does not contain the actual assessments themselves, we must inspect the articles' text to determine if either assessment method is suitable for the Case Study. Both studies used a single, self-reported question to assess their participants' GIS proficiency. Participants from Hope and Hunter (2007, 202) "subjectively classified themselves into one of three experience levels, which were described as: (1) novice (less than six months' experience using GIS); (2) some experience (between six months and two years of experience of using GIS); and (3) experienced (more than two years' experience of using GIS)." Opach and Rød (2014, 420) had their participants rate their level of training from 1 (no training) to 5 (proficient). Note that neither study actually assessed GIS competency, but rather used this self-reported information to assign participants into groups. Supposing we prefer the Hope and Hunter method for its use of a concrete timeframe to determine each proficiency level, then that is the method we will implement in the Case Study.

Keep in mind that CartoSDB does not necessarily identify the *best* practices, but rather the best *available* practices according to the defined best practice criteria and contextual filters. If neither of the two identified assessment methods are determined to be suitable for the Case Study, then one should adjust the filters and run the process again. Upon completion of Step 4, the benchmarker will have identified all promising practices and selected one for implementation.

Step 5. Implement the promising practice

Implementing an identified promising practice is perhaps the most straightforward step in the CartoSDB procedure. The benchmarker will take the practice from Step 4 and ensure that it appropriately fits into the study design, including making any necessary modifications to ensure the practice conforms to the requisite study format or language. Because the selected assessment method uses a single measurement (GIS proficiency levels are based on time spent using GIS), it should be very easy to fit into the current Case Study procedure. Ideally, CartoSDB will occur before the user study is in progress, allowing the researcher to review all components of the study prior to deployment, including the selected practice. Upon completion of Step 5, the benchmarker will have the promising practice embedded in the user study.

Step 6. Report the outcome of the selected practice

The final step requires the benchmarker to report on the implementation and outcome of the selected practice, and whether it had a positive or negative impact on the user study. Communicating this information to the rest of the cartographic research community is an essential step if we want researchers to make use of or avoid a particular practice in future studies. I suggest two ways for sharing this information: adding it to CartoBase, and reporting it in a refereed publication.

Study details can be added to CartoBase by accessing the database website[5] and entering the study details using the provided submission forms. The added information should include the result of having used the transferred practice (in this example, the modified assessment of map literacy), if other researchers are to know if the practice in question led to positive or negative results. The study authors should also include a statement in the published article which explains how the assessment method was selected, whether it produced the intended results, and if that method is recommended for other user studies. These two suggestions are rooted in a philosophy of continuous improvement, whereby a benchmark is recalibrated and a new search for *better* practices, or improvements to the implemented best practice, commences. Thus, when a study ends, the researcher is able to provide evidence or some measureable data which may serve as a point of comparison for future researchers. What this means for the Case Study is that we would

[5] The database has two online locations: osf.io/ur6xn/ contains the LibreOffice Base data files, and app.zohocreator.com/twhite10/cartobase is a temporary web-based version containing the same user study content.

include in the published article a statement explaining how the assessment method was selected, whether it produced the intended results, and if that method is recommended for other user studies.

4.7 Conclusion

Cartographers have prioritized a research agenda to develop tools and establish guidelines to support decision-making when designing user studies (Roth et al. 2017). CartoSDB, short for Cartographic user study Solution-Driven Benchmarking, is designed specifically for this purpose. The CartoSDB procedure expedites the process of searching and comparing the details of published user studies to identify the best available methods for any study context. Compared to traditional benchmarking, which follows a slow-paced, resource-intensive, complex procedure to identify best practices, solution-driven benchmarking is designed to be fast-paced, cheap, and relatively easy to perform. CartoSDB enables this by focusing on one individual study design problem at a time, and harnesses a purpose-built online database, CartoBase, to identify promising solutions.

CartoSDB follows six sequential steps: (1) define a specific user study design problem or challenge, (2) establish the criteria for the best practice, (3) define the relevant contextual information for the study, (4) search CartoBase for potential best practices, (5) implement the most promising practice, then (6) report on the outcome of the selected practice to the rest of the cartographic research community. This procedure was purposely designed for simplicity and broad applicability. The steps are straightforward and easy to execute, regardless of research experience or tacit knowledge; in this respect, the procedure is especially valuable to graduate students or researchers new to the discipline. While its use in my research is limited to identifying assessment methods for user differences, both CartoSDB and CartoBase can accommodate any type of cartographic research.

There are two limitations of CartoSDB that need to be addressed in future research. First, the proposed CartoSDB model is conceptual and has not yet been pilot-tested or validated by myself or other researchers. In order to demonstrate its utility and gain wider acceptance among the cartographic community, CartoSDB must be applied in a real-world setting, by a variety of study authors, and in a variety of user study contexts. This will require a large-scale, international effort, likely with the support of a professional or academic cartography organization. Second, until CartoBase is truly comprehensive, containing the bulk of published cartographic user studies, CartoSDB will only lead researchers to the best available practice within the limited sample of published studies it currently contains. For example, Olson and Brewer (1997) developed an experimental procedure that researchers could use to assess color vision impairment and map reading abilities simultaneously; however, CartoBase does not include user studies prior to 2004, and none of the studies currently in CartoBase referenced this work, so the CartoSDB model cannot suggest this method.

CartoSDB is *not* intended for designing an entire user study, nor can it guarantee a study's success. The model was designed to focus on one component of a user study at a time. Although traditional benchmarkers (e.g., Camp 1989; 1995) do not criticize solution-driven benchmarking by name, they do warn that chasing individual problems as they arise is akin to applying band-aids. This points to the significant difference between the two approaches: traditional benchmarking emphasizes analyzing a problem to get at the root cause, whereas solution-driven benchmarking emphasizes looking at what others have done to solve a similar problem. CartoSDB may help individual researchers with individual study design problems, but it makes little difference if CartoSDB identifies the best way to do something if the rest of the study is fundamentally flawed. To this end, the cartographic community as a whole needs to prioritize the ICA's research agenda of establishing design guidelines to help ensure that all components of a user study are selected and implemented correctly.

CHAPTER 5. CONCLUSION

5.1 Summary of Findings

From inconsistent user study designs and reporting methods, to organizing studies into a searchable archive, and systematizing how we search existing studies to find best practices, my dissertation research demonstrates some of the ways in which best practices, or the lack thereof, impact cartographic research. By exploring where study deficiencies exist, specifically in the selection, assessment, and reporting of user differences, Chapters Two, Three, and Four provide sufficient evidence that cartographers stand to benefit by making immediate improvements to the user study design process and have the means to do so. These chapters document some best practices identification tools and procedures already used in other fields of research, and demonstrate how the cartographic community can adopt them to achieve similar end goals. If adopted, my proposed strategies can contribute not only to improved user study designs, but also to improved levels of communication and collaboration among cartographers. Additionally, the findings of each chapter summarized here contribute to the long-term research agendas outlined by the International Cartographic Association (ICA), offer practical tools and solutions that can be scaled up to benefit *all* cartographic researchers, and build on theory and conceptual frameworks for future applications of cartographic user study design.

5.1.1 Chapter 2

The user study review performed in Chapter Two indicates that cartographic user studies are inconsistently designed and reported on, regardless of study topic or context. Two notable findings emerged from this review. First, the review provided sufficient evidence that some user differences, and methods for assessing those differences, are more preferred than others. However, frequency of use does not imply appropriateness of use, suggesting that cartographers need to clarify or make improvements to existing user study design procedures. This can be achieved by developing guidelines to determine which user differences should be assessed in a given study context, and then select appropriate assessment methods for those study contexts. Second, the findings indicate that many authors gather and report only those study details that are immediately relevant to their own study topic, and withhold information that may seem extraneous or nonessential to the study topic. While understandable, withholding any study data from a published article can diminish the transparency and quality of that research, and can prevent others cartographers from making use of those data or connections with their own or other published research. For these reasons, the cartographic community needs guidelines that improve the quality and consistency of study reporting, as well as an online database or repository that contains study data or information that do not make it into published articles.

5.1.2 Chapter 3

Chapter Three examines developments in open access research to underscore how fields related to cartography are handling their own research data and sharing it or making it discoverable within their communities. The cartographic community must keep pace with contemporary data collection and storage methodologies. To wit, I created CartoBase, the first online, searchable archive of published cartographic user studies. The database currently contains the participant assessment details of 110 user studies published in five refereed journals between 2005 and 2014. The database has been designed so that it can be expanded without changing its underlying data structure. As more studies and more study data categories are added, the audience for this database will also expand. And while it is likely that most researchers already have access to these published studies through their institutions' libraries, CartoBase gathers details of these studies in one location. An added advantage of CartoBase is that it archives user study details categorically, enabling efficient category-based search and query operations. If cartographers are to establish best practices guidelines and make improvements to their research designs, then they need to make existing study data more accessible and easier to peruse. Given how prevalent and vital online databases are to other fields of research, cartographers need to make a concerted effort to develop their own database. CartoBase is intended to fill this role.

5.1.3 Chapter 4

As mentioned throughout my dissertation, the contemporary cartographic community has prioritized a research agenda to develop tools and establish guidelines to support decision-making for designing user studies. CartoSDB, the best practices identification procedure presented in Chapter Four, is designed for this specific purpose. CartoSDB expedites the search for cartographic best practices by systematizing how searches are performed. The procedure was designed for ease-of-use and broad applicability. Utilizing six straightforward steps, a researcher identifies a specific user study design problem and then performs a focused search of CartoBase for the best available practice that can solve that problem.

One important issue addressed in Chapter Four concerns what CartoSDB is *not* intended to do, namely design entire user studies. The procedure helps identify best practices, but it cannot guarantee their success, nor the success of the entire study. For this reason, while CartoSDB is a valuable tool, it does not supplant the community's need for best practices guidelines. Thus, while I encourage the community to adopt my best practices identification method as a way to make immediate improvements to user study designs, I also encourage support for the ICA research agenda to establish design guidelines (Roth et al. 2017) that will help ensure that all components of a user study are selected and implemented correctly.

5.2 Summary of Research Contributions

There are many effective ways to design, perform, and report on participant assessments in cartographic user studies. Presuming that best practices do exist and are worth adhering to, cartographers should advocate for the "best" ways to perform actions for any component of their user studies. Importantly, the process of designing any user study component must be easy to follow, easy to justify, and easy to report. These characteristics are necessary if we hope to get the entire cartographic research community to actively pursue best practices in their own research. My primary contributions to the discipline are a practical, functioning tool and systematic procedure that organize our user studies in a central location, thus enabling cartographers to search among and identify best practices from those studies.

Chapter Two identified and established a general awareness of consistent flaws in user studies. Based on those recognized weaknesses, I believe there are two sets of best practices guidelines that need to be developed: one for the design of user studies and another for the reporting of user studies. My Chapter Two findings laid the groundwork for the next logical step in establishing a set of user study best practices: rather than allow inconsistent or sub-optimal methods to perpetuate, we should create a user study database that allows us to keep track of all existing and emerging study methods. Such a database would help monitor the use of each method, examine their use within their own study contexts, and compare methods with one another. In doing so, we can keep track of both completed user studies and the relative effectiveness or quality of individual study methods. In Chapter Three I presented CartoBase, the database I created to achieve these purposes.

CartoBase makes four significant contributions to cartography. First, CartoBase addresses an immediate communal need by aggregating and organizing existing cartographic user studies, which makes them more discoverable. Second, CartoBase can help facilitate decision-making during the planning phases of new studies, particularly for researchers investigating similar phenomena or employing similar methods. Third, by allowing collaborative editing, CartoBase encourages increased interaction among cartographic researchers in an open-access environment, by relying on the research community to contribute to, maintain, and enhance the database itself. Fourth, and perhaps most importantly, CartoBase provides a transparent foundation on which to establish best practices guidelines for the design of user studies by allowing everyone to view the specific research, methods, and outcomes on which those guidelines will be established.

Although CartoBase aggregates user studies and encourages researchers to examine existing study designs, the database alone cannot lead someone to a best practice. Therefore, the development of CartoSDB is the other major contribution of my research. CartoSDB takes an abstract and overgeneralized concept, the "best practice," and follows a methodical procedure to identify actual, applicable best study practices that can be compared and selected for implementation. The processes of comparison and selection are based on specific requirements of the researcher, and the context of their research. This methodical procedure has two uses: first, the cartographic community can adopt the procedure to establish a formal set of user study best practices; second, until those guidelines have been established, individual researchers can use CartoSDB situationally to find best practices for their own user studies.

5.3 Recommendations for Future Research

I recommend five principal challenges and opportunities for future research.

Recommendation 1. Establish use and growth strategies for CartoBase

My first recommendation is to develop strategies that raise the visibility of CartoBase within the cartographic community, encourage its use, contribution by others, and addresses long-term issues associated with hosting and maintaining an online database. Currently, CartoBase has limited utility to the cartographic community because it contains a relatively small sample of cartographic user studies and focuses solely on the topics of user differences and participant assessments. The ICA Commissions on Use, Users, and Usability and Cognitive Visualization have expressed interest in supporting this recommendation, and plan to make it an agenda item for the Spring 2018 meeting of these commissions.

Recommendation 2. Validate the CartoSDB procedure

The CartoSDB model is conceptual and has not yet been pilot-tested or validated by myself or other researchers. In order to more completely demonstrate its utility and gain wider acceptance among the cartographic community, CartoSDB must be applied in a real-world setting, by a variety of study authors, and in a variety of user study contexts. This will require a large-scale, international effort, likely with the support of a professional or academic cartography organization.

Recommendation 3. Promote Open Access research

Cartographers need to embrace open access research and increase the magnitude of overall data discovery and sharing. CartoBase is a functioning example of how a collaborative, open-access database can help the cartographic community aggregate and organize existing research, and use that information to improve the design of emerging cartographic research. Encouraging cartographers to share their data and research designs in an OA environment such as CartoBase is a logical—and valuable—step toward improving the overall quality and consistency of cartographic research.

Recommendation 4. Establish guidelines for the reporting of user studies

The cartographic community needs to establish a set of best practices guidelines for the reporting of user studies. Just as guidelines can help improve the design of user studies, a similar set of reporting guidelines can improve the clarity, conciseness, and quality of detail of the community's final publications. I have already drafted preliminary guidelines with members of the ICA Commissions on Use, Users, and Usability, and I and other ICA members will continue working on them through 2018.

Recommendation 5: Develop appropriate qualitative and quantitative measures of best practices

Cartographers need to come to a consensus regarding which qualitative *and* quantitative criteria should be used to identify or judge potential best practices for user studies. Currently, CartoSDB only uses qualitative, contextual information to help establish best practices criteria. I argue that we need to determine which *measurable* components of a user study can provide sufficient evidence to guide our best practices decisions, and lead to the establishment of a comprehensive set of best practices guidelines that are applicable to all user study topics and methods.

REFERENCES

Abowd, G., Dey, A., Brown, P., Davies, N., Smith, M., and Steggles, P. 1999. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium, HUC'99, Karlsruhe, Germany, September 27-29, 1999: Proceedings* (Lecture notes in computer science, 1707). Gellersen, H. (Ed.), 304-307. Berlin: Springer.

Abras, C., Maloney-Krichmar, D., and Preece, J. 2004. User-centered design. In *Encyclopedia of Human-Computer Interaction.* Thousand Oaks: Sage Publications 37 (4): 445-456.

Alstete, J. W. 1995. Benchmarking in higher education: Adapting best practices to improve quality, *ASHE-ERIC Higher Education Reports No 5*, Washington. DC: The George Washington University Graduate School of Education and Human Development.

American Productivity and Quality Center. (1999). What is benchmarking?. Available online: www.apqc.org. (Last accessed 17 May 2017).

Anand, G., and Kodali, R. 2008. Benchmarking the benchmarking models. *Benchmarking: An International Journal*, 15 (3): 257-291.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M.-J., and MacEachren, A. 2007. Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, 21 (8): 839–857.

Baker, K. M., Petcovic, H., Wisniewska, M., and Libarkin, J. 2012. Spatial signatures of mapping expertise among field geologists. *Cartography and Geographic Information Science*, 39 (3): 119-132.

Baker, M. 2016, May 26. 1,500 scientists lift the lid on reproducibility. *Nature*, 533 (7604): 452–454.

Balciunas, A. 2011. The possibilities for functionality research on interactive Internet maps. *Cartographica*, 46 (2): 66-73.

Battersby, S., R. G. Golledge, and Marsh, M. J. 2006. Incidental learning of geospatial concepts across grade level: Map overlay. *Journal of Geography*, 105 (4): 139–146.

Berlin Declaration on open access to knowledge in the sciences and humanities. October 22, 2003. Updated 2013. Available online: http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung (Last accessed 15 May 2017).

Bernard, H. R. 2006. *Research methods in anthropology: Qualitative and quantitative approaches*, 4th ed. Lanham, MD: AltaMira Press.

Bethesda statement on open access publishing. June 20, 2003. Available online: http://dash.harvard.edu/bitstream/handle/1/4725199/suber _bethesda.htm?sequence=1 (Last accessed 15 May 2017).

Bhowmick, S., Chua, H., Choi, B., and Dyreson, C. 2017. VISUAL: Simulation of visual subgraph query formulation to enable automated performance benchmarking. *IEEE Transactions on Knowledge and Data Engineering*, 29 (8): 1765-1778.

Bifulco, R., and Bretschneider, S. 2001. Estimating school efficiency: A comparison of methods using simulated data. *Economics of Education Review*, 20: 417-29.

Boardman, D. 1990. Graphicacy revisited: Mapping abilities and gender differences. *Educational Review*, *42* (1): 57-64.

Bogan. C. E., and English. M. J. 1994. *Benchmarking for best practices: Winning through innovative adaptation*. New York: McGraw-Hill.

Bowman, D.A., Gabbard, J.L., and Hix, D. 2002. A survey of usability evaluation in virtual environments: Classification and comparison of methods. *Presence*, 11: 404–424.

Bretschneider, S., Marc-Aurele, F., and Wu, J. 2005. Best practices research: A methodological guide for the perplexed. *Journal of Public Administration Research and Theory*, 15 (2), 307-323.

Buckingham, B., and M. Harrower. 2007. The role of color saturation in maps for children. *Cartographic Perspectives*, 58: 28-47.

Budapest open access initiative. February 14, 2002. Available online: http://www.budapestopenaccessinitiative.org/read (Last accessed 15 May 2017).

Camp, R. 1989. *Benchmarking: The search for industry best practices that lead to superior performance*. Milwaukee, Wisconsin: ASQC Quality Press.

Camp, R. 1995. *Business process benchmarking: Finding and implementing best practices*. Milwaukee, Wisconsin: ASQC Quality Press.

Camp. R. 1998. *Global cases in benchmarking: Best practices from organizations around the world*. Milwaukee, Wisconsin: ASQC Quality Press.

Cartwright, W., Crampton, J., Gartner, G., Miller, S., Mitchell, K., Siekierska, E. and Wood, J. 2001. Geospatial information visualization user interface issues. *Cartography and Geographic Information Science*, 28: 45–60.

Chang, K. T., and Antes, J. R. 1987. Sex and cultural differences in map reading. *The American Cartographer*, 14 (1): 29-42.

Chavan, V., and Penev, L. 2011. The data paper: A mechanism to incentivize data publishing in biodiversity science. *BMC bioinformatics*, 12 (15): S2.

Çöltekin, A., Heil, B., Garlindini, S., and Fabrikant, S. I. 2009. Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, 36 (1): 5–17.

Çöltekin, A. 2015. Mix well before use: Understanding the key ingredients of user studies. Presented at the International Cartographic Association Workshop on Envisioning the Future of Cartographic Research, Curitiba, Brazil, 21 August.

Çöltekin, A., Lokka, I., and Zahner, M. 2016. On the usability and usefulness of 3D (geo)visualizations -- A focus on virtual reality environments. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B2*, 387–392.

Çöltekin, A., Bleisch, S., Andrienko, G., and Dykes, J. 2017. Persistent challenges in geovisualization – a community perspective. *International Journal of Cartography*, (2017): 1-25.

Committee on Support for Thinking Spatially. 2006. *Learning to think spatially*. Washington, D.C.: The National Academies Press.

Cresswell, J., Klassen, A., Clark, V., Smith, K. 2011. Best practices for mixed methods research in the health sciences. National Institutes of Health Office of Behavioral and Social Sciences Research.

Cromley, R. G., and Ye, Y. 2006. Ogive-based legends for choropleth mapping. *Cartography and Geographic Information Science*, 33 (4): 257-268.

Cruz-Toledo, J., Mckeague, M., Zhang, X., Giamberardino, A., Mcconnell, E., Francis, T., DeRosa, M., and Dumontier, M. 2012. Aptamer Base: A collaborative knowledge base to describe aptamers and SELEX experiments. *Database: The Journal of Biological Databases and Curation,* (2012): Bas006.

Deeb, R., Ooms, K. and De Maeyer, P. 2012. Typography in the eyes of Bertin, gender and expertise variation. *The Cartographic Journal*, 49 (2): 176–185.

Deeb, R., Ooms, K., Vanopbroeke, V., and De Maeyer, P. 2014. Evaluating the efficiency of typographic design: Gender and expertise variation. *The Cartographic Journal*, 51 (1): 75-86.

Demšar, U. 2007. Combining formal and exploratory methods for evaluation of an exploratory geovisualization application in a low-cost usability experiment. *Cartography and Geographic Information Science*, 34 (1): 29-45.

Dent., B. D., Torguson, J. S., and Holder, T. W. 2008. Cartography: Thematic map design. New York: McGraw Hill.

Dillemuth, J. A. 2005. Map design evaluation for mobile display. *Cartography and Geographic Information Science*, 32 (4): 285-301.

Dillemuth, J. A. 2009. Navigation tasks with small-display maps: The sum of the parts does not equal the whole. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 44 (3): 187-200.

Donert, K. 2009. Benchmarking GIS–A charter for European education. In *Learning with Geoinformation IV*. Heidelberg: Wichmann.

Dykes, J., Lloyd, D. and Radburn, R. 2007. Understanding geovisualization users and their requirements – a user-centred approach. (pp. 209–214). Maynooth: GIS research UK.

Elzakker, C. P. J. M. v., and Wealands, K. 2007. Use and users of multimedia cartography. In W. Cartwright, M. Peterson, and G. Gartner (Eds.), *Multimedia cartography* (pp. 487–504). Heidelberg: Springer.

Elzakker, C. P. J. M. v. 2015. Usability of methods and techniques of use and user research in cartography and geoinformation science. Presented at the International Cartographic Association Workshop on Envisioning the Future of Cartographic Research, Curitiba, Brazil, 21 August.

European Research Advisory Board. 2006. Scientific publication: Policy on Open Access. Available online: http://ec.europa.eu/research/eurab/pdf/eurab_scipub_report_recomm_dec06_en.pdf (Last accessed 26 October 2017).

Evans, B. J. 1997. Dynamic display of spatial data-reliability: Does it benefit the map user?. *Computers and Geosciences*, 23 (4): 409-422.

Fabrikant, S. I., Hespanha, S. R., and Hegarty, M. 2010. Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making. *Annals of the Association of American Geographers*, 100 (1): 13-29.

Field, K., and Demaj, D. 2012. Reasserting design relevance in cartography: Some concepts. *The Cartographic Journal*, 49 (1): 70-76.

Finn, M. P., and Thunen, D. 2014. Recent literature in cartography and geographic information science. *Cartography and Geographic Information Science*, 41(2): 179-192.

Fish, C., Goldsberry, K.P. and Battersby, S. 2011. Change blindness in animated choropleth maps: An empirical study. *Cartography and Geographic Information Science*, 38 (4): 350–362.

Fuhrmann, S., Ahonen-Rainio, P., Edsall, R. M., Fabrikant, S. I., Koua, E. L., Tobon, C., Ware, C., and Wilson, S. 2005. Making useful and useable geovisualization: Design and evaluation issues. In *Exploring geovisualization*, Dykes, J., MacEachren, A. M., and M. J. Kraak, Eds. 553-566. Amsterdam: Elsevier Science.

Fram, E. and Camp, R., 1995. Finding and implementing best practices in higher education. *Quality Progress*, 28 (2): 69.

Gabbard, J. L., Hix, D., and Swan, J. E. 1999. User-centered design and evaluation of virtual environments. *IEEE Computer Graphics and Applications*, 19 (6): 51-59.

Ghosh, S., Lohani, B., and Misra, N. 2014. A study-based ranking of LiDAR data visualization schemes aided by georectified aerial images. *Cartography and Geographic Information Science*, 41 (2): 138-150.

Gilmartin, P. P., and Patton, J. C. 1984. Comparing the sexes on spatial abilities: Map-use skills. *Annals of the Association of American Geographers*, 74 (4): 605-619.

Github: Build software better, together. Available online: https://github.com/ (Last accessed 27 October 2017).

Golledge, R., Marsh, M., and Battersby, S. 2008. A conceptual framework for facilitating geospatial thinking. *Annals of the Association of American Geographers*, 98 (2): 285-308.

Greenberg, S. 2001. Context as a dynamic construct. *Human-Computer Interaction*, 16 (2): 257–68.

Greiner, B. 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1 (1): 114-125.

Griffin, A. 2015. Towards usable user studies. Presented at the International Cartographic Association Workshop on Envisioning the Future of Cartographic Research, Curitiba, Brazil, 21 August.

Griffin, A., White, T., Fish, C., Tomio, B., Huang, H., Sluter, C. R.,…Picanço, P. 2017. Improving Our ability to Design across Map Use Contexts: A Research agenda for the International Cartographic Association on Use, Users, and Usability Issues. *International Journal of Cartography*, (2017): 1-25.

Haklay, M. and Tobon, C. 2003. Usability evaluation and PPGIS: Towards a user-centred design approach. *International Journal of Geographical Information Science*, 17 (6): 577-592.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., and Subbiah, I. 2002. Development of a self-report measure of environmental spatial ability. *Intelligence*, 30: 425-448.

Hegarty, M., Smallman, H., Stull, A., and Canham, M. 2009. Naïve Cartography: How Intuitions about Display Configuration Can Hurt Performance. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 44 (3): 171–186.

Hey, T., Tansley, S., and Tolle, K. M. 2009. *The fourth paradigm: Data-intensive scientific discovery* (Vol. 1). Redmond, WA: Microsoft research.

Hope and Hunter 2007: Testing the effects of thematic uncertainty on spatial decision making. *Cartography and Geographic Information Science*, 34 (3): 199-214.

International Cartographic Association Commission on Cognitive Issues in Geographic Information Visualization. Available online: http://cogvis.icaci.org/ (Last accessed 27 October 2017).

International Cartographic Association Commission on Use, User, and Usability Issues. Available online: http://use.icaci.org/ (Last accessed 27 October 2017).

Ishikawa, T., Barnston, A. G., Kastens, K. A., Louchouarn, P., and Ropelewski, C. F. 2005. Climate forecast maps as a communication and decision-support tool: An empirical test with prospective policy makers. *Cartography and Geographic Information Science*, 32 (1): 3-16.

Ishikawa, T. 2013. Geospatial thinking and spatial ability: An empirical examination of knowledge and reasoning in geographical science. *The Professional Geographer* 65 (4): 636-646.

Ishikawa, T. 2016. Spatial thinking in Geographic Information Science: Students' geospatial conceptions, map-based reasoning, and spatial visualization ability. *Annals of the American Association of Geographers*, 106 (1): 76-95.

ISO-9241-11. 1998. Ergonomic requirements for office work with visual display terminals (VDTs) – Part II, Guidance on usability. International Organisation for Standardisation.

Keehley, P 1997. *Benchmarking for best practices in the public sector: Achieving performance breakthroughs in federal, state, and local agencies.* San Francisco: Jossey-Bass.

Keehley, P., and Abercrombie, N. 2008. *Benchmarking in the public and nonprofit sectors : Best practices for achieving performance breakthroughs* (2nd ed.). San Francisco: Jossey-Bass.

Kinkeldey, C., MacEachren, A. M., and Schiewe, J. 2014. How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51 (4): 372-386.

Klippel, A., Hardisty, F., and Li, R. 2011. Interpreting spatial patterns: An inquiry into formal and cognitive aspects of Tobler's First Law of Geography. *Annals of the Association of American Geographers*, 101 (5): 1011-1031.

Klein, D. E., Woods, D. D., Klein, G., Perry, S. J., Mosier, K., Militello, L. G. 2016. Can we trust best practices? Six cognitive challenges of evidence-based approaches. *Journal of Cognitive Engineering and Decision Making*, 10 (3), 244-254.

Korpi, J. A., Haybatollahi, M., and Ahonen-Rainio, P. 2013. Identification of partially occluded map symbols. *Cartographic Perspectives*, 76: 19-32.

Kramers, R. E. 2008. Interaction with maps on the internet: A user centred design approach for the Atlas of Canada. *The Cartographic Journal*, 45 (2): 98-107.

Lee, J., and Bednarz, R. 2009. Effect of GIS learning on spatial thinking. *Journal of Geography in Higher Education*, *33* (2), 183-198

Lee, J., and Bednarz, R. 2012. Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography*, 111 (1): 15-26.

Li, Z., and Qin, Z. 2014. Spacing and alignment rules for effective legend design. *Cartography and Geographic Information Science*, 41 (4): 348-362.

LibreOffice: The Document Foundation. Available online: https://www.libreoffice.org/ (Last accessed 27 October 2017).

Lincoln, Y. S., and Guba, E. G. 1985. *Naturalistic Inquiry.* Thousand Oaks, CA: SAGE Publications.

Lloyd, R. E., and Bunch, R. L. 2003. Technology and map learning: Users, methods, and symbols. *Annals of the Association of American Geographers*, 93 (4): 828– 850.

Lloyd, R. E., and Bunch, R. L. 2005. Individual differences in map reading spatial abilities using perceptual and memory processes. *Cartography and Geographic Information Science*, 32(1), pp.33–46.

Lobben, A. K. 2007. Navigational map reading: Predicting performance and identifying relative influence of map-related abilities. *Annals of the Association of American Geographers*, 97 (1): 64-85.

Lobben, A. K. 2008. Influence of data properties on animated maps. *Annals of the Association of American Geographers*, 98 (3): 583-603.

Lobben, A. K., Lawrence, M., and Pickett, R. 2014. The map effect. *Annals of the Association of American Geographers*, 104 (1): 96-113.

Longbottom, D. 2000. Benchmarking in the UK: An empirical study of practitioners and academics. *Benchmarking: An International Journal*, 7 (2): 98-117.

MacDonald, A. P. 1972. Characteristics of volunteer subjects under three recruiting methods: Pay, extra credit, and love of science. *Journal of Consulting and Clinical Psychology*, 39 (2): 222.

MacEachren, A. M., and Kraak, M. J. 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science*, 28 (1): 3-12.

Marsh, S.L., and Haklay, M. 2010. Evaluation and deployment. In *Interacting with Geospatial Technologies*, ed. Haklay, M., 199-221. West Sussex, UK: Wiley-Blackwell.

Martin, G. N., and Clarke, R. M. 2017. Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology*, 8: 523.

McGuinness, C. 1994. Expert/novice use of visualization tools. In *Visualization in modern cartography*, eds. A. M. MacEachren and D. R. Fraser Taylor, 185-200. New York: Elsevier.

McMaster, R., and McMaster, S. 2002. A history of Twentieth-Century American academic cartography. *Cartography and Geographic Information Science*, 29 (3): 305-321.

Meirelles, I. 2013. *Design for information: An introduction to the histories, theories, and best practices behind effective information visualizations*. Beverly, MA: Rockport Publishers.

Moellering, H. 2012. The International Cartographic Association research agenda: Review, perspectives, comments, and recommendations. *Cartography and Geographic Information Science*, 39 (1): 61-68.

Montello, D. R. 2002. Cognitive map-design research in the twentieth century: Theoretical and empirical approaches. *Cartography and Geographic Information Science*, 29 (3): 283-304.

Murphy K. 2016, August 28. Do you believe in God, or is that a software glitch?. *New York Times.*

National Institutes of Health. 2006. Data Sharing Policy. Available online: http://grants.nih.gov/grants/policy/data_sharing/ (Last accessed 26 October 2017).

National Research Council. 2005. Measuring performance and benchmarking project management at the Department of Energy. Available online: https://doi.org/10.17226/11344 (Last accessed 27 October 2017).

National Science Foundation. 2005. Long-lived digital data collections enabling research and education in the 21st Century, NSB-05–40 Available online: http://www.nsf.gov/pubs/2005/nsb0540/ (Last accessed 26 October 2017).

National Science Foundation. 2011. National Science Foundation Division of Ocean Sciences Sample and Data Policy 2011, NSF11060. Available online: http://www.nsf.gov/pubs/2011/nsf11060/nsf11060.pdf (Last accessed 26 October 2017).

Neylon, C. October 27, 2011. Building the perfect data repository…or the one that might get used. [Blog post]. Available online: http://cameronneylon.net/blog/building-the-perfect-data-repository-or-the-one-that-might-get-used/ (Last accessed 15 May 2017).

Nielsen, J. 1993. *Usability engineering.* San Francisco: Morgan Kaufmann.

Nielsen, J., 2000. Why you only need to test with 5 users. Available online: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (Last accessed on 5 January 2017).

Nielsen, J., and Sova, D. H. 2003. How to recruit participants for usability studies. Available online: https://www.nngroup.com/reports/how-to-recruit-participants-usability-studies/ (Last accessed on 5 January 2017).

Norman, D.A. 1988. *The design of everyday things*. New York: Basic Books.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G. and Contestabile, M. 2015. Promoting an open research culture. *Science*, 348 (6242): 1422-1425.

Nosek, B., and Bar-Anan, Y. 2012. Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, 23 (3): 217-243.

Nosek, B., Spies, J., and Motyl, M. 2012. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7 (6): 615-631.

O'Dell, C., and Grayson, C. J. 1998. If only we knew what we know: Identification and transfer of internal best practices. *California management review*, 40 (3): 154-174.

Oakley, N., and Daudert, B. 2016. Establishing best practices to improve usefulness and usability of web interfaces providing atmospheric data. *Bulletin of the American Meteorological Society*, 97 (2): 263-274.

Olson, J., and Brewer, C. 1997. An Evaluation of color selections to accommodate map users with color-vision impairments. *Annals of the Association of American Geographers*, 87 (1): 103-134.

Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., and Witlox, F. 2012. Investigating the effectiveness of an efficient label placement method using eye movement data. *The Cartographic Journal*, 49 (3): 234–246.

Ooms, K., De Maeyer, P., and Fack, V. 2014. Study of the attentive behavior of novice and expert map users using eye tracking. *Cartography and Geographic Information Science*, 41 (1): 37-54.

Ooms, K., and L. Lapon. 2015. Collaborative, remote, digital, virtual, and mixed research methods in cartography and geovisualization. Presented at the International Cartographic Association Workshop on Envisioning the Future of Cartographic Research, Curitiba, Brazil, 21 August.

Opach, T., Gołebiowska, I., and Fabrikant, S. I. 2014. How do people view multi-component animated maps? *The Cartographic Journal*, 51 (4): 330–342.

Open Access Scholarly Publishers Association. Available online: https://oaspa.org (Last accessed 26 October 2017).

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science*, 349 (6251), aac4716.

Open Science Framework. Available online: https://osf.io (Last accessed 27 October 2017).

Organization for Economic Co-operation and Development. 2007. Principles and guidelines for access to research data from public funding. Available online: http://www.oecd.org/dataoecd/9/61/38500813.pdf (Last accessed 26 October 2017).

Osborne, J. W. 2008. *Best practices in quantitative methods*. Los Angeles: Sage Publications.

Pampel H, Vierkant P, Scholze F, Bertelmann R, Kindling M, Klump J, and Dierolf, U. 2013. Making research data repositories visible: The re3data.org Registry. *PLoS ONE*, 8 (11): e78080.

Panter, A., and Sterba, S. K. 2011. *Handbook of ethics in quantitative methodology* (Multivariate applications book series). New York: Routledge.

Paton, N. 2008. Managing and sharing experimental data: Standards, tools and pitfalls. *Biochemical Society Transactions,* 36 (Pt 1): 33-6.

Patterson, T., and Jenny, B. 2014. Evaluating cross-blended hypsometric tints: A user study in the United States, Switzerland, and Germany. *Cartographic Perspectives*, 75: 5-17.

Pingel, T. J., and Schinazi, V. R. 2014. The relationship between scale and strategy in search-based wayfinding. *Cartographic Perspectives*, 77: 33-45.

Plaisant, C. 2004. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced Visual Interfaces*, 109-116. ACM.

Plaisant, C., Fekete, J. D., and Grinstein, G. 2008. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. *IEEE transactions on visualization and computer graphics*, 14 (1): 120-134.

Popelka, S., and Brychtova, A. 2013. Eye-tracking study on different perception of 2D and 3D terrain visualisation. *The Cartographic Journal*, 50 (3): .240–246.

Prestopnik, J. L., and Roskos–Ewoldsen, B. 2000. The relations among wayfinding strategy use, sense of direction, sex, familiarity, and wayfinding ability. *Journal of Environmental Psychology* 20 (2): 177-191.

Public Library of Science. Available online: https://plos.org (Last accessed 26 October 2017).

Publishing Research Consortium. 2010. Access vs. importance. Available online: http://publishingresearchconsortium.com/ (Last accessed 26 October 2017).

Registry of Research Data Repositories. Available online: https://www.re3data.org (Last accessed 26 October 2017).

Reichenbacher, T. 2003. Adaptive methods for mobile cartography. In *Proceedings of the 21st International Cartographic Conference,* Durban, South Africa. CD-ROM. pp. 1311-21.

Reichman, O. J., Jones, M. B., and Schildhauer, M. P. 2011. Challenges and opportunities of open data in ecology. *Science*, 331 (6018): 703-705.

Resch, B., Hillen, F., Reimer, A., and Spitzer, W. 2013. Towards 4D cartography – Four-dimensional dynamic maps for understanding spatio-temporal correlations in lightning events. *The Cartographic Journal* 50 (3): 266-275.

Robinson, A. C., Chen, J., Lengerich, E. J., Meyer, H. G., and MacEachren, A. M. 2005. Combining usability techniques to design geovisualization tools for epidemiology. *Cartography and Geographic Information Science*, 32 (4): 243-255.

Robinson, A. C., Demšar, U., Moore, A. B., Buckley, A., Jiang, B., Field, K., Kraak, M-J., Camboim, S. P., and Sluter, C. R. 2017. Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *International Journal of Cartography* (2017): 1-29.

Rohrer, C. *When to use which user-experience research methods*. Nielsen Norman Group. Available online: http://www.nngroup.com/articles/which-ux-research-methods/ (Last accessed 11 November 2015).

Roth, R. E. and Harrower, M. 2008. Addressing map interface usability: Learning from the Lakeshore Nature Preserve interactive map. *Cartographic Perspectives*, 60: 46-66.

Roth, R. E. 2009. The impact of user expertise on geographic risk assessment under uncertain conditions. *Cartography and Geographic Information Science*, 36 (1): 29-43.

Roth, R. E. 2015. Challenges for Human Subjects Research in Cartography. Presented at the International Cartographic Association Workshop on Envisioning the Future of Cartographic Research, Curitiba, Brazil, 21 August.

Roth, R. E., Ross, K. S., and MacEachren, A. M. 2015. User-centered design for interactive maps: A case study in crime analysis. *ISPRS International Journal of Geo-Information*, 4 (1): 262-301.

Roth, R., Çöltekin, A., Delazari, L., Filho, H., Griffin, A., Hall, A., Korpi, J., Lokka, I., Mendonça, A., Ooms, K., and Elzakker, C P. J. v. 2017. User studies in cartography: opportunities for empirical research on interactive maps and visualization. *International Journal of Cartography*. (2017): 1-27.

Saunders, E. K. S. 1999. The applicability of benchmarking for best practices to improve the process of recruiting and admitting students into undergraduate preservice teacher education programs. (Doctoral dissertation). Illinois State University.

Schlichtmann, H. 2009. Overview of the semiotics of maps. In *Proceedings, 24th International Cartographic Conference, Santiago, Chile, 15-21 November 2009 - refereed papers, themes 30 (theory)*.

Schönbrodt, F., Gollwitzer, M., and Abele-Brehm, A. 2016. Data management in psychological science: specification of the DFG guidelines on behalf of the DGPs executive board (adopted Sep 17, 2016; current version: Nov 9, 2016). Available online: https://psyarxiv.com/vhx89/ (Last accessed 26 October 2017).

Schulze, U., Kanwischer, D., and Reudenbach, C. 2013. Essential competences for GIS learning in higher education: A synthesis of international curricular documents in the GISandT domain. *Journal Of Geography In Higher Education*, 37 (2): 257-275.

Self, C. M., and Golledge, R. G. 1994. Sex-related differences in spatial ability: What every geography educator should know. *Journal of Geography*, 93 (5): 234-243.

Sharp, E. C., Pelletier, L. G., and Lévesque, C. 2006. The double-edged sword of rewards for participation in psychology experiments. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *38* (3): 269.

Sholl, M. J., Acacio, J. C., Makar, R. O., and Leon, C. 2000. The relation of sex and sense of direction to spatial orientation in an unfamiliar environment. *Journal of Environmental Psychology*, 20 (1): 17-28.

Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrmann, S., and Hedley, N. R. 2001. Cognitive and usability issues in geovisualization. *Cartography and Geographic Information Science*, 28 (1): 61-75.

Slocum, T. A., Cliburn, D., Feddema, J., and Miller, J. 2003. Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*, 30 (1): 299–317.

Slocum, T. A., Sluter, R. S., Kessler, F. C. and Yoder, S. C. 2004. A qualitative evaluation of MapTime, a program for exploring spatiotemporal point data. *Cartographica*, 59 (3): 43-68.

Slocum, T. A., McMaster, R. B., Kessler, F. C., and Howard, H. H. 2009. *Thematic cartography and geovisualization*. Upper Saddle River, NJ: Pearson Prentice Hall.

Society of Professional Journalists. 1996. Code of Ethics, updated 2014. Available online: https://www.spj.org/ethicscode.asp (Last accessed 17 May 2017).

Soegaard, M., and Dam, R. F. 2012. *The encyclopedia of human-computer interaction*. The Interaction Design Foundation. Available online: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed (Last accessed 27 October 2017).

Spendolini, M. 1992. *The benchmarking book*. New York: Amacom.

Stewart, J., and Kennelly, P. 2010. Illuminated choropleth maps. *Annals of the Association of American Geographers*, 100 (3): 513-534.

Suber, P. 2012. *Open Access*. Cambridge, MA: MIT Press.

Swienty, O. 2005. The development of explorative geovisualization processes. In *Proceedings of the 22nd International Cartographic Conference*, A Coruña, Spain.

Taylor, W. A. 2004. Computer-mediated knowledge sharing and individual user differences: an exploratory study. *European Journal of Information Systems*, 13 (1): 52–64.

Taylor, W., and Plewe, B. 2006. The effectiveness of interactive maps in secondary historical geography education. *Cartographic Perspectives* 55: 16-33.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., and Frame, M. 2011. Data sharing by scientists: practices and perceptions. *PLoS ONE*, 6 (6): e21101.

Tobon, C. 2002. Usability testing for improving interactive geovisualization techniques. Centre for advanced Spatial analysis, University College London. Available online at: http://www.bartlett.ucl.ac.uk/casa/pdf/paper45.pdf (last accessed 1 Nov 2015).

Tobon, C. 2005. Evaluating geographic visualization tools and methods: An approach and experiment based upon user tasks. In *Exploring Geovisualization, Volume One*, ed. J. Dykes, A. M. MacEachren, and M. J. Kraak, 645-666. New York: Pergamon.

Toker, D., Conati, C., Carenini, G., and Haraty, M. 2012. Towards adaptive information visualization: on the influence of user characteristics. In *International Conference on User Modeling, Adaptation, and Personalization*, 274-285. Berlin: Springer.

Tomaszewski, B., and MacEachren, A. M. 2012. Geovisual analytics to support crisis management: Information foraging for geo-historical context. *Information Visualization*, 11, 4: 339-359.

Tsou, M. H. 2011. Revisiting web cartography in the United States: The rise of user-centered design. *Cartography and Geographic Information Science*, 38 (3): 250-257.

Tsou, M. H., and Curran, J. M. 2008. User-centered design approaches for web mapping applications: A case study with USGS hydrological data in the United States. In *International Perspectives on Maps and the Internet*, ed. M. P. Peterson, 301-321. Berlin: Springer.

Usability.gov. Available online: http://www.usability.gov/ (last accessed on 11 November 2015).

Vandenberg, S., and A. Kuse. 1978. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 48: 599-604.

Virrantaus, K., Fairbairn, D., and M. J. Kraak. 2009. ICA Research agenda on Cartography and GI Science. *The Cartographic Journal*, 46 (2): 63-75.

Ward, M., Grinstein, G., and Keim, D. 2015. *Interactive data visualization: Foundations, techniques, and applications* (2nd ed.). CRC Press.

Ward, M., and Theroux, K. 1997. Perceptual benchmarking for multivariate data visualization. *Scientific Visualization Conference, 1997*, 314.

Ware, M., and Mabe, M. 2015. The STM Report: An overview of scientific and scholarly journal publishing. The Hague: International Association of Scientific, Technical and Medical Publishers.

White, T. 2017a. Under-designed or under-reported? Participant assessment methods in cartographic user studies. (Doctoral dissertation). University of Kansas.

White, T. 2017b. Constructing a cartographic user study database to identify best practices. (Doctoral dissertation). University of Kansas.

Wicherts, J. M., Kievit, R. A., Bakker, M., and Borsboom, D. 2012. Letting the daylight in: Reviewing the reviewers and other ways to maximize transparency in science. *Frontiers in Computational Neuroscience,* 6:1662-5188.

Wilkening, J., and Fabrikant, S. I. 2013. How users interact with a 3D geo-browser under time pressure. *Cartography and Geographic Information Science*, 40 (1): 40-52.

Wilkinson, M. D. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 2016 (3): 160018.

Wilson, G., Aruliah, D., Brown, C., Hong, N.,…Wilson, P. 2014. Best practices for scientific computing. *PLoS Biol*, 12 (1): e1001745.

Yarkoni, T., Poldrack, R. A., Van Essen, D. C., and Wager, T. D. 2010. Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends in Cognitive Sciences*, 14 (11): 489-496.

Zenodo. Available online: https://zenodo.org/ (Last accessed 27 October 2017).

APPENDICES

Appendix A. CartoBase Data Table Organization and Field Descriptions

Study Identification Group

Study Data Category	**Study Data Variable**	
Study Identification	DOI	Digital Object Identifier
	Journal	Refereed journal of publication.
	Year	Year of publication.
	Volume	Journal volume of publication.
	Issue	Journal issue of publication.
	Author 1	First author's last name.
	Title	Article title.
	ExperimentTotal	Total number of user studies (experiments) in the article.
	ExperimentNumber	User study (experiment) of current record.
	Author 1 Institution	First author's institution or professional affiliation at time of publication.
	Author 2	Second author's last name.
	Author 2 Institution	Second author's institution or professional affiliation at time of publication.
	Author 3	Third author's last name.
	Author 3 Institution	Third author's institution or professional affiliation at time of publication.
Study TAR	Title	Article title (duplicate field for easy reference when reading the abstract
	Abstract	Published article abstract
	References	All references cited in the article
Study Notes	Purpose	Stated purpose of article as written by the study authors.
	Notes	Miscellaneous information about the study that was deemed noteworthy.

Study Parameters

Study Data Category	**Study Data Variable**	**Description**
Study Parameters	StudyDesign	Experiment design. Currently three options: (1) Between Groups, (2) Within Subjects, (3) Mixed.
	Data Gathering Method	How study data were gathered. Currently three options: (1) Researcher Observed, (2) Self Reported, (3) Mixed.
	StudySetting1	Reporting status of study setting. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	StudySettingDetails	Specific setting of the user study. Currently ### options: (1) Laboratory, (2) Classroom, (3) Online, (4) Field, (5) Multiple sites.
	Length1	Reporting status of study length. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	LengthDetails	Study length in minutes , if reported.
	IV_total	Total number of independent variables in the experiment.
	IV_UserDiff1	This field indicates if a user difference was used as an independent variable.
	DV_total	Total number of dependent variables in the experiment.
Study IVs	IV_Study	The name of the independent variable. A separate table was necessary because each study could have multiple entries.
	IV User Diff Category	If the independent variable is a user difference, this field indicates what kind. Four possible options: (1) Demographic, (2) Expertise / Education, (3) Physiological, (4) Cognitive.
Study DVs	DV_Study	The name of the dependent variables. A separate table was necessary because each study could have multiple entries.
Study Procedure	Procedure	This field contains information regarding the study (experiment) procedure. A separate table was necessary because each study could have multiple entries.
Study Methods	Method	This field contains any notable quantitative or qualitative methods that were used in the study. A separate table was necessary because each study could have multiple entries.

Participant Sample Data

Study Data Category	Study Data Variable	Description
Sample Information	Sample Size 1	Reporting status of the participant sample size. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Sample Size 2	Specific participant sample size, if reported.
	Recruit Method 1	Reporting status of the participant recruitment method. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Incentive 1	Reporting status of any incentives or remunerations given for participation. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Incentive 2	Specific incentive or remuneration, if reported. Most were either class credit or monetary.
	Participant Paragraph	This field indicates if the authors reported the pertinent details of the participant sample in a separate paragraph. The reasoning is that organized articles are easier to compare and evaluate.
	Participant Subsection	This field indicates if the article contained a separate section or sub-section for the study participants. The reasoning is that organized articles are easier to compare and evaluate.
	University Students	This field indicates whether the study sample was composed of university students. In general, university students form convenience samples.
Sample Recruit Method	Recruit Method Details	Specific methods used to recruit participants. A separate table was necessary because each study could have multiple entries.

User Differences Information

Study Data Category	**Study Data Variable**	**Description**
User Differences Overview	Demographic	This field indicates if the study assessed or gathered demographic information about the participants.
	Expertise / Education	This field indicates if the study assessed or gathered expertise or education information about the participants.
	Physiological	This field indicates if the study assessed or gathered physiological information about the participants.
	Cognitive	This field indicates if the study assessed or gathered cognitive information about the participants.
	Sex	Reporting status of participant sex. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Age	Reporting status of participant age. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Demographic Other	Reporting status of other demographic information. The specific demographic variable, and its method of assessment, appear in the "Assessments Details" table. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Education Level 1	Reporting status of participant education levels. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Education Level 2	This field contains the general education level of participant sample. Current options are: (1) primary, (2) secondary, (3) university, (4) professional, and (5) various.
	Major Field 1	Reporting status of participant academic major or professional field. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Level of Expertise	The level of expertise of all participants. If reported, options include: (1) Novice, (2) Intermediate, (3) Expert, and (4) Novice vs. Expert.
	Expertise Definition	Reporting status of a published definition for a "novice," "expert," or "expertise" in general. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Expertise Definition Basis	If reported, this field indicates what the basis of the definition of expertise. Limited to (1) education, (2) experience, or (3) expertise in a specific domain.
	Expertise Definition Style	This field indicates whether the study authors provided an explicit definition of expertise, or if the definition was inferred through the article text.
User Education Lvl	Education Level Details	Specific education level of the participant sample, if reported.
User Major Field	Major Field Details	Specific academic majors or professional fields of the participants, if reported.
Expertise Def Quotes	Definition Pull Quotes	Direct quotes or excerpts from the article regarding the definition of expertise

Participant Assessment Methods

Study Data Category	Study Data Variable	Description
Assessments Overview	Assessment Method 1	Reporting status of the participant assessment method. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Assessment Medium 1	Reporting status of the participant assessment medium or format. Four options: (1) Reported as Not Gathered, (2) Reported Details Provided, (3) Reported No Details, (4) Not Reported.
	Assessment Medium 2	If reported, the specific medium of the assessment. Three options: (1) Paper, (2) Computer, (3) Smartphone, (4) Multiple
	Assessment Seq 1	Reporting status of where in the study an assessment occurred.
	Assessment Seq 2	If reported, where in the user study the participant assessment occurred. Four options: (1) Pre-experiment, (2) Post-experiment, (3) Pre- and Post-, and (4) Throughout.
	Self Assess	This field indicates if the assessment method was self-administered.
	Controlled Assess	This field indicates if the assessment method was controlled by the study author or experimenter.
	Assessment Sources	This field indicates if the study author provided the source of the original assessment method, or if they provided evidence of its effectiveness.
Assessments Details	User Diff Category	This field displays the category of the assessed difference. Four options: (1) Demographic, (2) Expertise or Education, (3) Physiological, (4) Cognitive.
	User Diff Detail	The specific user difference that was assessed.
	Assessment Description	This field indicates if the study authors provided a description of how the assessment method worked.
	Assessment Method	This field lists which method was used to assess the difference listed in the "User Diff Detail" field.
	Assessment Source	This field indicates if the study author provided a specific source of the method used to assess the difference listed in the "User Diff Detail" field.

Appendix B. CartoBase Tutorial for LibreOffice Base

LibreOffice (LO) is a free and open source office productivity suite similar to Microsoft (MS) Office. Each download includes six applications, including Base (similar to MS Access), writer (similar to MS Word), Calc (MS Excel), and Impress (MS PowerPoint). CartoBase was created in Base. This brief tutorial walks you through the processes of downloading and installing LO, then downloading and using CartoBase. Aside from the LO software, all documents and files are available at the CartoBase Open Science Foundation (OSF) project page: https://osf.io/ur6xn/

Download LibreOffice

To use CartoBase, you must first install LibreOffice on your system.

1. Navigate to the LO download page: https://www.libreoffice.org/download/download/. As of November 6, 2017, two versions of LO are available, 5.4.3 and 5.3.7. For this tutorial, download **v5.4.3**.

2. Select your operating system from the "Choose your operating system" dropdown. The software is available for multiple operating systems, including MS Windows, Mac OS X, and GNU/Linux.

3. Select DOWNLOAD. Depending on your operating system, the download package size is 200-250 MB. You will be invited to give an optional donation, which you may ignore.

4. Once the download is complete, Install the software. Follow the standard installation procedure of your specific operating system. For example: Mac users will download a .DMG file that they drag and drop into the Applications folder; MS Windows users will download a .MSI file that will open an Installation Wizard dialog box to guide you through the process. If necessary, installation instructions for each operating system are available at: https://www.libreoffice.org/get-help/install-howto/.

Download CartoBase

This version of CartoBase uses an .ODB (Open Document Database) file format which contains an internal HSQLDB (**H**yper **SQL** **D**ata**b**ase) engine. All tables and records are stored internally within the .ODB file rather than in a shared, external database, meaning that any changes you make to the database will only affect your downloaded copy of CartoBase. No one else will see those changes.

1. Navigate to the CartoBase OSF project page: https://osf.io/ur6xn/

2. Select "CartoBase_LibreOfficeBase.obd", which will open a new page.

3. Select Download from the toolbar. The current CartoBase file size is only 126 KB. Save the file to your preferred location. The file is ready for use.

Working with CartoBase

LO Base uses a relational database, which means that all fields in the database have specific relationships with each other. To borrow an example from the LO "Getting Started" guide, CartoBase contains fields for the names of the study authors, and another field for each individual user study. There is an obvious relationship between the authors and the studies they have published. CartoBase may contain more than one study by the same author. Similarly, each author may publish in multiple journals over multiple years; CartoBase keeps track of these relationships. If you want additional information about LO Base, you may view or download the LO Base handbook from the OSF project page: https://osf.io/8sc6a/

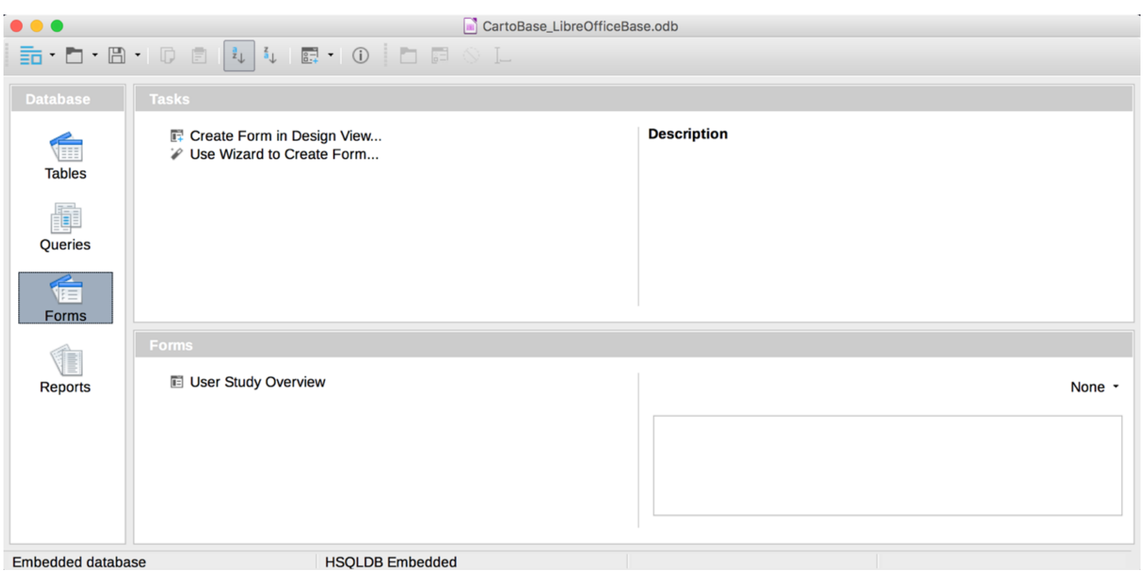
There are four basic database components:

- **Tables**: Stores information in a group of fields (columns) and records (rows). For example, the Study DVs table contains three fields: TableID (a unique identifier for each record in this specific table), StudyID (a unique identifier for each user study in CartoBase), and DV (the dependent variable used in the study, such as completion time or accuracy of response).
- **Queries**: Used to retrieve specific information from a database. Query results create special tables within the database. I will demonstrate Queries below.
- **Forms**: The front end of the data used to enter data into the database and edit data.
- **Reports**: Compiles and arranges information from a single table, view, or query.

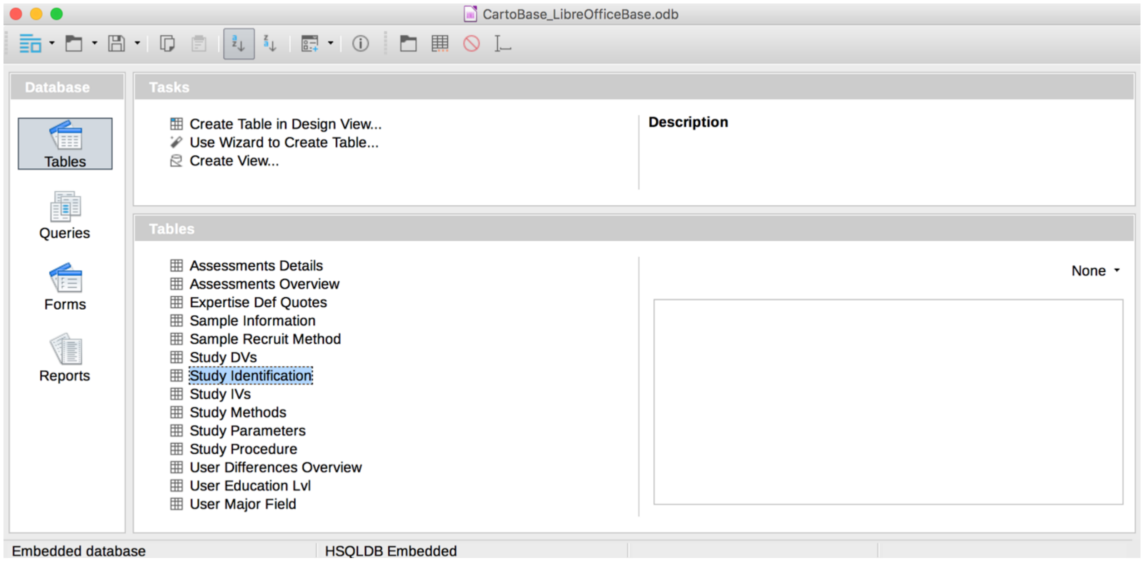
In this tutorial, I will provide overviews of the CartoBase tables and entry form, discuss two existing queries, and provide a step-by-step guide to create a new query. I did not use or create any reports and will not cover this component in the tutorial.

CartoBase Tables & Forms

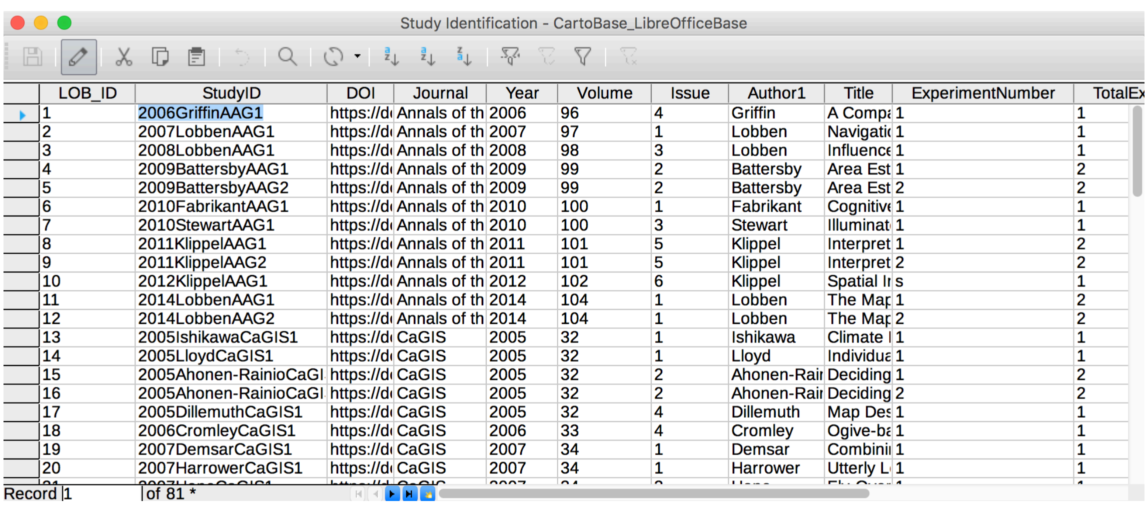
1. Open CartoBase_LibreOfficeBase.odb. The default screen view will look like the image below. Notice that the four database components are listed in the left-most column.



2. Select the Tables icon. This will display fourteen separate tables which constitute CartoBase, as shown in the image below. Each of these tables are available for individual download from the OSF project page. Note that two tables ("Study TAR" and "Study Notes") do not appear in this version of CartoBase; these tables contain large quantities of text data (e.g., article abstracts, full bibliographies) that impede functionality.



3. Open the "Study Identification" table (double-click). This table contains many of the identifying information of each study, including author last names, and journal, volume, issue, and year of publication. Importantly, this also contains the "StudyID" field, the unique identifier or "Key" for each individual user study. StudyID is created by combining the year of publication, the first authors' last name, the journal of publication, and a number to indicate which study experiment the StudyID refers to. For example, Battersby 2009 reported on two separate user studies: 2009BattersbyAAG1 refers to the first study experiment and 2009BattersbyAAG2 refers to the second experiment.

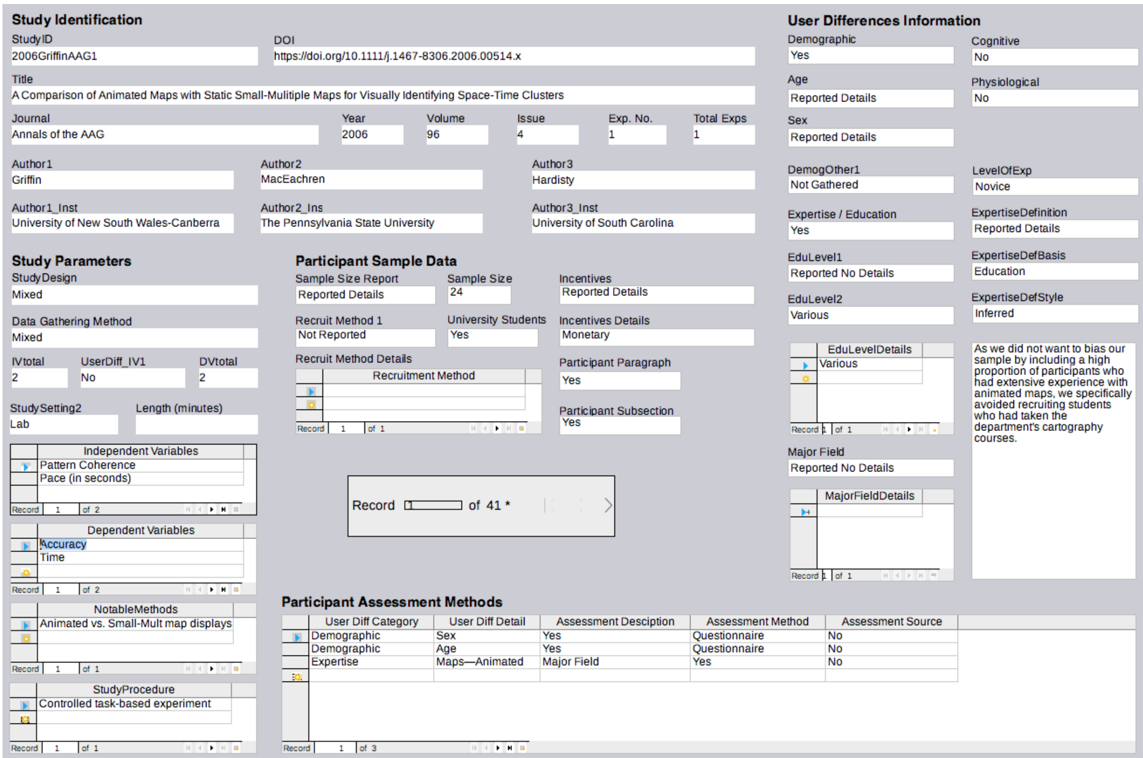

Study Identification - CartoBase_LibreOfficeBase

LOB_ID	StudyID	DOI	Journal	Year	Volume	Issue	Author1	Title	ExperimentNumber	TotalEx
1	2006GriffinAAG1	https://d	Annals of th	2006	96	4	Griffin	A Compa	1	1
2	2007LobbenAAG1	https://d	Annals of th	2007	97	1	Lobben	Navigati	1	1
3	2008LobbenAAG1	https://d	Annals of th	2008	98	3	Lobben	Influence	1	1
4	2009BattersbyAAG1	https://d	Annals of th	2009	99	2	Battersby	Area Est	1	2
5	2009BattersbyAAG2	https://d	Annals of th	2009	99	2	Battersby	Area Est	2	2
6	2010FabrikantAAG1	https://d	Annals of th	2010	100	1	Fabrikant	Cognitive	1	1
7	2010StewartAAG1	https://d	Annals of th	2010	100	3	Stewart	Illuminat	1	1
8	2011KlippelAAG1	https://d	Annals of th	2011	101	5	Klippel	Interpret	1	2
9	2011KlippelAAG2	https://d	Annals of th	2011	101	5	Klippel	Interpret	2	2
10	2012KlippelAAG1	https://d	Annals of th	2012	102	6	Klippel	Spatial I	s	1
11	2014LobbenAAG1	https://d	Annals of th	2014	104	1	Lobben	The Map	1	2
12	2014LobbenAAG2	https://d	Annals of th	2014	104	1	Lobben	The Map	2	2
13	2005IshikawaCaGIS1	https://d	CaGIS	2005	32	1	Ishikawa	Climate I	1	1
14	2005LloydCaGIS1	https://d	CaGIS	2005	32	1	Lloyd	Individua	1	1
15	2005Ahonen-RainioCaGI	https://d	CaGIS	2005	32	2	Ahonen-Rai	Deciding	1	2
16	2005Ahonen-RainioCaGI	https://d	CaGIS	2005	32	2	Ahonen-Rai	Deciding	2	2
17	2005DillemuthCaGIS1	https://d	CaGIS	2005	32	4	Dillemuth	Map Des	1	1
18	2006CromleyCaGIS1	https://d	CaGIS	2006	33	4	Cromley	Ogive-ba	1	1
19	2007DemsarCaGIS1	https://d	CaGIS	2007	34	1	Demsar	Combini	1	1
20	2007HarrowerCaGIS1	https://d	CaGIS	2007	34	1	Harrower	Utterly L	1	1

Record 1 of 81 *

4. Close the table. Every table contains a table ID field and the StudyID field; the table ID is specific to each table, whereas the StudyID is used to relate all table records and fields.

5. Select the Forms icon. There is only one form for CartoBase, "User Study Overview." Double-click to open. The default view will be similar to the image below. This is the entry form one can use to add new studies to CartoBase, although that feature is currently locked to prevent anyone from inadvertently altering the database contents.



6. To scroll through individual records, enter a record number or use the arrows in the navigation pane. Some data fields appear on a single line, while others are displayed in small tables. Note that you cannot query or search the database using a form. To do this, you must close the form, open an existing query, or create a new query.

CartoBase Queries

As explained in Chapters Three and Four, a user can filter or query CartoBase using one data field or a combination of data fields. Queries allow the user to combine records from multiple tables, filter a specific record from a mass of data, or group records based on one or more data attributes. The CartoBase_LibreOfficeBase.odb file contains ten pre-loaded database queries which were used as part of the CartoBase demonstration in Chapter Three. This section of the tutorial examines the basic structure underlying a couple of these queries, and then builds new query from scratch.

1. Select the Queries icon. This will display the ten existing queries, as shown in the image below. Each query name lists the corresponding table, the query, and a brief description of the data attributes sorted or filtered by the query.

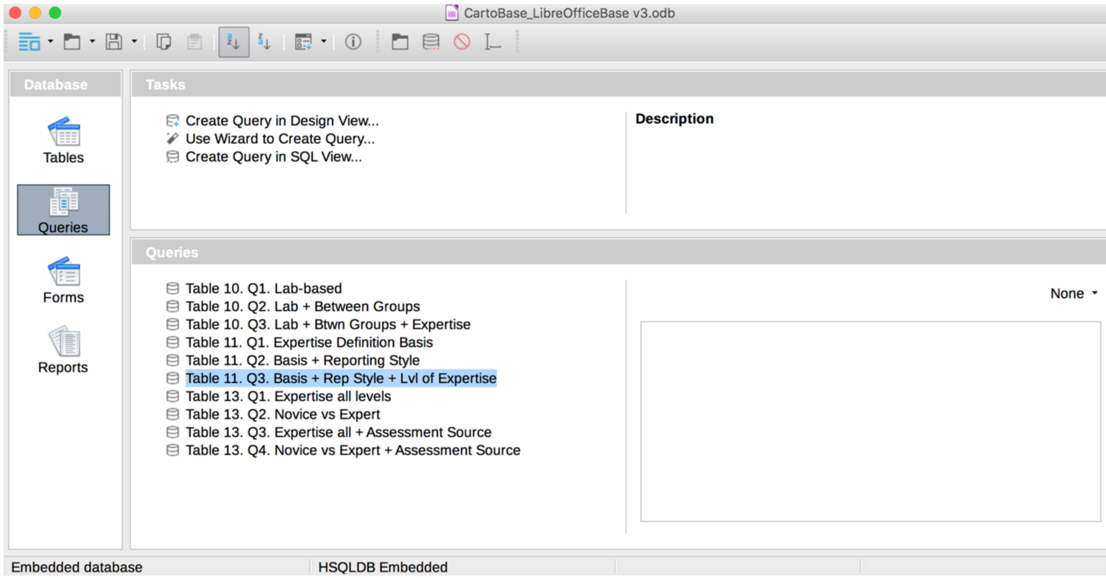


Table 10, found on page 70 and reprinted below, displays the subset of studies matching three filtering parameters. The filters are actually three separate query functions, each with a separate set of filtering parameters.

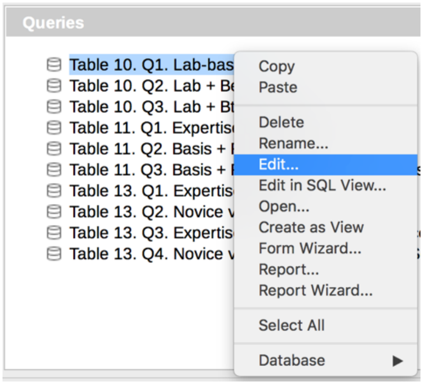
Table 10. Subset of studies matching three filtering parameters.

Filter	*n*	*%*	*Study subset*
All studies	110	100%	
Lab-based	53	48%	
Lab-based + Between-groups	13	12%	
Lab-based + Between-Groups + Novice vs. Experts	3	3%	Lloyd 2005; Bearman 2010; Ooms 2014

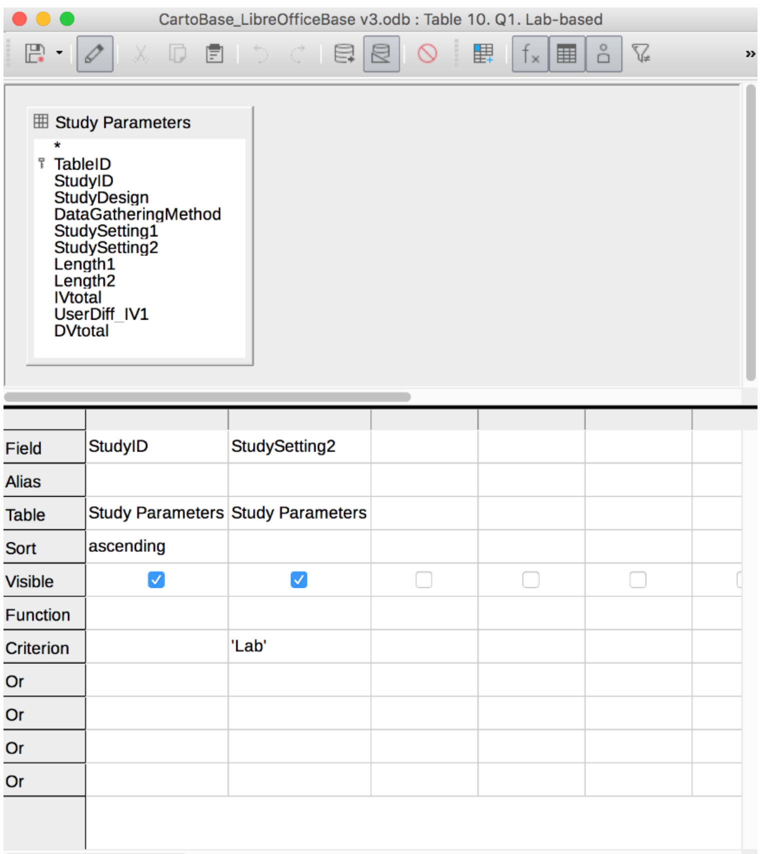
2. Double-click to view the results of the query “Table 10. Q1. Lab-based”. The query filtered CartoBase to isolate all lab-based user studies. The results table displays two fields: the StudyID of each of the 53 lab-based studies, and the StudySetting2 field, all of which display “Lab.”

3. Close the query results table.

4. Now right-click on the same query and select “Edit” to open the Query Design dialog.



As the image below shows, the query was constructed using just two fields from the Study Parameters table (StudyID and StudySetting2), and one filtering criterion (StudySetting2 = 'Lab'). All retained records are sorted in ascending order by StudyID.

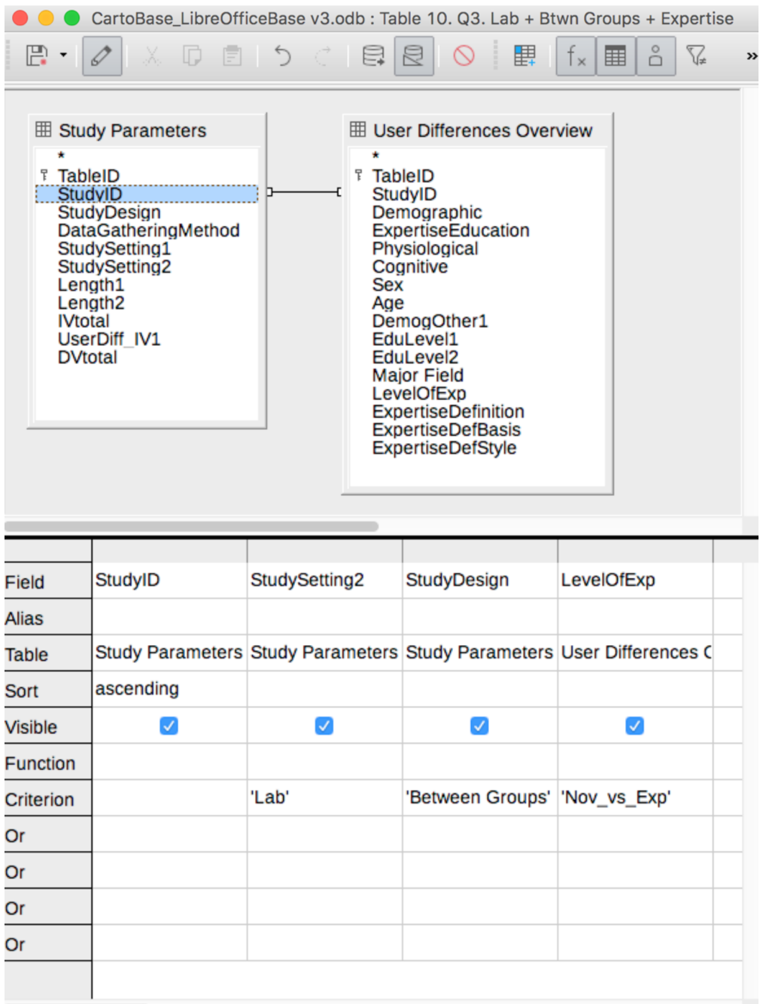


5. Close the query.

6. Double-click to view the results of the query "Table 10. Q3. Lab + Btwn Groups + Expertise". This displays the three user studies from Table 10 that match the three filtering criteria. The results table displays four fields: the StudyID of each study, StudySetting2 StudyDesign, and LevelOfExp.

7. Close the results table.

8. Right-click on the query and select Edit. The Query Design Dialog will look like the image below.



To combine data from two separate tables, a query must identify which fields from each table are "related." The image above uses a line with extenders to indicate which two fields (StudyID and StudyID) are joined to establish the relationship. Most CartoBase relationships are based on the StudyID field, which appears in every data table. Filter fields from either data table may be selected once they are joined. This query uses two additional filtering criteria besides StudySetting2 = 'Lab': StudyDesign = 'Between Groups' and LevelOfExp = 'Nov_vs_Exp'.

9. Close the query.

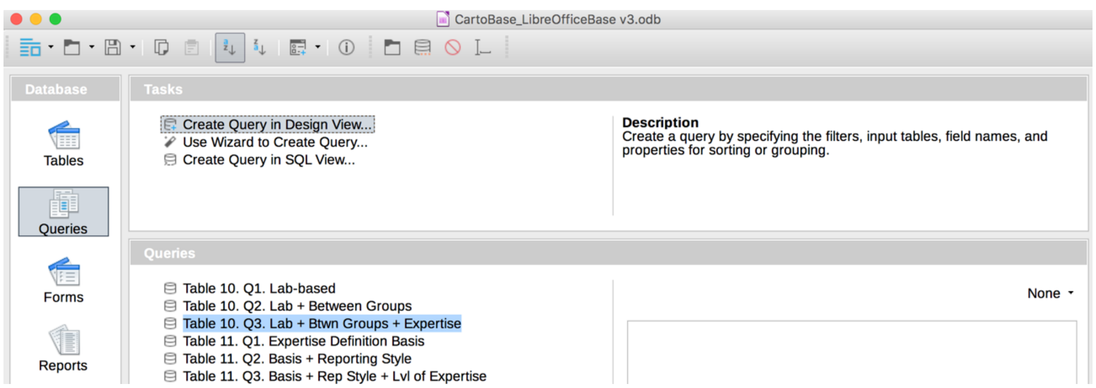
10. Now the tutorial will demonstrate how to build a new query from scratch. This query will produce the results that appear in Table 16 from Chapter Four, page 110, reprinted below.

Table 16 lists the subset of CartoBase studies matching three Case Study filters. Here we will create the query which identified those final two studies.

Table 16. Subset of CartoBase studies matching the Case Study filters.

Filter	*n*	*%*	*Study subset*
All CartoBase studies	110	100%	
Best Practice Criteria (assesses GIS proficiency)	16	15%	
Relevant Contextual Information (college-aged; approx. half hour length; novice vs. experts)	13	12%	
Best Practice Criteria + Relevant Contextual Information	2	2%	Hope and Hunter 2007; Opach et al. 2014

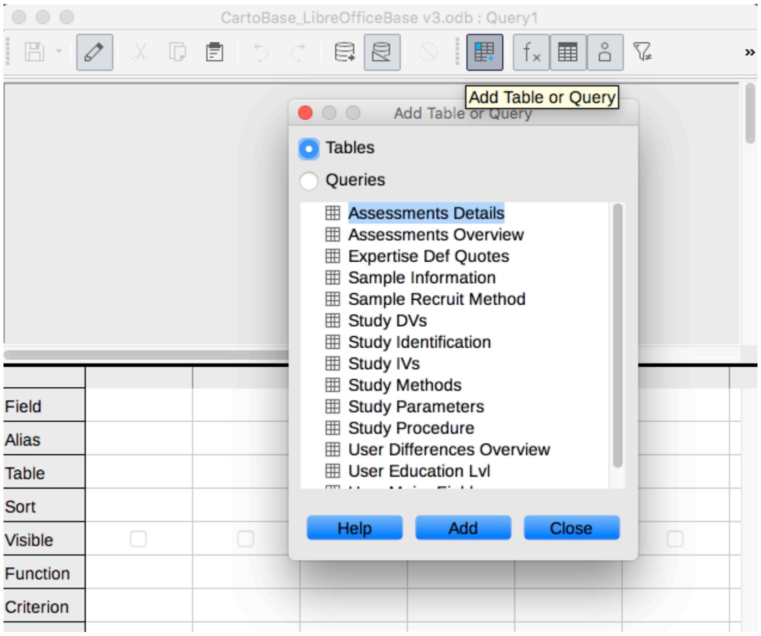
11. From the Tasks submenu, select "Create Query in Design View…" which will open up a blank query template.



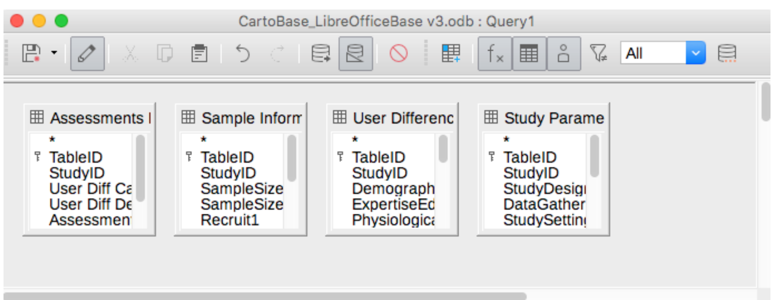
The query will use the following data tables, fields, and filtering criteria:

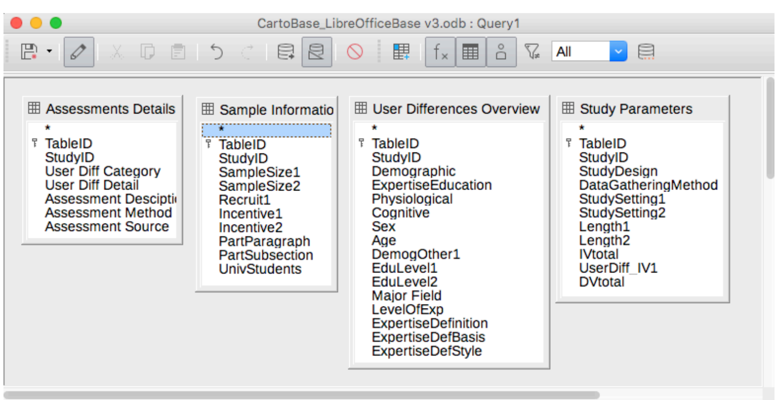
Tables	**Fields**	**Filtering Criteria**
Assessments Details	Study ID	n/a
Assessments Details	User Diff Category	Expertise
Assessments Details	User Diff Detail	GIS
Sample Information	UnivStudents	Yes & Partial
User Differences Overview	LevelOfExp	Nov_vs_Exp
Study Parameters	Length2	> 10 & < 59

12. First, add four tables to the Query Form: Assessments Details, Sample Information, User Differences Overview, and Study Parameters. To add each table, select it from the menu and click "Add".

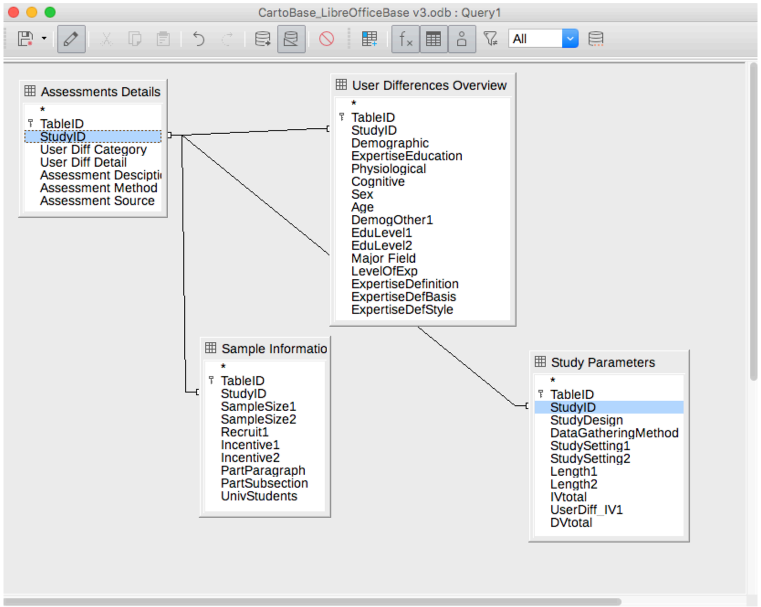


13. Once the tables are added to the graphical table pane, click-and-drag to move or resize them and make all available data fields visible.

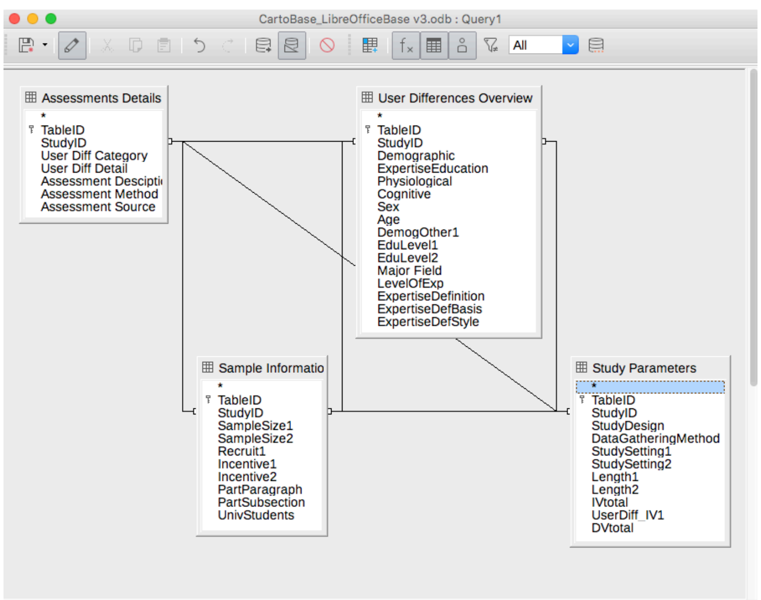




14. Now define the table relationships by joining the related fields. To achieve this, click and drag the mouse cursor from the StudyID of one table and release on the StudyIDs of the other three tables. The result will look like this:



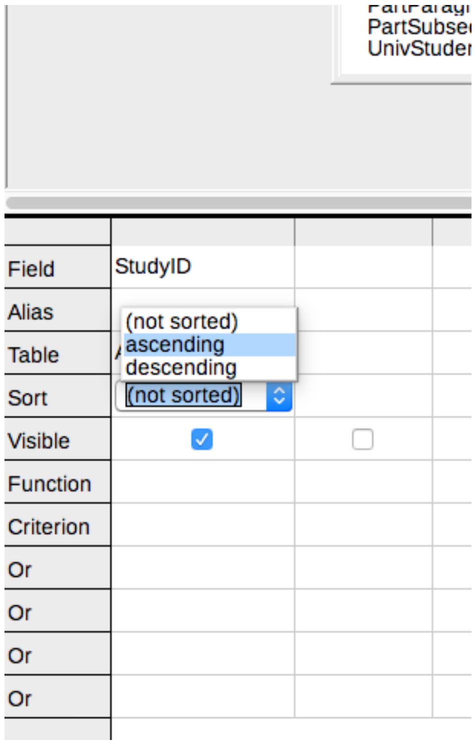
15. Repeat this until all four StudyID fields are joined. The result will look something like this:



16. Once the tables are all linked, add the desired fields to the query. In the subsequent steps, you will click and drag each of the six data fields into the blank columns of the Query Dialog Box.

17. Start with the StudyID from the Assessments Details. Drag and drop it into the empty field cell of the first column.

18. Change the Sort option for StudyID to "ascending". This field will remain blank until you click on it.



19. Next, drag and drop each of the remaining five query fields:

Tables	**Fields**	**What it queries**
Assessments Details	User Diff Category	All assessments of expertise
Assessments Details	User Diff Detail	All assessments of GIS proficiency
Sample Information	UnivStudents	All studies that used university students
User Differences Overview	LevelOfExp	Comparisons of novice vs. expert users
Study Parameters	Length2	Average time of study in minutes

The result will look like this:

Field	StudyID	User Diff Category	User Diff Detail	UnivStudents	LevelOfExp	Length2
Alias						
Table	Assessments Deta	Assessments Deta	Assessments Deta	Sample Informatio	User Differences (	Study Parameters
Sort	ascending					
Visible	☑	☑	☑	☑	☑	☑
Function						

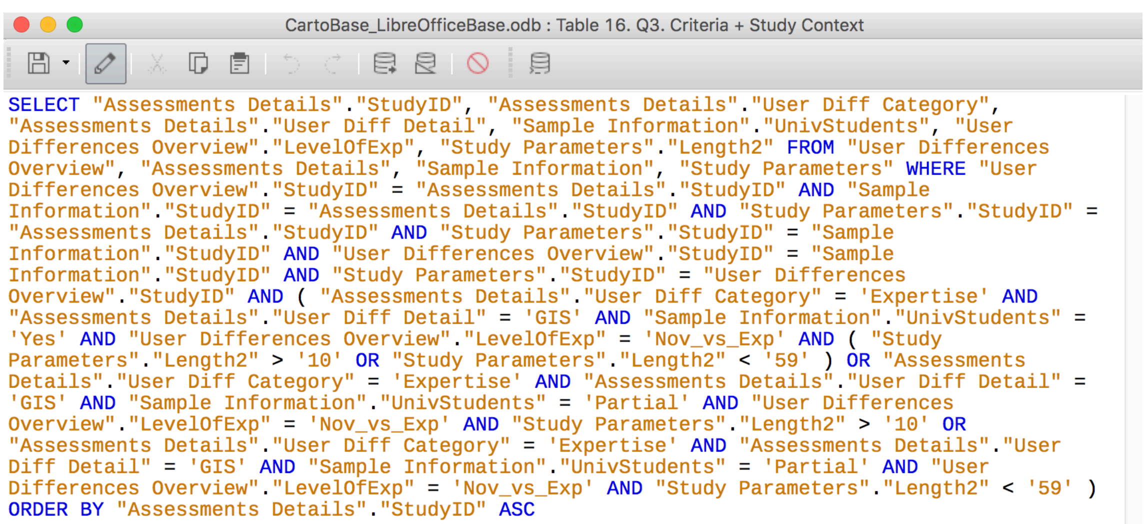
20. Type in the filtering criteria displayed in the image below. LibreOffice automatically adds tic marks once a field has been filled. There are multiple conditions, so multiple criterion rows will be filled.

Fields	Criterion	Purpose of filter
User Diff Category	Expertise	retains studies that assessed expertise
User Diff Detail	GIS	retains assessments of GIS proficiency
UnivStudents	Yes & Partial	retains studies that used university students
LevelOfExp	Nov_vs_Exp	retains studies that compared novice and expert users
Length2	> 10 and < 59	retains studies that averaged between 10 and 59 minutes

Field	StudyID	User Diff Category	User Diff Detail	UnivStudents	LevelOfExp	Length2
Alias						
Table	Assessments Deta	Assessments Deta	Assessments Deta	Sample Informatio	User Differences (	Study Parameters
Sort	ascending					
Visible	☑	☑	☑	☑	☑	☑
Function						
Criterion		'Expertise'	'GIS'	'Yes'	'Nov_vs_Exp'	> '10'
Or		'Expertise'	'GIS'	'Yes'	'Nov_vs_Exp'	< '59'
Or		'Expertise'	'GIS'	'Partial'	'Nov_vs_Exp'	> '10'
Or		'Expertise'	'GIS'	'Partial'	'Nov_vs_Exp'	< '59'

21. Click on the Save icon (the disk in the upper-left of the window) to save the query. Name it “Table 16. Q3. Criteria + Study Context v2” for easy identification.

22. While this query was created using the graphical user interface, the exact same query can be constructed and/or edited using SQL. To view this, select the “Switch Design View On/Off” button found on the middle of the top toolbar:




```
SELECT "Assessments Details"."StudyID", "Assessments Details"."User Diff Category",
"Assessments Details"."User Diff Detail", "Sample Information"."UnivStudents", "User
Differences Overview"."LevelOfExp", "Study Parameters"."Length2" FROM "User Differences
Overview", "Assessments Details", "Sample Information", "Study Parameters" WHERE "User
Differences Overview"."StudyID" = "Assessments Details"."StudyID" AND "Sample
Information"."StudyID" = "Assessments Details"."StudyID" AND "Study Parameters"."StudyID" =
"Assessments Details"."StudyID" AND "Study Parameters"."StudyID" = "Sample
Information"."StudyID" AND "User Differences Overview"."StudyID" = "Sample
Information"."StudyID" AND "Study Parameters"."StudyID" = "User Differences
Overview"."StudyID" AND ( "Assessments Details"."User Diff Category" = 'Expertise' AND
"Assessments Details"."User Diff Detail" = 'GIS' AND "Sample Information"."UnivStudents" =
'Yes' AND "User Differences Overview"."LevelOfExp" = 'Nov_vs_Exp' AND ( "Study
Parameters"."Length2" > '10' OR "Study Parameters"."Length2" < '59' ) OR "Assessments
Details"."User Diff Category" = 'Expertise' AND "Assessments Details"."User Diff Detail" =
'GIS' AND "Sample Information"."UnivStudents" = 'Partial' AND "User Differences
Overview"."LevelOfExp" = 'Nov_vs_Exp' AND "Study Parameters"."Length2" > '10' OR
"Assessments Details"."User Diff Category" = 'Expertise' AND "Assessments Details"."User
Diff Detail" = 'GIS' AND "Sample Information"."UnivStudents" = 'Partial' AND "User
Differences Overview"."LevelOfExp" = 'Nov_vs_Exp' AND "Study Parameters"."Length2" < '59' )
ORDER BY "Assessments Details"."StudyID" ASC
```

For legibility, I have reformatted the code below:

```
SELECT "Assessments Details"."StudyID", "Assessments Details"."User Diff
  Category", "Assessments Details"."User Diff Detail", "Sample
  Information"."UnivStudents", "User Differences Overview"."LevelOfExp", "Study
  Parameters"."Length2"

  FROM "User Differences Overview", "Assessments Details", "Sample Information",
    "Study Parameters"

    WHERE "User Differences Overview"."StudyID" = "Assessments
      Details"."StudyID"

    AND "Sample Information"."StudyID" = "Assessments Details"."StudyID"
    AND "Study Parameters"."StudyID" = "Assessments Details"."StudyID"
    AND "Study Parameters"."StudyID" = "Sample Information"."StudyID"
    AND "User Differences Overview"."StudyID" = "Sample Information"."StudyID"
    AND "Study Parameters"."StudyID" = "User Differences Overview"."StudyID"
    AND ( "Assessments Details"."User Diff Category" = 'Expertise'
        AND "Assessments Details"."User Diff Detail" = 'GIS'
        AND "Sample Information"."UnivStudents" = 'Yes'
        AND "User Differences Overview"."LevelOfExp" = 'Nov_vs_Exp'
        AND ( "Study Parameters"."Length2" > '10' OR "Study Parameters"."Length2"
          < '59' )

    OR "Assessments Details"."User Diff Category" = 'Expertise'
        AND "Assessments Details"."User Diff Detail" = 'GIS'
        AND "Sample Information"."UnivStudents" = 'Partial'
        AND "User Differences Overview"."LevelOfExp" = 'Nov_vs_Exp'
        AND "Study Parameters"."Length2" > '10'

    OR "Assessments Details"."User Diff Category" = 'Expertise'
        AND "Assessments Details"."User Diff Detail" = 'GIS'
        AND "Sample Information"."UnivStudents" = 'Partial'
        AND "User Differences Overview"."LevelOfExp" = 'Nov_vs_Exp'
        AND "Study Parameters"."Length2" < '59' )

ORDER BY "Assessments Details"."StudyID" ASC
```

23. Run the Query by clicking on the "Run Query" icon at the top of the toolbar. A new window will open displaying the two user studies matching the selected filtering criteria. As the image below shows, these studies match the ones identified in Table 16.