Data Visualization for Psychotherapy Progress Tracking

Kelly Koerner
Evidence-Based Practice Institute, LLC
206.265.2507
k.koerner@comcast.net

Dharma Dailey
University of Washington HCDE
dharma.dailey@gmail.com

Mike Lipp
University of Washington HCDE
mikeli77@uw.edu

Heidi Connor
University of Washington HCDE
hidcon@uw.edu

Rohit Sharma
University of Washington HCDE
rohitsh@uw.edu

ABSTRACT
In this experience report, we recount how we designed and built data visualization tools for clinical decision making in psychotherapy. We describe how a combination of three factors enabled us to build a high-fidelity prototype within eight-weeks: 1) a multi-disciplinary team; 2) an agile methodology that incorporated participatory user-centered research into the design approach; and 3) a coherent conceptual framework for designing data visualization for decision making [1]. Elements of our approach and the lessons learned may be useful to others who must design tools to display multivariate data for users who work under tight time constraints and high cognitive loads, and whose skills using data visualization vary widely.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces—User centered design, Graphical user interfaces (GUI); I.3.6 [ComputerGraphics]: Methodologies and Techniques—Interaction Techniques; H.4.8 [Information Systems Applications]: Types of Systems—Decision Support

Keywords
Data Visualization, Agile User Research

1. INTRODUCTION
In this case study, we describe our experience designing and building a data visualization tool for clinical decision making in psychotherapy. We describe how our multi-disciplinary team successfully combined agile methodology with participatory user-centered design to understand and address the needs of users who vary widely in their ability to use data visualizations and who make decisions under tight time constraints and high cognitive-loads. We used a conceptual framework that maps decision makers' tasks onto the data visualization tasks [1], giving us a model that conceptually directed not only development but also our strategy for evaluating the prototype's efficacy. These combined factors allowed us to design and build a working experiment-ready prototype within eight weeks.

The final data visualization prototype features two components: an individual client view (Client Dashboard, Figure 1) and a caseload view (Multi-Client Dashboard, Figure 2). Each dashboard helps the therapist rapidly interpret scores from patient reported symptom measures and compare those scores to relevant benchmarks to aid clinical decision making. In both dashboards, heads up displays and filters appear on the left side. Graphical displays of symptom monitoring for an individual patient over time (Figure 1) or a population of patients and therapists (Figure 2) appear to the right. The prototype we built is part of a larger performance monitoring and feedback system, Online Progress Tracking (*OPT-beta,* [2]), that helps mental health therapists systematically monitor the progress of patients in psychotherapy while they learn and use evidence-based mental health care practices. Monitoring patients' progress is important as it has been demonstrated that therapists who do so have improved client outcomes [3].

Figure 1. Client Dashboard

Figure 2. Multi-Client Dashboard

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SIGDOC'12, October 3–5, 2012, Seattle, Washington, USA.
Copyright 2012 ACM 978-1-4503-1497-8/12/10...$15.00.

1.1 Our Team

Our five-member team divided responsibility for the design process into primary roles with two programmers (Sharma and Lipp), one designer (Connor), one user experience researcher (Dailey) and one domain expert and data collector (Koerner). But in practice, each took on secondary roles as described below. The clear yet overlapping skills and frequent communication made for a strong agile process.

1.2 A Conceptual Model for Decision Making Guides Our Design

The purpose of our visualization design is to aid decision making. To help us think categorically about how to design for decision making, we used a heuristic framework devised by Bautista and Carenini [1] for designing visual interfaces. This framework integrates task analysis of the general decision-analysis process with task taxonomies for data visualization, making it more likely that a visualization supports the user's efforts to make good decisions from the data. It also conceptually divides decision making interactions into three separate, yet interwoven elements: construction, inspection and sensitivity analysis.

Our user research and testing focused on application of these elements as follows. In keeping with Bautista and Carenini's concept of *construction*, our user tests evaluated whether our visual displays supported users ability to quickly yet accurately *construct* data models that readily aid the most common clinical decisions. In keeping with the concept of *inspection*, we tested if the user could easily retrieve details on demand that facilitate forming and testing hypotheses relevant to their work. Along these lines, we also evaluated whether our displays reduced ambiguity and the likelihood that users would misinterpret data. Finally, in keeping with Bautista and Carenini's concept of *sensitivity analysis*, we tested if the therapist can use our tool to easily explore various "what if..?" scenarios to better understand patients' progress. Because users often quickly transition through these aspects of decision making, our design aims to support fluid movement between them. For example, if a user discovers new data that might alter their working hypothesis on the treatment course for a patient, our design should readily support a flow from inspection of current data to sensitivity analysis exploring alternative hypotheses to construction of a new working hypothesis of the best treatment options for that patient.

Following Bautista & Carenini, we then mapped key decision making tasks to specific visualization techniques. When a user selects a single patient or a group of patients, urgent status indicators are clearly distinguishable based on their color, shape and location. Preattentive attention grabbing helps users to rapidly recognize the most important information on the screen. For example, we used a red dot to signal when a patient's symptom scores indicate that they are suicidal. Once alerted to such information, we support a therapist's ability to construct decision models by employing zoom and filter interactions that afford users the ability to explore the details behind alerts.

Our visual design supports the decision making task of inspection with details-on-demand and focus-plus-context interactions. These visualization interactions help therapists rapidly form hypotheses about the client's status, find multivariate explanations, and formulate cause and effect relationships among variables over time. During inspection, data visualization must expose uncertainty. For example, when the therapist sees scores with large variability from week to week, the data visualization tool should help the therapist to discern whether the pattern observed is clinically relevant or due to normal score fluctuations. Finally, upon seeing data that informs choice, therapists will want to add or modify their objectives as well as add or modify the alternative actions about which they want information (sensitivity analysis). For example, after identifying a recent relapse in symptoms, the therapist may change the selected time period to scan across the entire course of treatment looking for other times where the patient has relapsed or explore how this patient's course in therapy compares to other patients with similar characteristics.

To adapt the conceptual decision making framework for usability testing, we designed questions directly based on each of its three components: construction, inspection, and sensitivity analysis. For construction, we tested if our tool can enable users to quickly find the individual or group of patients, or variables relevant to construct models of the most common clinical decisions. In addition, we tested if the therapist could rapidly gain needed information under strict time constraints. For inspection, we tested if the user could easily retrieve details on demand that facilitate forming and testing hypotheses. In addition, we presented tasks that include 'traps' where the therapist may be prone to misinterpret data, and tested whether the data visualization tool provided sufficient help to expose uncertainty. Finally, for sensitivity analysis, we tested whether the therapist could use our tool to easily explore various "what if..?" scenarios to better understand patients' progress. We also tested if the transition between all three phases was sufficiently fluid and if the user could change objectives upon discovering new data and easily pursue the next hypothesis.

1.3 Challenges Our Users Face

Our users face several challenges that constrained our preliminary design. First, therapists must fill each work day with as many sessions as possible, with 5-10 minutes between sessions to document the last session and then prepare for the next. Therapists work under high cognitive load as they manage multiple tasks, under time pressures, with frequent interruptions.

Second, therapists vary widely in their ability to correctly read and interpret graphically displayed data. For some, interpreting basic line graphs is challenging. For most, general concepts about reliable change are unfamiliar. For example, when asked to read patients' graphically displayed data, most therapists do not consider how clients' ratings on symptom measures can vary by chance without actually representing reliable change and end up incorrectly assuming that changes are significant when they are not. Tight workflow, high cognitive load, and low data interpretation skills can make therapists prone to dismiss emotionally challenging information. For example, therapists may be prone to incorrectly deduce that a particular patient is making progress when in fact he is not.

Finally, therapists need to quickly and accurately analyze both episodic and weekly measurements relative to benchmarks. The user needs to understand symptom severity and see patterns across different measures easily, with a single glance. The challenge is that each measure has its own scoring system, with different norms and cutoffs. This makes it difficult to display all measures on the same graph while simultaneously providing visual cues that quickly indicate clinical norms (e.g., how severe the symptom score is).

2. USABILITY RESEARCH AND DESIGN PROCESSES

2.1 An Agile Methodology

Our team employed aspects from agile software development around teamwork and rapid iterations. We averaged two design iterations per week. This quick pace provided a constant stream of incremental feedback that we used to validate our design or make adjustments to improve observed issues. In terms of teamwork, we shared information and sought out each other's input. Perhaps, most helpful was that tasks were assigned to small groups rather than individuals, and therefore every task afforded collaboration. Because our team was interdisciplinary this assured that knowledge was shared across roles.

Each design iteration consisted of these components:

1. usability research and evaluation
2. regular consultations with domain experts as needed to resolve questions
3. confirm or improve design (and prototype) based on findings.

In tandem with early design iterations, we also built use cases, a simulated data set, and selected a visualization library.

2.2 Usability Research and Evaluation

Typically, to facilitate rapid design iterations, we conducted tests and interviews with one participant at a time, enabling us to incorporate any changes into the design prototype before the next session. One of our priorities was to quickly verbally report the take-aways and subsequent designs with the entire team. Research notes and design iterations were also available to the whole team via shared online folders. This constant communication enabled us to absorb user research findings and consider design options as a team.

Usability testing included both formal and informal components. Formal components included structured usability evaluations which presented therapists with realistic tasks, such as using the graphical display to explain how well a patient is doing. Similarly, to test the intuitiveness of individual components, we asked therapists to tell us the meaning of several unlabeled design elements presented within the dashboard. These more formal evaluations were followed by informal semi-structured interviews, which involved returning to the same users with new iterations and asking for additional feedback and input. Because of the ongoing participatory nature of the testing, users could quickly tell us if our design was improving or getting worse.

2.3 Regular Consultations with Domain Experts

For this design cycle, user research was conducted with domain experts who had specific and complementary knowledge on aspects of the type of decision making we are trying to represent. For example, we conducted a user test with a domain expert who was both a therapist and a statistician, followed by a semi-structured interview. We were then able to circle back later, with specific questions about our design, through email on an ad-hoc basis. One of our team members was a domain expert herself and had access to a number of other domain experts. These experts were all trained in the best practices of progress monitoring but had various degrees of experience with the system in development. The diversity of these users gave us a nice combination of opinions and experiences to evaluate our designs. This rich pool of experts worked well with our rapid iteration technique as it provided a nearly continuous feedback into our design iterations, keeping our agile process user-centered throughout development.

2.4 Confirm or Improve Design (and Prototype) Based on Findings

User testing results either confirmed the design or led to better definition of problems with the design. Problems were identified and shared back to the full team, and all team members sketched possible solutions. Because our designer sat in on most user tests and interviews, she could rapidly translate the team's sketches into new design options. This helped build a shared understanding of general user experience and consensus on specific solutions. This process typically resolved some design questions, but led us back to user research and testing for others.

2.5 Supporting the Prototypes with Use Cases and Data Set

Concurrent with the development of the interactive prototypes, we created two use cases. These were used for formalized scenario based user testing and a cognitive walkthrough, emphasizing anticipated interactions with the two dashboards. In concert with these narratives, a sample data set was created to populate the prototypes, ensuring that key analytical and visual elements of our design would be evaluated.

2.6 Criteria and Process for Selecting a Visualization Library

To guide the selection process for a visualization library we would use to build our prototype, we established a set of criteria to describe important elements. These criteria proved valuable, serving as common points of evaluation from which we could make our final choice. The criteria consisted of the following items:

1. easy enough to learn and implement by novice programmers over a period of eight weeks, but robust enough to handle the requirements laid out in our user research
2. flexible enough to be expanded and used in an actual clinic
3. free to use for a non-profit

Utilizing the selection criteria, we conducted an informal survey of available visualization platform and narrowed our choices down to the D3 or HighStock JavaScrip libraries. While both platforms satisfied our criteria, we decided to use HighStock because of its relative ease of use, and a wide variety of sample visualizations provided, which were similar to what we intended to build.

3. LESSONS LEARNED

3.1 Displaying Multiple Measures with Different Norms is Challenging

We faced one particularly tough design challenge due to the users' need to see multiple measures on the same graph while indicating symptom severity. Each of the measures used has different norms and cut offs for mild/moderate/severe. For statistical reasons, normalizing all measures to the same scale so that the measures could be displayed against a common severity range was not possible without distorting the underlying data. Additionally, we needed to display both weekly and periodic measures. In the inspection phase, our users wanted to quickly see two types of data: the range in which the score falls and the individual question responses that make up each score. At first, we created a simple table that displayed the actual scores in each cell. This displayed the data clearly, but took up a lot of space. Additionally, our time constrained users had difficulty quickly identifying critical issues and patterns, since they had to read and interpret values from each cell individually.

To reduce the footprint required to present these two types of data and to make the data easier to scan and interpret, we decided to use a two tiered expanding/collapsing heatmap (Figures 3 and 4).

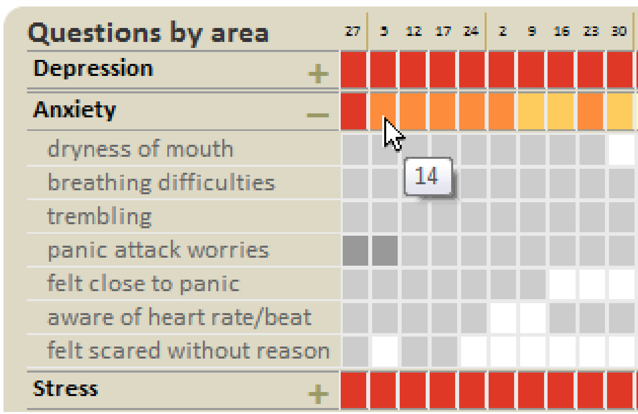

Figure 3. Heatmap expanded to show Anxiety scores

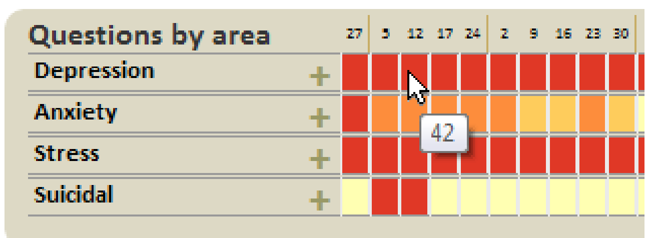

Figure 4. Collapsed heatmap showing symptom categories

The heatmap solution offered a number of improvements over the simple table including:

1. encoding numeric values into colors that could be mapped to severity and that were easier and quicker to analyze by the user
2. implementation of collapsed/expanded modes, the heatmap displays key information without cluttering the UI
3. making scores for individual questions were available as details on demand by expanding rows of the heatmap or on mouseover.

3.2 Select a Visualization Platform with an Established User Base

One thing that we did not anticipate was the time it would take for us to become familiar enough with JavaScript to utilize our selected visualization platform HighStock. To quickly learn JavaScript, CodeAcademy proved to be a great resource. We also did not realize that the number of current users of a platform is an important selection criterion. We learned that this is the case, as more users result in a richer user community and better support mechanisms. In this way, it may have ended up being easier to debug our code if it were created using D3 instead of HighStock, since the former is a more popular platform. However, there were ample resources available that enabled us to tackle each HighStock programming challenge and achieve the goals laid out by our design.

3.3 Simulating Data is Harder than You'd Think

We created a dataset that would allow us to test key analytical and visual elements of our design. Data from real patients were used (with personal details removed) as the basis for the creation of fictitious cases. Cases were then replicated for a final 86 x 1881 matrix of 30 therapists and 200 clients assessed with multiple measures and seen for differing lengths of treatment. The end result accurately represented the distribution of treatment response rates and typical missing data rates. What proved difficult, however, was that when therapist users explore, they very rapidly begin moving between construction, inspection and sensitivity analysis, and expect the data to yield a deep and coherent backstory for any case they choose to examine. For these users, realistic simulation is critical to creating a high-fidelity interactive prototype. As we attempted to build out our data set, we realized how difficult and time consuming it is to create a large simulated multivariate data set.

3.4 Users Need a High-Fidelity Prototype to Conceive of How to Use Big Data

Even though the users we worked with when developing the visualization prototypes were all domain experts in progress monitoring, data visualization of big data is new to them. They had a hard time conceiving of what they would want to do if they could explore hundreds or thousands of cases of multivariate data. Low fidelity prototypes did not provide enough interactivity for them to imagine what they might be able to do with such data. It became clear that users require a high-fidelity prototype with realistic data to explore before they begin to understand what is possible. We expect it will take several more iterations with high-fidelity prototypes before we can confidently understand users' needs when exploring big multivariate data for clinical decision making.

3.5 Agile User Research and Rapid Prototyping Worked Well for Us

We had eight weeks to create a functioning prototype and each of us had limited hours per week for project development. We balanced the workload across the team on the front end by learning each other's schedules so that we timed the handoff of tasks to team mates' availability. Individual team member's responsibilities were clear, but flexible. We informally defined and refined goals and milestones by consensus, with each team member taking on some of the project management work. Where possible we shared functionality. For example, the designer was present for all user research interviews and the developers contributed drawings based on user research during the ideation phase. The incremental and participatory user-centered design approach resulted in a nearly continuous pipeline of feedback from users and domain experts. The agile team work, with multiple disciplines and overlapping responsibilities let us maintain a rapid pace and come in with a working high fidelity prototype on deadline.

In retrospect, we could have benefited from more structured project management and monitoring of schedules and progress. Half way through the project we ran into some challenges including those mentioned in previous sections. Our team was able to leverage our experience by making good decisions on how to spend our remaining time, but we were still rushed at the end and making last minute adjustments and finishing touches. Paying closer attention to actual progress and schedules may have allowed us to recognize the impact of unforeseen challenges and recognize we would need to adjust future commitments in order to stay on track with the originally scheduled delivery date.

4. CONCLUSION

It was a rewarding experience to collaborate as a multi-disciplinary team and, in eight weeks, successfully design and build a data visualization prototype for clinical decision making. A key learning from this experience was that it is possible to adapt a particular framework for understanding decision making - the conceptual decision making framework of Bautista & Carenini - in order to design a visualization tool that supports better decision making. This framework usefully focused our initial design and user research, and provided a heuristic standard against which to test design decisions—all elements should strengthen the user's ability to complete decision tasks and result in a smooth flow between construction, inspection, and sensitivity analysis phases of the decision making process. Our work now is to assist the Evidence-Based Practice Institute in experimentally evaluating whether the prototype works as intended. We will be releasing our code publicly via GitHub and then assisting the team at the EBPI by sharing our knowledge and helping them to deploy the tool.

5. REFERENCES

[1] Bautista, J., & Carenini, G. (2006) An Integrated Task-Based Framework for the Design and Evaluation of Visualizations to Support Preferential Choice. AVI '06, May 23–26, 2006, Venezia, Italy. http://www.cs.ubc.ca/~carenini/PAPERS/AVI06.pdf

[2] Koerner, K. & Persons, J. P. (2010-12) National Institute of Mental Health, 2010. Project title: Online Platform to Help Mental Health Practitioners Implement EBPs (1R43MH093993 – 01).

[3] Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298-311.

6. ACKNOWLEDGMENTS

This research was partially supported by National Institute of Mental Health (1R43MH093993-01) to Kelly Koerner and Jacqueline B. Persons. We gratefully acknowledge Cecilia Aragon, PhD, for her mentorship and expertise during this project.