Cogsci 220: Information Visualization

Jim Hollan
Design Lab
Department of Cognitive Science
Department of Computer Science & Engineering

Email: hollan@ucsd.edu
Web: hci.ucsd.edu/hollan & designlab.ucsd.edu
Welcome to week four. 
Hope everyone continues to be safe in these challenging times.

Be sure you are keeping up with readings and going through the UW notebook introduction to Vega-Lite. There are also nice introductions to Observable notebooks and JavaScript.

Today:
Main goal today is to ensure project groups are formed and your initial topic is decided

We will start with hearing from your groups

Briefly talk about writing and especially about writing proposals

End with an interview with this week’s Researcher of Week: Arvind Satyanarayan

Your group should get together soon to coordinate and prepare a note to sent to 220-g@ucsd.edu by Friday evening
What is the point of a thesis?

What makes a great thesis?

How do you write a great thesis?
What, Why, and How of Visualization

What is shown?
Data abstraction

Why is the user looking at it?
Task abstraction
Increasingly important to look beyond task to encompassing activities.

How is it shown?
Idiom: visual encoding and interaction
Goal: Reduce Task and Activity Complexity

Don’t just draw what you’re given

- Decide what the right thing to show is
- Create it with a series of transformations from the original dataset
- Draw that “right thing”

Increasing need to take context into account, especially as part of interactivity and dynamic content.

Original Data

Derived Data

trade balance = exports – imports
Lu, Casey, and I are interested in working on visualizations related to data privacy. We're envisioning a tool that could track the kind of data that you share with different applications over time. You could use different lenses to focus on specific kinds of data; for example, you might be interested in specifically looking at location data that you have shared over time.
We're also interested in finding a way to link your data as an individual to the broader collective issue of online data collection. For example, this visualization from the anti-eviction mapping project gives the viewer an idea of the "big picture" of the problem of evictions in the SF area. However, you also have the ability to "zoom in" and click on individual points on the map to hear individual stories. We'd like to do something similar to connect your privacy as an individual to the broader landscape of mass data collection.
Mary Anne Smart, Casey Meehan, Lu Sun
Martha, Michael, and I are interested in representation and behavior of information entities through interactive interfaces. We want to provide customized representations of data narrative articles. These will give more context to potentially confusing or specific information on a website.
For example, the New York Times "Upshot" articles have interactive data visualizations and websites like these could be expanded on and tailored for readers based on their background knowledge and interests they provide. Potential mediums could be a Jupyter notebook or desktop plugin or website but we will figure that out as we go!
Matin, Tommy, Naba and I are interested in exploring the visualization of podcast audio for efficient navigation and information retrieval.

Podcast is a common source of information with many genres. People use podcasts while doing other activities, such as exercising. Cognitive load theory suggests that multi tasking diminishes understanding and recall of the content. As such, people tend to re-listen to specific bits of podcasts afterwards, such as specific comedic or informative bits.

However, navigating through lengthy auditory pieces is challenging and time-consuming. Thus, it is important to visually represent the audio information for efficient information retrieval.
In this project, we aim to address the navigability of podcasts via a personalized and multi-layer visual representation: user's interaction with the podcast (e.g., skipping ahead for 15 sec) influences the weight of bits of the audio that the user might be more interested in reviewing. The multilayer structure starts from coarse content (i.e., words) to more granular phrases.

Some References:

CrossCast, UIST '20: https://urldefense.com/v3/__https://dl.acm.org/doi/10.1145/3379337.3415882__;!!Mih3wA!RwPOrUI5RGVX_ZdfhVS4AlMM2ByM5BE0gdR_lAvsPzG3-Z1UVi_vlxhV_HZxSA$

TalkTraces, 'CHI '19: https://urldefense.com/v3/__https://doi.org/10.1145/3290605.3300807__;!!Mih3wA!RwPOrUI5RGVX_ZdfhVS4AlMM2ByM5BE0gdR_lAvsPzG3-Z1UVi_vlxixEm-DRw$

Quality'Alone'Time through Conversations and Storytelling: Podcast Listening Behaviors and Routines, https://urldefense.com/v3/__https://dl.acm.org/doi/abs/10.20380/Gl2018.11__;!!Mih3wA!RwPOrUI5RGVX_ZdfhVS4AlMM2ByM5BE0gdR_lAvsPzG3-Z1UVi_vlxgB-i0unA$
We interested in exploring ways of representing activity histories particularly focusing on a more intuitive organization of tabs. The core idea being to cluster tabs into related ideas or categories. And perhaps automating that feature when examining a browser history. Let us know if you would like to join us!
Sloan Hill-Lindsay, Garrett Wolfe, Xuhan Yang
Brinley Stringer, Talia Latona-Tequida, Isabel White, Kristin Tenney

We will be focusing on behaviors within the information space of Teacher Desmos. We will be exploring ways that behaviors can be changed so that it is more intuitive for teachers to find relevant activities for their math classrooms, as well as the relevant research that supports the use of those activities so that teachers have a one stop shop where they can find good resources and information to feel confident about the practices they implement in their classrooms. There is also potential to explore activity history within Teacher Desmos as well as lenses… we will narrow that down as we get further along! The emphasis will be on behaviors though.
Brinley Stringer, Talia Latona-Tequida, Isabel White, Kristin Tenney
Proposal vs Paper

Would you be able to clarify the scope of the project topics? We are not sure if our project should align with a paper publication (narrow topic followed by a concrete prototype) or a grant proposal (broad topic followed by multiple research directions with viable prototypes).

For instance, a narrow topic of our group is to visualize podcasts personalized by user's interactions (e.g., skip ahead for 15 sec). A broader topic would be to visualize all auditory content (including conversations).

Best wishes,
Matin
Why a proposal rather than a paper?

What your group should produce for the course is a draft of a proposal. Of course the more finished it is the better. Writing and research projects are never finished. One just runs out of time.

It may well include a description of a running implementation that could itself be the focus of a paper. A proposal fits better because you will likely not be finished nor have time to do an evaluation but you can (and it is important to) include more about what you didn’t have time to do as well as a plan for evaluation.

Your ideas are likely to be much more ambitious than you have time to explore. A proposal will allow you to include them.

It is very useful to get experience with and develop expertise in proposal writing.
Donald Schön: Reflective conversation with materials

Rewriting and Iterations
Moving from Writer-Centric -> Reader-Centric

Takes Time; Learning Genre; Different Levels; Trip Metaphor; Style of Work; Learn to share early drafts and structure feedback; Tools (Dave Wroblewski);

Writing from the Inside Out
  What are the main points you want to make?
  What does the reader need to know to understand those points?

Backward-Outline
Proposal Evolution

Ideas to Ideas Informed by Literature to Draft Proposal to Many Iterations of Draft to Final Proposal

Writing always needs to move from something to help you think to something that communicates to others

Initial Focus
  Clear formulation of question you are addressing and why it is important
  Informing the question based on the literature

Later Focus
  Specifics of how you will address the question and
  Why your approach is promising. An initial implementation or preliminary data strengthens.
Proposal Writing

Writing about what you plan to do rather than what you’ve done is different and challenging.

Avoid vagueness and handwaving
If you don’t know how you are going to accomplish some aspect of the project, describe how you are going to find out, describe alternative approaches, and how you will make a decision. Similar to talking about possible outcomes of experiments.

Get feedback from others
Difficulty of allowing others to see early drafts
Structure your interactions to get the right level of comments.
Proposal Heuristics

Look at instructions reviewers are given. Make their job easy.

Summaries are particularly important

Show you really know the literature (sometimes via the appropriateness of citations)

Placement of related work

In information visualization area, one challenge is how to depict dynamic interactive systems. Unlike for NSF your proposal can have links dynamic depictions or might even be authored as a computational notebook or website.

List possible criticisms and how you respond to them
Proposal Heuristics

Don’t wait for funding, get started on the project, data and early prototypes will make proposal much stronger

Know the genre of writing for funding agency

Get to know and consult program directors; brief concept papers

Participate on review panels

Appreciate admin and contracts & grants staff

Involve your students and others
What is the question?
Why is the question important?
Who cares? If you’re successful, what difference will it make?
What are you trying to do?
Articulate your objectives without jargon
How is it done today?
What are limits of current practice? What is your approach?
How will you determine if you are successful?
What are the risks and payoffs?
How much will it cost and how long will it take? Importance of balanced portfolio of research projects.
What are midterm and final “exams” to check for success?
Proposals

Program Solicitation
Due Dates: department, contracts and grants, agency

Formal Proposal

Merit Review
(1) the intellectual merit of the proposed activity
(2) the broader impacts resulting from the proposed activity

Get involved in this process
Get to know program directors

Decision
At least 6 months and often longer

Faculty Early Career Development Program

Office of Advanced Cyberinfrastructure (OAC)
Computing and Communication Foundations (CCF)
Computer and Network Systems (CNS)
Information & Intelligent Systems (IIS)
<table>
<thead>
<tr>
<th>Org (Drill to Next Level)</th>
<th>FY</th>
<th>Number of Proposals</th>
<th>Number of Awards</th>
<th>Funding Rate</th>
<th>Average Decision Time (months)</th>
<th>Mean Award Duration (years)</th>
<th>Median Annual Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NSF</strong></td>
<td>2019</td>
<td>41,032</td>
<td>11,252</td>
<td>27%</td>
<td>5.76</td>
<td>2.67</td>
<td>$134,999</td>
</tr>
<tr>
<td></td>
<td>2018</td>
<td>48,334</td>
<td>11,716</td>
<td>24%</td>
<td>5.63</td>
<td>2.74</td>
<td>$130,000</td>
</tr>
<tr>
<td></td>
<td>2017</td>
<td>49,425</td>
<td>11,457</td>
<td>23%</td>
<td>5.53</td>
<td>2.79</td>
<td>$111,402</td>
</tr>
<tr>
<td></td>
<td>2016</td>
<td>49,307</td>
<td>11,894</td>
<td>24%</td>
<td>5.39</td>
<td>3.10</td>
<td>$100,000</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>49,631</td>
<td>12,017</td>
<td>24%</td>
<td>5.75</td>
<td>3.24</td>
<td>$92,095</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>48,200</td>
<td>11,115</td>
<td>23%</td>
<td>5.75</td>
<td>3.35</td>
<td>$90,607</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>49,152</td>
<td>10,983</td>
<td>22%</td>
<td>5.77</td>
<td>3.37</td>
<td>$88,213</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>48,716</td>
<td>11,627</td>
<td>24%</td>
<td>5.64</td>
<td>3.34</td>
<td>$84,247</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>51,568</td>
<td>11,201</td>
<td>22%</td>
<td>5.59</td>
<td>3.37</td>
<td>$81,463</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>55,559</td>
<td>13,015</td>
<td>23%</td>
<td>5.58</td>
<td>3.34</td>
<td>$85,232</td>
</tr>
<tr>
<td><strong>BIO</strong></td>
<td>2019</td>
<td>3,115</td>
<td>1,051</td>
<td>34%</td>
<td>5.50</td>
<td>2.99</td>
<td>$175,108</td>
</tr>
<tr>
<td></td>
<td>2018</td>
<td>4,767</td>
<td>1,192</td>
<td>25%</td>
<td>6.09</td>
<td>2.96</td>
<td>$166,460</td>
</tr>
<tr>
<td></td>
<td>2017</td>
<td>5,010</td>
<td>1,147</td>
<td>23%</td>
<td>5.71</td>
<td>2.99</td>
<td>$136,922</td>
</tr>
<tr>
<td></td>
<td>2016</td>
<td>5,211</td>
<td>1,335</td>
<td>26%</td>
<td>5.33</td>
<td>3.28</td>
<td>$125,767</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>5,123</td>
<td>1,383</td>
<td>27%</td>
<td>5.39</td>
<td>3.44</td>
<td>$111,927</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>4,792</td>
<td>1,280</td>
<td>27%</td>
<td>5.41</td>
<td>3.65</td>
<td>$108,679</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>5,938</td>
<td>1,254</td>
<td>21%</td>
<td>5.11</td>
<td>3.65</td>
<td>$102,941</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>5,272</td>
<td>1,296</td>
<td>25%</td>
<td>5.27</td>
<td>3.51</td>
<td>$98,256</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>7,439</td>
<td>1,310</td>
<td>18%</td>
<td>5.53</td>
<td>3.65</td>
<td>$108,445</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>8,056</td>
<td>1,553</td>
<td>19%</td>
<td>5.37</td>
<td>3.65</td>
<td>$100,790</td>
</tr>
<tr>
<td><strong>CSE</strong></td>
<td>2019</td>
<td>8,616</td>
<td>2,009</td>
<td>23%</td>
<td>5.73</td>
<td>2.66</td>
<td>$158,728</td>
</tr>
<tr>
<td></td>
<td>2018</td>
<td>9,149</td>
<td>2,097</td>
<td>23%</td>
<td>5.72</td>
<td>2.74</td>
<td>$152,809</td>
</tr>
<tr>
<td></td>
<td>2017</td>
<td>8,769</td>
<td>1,865</td>
<td>21%</td>
<td>5.47</td>
<td>2.86</td>
<td>$133,455</td>
</tr>
<tr>
<td></td>
<td>2016</td>
<td>8,320</td>
<td>1,938</td>
<td>23%</td>
<td>5.23</td>
<td>3.16</td>
<td>$112,500</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>8,076</td>
<td>1,925</td>
<td>24%</td>
<td>5.01</td>
<td>3.35</td>
<td>$100,000</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>7,456</td>
<td>1,708</td>
<td>23%</td>
<td>4.85</td>
<td>3.55</td>
<td>$104,252</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>7,838</td>
<td>1,633</td>
<td>21%</td>
<td>5.61</td>
<td>3.62</td>
<td>$100,000</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>7,695</td>
<td>1,741</td>
<td>23%</td>
<td>5.44</td>
<td>3.52</td>
<td>$100,367</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>6,714</td>
<td>1,541</td>
<td>23%</td>
<td>5.20</td>
<td>3.58</td>
<td>$96,618</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>7,045</td>
<td>1,753</td>
<td>24%</td>
<td>5.46</td>
<td>3.53</td>
<td>$92,973</td>
</tr>
</tbody>
</table>
Proposal Review

When evaluating NSF proposals, reviewers will be asked to consider what the proposers want to do, why they want to do it, how they plan to do it, how they will know if they succeed, and what benefits could accrue if the project is successful.

These issues apply both to the technical aspects of the proposal and the way in which the project may make broader contributions. To that end, reviewers will be asked to evaluate all proposals against two criteria:

**Intellectual Merit:** The Intellectual Merit criterion encompasses the potential to advance knowledge.

**Broader Impacts:** The Broader Impacts criterion encompasses the potential to benefit society and contribute to the achievement of specific, desired societal outcomes.
Unprecedented scales of data and previously unimaginable computing power are not only transforming science and engineering but also the data-driven information systems that now shape our world. Designing these increasingly intelligent systems requires both technological sophistication and understanding the complex socio-technical environments in which they operate. One educational consequence is that students in every STEM discipline now require training in both data science and design of systems involving people. In response to this and to help realize NSF’s vision of harnessing the data revolution, faculty from two new centers at UC San Diego – the Haliciog˘lu Data Science Institute and the Design Lab – have come together around a theme of people-centered data science and data-driven design with the shared goal of training a new generation of graduate students in the theory and research methods required to create effective, ethical, and humane systems. Although open to all Ph.D. students across social sciences, physical sciences, and engineering, the primary focus will be on students in a new Ph.D. Design Specialization. We estimate there will be 15 students in each cohort, a total of 75 over the five years of the program, and 30 will receive NRT funded traineeships.
Summary

Intellectual Merit

We propose to develop, evaluate, and evolve a new module-based problem-centered approach to data science and design education that will enrich existing courses, provide a mechanism to integrate new topics, be synergistic with on-going research efforts across the university, and be shared with the wider community via MOOCs and online resources. The goal of the Halicioglu Data Science Institute (HDSI) is to advance research and training in the new field of data science. The goal of the UC San Diego Design Lab is to provide a research and educational foundation for understanding and designing complex socio-technical systems. Just as the Qualcomm Institute, the UC San Diego campus of the California Institute for Telecommunications and Information Technology (Calit2), has catalyzed transdisciplinary research across the campus by focusing on problems rather than disciplines, we propose to catalyze graduate education and transdisciplinary research by focusing on the exigent combination of data science and design of complex socio-technical systems.

Overview

Unprecedented scales of data and previously unimaginable computing power are not only transforming science and engineering but also the data-driven information systems that now shape our world. Designing these increasingly intelligent systems requires both technological sophistication and understanding the complex socio-technical structures in which they operate. One educational consequence is that students in every STEM discipline now require training in both data science and design of systems involving people. In response to this and to help realize NSF’s vision of ‘harmonizing the data revolution,’ faculty from two new centers at UC San Diego—the Halicioglu Data Science Institute and the Design Lab—have come together around a theme of people-centered data science and data-driven design with the shared goal of training a new generation of graduate students in the theory and research methods required to create effective, ethical, and humane systems. Although open to all Ph.D. students across social sciences, physical sciences, and engineering, the primary focus will be on students in a new Ph.D. Design Specialization. We estimate there will be 15 students in each cohort, a total of 75 over the five years of the program, and 30 will receive NRT-funded traineeships.

Intellectual Merit

We propose to develop, evaluate, and evolve a new module-based problem-centered approach to data science and design education that will enrich existing courses, provide a mechanism to integrate new topics, be synergistic with on-going research efforts across the university, and be shared with the wider community via MOOCs and online resources. The goal of the Halicioglu Data Science Institute (HDSI) is to advance research and training in the new field of data science. The goal of the UC San Diego Design Lab is to provide a research and educational foundation for understanding and designing complex socio-technical systems. Just as the Qualcomm Institute, the UC San Diego campus of the California Institute for Telecommunications and Information Technology (Calit2), has catalyzed transdisciplinary research across the campus by focusing on problems rather than disciplines, we propose to catalyze graduate education and transdisciplinary research by focusing on the exigent combination of data science and design of complex socio-technical systems.

Broader Impacts

The broader impacts of this effort derive from (1) strong UC San Diego commitments, exemplified by providing seven fellowships to the NRT program to assist in attracting diverse cohorts of trainees and by helping to ensure sustainability beyond the award period by committing 15 tenure track faculty positions for the Halicioglu Data Science Institute and UC San Diego Design Lab, (2) established industrial liaison programs of HDSI and Design Lab to foster partnerships with companies, nonprofits, and government organizations, which are critical career pipelines for NRT trainees as well as avenues for future program support, (3) integration of the NRT program with a new Ph.D. Design Specialization, and (4) the potential to exploit curriculum materials in a new graduate degree program in Computational Social Sciences, in existing courses outside the NRT program and Design Specialization, and more widely by availability online and in MOOCs.

Keywords: data-driven design, people-centered data science, problem-centered education, socio-technical systems

1In the University of California system, a specialization is similar to an emphasis for undergraduate students but for graduate students. Students who complete the specialization receive a Ph.D. in the major discipline augmented to indicate the specialization, e.g. a Ph.D. in Computer Science with a Specialization in Design.
Summary

Broader Impacts

The broader impacts of this effort derive from (1) strong UC San Diego commitments, exemplified by providing seven fellowships to the NRT program to assist in attracting diverse cohorts of trainees and by helping to ensure sustainability beyond the award period by committing 15 tenure track faculty positions for the Halicioglu Data Science Institute and UC San Diego Design Lab, (2) established industrial liaison programs of HDSI and Design Lab to foster partnerships with companies, nonprofits, and government organizations, which are critical career pipelines for NRT trainees as well as avenues for future program support, (3) integration of the NRT program with a new Ph.D. Design Specialization, and (4) the potential to exploit curriculum materials in a new graduate degree program in Computational Social Sciences, in existing courses outside the NRT program and Design Specialization, and more widely by availability online and in MOOCs.
Proposal Review

The following elements should be considered in the review for both criteria:

1. What is the potential for the proposed activity to:
   a. Advance knowledge and understanding within its own field or across different fields (Intellectual Merit); and
   b. Benefit society or advance desired societal outcomes (Broader Impacts)?

2. To what extent do the proposed activities suggest and explore creative, original, or potentially transformative concepts?

3. Is the plan for carrying out the proposed activities well-reasoned, well-organized, and based on a sound rationale? Does the plan incorporate a mechanism to assess success?

4. How well qualified is the individual, team, or organization to conduct the proposed activities?

5. Are there adequate resources available to the PI (either at the home organization or through collaborations) to carry out the proposed activities?
Brief Group Coordination Meeting

By Friday evening, send a note to 220-g@ucsd.edu

1. Describe and motivate your project focus. Why is it interesting and important?

2. Detail the central question(s) you are addressing. Not only what the questions are but how you plan to address them.

3. Provide a timeline for the rest of the quarter. Convince us (and yourselves) your plan is doable. Be as specific as you can. Of course your plan will change.

4. Suggest a couple of papers (2 or 3; coordinate with me) relevant to your project for us all to read. These should be selected to provide background and help us help you with your project. You will lead discussion of these papers in class. Provide links to the papers.