Cogsci220: 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

Cogsci220: Information Visualization

Welcome to the class and a new year.

We need to acknowledge that this is an incredibly challenging, uncertain, and deeply troubling time.

We are still in the midst of a global pandemic.

Mounting deaths total over 350,000 in the US and approach 2 million worldwide.

The political situation is highly polarized.

We are grappling with monumental changes in our lives at work, at home, and in our communities.

These are not normal times, and we should not pretend they are.

Cogsci220: Information Visualization

In this challenging and unprecedented time, we all need to support each other.
The highest priority is for everyone to remain safe and healthy.

We are all displaced from our normal context and no one knows the time course of this pandemic. Everyone is stressed and each of us is adjusting our lives to respond to complex rapidly changing situations.

My goal in the course is to be flexible and response to issues and problems that may arise. Communication is central to provide the feedback required to make adjustments, minimize stress, and make the best of this challenging situation.

This course has always been an interactive seminar.

The hope is the course will be a respite from the pandemic, an opportunity to think about research ideas, be creative, and develop useful visualization skills.

Before we begin

The class is too large for an interactive seminar class.

Do consider dropping the class. It involves considerable reading, writing, active participation in class, and the challenge of developing an NSF style research proposal with visualization demonstrations.

You will work in a project team of three or four students. Group work is challenging. It can be delightful but also can be maddening. One possible reason to drop the course is you can't find the right people to work with or a group that shares a common project focus you find of interest. Another is your situation during the pandemic. Consider taking the course at another time when things are more normal.

As with the start of every class (and especially appropriate at the start of a new year), I will begin with some advice.

Advice for Undergrads

Why do so few people make significant contributions?

What is the difference between those who make significant impact and those who don't?

- One factor is expectations: If you think you can't almost certainly you won't
- Another is that to do significant things you have to neglect other things

Be careful about your commitments but when you commit really commit

Be thoughtful about who you spend time with

Take time to think important thoughts

You can't win the battle every time but try hard not to let the urgent drive out the important

Advice for Grad Students

I make no claim to originality, only to good taste in borrowing excellent advice from others.

Special appreciation for wisdom from Richard Hamming, Ivan Sutherland, and Dave Patterson, from my personal mentors Pat Suppes, Dave Rumelhart, and Don Norman, and especially from my students and collaborators from whom I have learned much.

Why listen to me? I have over four decades of experience in both university and industrial research labs, honored with the ACM SIGCHI Lifetime Research Award, promoted to Distinguished Professor, and have collaborated with amazing researchers (including four MacArthur Fellows and six Rumelhart Prize winners), and mentored many young faculty, grad students, and postdocs.

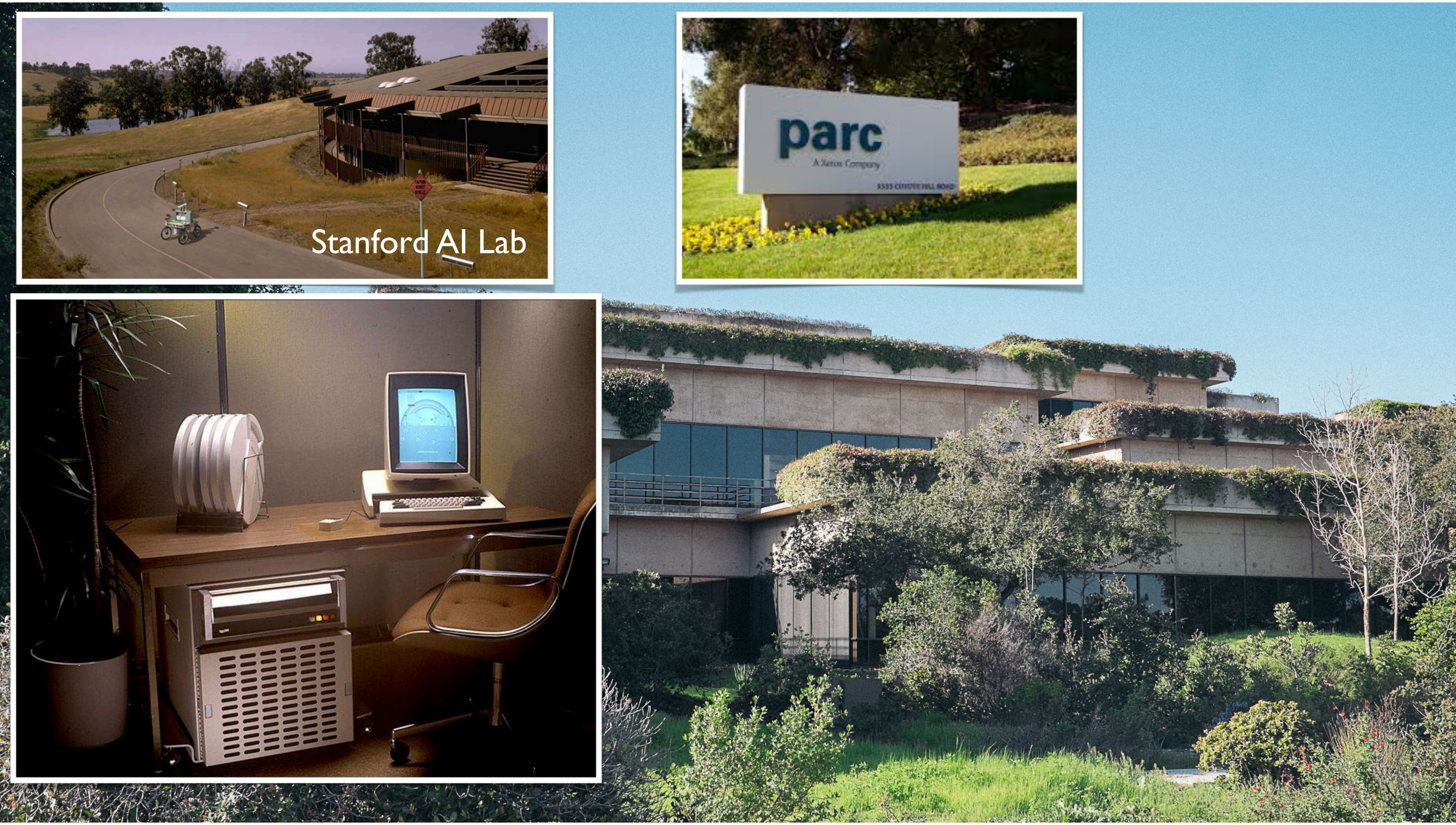
Advice for Grad Students

Over a research career one is fortunate to witness first-hand one paradigm shift.

I have been extremely fortune to witness four.

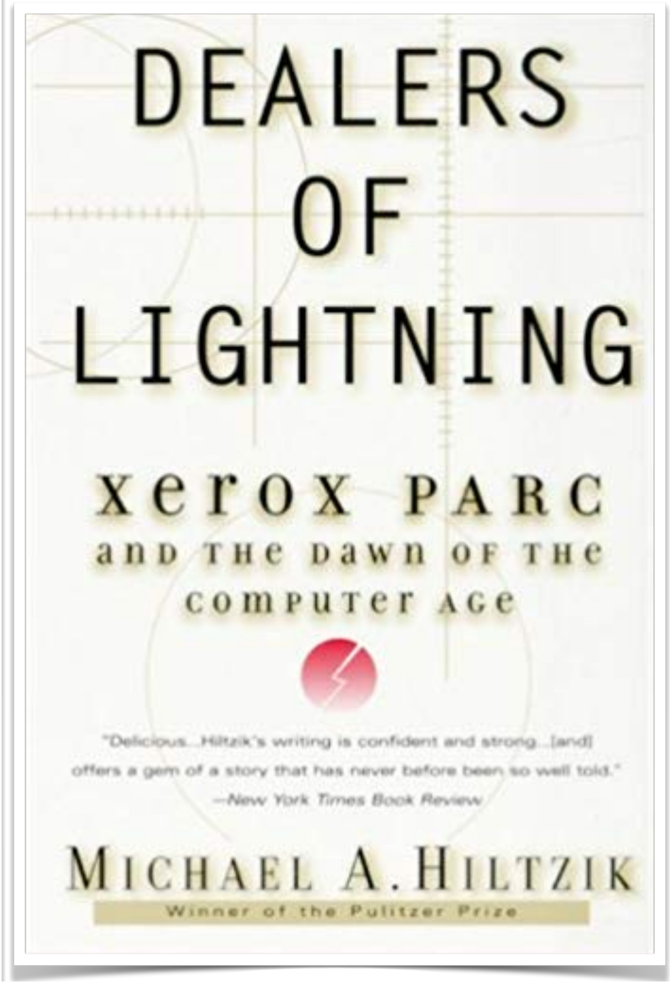
The first two were in computation: the exciting beginnings of personal computing and the Internet.

I was a postdoc in the AI Lab at Stanford when I first sat down at an Alto at Xerox Parc connected to the Internet. This was profoundly influencing.


Stanford AI Lab
parc
A Xerox Company

Xerox Parc and Alto

Chuck Thacker and colleagues at Parc develop the Alto in 1973. Parc approach: hire the best researchers and leave them unburdened by directives, instructions, or deadlines.



Paradigm Shifts

Just as important for me as personal computing and the Internet paradigm shifts was arriving at UC San Diego and being in the midst of **a third paradigm shift: a cognitive revolution** was just emerging as part of a rapidly developing zeitgeist shared across the social sciences that has radically transformed our view of cognition and created the discipline of cognitive science.

A fourth paradigm shift didn't arrive until the 90's with t**he marvelous explosion of the web and all that it has enabled for both good and for ill**, such as social networks, collaborative filtering, instantaneous access to massive ever expanding information resources and to previously unimaginable computational power, and an ecosystem driven primarily by advertising.

As just one example of this shift, consider that already today there has been over a billion Google searches and each search took about the equivalent computing to answer as all the computing done – in flight and on the ground – for the Apollo program that landed the first humans on the Moon.

These paradigm shifts have fundamentally changed the way we interact, communicate, and think.

Even with the current pandemic and political situation, and the ominous consequences of climate change, it still is an absolutely amazing time. It is a good time to be in graduate school. I am envious.

How to succeed in graduate school

The short answer: **Don't think of yourself as being a grad student**

Of course much to learn but grad education is unlike undergrad education. The most valuable skills from undergraduate days are probably learning to write well, to work hard, and to multitask.

Courses are important but the real challenge is you must deeply understand the research literature of your area so you can begin to contribute to it. The literature is massive and only a tiny fraction of it will be covered in courses. In addition, it is expanding at an ever accelerating rate.

Your immediate and continuing tasks are to: (1) develop a portfolio of research questions, and (2) master the skills needed to pursue and address the questions you care about.

Grad school at its best is an apprenticeship and the beginning of many research collaborations.

The good news (or perhaps the bad news) is: during graduate school you will have more discretionary and flexible time than you will ever have again.

Transition from Student to Researcher is Difficult

In classes, the world is rigged. There's a simple correct answer and the problem and activities are structured to let you come to that answer. You even get feedback and the correct answer soon after you submit your response.

Research is different

No one tells you the right answer. We don't often even know if there is a right answer. Frequently what you find out doesn't answer the question but only leads to more questions. Sometimes what you find out is that you are asking the wrong question.

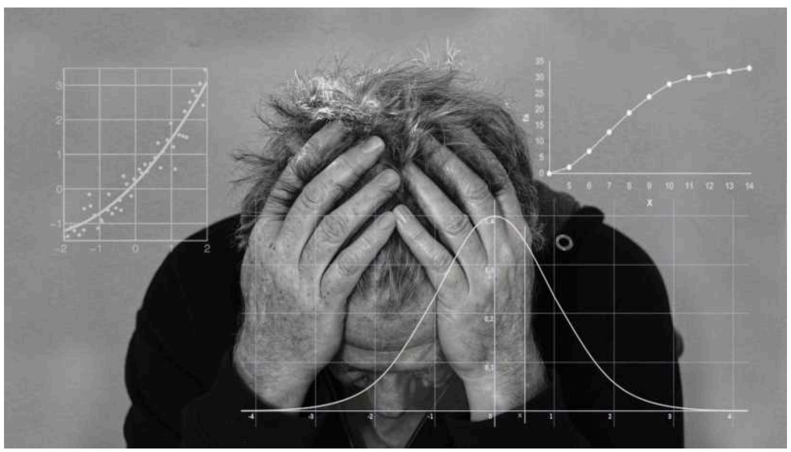
Skills associated with being a stellar student may even impede your becoming a stellar researcher

Can't depend on learning everything in courses. You often need to **learn just enough** about X so you can **make research progress.** Every topic is infinitely rich but time is limited.

Advice for Grad Students

Research Requires Courage Because It Involves Risks

- Social and emotional
- Risks to reputation and pride
- High probability any particular research project, especially if challenging, will fail



"Dealing with failure is easy: Work hard to improve."

"Success is also easy to handle: You've solved the wrong problem. Work hard to improve."

Alan Perlis

Advice for Grad Students

Reputations start early

Learn how to balance multiple things

To enjoy research you really need to accept always present ambiguity

Be smart in using your time

Get to know and be known by people in your field (email, conferences, talks, visits, …)

Advice for Grad Students

Develop a Research Portfolio

Portfolio will continue to develop and evolve over your whole career

Your time is the investment currency

Work on important problems

What makes a problem personally important? You have a way to attack it.

Portfolio should be a mix of differing risk/payoff projects.

Need to continually assess and adjust the mix.

Advice for Grad Students

Goal is to have impact

Solve problems people care about

Impactful projects are likely longer term projects

Need to communicate your ideas and work to others

Not only publish but people must read

As much (or more) time on communicating as on work itself

<u>Know the literature!</u> Ideas have histories

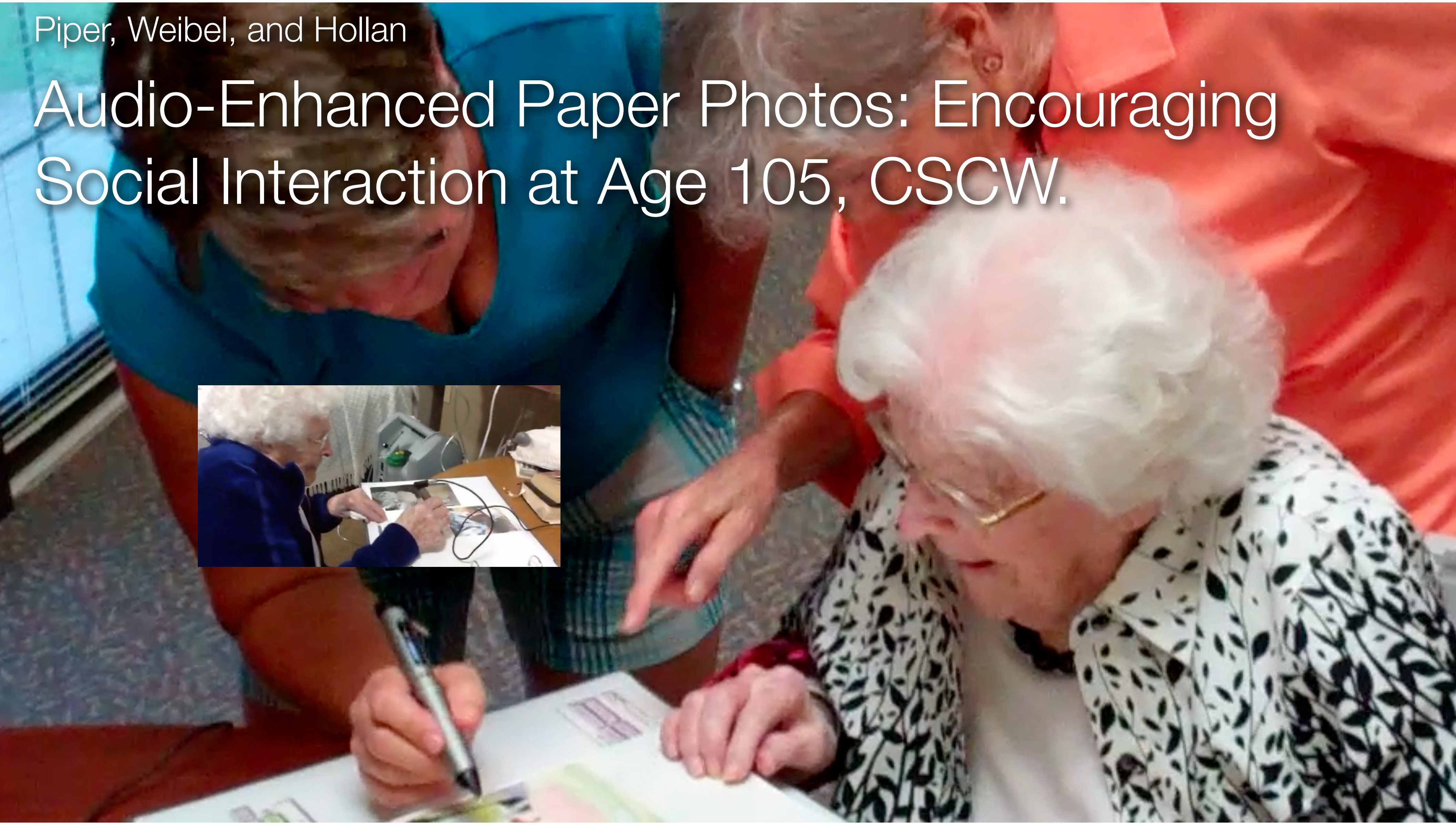
Feedback is key: seek out and value critics

Your students and other collaborators are a key part of your impact and the real multiplier of your impact

Develop taste and learn to be opportunistic in selecting research problems. An example:

Piper, Weibel, and Hollan

Audio-Enhanced Paper Photos: Encouraging Social Interaction at Age 105, CSCW.



Challenges

Science is extremely competitive

Have to work hard

Need to avoid burning out

And again: One of the biggest challenges is learning to deal with ambiguity

Build a repertoire of research skills. Don't avoid those skills you know you need but wish you could avoid. Easy to always do more of what you are already good at doing. [Jay]

Heilmeier Catechism

- George Heilmeier
- 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 promising about your approach?
- How will you determine if you are successful?
- What are the risks and payoffs?
    - How much will it cost (especially in terms of your time) and how long will it take?
    - What are midterm and final “exams” to check for success?

Science is a Social Activity

Doing

Collaborate, collaborate, collaborate

Talking

Importance of both formal and informal interactions
(Hinton, DER)

From water cooler chats to lab meetings to classes/seminars to conference talks

Approach TAing and teaching as valuable practice opportunities

Seek out opportunities to give talks (both formal and informal)

Writing

A constant activity and continual developing skill (Knuth, Don, Jeff)

Set aside regular time to write

A new challenge is learning how to write proposals

Science is About Stories

Peter Medawar on science: ***"It begins as a story about a Possible World—a story which we invent and criticize and modify as we go along, so that it ends by being, as nearly as we can make it, a story about real life."***

One of the wonderful things about areas such as HCI, Design, and Visualization is that we can make some of our stories true by writing code and building things.

For me doing science is all about stories and people.

Especially important and rarely appreciated are those stories we use to convince ourselves to invest our time and energy in projects long before we have sufficient data, often before we have any data at all.

Cogsci220: Information Visualization

Today an overview of the course

Website: hci.ucsd.edu/220

Introductions:

Amy Rae Fox and Zhutian Chen

It is always challenging to get to know others in a class and even more so in a virtual class. To start the process of getting to know each other please fill out this Google form https://forms.gle/oxvon2JcWEqZAFsB6 by Friday (1/8). We may also try an experiment next week with some experimental software.

COGSCI 220: Information Visualization (Winter 2021)

INSTRUCTOR
Jim Hollan
By appointment

POSTDOC
Zhutian Chen

GRAD STUDENT
Amy Rae Fox

LECTURE
Location: Zoom
Time: Tues 9:00-11:50

ABOUT

TEXTBOOKS

SCHEDULE
1 2 3 4 5 6 7 8 9 10

What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. -- Herb Simon

A graphic is never an end in itself; it is a moment in the process of decision making.-- Jacques Bertin

This seminar provides a gentle introduction to the new field of information visualization. Rather than a comprehensive survey of the field the goal is to provide sufficient background for you to develop an original visualization research proposal. Depending on your implementation skills, visualizations in the proposal may vary from sketches and descriptions of proposed visualizations to running demonstrations of the visualizations. The primary focus is on designing and building interactive visualizations for the web using Vega-Lite and D3.js (Data-Driven Documents) frameworks. In addition, you are encouraged to make use of Observable interactive JavaScript notebooks.

The course is run as a seminar. In addition to class discussions, participants will complete visualization assignments and a group visualization proposal. The ideal group size is three or four students. With approval, larger groups are possible. Your project group must be formed by the beginning of week four and your project or proposal topic must be approved by the end of week four. Project groups will share the results of their effort as both an interactive website and a video presentation.

The seminar meeting each week is devoted to discussing readings, tools and techniques, and your group's developing proposal ideas. In the beginning of the quarter, we will cover selected general background readings and discuss possible research proposal topics. As the quarter progresses, we will focus on readings chosen by groups to provide background information specific to their proposal.

By the end of the course, students will understand key visualization techniques and theory, have gained practical experience developing a visualization proposal, and advanced their skill with modern visualization toolkits.

Class Website
hci.ucsd.edu/220

This seminar provides a *gentle introduction* to the important new field of information visualization.

Rather than a survey of the field the goal is to help you acquire sufficient background to develop a visualization proposal.

COGSCI 220: Information Visualization (Winter 2021)

INSTRUCTOR
Jim Hollan
By appointment

POSTDOC
Zhutian Chen

GRAD STUDENT
Amy Rae Fox

LECTURE
Location: Zoom
Time: Tues 9:00-11:50

ABOUT
TEXTBOOKS
SCHEDULE
1 2 3 4 5 6 7 8 9 10

What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. -- Herb Simon

A graphic is never an end in itself; it is a moment in the process of decision making.-- Jacques Bertin

This seminar provides a gentle introduction to the new field of information visualization. Rather than a comprehensive survey of the field the goal is to provide sufficient background for you to develop an original visualization research proposal. Depending on your implementation skills, visualizations in the proposal may vary from sketches and descriptions of proposed visualizations to running demonstrations of the visualizations. The primary focus is on designing and building interactive visualizations for the web using Vega-Lite and D3.js (Data-Driven Documents) frameworks. In addition, you are encouraged to make use of Observable interactive JavaScript notebooks.

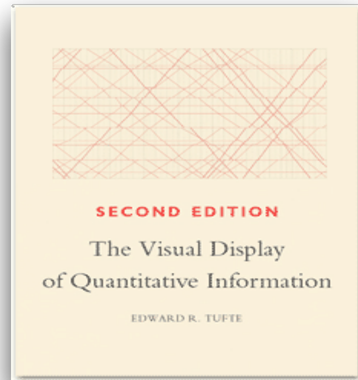
The course is run as a seminar. In addition to class discussions, participants will complete visualization assignments and a group visualization proposal. The ideal group size is three or four students. With approval, larger groups are possible. Your project group must be formed by the beginning of week four and your project or proposal topic must be approved by the end of week four. Project groups will share the results of their effort as both an interactive website and a video presentation.

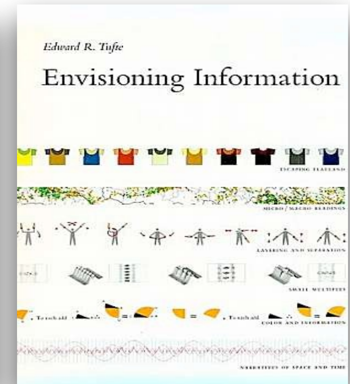
The seminar meeting each week is devoted to discussing readings, tools and techniques, and your group's developing proposal ideas. In the beginning of the quarter, we will cover selected general background readings and discuss possible research proposal topics. As the quarter progresses, we will focus on readings chosen by groups to provide background information specific to their proposal.

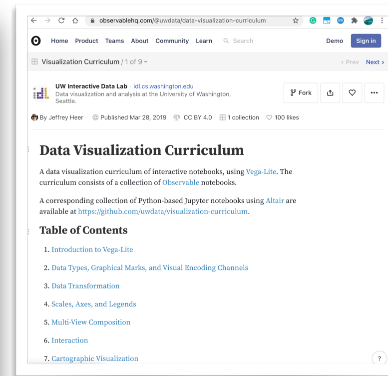
By the end of the course, students will understand key visualization techniques and theory, have gained practical experience developing a visualization proposal, and advanced their skill with modern visualization toolkits.

Approach

Read two of Tufte's books as a gentle introduction to the field of information visualization. Order now, consider sharing. Also complete the excellent UW IDL data visualization interactive notebook curriculum using Vega-Lite. For projects using D3, read Scott Murray's book.

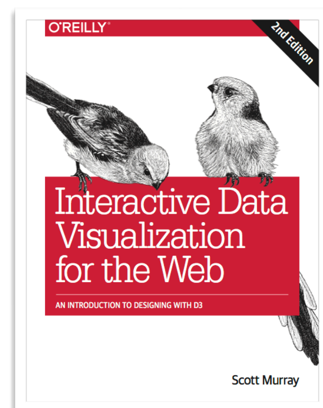






Algorithm for Course:

1. Find partners to work with. Three or four person groups are ideal. You should start to form groups as soon as possible. Both the topic and people are important. People are likely more important than topic.
2. Select a visualization area you are mutually interested in. Choice is constrained to personal information space ideas as described in recent NSF proposal.
3. Read selected literature in the area and identify a small set of excellent readings for class.
4. **Write an NSF-style research proposal.**



Along the way each group will share with the seminar their project idea, plans, and the best of the associated literature. A major goal of the seminar is for all of us to work together to help each group to do an excellent proposal.

Google Group

This is a reading and writing intensive seminar.

In each class you should be prepared to summarize the readings and to participate in a discussion of them. I may call on people to briefly summarize readings and to help lead discussions. Come prepared.

You should be an active participant in our Google Group.

It is the place for critical commentary about the readings, things you find especially interesting or problematic, or that stimulate a research idea. It is also the place to ask questions, discuss topics, and ask for help and advice.

Your participation in the group will count for 25% of your grade.

Proposal and Group Presentations

Your group of three or four people will jointly write an NSF style research proposal.

In week 5 each group will give a brief initial presentation about their project and select one or two readings for the class relevant to their project. This presentation is informal and the goal is for the class to provide feedback and help focus the project.

In weeks 6 to 9, groups will provide updates on their projects as well as lead discussion of their selected papers.

A video presentation of each project and the written proposal are due during finals week.

Initial Readings
(Linked on hci.ucsd.edu/220)

Schedule & Readings

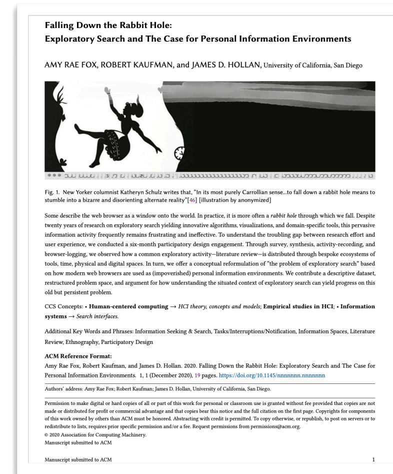
Week 1

Tues 1/5 **Information Visualization Course Overview and Introductions** Slides

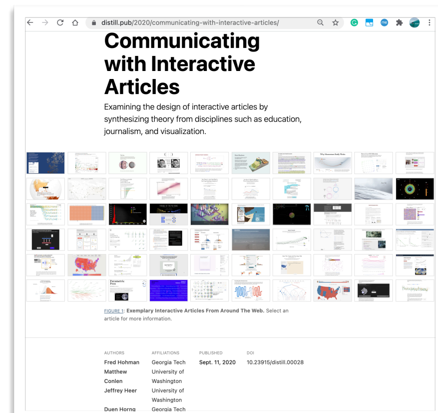
The first step in Selecting A Project and Group is to read the proposal and paper mentioned above in order to provide background and context for your selection of a proposal topic. This should be completed before 1/12.

- REQUIRED Communicating with Interactive Articles, Fred Hohman, Matt Conlen, Jeffrey Heer, Duen Horng (Polo) ChauDistill, 2020 Journal.
- Optional Human-Centered Interactivity of Visualization Tools: Micro- and Macro-level Considerations, Kamran Sedig, Paul Parsons, Mark Dittmer, and Robert Haworth.
- Optional A Tour Through the Visualization Zoo, Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky.
- Optional Visualizing Algorithms, Mike Bostock
- Optional Chapter 3: The Power of Representation, in *Things That Make Us Smart*. Don Norman.


CHS: Medium: A Human-Centered Information Space: Designing Dynamic Personalized Visual Information


Falling Down the Rabbit Hole: Exploratory Search and The Case for Personal Information Environments

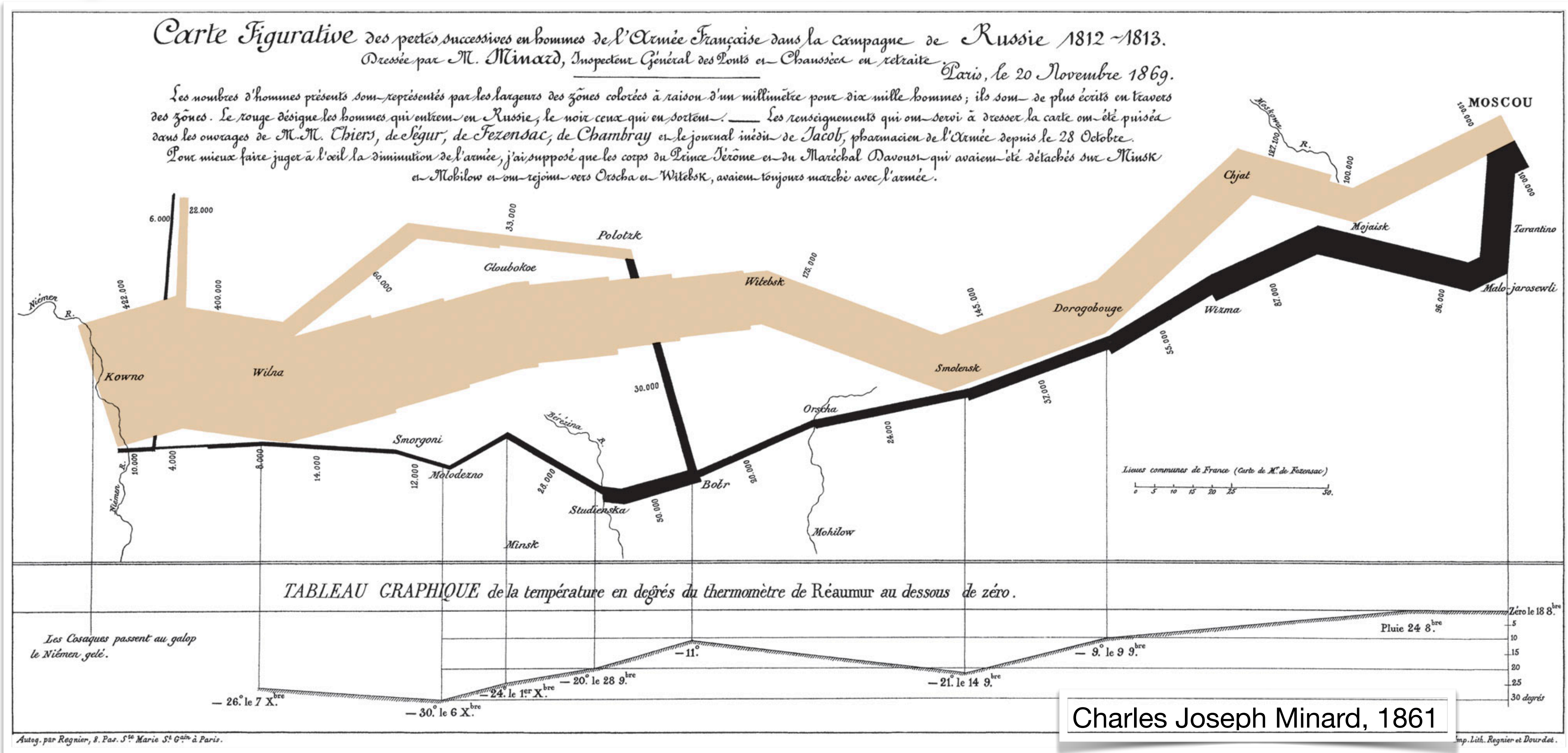
AMY RAE FOX, ROBERT KAUFMAN, and JAMES D. HOLLAN, University of California, San Diego


Communicating with Interactive Articles

Examining the design of interactive articles by synthesizing theory from disciplines such as education, journalism, and visualization.

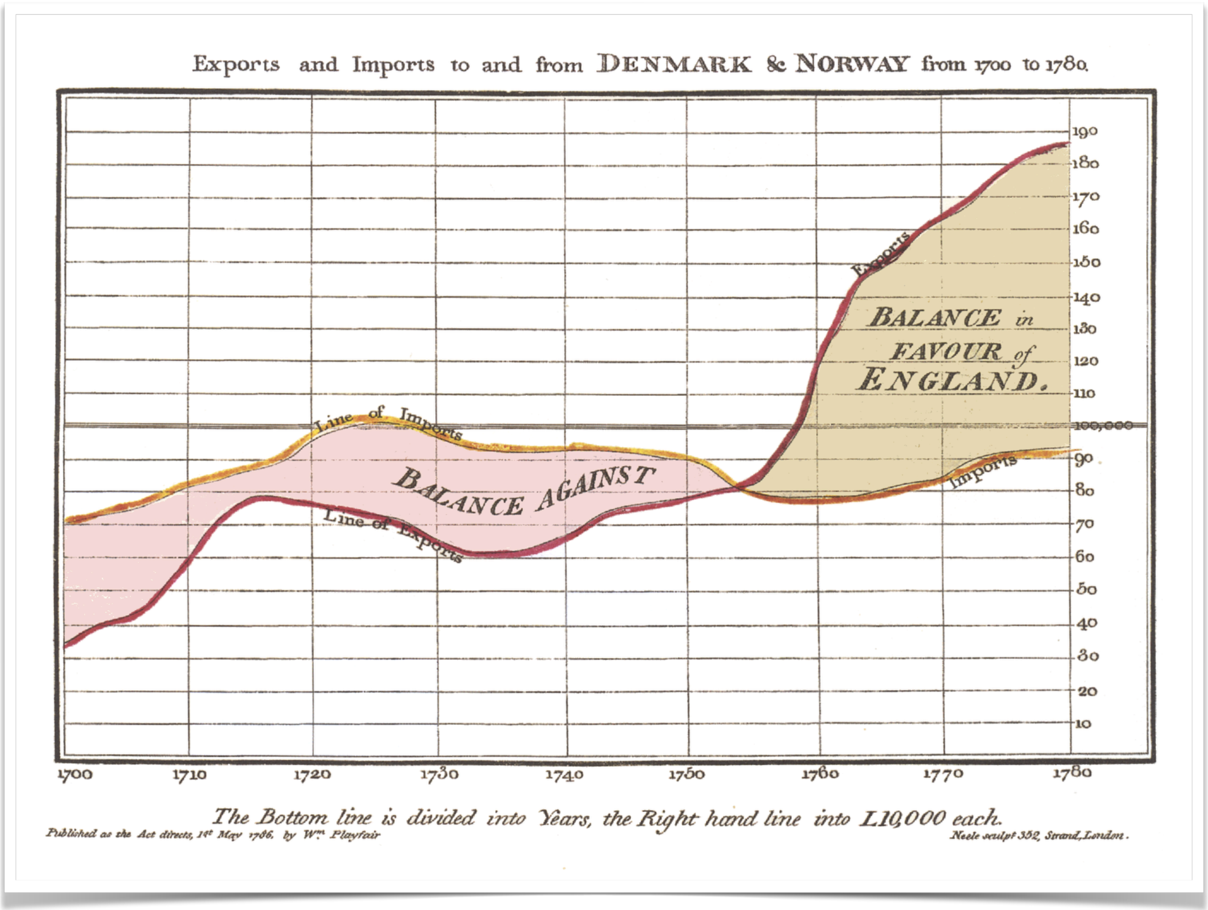
Optional Readings: Ideally you will read these too if you have time after reading required readings. You should at least skim and read parts that are interesting. Many will be relevant to your proposal and great sources of ideas.

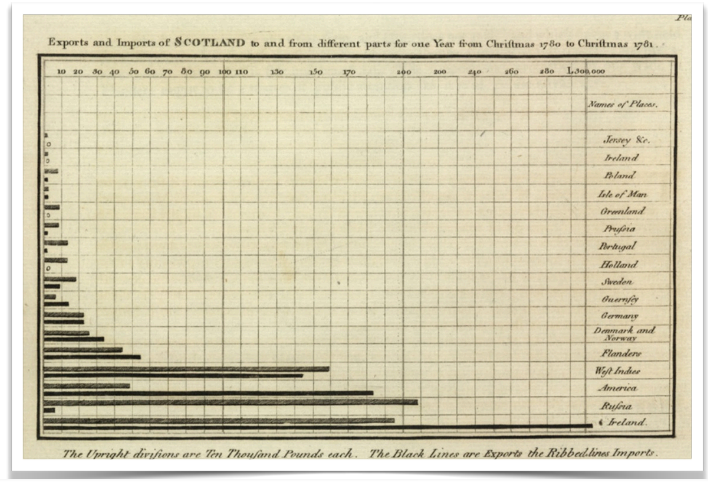
Cogsci 220: Information Visualization Seminar

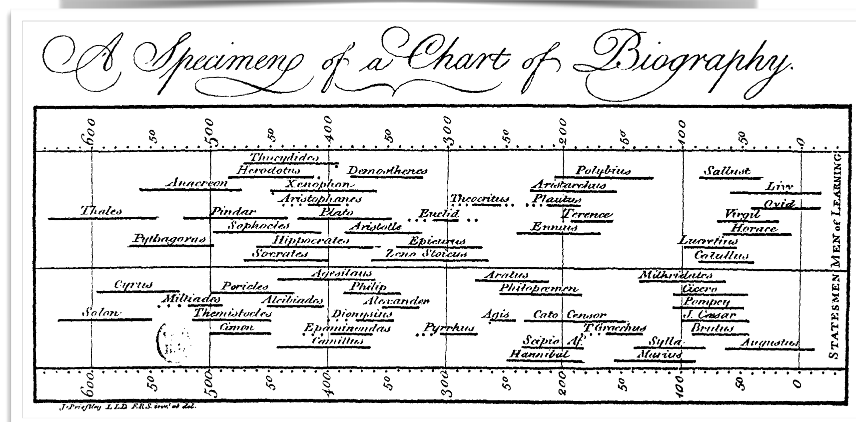


Charles Joseph Minard, 1861

Statistical graphics did not emerge until the late 18th century Most notably with William Playfield

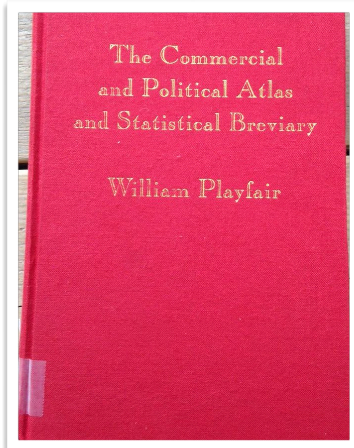


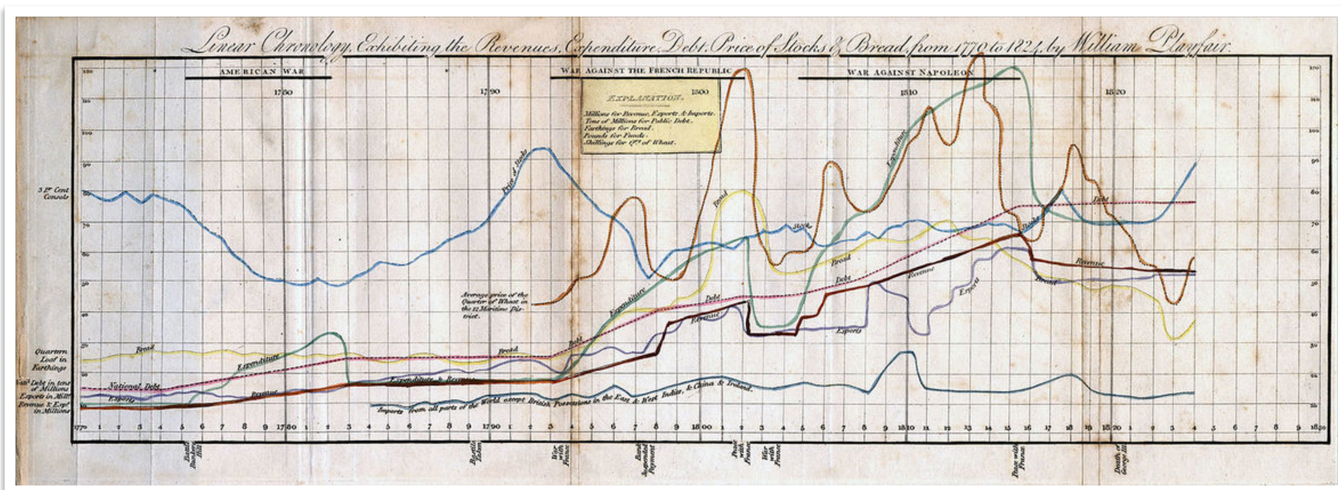


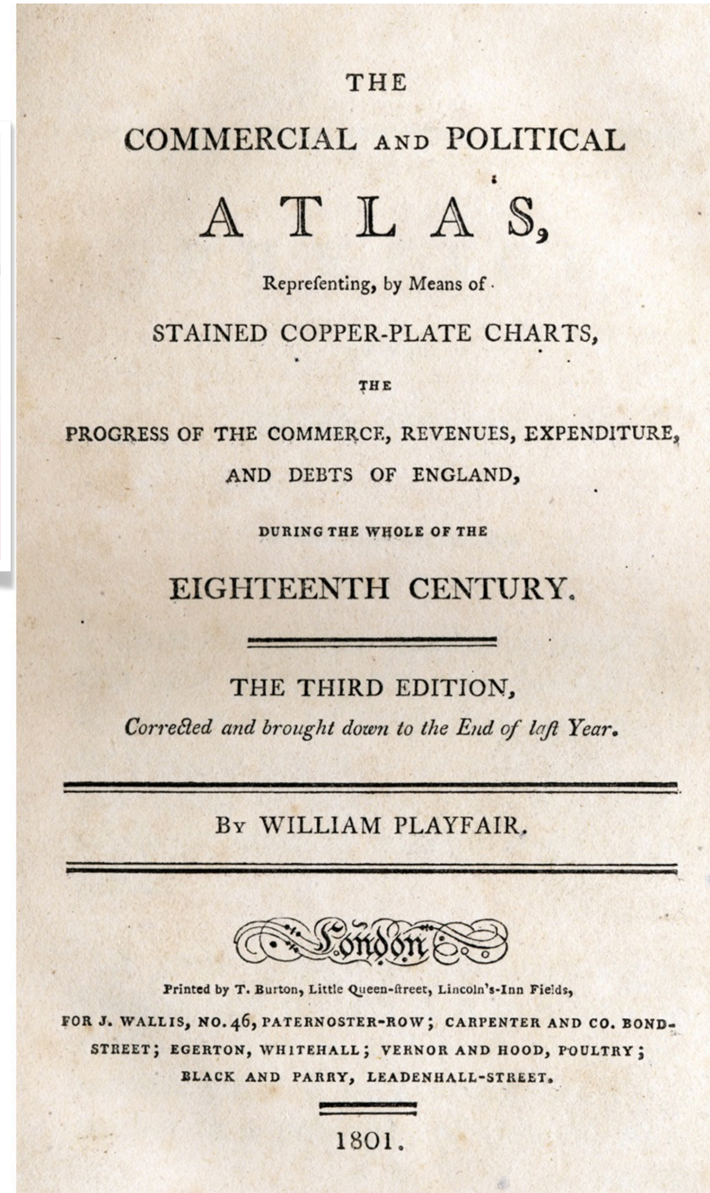


William Playfield

A scientific revolution began at the end of the 18th century with the invention and popularization of the graphic display of data by the remarkable Scot William Playfair. His marvelous *Atlas* showed how much could be learned if one plotted data atheoretically and looked for suggestive patterns. Those patterns provide evidence, albeit circumstantial, on which to build new science. Playfair's work has much to teach us, but finding a copy of it is very difficult. This full-color reproduction of two of his classic works, with new explanatory material, makes Playfair's wisdom widely available for the first time in two centuries.







THE

COMMERCIAL AND POLITICAL

ATLAS,

Reprefenting, by Means of

STAINED COPPER-PLATE CHARTS,

THE

PROGRESS OF THE COMMERCE, REVENUES, EXPENDITURE, AND DEBTS OF ENGLAND,

DURING THE WHOLE OF THE

EIGHTEENTH CENTURY.

THE THIRD EDITION,

Corrected and brought down to the End of laft Year.

BY WILLIAM PLAYFAIR.

London

Printed by T. Burton, Little Queen-ftreet, Lincoln's-Inn Fields,

FOR J. WALLIS, NO. 46, PATERNOSTER-ROW; CARPENTER AND CO. BOND-STREET; EGERTON, WHITEHALL; VERNOR AND HOOD, POULTRY; BLACK AND PARRY, LEADENHALL-STREET.

1801.

Data, Data, Data, and more Data

The availability of data and massive computational power is changing the world, as all the recent deep learning advances have make clear and as micro-targeting on social media has too.

It is certainly a different world. As I mentioned earlier, already today there have been a billion Google searches and <u>each</u> search took the equivalent computing of all the computing of the Apollo project that landed humans on the moon.

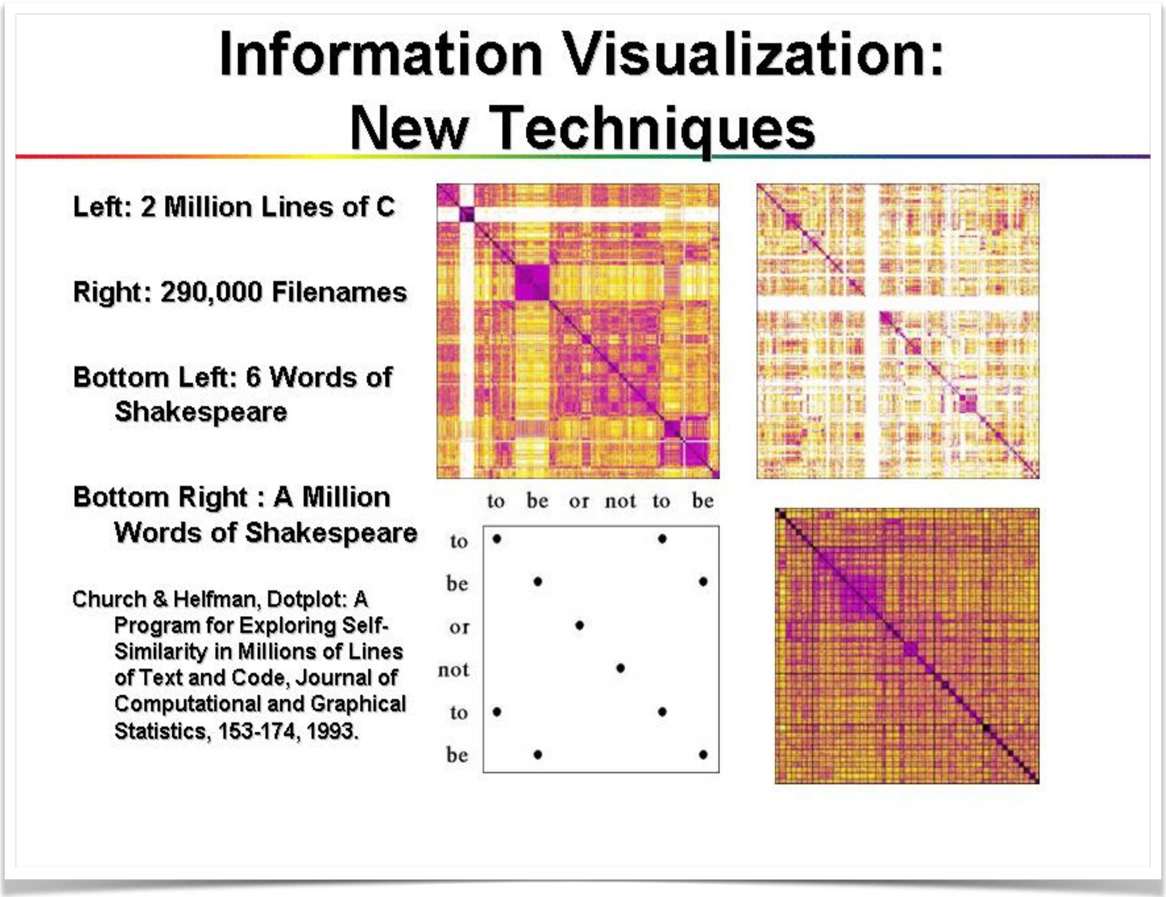
Google Data Center

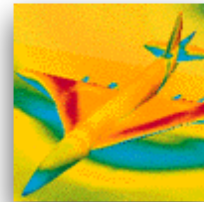
Something on the order of 5000 exabytes of data are produced each year and this is always increasing. 10x increase in the last five years.

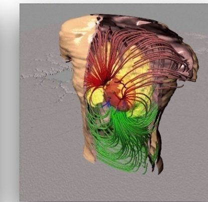
My laptop has 16 gigabytes of storage and 500gb of disk. A 1,024 gigabytes is a terabyte. A 1,024 terabytes is a petabyte. A 1,024 petabytes is a exabyte. One exabyte could hold the entire Library of Congress 3,000 times over.

In 2013 it was estimated that Google's data centers had about 10 exabytes of disk.

Scientific Visualization versus Information Visualization



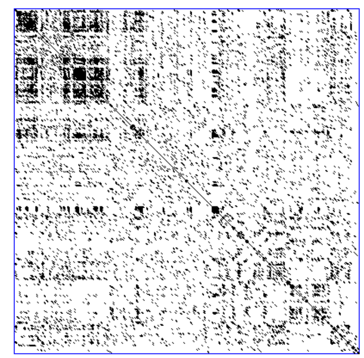




Scientific Visualization Examples:

Airflow over aircraft wing

Electrical current in thorax



A DNA dotplot

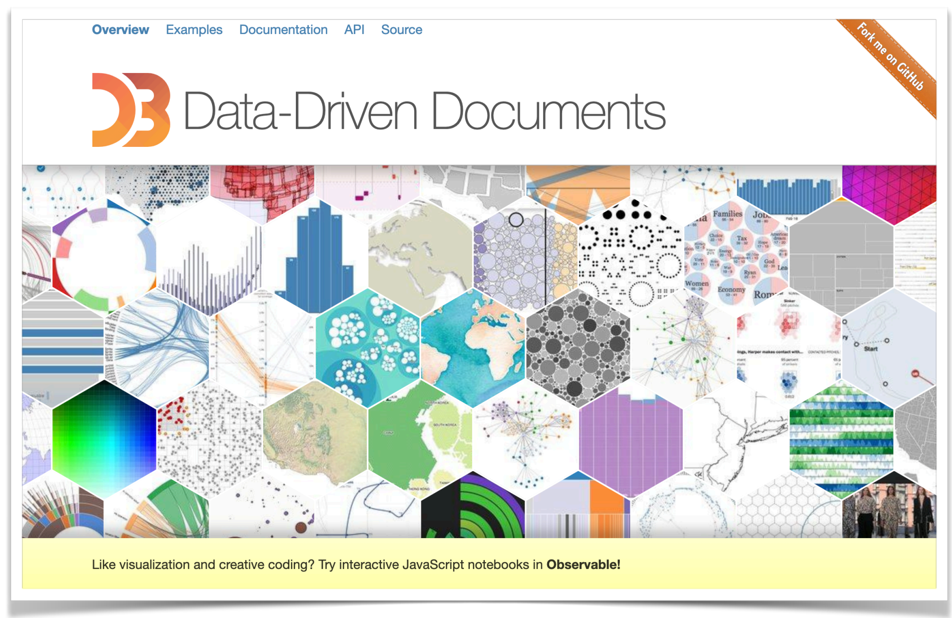
Visualization Software

One goal of the course is for you develop expertise with some visualization software to augment your research

Different research communities have selected different software systems (Python, R, Javascript, …)

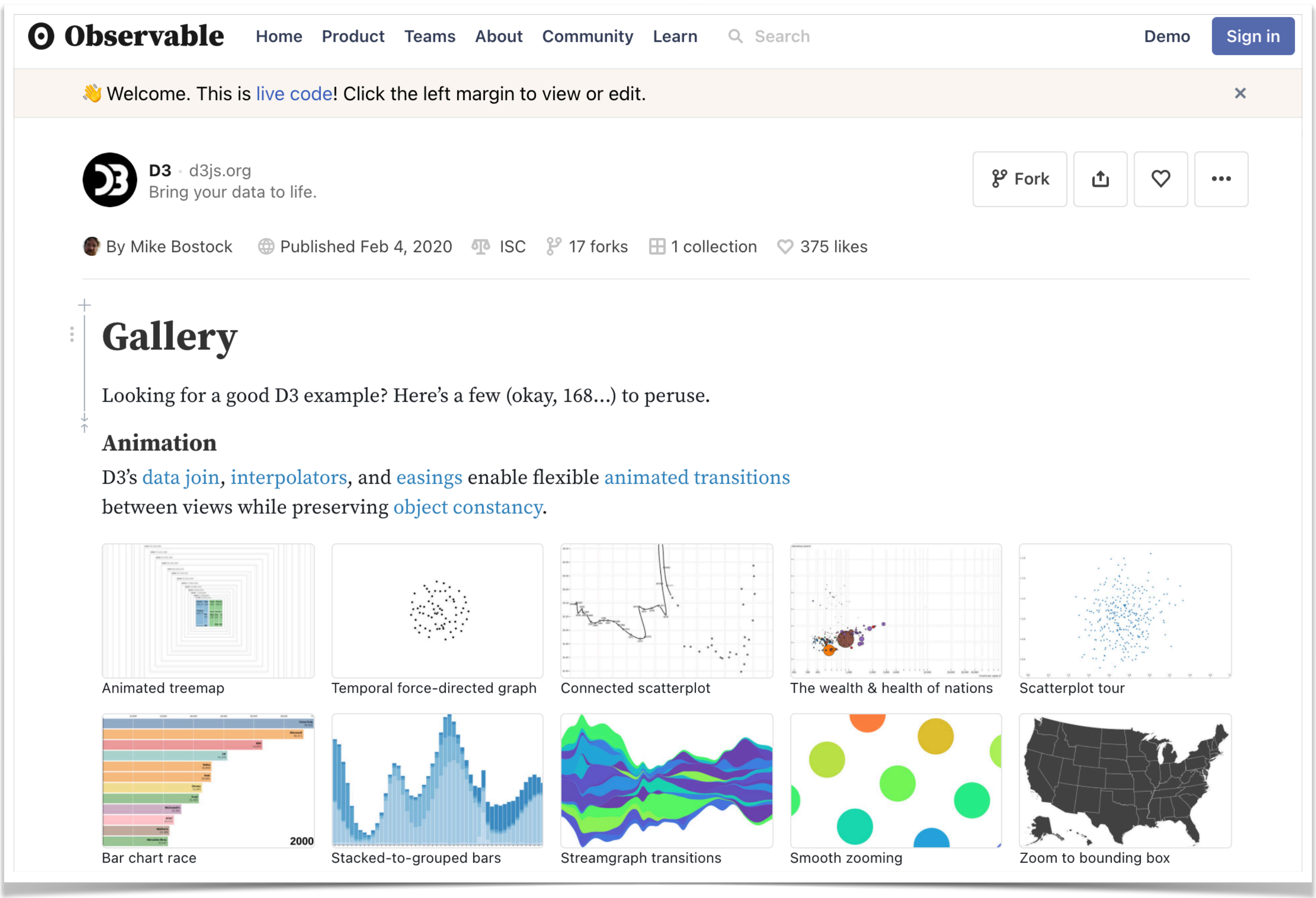
Our focus is on Vega-Lite and D3 (Data Directed Documents).

Very active communities.



D3 (Data-Driven Documents) https://d3js.org/

https://observablehq.com/@d3/gallery



All the O’Reilly books are available for free on the ucsd library site. Must be at ucsd or on via vpn.

News & Events

The Library UC SAN DIEGO

Home | Events | Exhibits | Press Releases | Library Updates | Search

O’Reilly for Higher Education (formerly Safari/O’Reilly Tech Books)

June 11, 2020

UC San Diego now has O’Reilly for Higher Education, replacing the Safari/O’Reilly Tech book collection that is being retired. We now have access to **more IT/tech publisher content** (including videos) than just O’Reilly. The new platform has unlimited simultaneous users. ProQuest has a user guide to introduce you to the interface, how to browse and search, and how to create your own playlists and follow playlists created by others.

How to Access OHE:

Unlike the other library-licensed electronic resources, OHE does not require VPN. However, you are still required to login, and if you are

O'REILLY

Welcome! Get instant access through your library.

At a Glance

Connect With Us

Library Publications

Newsletter Subscription

Categories

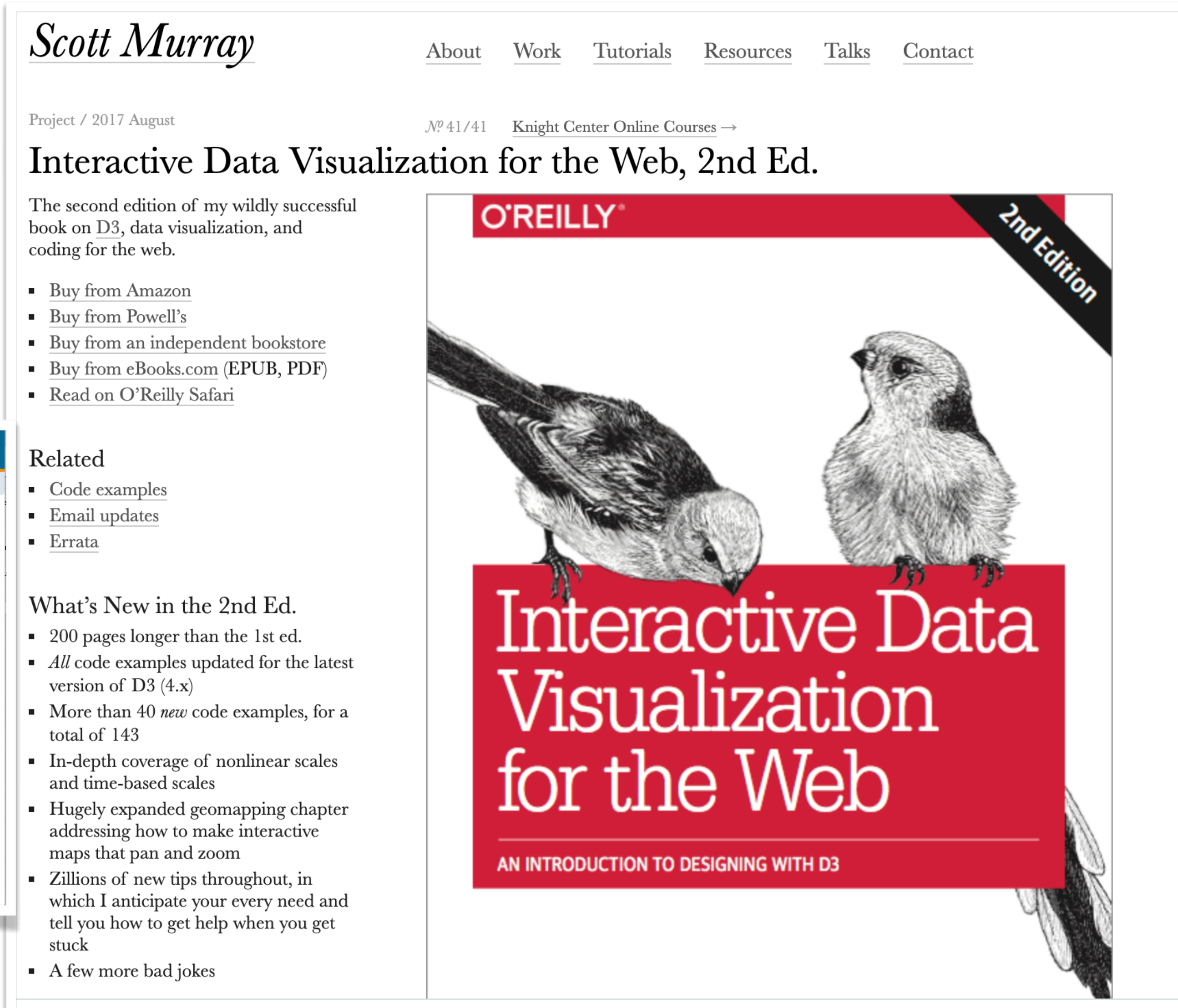
Library Services

Outage Information & Hours

Research & Collections

Technology News & Updates

Support the Library



Scott Murray

About | Work | Tutorials | Resources | Talks | Contact

Project / 2017 August

№ 41/41 Knight Center Online Courses →

Interactive Data Visualization for the Web, 2nd Ed.

The second edition of my wildly successful book on D3, data visualization, and coding for the web.

- Buy from Amazon
- Buy from Powell’s
- Buy from an independent bookstore
- Buy from eBooks.com (EPUB, PDF)
- Read on O’Reilly Safari

Related

- Code examples
- Email updates
- Errata

What’s New in the 2nd Ed.

- 200 pages longer than the 1st ed.
- *All* code examples updated for the latest version of D3 (4.x)
- More than 40 *new* code examples, for a total of 143
- In-depth coverage of nonlinear scales and time-based scales
- Hugely expanded geomapping chapter addressing how to make interactive maps that pan and zoom
- Zillions of new tips throughout, in which I anticipate your every need and tell you how to get help when you get stuck
- A few more bad jokes

Tableau

Tableau for Students
-get a free Tableau license as a student

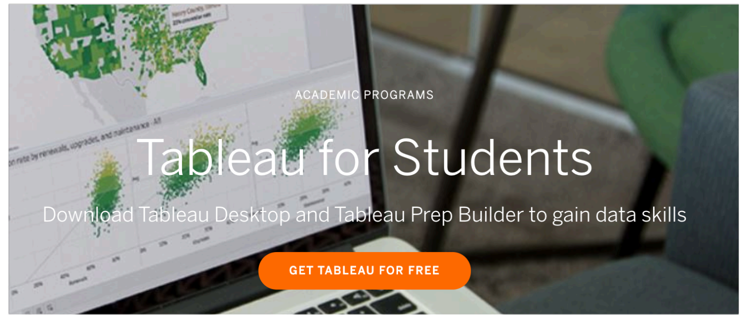
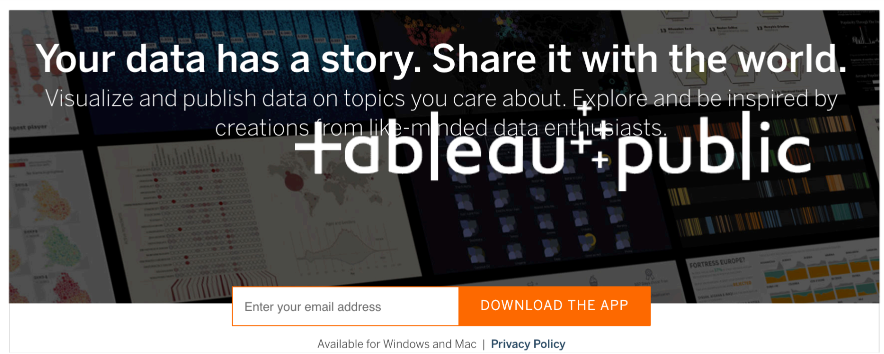


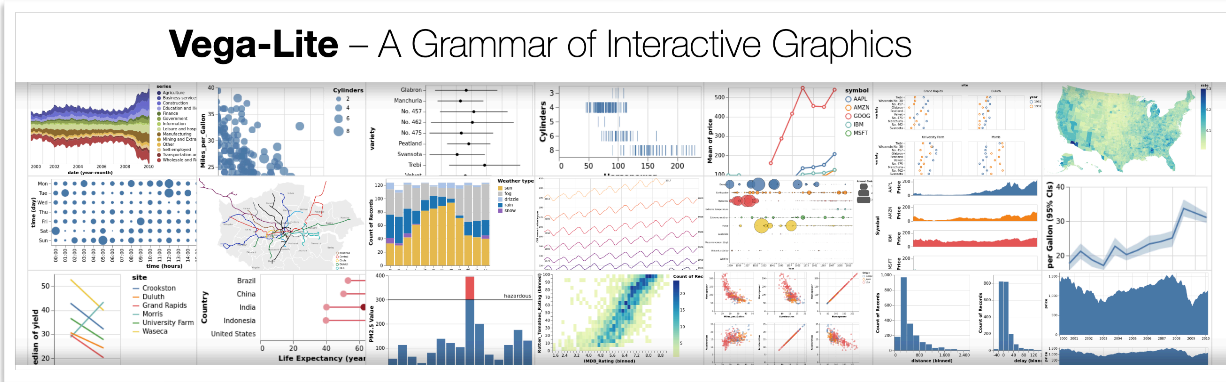
tableau.com/academic/students

Tableau Public
- a free version of Tableau which publishes to the web
public.tableau.com



Vega-Lite: A Grammar for Interactive Graphics

Arvind Satyanarayan

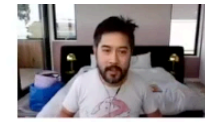


D3 -> Vega -> Vega-Lite

Vega-Lite specifications describe visualizations as mappings from data to **properties of graphical marks** (e.g., points or bars).

The Vega-Lite compiler **automatically produces visualization components** including axes, legends, and scales. It then determines properties of these components based on a set of **carefully designed rules**. This approach allows specifications to be succinct and expressive, but also provide user control.

As Vega-Lite is designed for analysis, it supports **data transformations** such as aggregation, binning, filtering, sorting, and **visual transformations** including stacking and faceting. Moreover, Vega-Lite specifications can be **composed** into layered and multi-view displays, and made **interactive with selections**.



Vega-Lite: A Crash Course

"Vega-Lite is a high-level grammar of interactive graphics. It provides a concise JSON syntax for rapidly generating visualizations to support analysis."

D3 → Vega → Vega-Lite

Visnu Pitiyanuvath
Engineer
observablehq.com

Vega-Lite Crash Course
https://www.youtube.com/watch?v=ZV_Yjcs5WtM