Information Visualization
Logistics, Intro

Tamara Munzner
Department of Computer Science
University of British Columbia

Week 1, Tue 12 Jan 2021

https://www.students.cs.ubc.ca/~cs-436v/21Jan/
Logistics
Course staff

• Instructor:
  – Tamara Munzner (she)

• TAs:
  – Steve Kasica (he)
  – Francis Nguyen (he)

• Piazza is the best way to reach us
  – use for all discussion and questions (not email)
  – https://piazza.com/class/k41qv94wb3r4uq
Course structure

• theoretical foundations, all term
  – hybrid sync / async
  – async pre-class: video lectures once/week (no Tue sync session)
  – sync in-class: Thu 2-3:20pm once/week
    • additional lectures & in-class exercises leading into foundations exercises
  – post-class: finish foundations exercises

• D3 programming, weeks 1-6
  – fully asynchronous
  – each week: work through written tutorials
  – end of week: tutorial quizzes, do by 5pm Fridays
  – individualized help: TA office hours (Friday labs & more times TBD)
  – every other week: finish programming exercises at home, to hand in
Course structure

• final projects, weeks 8-14
  – integrate programming and foundations
  – self-chosen teams of 3
  – stages
    • milestone 1: proposal (due Mar 10)
    • milestone 2: work in progress (due Mar 30)
    • milestone 3: final version (due Apr 13)

• exams
  – midterm (Mar 25)
  – primary focus will be on foundations
  – no final exam

• participation
  – in-class exercises, Piazza discussion
Grading scheme

• Final Project: 40%
  – Milestone 1: 6%
    • Foundations Design 60%
    • Project Management 15%
    • Writeup 25%
  – Milestone 2: 14%
    • Programming Achievement: 80%
    • Project Management: 5%
    • Process Log Writeup: 15%
  – Milestone 3: 20%
    • Programming Achievement: 40%
    • Foundations Design: 40%
    • Process Log Writeup: 15%

• Midterm Exam: 15%

• Programming Assignments: 15%
  – P0 3%, P1 6%, P2 6%

• Foundations Assignments: 15%
  – 3 instances, 5% each

• Participation: 10%
  – in-class exercises, Piazza engagement

• Post-Tutorial Check Quizzes: 6%
  – 6 quizzes, 5 of them count 1% each
    (worst score dropped)
Rhythm of the week

• async
  – video lecture on foundations
  – written tutorials on D3 tooling
• Tue: Tamara office hours 2-3:20
• Wed 11:59pm: assignments usually due
  – foundations or programming or project milestone
  – exception: midterm week
• Thu 2-3:20: sync class session
• Fri & other days TBD: TA office hours
• Fri 5pm: quiz on tutorials due
Delivery mechanisms

- **Web**: syllabus & all instructional materials, slides, videos
  - [https://students.cs.ubc.ca/~cs-436v/21Jan](https://students.cs.ubc.ca/~cs-436v/21Jan)
  - main reference point! don’t forget to refresh, frequent updates
- **Zoom**: synchronous sessions, sync session polls, office hours
- **Piazza**: all asynchronous discussion & questions, project team formation discussion, logistics and other updates
- **Canvas**: quizzes, marks, project team signup, office hours and marking demos signup, submit foundations & prog assignment writeups & project writeups
- **Github/Classy**: distribute, revision control, submit for programming assignments & tutorials & examples & case studies. Revision control & submit for project.
- **Codesandbox.io**: web IDE for D3 examples and case studies
- **Gradescope**: midterm exam marking/handback
Getting help

• TA office hours, by Zoom
  – 3 lab slots on Fridays: 9-10, 11-12, 4-5
  – additional times TBD after time zones poll
  – first lab: Jan 22 (week 2)
  – consultation on D3 exercises and final project
  – 15 min slots bookable in advance
  – fine to just drop in if nobody is signed up

• my office hours Tue in lecture slot (2-3:30pm)
  – or by appointment, Piazza private message to arrange
  – by Zoom, see Piazza post for link
Resources

• optional textbook for further reading
  – Tamara Munzner.
    Visualization Analysis and Design.
    AK Peters Visualization Series. CRC Press, 2014.
    • https://www.cs.ubc.ca/~tmm/vadbook/
  – UBC library has multiple free ebook copies
  – content will be covered in async lecture videos & sync in-class sessions
Academic honesty

• short version
  – don't cheat!
    • bad consequences: you won't learn the material
    • bad consequence: I will prosecute, to keep playing field fair for others
  – do ask for help, we want you to succeed!

• long version
  – read course-specific page
  – read my general page
  – read department page
  – Quiz 1 Question 1 will require you to sign off that you've done so
Waitlist

• Very likely to clear, just act as if you're in
  – It's always cleared in the past

• Stay calm and carry on
  – Stay caught up: attend class, keep up with work
  – Canvas access is only an issue for quizzes and handin, most things on web page
  – You'll be allowed to take quizzes without penalty once you're registered
  – We'll do workaround for Foundations 1 handin if need be
Todo this week

• Sync class session Thu Jan 14 2-3:30
• Refresher tutorial: web development (if you need it, no quiz)
• D3 tutorial: Intro to D3
• Quiz 1: due by Fri Jan 15, 5pm
• remember, no labs (TA office hours) this week

• Foundations Exercise 1 will be released Thu Jan 14, due Wed Jan 20
• Programming 0 will be released Thu Jan 14, due Wed Jan 27

• my office hours start next week, Tue 2-3:30
  — or by appointment, send Piazza private message to schedul
Introduction
Why create visualizations?

• analyze data to support reasoning
• answer questions
• communicate ideas to others
• confirm hypotheses
• expand memory
• find/reveal patterns
• generate hypotheses
• inspire
• make decisions

• record information
• see data in context
• support computational analysis
• tell a story
Reveal patterns

Mapping Migration in the United States

Where people who lived in each state in 2012 were born

Each circle represents the percentage of people living in a state who were born in a state, with larger circles indicating a larger share of the population.

1900 1950 2012

Northeast South Midwest West Outside the U.S.

Communicate ideas to others

The Upshot, Five Years In

Which subway map is better? Why?
Many definitions

• The purpose of visualization is insight, not pictures
• Visualization is really about external cognition, that is, how resources outside the mind can be used to boost the cognitive capabilities of the mind
• Good data visualization...
  – makes data accessible
  – combines strengths of humans and computers
  – enables insight
  – communicates
• visualization = human data interaction
My own favorite definition

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
Visualization: definition & motivation

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- human in the loop needs the details & no trusted automatic solution exists
  - doesn't know exactly what questions to ask in advance
  - exploratory data analysis
    - speed up through human-in-the-loop visual data analysis
  - present known results to others
  - stepping stone towards automation
    - before model creation to provide understanding
    - during algorithm creation to refine, debug, set parameters
    - before or during deployment to build trust and monitor
Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• external representation: replace cognition with perception

Why depend on vision?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• human visual system is high-bandwidth channel to brain
  – overview possible due to background processing
    • subjective experience of seeing everything simultaneously
    • significant processing occurs in parallel and pre-attentively

• sound: lower bandwidth and different semantics
  – overview not supported
    • subjective experience of sequential stream

• touch/haptics: impoverished record/replay capacity
  – only very low-bandwidth communication thus far

• taste, smell: no viable record/replay devices
Why represent all the data?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, details matter — confirm expected and find unexpected patterns
- assess validity of statistical model

Anscombe’s Quartet

| Identical statistics |  
|----------------------|--------|
| x mean               | 9      |
| x variance           | 10     |
| y mean               | 7.5    |
| y variance           | 3.75   |
| x/y correlation      | 0.816  |

Datasaurus Dozen


Matejka & Fitzmaurice

https://www.youtube.com/watch?v=DbJyPELmhJc
Why focus on tasks and effectiveness?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• effectiveness requires match between data/task and representation
  – set of representations is huge
  – many are ineffective mismatch for specific data/task combo
  – increases chance of finding good solutions if you understand full space of possibilities

• what counts as effective?
  – novel: enable entirely new kinds of analysis
  – faster: speed up existing workflows

• how to validate effectiveness
  – many methods, must pick appropriate one for your context
Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

- computational limits
  - processing time
  - system memory
- human limits
  - human attention, cognition, and memory
- display limits
  - pixels are precious resource, the most constrained resource
  - information density: ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space, find sweet spot between dense and sparse
Why does visualization work?

- limits of memory & cognition
  - change blindness

Dan Simons, The "Door" Study

https://youtu.be/FWSxSQsspiQ
Credits

• Visualization Analysis and Design (Ch 1)
• Alex Lex & Miriah Meyer, http://dataviscourse.net/