# Information Visualization Intro

**Tamara Munzner**
Department of Computer Science
University of British Columbia

***Lect 1, 7 Jan 2020***

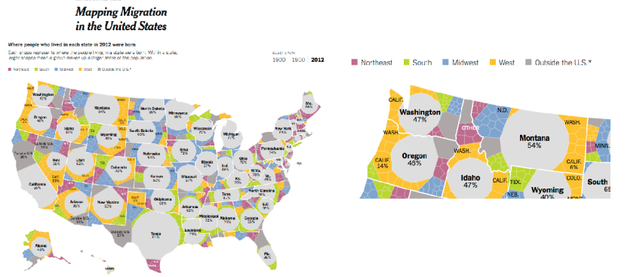
**http://www.cs.ubc.ca/~tmm/courses/436V-20**

## 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




https://www.nytimes.com/2014/08/16/upshot/mapping-migration-in-the-united-states-since-1900.html



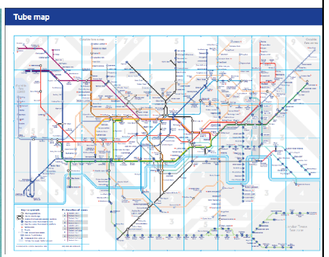
## Communicate ideas to others


https://www.nytimes.com/interactive/2019/04/22/upshot/upshot-at-five-years.html



## Which subway map is better? Why?

1927 2019




## 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.**

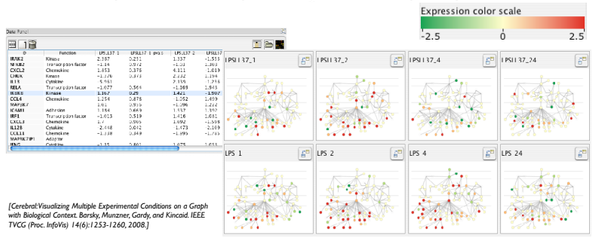
- human in the loop needs the details & no trusted automatic solution exists
  - doesn't know exactly what questions to ask in advance
  - exploratory data analysis
    - ***speed up*** through human-in-the-loop visual data analysis
  - present known results to others
  - stepping stone towards automation
    - 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




*[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]*



## 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



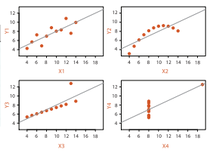
## Why represent all the data?

**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

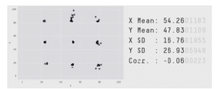
**Anscombe's Quartet**

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



https://www.youtube.com/watch?v=DbIyPELmhJc

**Datasaurus Dozen**



Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. CHI 2017. Matejka & Fitzmaurice



## 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



## What resource limitations are we faced with?

**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



## Why does visualization work?

- limits of memory & cognition
  - change blindness
- power of perception to reveal
  - how many V's?

MTHIVLWYADCEQGHKILKMTWYN

ARDCAIREQGHLVKMFPSTWYARN

GFPSVCEILQGKMFPSNDRCEQDIFP

SGHLMFHKMVPSTWYACEQTWRN



## Why does visualization work?

- limits of memory & cognition
  - change blindness
- power of perception to reveal
  - how many V's?

MTHI**V**LWYADCEQGHKILKMTWYN

ARDCAIREQGHL**V**KMFPSTWYARN

GFPS**V**CEILQGKMFPSNDRCEQDIFP

SGHLMFHKM**V**PSTWYACEQTWRN

## Why does visualization work?

- limits of memory & cognition
  - change blindness
- power of perception to reveal
  - how many V's?
  - which of these 50 numbers appears most often?

15 19 60 33 11 75 57 34 79 18 51 92 73 22 13 71 60 22
17 10 68 73 18 55 65 46 29 60 73 22 46 92 97 10 58 46
57 17 83 26 99 33 88 92 60 91 29 57 96 12 47



## Why does visualization work?

- limits of memory & cognition
  - change blindness
- power of perception to reveal
  - how many V's?
  - which of these 50 numbers appears most often?

15 19 60 33 11 75 57 34 79 18 51 92 73 22 13 71 60 22
17 10 68 73 18 55 65 46 29 60 73 22 46 92 97 10 58 46
57 17 83 26 99 33 88 92 60 91 29 57 96 12 47




## Exercise

- Which gender and income level shows a different effect of age on triglyceride levels?

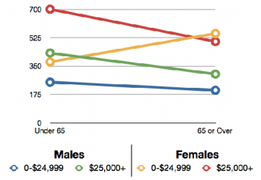
| | Males | | Females | |
|---|---|---|---|---|
| **Income Group** | Under 65 | 65 or Over | Under 65 | 65 or Over |
| **0-$24,999** | 250 | 200 | 375 | 550 |
| **$25,000+** | 430 | 300 | 700 | 500 |



## Exercise

- Which gender and income level shows a different effect of age on triglyceride levels?

| | Males | | Females | |
|---|---|---|---|---|
| **Income Group** | Under 65 | 65 or Over | Under 65 | 65 or Over |
| **0-$24,999** | 250 | 200 | 375 | 550 |
| **$25,000+** | 430 | 300 | 700 | 500 |

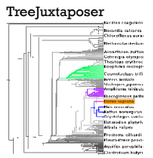




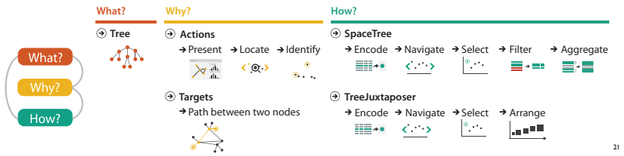

## Why analyze visualizations?

- imposes structure on huge design space
  - scaffold to help you think systematically about choices
  - analyzing existing as stepping stone to designing new
  - most possibilities ineffective for particular task/data combination


[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57–64.]




[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]






# Logistics



## Course staff

- Instructor:
  - Tamara Munzner
  - pronouns: she/her
- TAs:
  - Michael Oppermann
  - Zipeng Liu
  - pronouns: he/him

- 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
  - in-class: lectures twice/week, 2-3:20pm Tue/Thu
  - in-class: in-class exercises leading into foundations exercises
  - post-class: finish foundations exercises
- D3 programming, weeks 1-8
  - partially flipped
  - pre-class: watch videos (plus a few readings)
  - pre-class: pre-lab quizzes, do by 8am Fridays
  - in-class: work on programming exercises in Friday labs
    - individualized consultation with TAs
  - post-class: finish exercises at home, to hand in



## Course structure

- final projects, weeks 6-14
  - integrate programming and foundations
  - self-chosen teams of 3
  - stages
    - milestone 1: pitch (due Mar 6)
    - milestone 2: work in progress (due Mar 25)
    - milestone 3: final version (due Apr 8)
- exams
  - midterm (Mar 12)
  - final (TBD)
  - primary focus will be on foundations
- participation
  - in-class exercises, Piazza discussion



## Grading Scheme

- Exams: 30%
  - Midterm Exam: 10%, Final Exam: 20%
- Final Project: 30%
  - Programming Achievement: 40% of project
  - Foundations Design: 40% of project
  - Process Log Writeup: 20% of project
- Programming Assignments: 12%
  - 3 instances, 4% each
- Foundations Assignments: 12%
  - 6 instances, 2% each
- Participation: 10%
  - in-class exercises, Piazza discussion
- Pre-Lab Prep Quizzes: 6%
  - 7 quizzes, 6 of them count 1% each (worst score dropped)



## Information

- web: course page is the vortex
  - mirror/temporary now up: https://www.cs.ubc.ca/~tmm/courses/436V-20/
  - permanent URL coming soon: https://www.students.cs.ubc.ca/~cs-436v/20Jan
  - don't forget to refresh, frequent updates
- Socrative: software clicker
  - https://api.socrative.com/rc/FwT2fa
- Canvas: pre-lab quizzes
  - https://canvas.ubc.ca/courses/44149
- github, classy
  - stay tuned



`

- lectures Tue/Thu
- labs Friday
  - watch videos before then
  - pre-lab quizzes due by Fri 8am
    - released by Wed morning
  - start/continue programming assignments
  - individualized help on projects
- assignments due Wed 6pm
  - foundations or programming or project milestone
  - exception: midterm week shift



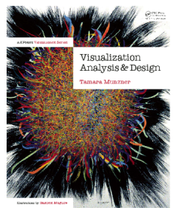
## Getting help

- labs with TAs
  - 3 slots on Fridays: 9-10, 11-12, 4-5
  - all in ICICS/CS Room 015
  - first lab: Jan 17
  - consultation on D3 exercises and final project
- my office hours Tue right after class (3:30-4:30pm)
  - or by appointment, email me to arrange (tmm@cs.ubc.ca)
    - unlikely to catch me by dropping by, I'm usually either in meeting or elsewhere
  - X661 (X-Wing of ICICS/CS bldg)



## 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 lecture






## Todo this week

- D3 videos to watch this week
  - refresher only if you need it: JS/HTML [90 min]
  - Intro to HTML/CSS/SVG [35 min]
  - Intro to D3.js [45 min]
- Quiz 1 to do this week, due by Fri Jan 10, 8am
- remember, no in-person labs this week!
- Foundations Exercise 1 will be released Thu Jan 9, due Wed Jan 15
- my office hours start today, right after class (X661)



## Credits

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