CS4460: **Introduction to Information Visualization**

Course **Overview** and **Intro**

Alex Endert

Spring, 2019

Agenda for Today

- Introductions of your instructor & GTAs
- a short intro to infovis (more extensive next time)
- Course information
  - Project information
  - Assignments (HW and P)
  - Expectations
  - Schedule
  - Grading

- *The point of this lecture is for you to know whether or not you should take this course this semester*

Introductions - Alex Endert

- Ph.D. from Virginia Tech in 2012
- Worked at Pacific Northwest National Labs for 2 years
- Joined GT in 2014

- Direct the **GT Visual Analytics Lab**
  - My students and I actively research and develop InfoVis tools and techniques.
  - i.e., *the topic of this course is my active area of research and career interest!*

hobbies

You also have a great group of TAs!

- Emily Wall
- Fred Hohman
- Julia Deeb-Swihart
- Matt Britton
- Rachel Chen

Emily Wall

contact: emilywall@gatech.edu

me: 4th year PhD student

research: mitigating *cognitive bias* in VA decision making

office hours: sometime Tuesday, Wednesday, or Thursday ☺

fun fact: one time I did a firewalk

Fred Hohman

fredhohman@gatech.edu
fredhohman.com
@fredhohman

Audiophile (i.e. music snob), frisbee thrower

Office hours TBA, likely in:
- Klaus 1st floor common area, or
- TSRB 335

GT CSE PhD Student 2015—
Human-centered & explainable AI,
visual analytics,
explorable explanations


Polo Chau


Alex Endert

Pacific Northwest National Lab 2016
NASA Jet Propulsion Lab 2017
Microsoft Research 2018

Pacific Northwest NATIONAL LABORATORY

NASA

Microsoft Research

Julia Deeb-Swihart

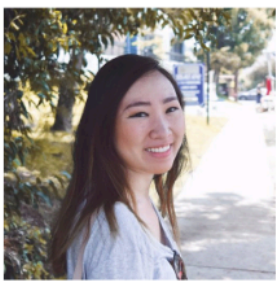
- Email: jdeeb3@gatech.edu
- Research: I build visual analytics tools for law enforcement to help combat human trafficking.
- Office Hours: TSRB 334 (& lab area outside)

Matt Britton

- M.S. student in Visual Analytics
- mbritton8@gatech.edu
- Office Hours are Thursday 1:30-2:30
- My research interests include:
  - Interpretable Machine Learning
  - Data Journalism
- In my spare time, I like cooking, hiking, and hanging with my 2-yr-old daughter.

Rachel Chen

- Second year MS-HCI student
- Work: UX research
- Interests: data journalism, crowdsourcing, information-sharing
- Spare-time activities: running & lettering



Information Visualization

a quick overview

What is Information Visualization?

- My elevator pitch
  - Presenting data via **interactive** charts, graphs, maps so that users can understand the data, answer questions about the data and gain insights from the data
  - NOT just printed (static) information presentation – *interaction is key*!

What is Information Visualization?

- The engineering definition
  - Binding data values and structure to graphical elements on a display
- Why do we do this?
  - Help people **think about their data**
  - Help people answer **questions** (and come up with new ones)
  - Data **Usability**!
- *Longer version == this whole course*

What is Information Visualization?

- Others' definitions
  - The use of computer-supported, interactive visual representations of data to **amplify cognition** (Card, Mackinlay Shneiderman '98)
- Alternative definitions of InfoVis, but what are they missing?
  - Transformation of the **symbolic into the geometric** (McCormick et al., 1987)
  - The **depiction of information using spatial or graphical representations**, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by **making use of the visual system** (Hearst 03).

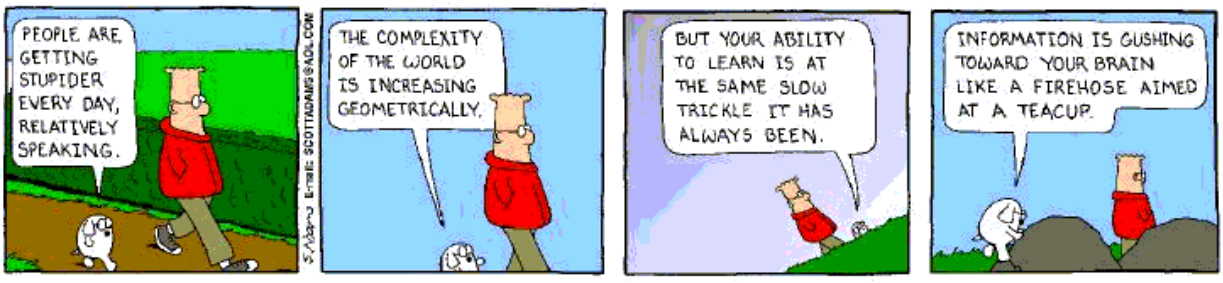
Some InfoVis Application Areas

- Financial/business data –“business intelligence”
- Internet information –traffic, topology
- Software development
- Intelligence analysis
- Social networking analysis
- Weather
- Sports
- Politics

- and increasingly more each year

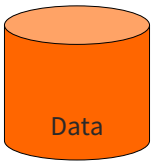
Data Overload

- How to make use of data?
  - How do we make sense of data?
  - How do we harness data in decision-making processes?
  - How do we avoid being overwhelmed?



The Problem

Web,
Books,
Papers,
Game scores,
Scientific data,
Biotech,
Shopping
People
Stock/finance
News



Data Transfer

How?

Vision: 100 MB/s
Ears: <100 b/s
Telepathy
Haptic/tactile
Smell
Taste

slide courtesy
of Chris North

Information Visualization

- What is “information”?
  - Items, entities, things which do not necessarily have a direct physical correspondence
  - Often collected from the “real world”, and through sensors, made into “information” (or “data”)
  - Notion of abstractness of the entities is important too
  - Examples: baseball statistics, stock trends, connections between criminals, car attributes...

The Challenge of InfoVis

- Transform data into knowledge (understanding, insight) thus making it useful to people
  - To **understand** the data
  - To **make decisions** based on that understanding

- Visual Analytics
  - "new" term - focus on decisions/action
  - **integrate ML and DM**
  - 1 week about this later in semester

Why not just use stats to get answers from data?

Anscombe's Quartet – Tables or Graph?

- **Statistics the same for each** of the four x-y tables (the quartet)
  - Mean = 9
  - Variance = 9
  - Correlation = 0.816
  - Linear regression
  - y = 3 + 0.5x
- So **what's different** about the data?
- You could study the tables very closely
  - And make a little progress
- Or you could graph the data ☺

From F.J. Anscombe, "Graphs in Statistcal Analysis", American Statistician, February 1973, 17-21.

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

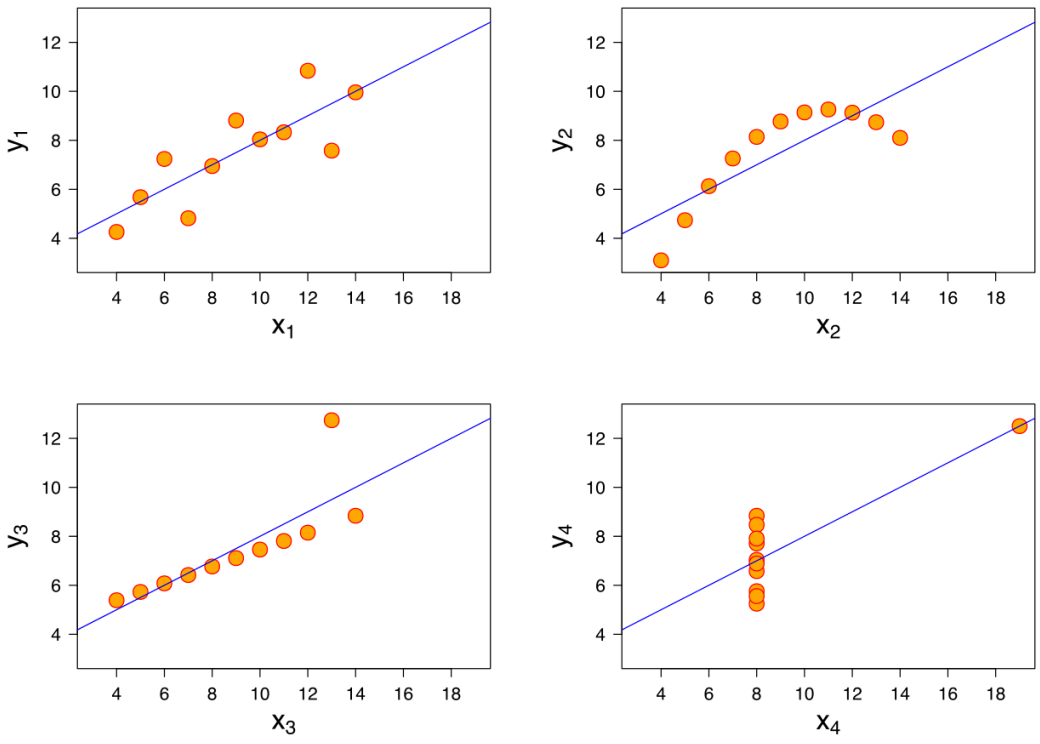


Table - StateData ()

Load | Snap

State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	33.9%	14821
Connecticut	33.8%	20189
Delaware	27.9%	15854
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
Idaho	25.2%	11457
Illinois	26.8%	15201
Indiana	20.9%	13149
Iowa	24.5%	12422
Kansas	26.5%	13300
Kentucky	17.7%	11153
Louisiana	19.4%	10635
Maine	25.7%	12957
Maryland	31.7%	17730
Massachusetts	34.5%	17224
Michigan	23.1%	13748
Minnesota	30.4%	14389
Mississippi	19.9%	9648
Missouri	22.3%	12989
Montana	25.4%	11213
Nebraska	26.0%	12452
Nevada	21.5%	15214
New Hampshire	32.4%	15959
New Jersey	30.1%	18714
New Mexico	25.5%	11246
New York	29.6%	16501
North Carolina	24.2%	12885
North Dakota	28.1%	11051
Ohio	22.3%	13461
Oklahoma	22.8%	11893
Oregon	27.5%	13418
Pennsylvania	23.2%	14068
Rhode Island	27.5%	14981
South Carolina	23.0%	11897
South Dakota	24.6%	10661
Tennessee	20.1%	12255
Texas	25.5%	12904
Utah	30.0%	11029
Vermont	31.5%	13527
Virginia	30.0%	15713
Washington	30.9%	14923

Questions:

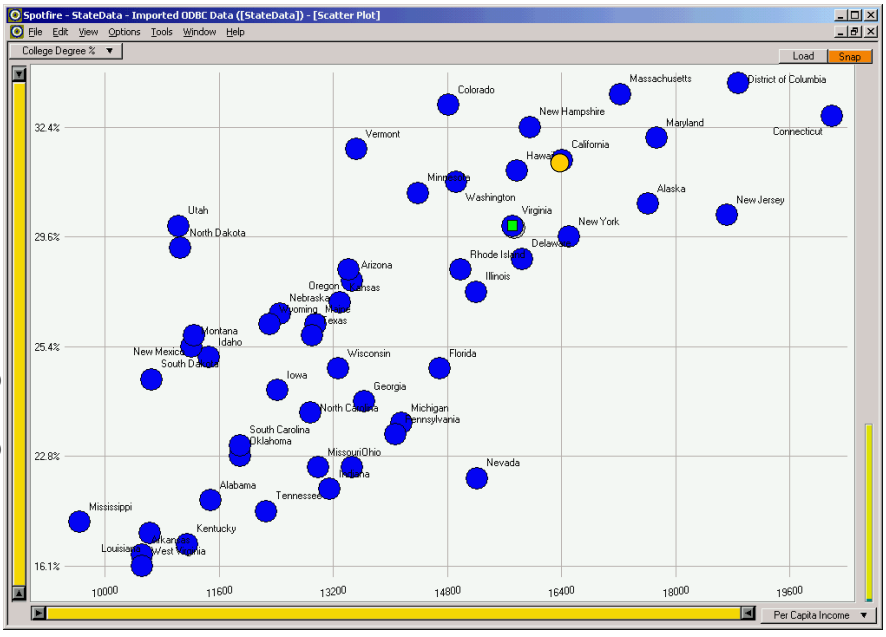
Which state has the highest income?

Is there a relationship between income and education?

Are there any outliers?

Example courtesy of Chris North

Visualize the Data!



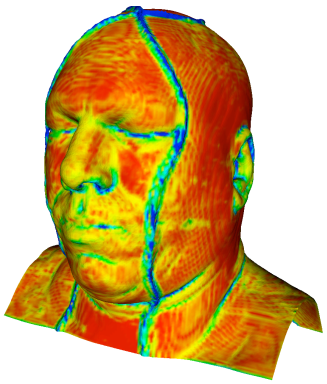
BTW

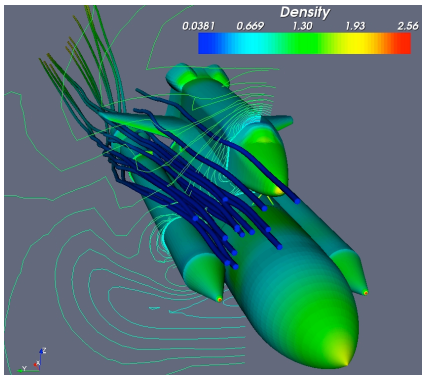
- Information Visualization aka
  - InfoVis
  - InfoViz *(I don't like the 'z' so much though…)*

- Other related terms
  - Infographics (we will talk about the different later)
  - Dashboards
  - Data Visualization

InfoVis ¬= SciVis or MedVis

- **SciVis**: Scientific Data Visualization
- **MedVis**: Medical Data Visualization (body)





SciVis & MedVis

- Data generally associated with physical positions in a 2D or 3D space – has a geometry
- InfoVis – data generally abstract
  - We have to create a geometry or visual metaphor with which the data is encoded
- These topics are not directly covered in this course.

Information Visualization

- Consists of:
  - Taking elements of information without a direct physical correspondence and mapping them to a 2-D or 3-D physical space.
  - Giving the information elements a visual representation that is useful for analysis and decision-making
  - Adding interactions to facilitate exploration
- Also, there are intersections of InfoVis and Computer Graphics, but this course will not focus on those (and also not on computer graphics specifically)
- Again, much more detail in this throughout the semester

Course Overview

Where to find info

- Course syllabus, assignment info, schedule, due dates, policies, etc.
  - http://va.gatech.edu/courses/cs4460/
- Grades, submitting assignments, announcements, (most) datasets will be on **Canvas**

Learning Objectives: Be Able to

- Know how to **design and implement** information visualizations.
- **Use design principles** to create effective information visualizations.
- **Choose appropriate types of visualizations** for various types of data and for different goals.
- Choose appropriate **interaction methods**.
- Apply understanding of **human perceptual and cognitive capabilities** to the design of information visualizations.
- **Critiquing different visualization techniques** in the context of user goals and objectives.
- **Implement** compelling Information Visualizations using d3

Course policy highlights

- **Late homework submissions will lose 25% each day** without officially documented personal issues (serious illness, family emergency, etc.)

- Review the Georgia Tech Academic Honor Code
  - http://www.deanofstudents.gatech.edu/Honor/
  - Individual HWs and Ps done individually (if collaboration is allowed, it will be explicitly specified)

- Attendance expected — there are pop quizzes.

Device Usage in Class

- During lectures please put your smart phones away, and use your computers and tablets for course-relevant work only.
- Games, browsing, emailing is distracting not only to you, but to those around you.
  - Remember, messengers and notifications are designed to grab your attention and are de-facto irresistible.
- BTW, even note-taking by hand versus on your computer has been shown to be more efficient for learning (also see this news story). But that's your call to make.

Texts and Readings

- There is no required textbook for this course.
- For some lectures, we will read articles, try out online visualizations, or watch videos before the lectures.
- Reading material to be look at before class is available online.
  - Note that some material will require you to be on the GT network.
- mostly, showing up and regularly reviewing your notes and slides is a good habit to get into to succeed in this course

Grading

- Point distribution - let's look at the website
- Cutoffs start with normal grading scale
  - >=90 —> A
  - 80..<90 —> B
  - …
- Final grades may be curved (only to your advantage, not down). CURVING IS NOT GUARANTEED!
- Final grades in past classes do include Cs, Ds, and Fs…
- Participation and attendance matter if you are close to a boundary

Web Sites

- Class schedule, assignments, PPts:
  - http://va.gatech.edu/courses/cs4460/
- Piazza for conversation, questions, etc.
  - Not only between you and the instructor of TAs, but **among your classmates**
  - **Use it :)**

Lecture Notes and Slides

- Slides will be posted to Canvas (under the resources tab)
  - within the week of the lecture (usually on Fridays)
- Word of caution
  - Slides often show visualizations, not the discussion about the visualization technique or guideline we're covering.
  - Take notes about our in-class discussion and lecture.
  - **Only the slides will not be enough** for you to learn the material (and do well on the exams).
  - In short, come to class, listen, participate, learn, get good at InfoVis!

Homework (HW) and Programming Homework (P)

- 5 HW assignments
  - design questions, surveys, trying out a vis
- 5 Programming assignments
  - help you learn how to program using d3
- *not all assignments require the same amount of time/effort (not all are worth the same amount of point). Look and plan ahead.*

Schedule

- Due dates and topics for the course assignments may change slightly this week (except for HW1).
- Exam schedule will not change!
  - You have to be here at the scheduled exam time! DO NOT BOOK YOUR TICKETS UNTIL YOU CHECK!

Active Learning

- This course leverages in-class activities and discussion
- Research studies on active learning show greater student retention and understanding of material
- Size of this classroom is challenging, but we will make it work

Contacting us

- To contact me, send me an email
  - endert@gatech.edu
  - **Include in the subject of the email: [CS4460]**
- To ask a general question”
  - Post a public post to Piazza
  - The whole class can see these
- To ask a specific question (about an assignment, for example)
  - Post a private post to Piazza
  - All TAs and instructor can see these.

Warning

- Do not wait until a few days before an assignment it due to ask questions
  - Instructors and TAs may not be able to answer them before the deadline. This is not a valid reason for a deadline extension.
  - Start early, ask questions early, finish early, lower your stress!
- Do not wait until the end of the semester to realize that your grade in the class is not what you need.
  - This is especially important for those of you graduating this semester!

Let's take a look at the website

- va.gatech.edu/courses/cs4460
-

Active Learning

- Talk with your neighbors about course information.
  - What is one aspect of the course structure (HW, programming, exams, …) that both of you expected the class to have?
  - What is the part of the class that you think will take the most amount of work on your part, and why?
- Take 2 minutes.
  - I will call on a few of you, so be ready.

- *BTW, there will be many of these in-class discussions and conversations. Come to class!*
  - *The content of these conversations will be on the test, and often do not show up explicitly in the slides*

Next time

- InfoVis Overview
- No assignments due