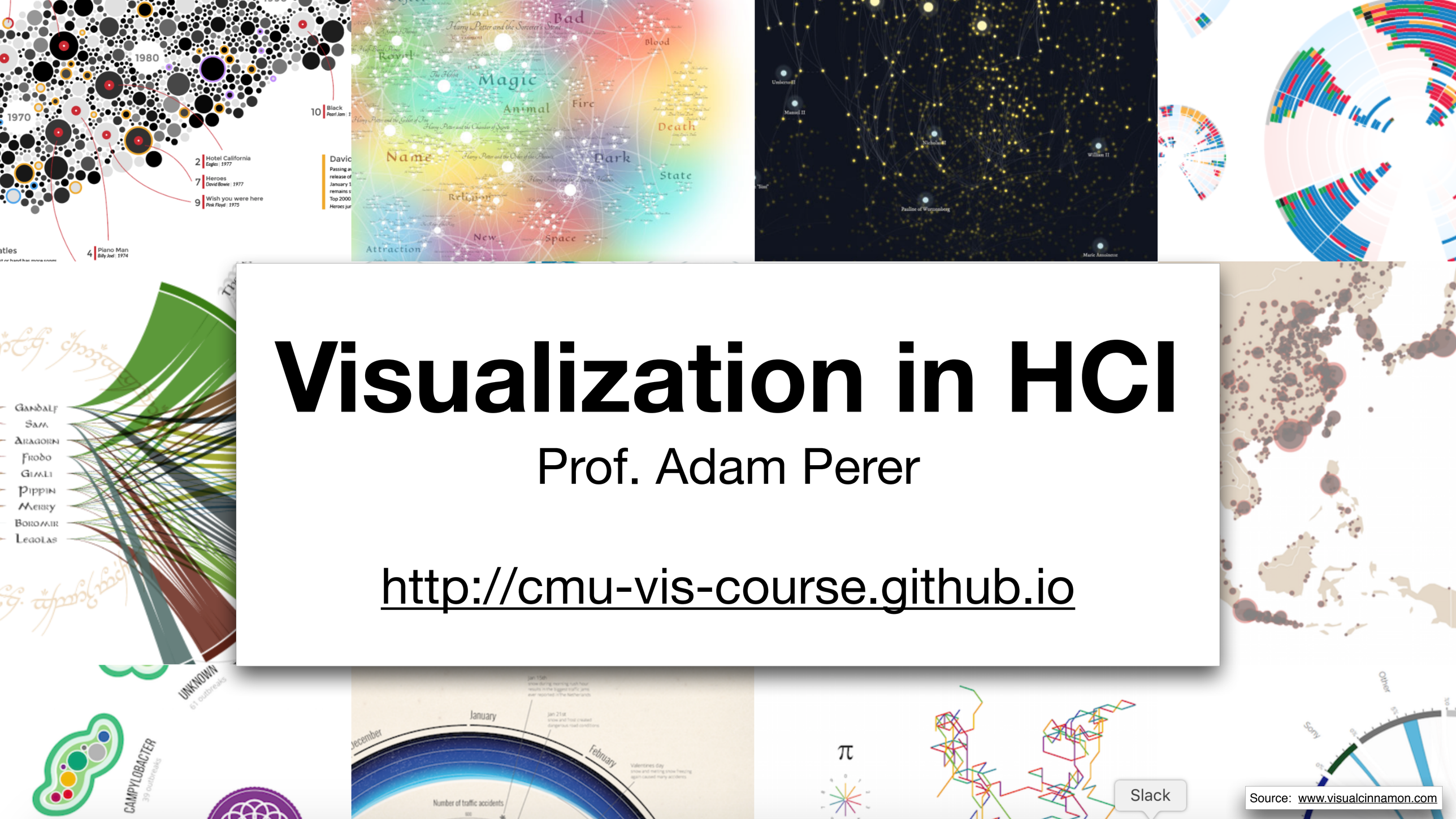


Visualization in HCI

Prof. Adam Perer

http://cmu-vis-course.github.io

Source: www.visualcinnamon.com

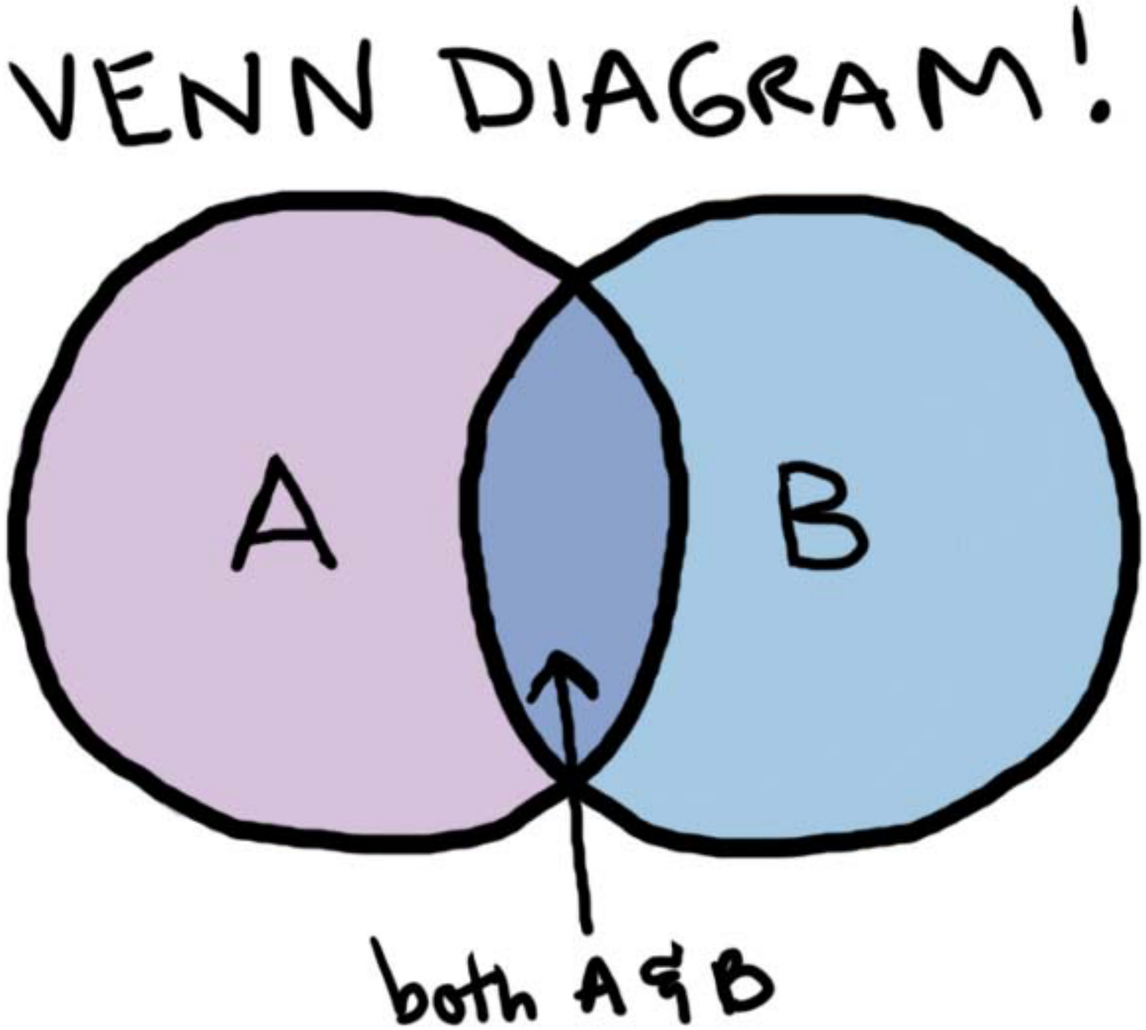
visualization *pictures*

The purpose of computing is insight, not numbers.

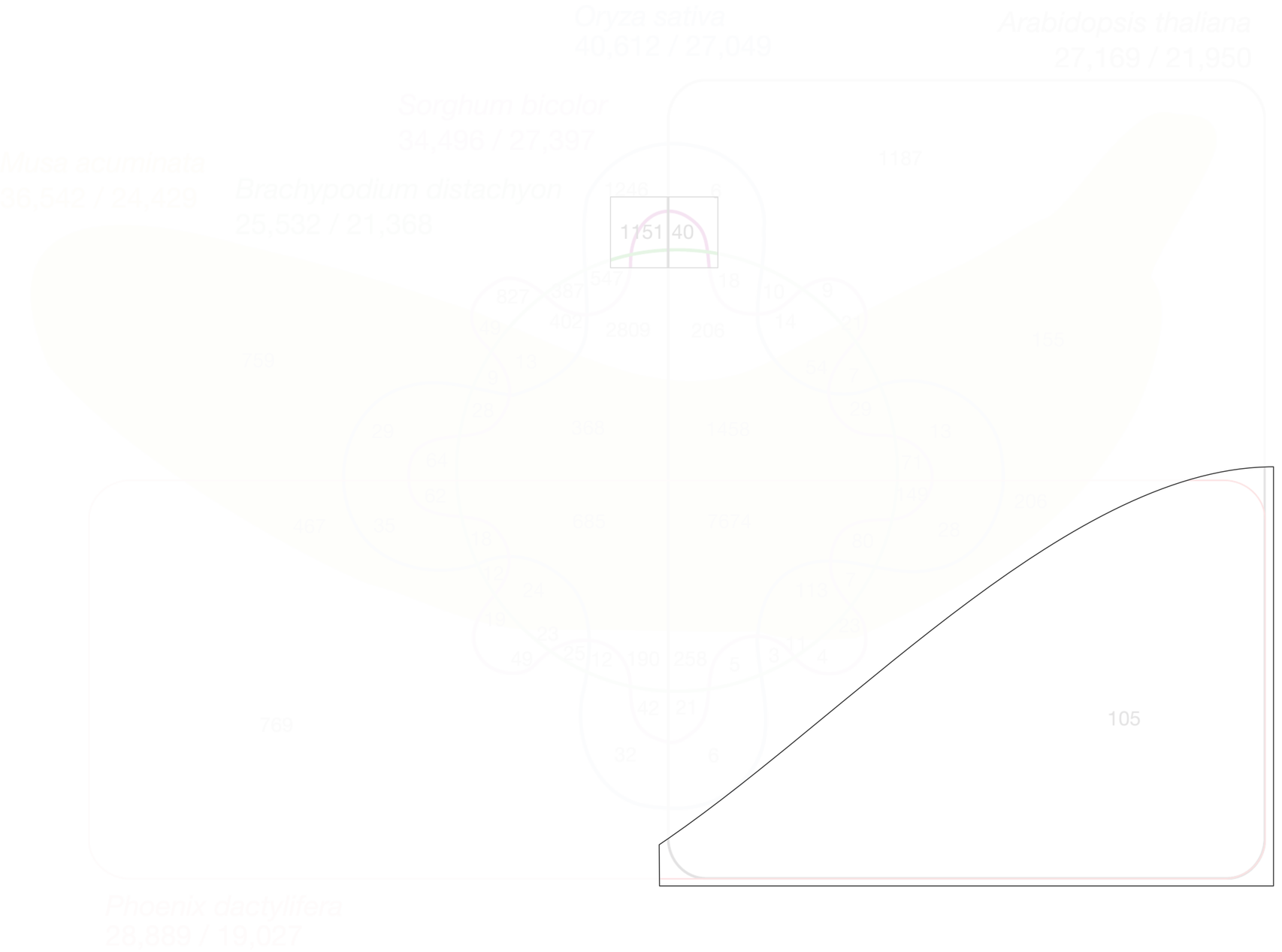
- Richard Wesley Hamming

- Card, Mackinlay, Shneiderman

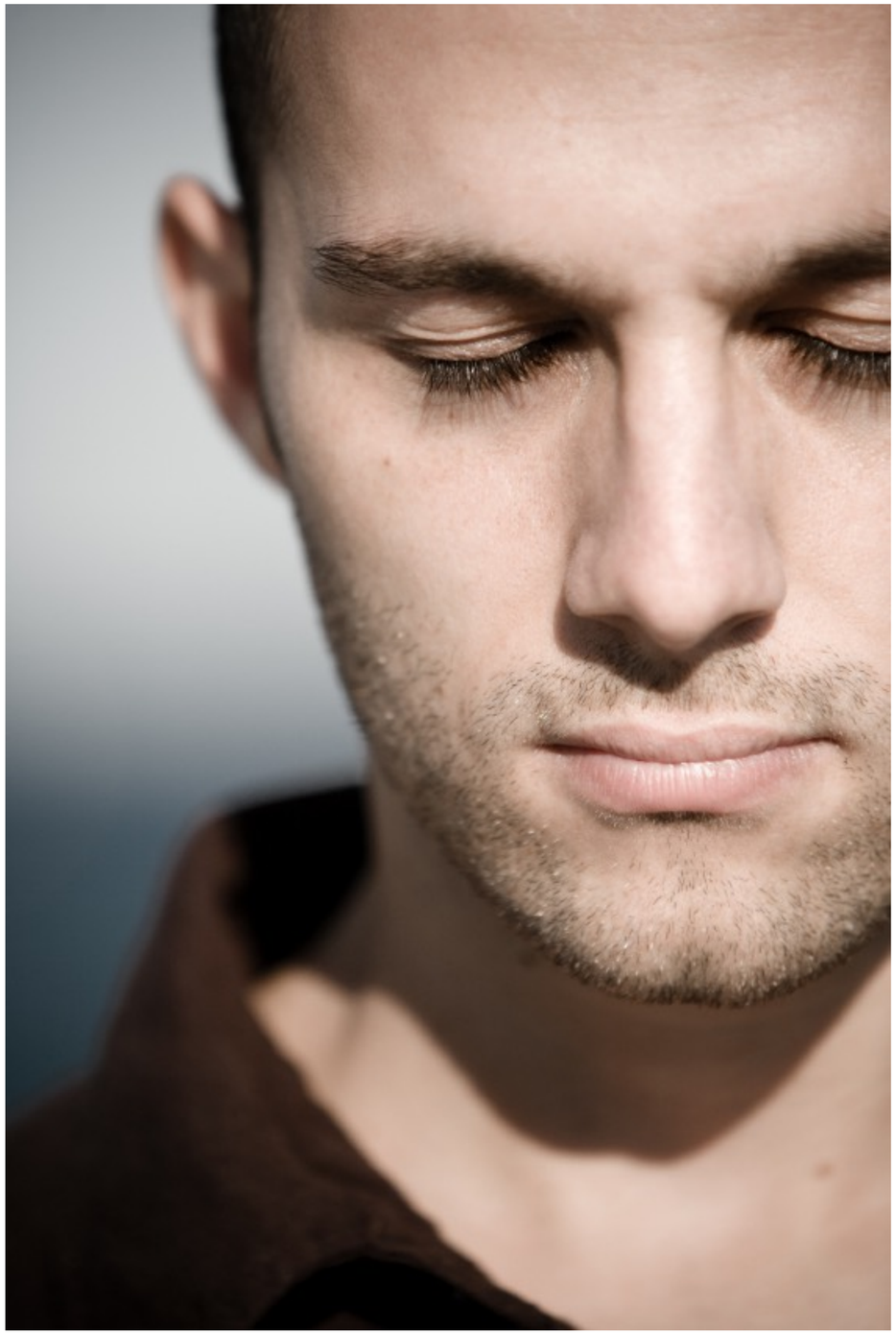
Banana	***M. acuminata***
Date	***P. dactylifera***
Cress	***Arabidopsis thaliana***
Rice	***Oryza sativa***
Sorghum	***Sorghum bicolor***
Brome	***Brachypodium distachyon***



Source: http://math.stackexchange.com/questions/682117/venn-diagram-of-conditional-probability-and-extending-to-bayes-theorem



[D'Hont et al., Nature, 2012]



vi · su · al · i · za · tion

1. Formation of mental visual images
2. The act or process of interpreting in visual terms or of putting into visible form

The American Heritage Dictionary

Visualization Definition

Visualization is the process that **transform**s
(abstract) **data** into
interactive graphical representations for the purpose of
exploration, confirmation, or presentation.

Good Data Visualization

… makes data accessible

… combines strengths of humans and computers

… enables insight

… communicates

Visualization

“Visualization is really about external cognition, that is, how resources outside the mind can be used to boost the cognitive capabilities of the mind.”

Stuart Card

Why Visualize?

To inform people: **Communication**

How is X ahead in the election polls?

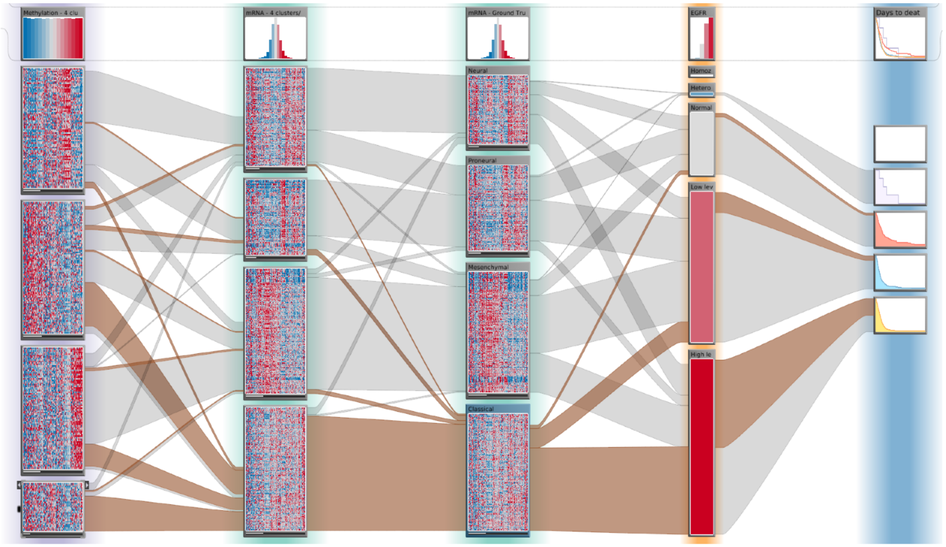
When questions are not well defined: **Exploration**

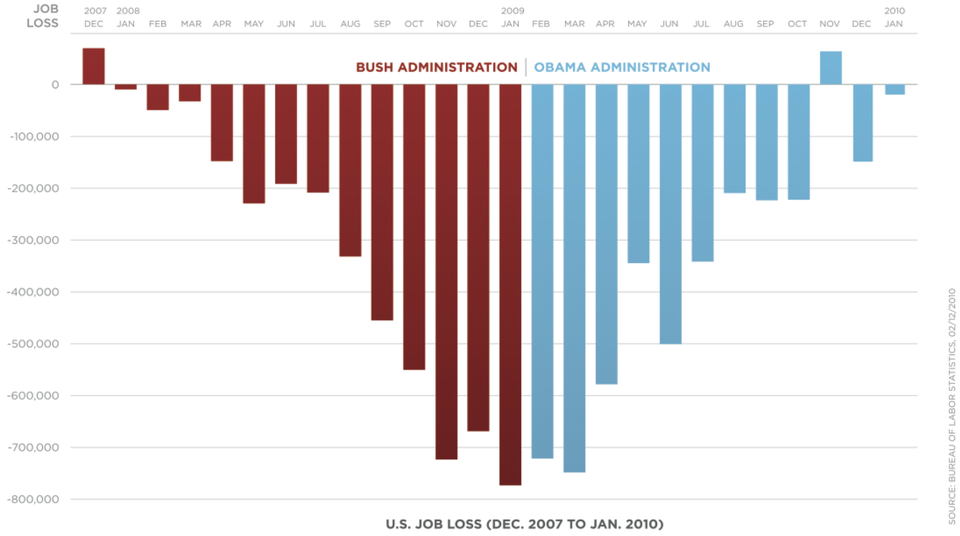
What is the structure of a terrorist network?

Which drug can help patient X?

Purpose of Visualization

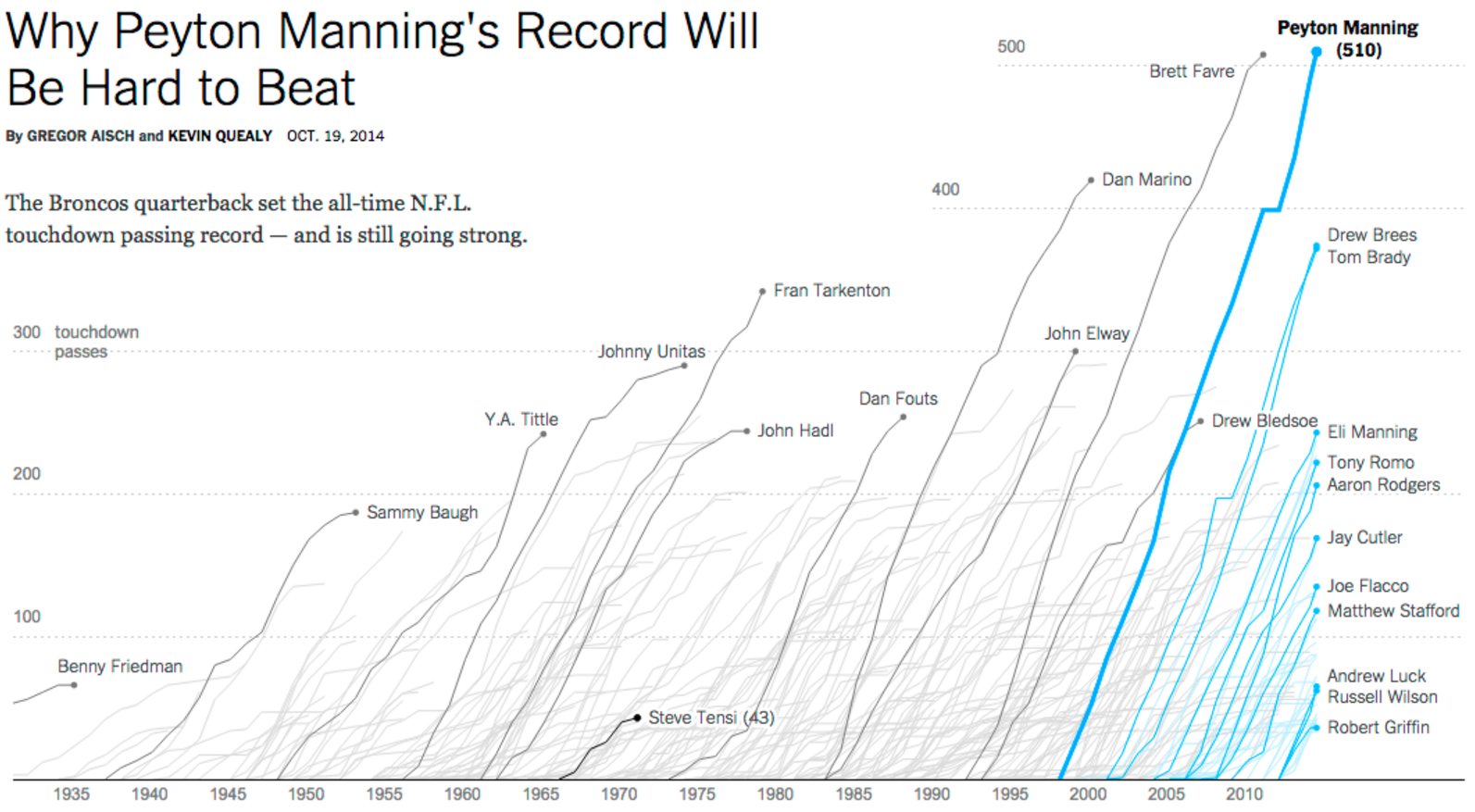
[Obama Administration]





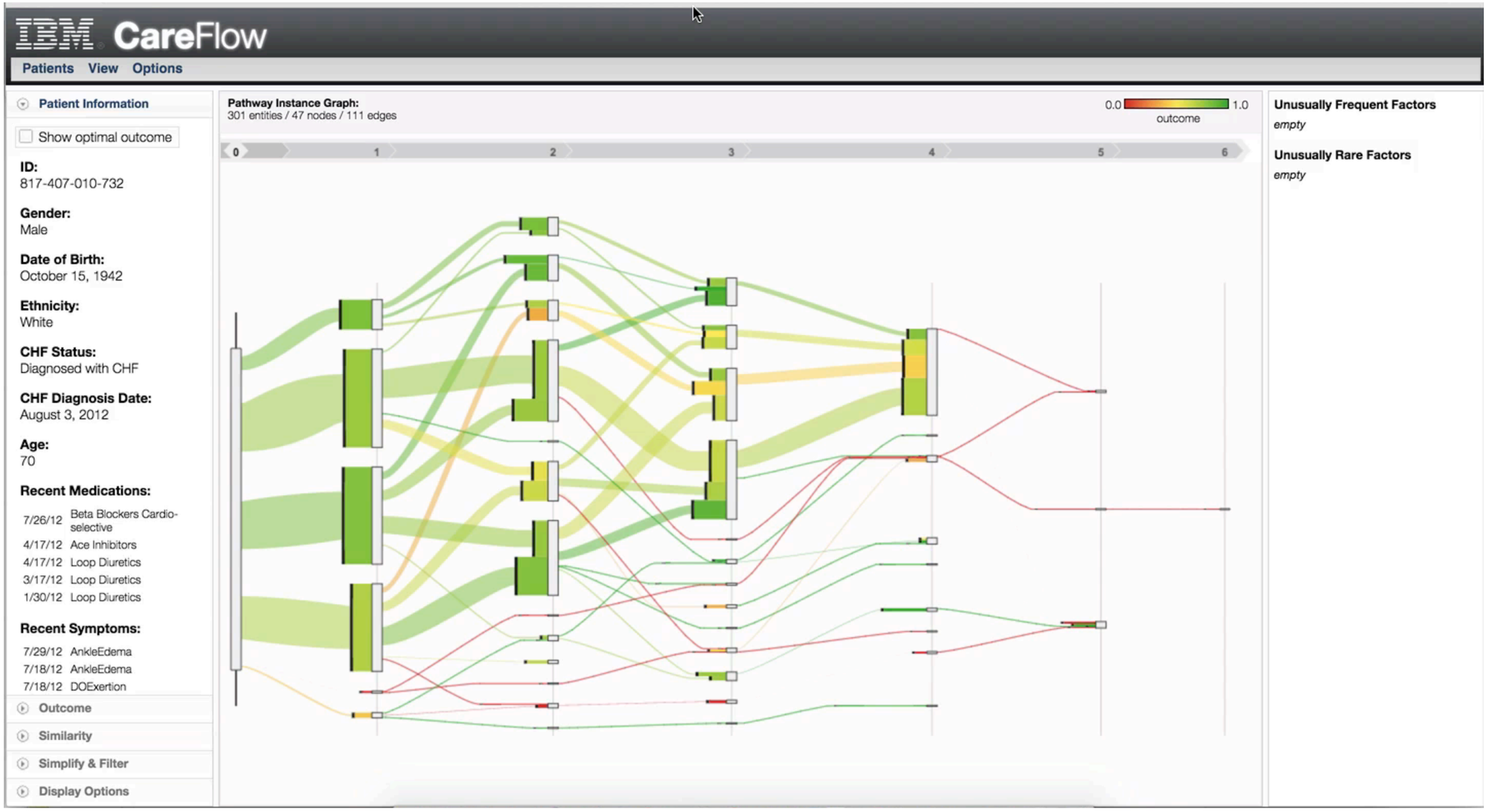
Open Exploration — Confirmation — Communication

Example Communication



[New York Times]

Example Exploration:



[CareFlow, CHI 2013]

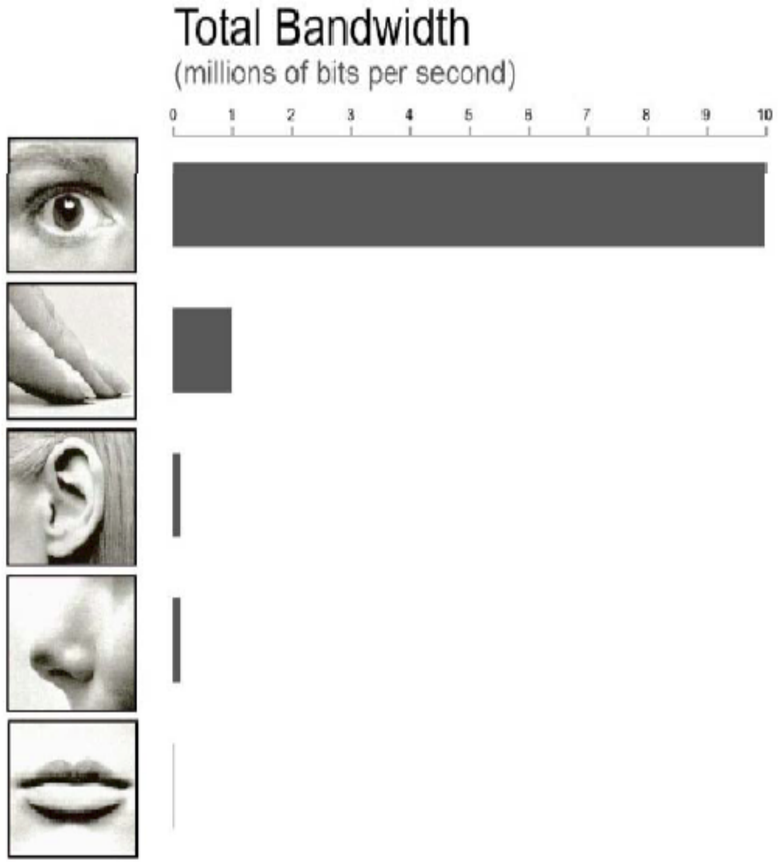
Why Graphics?

Figures are richer; provide more information with less clutter and in less space.

Figures provide the *gestalt* effect: they give an overview; make structure more visible.

Figures are more accessible, easier to understand, faster to grasp, more comprehensible, more memorable, more fun, and less formal.

list adapted from: [Stasko et al. 1998]



Campanella then switched to an identically constructed map, only this time based on 2010 census data, and in bits and pieces on the screen there was a simple and arresting picture of what Katrina meant. In the neighborhoods that were once a dense black, many of the little squares had thinned and turned gray. The sharp lines that once separated the teapot from Central City were now blurry: the white areas of the city were pushing north, into the vacuum left by the exodus. The Bywater was graying, as it gentrified still further. "Before Katrina, an American Community Survey estimate of New Orleans Parish population was four hundred and fifty-five thousand, and about sixty-eight per cent black," Campanella said. "Now the latest estimate is three hundred and eight-four thousand, and it's about sixty per cent black. The white population has largely reconstituted itself numerically. So if you do that math, we're talking about seventy-nine to eighty thousand fewer African-Americans." In the most traumatic manner possible, Katrina caused what social scientists normally consider a good thing—a major move—and that is why, a decade after it devastated the Gulf Coast, the storm continues to defy simple categorization.

New Yorker
http://www.newyorker.com/magazine/2015/08/24/starting-over-dept-of-social-studies-malcolm-gladwell

When not to visualize?
When to automate?

Well defined question on well-defined dataset

Which gene is most frequently mutated in this set of patients?

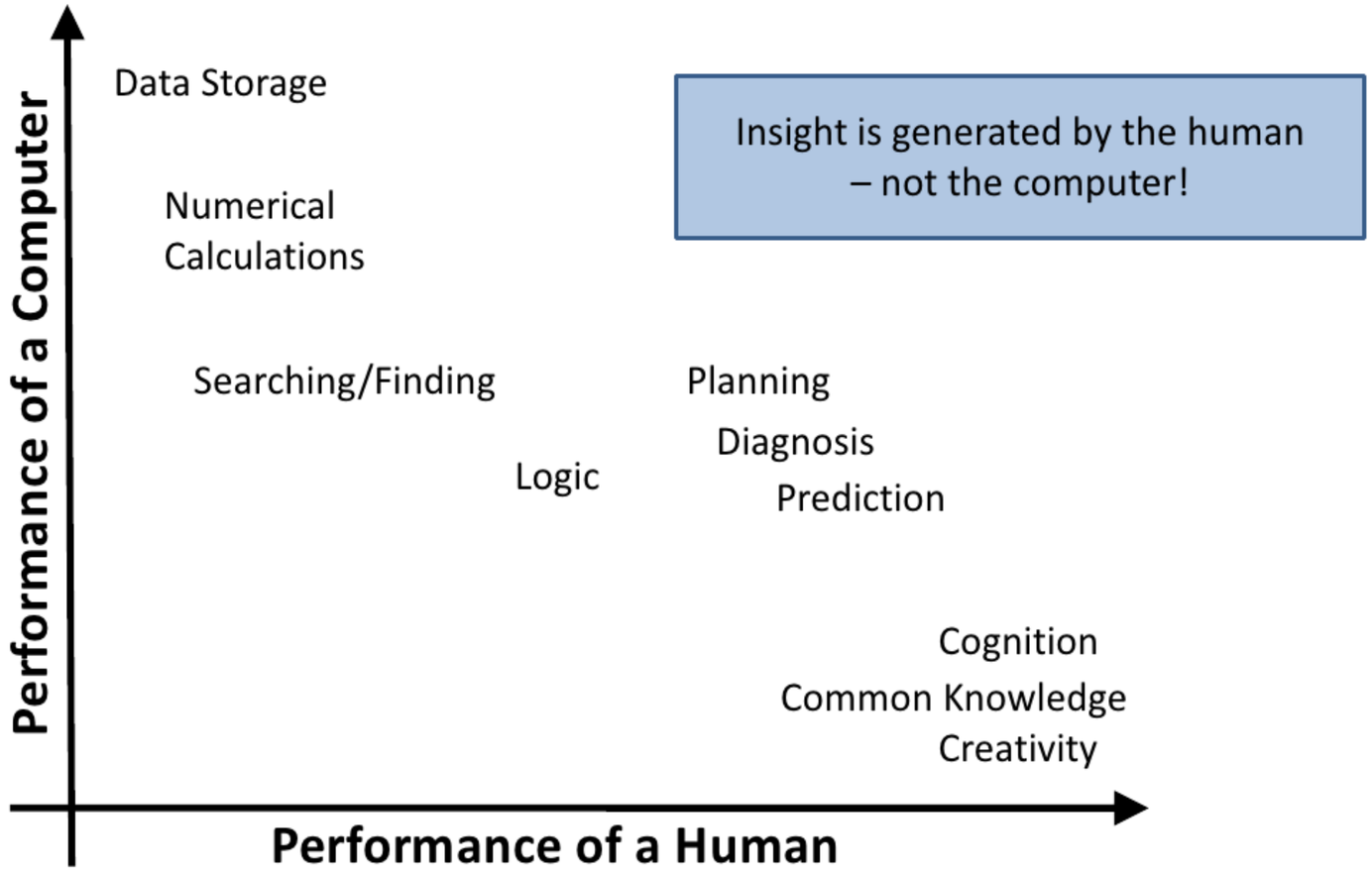
What is the current unemployment rate?

Decisions needed in minimal time

High frequency stock market trading: which stock to buy/sell?

Manufacturing: is bottle broken?

The Ability Matrix



Why Use Computers?

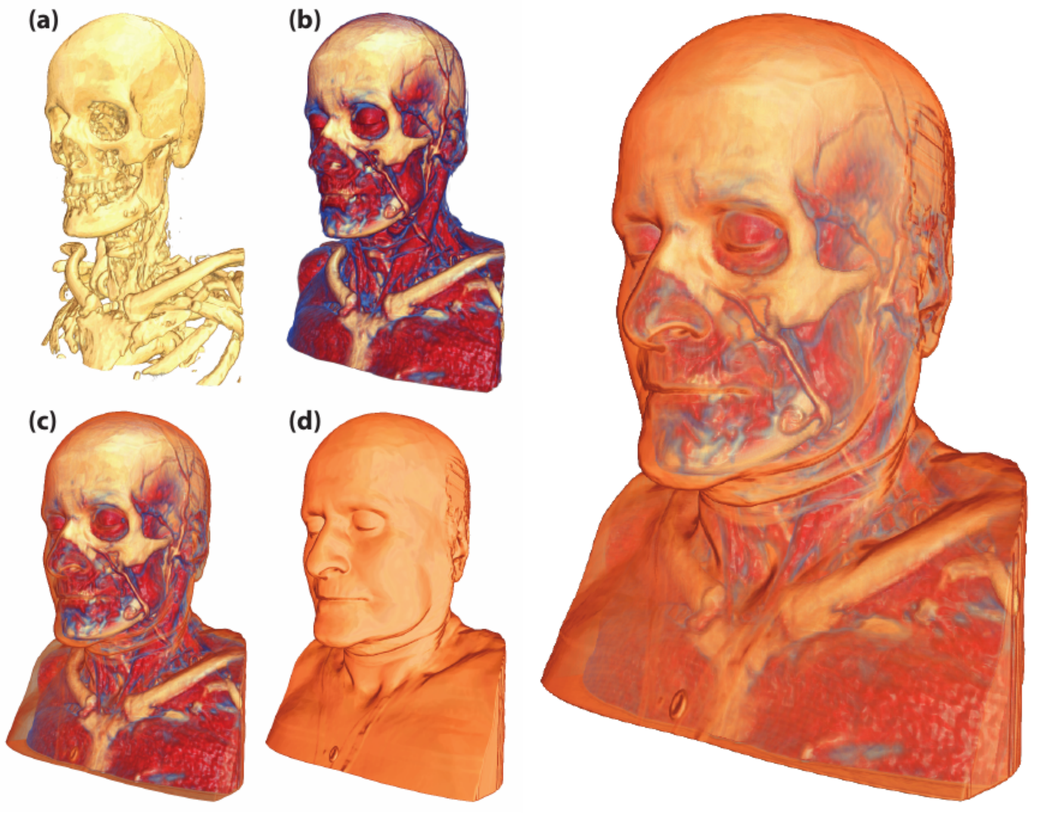
Scale

Drawing by hand (or Illustrator)

infeasible

inflexible (updates!)

How to draw an MRI scan?



[Bruckner 2007]

Why Use Computers?

Interaction

Interaction allows to “drill down” into data

Integration

Integration with algorithms

Make visualization part of a data analysis pipeline

[Sunburst by John Stasko, Implementation in Caleydo by Christian Partl]

Why Use Computers?

Efficiency

Re-use charts / methods for
different datasets

Quality

Precise data driven rendering

Storytelling

Use time

Tell Stories

[New York Times]

Why not just use Statistics?

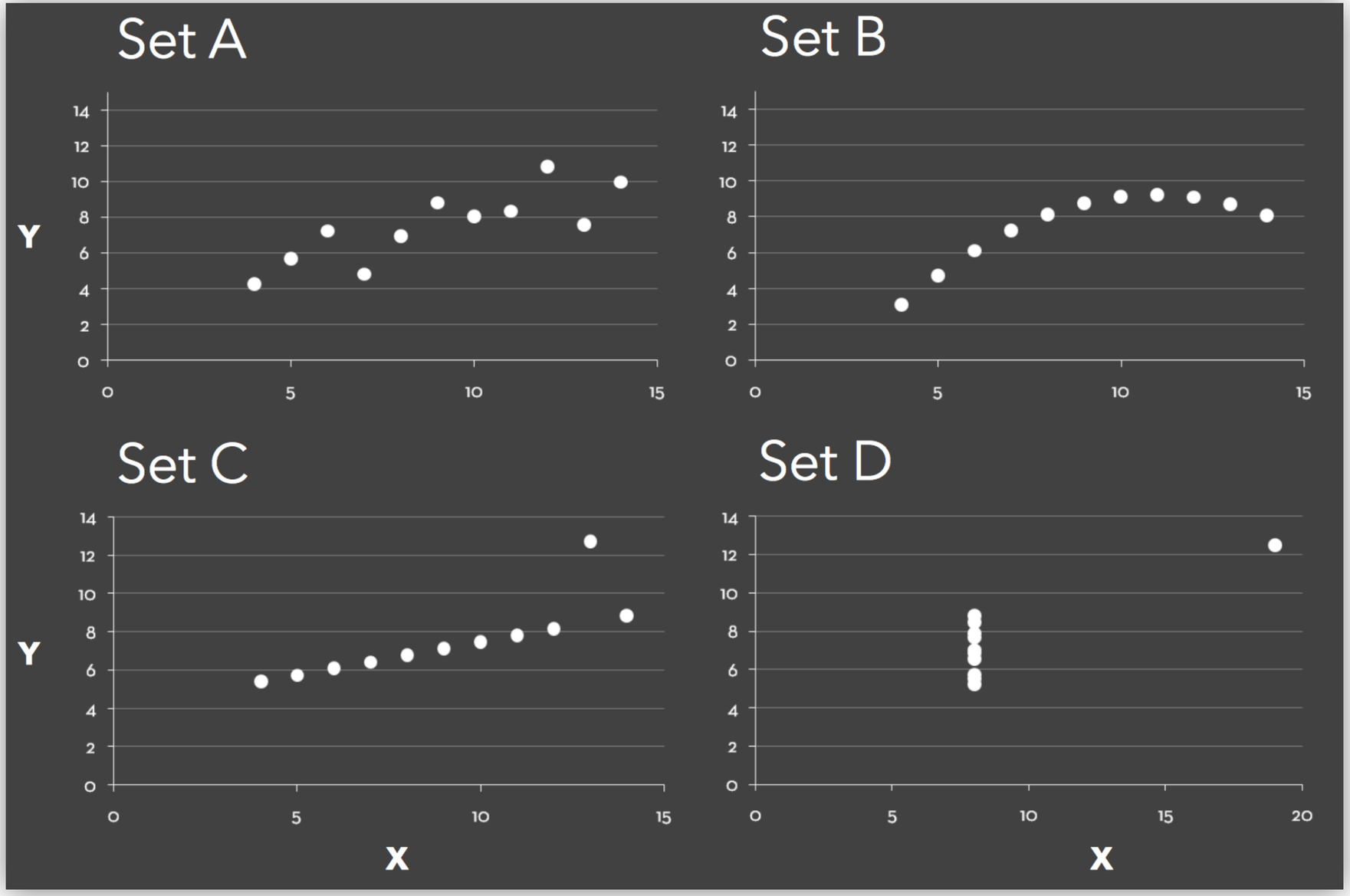
Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics	**Linear Regression**
$u_X = 9.0$ $\sigma_X = 3.317$	$Y = 3 + 0.5\ X$
$u_Y = 7.5$ $\sigma_Y = 2.03$	$R^2 = 0.67$

[Anscombe 1973]

Anscombe's Quartet



Data

Visualization in the Data Science Process

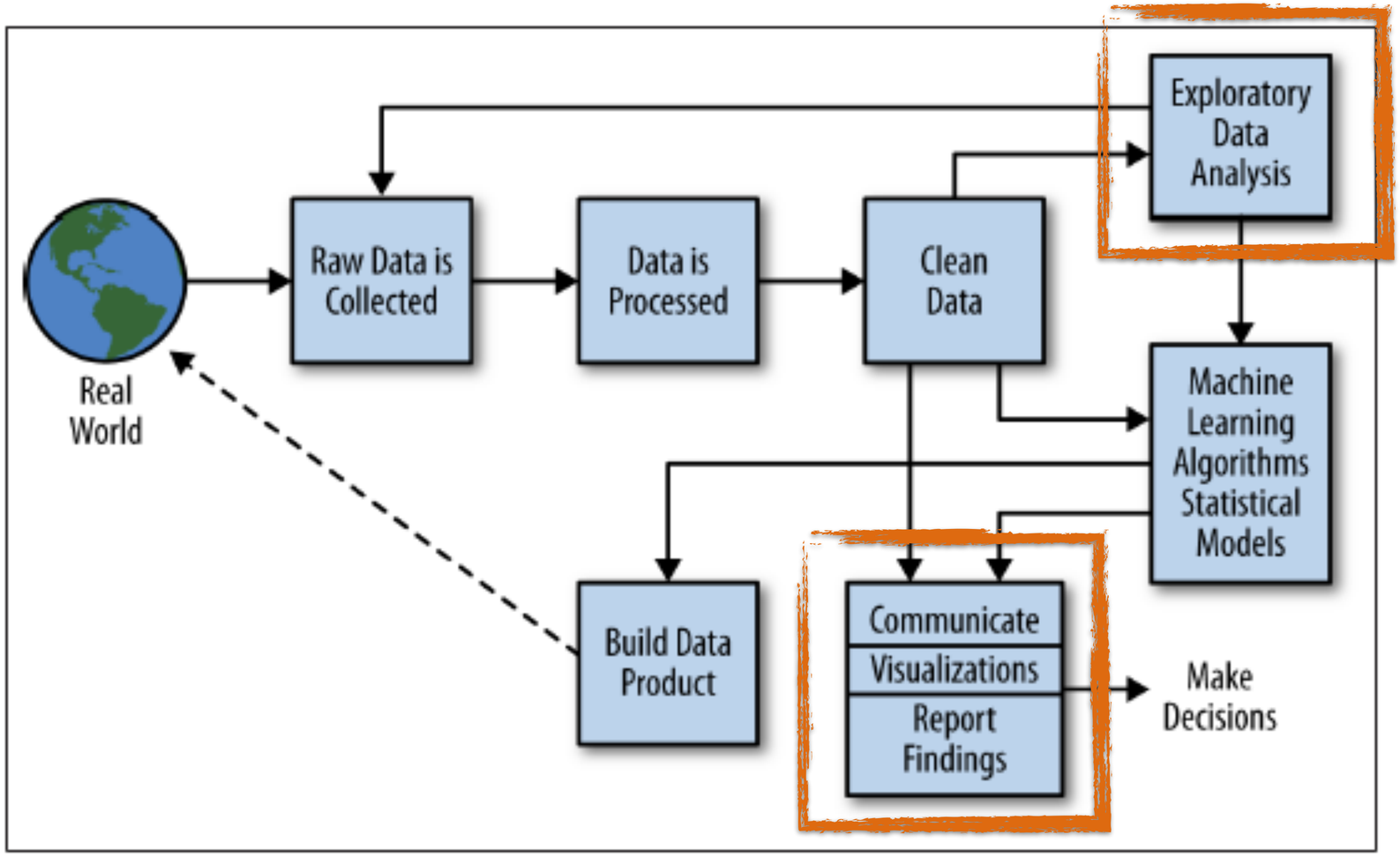


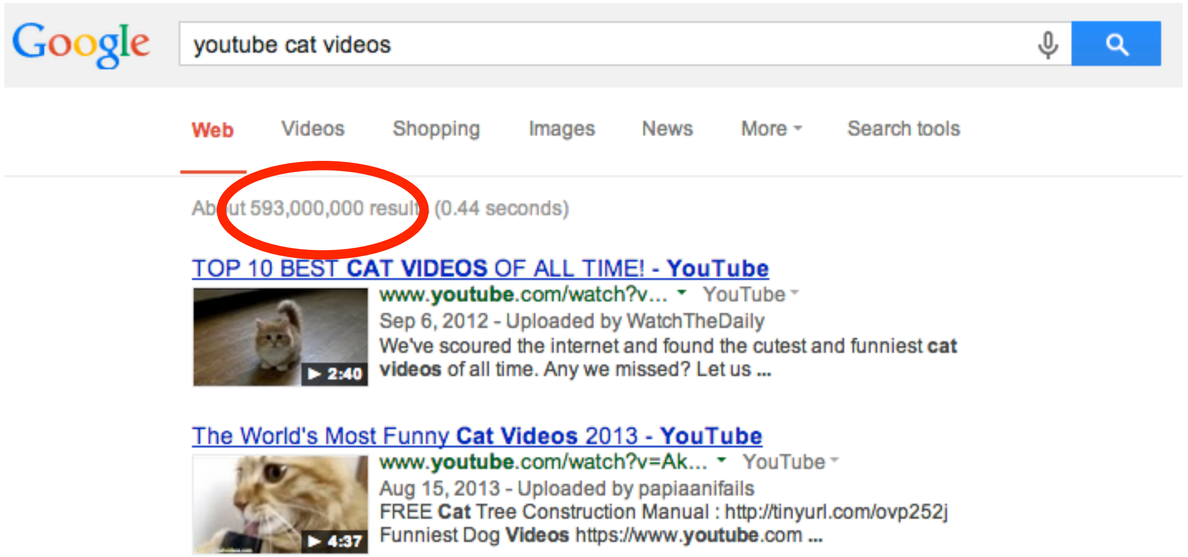
Figure 2-2. The data science process

Big Data

2010: 1,200 exabytes, largely unstructured

Google stores ~10 exabytes (2013)

Hard disk industry ships ~8 exabytes/year



15 Exabytes in Punch Cards:
4.5 km over New England

In one second on the Internet there are...

http://onesecond.designly.com/

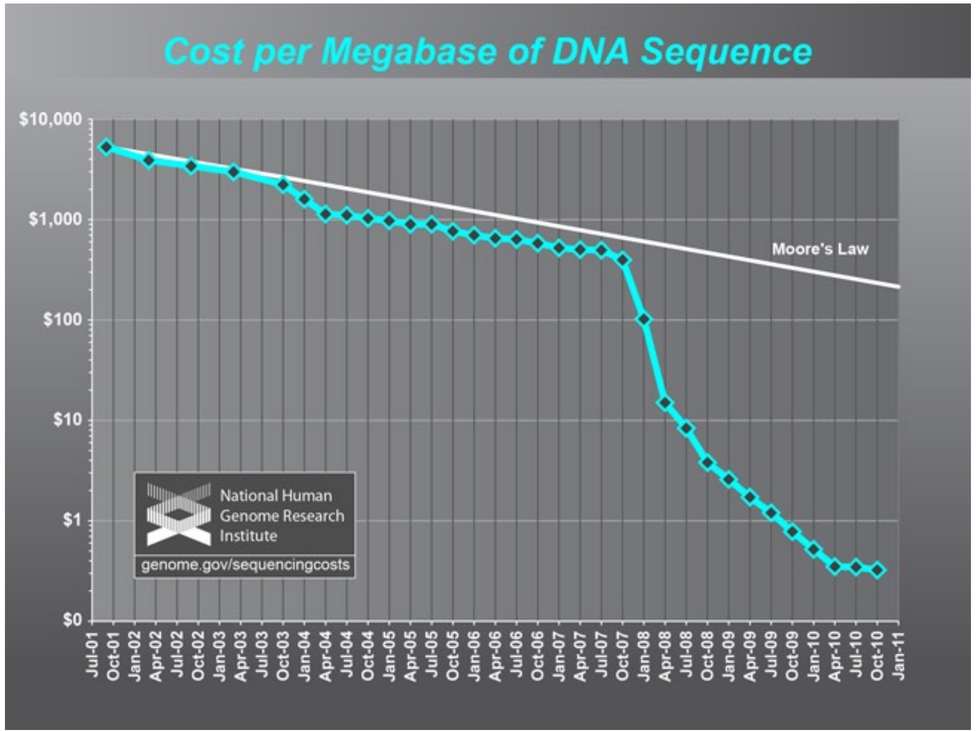
Big Data in Science and Engineering

“Big Data” hasn’t just transformed industry!

It’s also transformed science and engineering. Cheap sensors (e.g. imaging) have changed the way science and engineering are done.

Examples:

- Large physics experiments and observations
- Cheaper and automated genome sequencing
- Smart buildings / cities (blyncsy)
- Geophysical imaging



Example: CERN Large Hadron Collider Data

CERN has publicly released over 300TB of data: CERN Open Data Portal

How much is that?

- At 15 GB of storage a piece, you'd need 20,000 Gmail accounts to store the whole shebang. If you wanted to send that much data at the max attachment size of 25 MB, it would take you 12 million emails.
- A DVD-R holds 4.7 GB. You'd need 63,830 of them to hold 300 TB.
- Your Blu-ray collection wouldn't need to expand quite so much. 6,000 discs ought to hold it.
- It takes Pandora about a day and a half to burn through a gig of mobile data. So if the CERN data was an album, you could stream it in just over 1,230 years.
- At 350 MB per hour for 4K video streaming, so if the CERN data was a 4K movie it'd probably be about 857,142 hours, or about 98 years long.
- But its no big deal to the National Security Agency works with. Going by 2013 figures the agency released, the NSA's various activities "touch" 300 TB of data every 15 minutes or so

(Popular Mechanics Article)

NSA Utah Data Center

Storage Capacity?

estimates vary, but Forbes magazine estimates 12 exabytes (12,000 petabytes or 12 million terabytes)

“The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that’s going to be a hugely important skill in the next decades, … because now we really do have **essentially free and ubiquitous data**.”

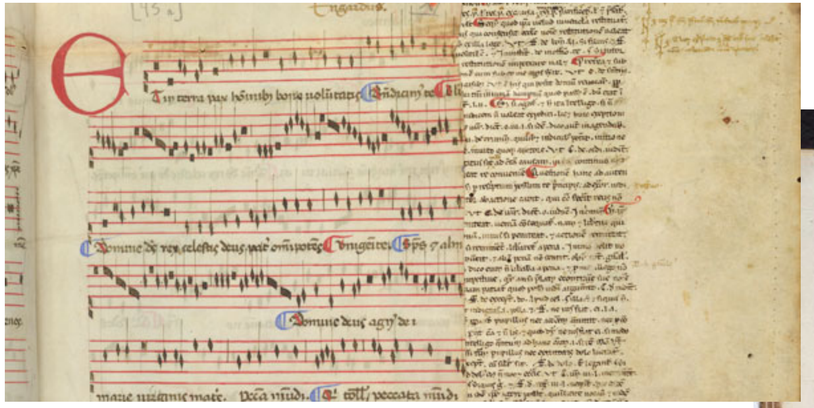
Hal Varian, Google’s Chief Economist
The McKinsey Quarterly, Jan 2009

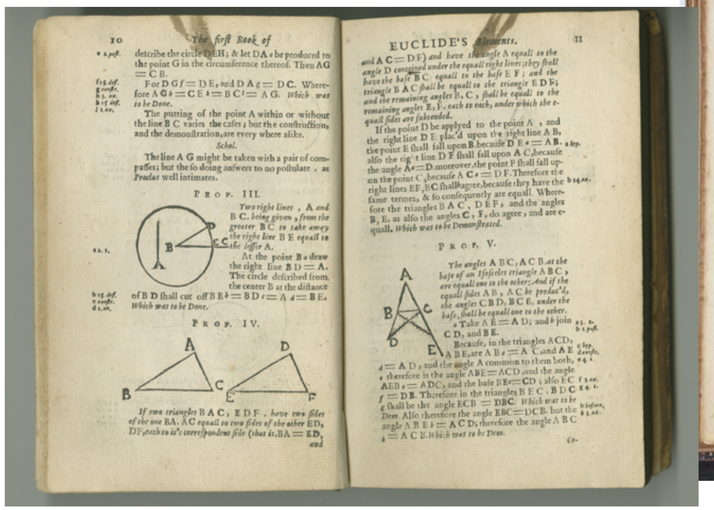
How did we get here?

A bit of history

“It is things that make us smart”

Donald A. Norman





10 *The first Book of*

detcribe the circle GEH; & let DA, e be produced to the point G in the circumference thereof. Then AG = CB.

For DG (= DE,) and DA (= DC. Wherefore AG (= CE (= BC (= AG. *Which was to be Done.*

The putting of the point A within or without the line BC varies the cases; but the construction, and the demonstration, are every where alike.

Schol.

The line AG might be taken with a pair of compasses; but the so doing answers to no postulate, as *Proclus* well intimates.

PROP. III.

Two right lines, A and BC, being given, from the greater BC to take away the right line BE equall to the other line A.

At the point B a draw the right line BD = A. The circle described from the center B at the distance of BD shall cut off BE (= BD (= A (= BE. *Which was to be Done.*

PROP. IV.

If two triangles BAC, EDF, have two sides of the one BA, AC equall to two sides of the other ED, DF, each to it's correspondent side (that is, BA = ED, and

EUCLIDE'S *Elements.* 11

and AC = DF) and have the angle A equall to the angle D contained under the equall right lines; they shall have the base BC equall to the base EF, and the triangle BAC shall be equall to the triangle EDF; and the remaining angles B, C, shall be equall to the remaining angles E, F, each to each, under which the equall sides are subtended.

If the point D be applyed to the point A, and the right line DE plac'd upon the right line AB, the point E shall fall upon B, because DE (= AB. Also the right line DF shall fall upon AC, because the angle A (= D, moreover, the point F shall fall upon the point C, because AC (= DF. Therefore the right lines EF, BC shall agree, because they have the same termes, & so consequently are equall. Wherefore the triangles BAC, DEF, and the angles B, E, as also the angles C, F, do agree, and are equall, *Which was to be Demonstrated.*

PROP. V.

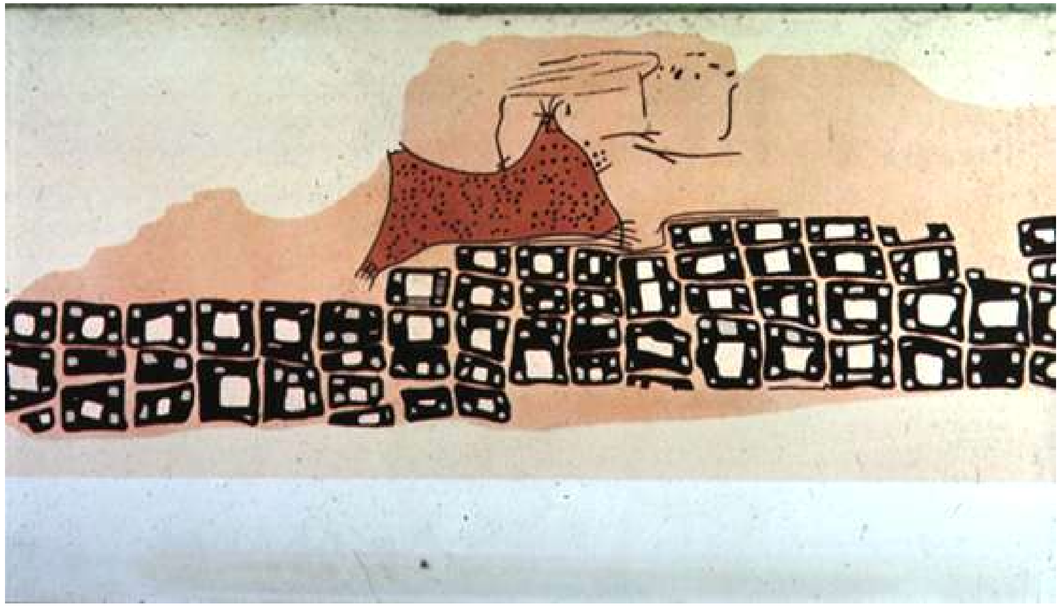
The angles ABC, ACB, at the base of an Isosceles triangle ABC, are equall one to the other; And if the equall sides AB, AC be produc'd, the angles CBD, BCE, under the base, shall be equall one to the other.

Take AE = AD; and join CD, and BE.

Because, in the triangles ACD, ABE, are AB (= AC, and AE (= AD, and the angle A common to them both, therefore is the angle ABE = ACD and the angle AEB (= ADC, and the base BE = CD; also EC (= DB. Therefore in the triangles BEC, BDC shall be the angle ECB = DBC. *Which was to be Dem.* Also therefore the angle EBC = DCB: but the angle ABE = ACD; therefore the angle ABC (= ACB. *Which was to be Dem.*

Co-

Record

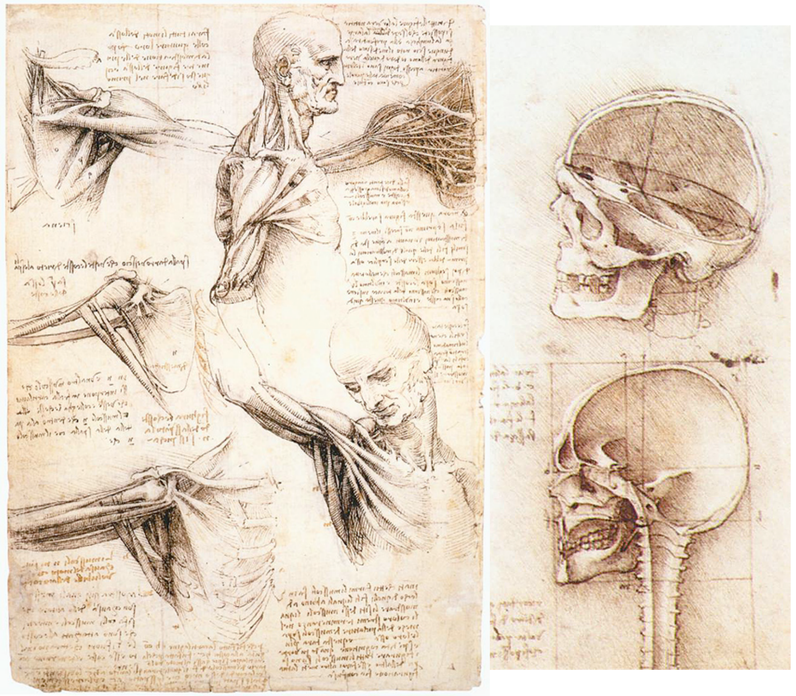


Konya town map, Turkey, c. 6200 BC

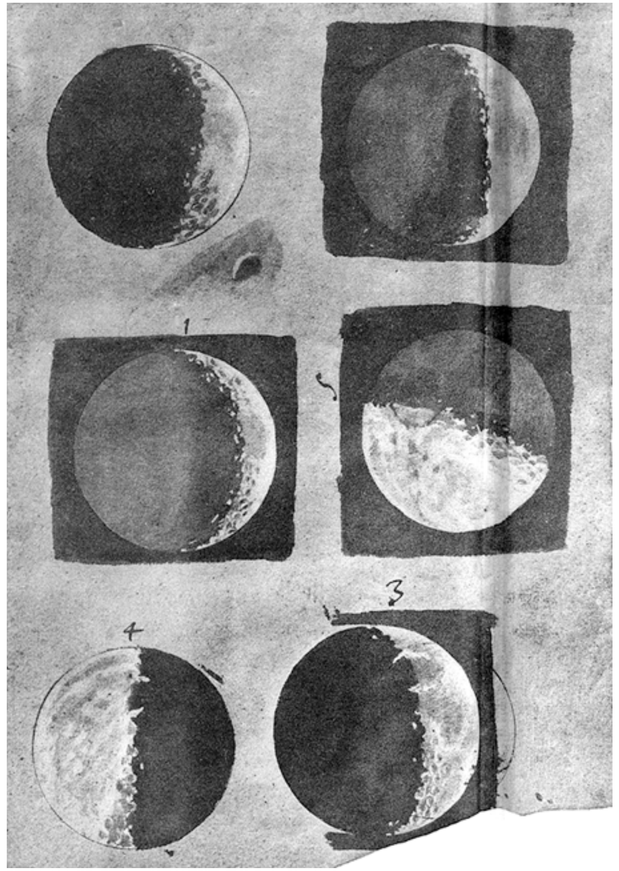
Anaximader's Map of the World

Anaximander of Miletus, c. 550 BC

Record



Leonardo Da Vinci, ca. 1500



Galileo Galilei, 1616

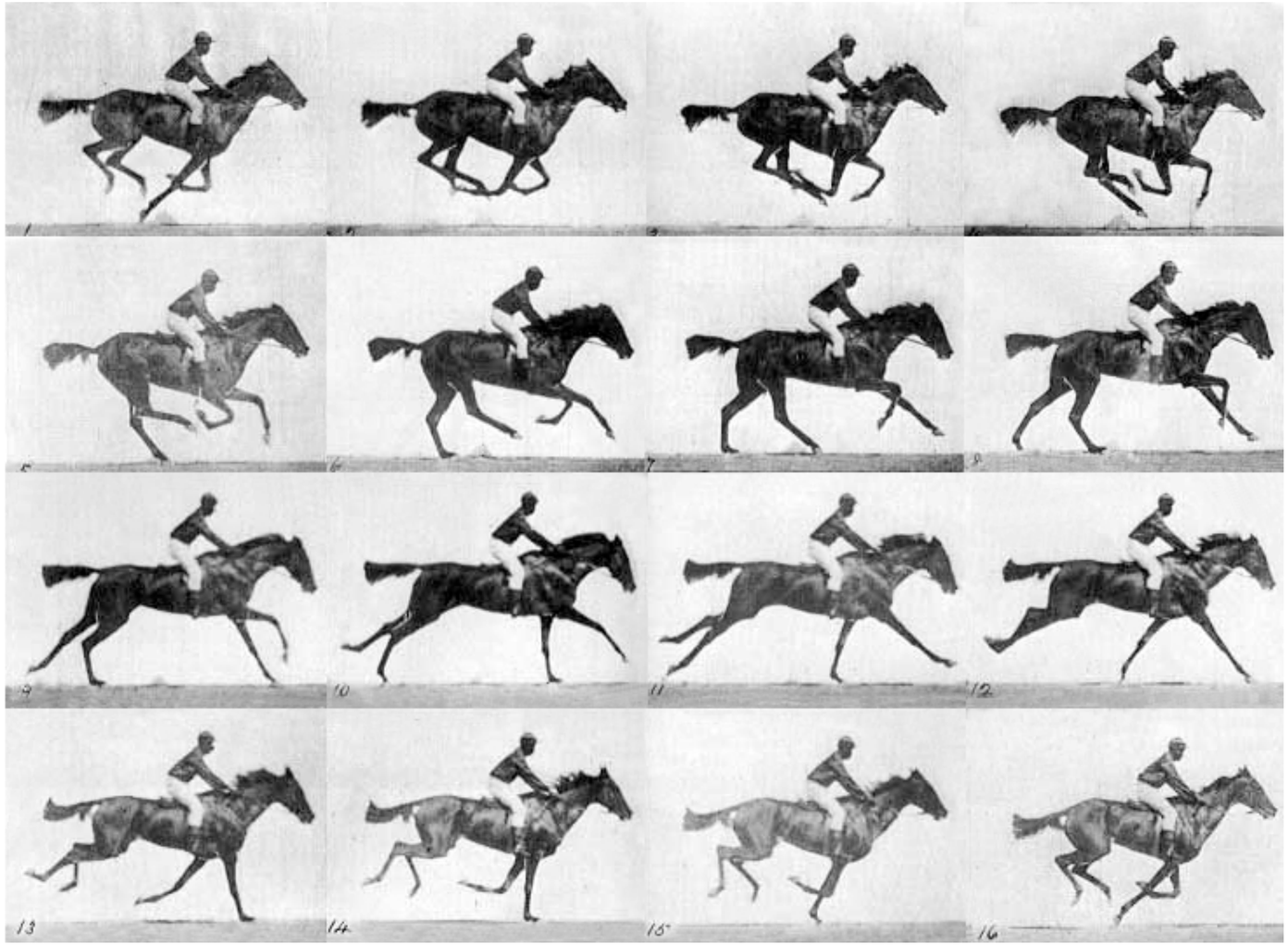
William Curtis (1746-1799)

Donald Norman

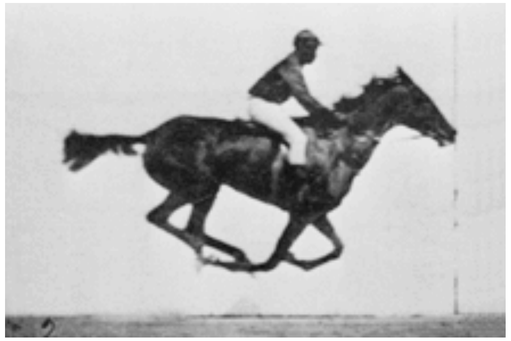
The History of Visual Communication

The Galileo Project, Rice University

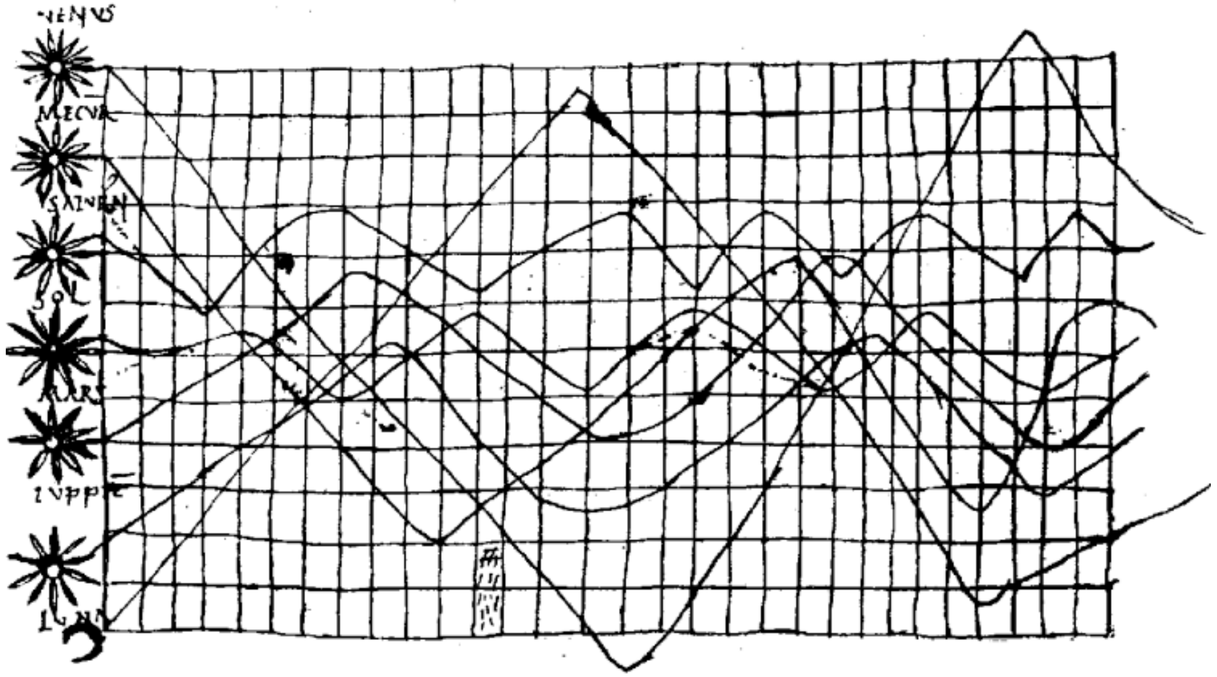
Record



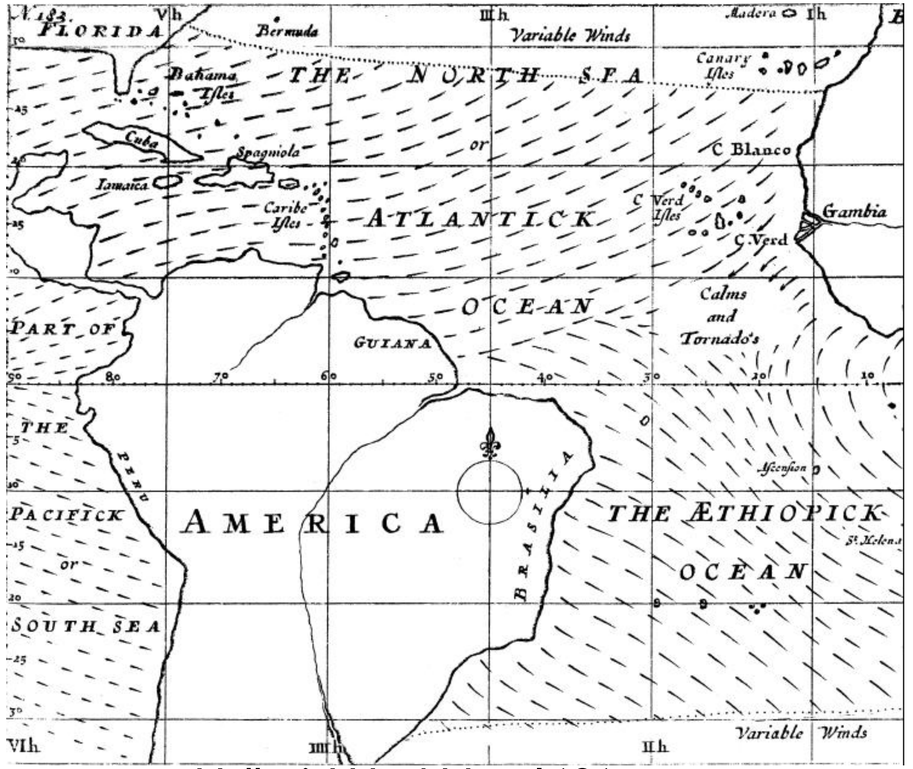
E. J. Muybridge, 1878



Analyze

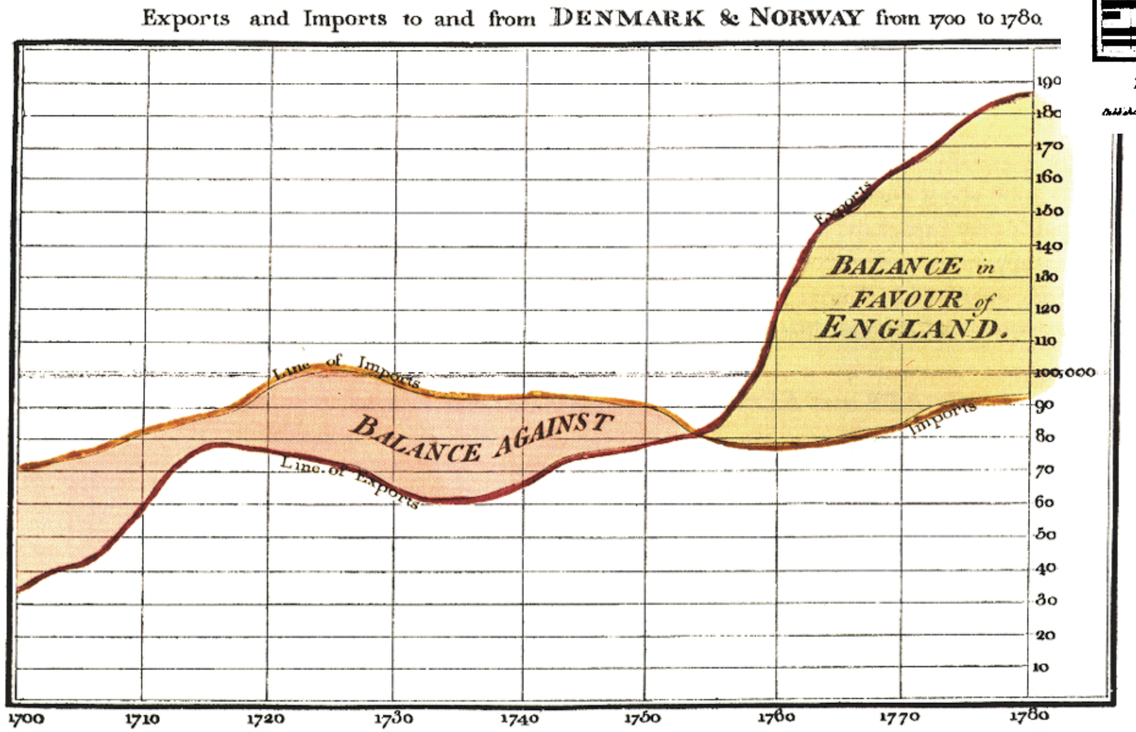


Planetary Movement Diagram, c. 950

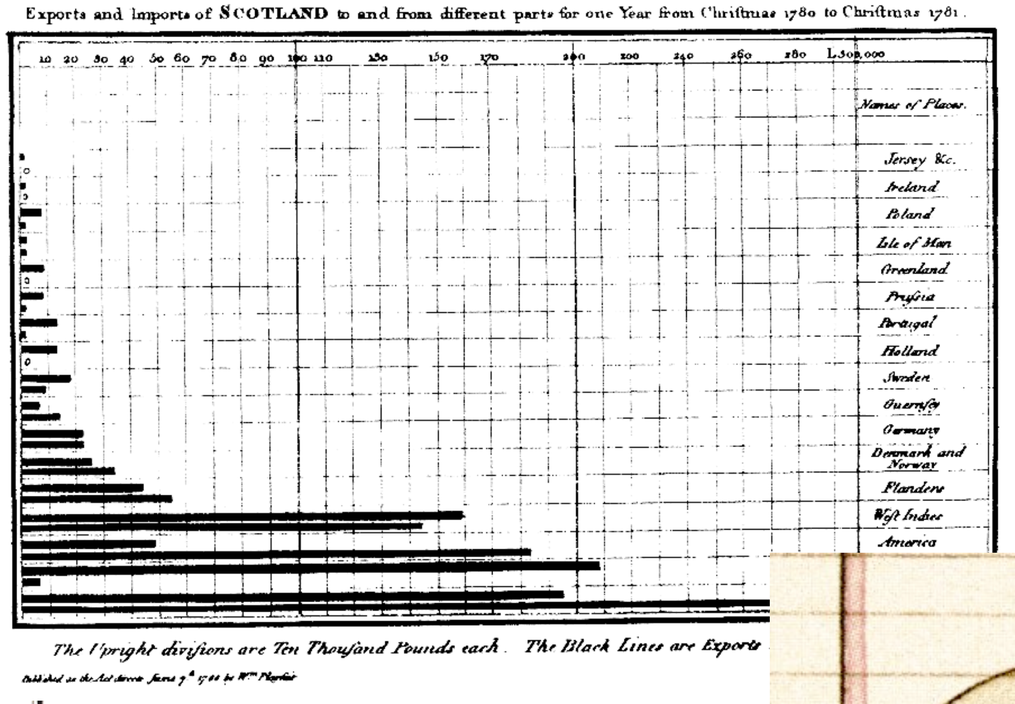


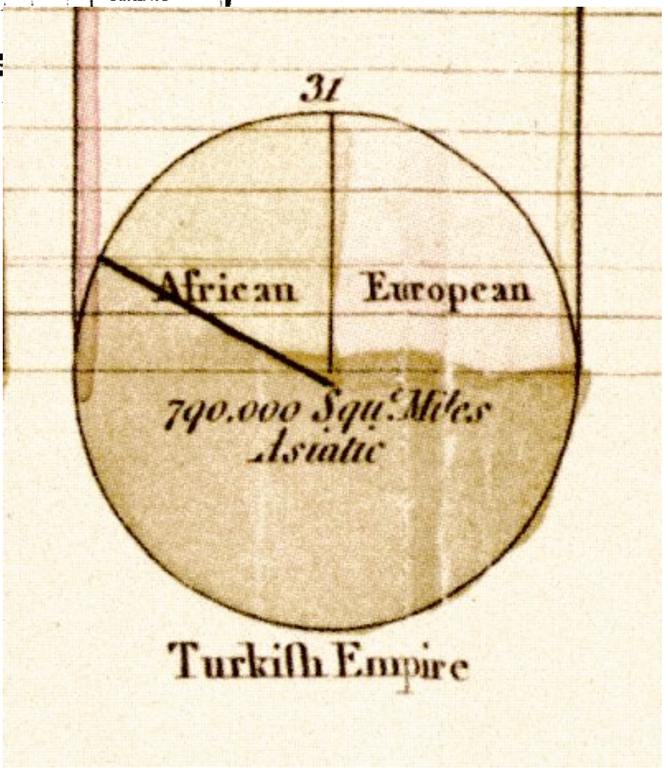
Halley's Wind Map, 1686

Analyze



W. Playfair, 1786

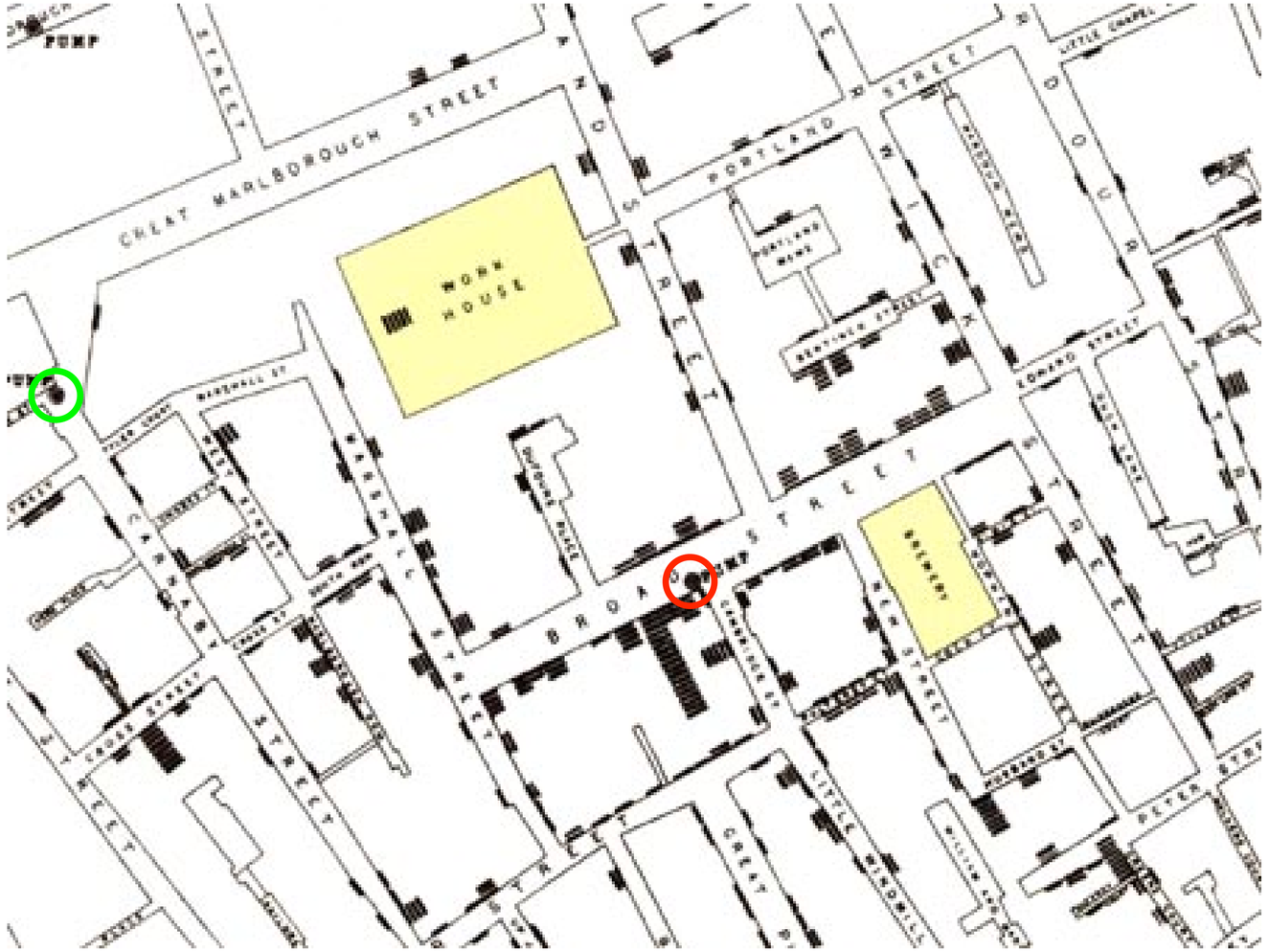




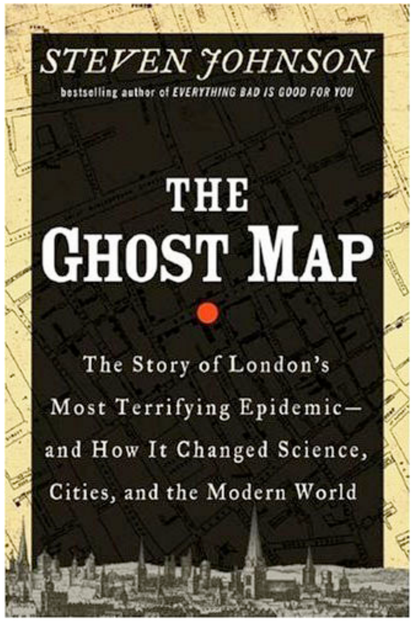
wikipedia.org

W. Playfair, 1801

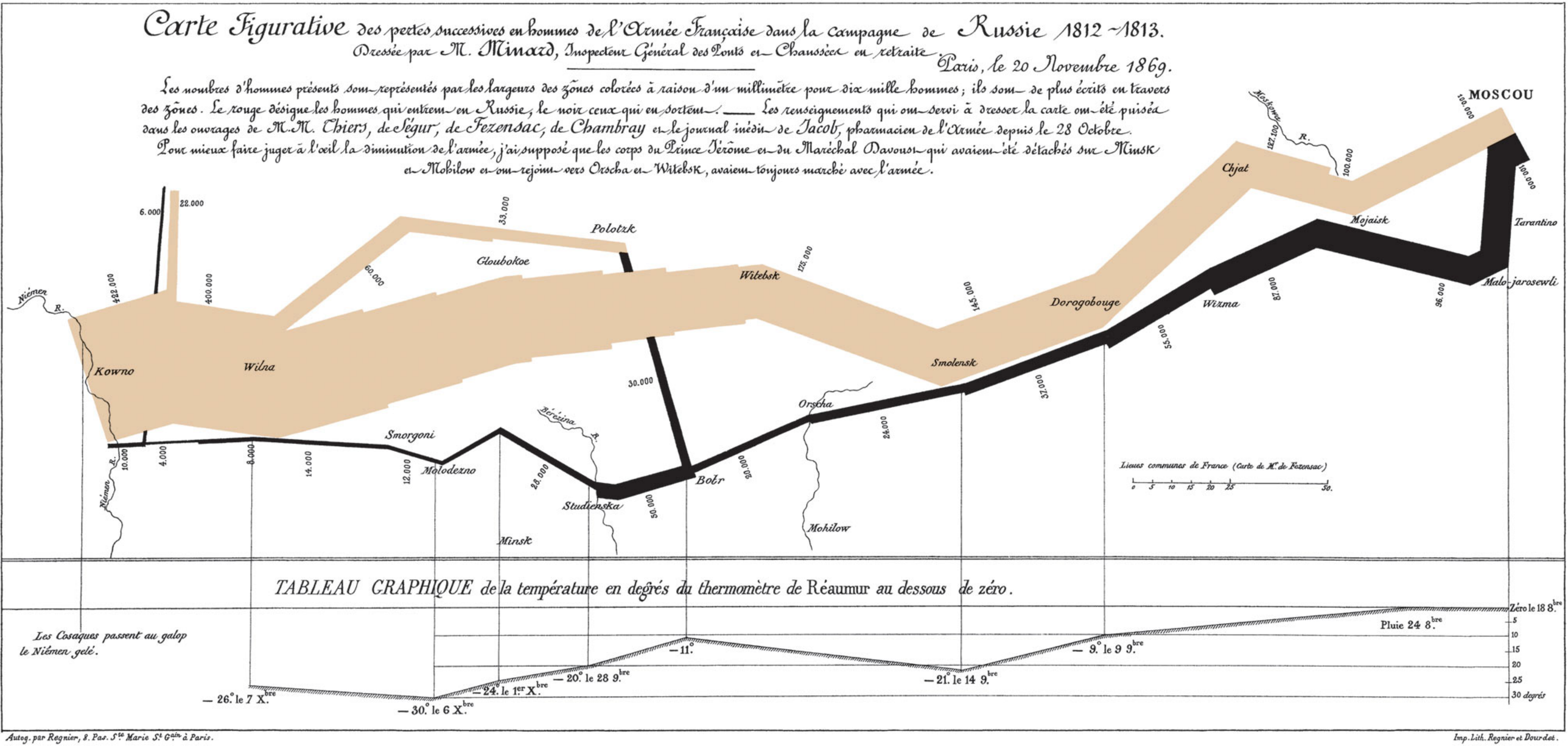
Find Patterns



John Snow, 1854



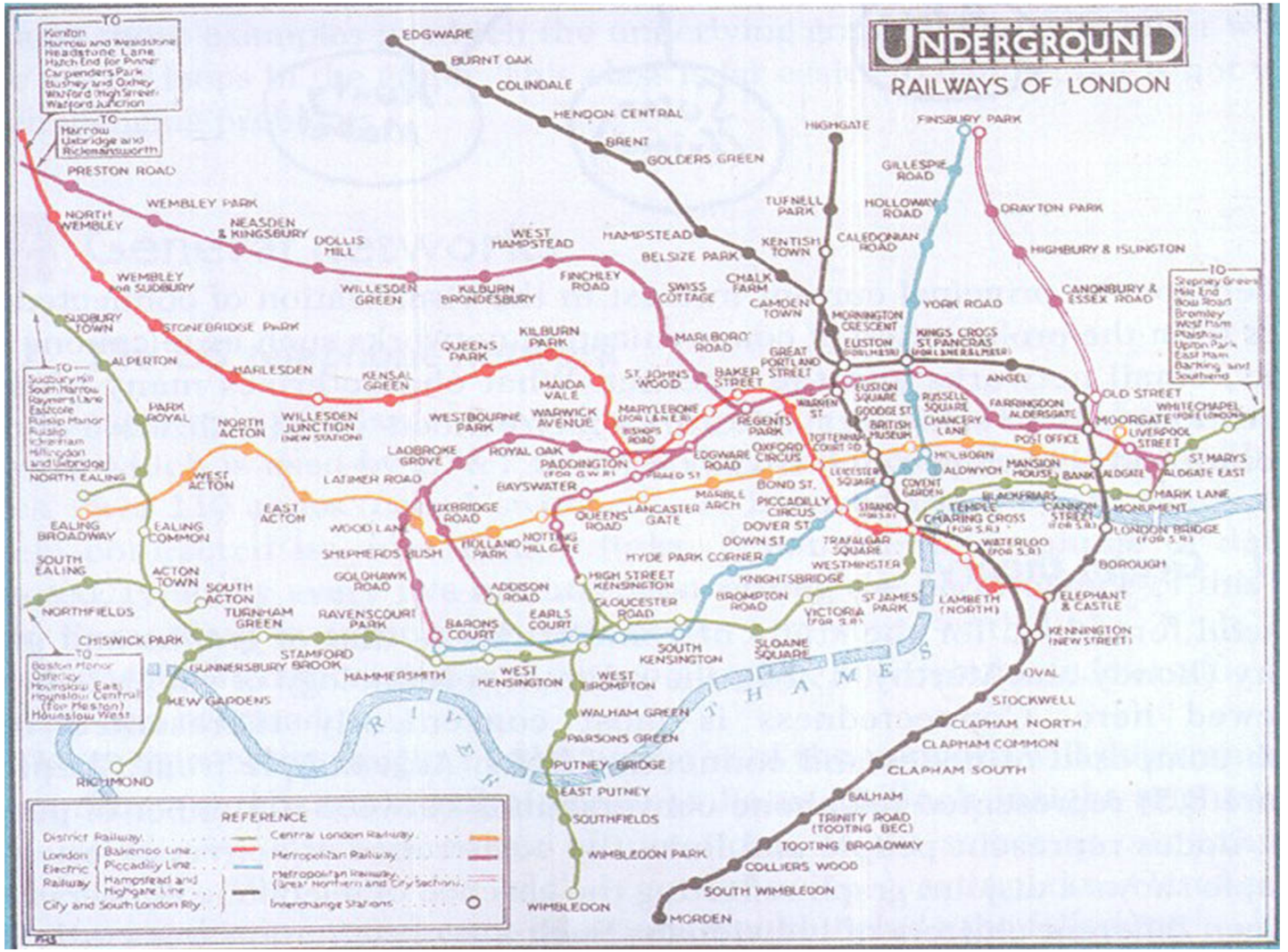
E. Tufte, Visual Explanations, 1997



C.J. Minard, 1869

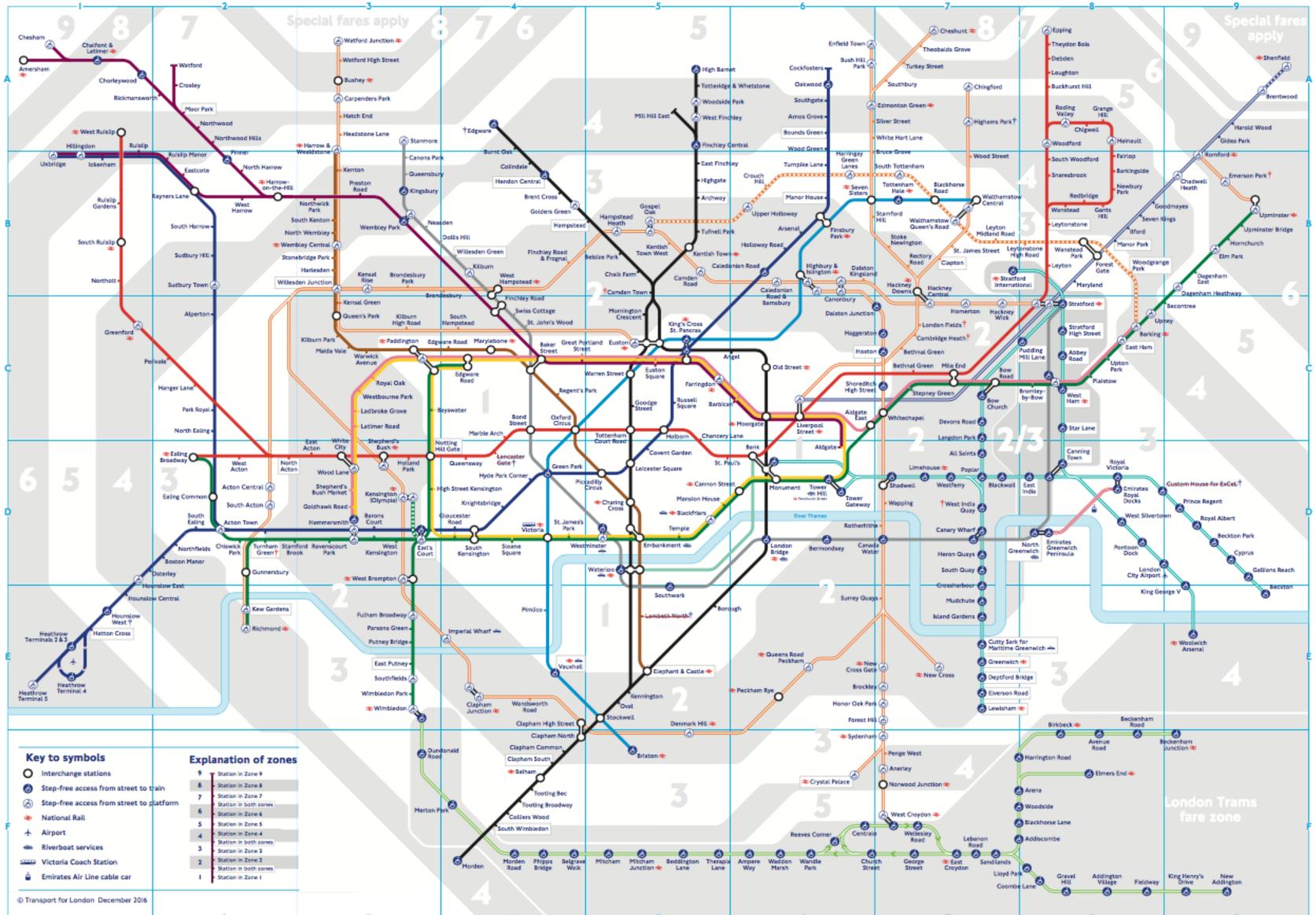
E. Tufte, Writings, Artworks, News

Communicate



London Subway Map, 1927

Communicate



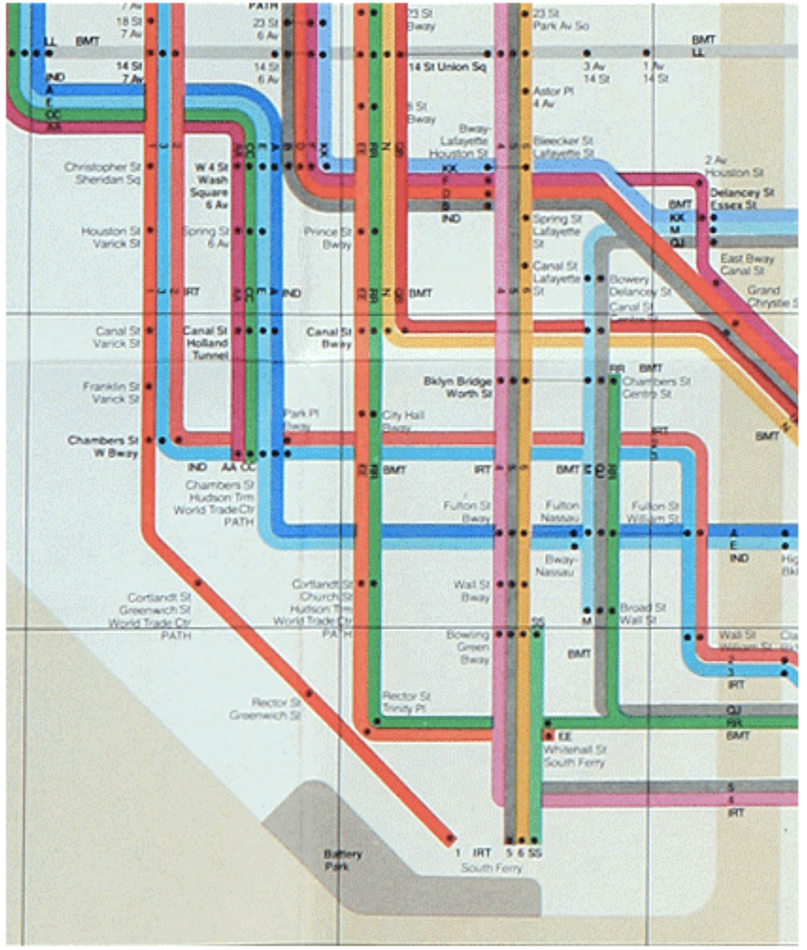
An Overhaul of an Underground Icon

Next month, the Metropolitan Transportation Authority will unveil a resized, recolored and simplified edition of the well-known map, its first overhaul in more than a decade. Related Article »

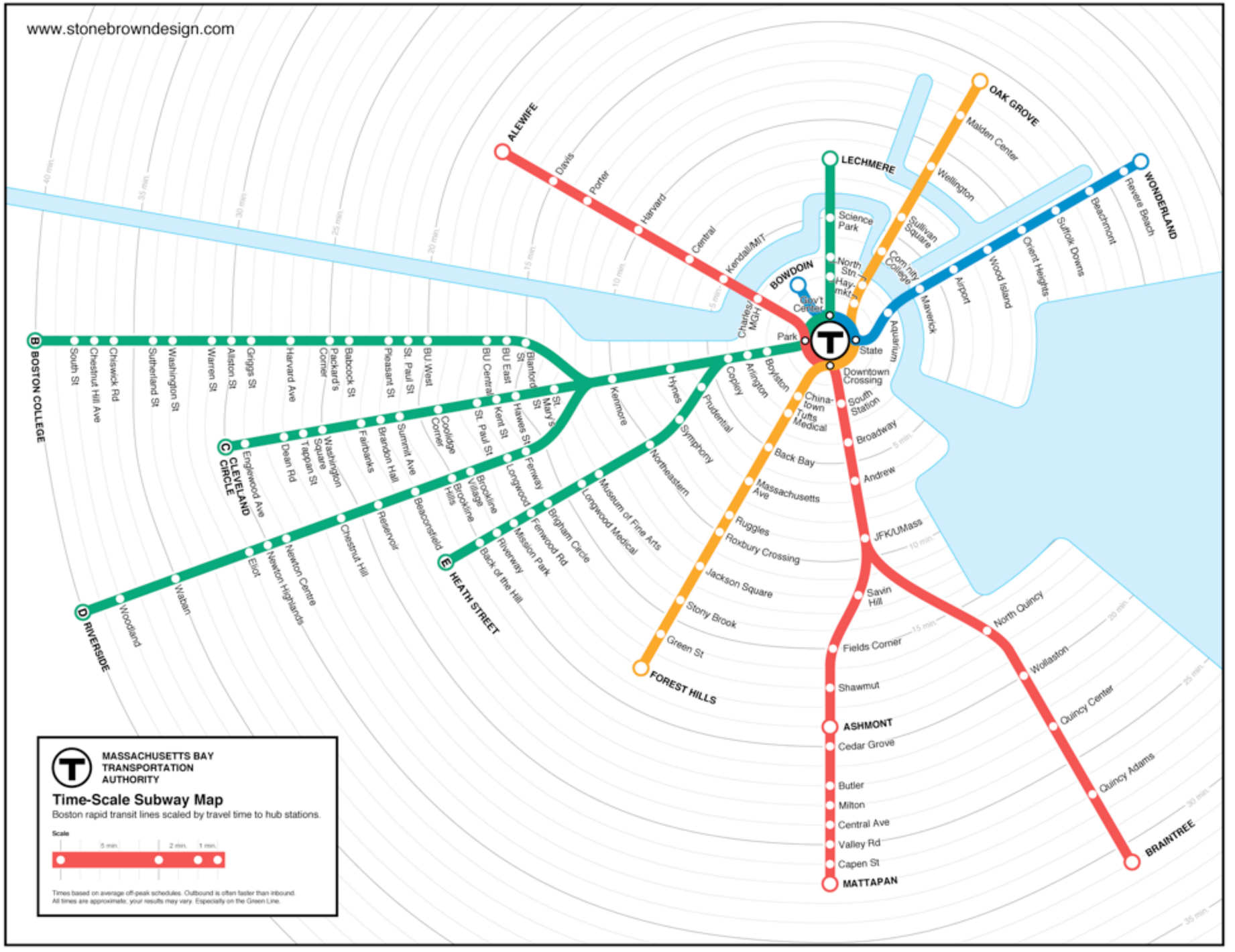
FULL MAP
BRONX
BROOKLYN
MANHATTAN
QUEENS

The New Map

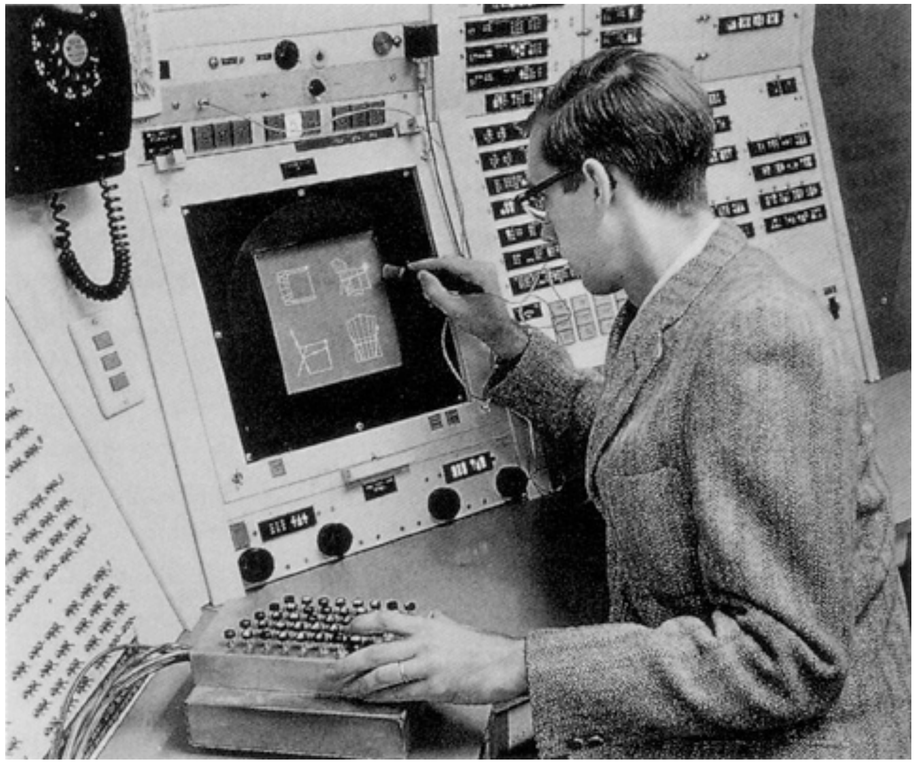
1972: Vignelli's Classic



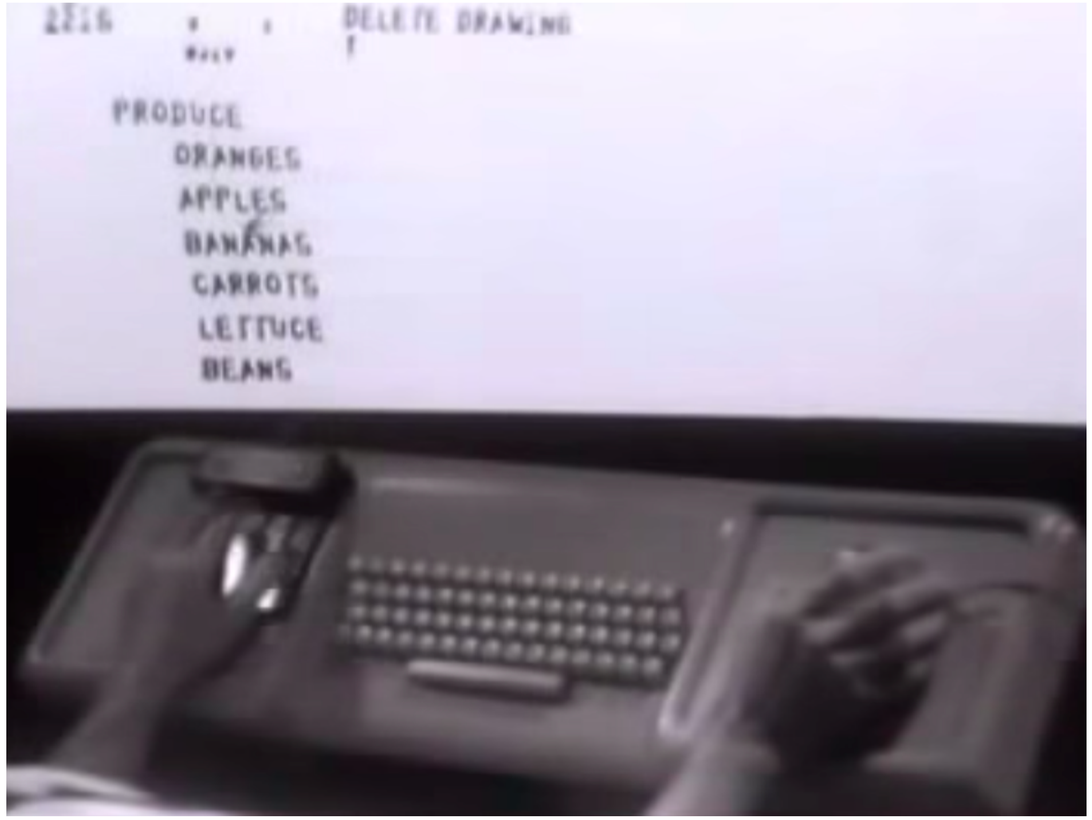
New York Times, 2010


www.stonebrowndesign.com
ALEWIFE
Davis
Porter
Harvard
Central
Kendall/MIT
LECHMERE
Science Park
North Stn
Haymarket
Gov't Center
BOWDOIN
Charles/MGH
Park
State
Downtown Crossing
OAK GROVE
Malden Center
Wellington
Sullivan Square
Comm'ty College
Maverick
Airport
Wood Island
Orient Heights
Suffolk Downs
Beachmont
Revere Beach
WONDERLAND
Aquarium
Boylston
Arlington
Copley
Prudential
Hynes
Kenmore
Symphony
Northeastern
Museum of Fine Arts
Longwood Medical
Brigham Circle
Fenwood Rd
Mission Park
Riverway
Back of the Hill
HEATH STREET
Longwood
Fenway
Hawes St
Kent St
St. Paul St
Coolidge Corner
Summit Ave
Brandon Hall
Fairbanks
Washington Square
Tappan St
Dean Rd
Englewood Ave
CLEVELAND CIRCLE
Brookline Village
Brookline Hills
Beaconsfield
Reservoir
Chestnut Hill
Newton Centre
Newton Highlands
Eliot
Waban
Woodland
RIVERSIDE
St. Mary's St
Blanford St
BU East
BU Central
BU West
St. Paul St
Pleasant St
Babcock St
Packard's Corner
Harvard Ave
Griggs St
Allston St
Warren St
Washington St
Sutherland St
Chiswick Rd
Chestnut Hill Ave
South St
BOSTON COLLEGE
China-town
Tufts Medical
Back Bay
Massachusetts Ave
Ruggles
Roxbury Crossing
Jackson Square
Stony Brook
Green St
FOREST HILLS
South Station
Broadway
Andrew
JFK/UMass
Savin Hill
Fields Corner
Shawmut
ASHMONT
Cedar Grove
Butler
Milton
Central Ave
Valley Rd
Capen St
MATTAPAN
North Quincy
Wollaston
Quincy Center
Quincy Adams
BRAINTREE
40 min
35 min
30 min
25 min
20 min
15 min
10 min
5 min
MASSACHUSETTS BAY TRANSPORTATION AUTHORITY
Time-Scale Subway Map
Boston rapid transit lines scaled by travel time to hub stations.
Scale
5 min.
2 min.
1 min.
Times based on average off-peak schedules. Outbound is often faster than inbound.
All times are approximate; your results may vary. Especially on the Green Line.

Interact



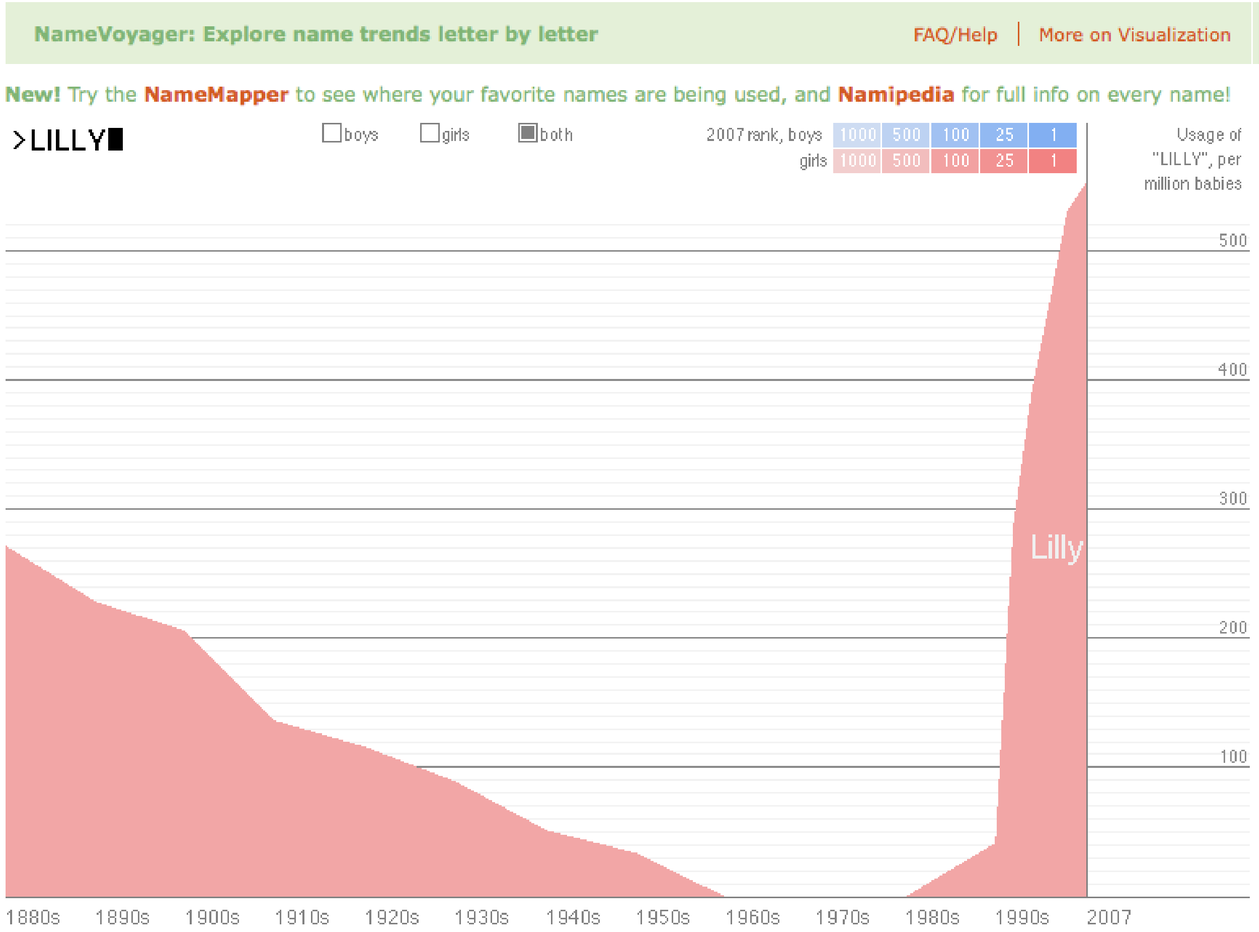
Ivan Sutherland, Sketchpad, 1963



Doug Engelbart, 1968

Modern Examples

Analyze



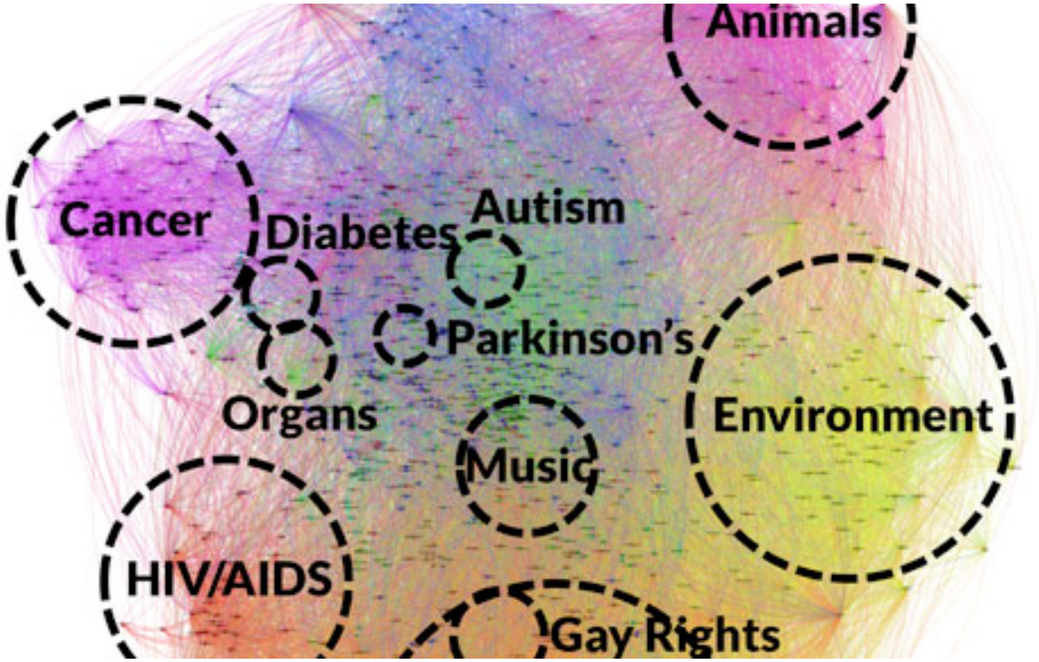
M. Wattenberg, 2005

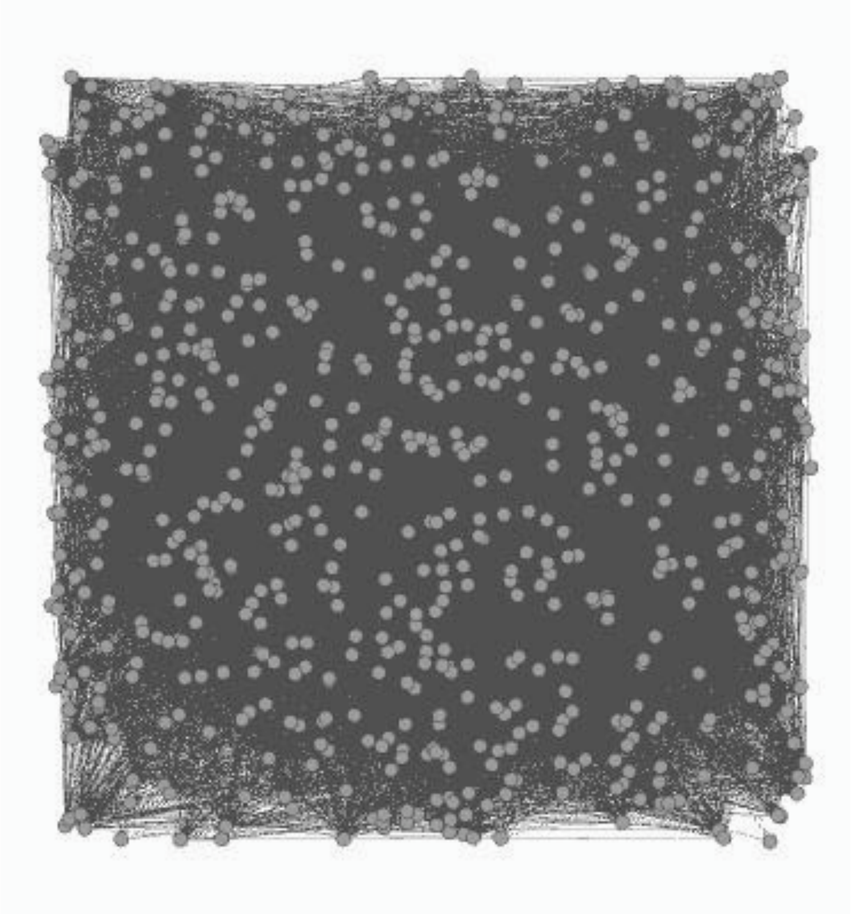
Communicate

Hans Rosling, TED 2006

It’s about Humans!

Not everything that can be drawn can be read!





Limits of Cognition

Daniel J. Simons and Daniel T. Levin, Failure to detect changes to people during a real world interaction, 1998

What is this course?

Adam Perer http://perer.org

perer.org

Adam Perer | Home | Bio | Events | Talks | Recent Projects | Older Research

Adam Perer, Ph.D.

Research Scientist, IBM Research

Adjunct Professor, Carnegie Mellon University

I blend **data visualization** and **data mining** techniques to create visual interactive systems to help users make sense out of big data. Lately, my research focuses on extracting insights from clinical data to support **data-driven medicine**. -@adamperer

Research Scientist
Healthcare Analytics Research Group
IBM T.J. Watson Research Center

Adjunct Professor
Human-Computer Interaction Institute
Carnegie Mellon University

firstname.lastname AT gmail.com
Twitter: @adamperer

Curriculum vitae

Talks & Travel

May 2016	ACM CHI 2016 San Jose, CA
April 2016	University of Pittsburgh Big Data Colloquium Pittsburgh, PA
December 2015	University of Pittsburgh DBMI Seminar Pittsburgh, PA
November 2015	AMIA San Francisco, CA
October 2015	IEEE VIS Chicago, IL
September 2015	CMU HCII Seminar Pittsburgh, PA
April 2015	Wellcome Trust Genome Hinxton, UK
April 2015	OpenVis Conference

Bio

Adam Perer is a Research Scientist at IBM's T.J. Watson Research Center in New York, where he is a member of the Healthcare Analytics Research Group. He is also an Adjunct Professor in the Human-Computer Interaction Institute at Carnegie Mellon University. His research in visualization and human-computer interaction focuses on the design of novel visual analytics systems. This work has been published at premier venues in visualization, human-computer interaction, and medical informatics (IEEE InfoVis, IEEE VAST, ACM CHI, ACM CSCW, ACM IUI, AMIA). He holds Ph.D. and masters degrees in Computer Science from the University of Maryland, College

About You

Structure & Goals

Course Goals. You will learn:

How to **efficiently visualize data**

- **Evaluate** and **critique** visualization designs
- **Apply** fundamental principles & techniques
- **Design** visual data analysis solutions

Implement interactive data visualizations

- Web development skills

Course Components

Lectures: introduce theory

Design Critiques: develop “an eye” for vis design, critique, learn by example

Labs: short coding tutorials, examples

- Based on a published script on website
- Strongly related to homework assignments

Homeworks help practice specific skills

Final Project gives you a chance to go through a complete vis project

Course Components

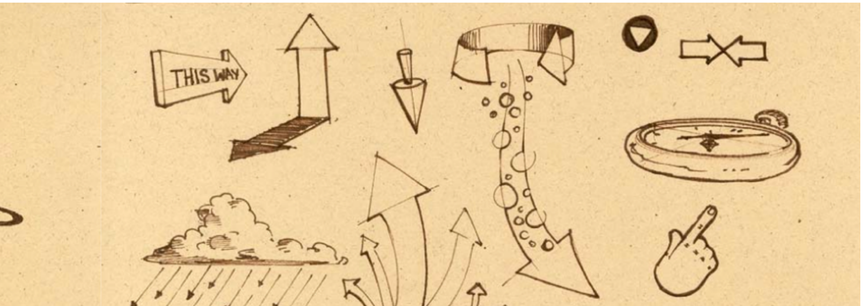
Lecture
Reading
Discussion

Theory

Design Lecture
Design Studios

Labs
D3 reading
Self-study
Office hours

Design Skills ⟷ Coding Skills




```
<!DOCTYPE html>
<meta charset="utf-8">
<style>

text {
  font: 10px sans-serif;
}

</style>
<body>
<script src="http://d3js.org/d3.v3.min.js"></script>
<script>
```

Schedule

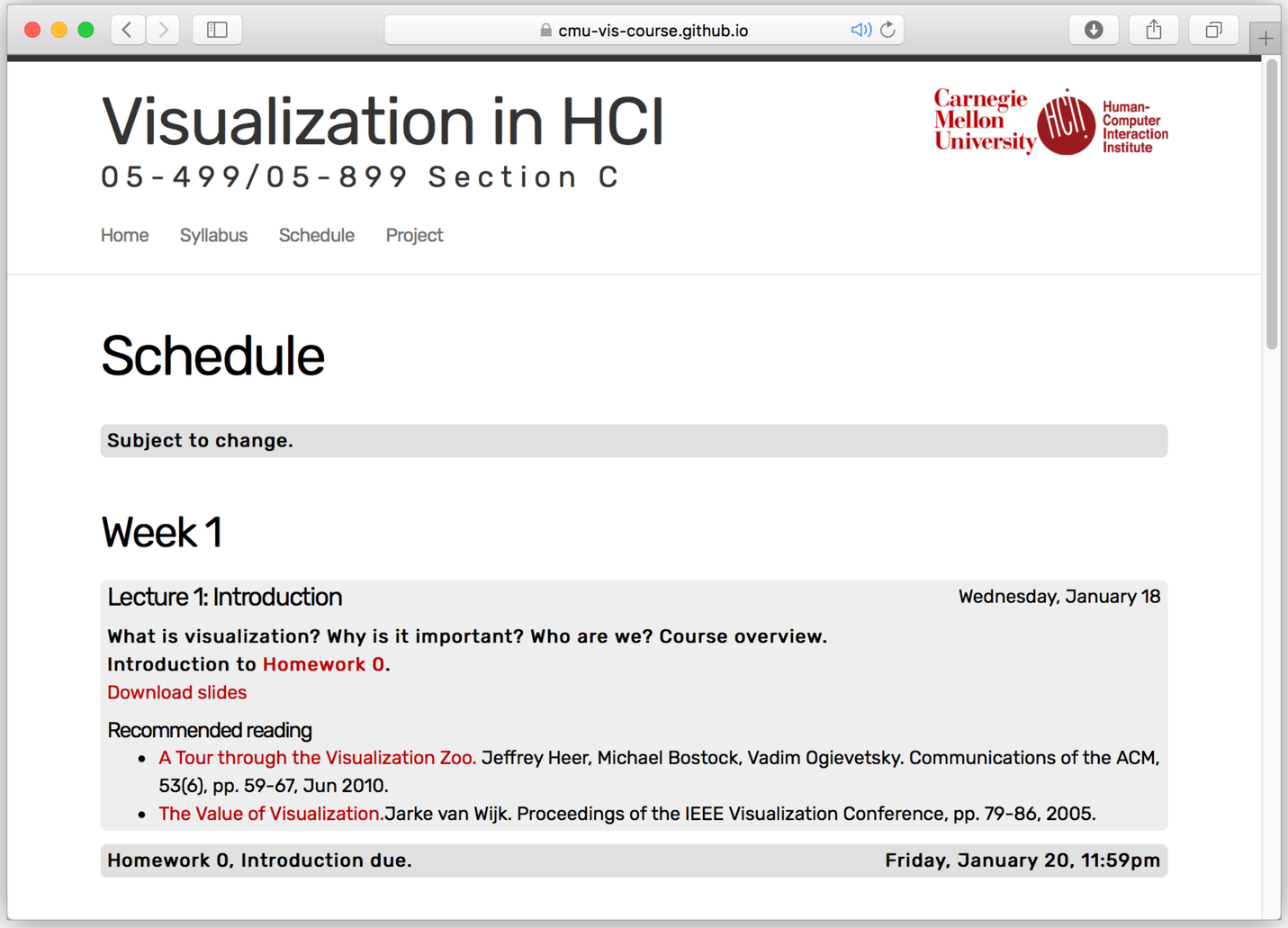
Lectures: Monday and Wednesdays
12:00-1:20 pm

Three Key Parts:

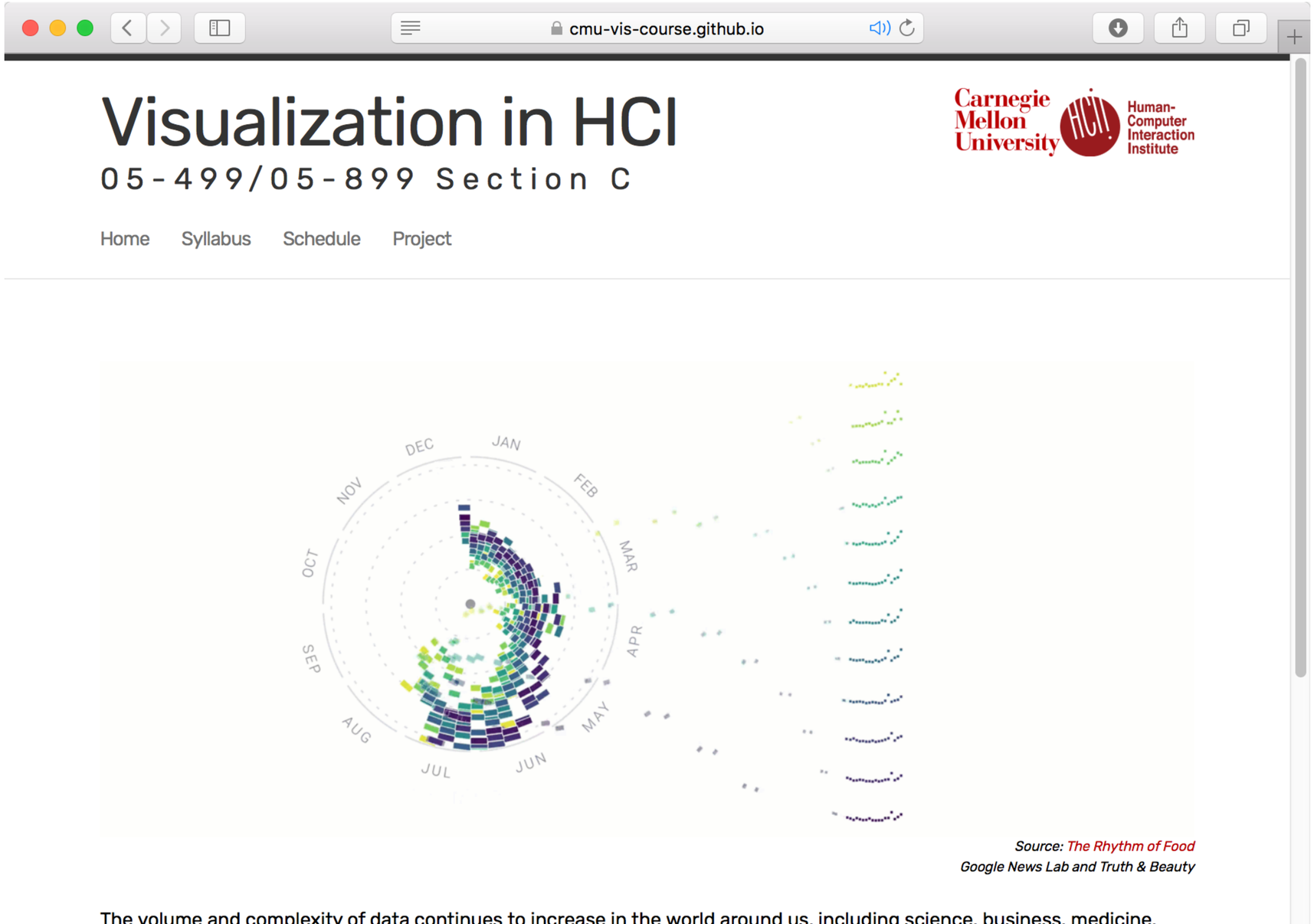
I. Technical Foundations
HTML, Javascript, D3

II. Visualization Fundamentals
Perception, Visual encodings, Design Guidelines, Tasks..

III. Abstract Data Visualization
Tables, Graphs, Maps



Information http://cmu-vis-course.github.io



Communicate

Slack

http://cmu-vis-course.slack.com

Please use Slack for all general questions - code, concepts, etc.

Only use e-mail for personal inquiries

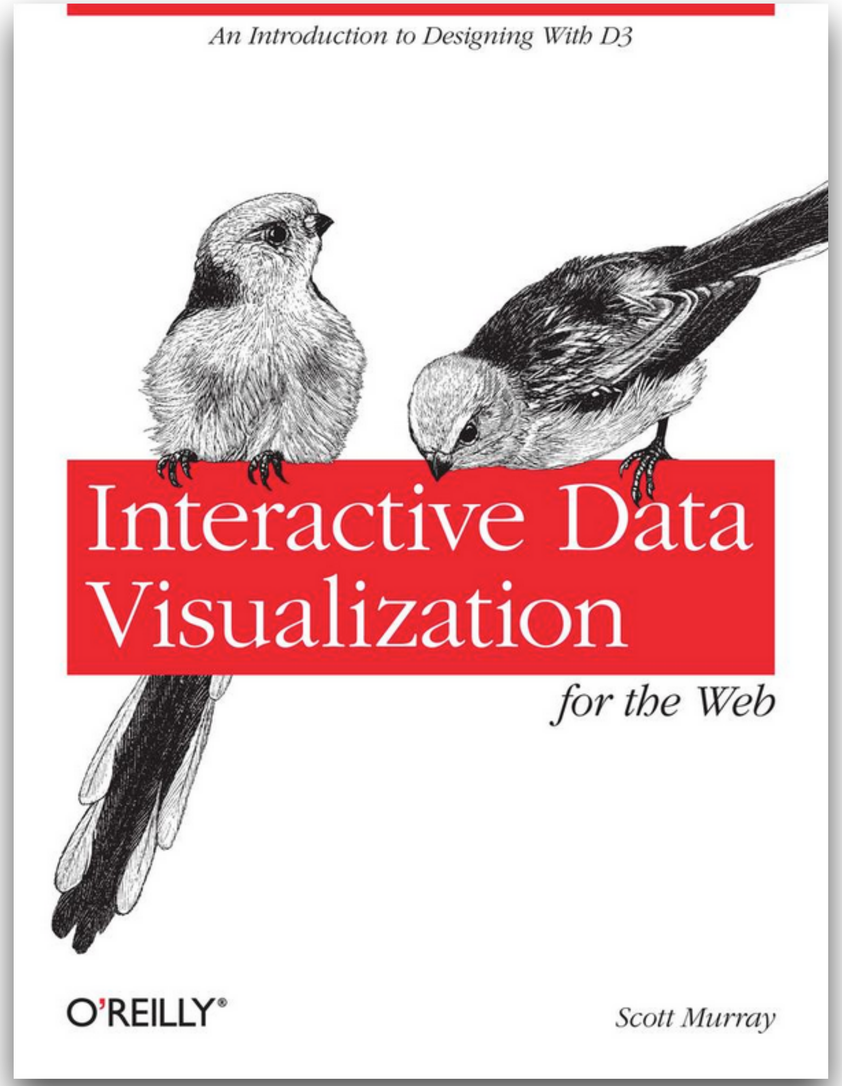
Office Hours

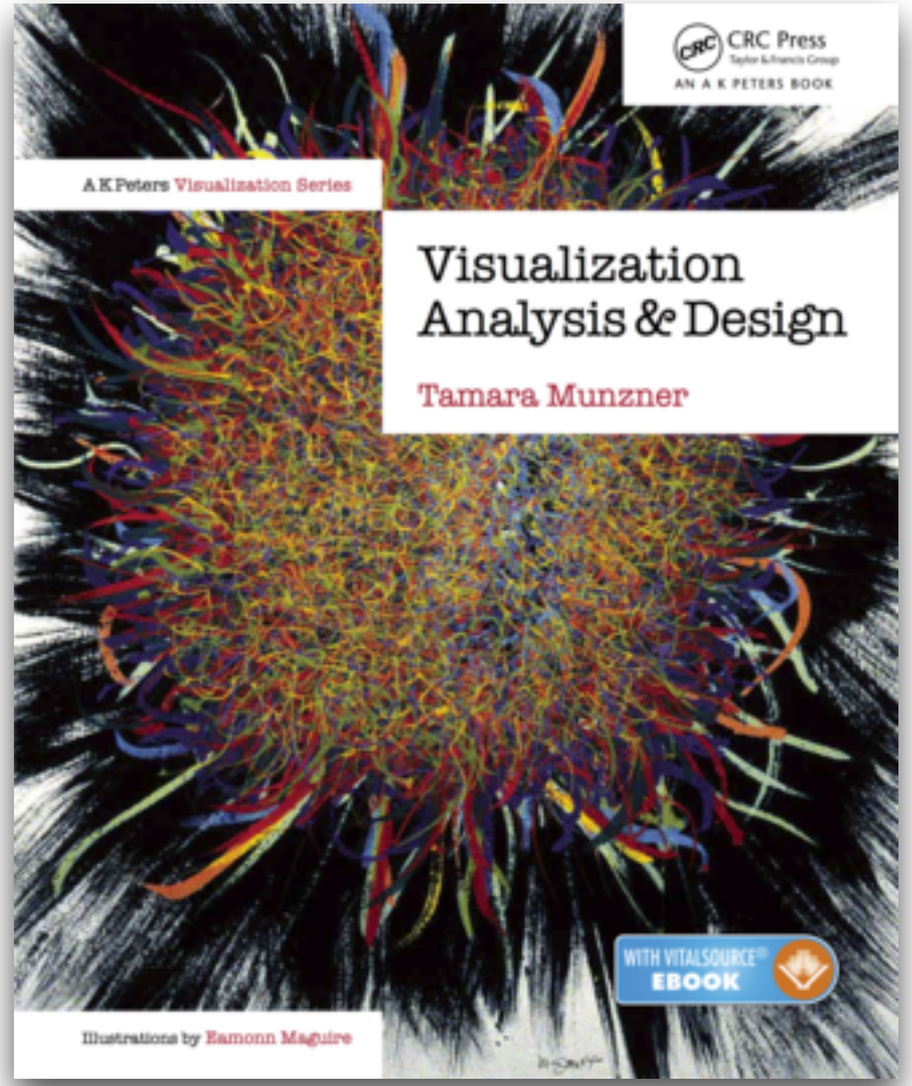
Today after class, future schedule TBA

E-Mail

cmu.vis.course@gmail.com

Required Books





Programming

HTML

JS

Data-Driven Documents

Prerequisites

- Programming experience
- Willingness to learn new software & tools
- You will need to build skills by yourself!

Formalities

How are you graded?

Homework Assignments: 40%

- Start early! Will take long if you don’t know JS/D3 yet

Final Project: 60%

- Teams, two milestones

Cheating

You are welcome to discuss the course's ideas, material, and homework with others in order to better understand it, but the work you turn in must be your own (or for the project, yours and your teammate's). For example, you must write your own code, design your own visualizations, and critically evaluate the results in your own words. You may not submit the same or similar work to this course that you have submitted or will submit to another. Nor may you provide or make available solutions to homeworks to individuals who take or may take this course in the future.

Will automatically check for plagiarism in all your submissions

This Week

Homework 0 due this **Friday** [HW0]

1. Sign up for Github (and apply for a free student upgrade)
2. Sign up for the course's Slack team, and introduce yourself
3. Complete the course survey

Next Week

Technological foundations

- Introduction to Git, HTML, CSS
- JavaScript, JSON, D3

Readings

- D3 Book, Chapters 1-3
- VDA Book, Chapter 1

HW1 due