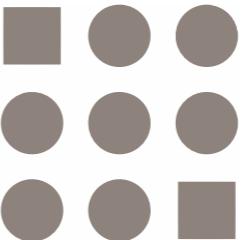


CS 171



Visualization

Hanspeter Pfister
pfister@seas.harvard.edu

Outline

- What?
- Why?
- Who?
- How?

Outline

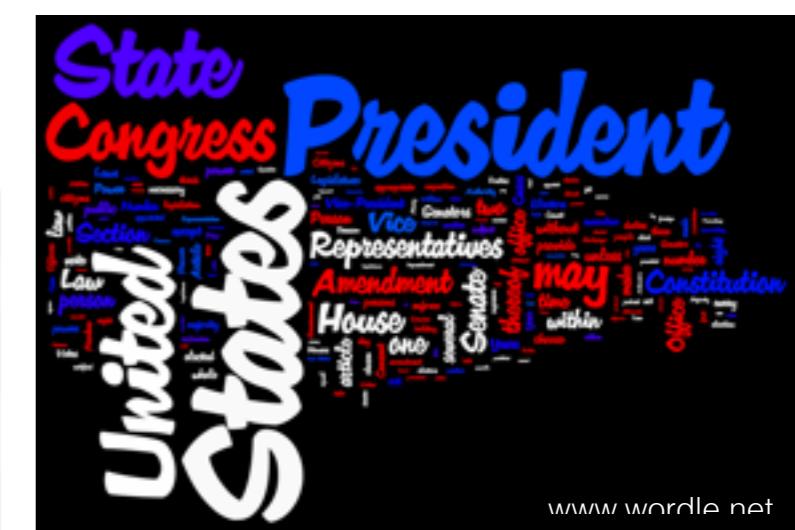
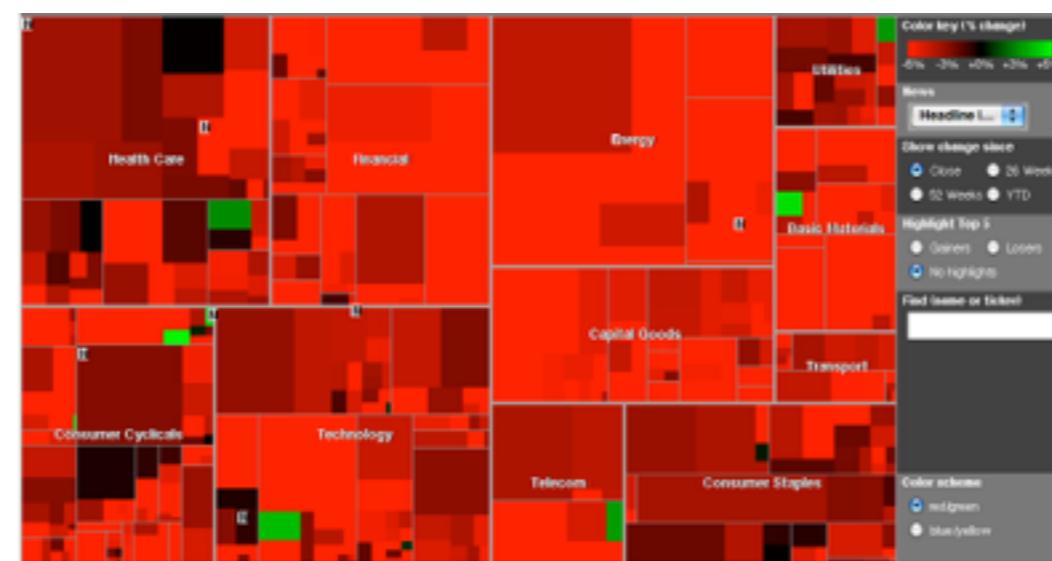
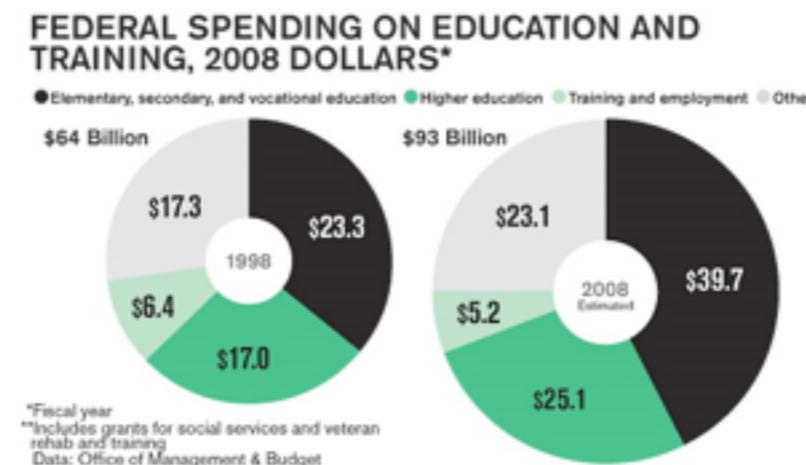
- What?
- Why?
- Who?
- How?



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

...to convey information through graphical representations of data



Visualization Goals

Record information

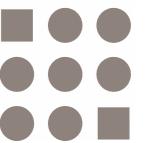
Analyze (explore) data

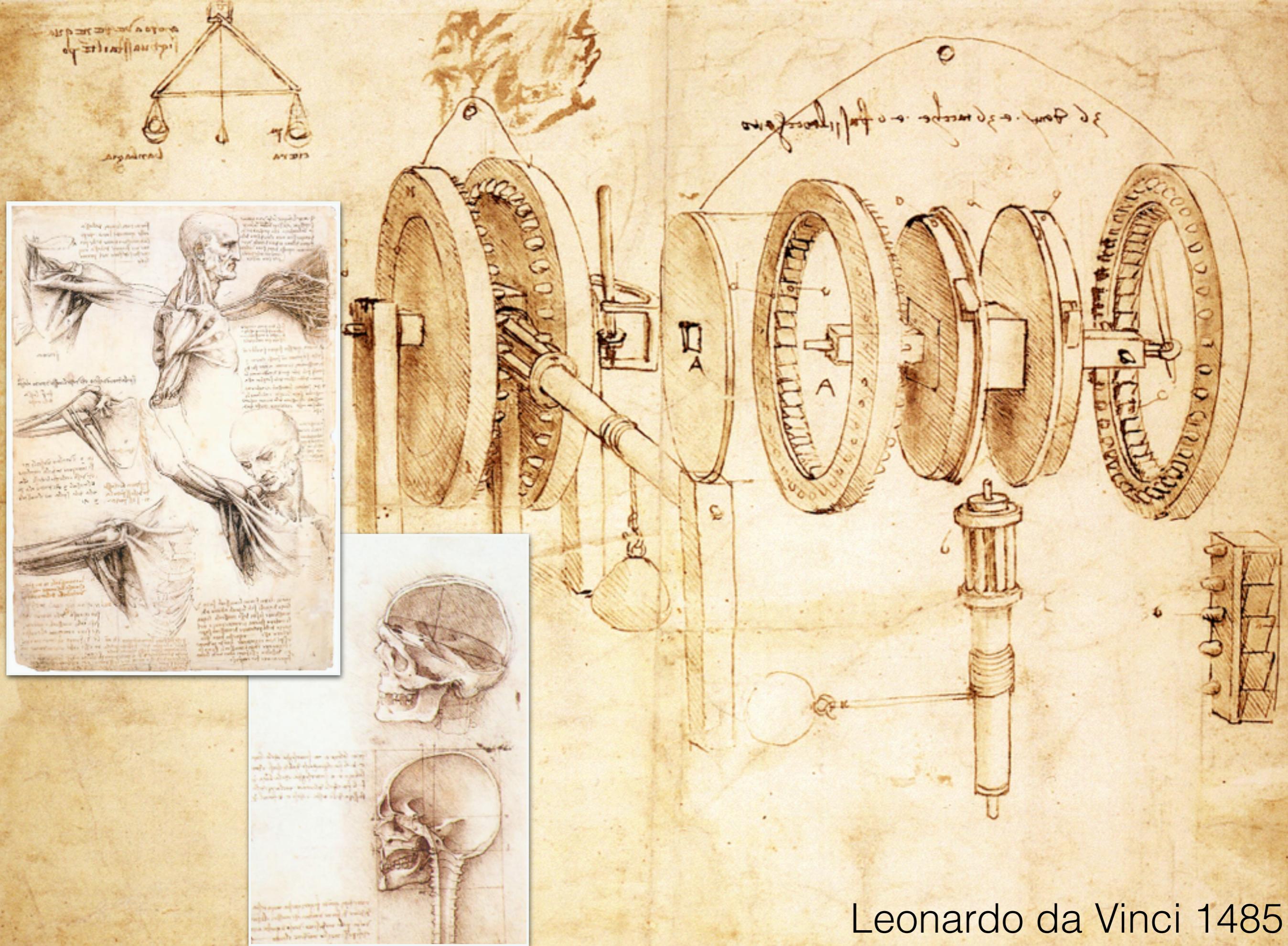
Communicate (explain) to others

Record

CS

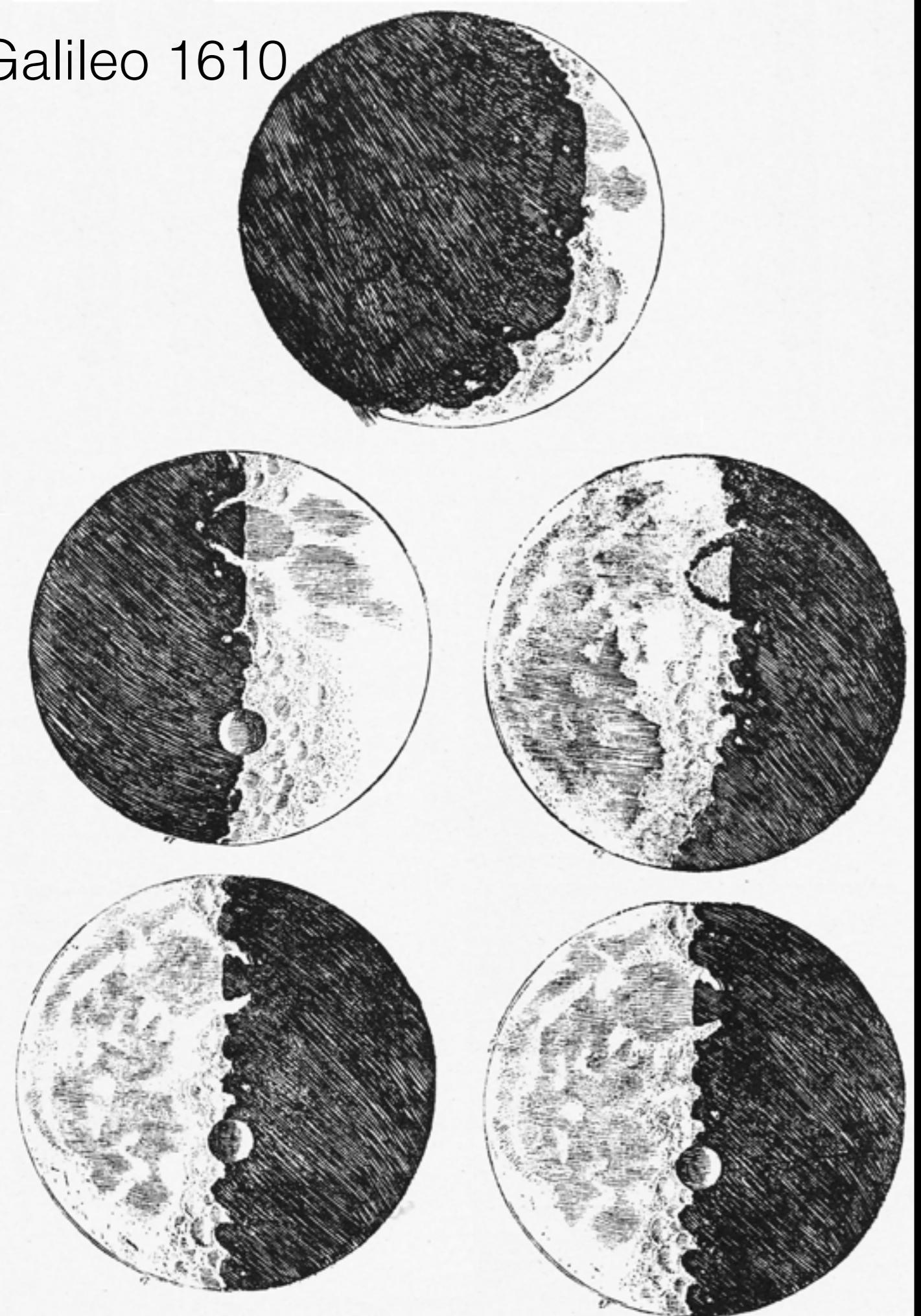
171



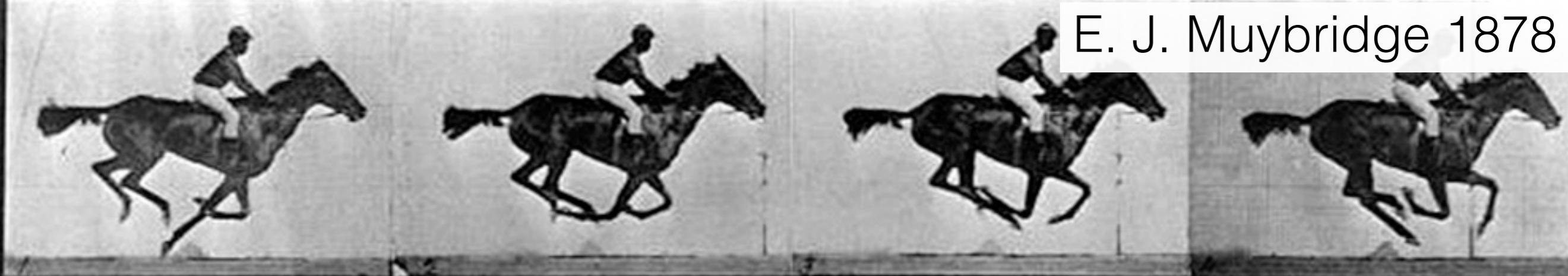


Leonardo da Vinci 1485

Galileo 1610



E. J. Muybridge 1878

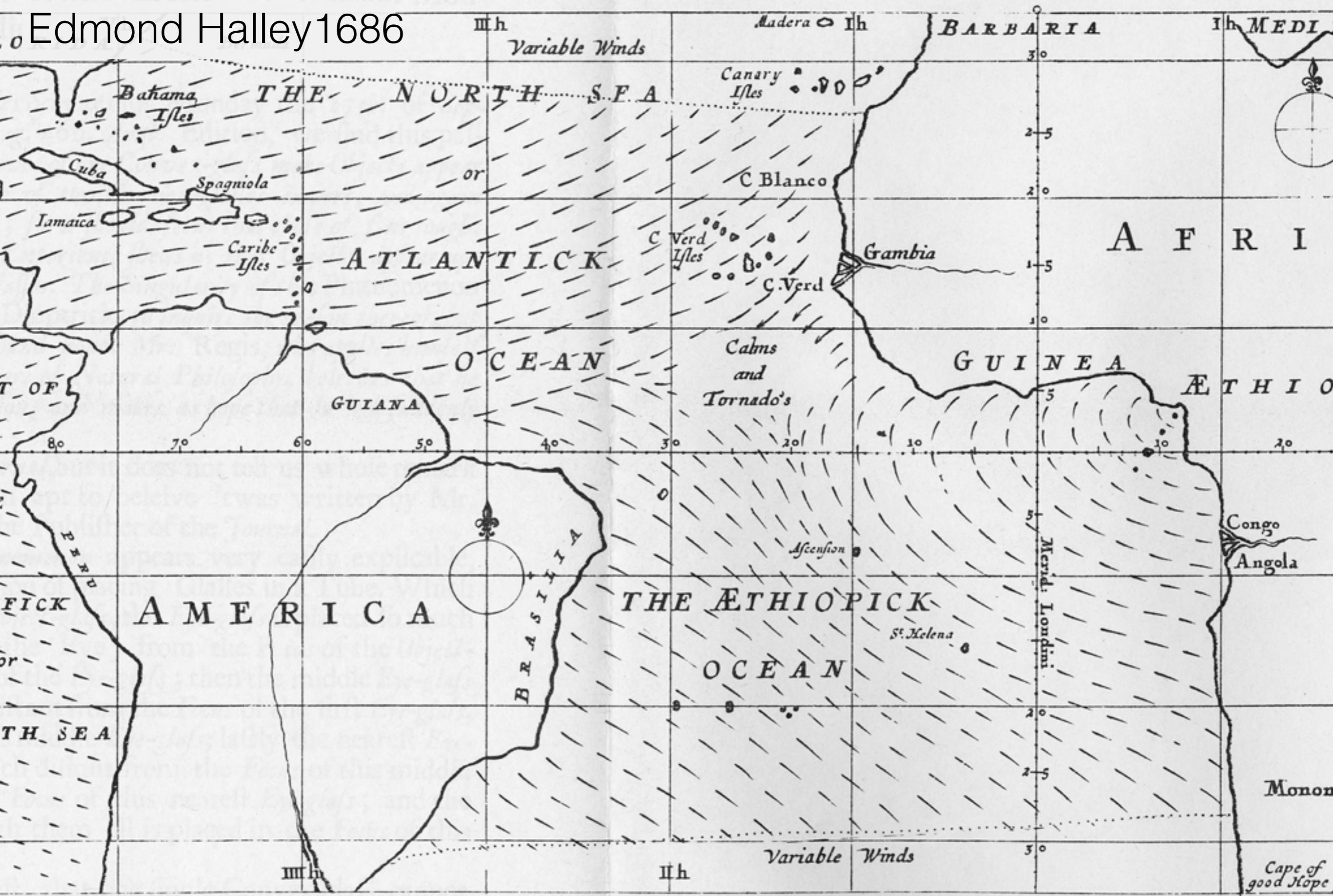


Analyze

CS
171



Edmond Halley 1686



17th Century

Nº3.

18th Century

William Playfair 1820

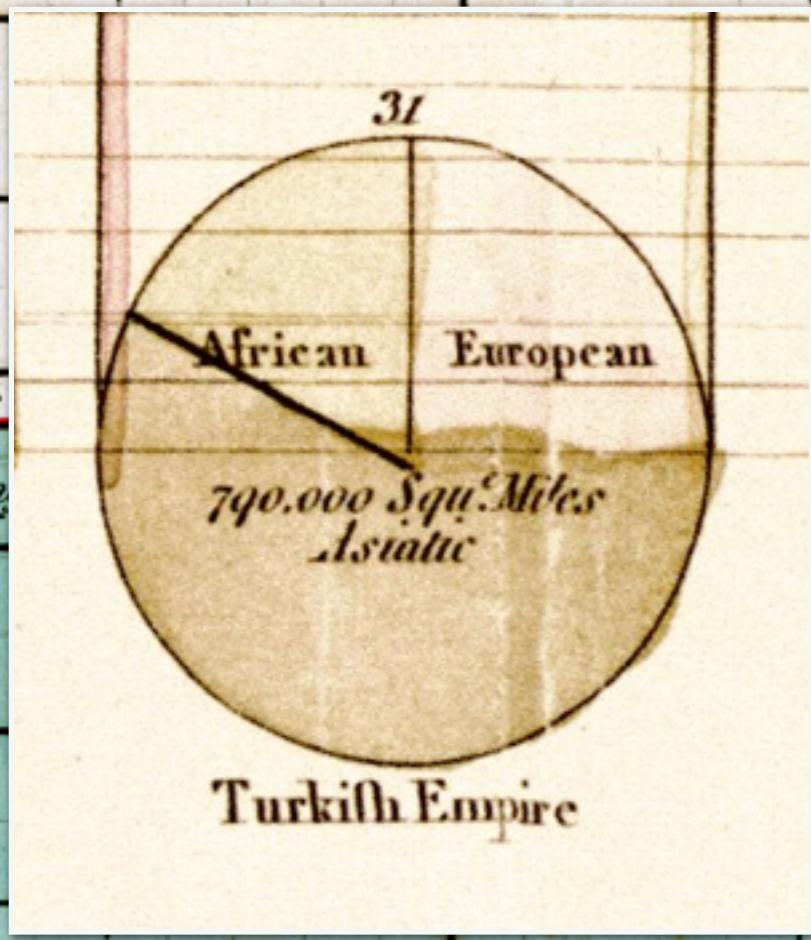


CHART
*Shewing the Value of the Quarter of Wheat
In Shillings & in Days Wages
OF A GOOD MECHANIC*
from
1565 to 1821.

Shillings

in Shillings

in Wages

in Shillings

in Wages

1625

1650

1675

1700

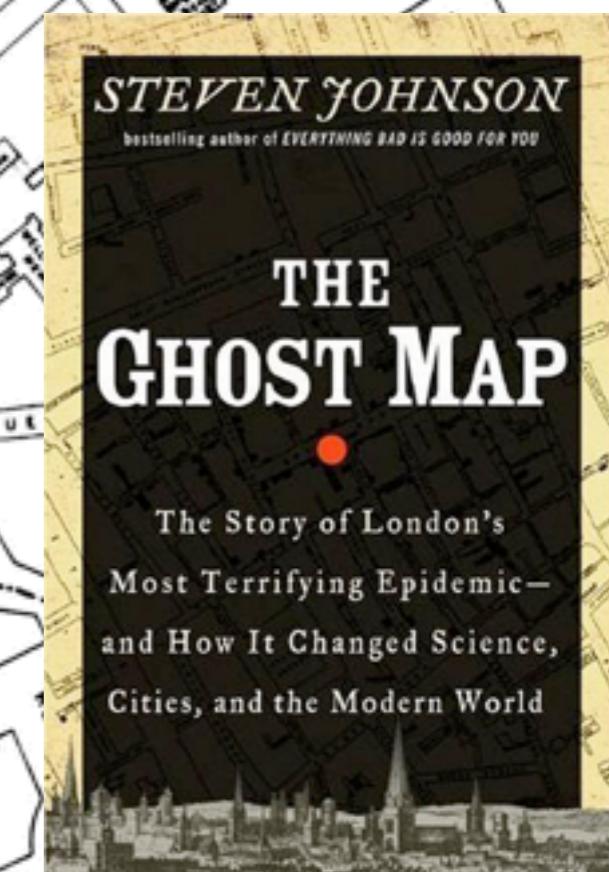
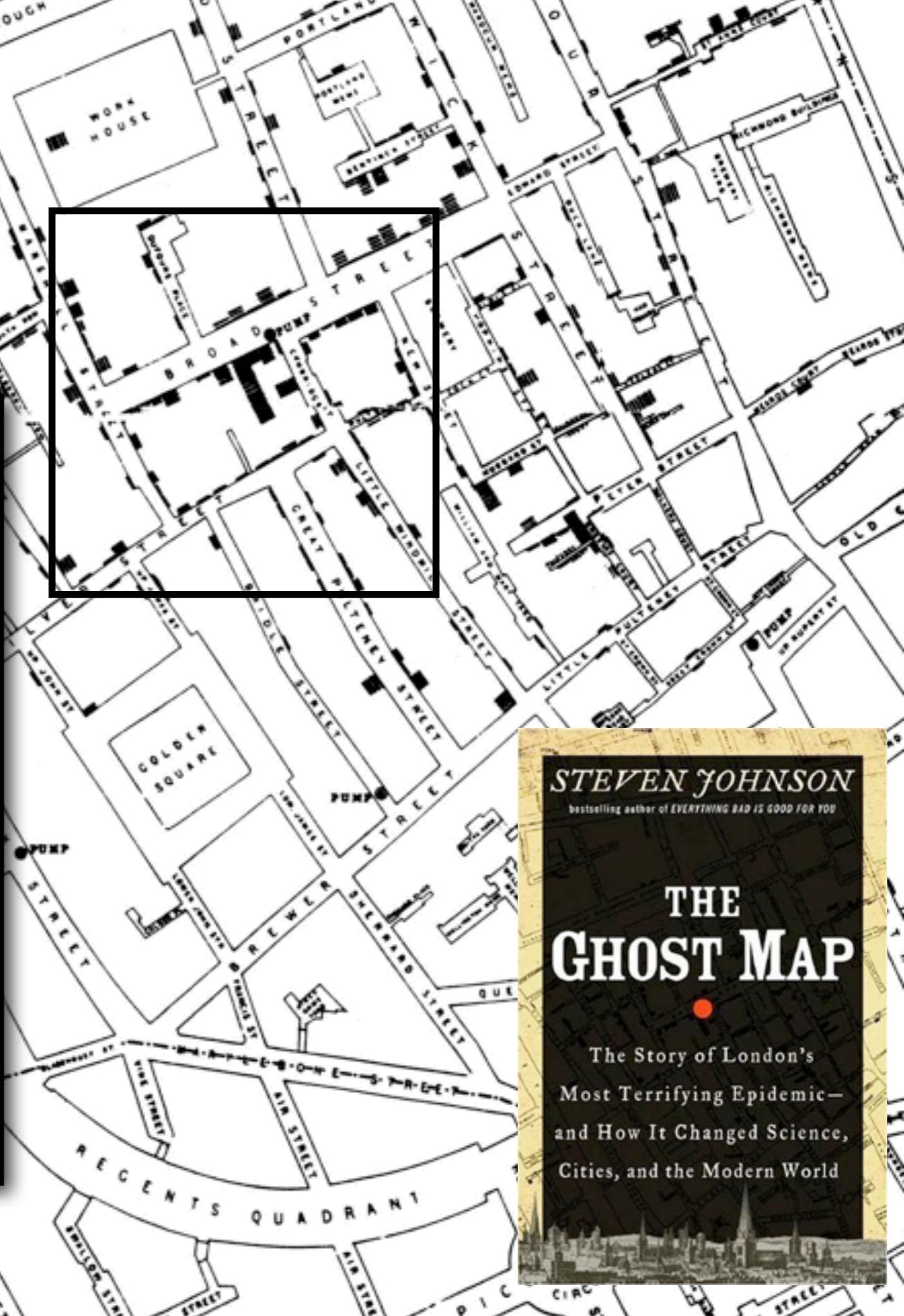
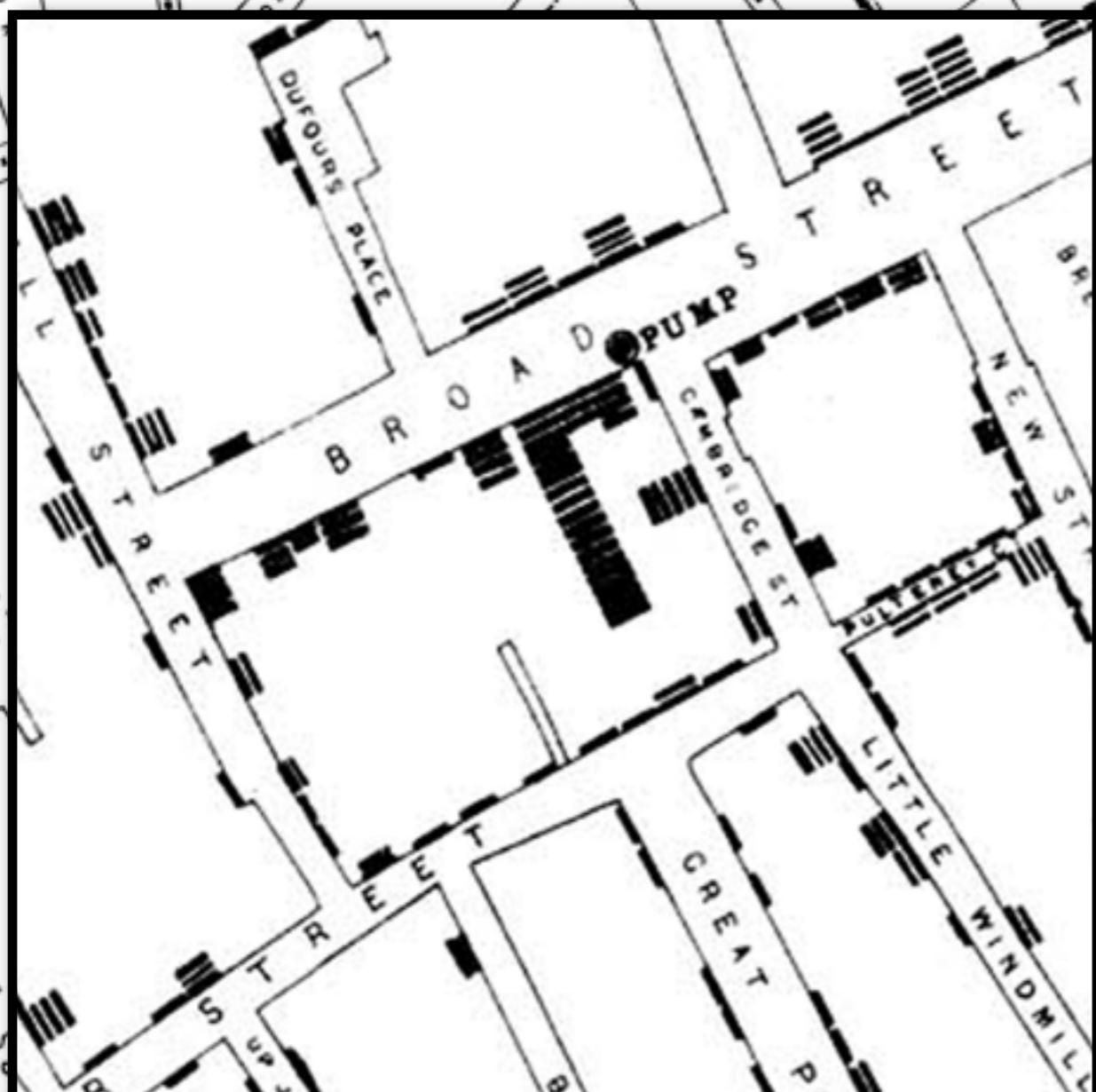
1725

1750

1775

1800

John Snow 1854



New! Try the [NameMapper](#) to see where your favorite names are being used, and [Namipedia](#) for full info on every name!

>LILLY■

 boys girls both

2007 rank, boys

1000	500	100	25	1
------	-----	-----	----	---

girls

1000	500	100	25	1
------	-----	-----	----	---

Usage of
"LILLY", per
million babies

500

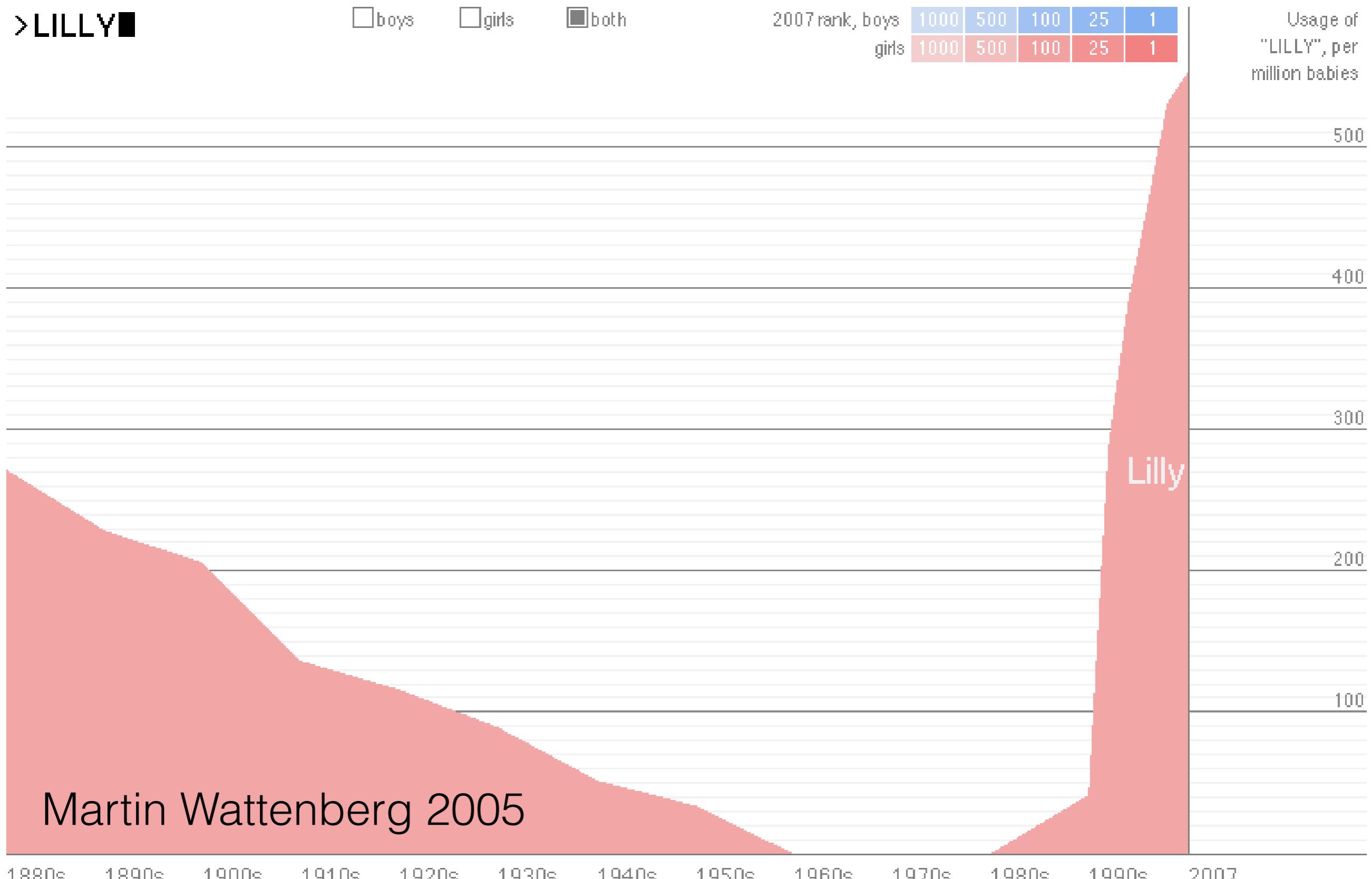
400

300

200

100

Lilly



Communicate

CS

171

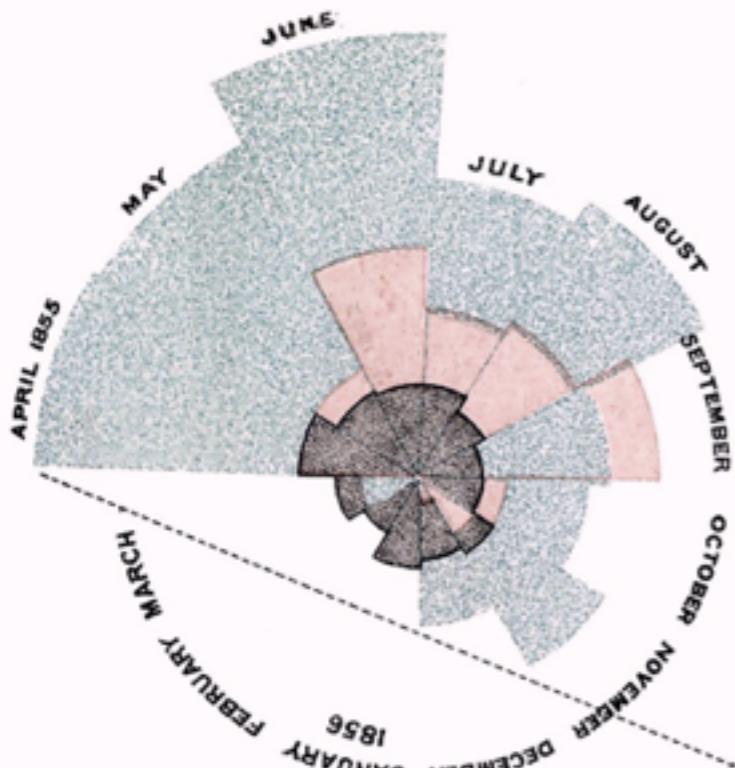


Konya town map, Turkey, c. 6200 BC



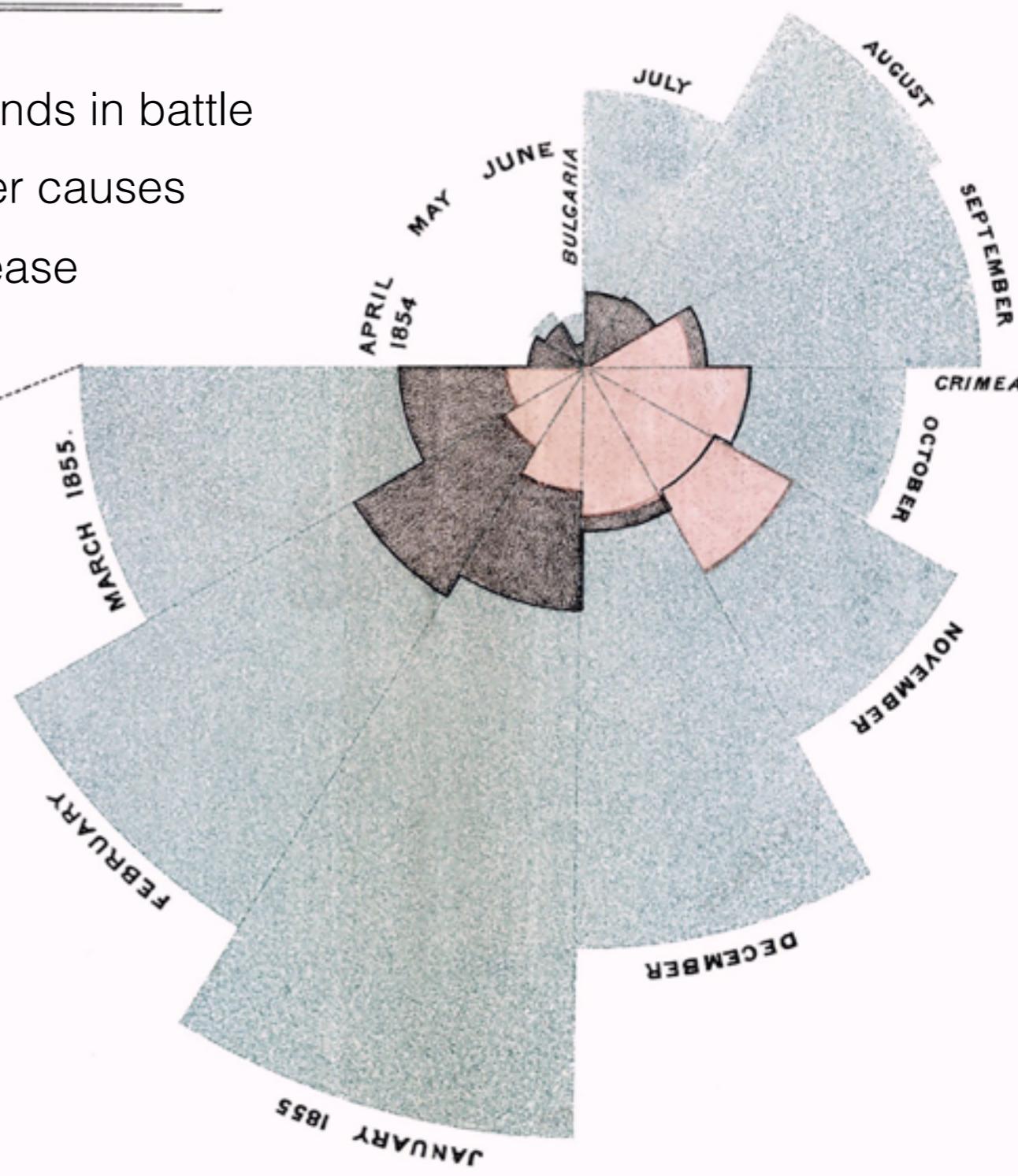
DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.



- Death from wounds in battle
- Death from other causes
- Death from disease

1.
APRIL 1854 to MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

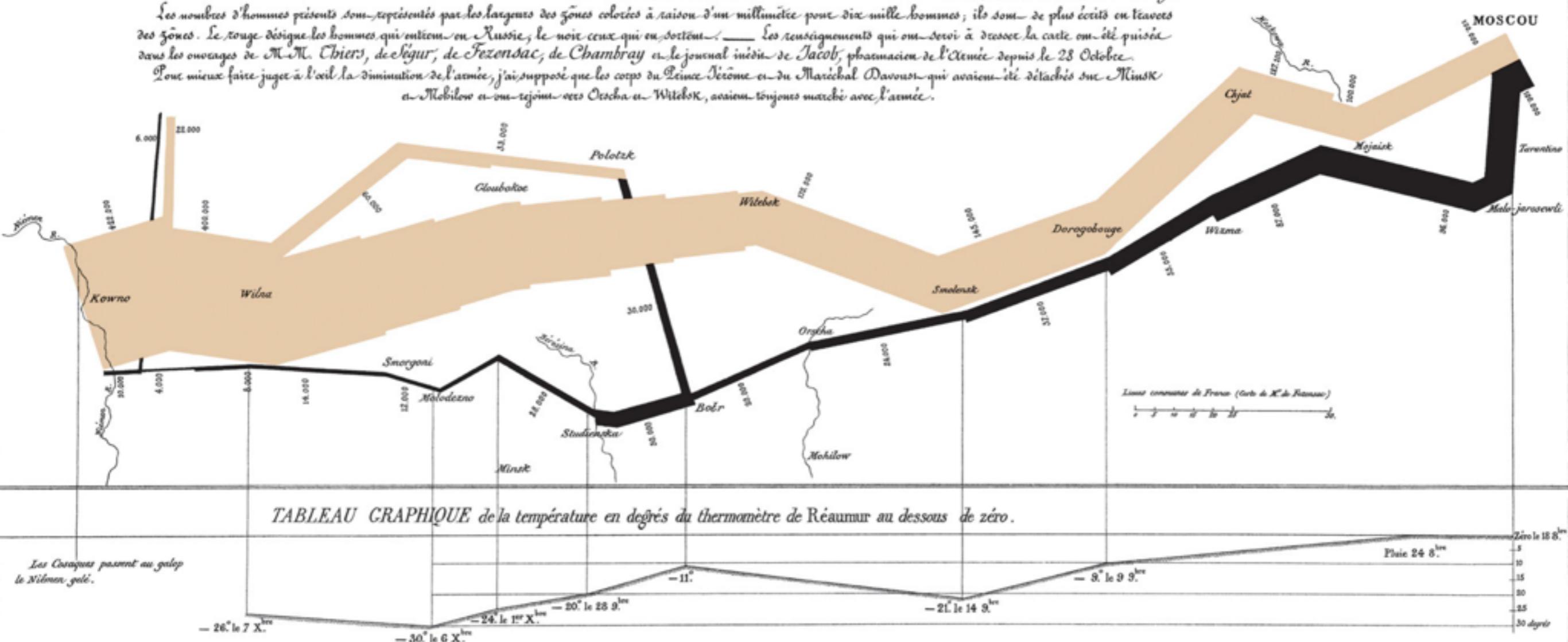
Joseph Minard 1861

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite
Paris, le 20 Novembre 1869.

Les nombres d'hommes perdus sont représentés par les larges des zones colorées à raison d'un millimètre pour dix mille hommes ; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui ont été en Russie, le noir ceux qui en sortirent. — Les renseignements qui ont servi à dresser la carte on été pris dans les ouvrages de M. Chiers, de Segur, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk à Mohilow et qui rejoignirent Orel et Wiltsk, avaient toujours marché avec l'armée.





Hans Rosling 2006

Evaluating Visualizations

CS
171





Massachusetts Bay Transportation Authority Rapid Transit/Key Bus Routes Map



Legend

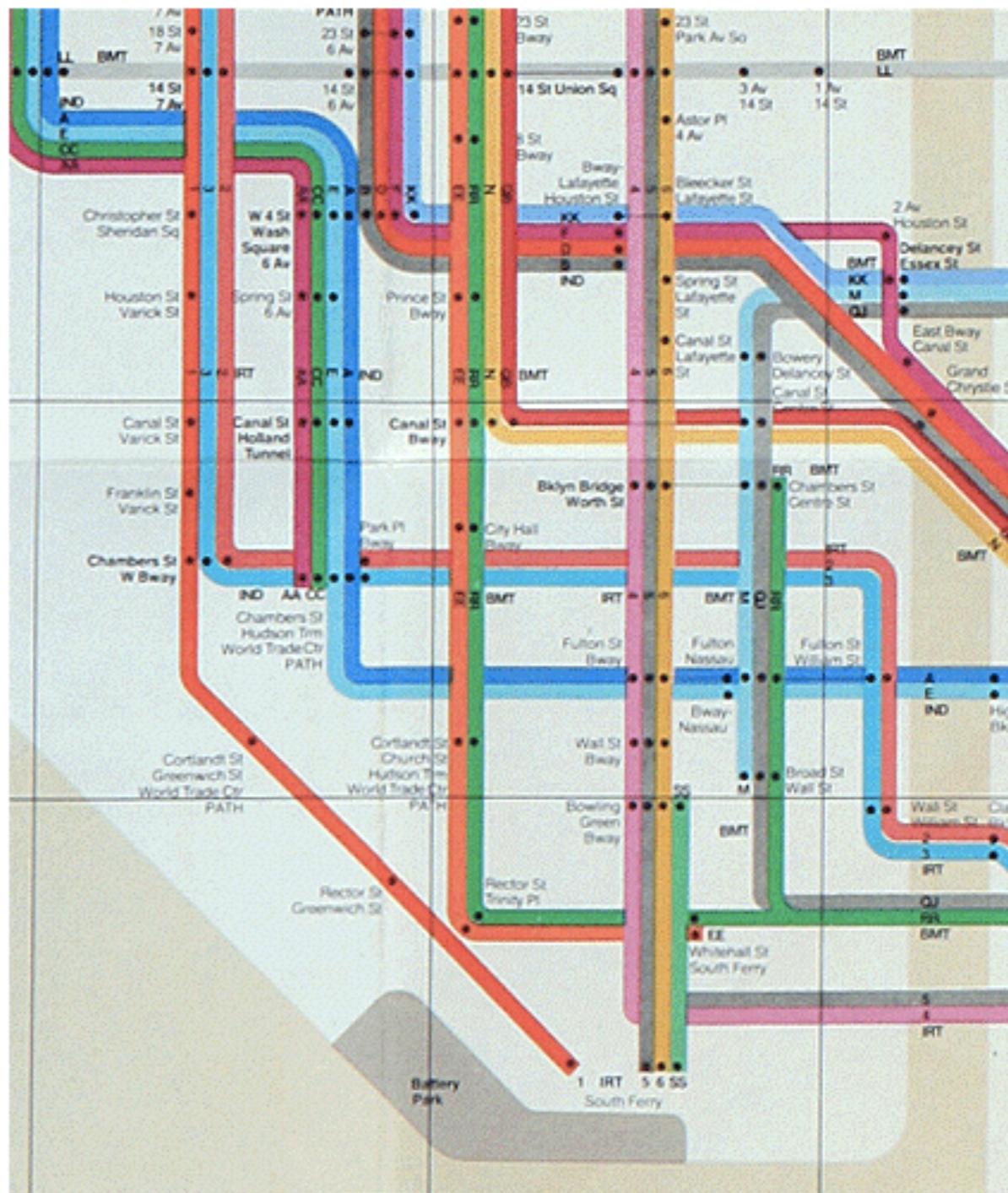
-

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 »](#)

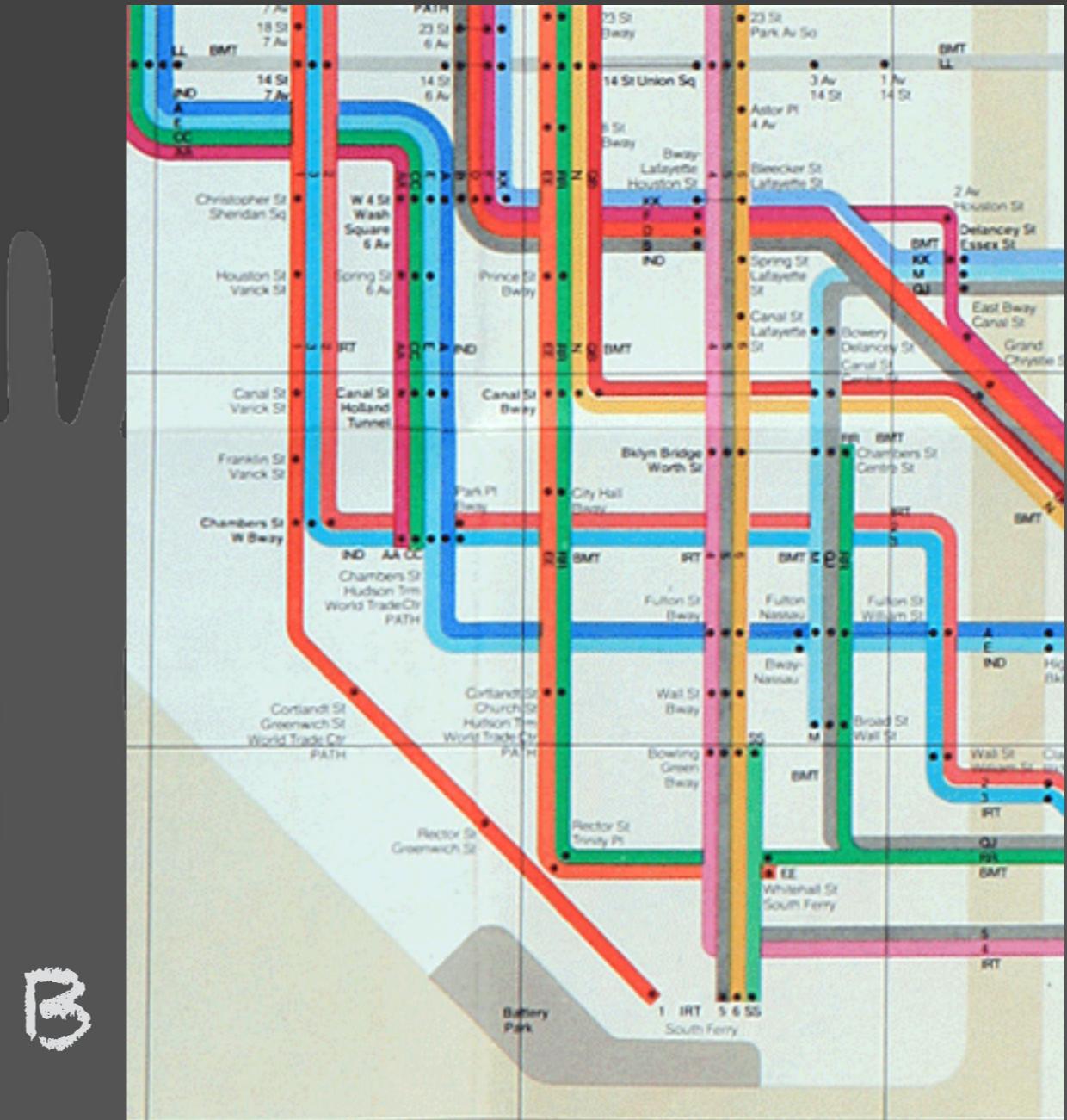


1972: Vignelli's Classic



Activity

Which of these two subway maps is better? Why? [1 min]



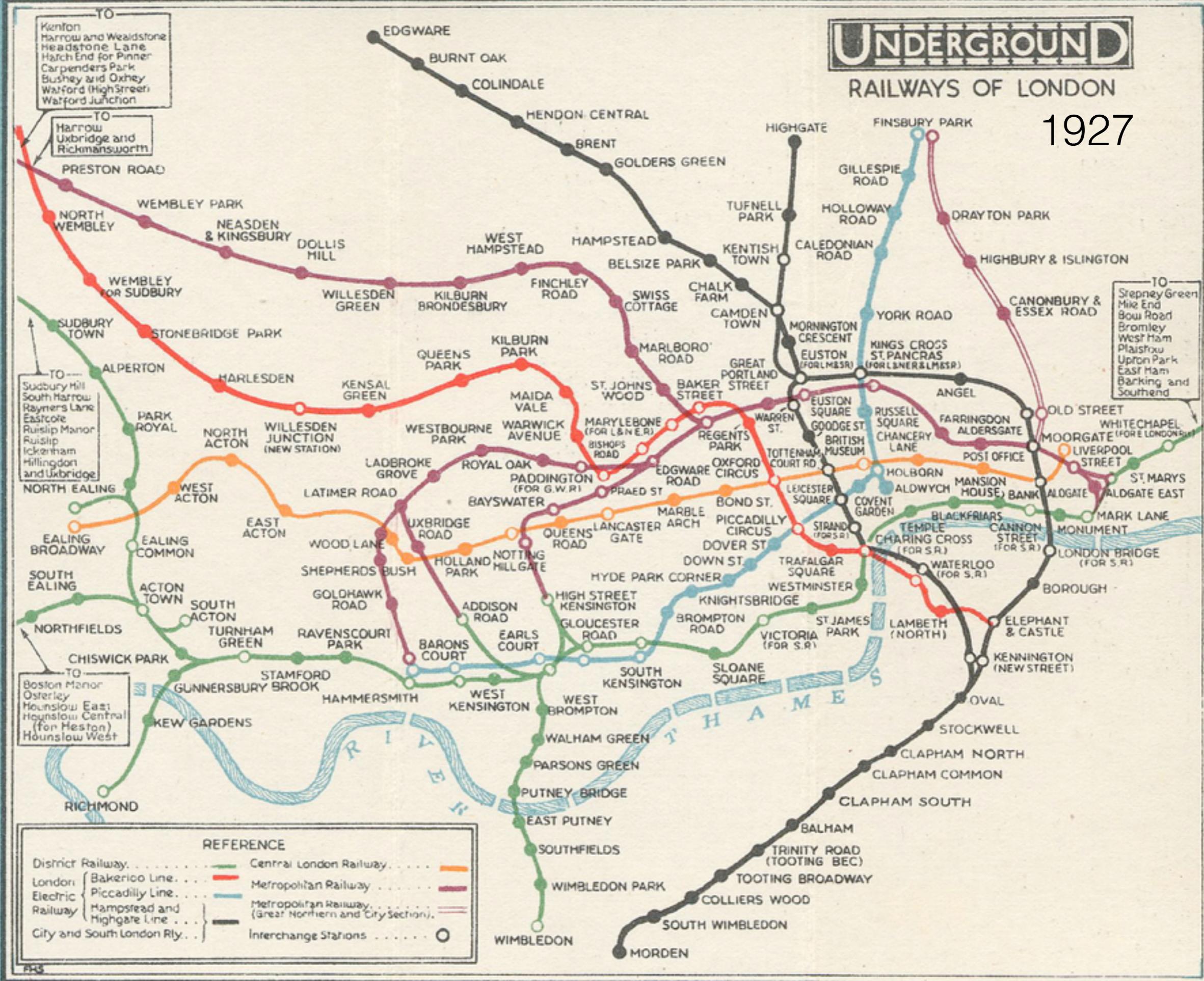
A

B

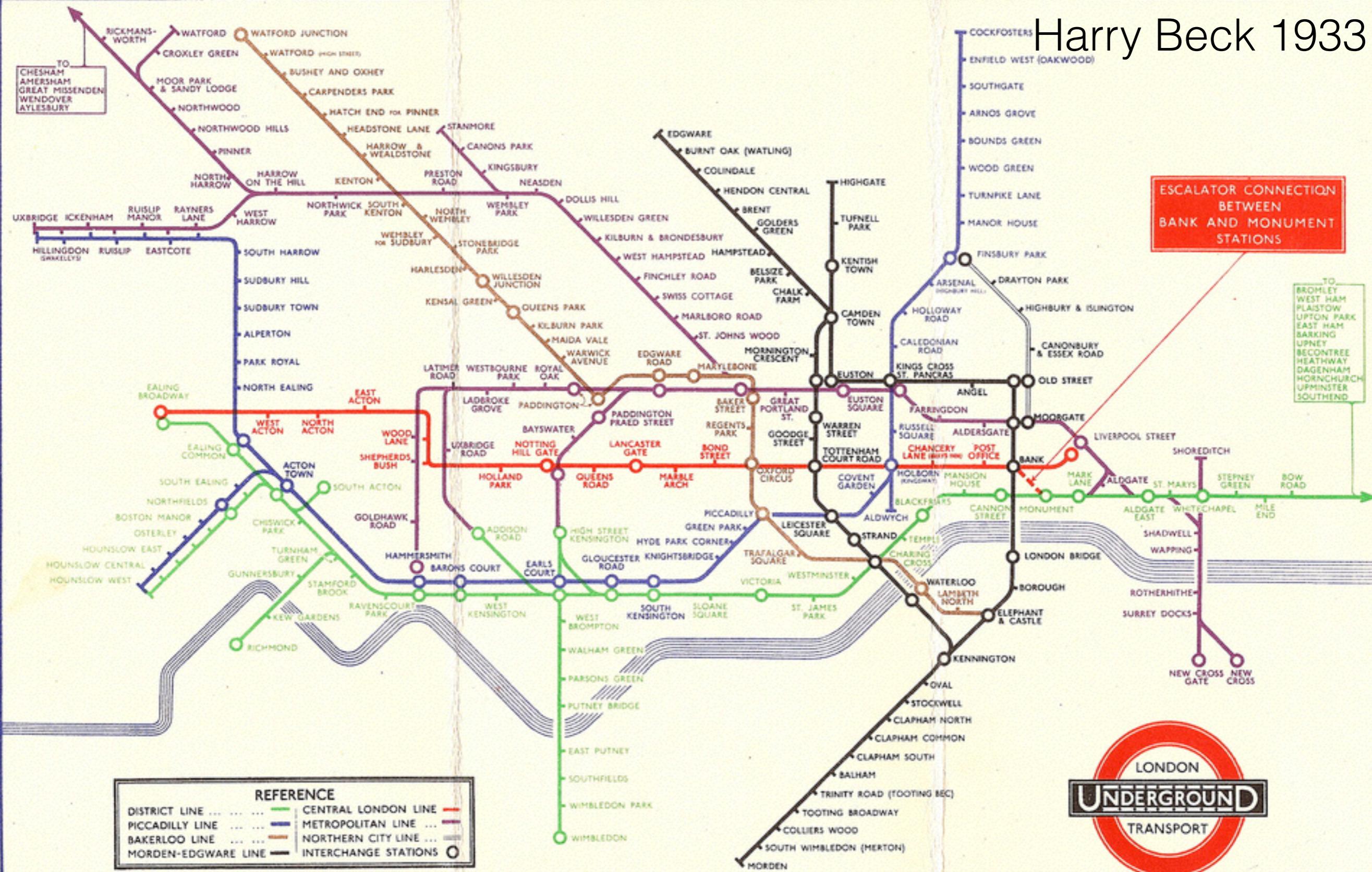
UNDERGROUND

RAILWAYS OF LONDON

1927



Harry Beck 1933



REFERENCE

DISTRICT LINE | CENTRAL LONDON LINE —
 PICCADILLY LINE | METROPOLITAN LINE
 BAKERLOO LINE | NORTHERN CITY LINE
 MORDEN-EDGWARE LINE —| INTERCHANGE STATIONS ○

WIMBLEDON PARK
 WIMBLEDON

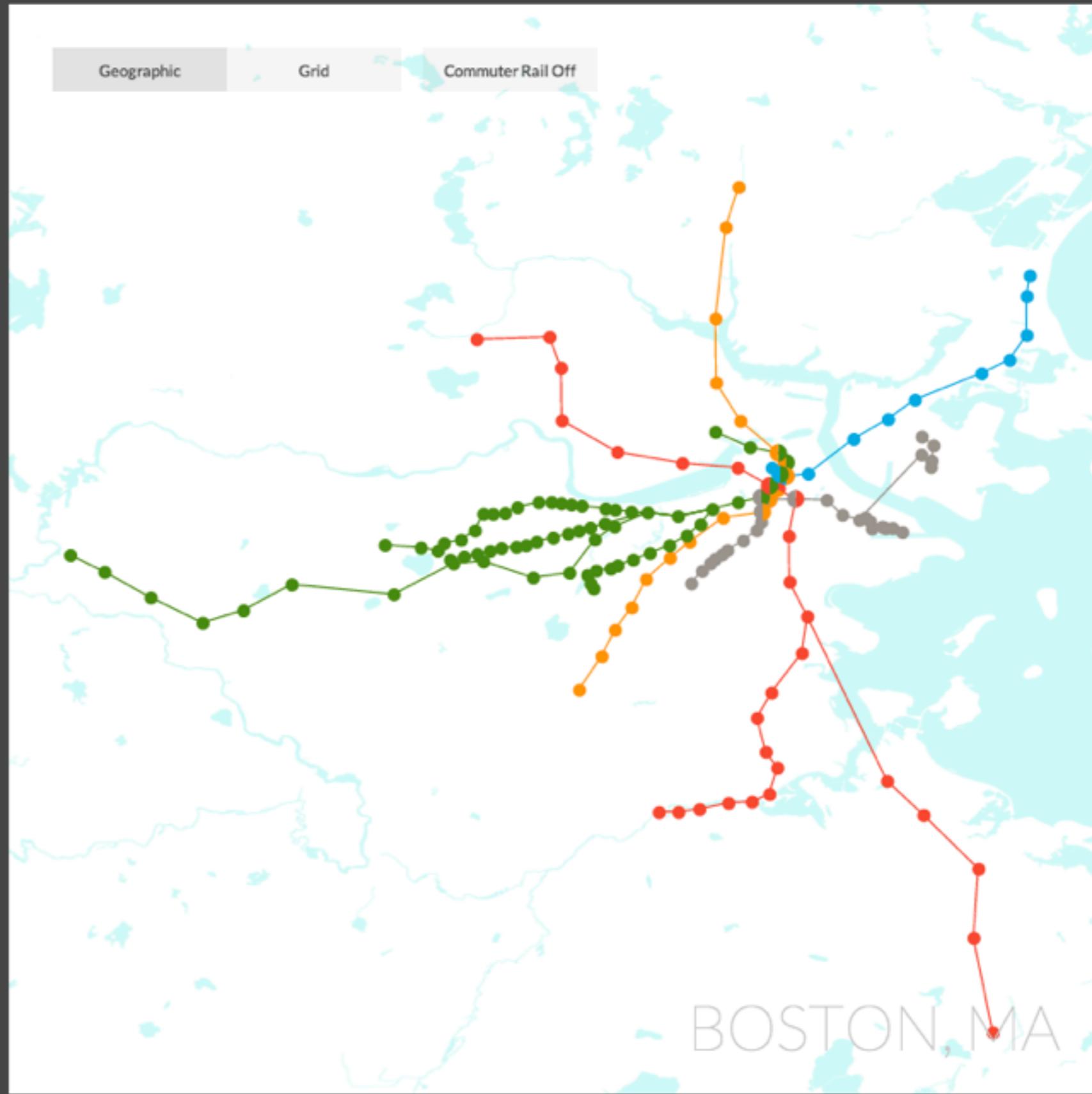
X C BECK

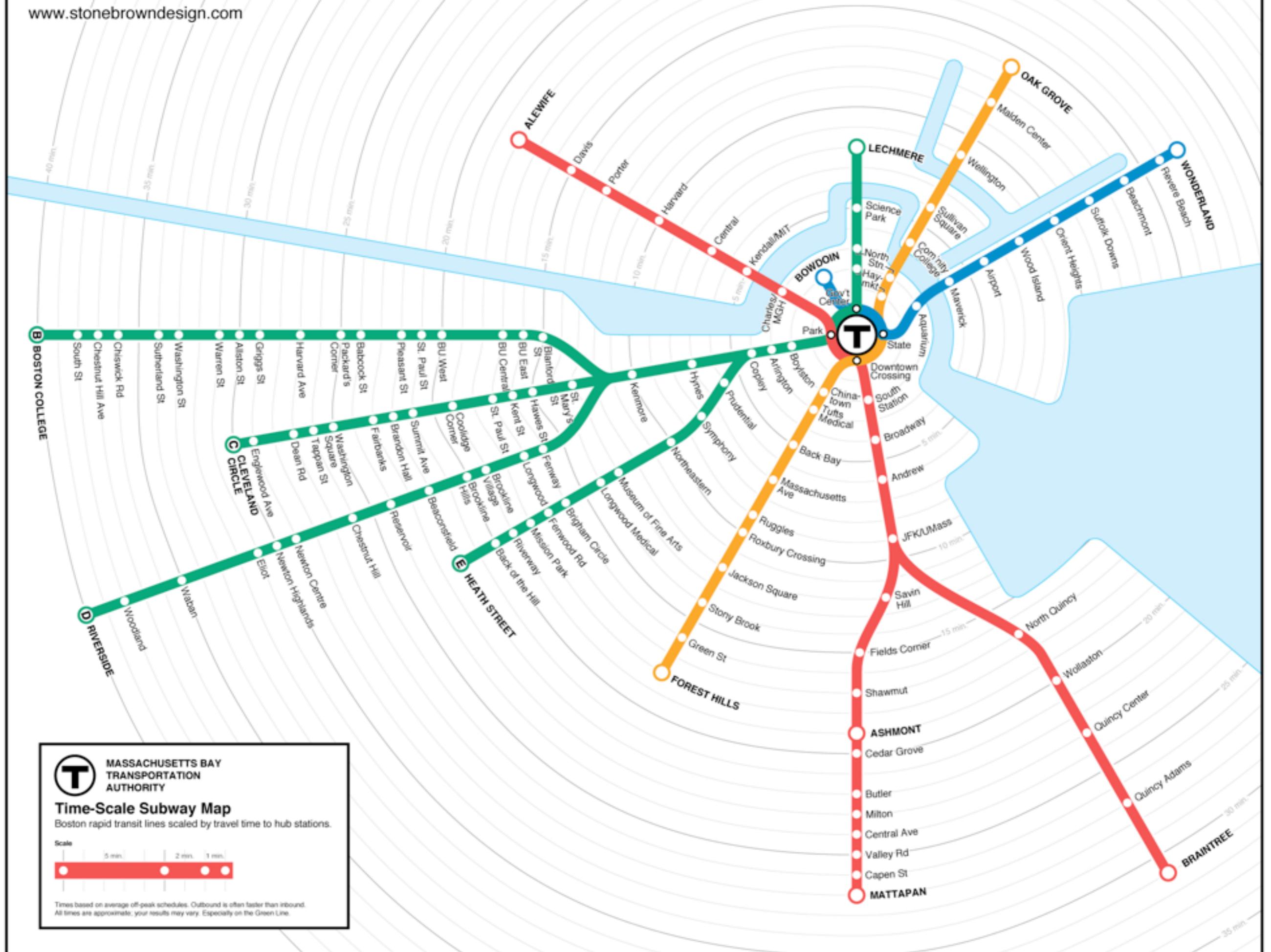
Geographic vs Topological Metro Map

FATHOM, 2013



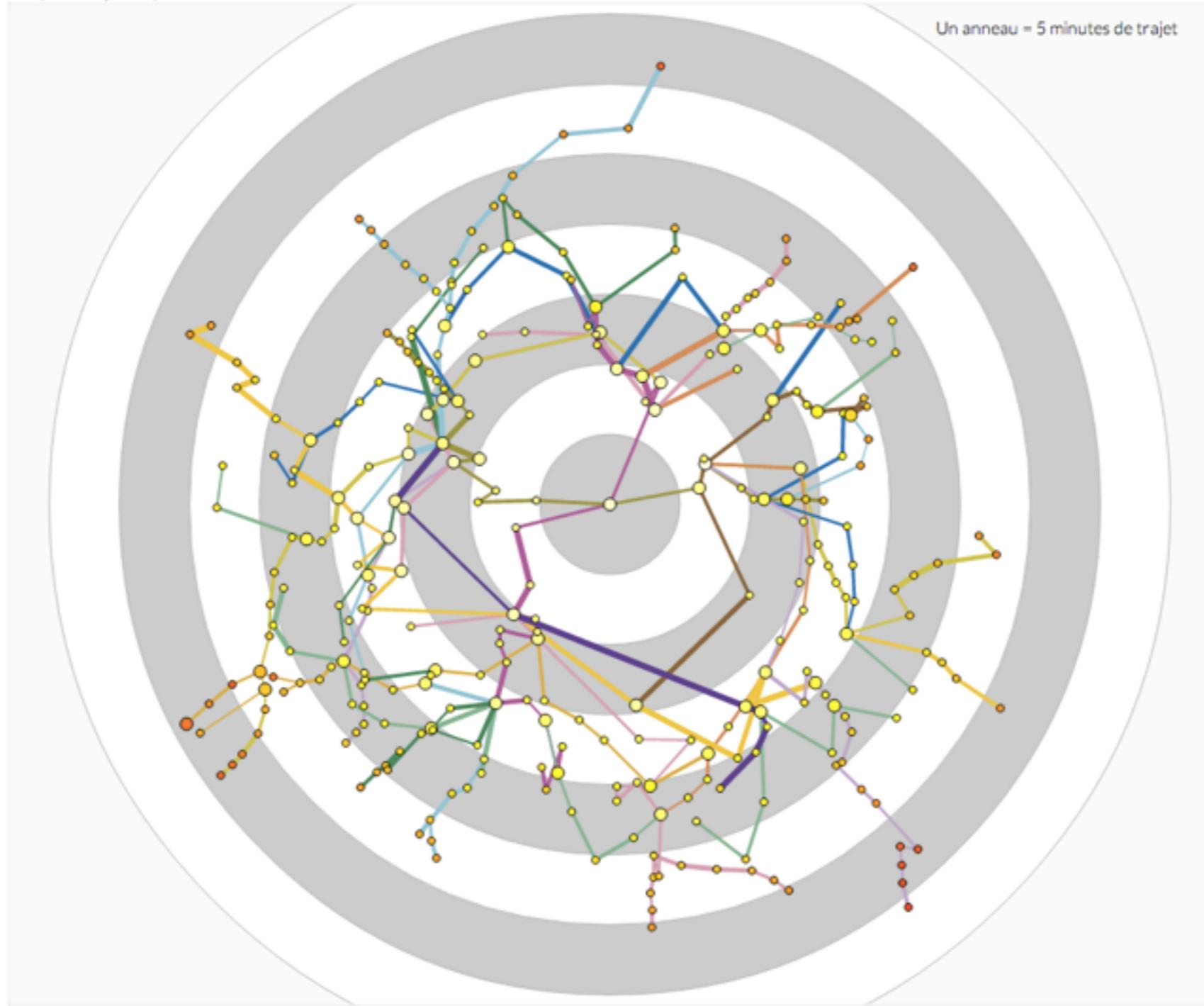
Terrence
Fradet
2013





Réaumur – Sébastopol

Temps de trajet moyen: 18 minutes 26 secondes



Utiliser les positions exactes des stations

Améliorez le plan!

Vous connaissez bien la station Réaumur – Sébastopol ? Cliquez dans les zones grises pour améliorer les estimations.

Quelle distance entre la sortie et les quais ?

Quelle est la longueur des correspondances ?

3

4

3

4

Outline

- What?
- Why?
- Who?
- How?



data

data government

SHAKESPEARE
QUARTERLY

INDUSTRIAL REVOLUTION OF DATA

Joe Hellerstein, UC Berkley, 2008



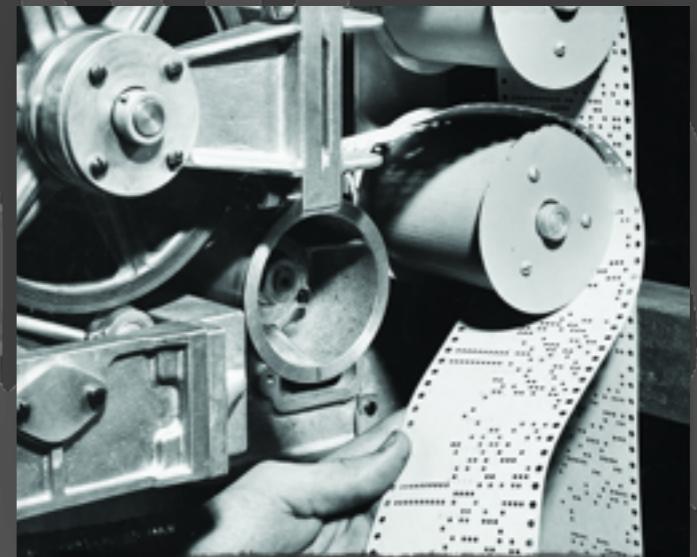
Why? An agenda for CS171

- Why Have A Human in the Loop?
- Why Have A Computer in the Loop?
- Why Use An External Representation?
- Why Depend on Vision?
- Why Show The Data In Detail?
- Why Use Interactivity?
- Why Is the Vis Idiom Design Space Huge?
- Why Focus on Tasks?
- Why Focus on Effectiveness?
- Why Are Most Designs Ineffective?
- Why Is Validation Difficult?
- Why Are There Resource Limitations?
- Why Analyze?

Activity

Imagine a system to analyze large amounts of data. Collect and discuss arguments answering the questions:

Why would you have a human in the loop?



Why would you have a computer in the loop?

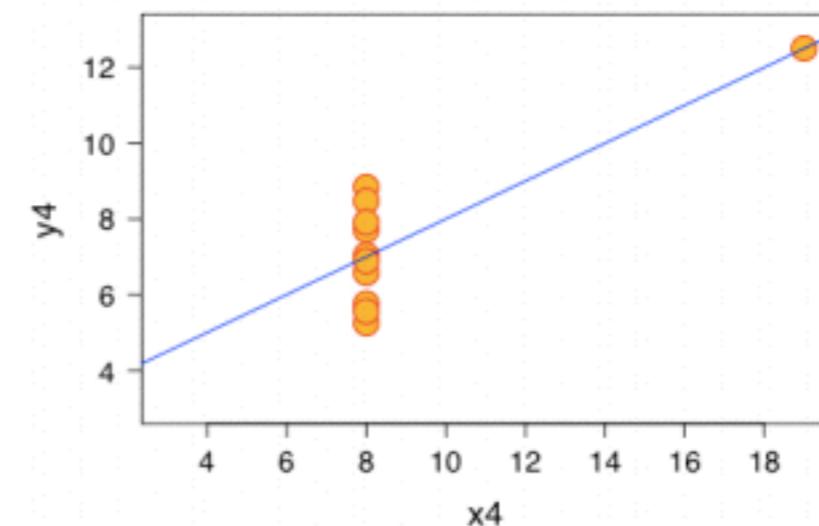
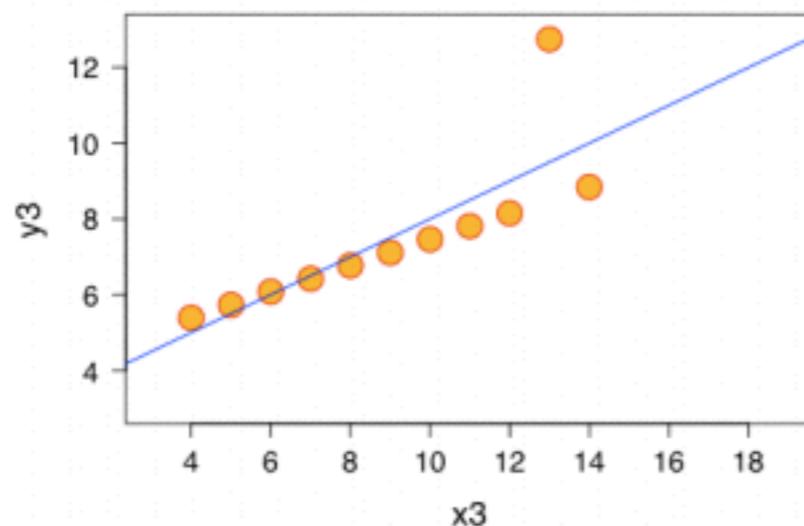
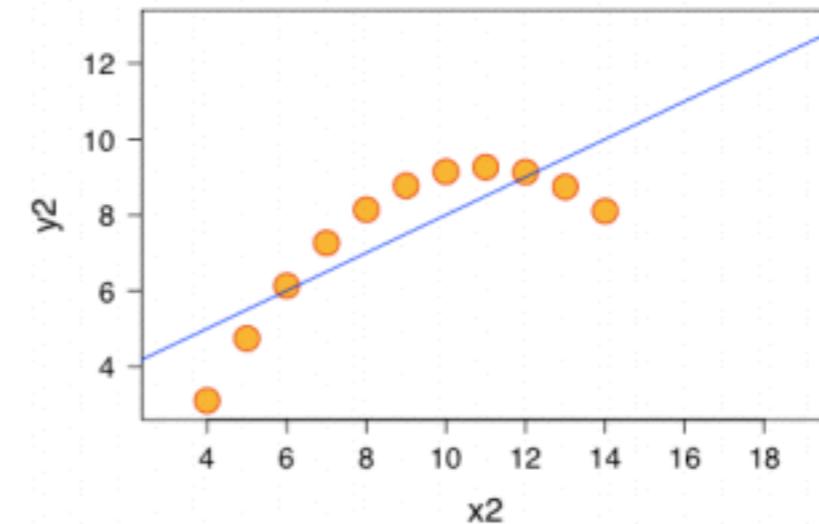
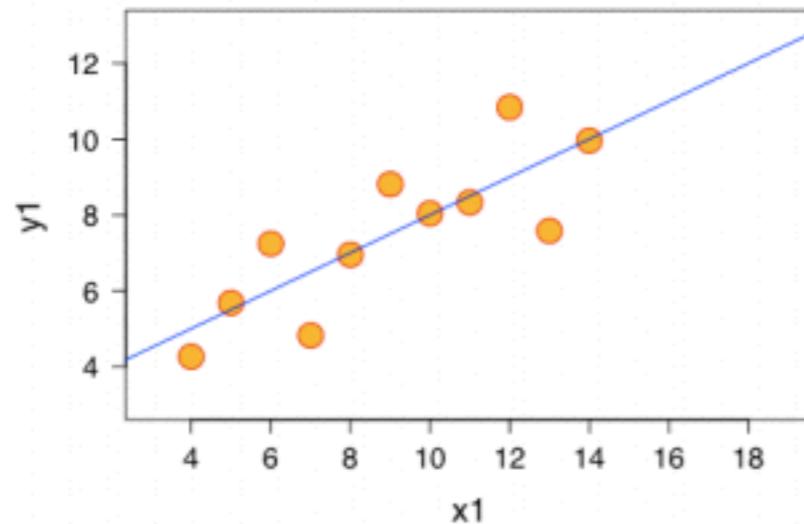
Anscombe's Quartet

Anscombe's Quartet: Raw Data

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	
mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
var.	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
corr.		0.816		0.816		0.816		0.816

Anscombe's Quartet

Same mean, variance, correlation, and linear regression line

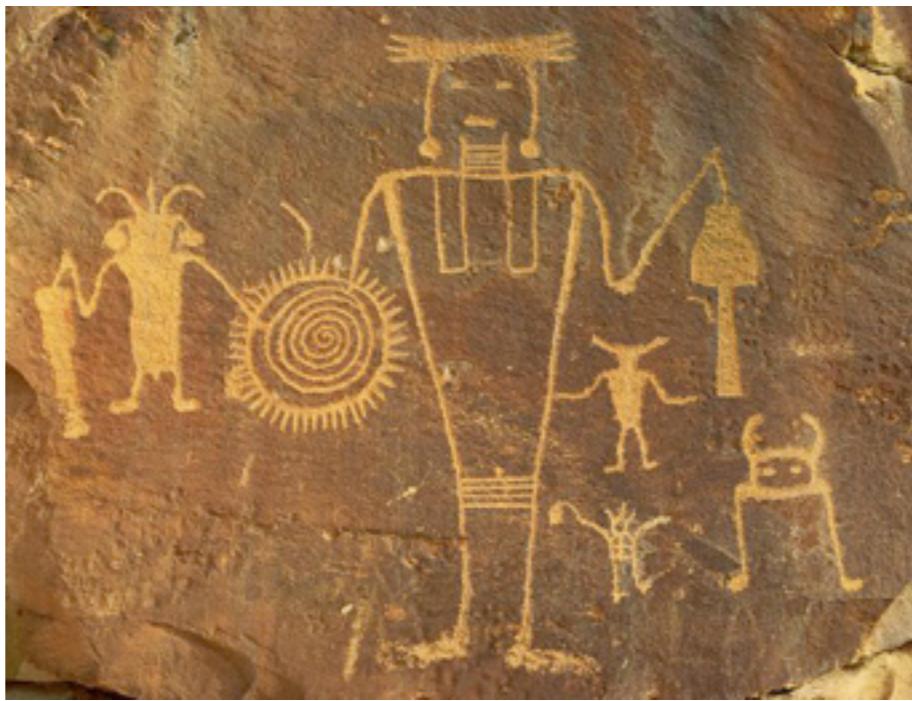


Daniel Simons 1998



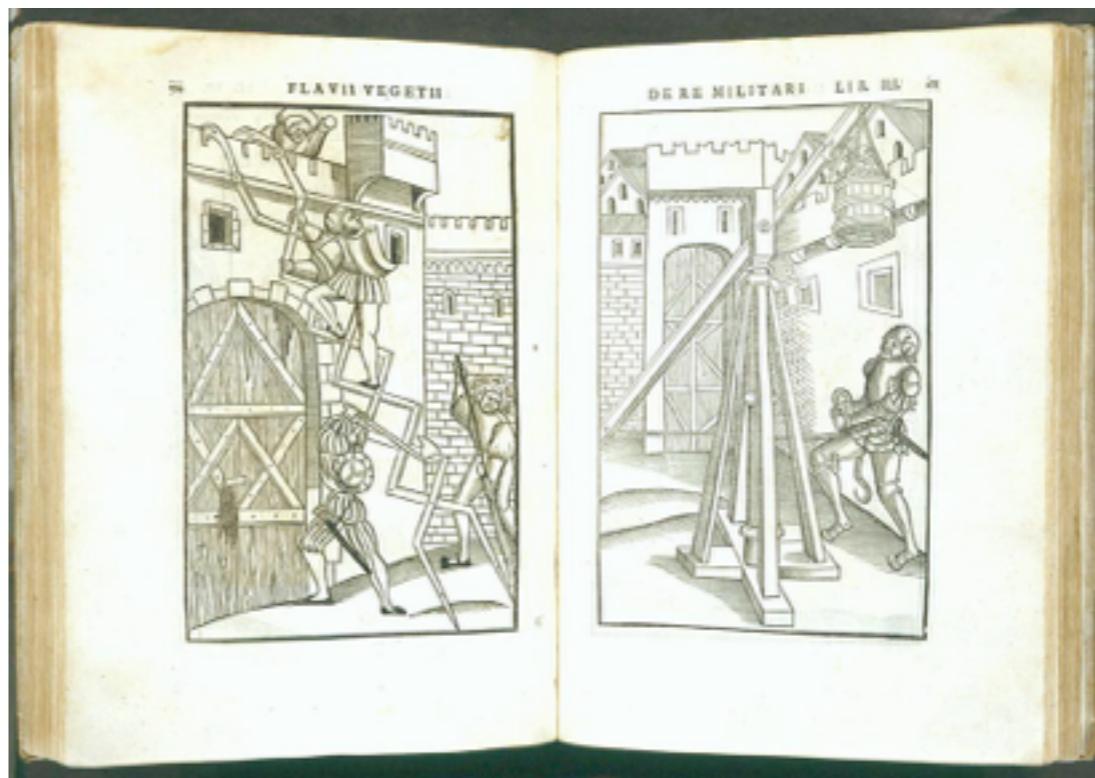
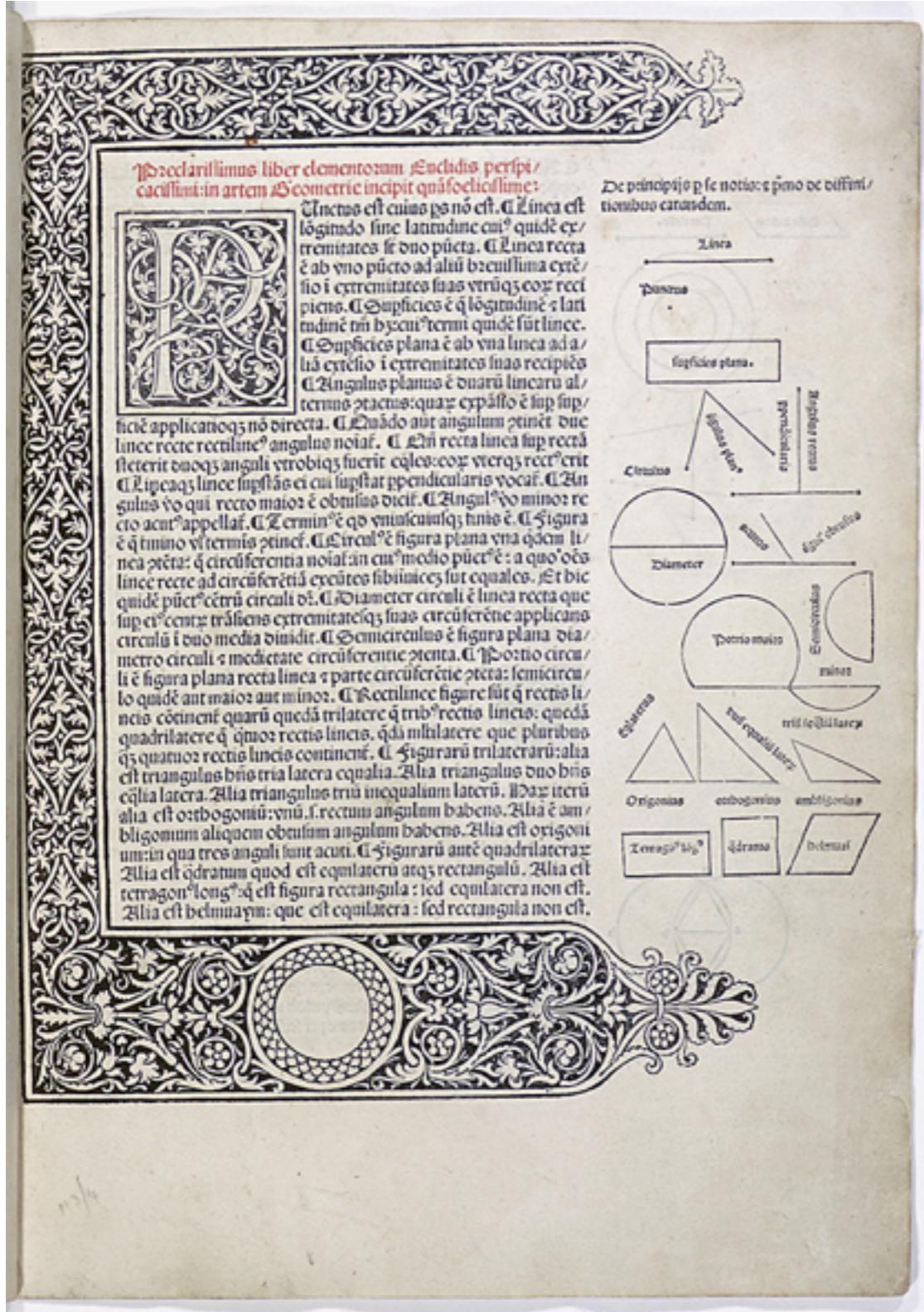
“It is things that make us smart”

Donald Norman



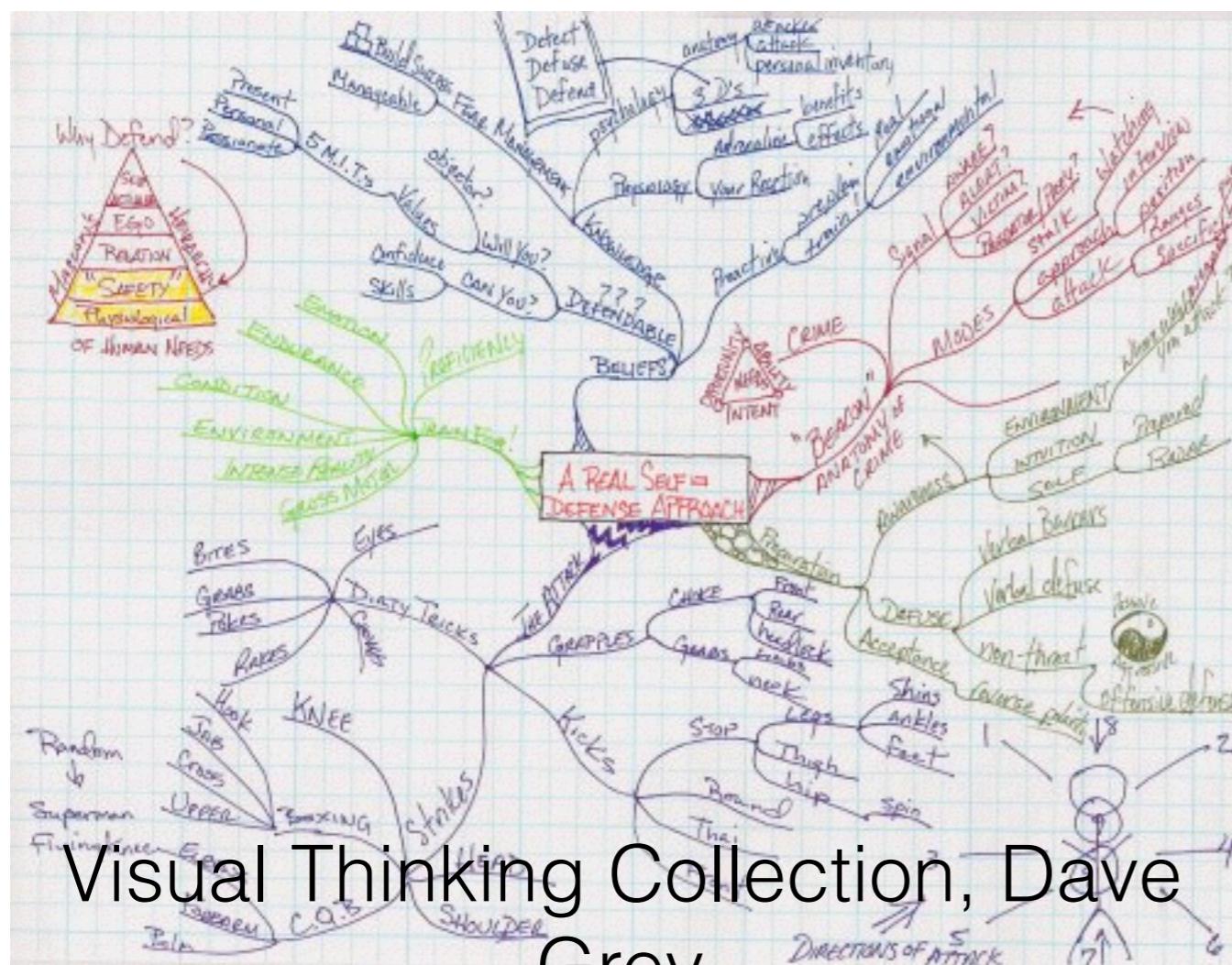
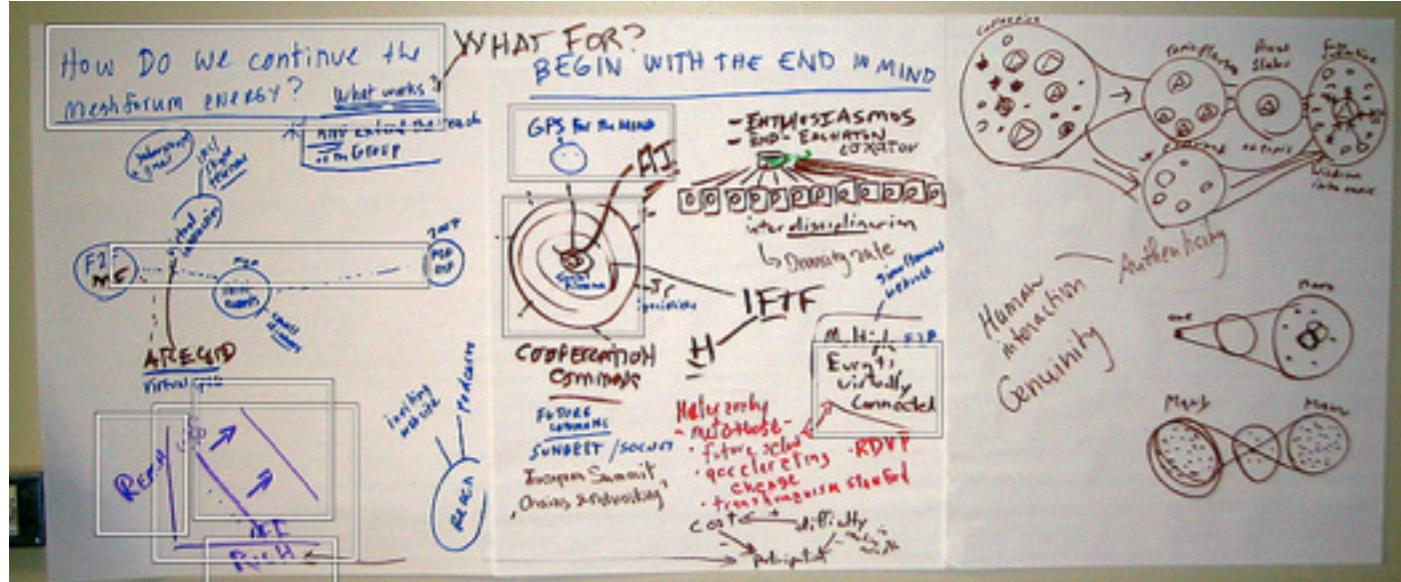
“It is things that make us smart”

Donald Norman



“It is things that make us smart”

Donald Norman



Why depend on vision?

MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHLVKMFPSTWYARN
GFPSVCEILQGKMFPSNDRCEQDIFP
SGHLMFHKMVPSTWYACEQTWRN

Why depend on vision?

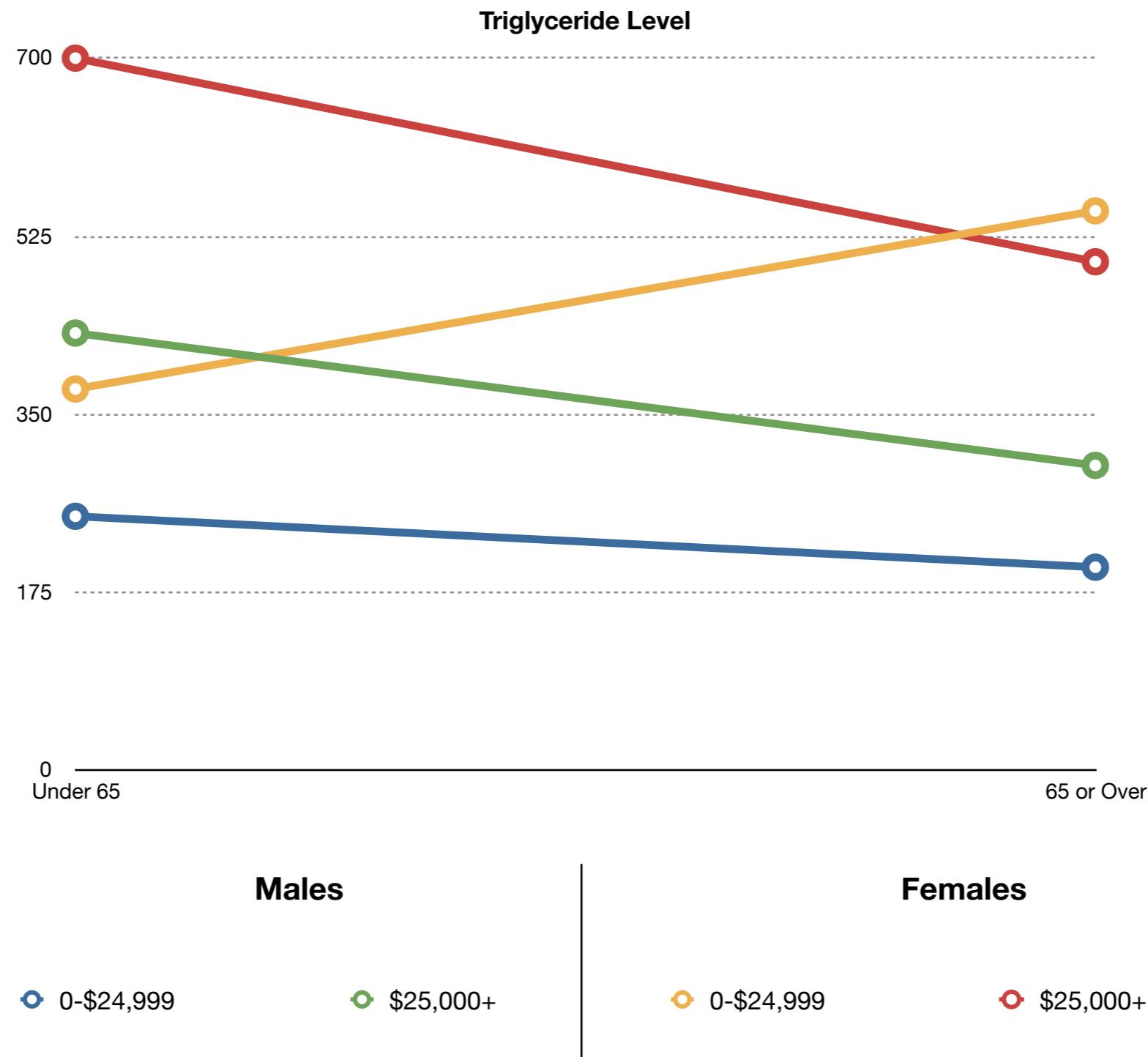
MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHLVKMFPSTWYARN
GFPSVCEILQGKMFPSNDRCEQDIFPS
GHLMFHKMVPSTWYACEQTWRN

Why use graphical representations?

Which gender or income level group shows different effects of age on cholesterol levels?

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

Why use graphical representations?



Stephen Kosslyn “Clear and to the Point”



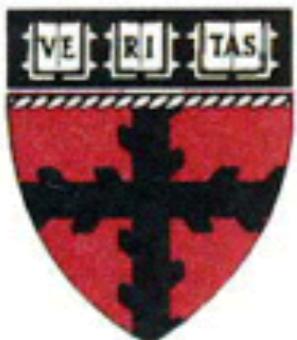
“Visualization is really about **external cognition**, that is, how resources outside the mind can be used to boost the cognitive capabilities of the mind.”

—Stuart Card

Outline

- What?
- Why?
- Who?
- How?

Hanspeter Pfister



Harvard
School of Engineering
and Applied Sciences

Hanspeter Pfister

An Wang Professor of Computer Science

Maxwell Dworkin 227
33 Oxford Street
Cambridge, MA 02138

Tel: 617.496.8269
Fax: 617.496.3012

pfister@seas.harvard.edu

gvi.seas.harvard.edu/pfister

Twitter: @hpfister

Visual Computing Group



Dr. Hanspeter Pfister

Dr. Dequing Sun

Dr. James Tompkin

Dr. Adi Peleg

Dr. Hendrik Strobelt

Dr. Kasper Dinkla

Michaela Kapp

Dr. Johanna Beyer

Daniel Haehn

Gaurav Bharaj

Nam Wook Kim

Michail Schwab

Dr. Verena Kaynig-Fittkau

Michael Oppermann

Dr. Ray Thouis Jones



Visual Computing Group

Professor Hanspeter Pfister



News Projects Publications Presentations People Code and Data Classes
Open Positions Contact



VISUAL COMPUTING

Our research in visual computing lies at the intersection of visualization, computer graphics, and computer vision. It spans a wide range of topics, including bio-medical visualization, image and video analysis, 3D fabrication, and data science.

OUR LAB

Our group belongs to Harvard's School of Engineering and Applied Sciences and the Center for Brain Science. We are located in the Maxwell Dworkin Building (33 Oxford St.) as well as the Northwest Laboratory (52 Oxford St.) on Harvard's main campus in Cambridge, Massachusetts.

OUR RESEARCH

Our goal is to combine interactive computer systems with the perceptual and cognitive power of human observers to solve practical problems in science and engineering. We are providing visual analysis tools and methods to help scientists and researchers better process and understand large, multi-dimensional data sets in various domains such as neuroscience, genomics, systems biology, astronomy, and medicine. And we are developing data-driven approaches for the acquisition, modeling, visualization, and fabrication of complex objects.

RECENT PUBLICATIONS

- [A Crowdsourced Alternative to Eye-tracking for Visualization Understanding](#)
- [State-of-the-Art in GPU-Based Large-Scale Volume Visualization](#)
- [Large-Scale Automatic Reconstruction of Neuronal Processes from Electron Microscopy Images](#)
- [Computational Design of Walking Automata](#)
- [Local High-order Regularization on Data Manifolds](#)
- [Semi-supervised Learning with Explicit Relationship Regularization](#)

[More ▶](#)

vcg.seas.harvard.edu

CS171 Staff



Johanna Beyer (Head TF) - Postdoctoral Fellow, Visual Computing Group

Kasper Dinkla - Postdoctoral Fellow, Visual Computing Group

Hendrik Strobelt - Postdoctoral Fellow, Visual Computing Group

James Tompkin - Postdoctoral Fellow, Visual Computing Group

Jack Birger - Civil Engineer with a passion for technology and visualization

Niamh Durfee - Chemistry and Computer Science Concentrator

Mirhee Kim - Junior, Molecular Cellular Biology (MCB) Concentrator

Nam Wook Kim - PhD Student, Visual Computing Group

Benjy Levin - Computer Science Concentrator

Lezhi Li - Master student, Harvard Graduate School of Design

Maria Lovett - Computer Science Concentrator

Ana Marinovic - Social Studies Concentrator

Kunal Mehta

Rachael Smith - Philosophy Concentrator, Secondary in Computer Science

Zona Kostic - Postdoctoral Researcher, Web-based Interactive Visualization

Alain Ibrahim - Web Developer

Andrew Reece - PhD Student in Psychology

Brandon Tineo

Gus Wezerek

Michael Oppermann - Lab & Studio master

About You

CS

171



Outline

- What?
- Why?
- Who?
- How?

CS 171 Goals

Evaluate and **critique** visualization designs

Implement interactive data visualizations

Apply fundamental principles & techniques

Develop a substantial visualization project

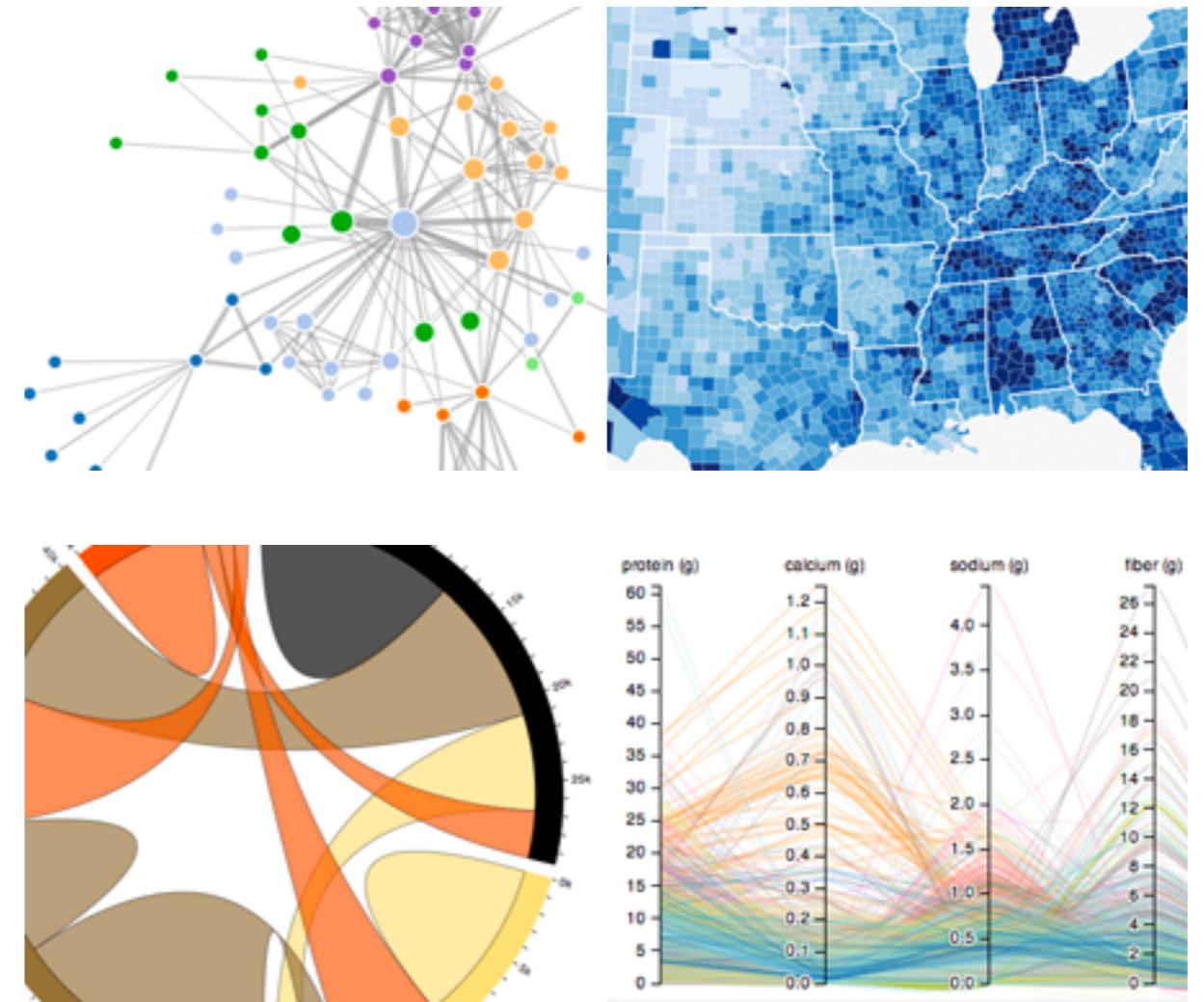
Principles

- Design
- Perception
- Cognition
- Interaction
- Process



Methods

- Temporal
- Hierarchical
- Relational
- Spatial
- Text
- High-dimensional

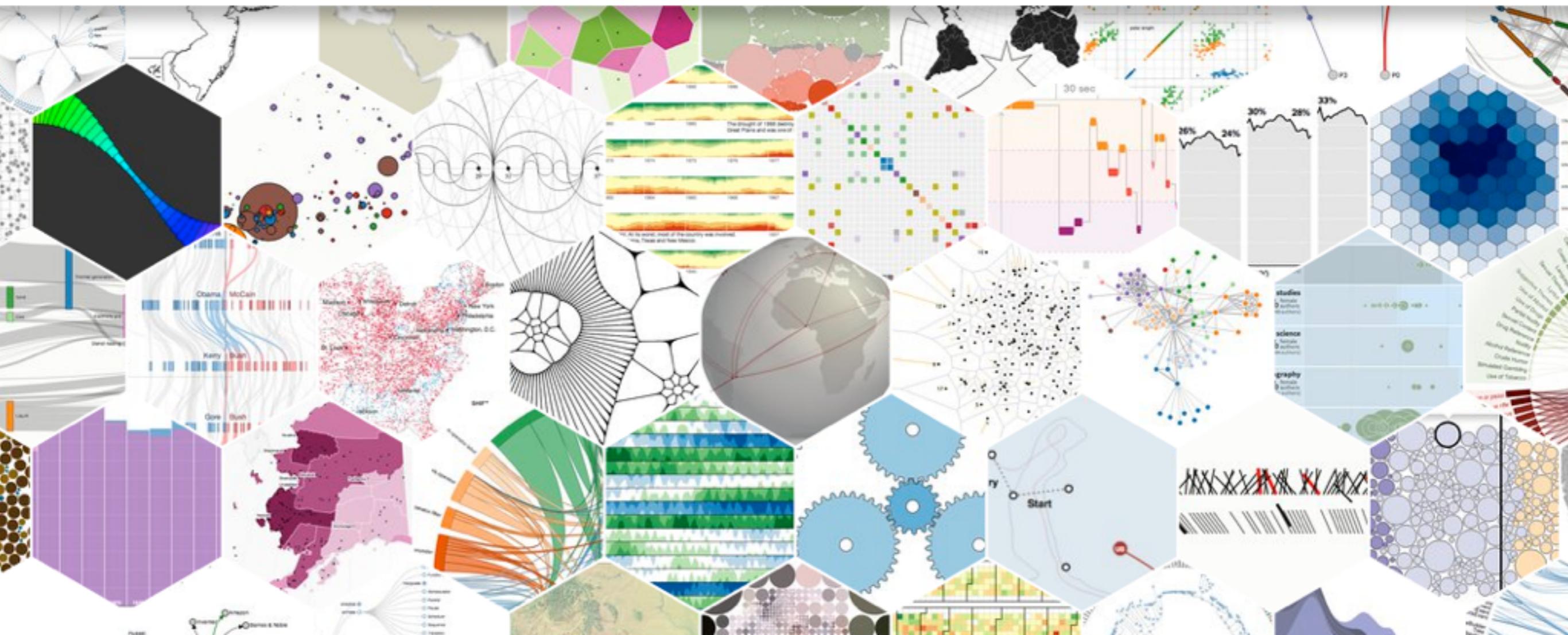


Implementation

[Overview](#) [Examples](#) [Documentation](#) [Source](#)



Data-Driven Documents



Course Structure

CS

171



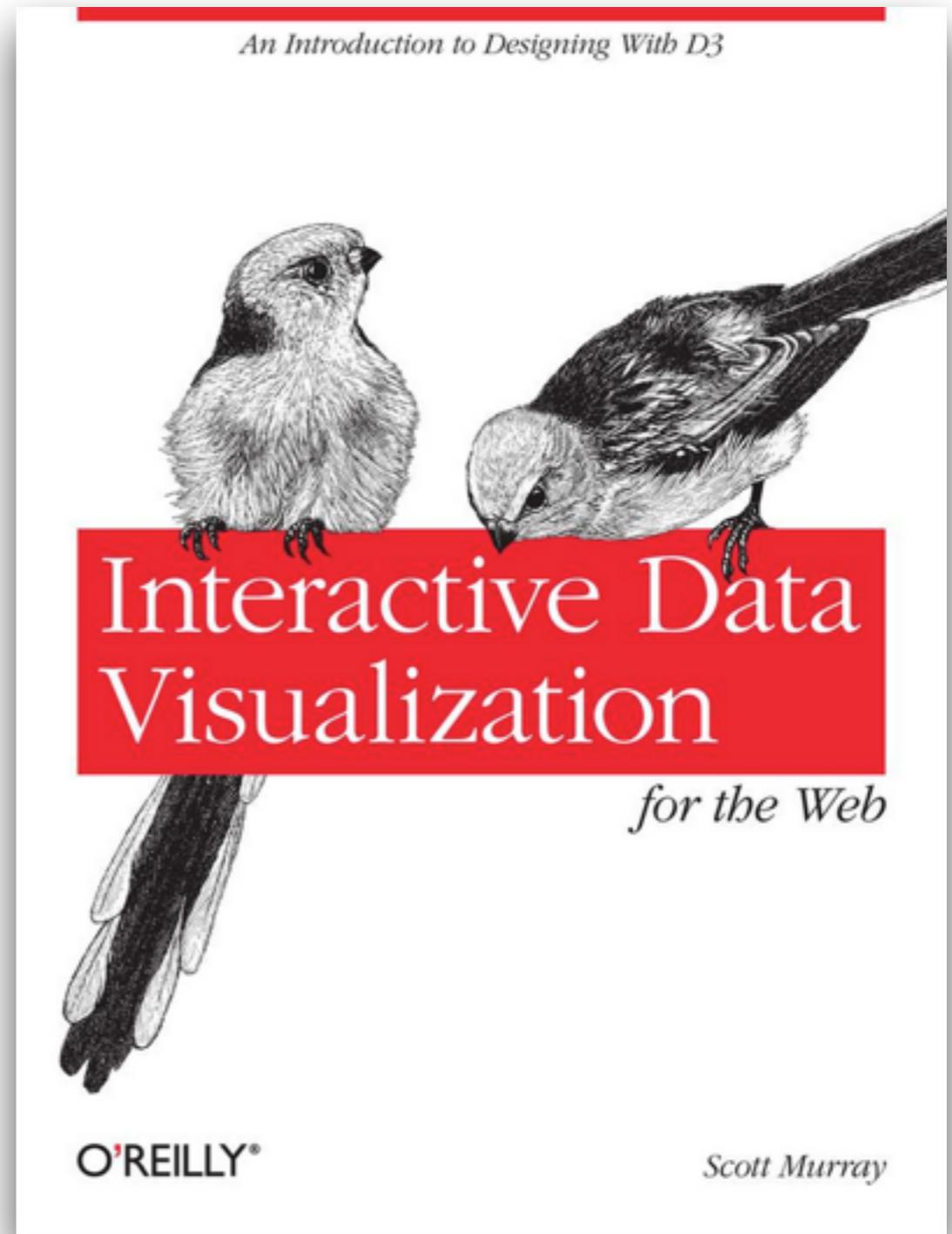
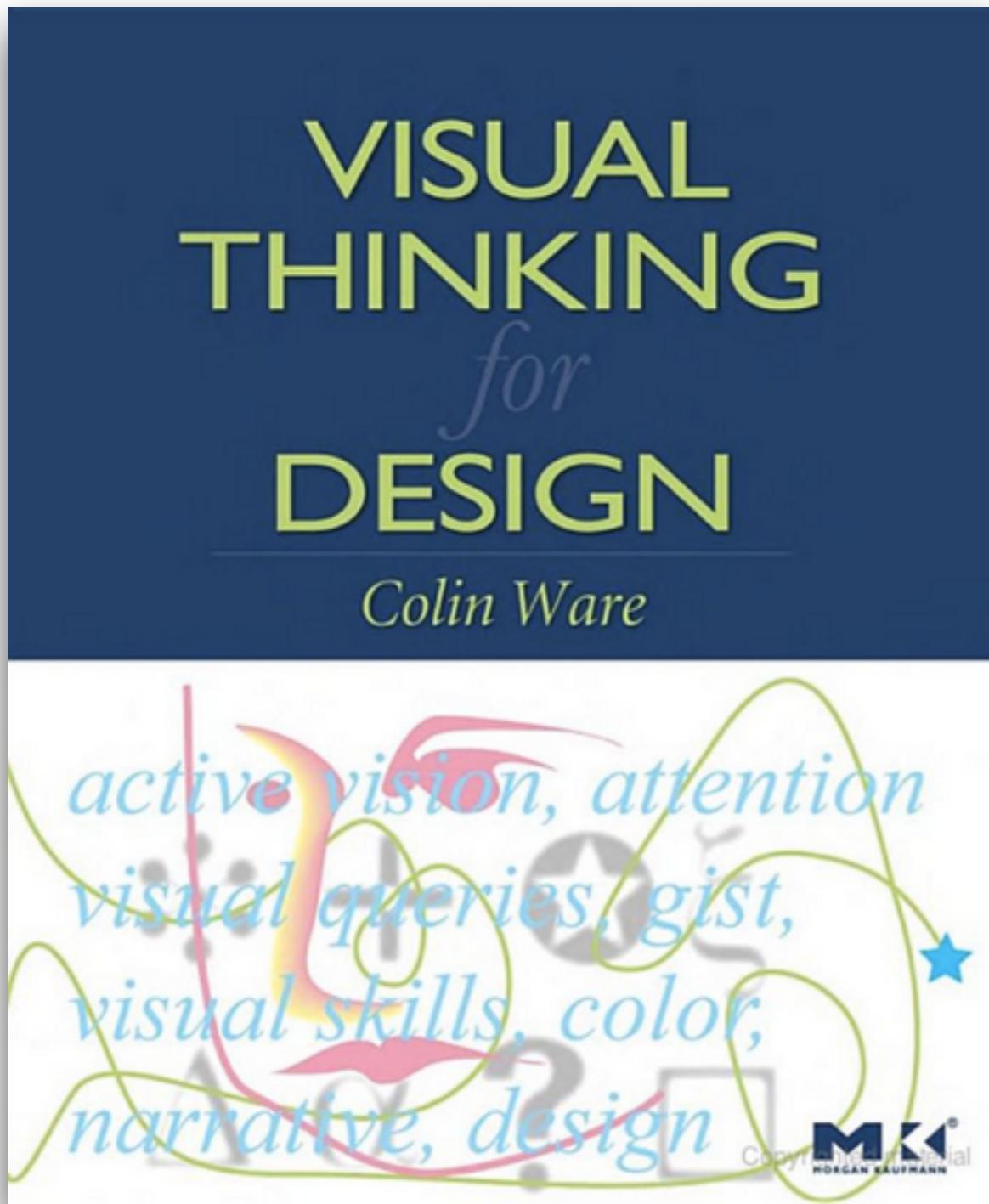
Lectures & Labs



Tu & Th 2:30-4 pm
MD G115 / G125



Readings



Studios

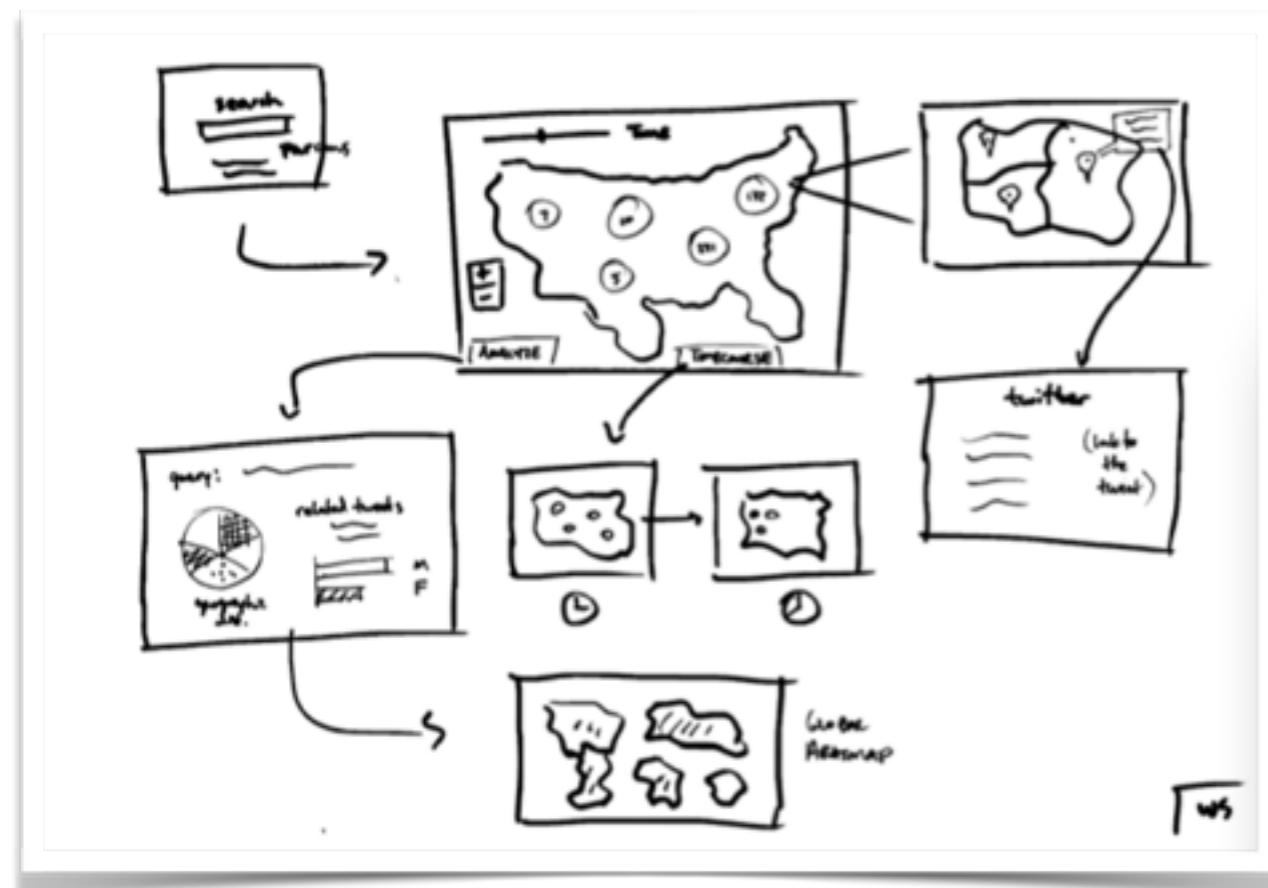


Weekly (TBA), 60-90 mins
Mandatory

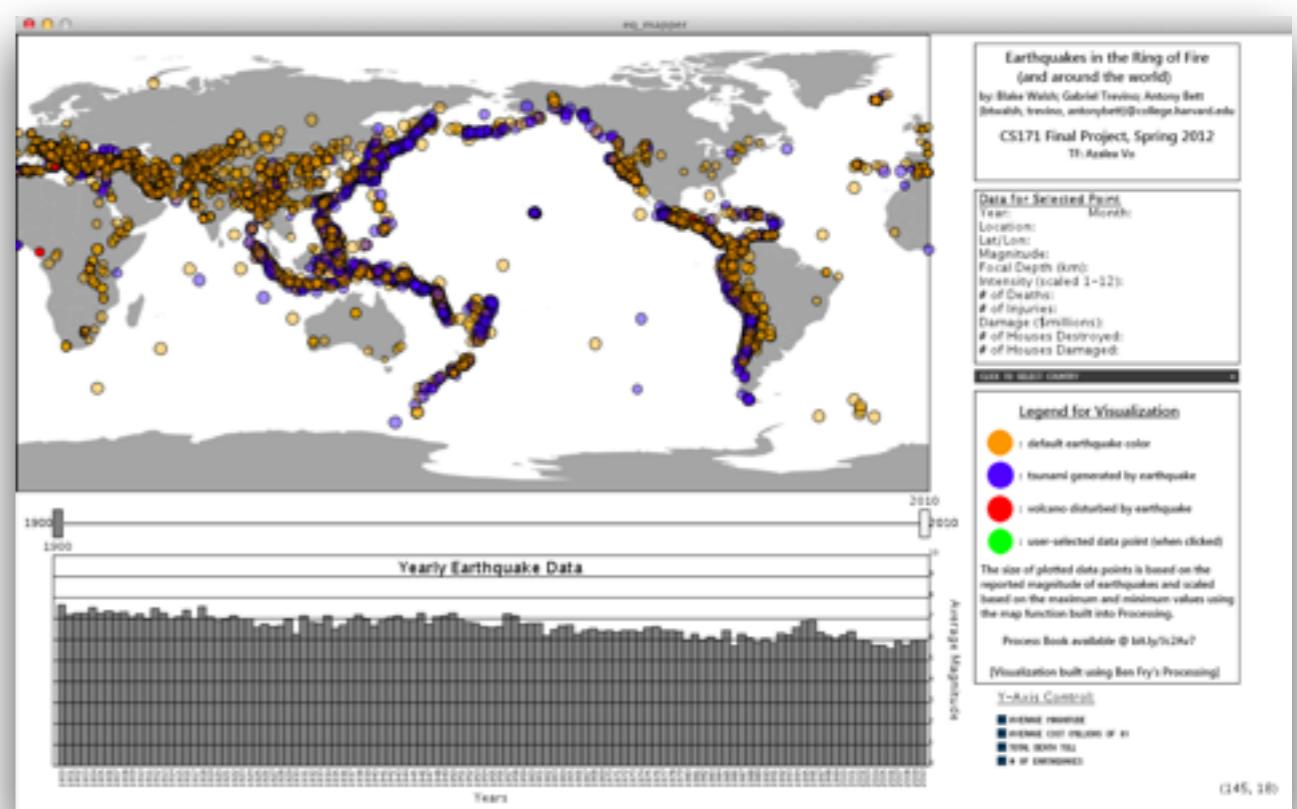
Weekly Schedule

old unit	new unit					
Mon	Tue	Wed	Thu	Fri	Sat	Sun
11:59pm Homework due	2:30-4pm Lecture (release of homework; release of lab for online students)		2:30-4:00pm Lab Studios			

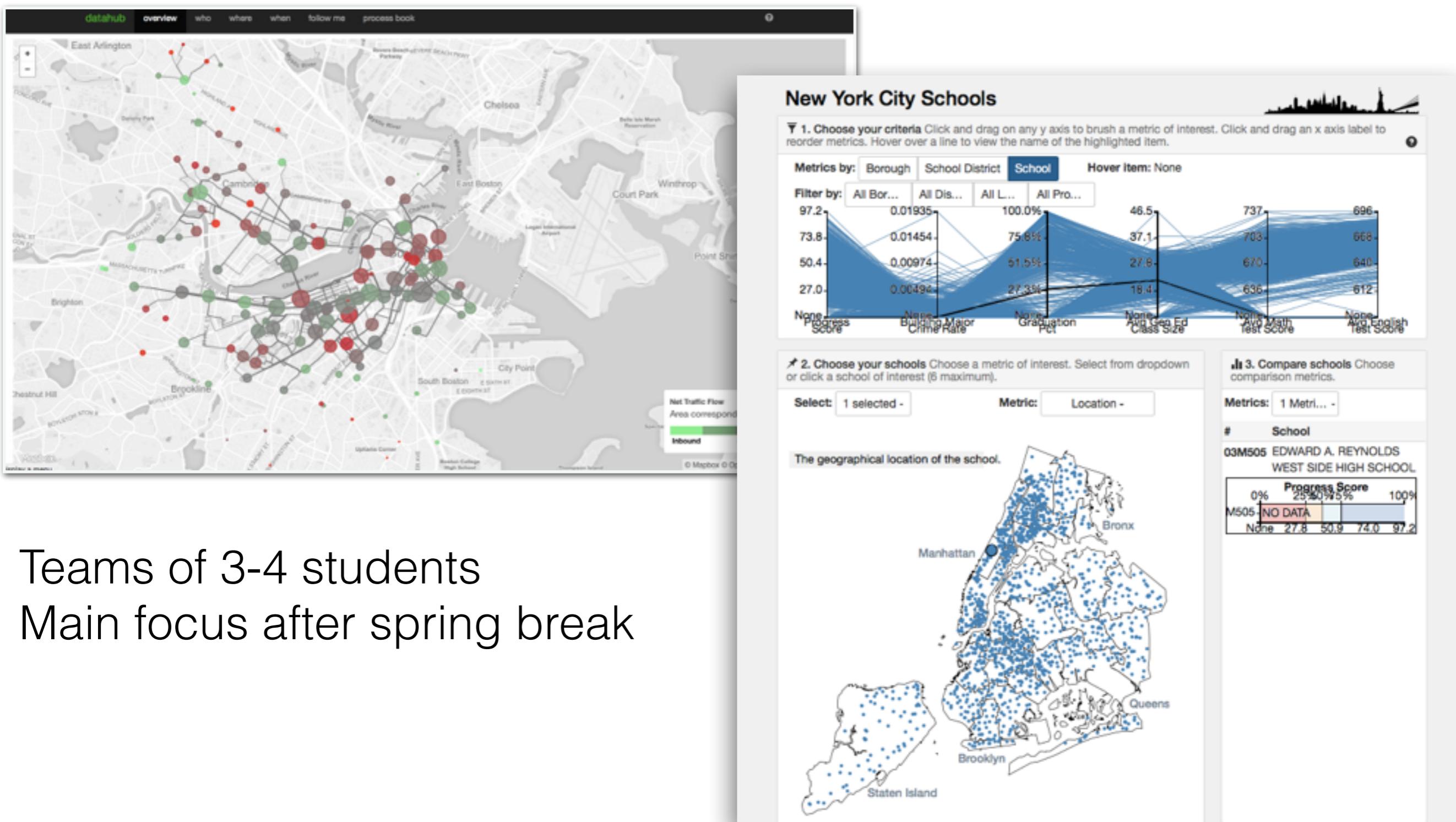
Midterm



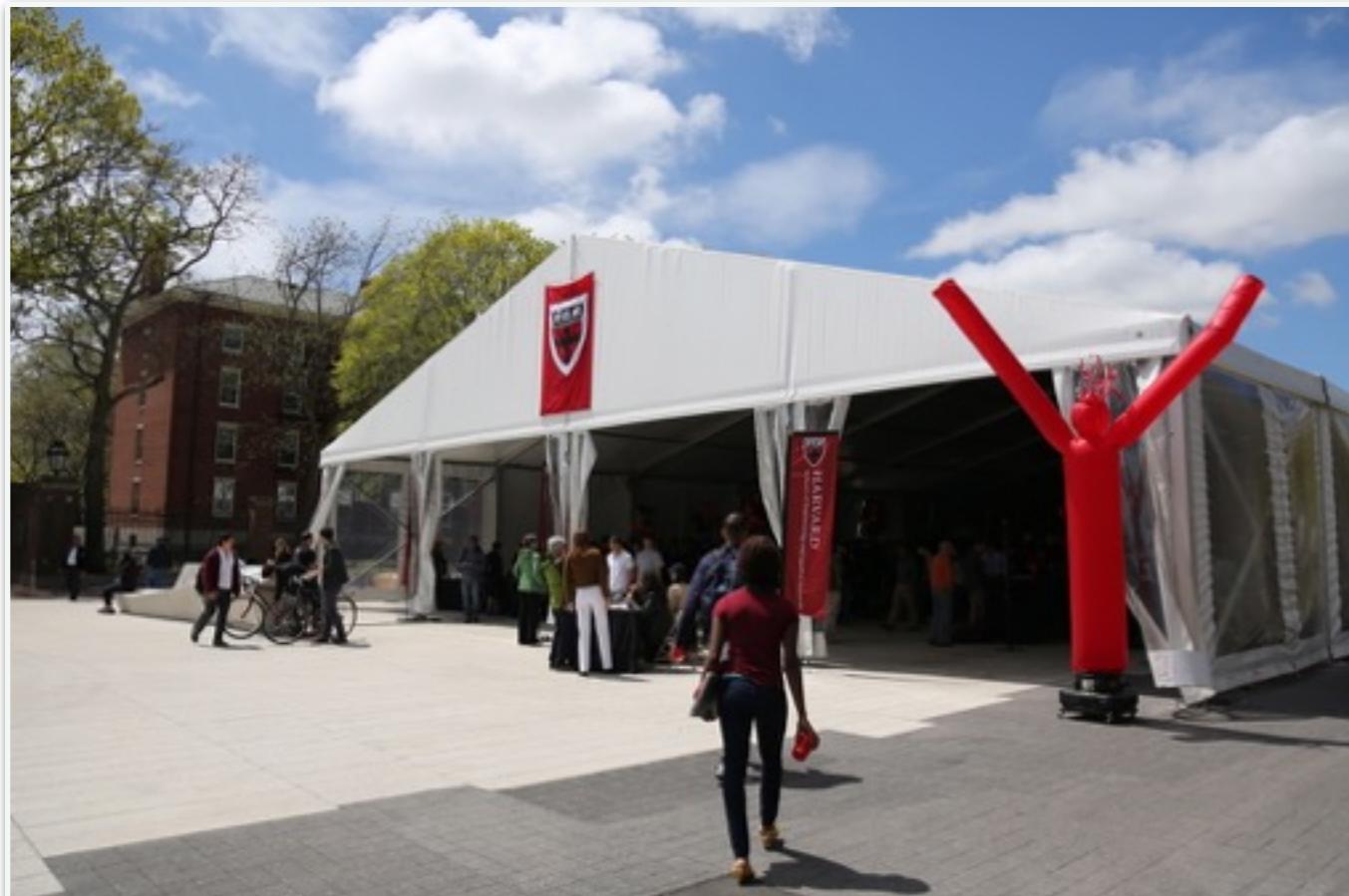
Part I: 90 minute design exercise
Part II: 5-10 hours D3 implementation



Group Project



SEAS Design Fair



Wednesday, May 4
“Best Of Show” Prizes



Grades

- Homework Assignments (25%)
- Midterm (20%)
- Group Project (40%)
- Quizzes (5%)
- Studios (5%)
- Participation (5%) (lectures, labs, Piazza)

Is this course for me ???



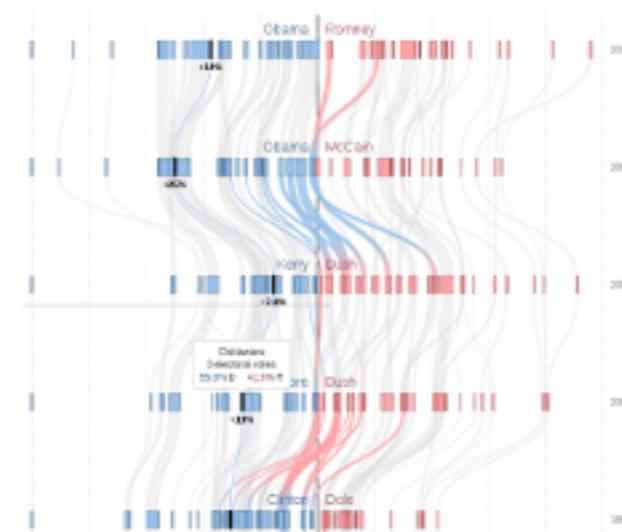
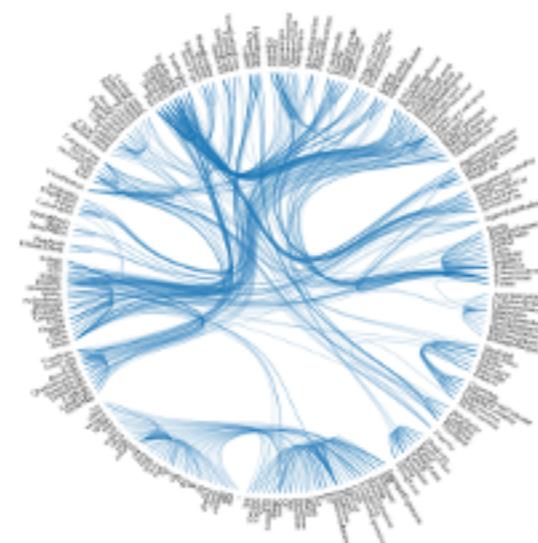
Prerequisites

Programming experience

- At least at the level of CS50
- C, C++, Java, Python, etc.

Willingness to learn new software & tools

- This can be time consuming
- You will need to build skills by yourself

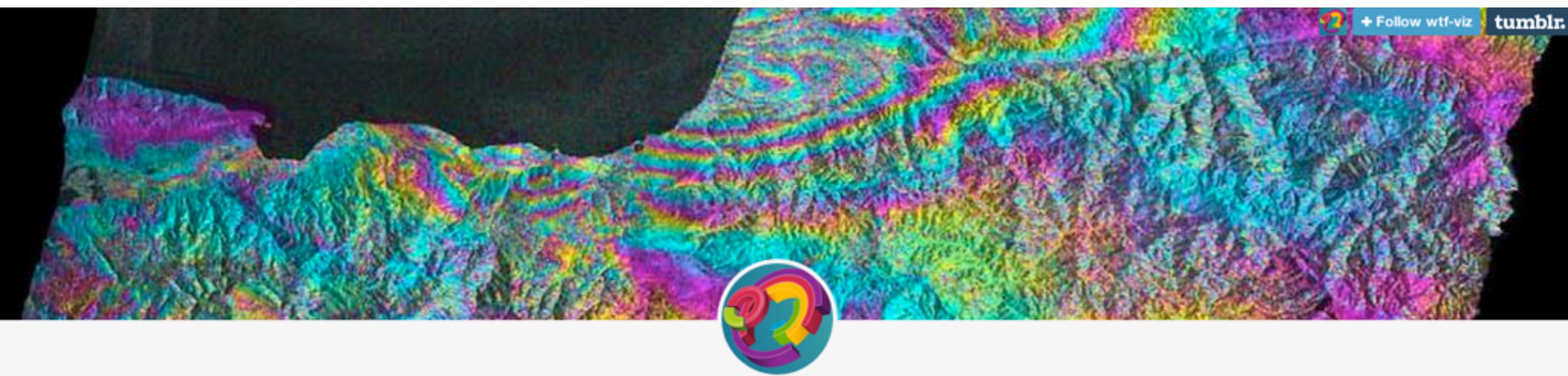


[Hierarchical edge bundling](#) | [Wind map](#) | [How states have shifted](#)

The amount and complexity of information produced in science, engineering, business, and everyday human activity is increasing at staggering rates. The goal of this course is to expose you to visual representation methods and techniques that increase the understanding of complex data. Good visualizations not only present a visual interpretation of data, but do so by improving comprehension, communication, and decision making.

In this course you will learn how the human visual system processes and perceives images, good design practices for visualization, methods for visualization of data from a variety of fields, and programming of interactive web-based visualizations using D3.

<http://wtfviz.net/>

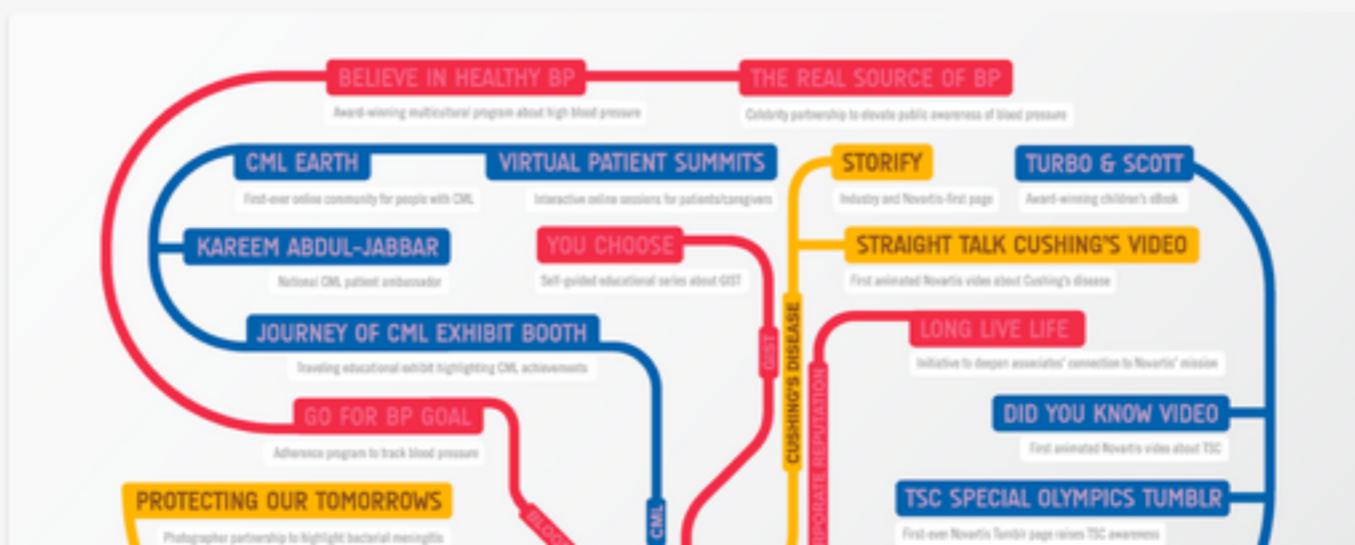


WTF Visualizations

Visualizations that make no sense.

For a discussion of what is wrong with a particular visualization, tweet at us [@WTFViz](#).

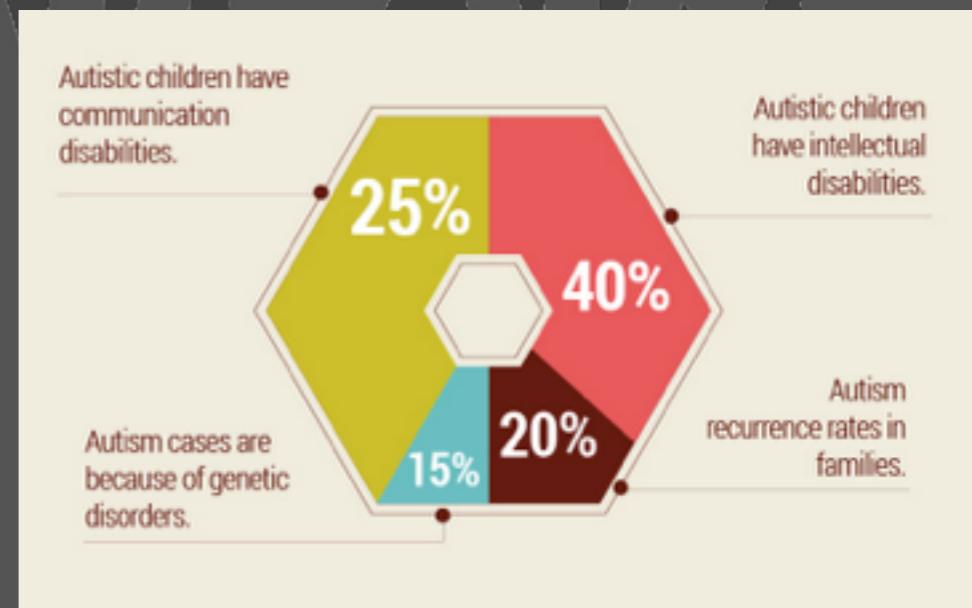
Check out our friends [Thumbs Up Viz](#) and [accidental aRt](#), or [submit](#).



Activity

Critique the following visualization answering the questions:

- What questions does this visualization answer?
- Why do you like / dislike this visualization?

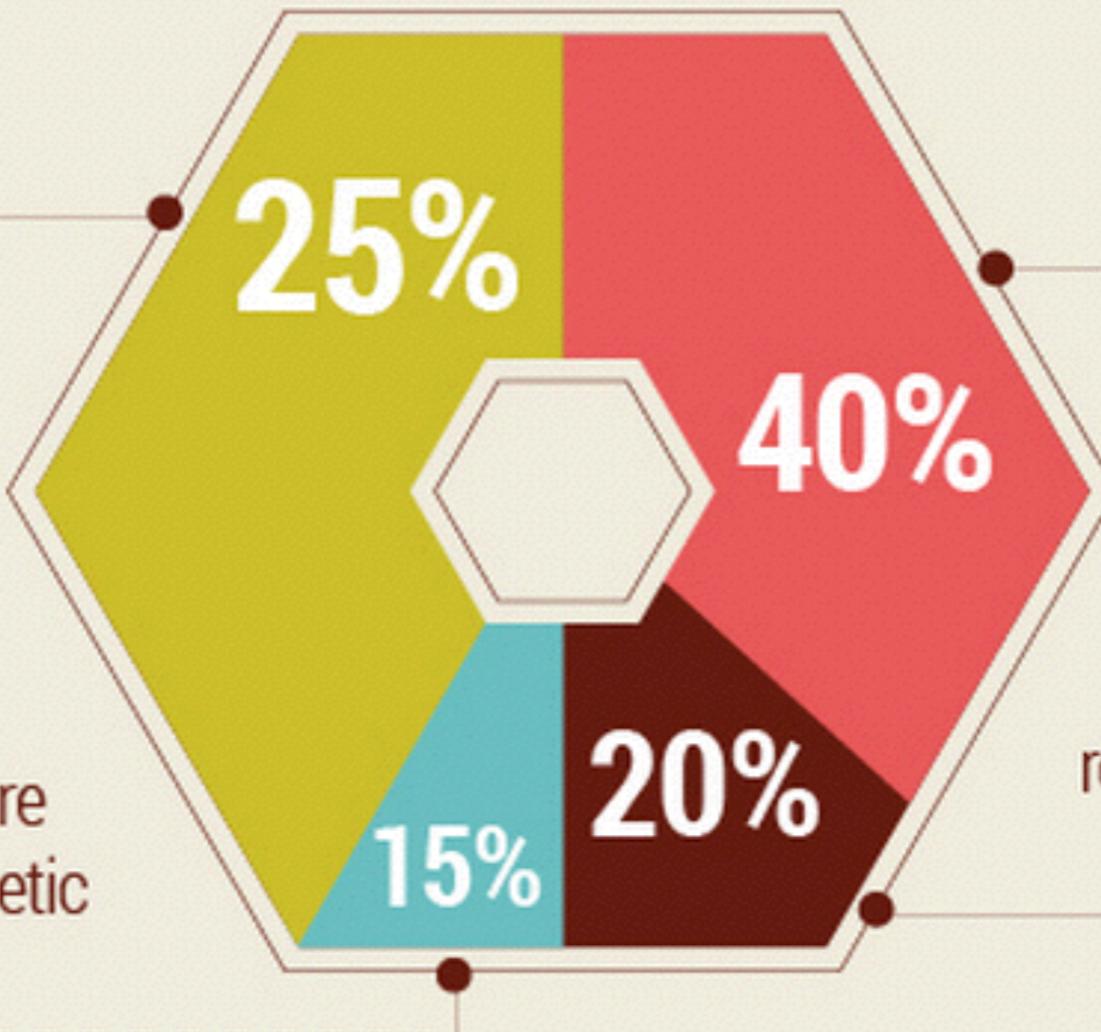


Autistic children have communication disabilities.

Autistic children have intellectual disabilities.

Autism cases are because of genetic disorders.

Autism recurrence rates in families.



What ?

“Design is a word that's come to mean so much that it's also a word that has come to mean nothing.”

–Johathan Ive



This Thursday...

- HTML / CSS / DOM
- Reading: Murray, Chapter 1-3



Next Tuesday...

- Design & Design Principles
- Reading: Ware, Chapter 8



Homework (due Monday)...

- HW 0 due **FRIDAY** including course survey
- HW 1 due Monday