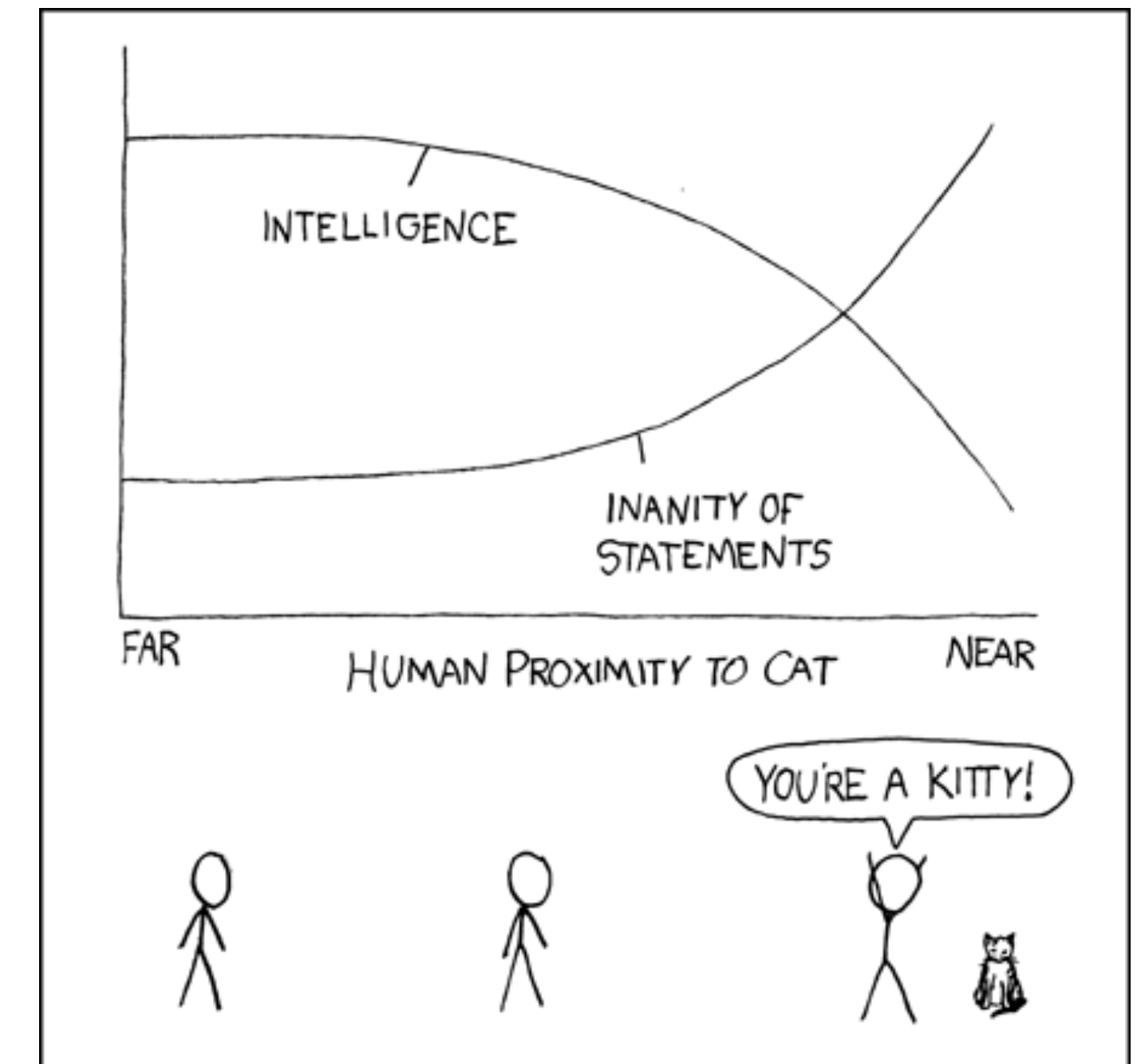


CS-5630 / CS-6630

Visualization for Data Science

Alexander Lex
alex@sci.utah.edu



visualization

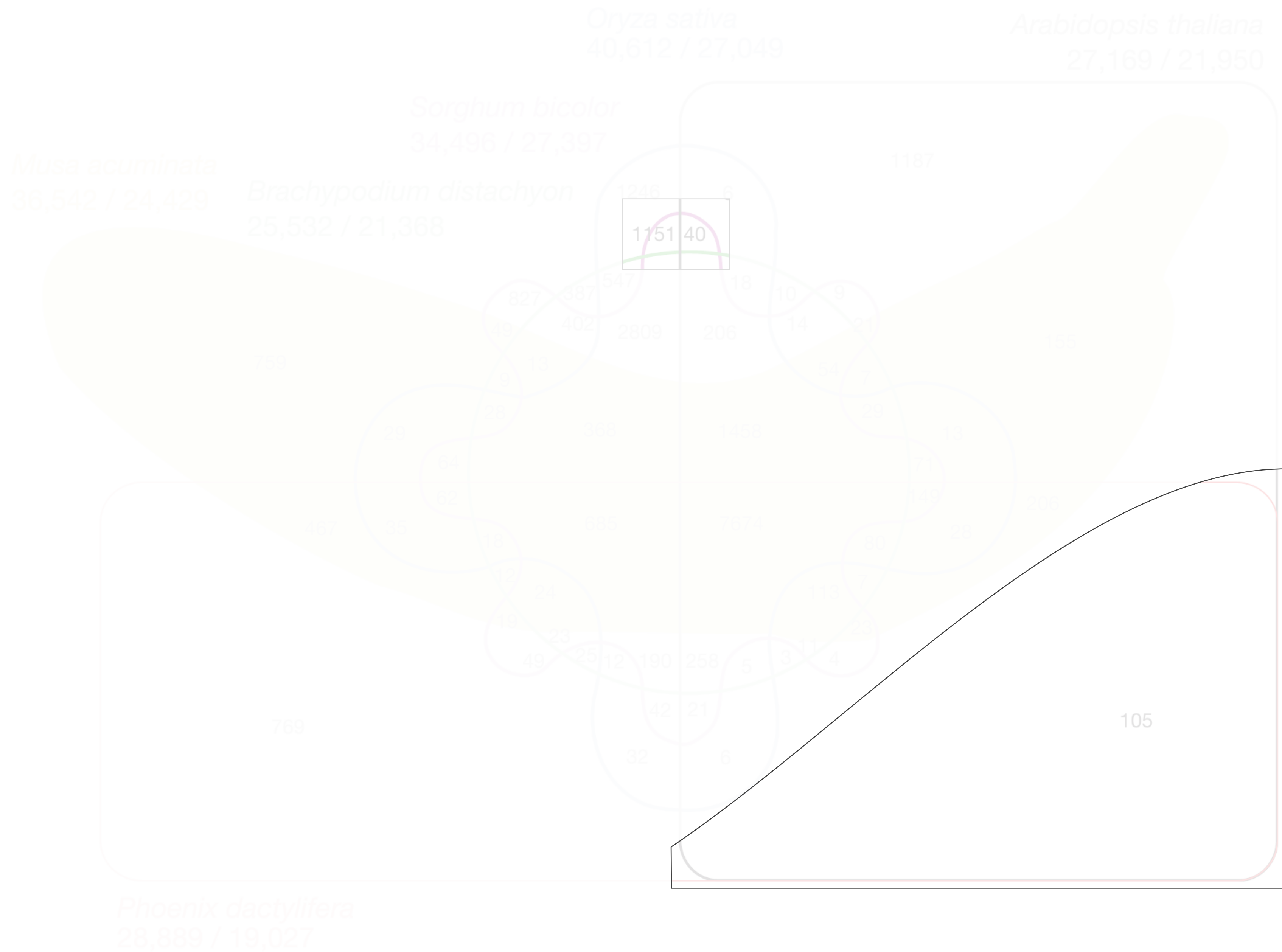
pictures

The purpose of computing is insight, not numbers.

- Richard Wesley Hamming

- Card, Mackinlay, Shneiderman

Banana	<i>M. acuminata</i>
Date	<i>P. dactylifera</i>
Cress	<i>Arabidopsis thaliana</i>
Rice	<i>Oryza sativa</i>
<i>Sorghum</i>	<i>Sorghum bicolor</i>
<i>Brome</i>	<i>Brachypodium distachyon</i>



[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

Visualization Definition

Visualization is the process that **transforms** (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 humans: Communication

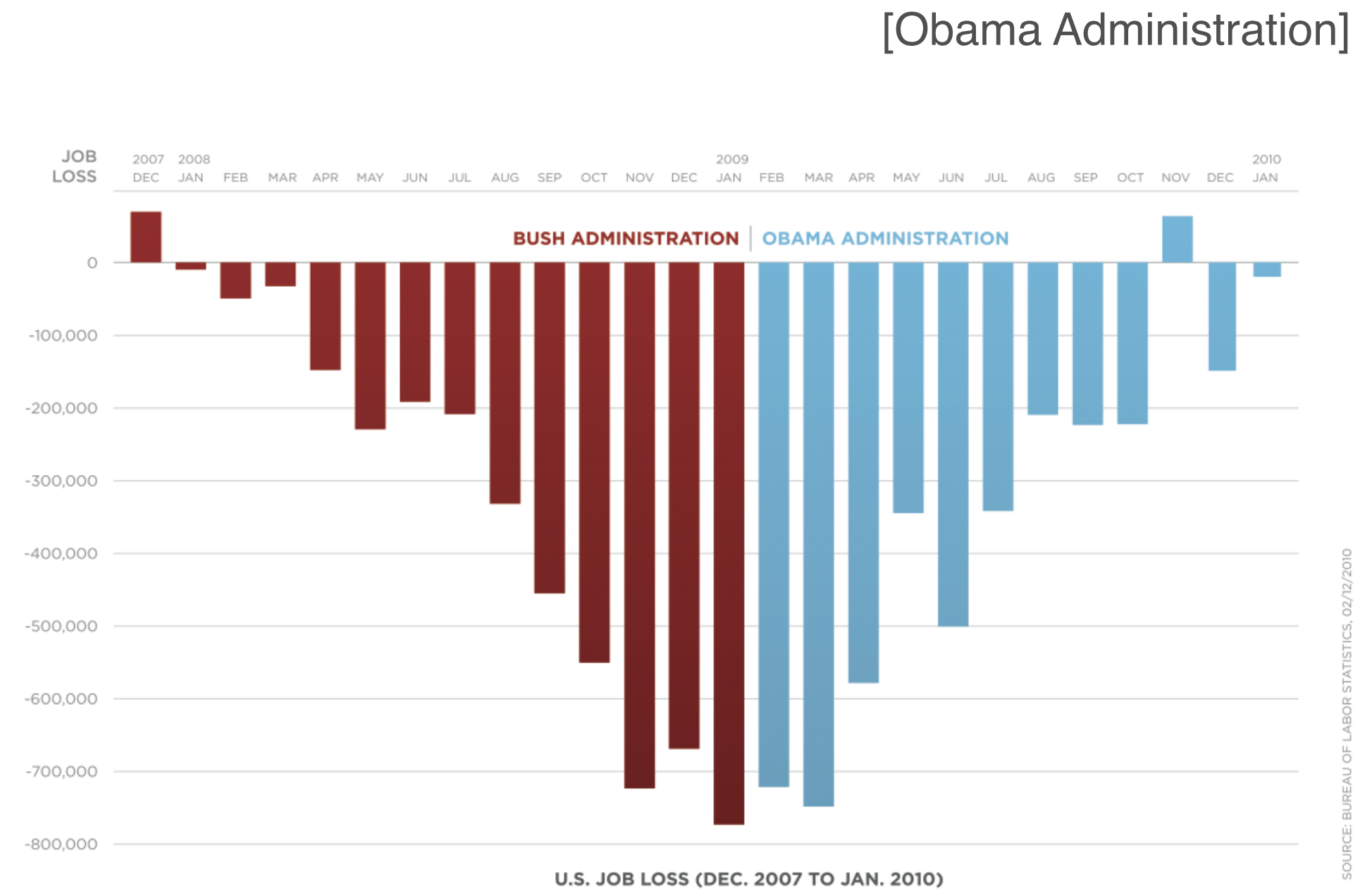
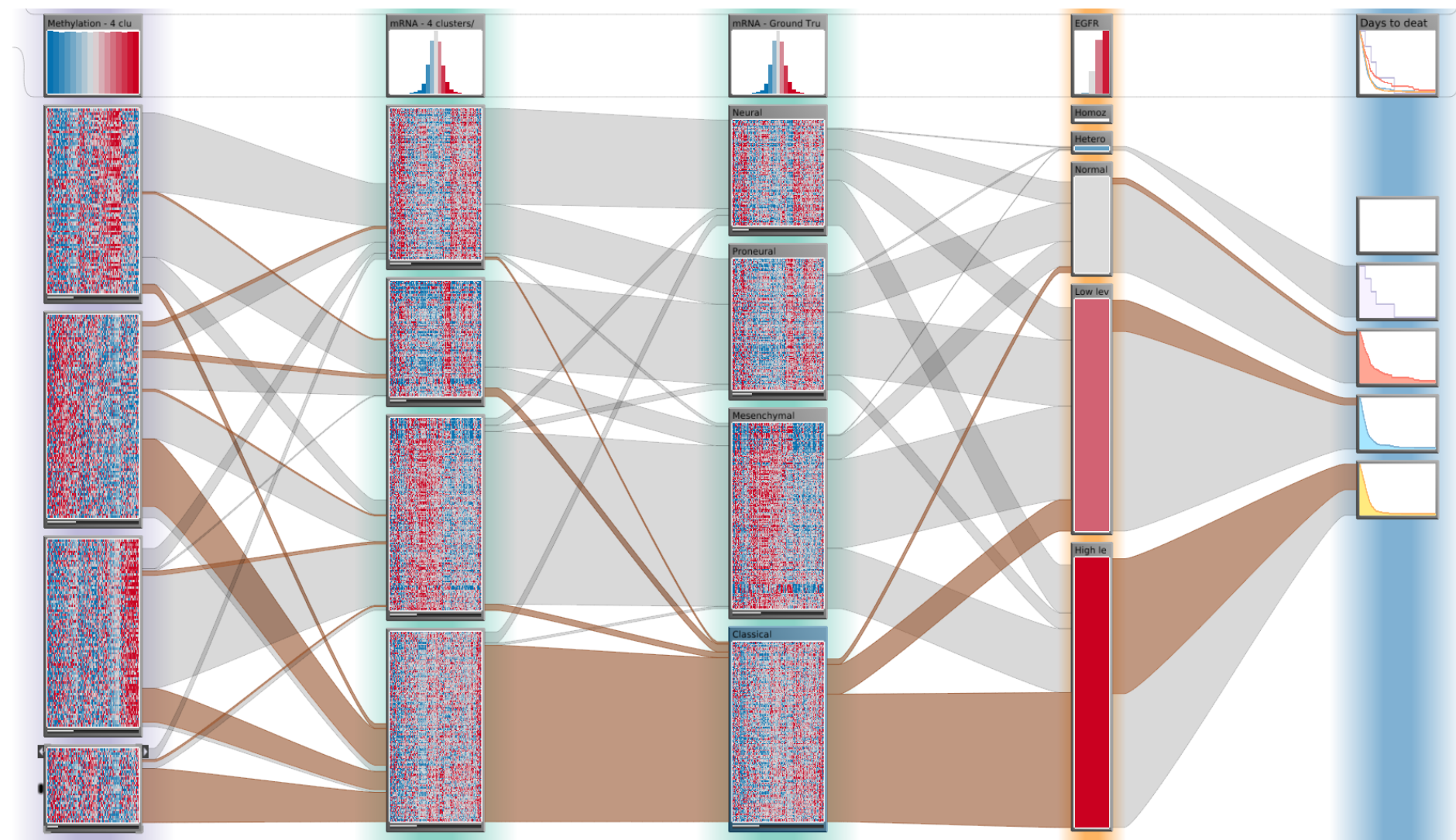
How is ahead in the election polls?

When questions are not well
defined: Exploration

What is the structure of a terrorist network?

Which drug can help patient X?

Purpose of Visualization



Open Exploration

Confirmation

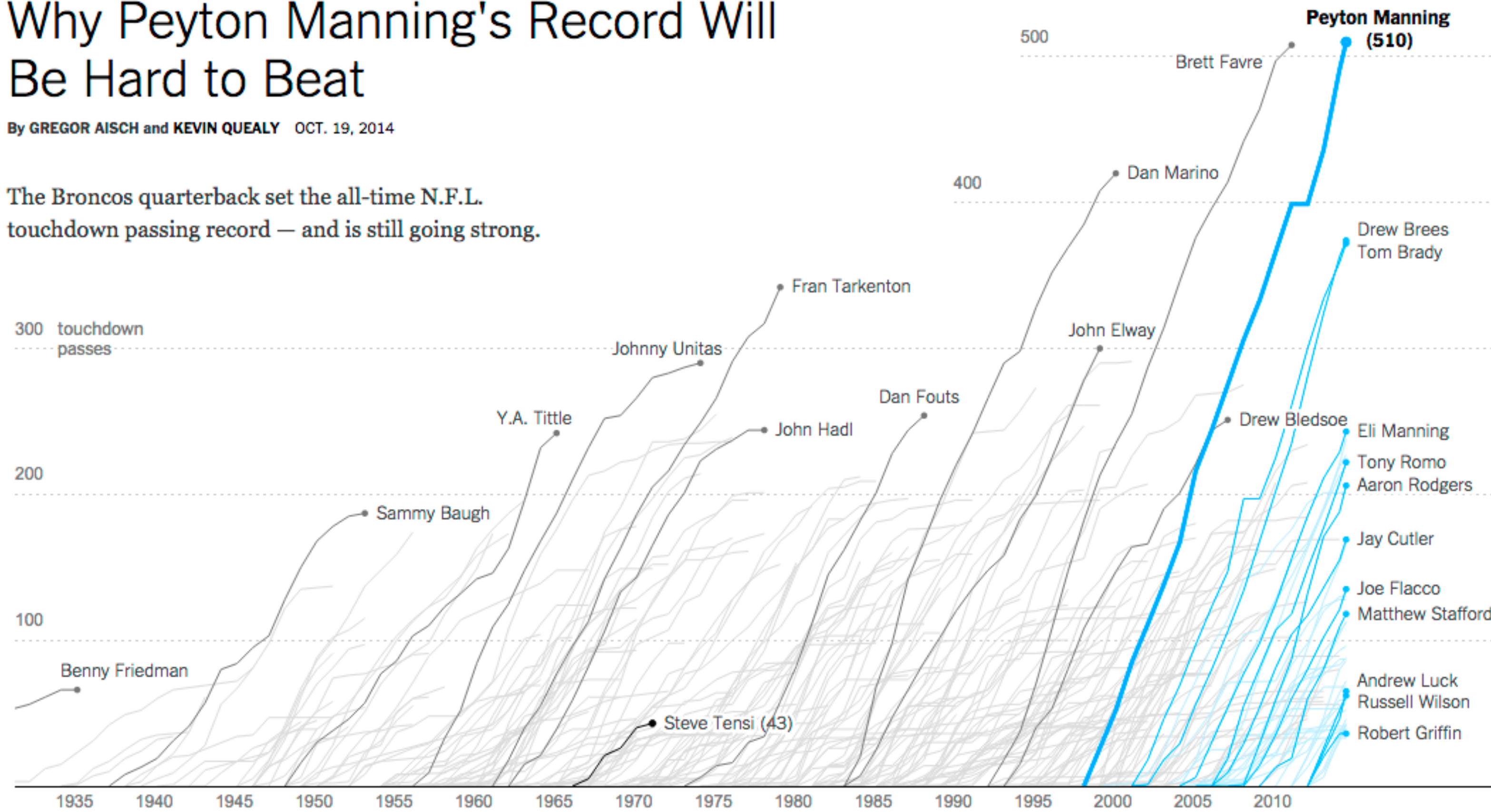
Communication

Example Communication

Why Peyton Manning's Record Will Be Hard to Beat

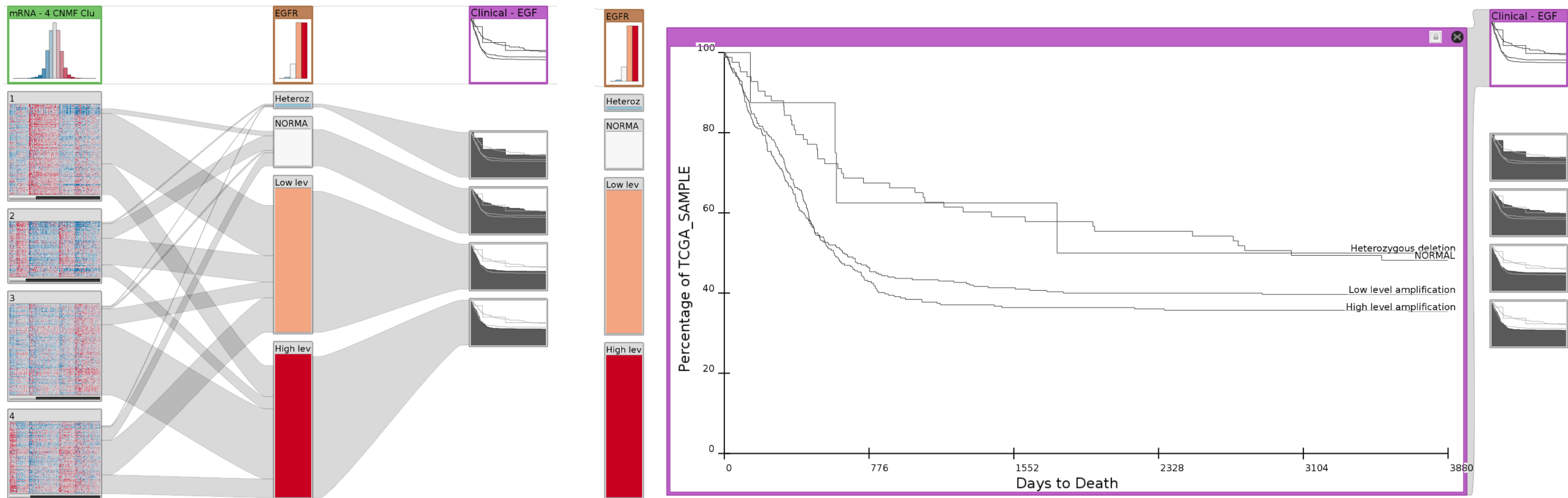
By GREGOR AISCH and KEVIN QUEALY OCT. 19, 2014

The Broncos quarterback set the all-time N.F.L. touchdown passing record — and is still going strong.



[New York Times]

Example Exploration: Cancer Subtypes



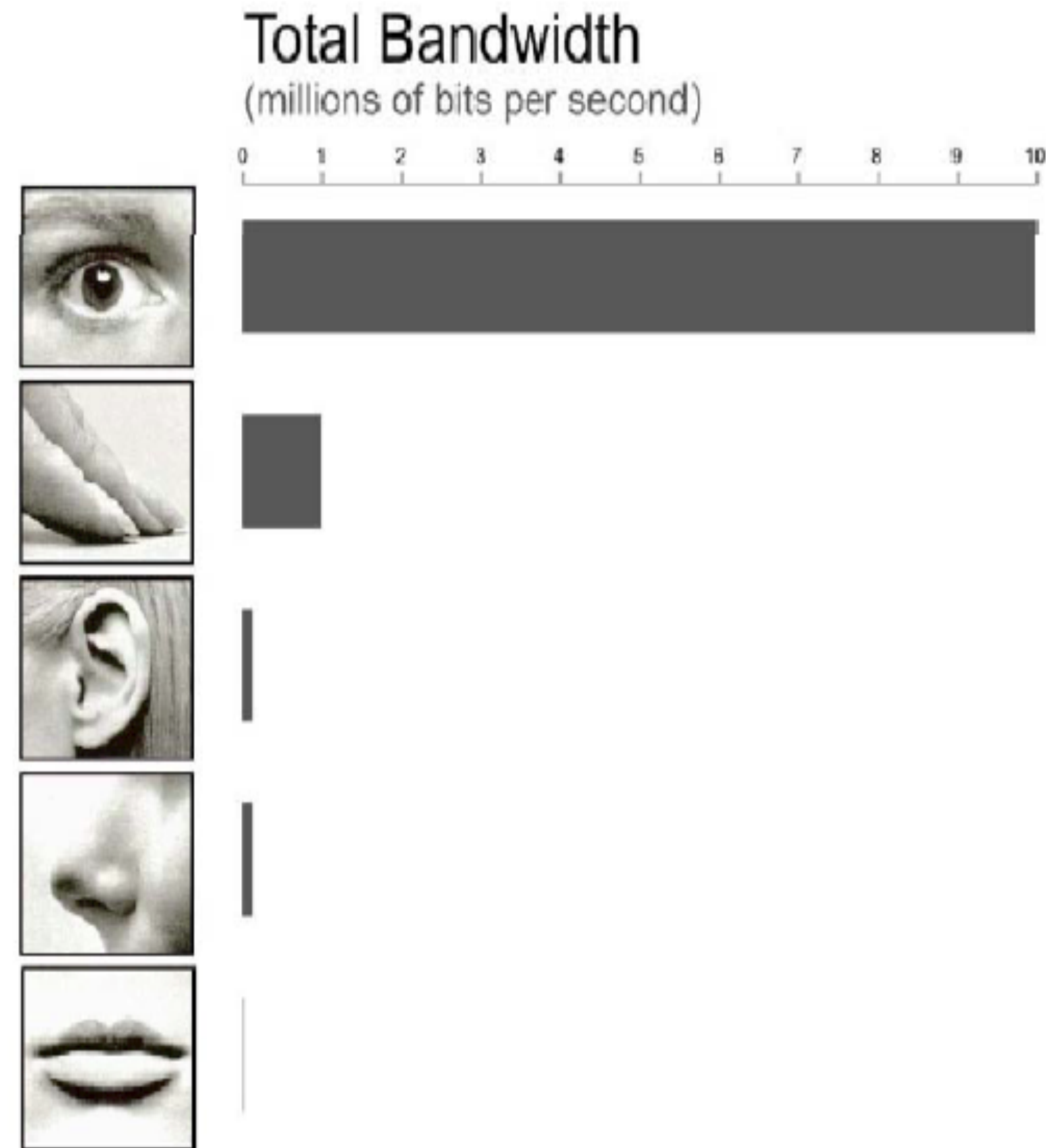
Why Graphics?

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]



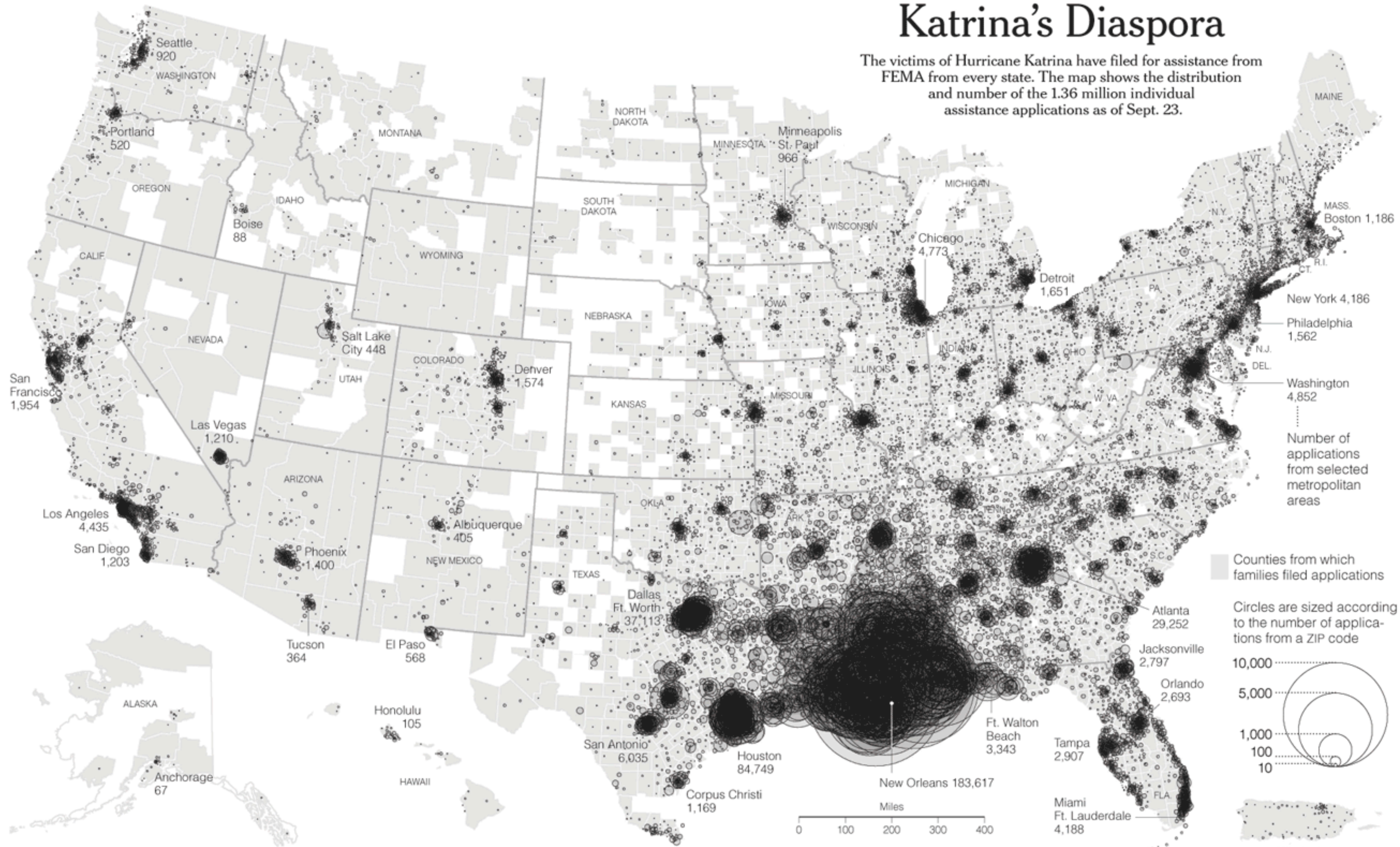
for city's main public hospital was a wreck, and the city's public-housing projects were shuttered.

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 eighty-four thousand, and it's about

Textual description of a map of the effects of hurricane Katrina on New Orleans. New Yorker, posted by Alberto Cairo

Katrina's Diaspora

The victims of Hurricane Katrina have filed for assistance from FEMA from every state. The map shows the distribution and number of the 1.36 million individual assistance applications as of Sept. 23.



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?

No human intervention possible/necessary

Decisions needed in minimal time

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

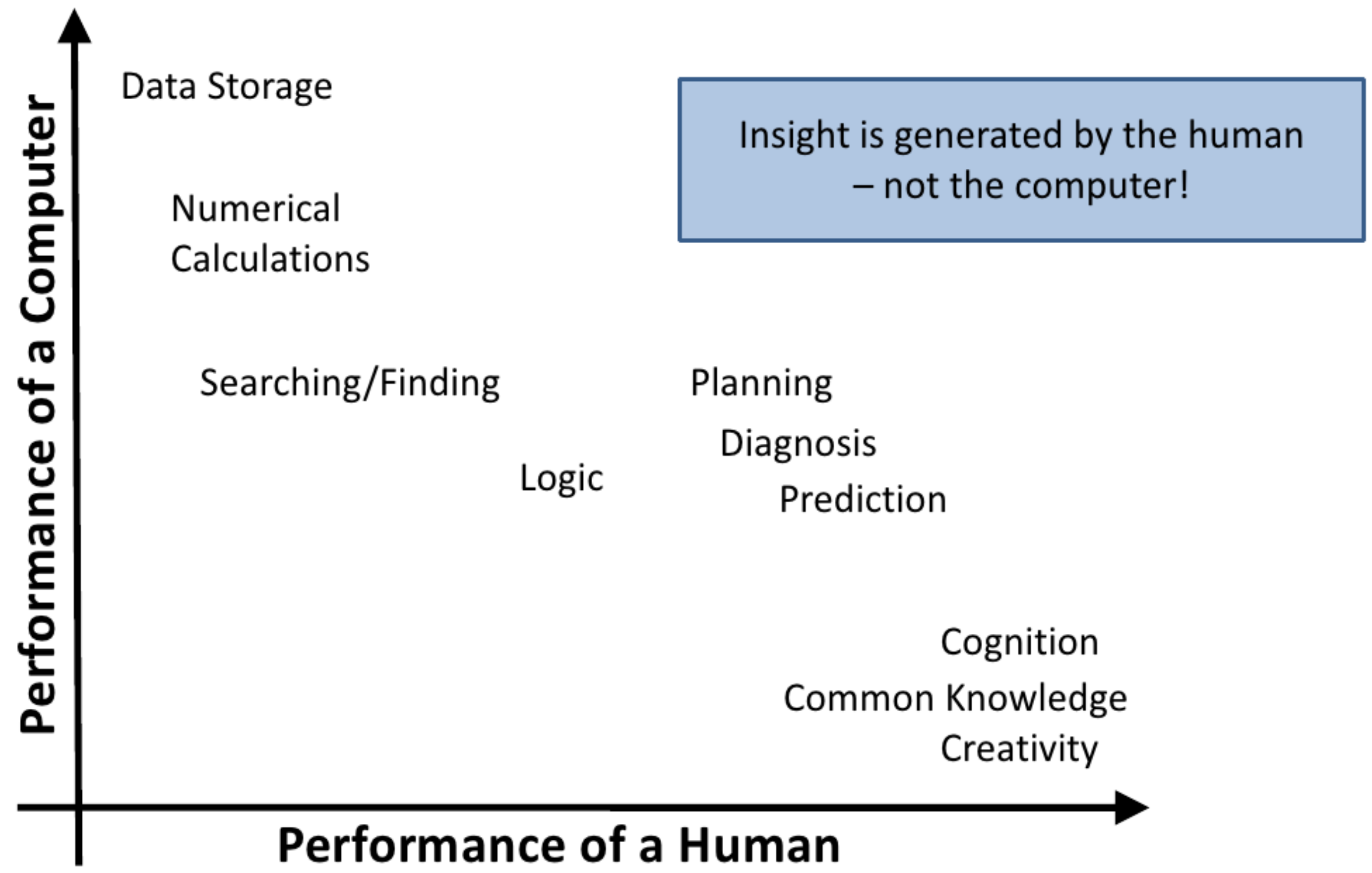
Manufacturing: is bottle broken?

Impractical for human to be involved

Automatic data products



The Ability Matrix



Why Use Computers?

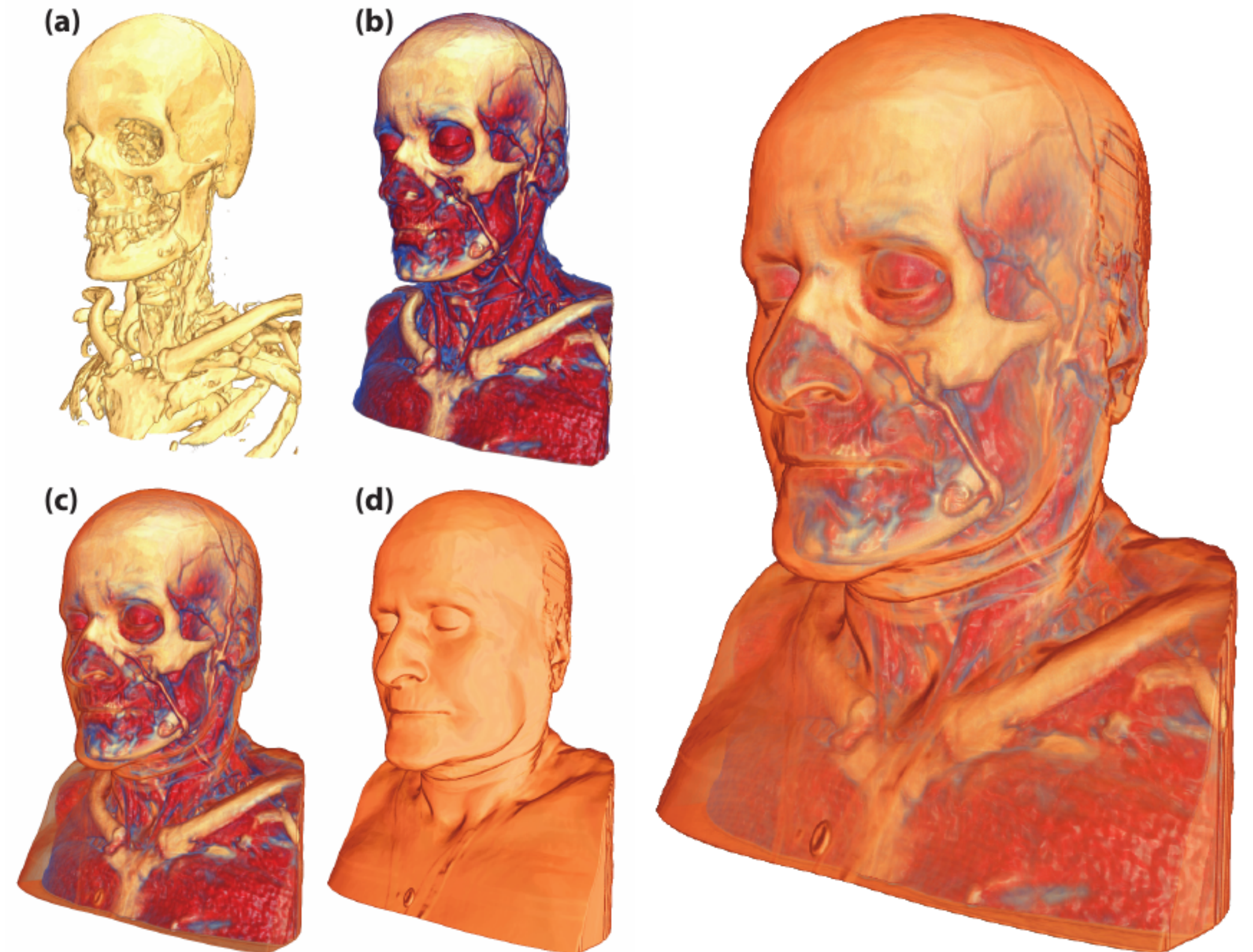
Scale

Drawing by hand (or Illustrator)

infeasible

inflexible (updates!)

How to draw an MRI scan?



[Bruckner 2007]

Why Use Computers?

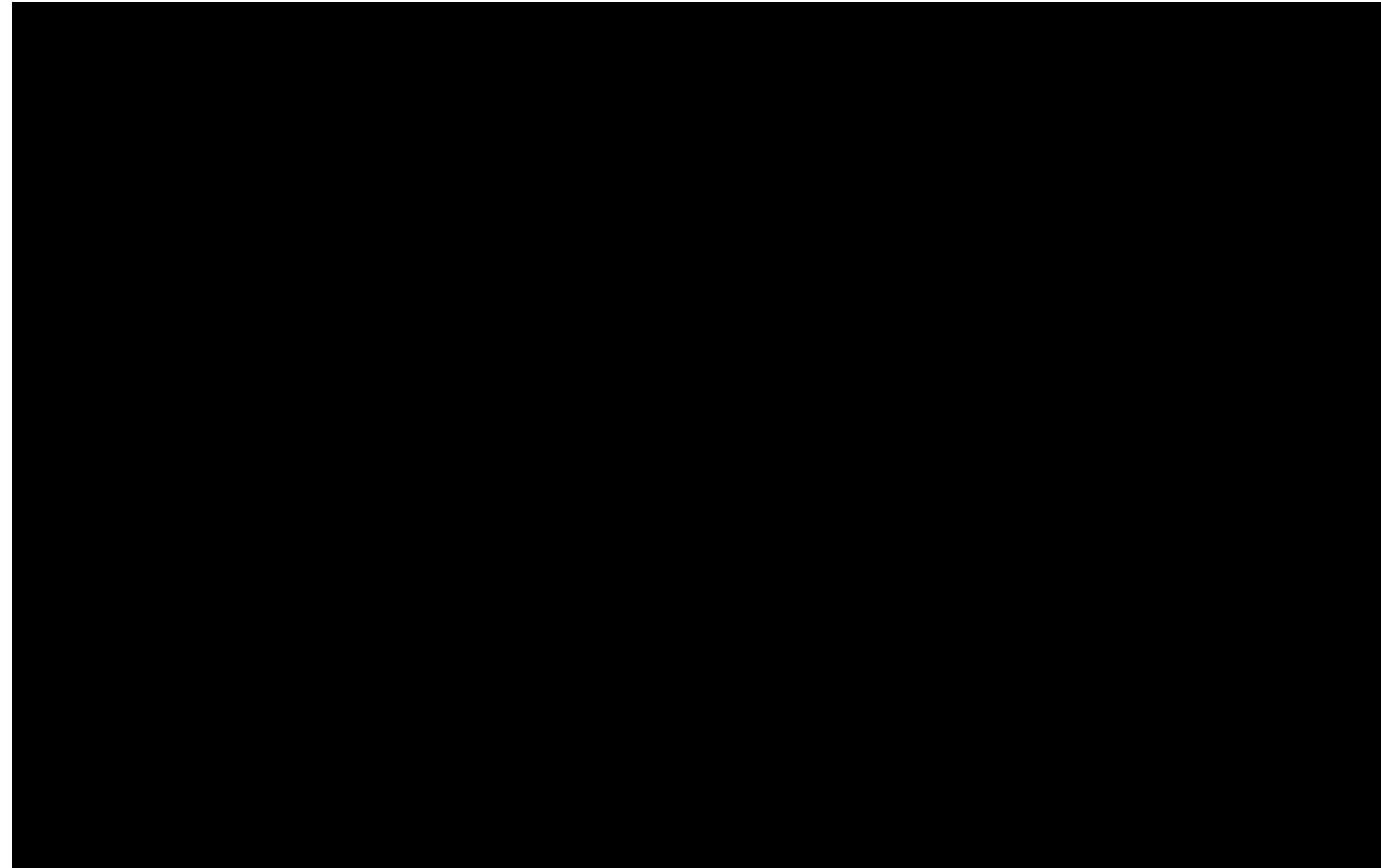
Interaction

Interaction allows to “drill down” into data

Integration

Integration with algorithms

Make visualization part of a data analysis pipeline



Why User 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?

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.0	10	9.1	10	7.4	8	6.5
8	6.9	8	8.1	8	6.7	8	5.7
13	7.5	13	8.7	13	12.	8	7.7
9	8.8	9	8.7	9	7.1	8	8.8
11	8.3	11	9.2	11	7.8	8	8.4
14	9.9	14	8.1	14	8.8	8	7.0
6	7.2	6	6.1	6	6.0	8	5.2
4	4.2	4	3.1	4	5.3	19	12.
12	10.	12	9.1	12	8.1	8	5.5
7	4.8	7	7.2	7	6.4	8	7.9
5	5.						6.8

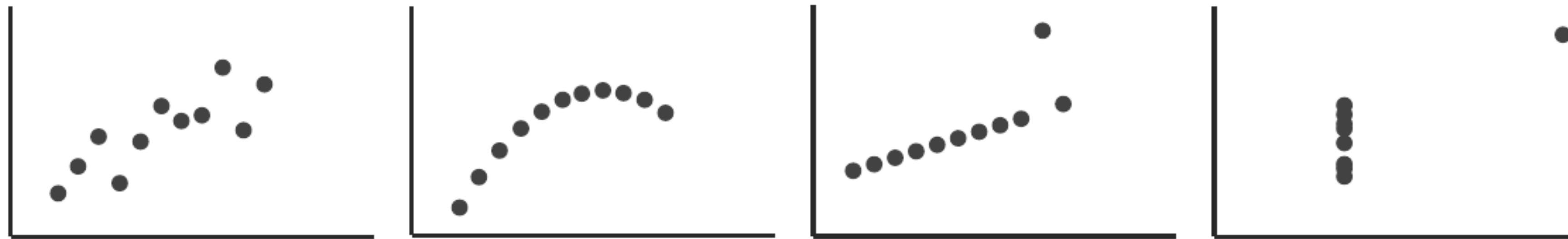
Mean x: 9 y: 7.50

Variance x: 11 y: 4.122

Correlation x - y: 0.816

Linear regression: $y = 3.00 + 0.500x$

Anscombe's Quartett

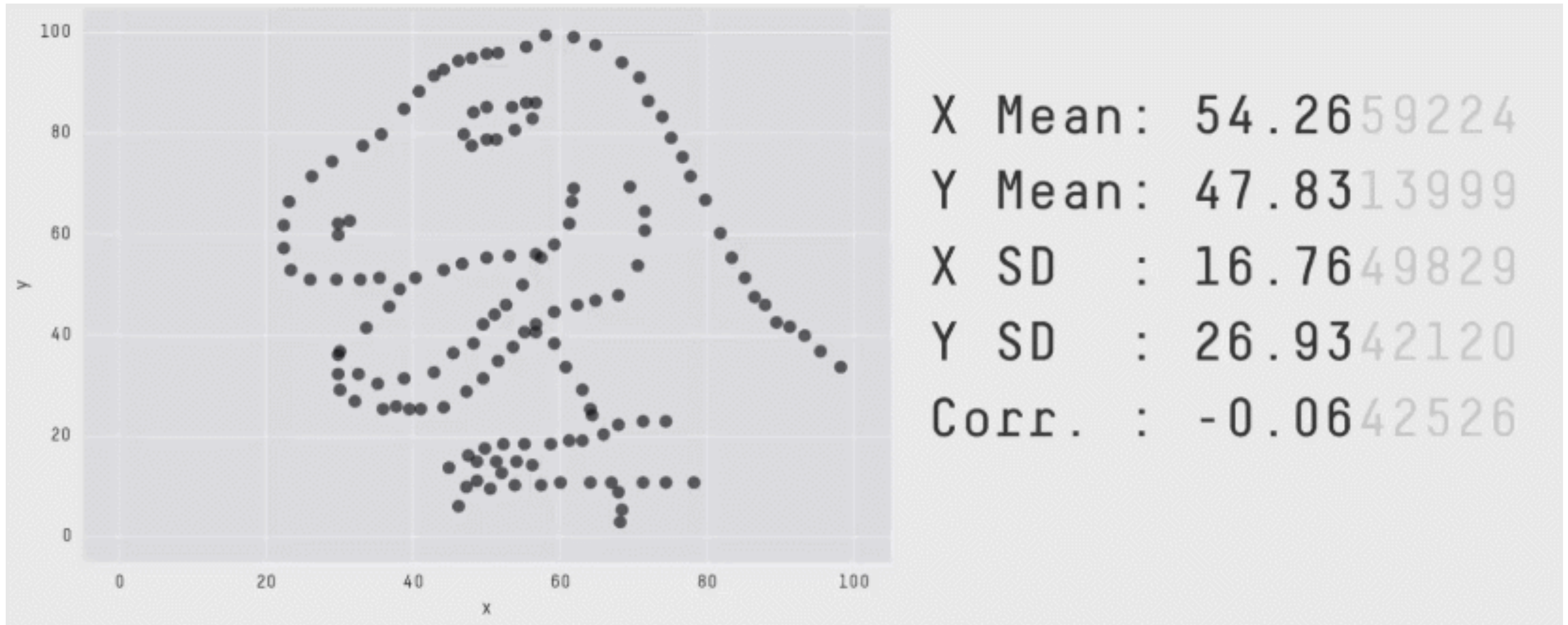


Mean x: 9 y: 7.50

Variance x: 11 y: 4.122

Correlation x - y: 0.816

Linear regression: $y = 3.00 + 0.500x$



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

Visualization =

Human Data Interaction

Data

Human-Data Interaction

Visualization in the Data Science Process

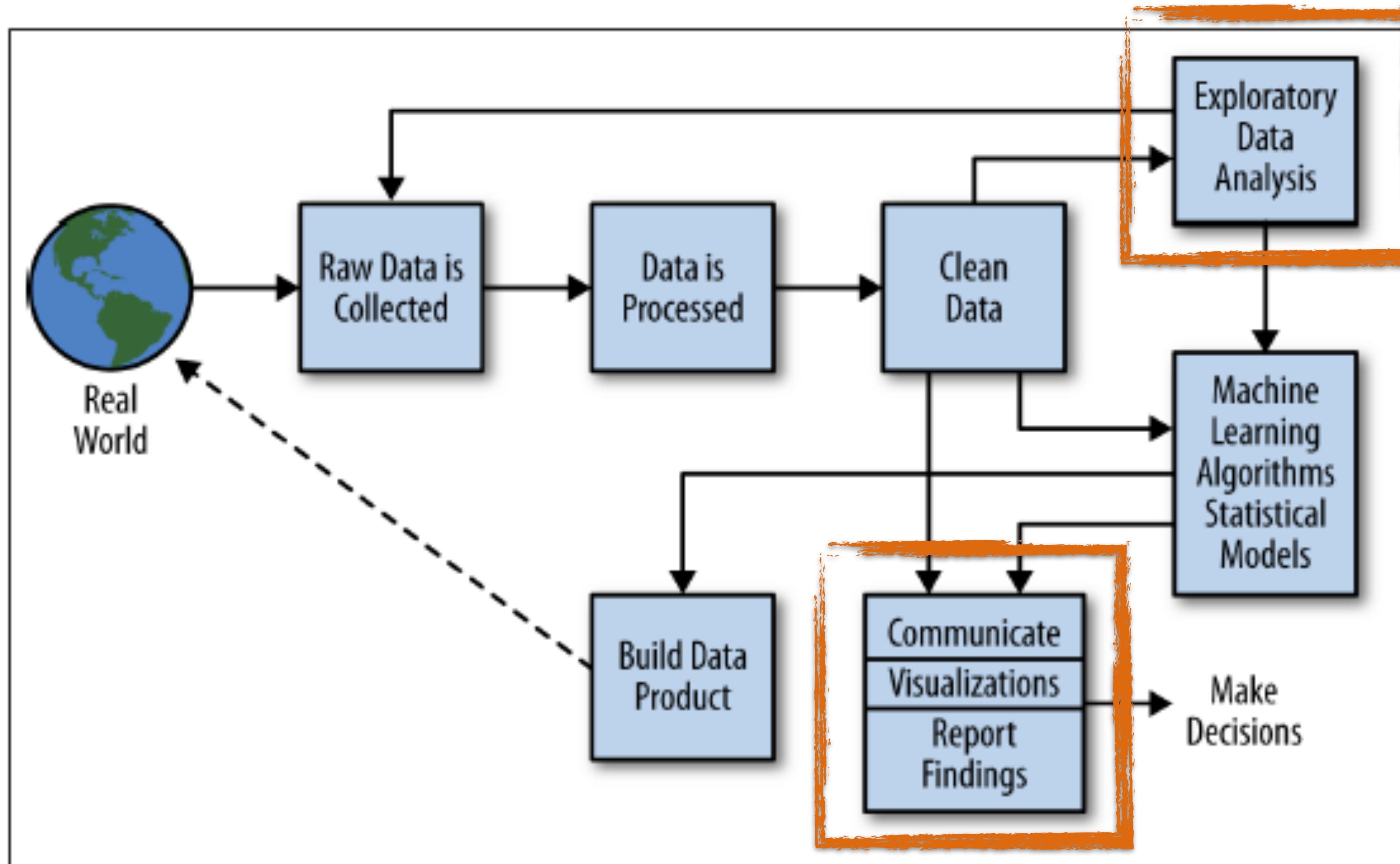


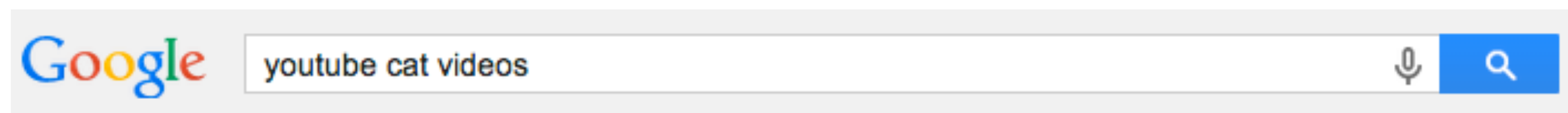
Figure 2-2. The data science process

Big Data

2017: 2.5 exabytes (quintillion bytes) of data per day, largely unstructured
90% of the data created in last two years

Source: IBM

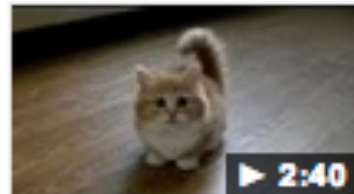
15 Exabytes in Punch Cards:
4.5 km over New England



Web Videos Shopping Images News More Search tools

About 593,000,000 results (0.44 seconds)

[TOP 10 BEST CAT VIDEOS OF ALL TIME! - YouTube](#)



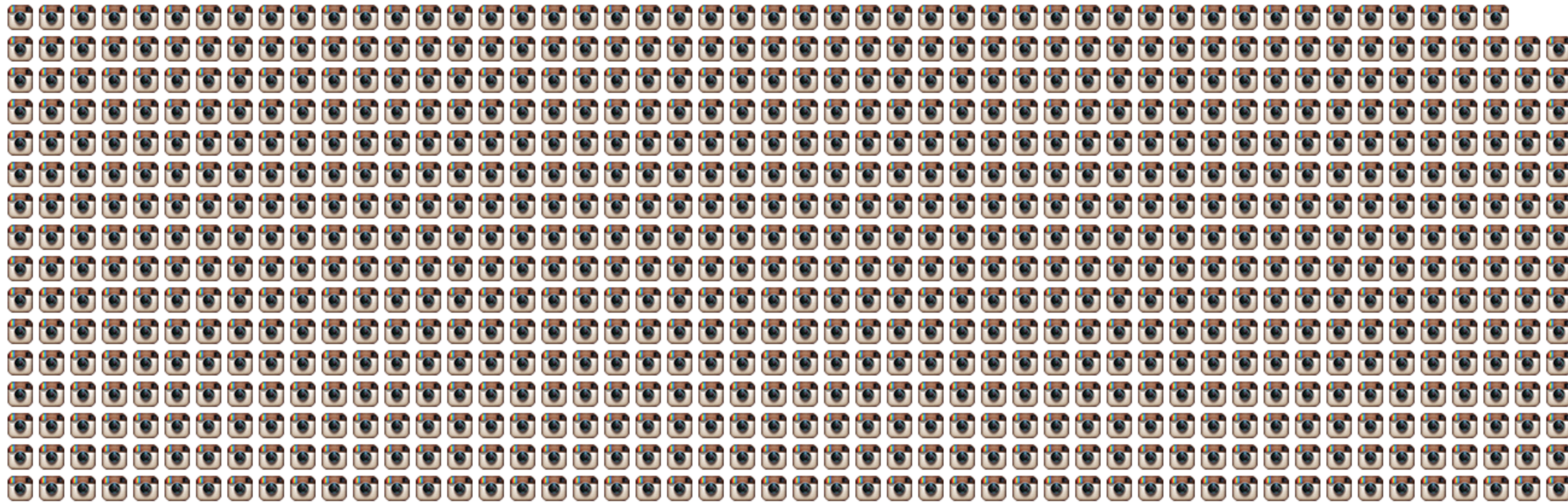
www.youtube.com/watch?v=... YouTube
Sep 6, 2012 - Uploaded by WatchTheDaily
We've scoured the internet and found the cutest and funniest cat videos of all time. Any we missed? Let us ...

[The World's Most Funny Cat Videos 2013 - YouTube](#)



www.youtube.com/watch?v=Ak... YouTube

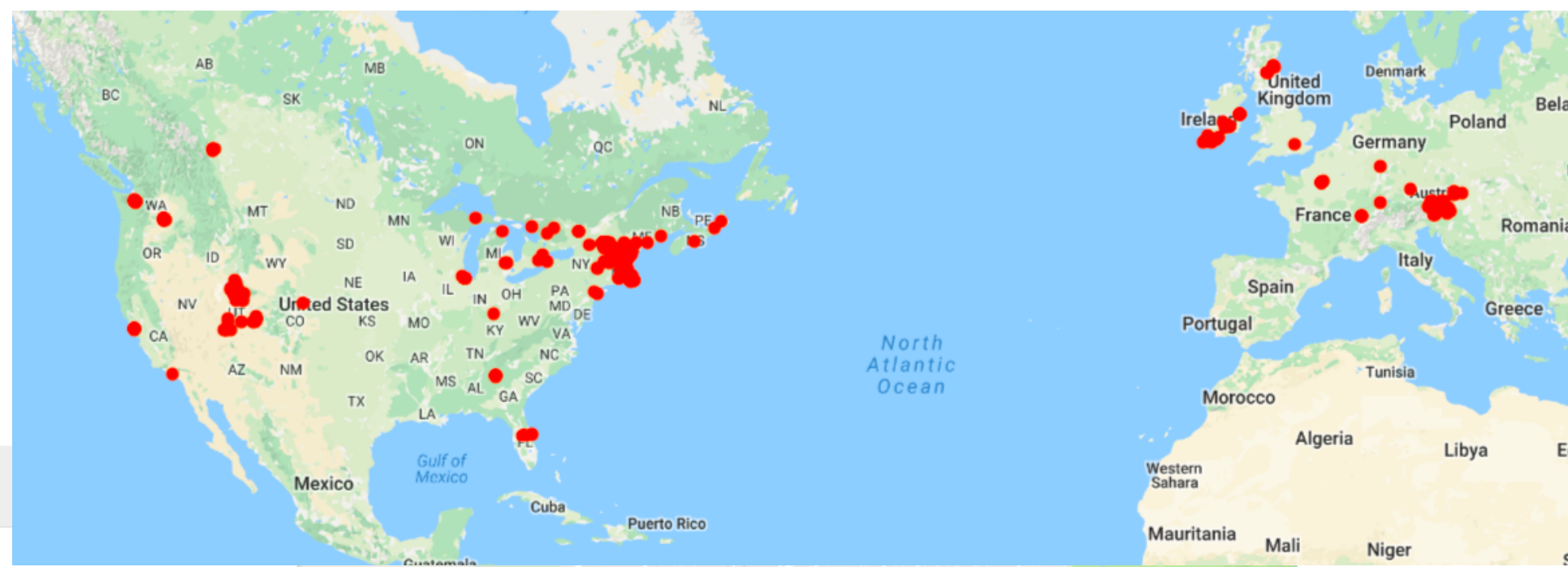
798 Instagram photos uploaded in 1 second



1,277 Tumblr posts in 1 second



Example: Personal Data



Timeline

2016 June 29

Wednesday, June 29, 2016

36.6 mi 1h 59m

11.4 mi 1h 48m

0.2 mi 11m

Home ? 8:51 AM

3125 Kennedy Dr, Salt Lake City, UT 84108

YES NO

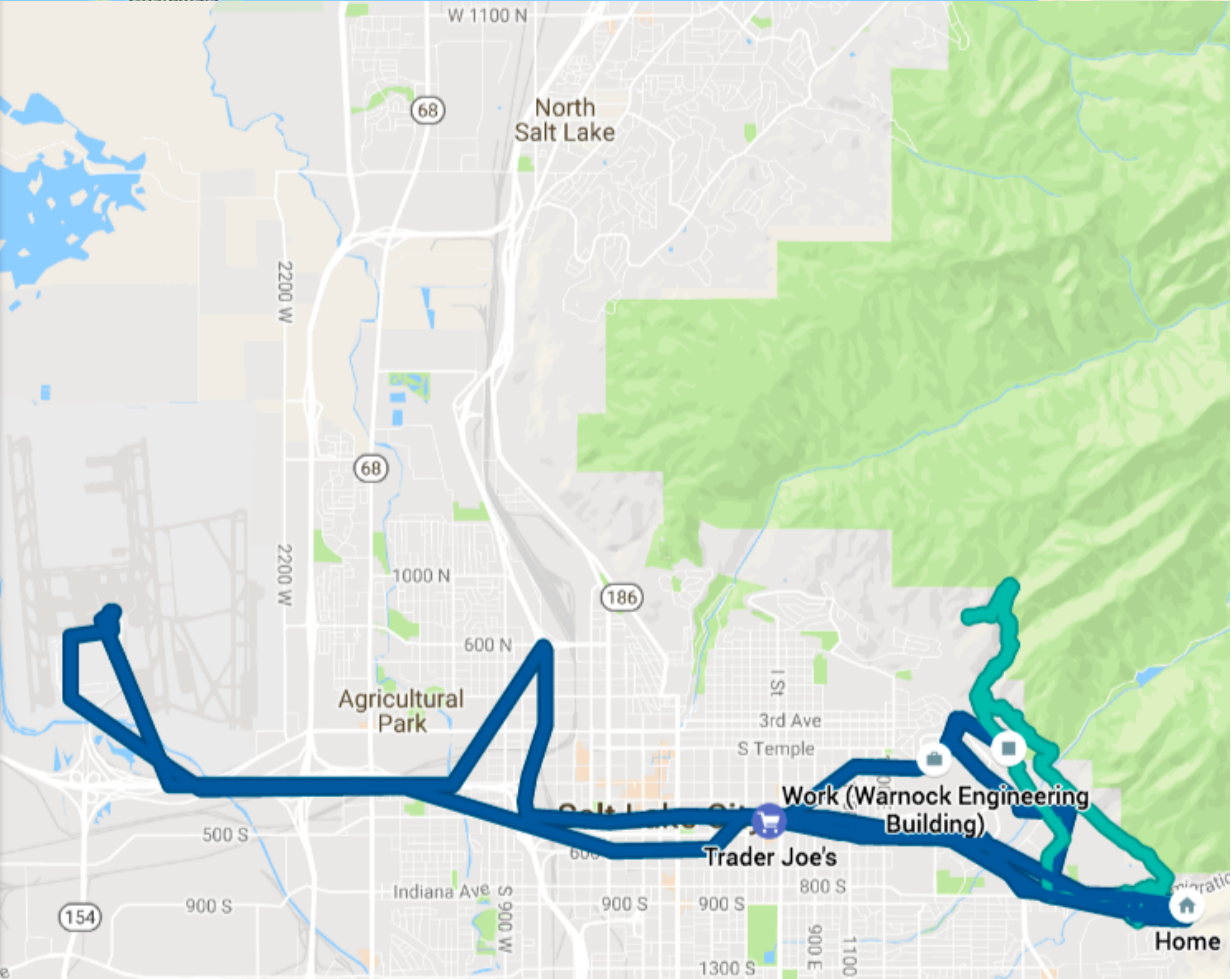
Driving - 3.4 mi 18 mins

Work (Warnock Engineering Bu...) 9:09 AM - 12:42 PM

72 Central Campus Dr, Salt Lake City, UT 84112

YES NO

20 S 2030 E ? 1:01 PM - 3:01 PM



Google My Activity

Bundle view

Item view

Delete activity by

Other Google activity

Activity controls

My Account

Help

Search

Filter by date & product

Only you can see this data. Google protects your privacy and security. [Learn more](#)

Today

Some activity may not appear yet

ITEMS 123

CHROME SEARCH ANDROID IMAGE SEARCH NOW

STRAVA Dashboard Training Explore Challenges

Go Premium

Overview

Analysis

Premium

Heart Rate

Est Power Curve

Est 25W Distribution

Alexander Lex - Ride

8:54 AM on Saturday, August 20, 2016

Wasatch Crest Trail

Add a description

40.7 km 2:34:29 442m

Distance Moving Time Elevation

148w 1,372kJ

Estimated Avg Power Energy Output

	Avg	Max
Speed	15.8km/h	74.2km/h
Elapsed Time	3:30:52	

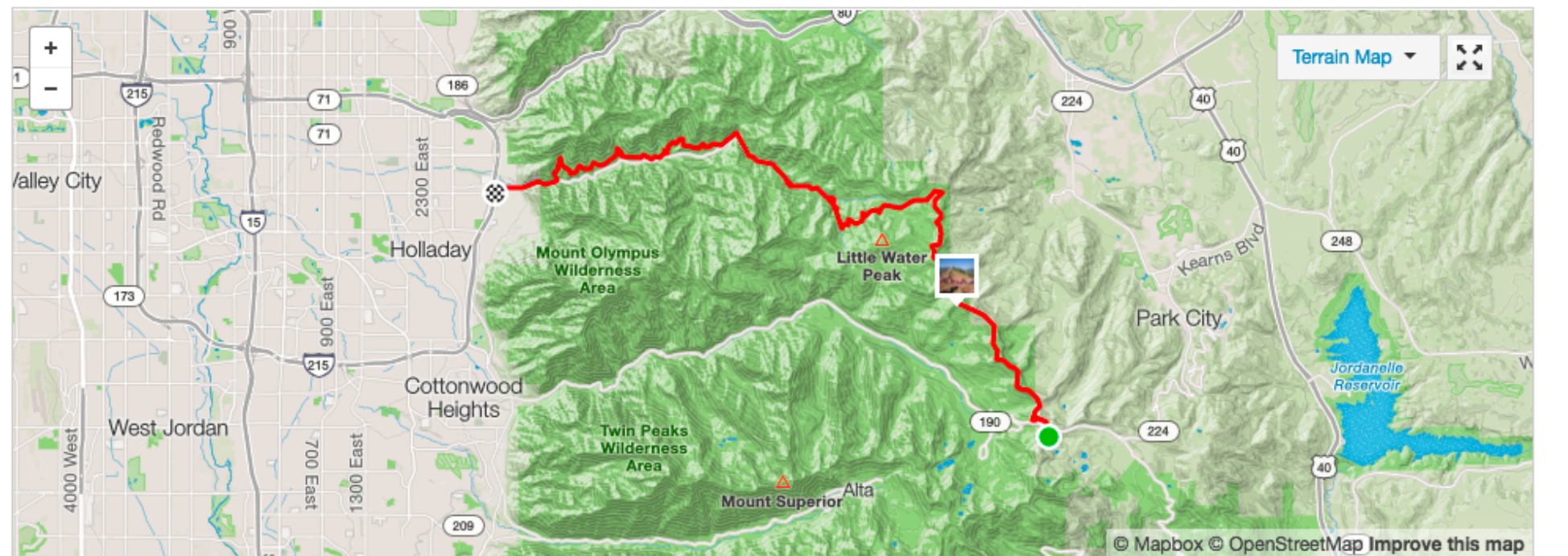
Show More

Device: [Strava Android App](#) Bike: --

TOP RESULTS

- PR on [rattlesnake ch](#) (6:39)
- PR on [Church Fork to Bottom of Rattlesnake](#) (18:45)
- PR on [Elbow to Birch](#) (16:13)

View all



Big Data in Science and Engineering

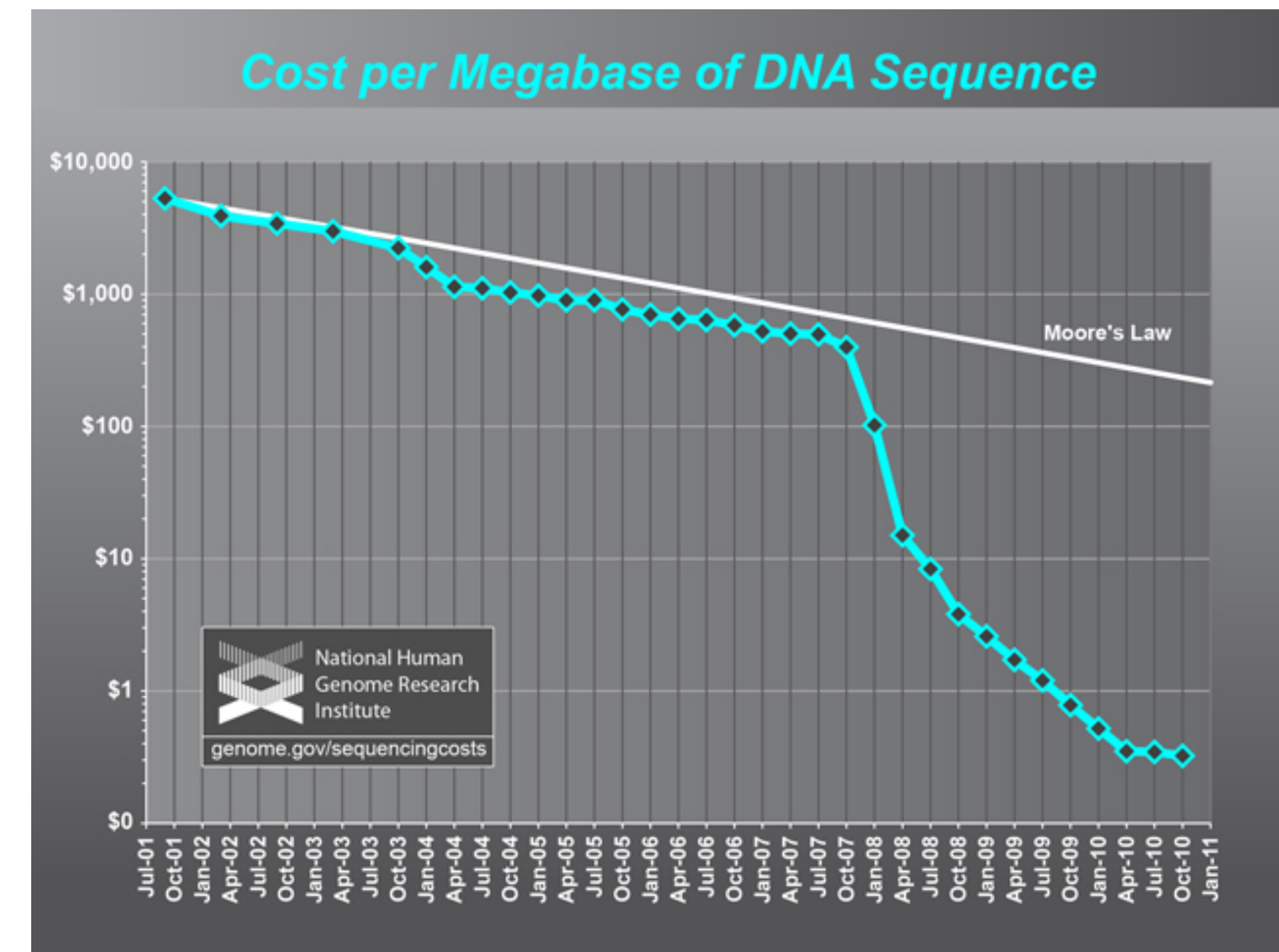
“Big Data” hasn’t just transformed industry!

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

Controversy: Hypothesis or data driven methods



Example: CERN Large Hadron Collider Data

CERN has publicly released over 300TB of data: [CERN Open Data Portal](#)

How much is that?

- A **DVD-R** holds 4.7 GB. You'd need **63,830** of them to hold 300 TB.
- 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 it ain't no thing compared to what 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

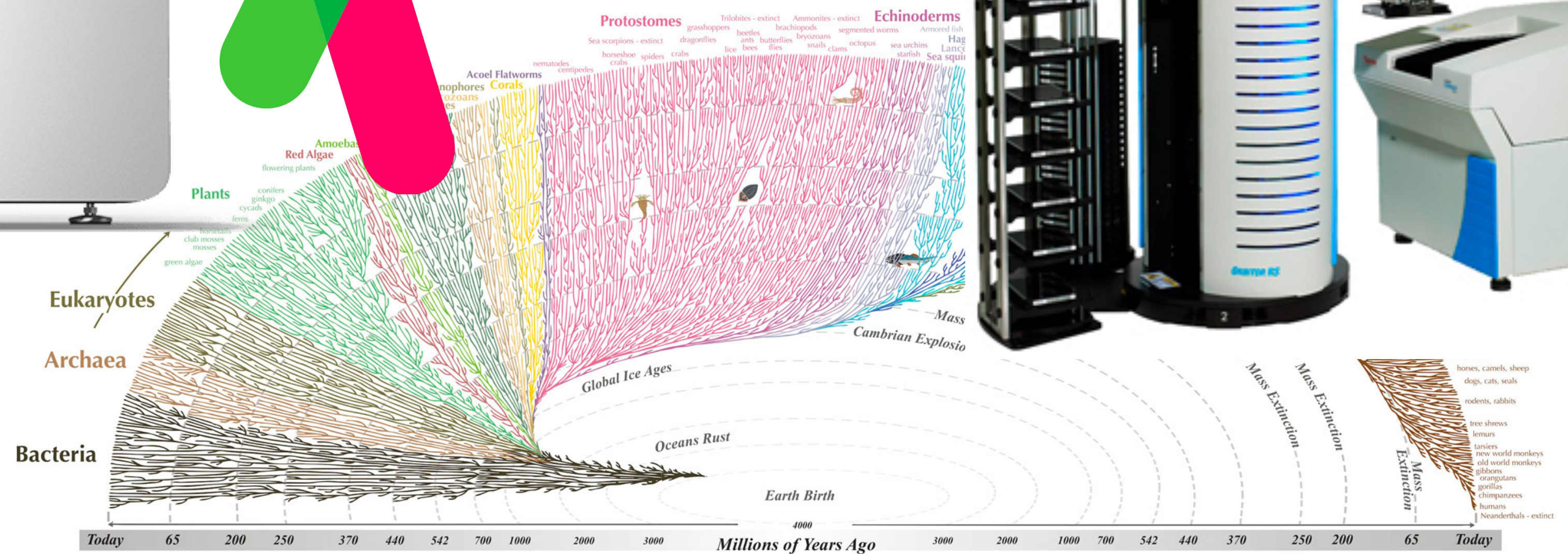
([Popular Mechanics Article](#))

Example: Genomics



23andMe

Example TCGA: 1 Petabyte



All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct

NSA Utah Data Center (Bluffdale, Utah)

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

Humans!

Human Data Interaction

Why Humans?

Leveraging human capabilities

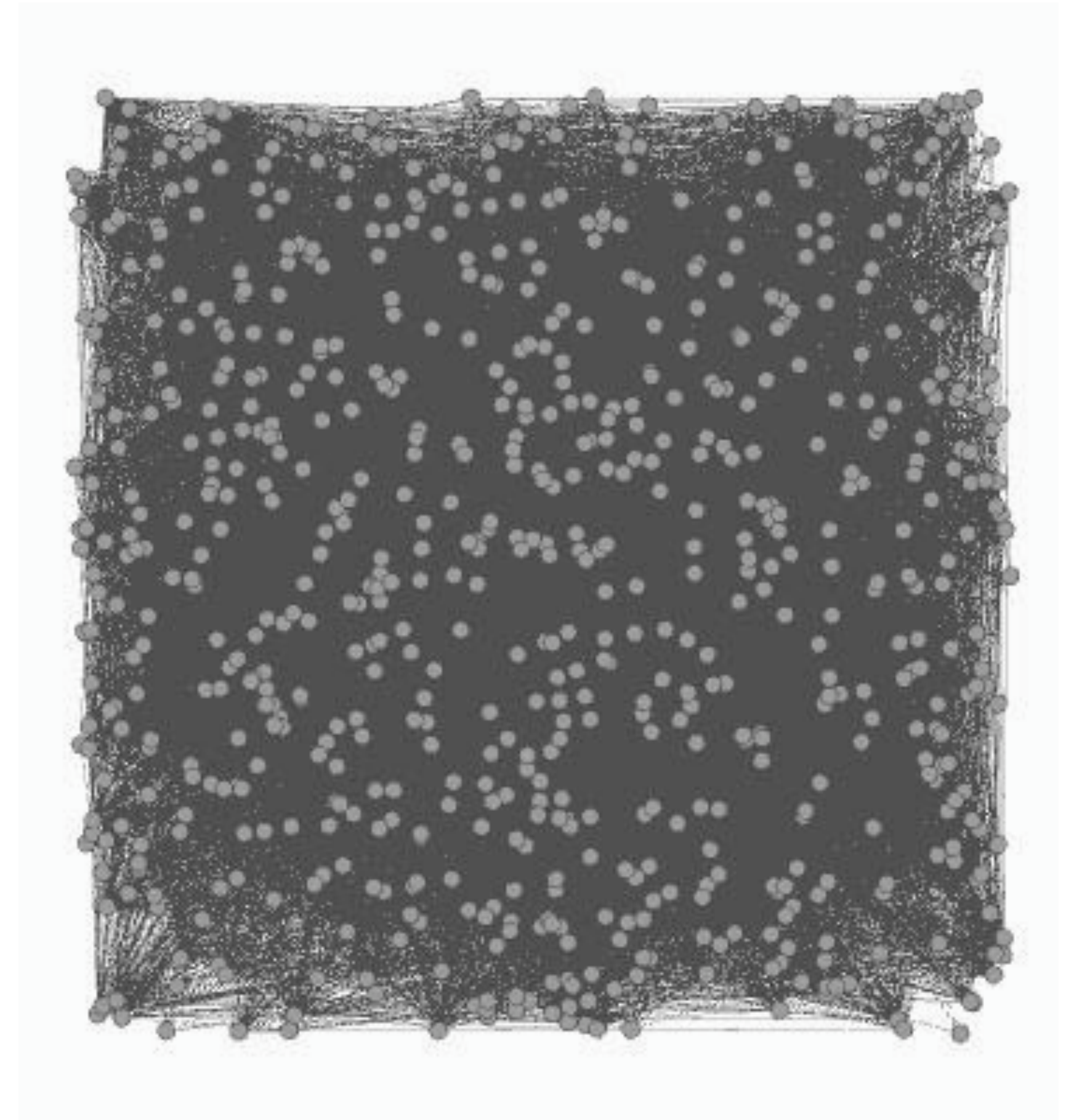
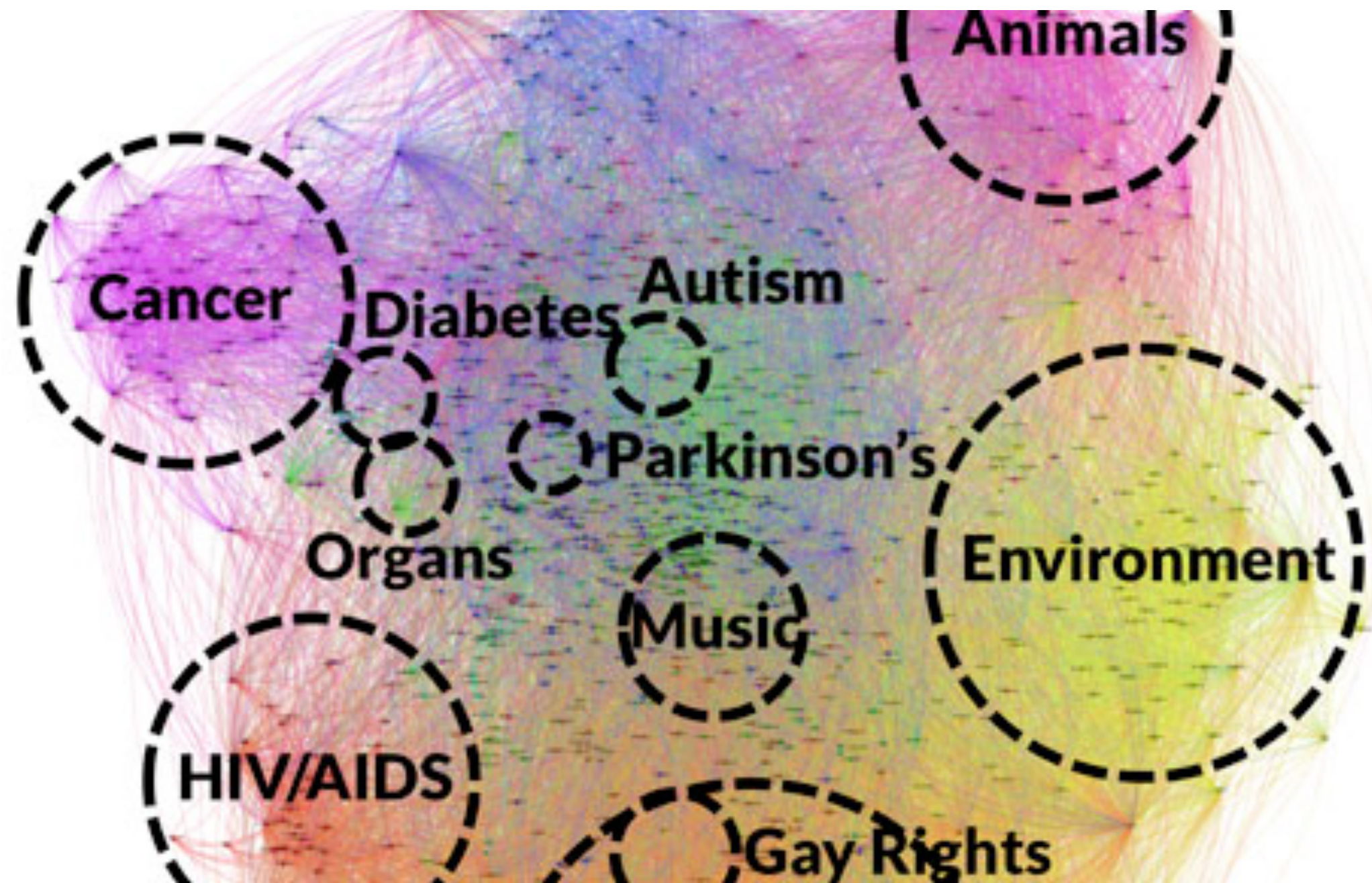
Pattern Discovery: clusters, outliers, trends

Contextual Knowledge: expectations for dataset, explanations for patterns

Action: humans learn and take action

But: we also have to **design for Humans and their limitations**

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



How did we get here?

A bit of history

“It is things that make us smart”

Donald A. Norman



15 a

E In terra pax hominibus bonae voluntatis. **Quidiam** re...

Domine deus rex caelestis deus pater omnipotens. **Dirige** re. **Es** p's al...

Domine deus agri de...

maue uiciniat maie. Dea mundi. **C** tolli peccata mundi

The first Book of

EUCLIDE'S ELEMENTS.

10 describe the circle DEH ; & let DA 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 B , being given, from the greater BC to take away the right line BE equal to the lesser A .

At the point B 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 equal to two sides of the other ED , DF ; each to its correspondent side (that is, $BA = ED$, and

11

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

If the point D be applied to the point A , and the right line DE placed 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 terms, & to consequently are equal. Wherefore the triangles BAC , DEF , and the angles B , C , as also the angles E , F , do agree; and are equal. Which was to be Demonstrated.

PROP. V.

The angles ABC , ACB at the base of an isosceles triangle ABC , are equal one to the other; and if the equal sides AB , AC be produced, the angles CBD , BCE , under the base, shall be equal one to the other.

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

Because, in the triangles ACD , AEB are $AC = AE$ and $AD = AE$, and the angle A common to them both; therefore is the angle $ADC = AEB$; and the angle $ACD = ABE$; and the base $CE = BE$; also $BC = BC$. Therefore in the triangles BCE , BCD shall be the angle $ECB = DBC$. Which was to be Done. Also therefore the angle $ABC = DCB$; but the angle $ABC = ACD$; therefore the angle $ABC = ACD$. Which was to be Done.

333

Incipit festiuitas sanctorum per ani er...

In festo sancti saturnini martyris. **Oratio.**

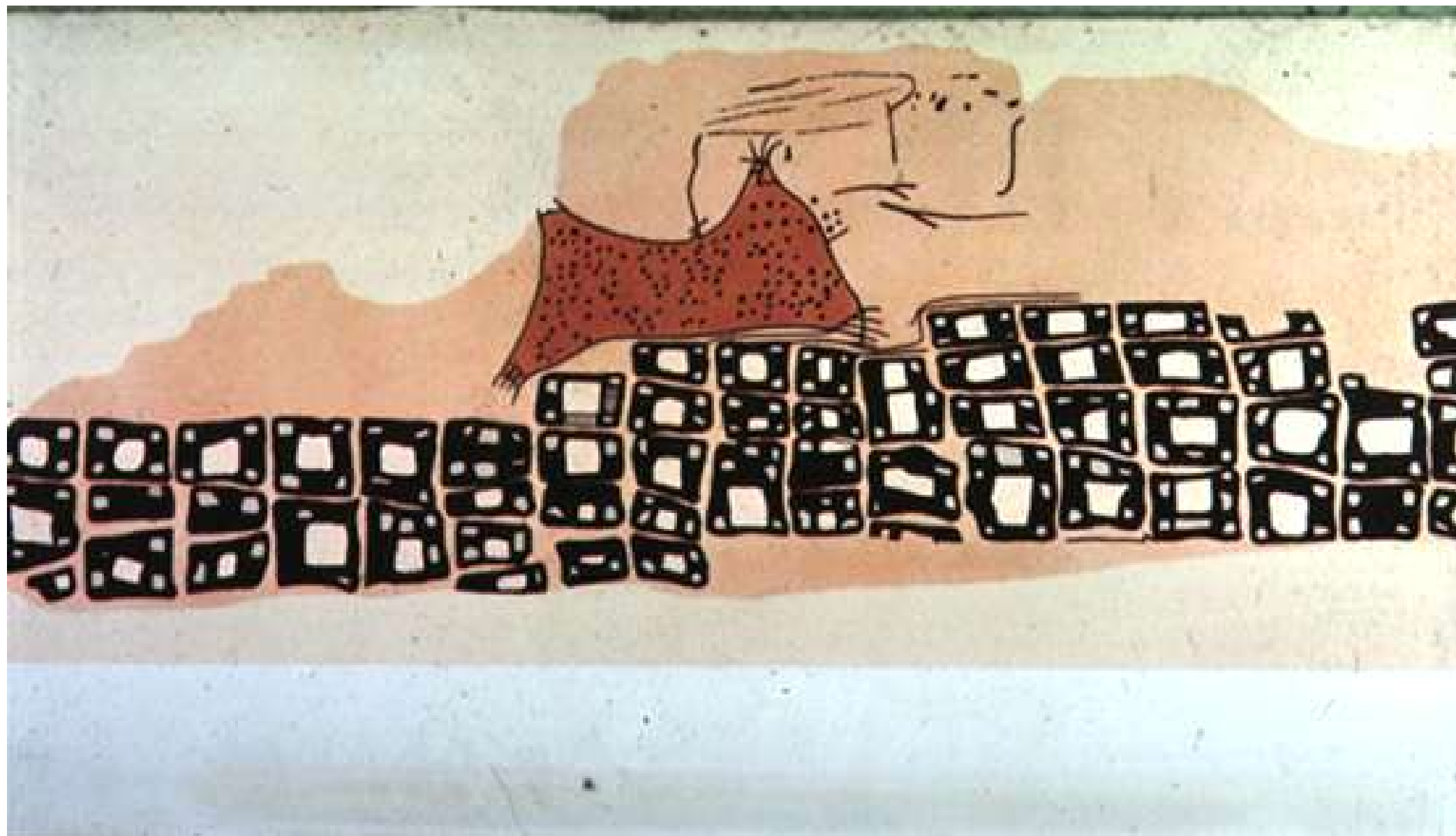
Eus qui nos beati saturnini martyris tui concedis natalicio p'sentibus nos tribue meritis adiuari. **Oratio.** **P**roter dominus nostrum. **Oratio.** **R**ome natale laueri saturnini martyris et sentio: et sicut viaconi sub maximo. Et quo primo iter altis seruas dei caritati sunt ad fodendam harenas ad faciendam thermas vicetianam. Quorum unum sibi presertim: cum interrogasset idem maximianus quis uocaretur, respondit: Ego peccator: sicut seruus seruorum domini nostri iesu christi. Eui cuius post alia dicitur maximianus: aut sacrificia deo h'rculit: aut carnes tuas cremato: respondit. Ego quidem semper hoc optaui: ueritate si meriti fuerit: corona desiderata accepta.

Maximus igitur laudato. **Actus secunda.** **P**refecto missus est in custodiam decem et

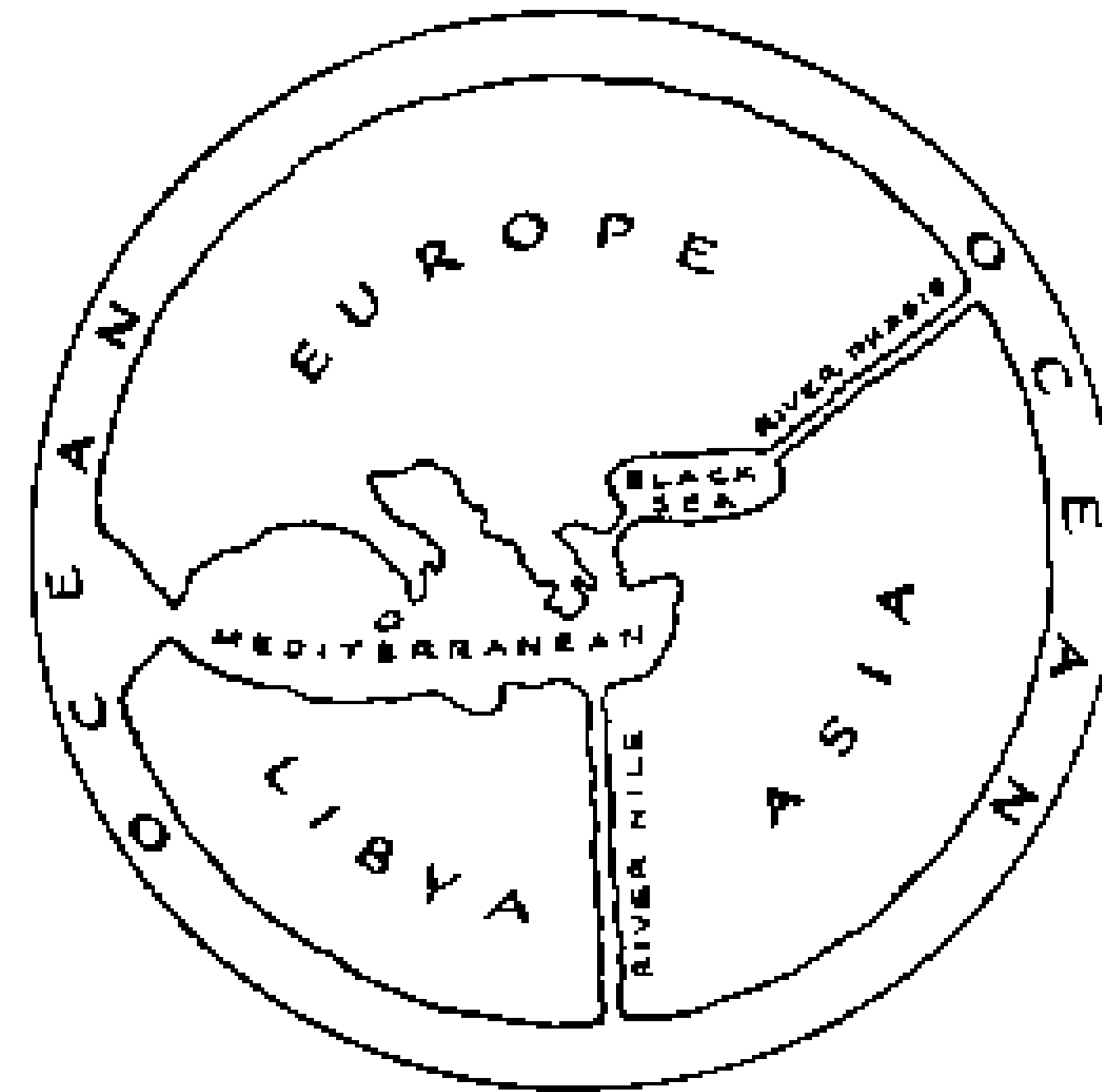
Ecce ego miras piscatores milios dicit dicit: piscabant eos. **Luc. 16.**

Mocavit iesus discipulos suos et cetera. **Mat. 23.** **E**t ipse quos aplos **Luc. 6.**

Record



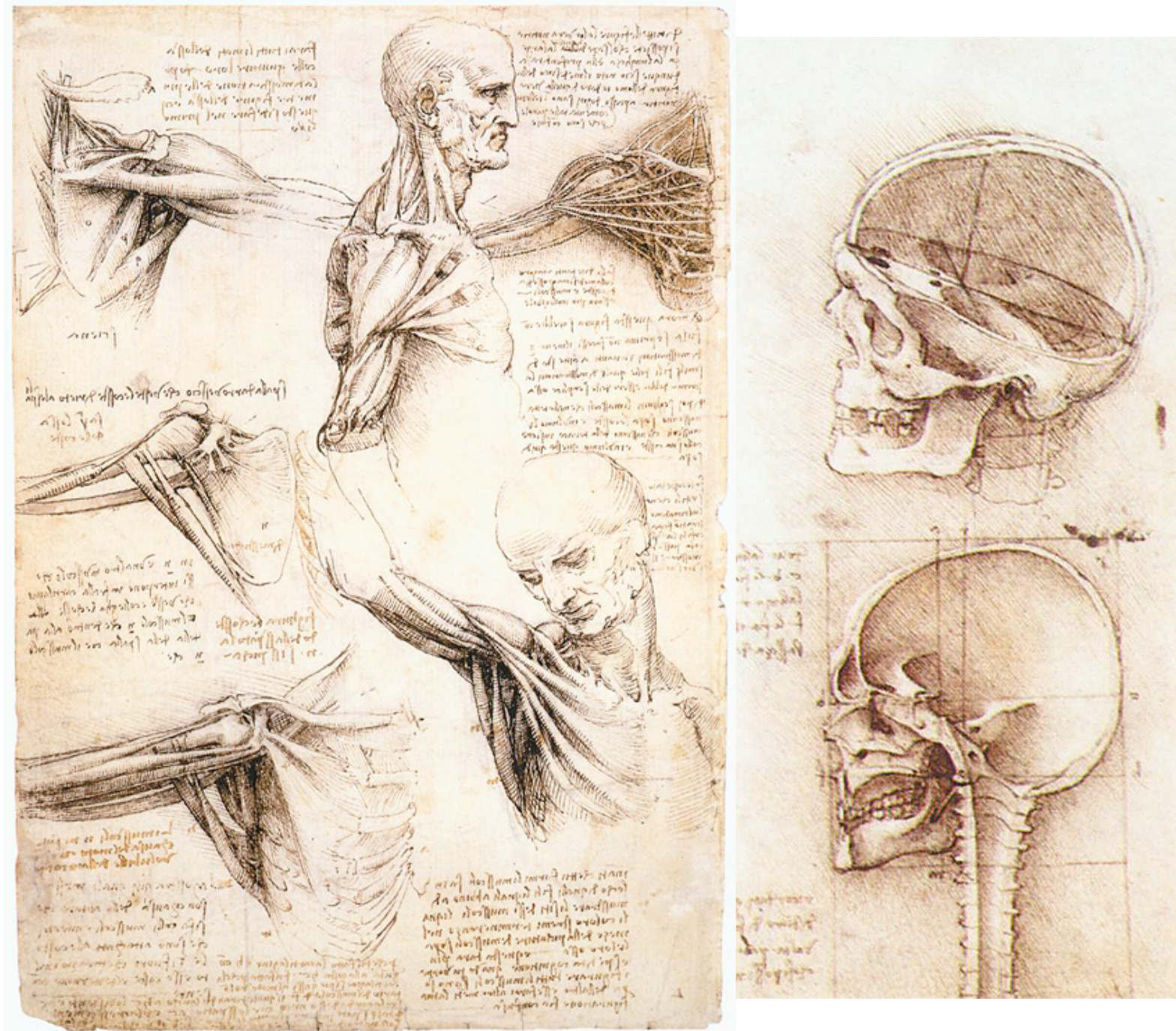
Konya town map, Turkey, c. 6200 BC



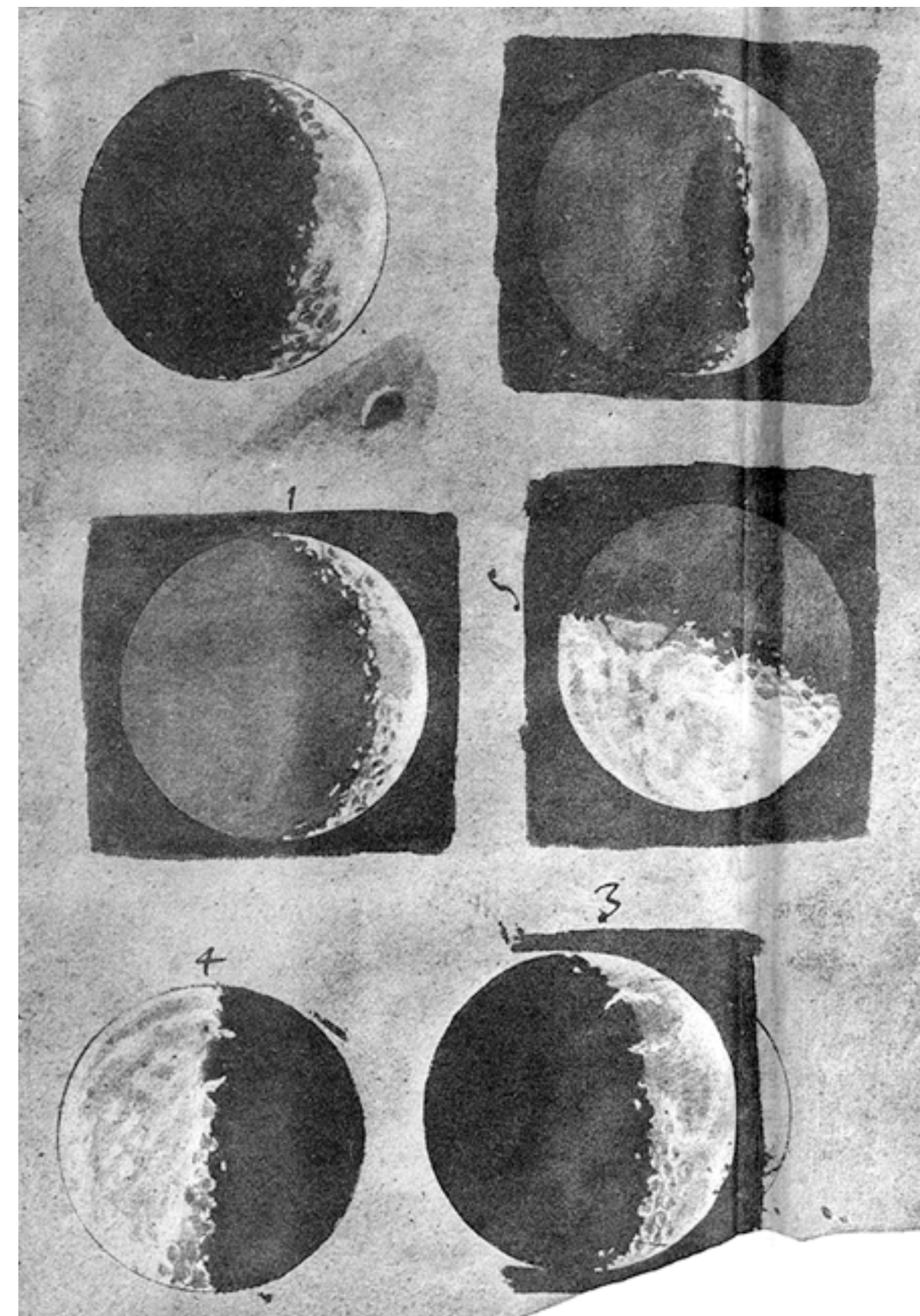
Anaximander's Map of the World

Anaximander of Miletus, c. 550 BC

Record



Leonardo Da Vinci, ca. 1500



Galileo Galilei, 1616

Donald Norman

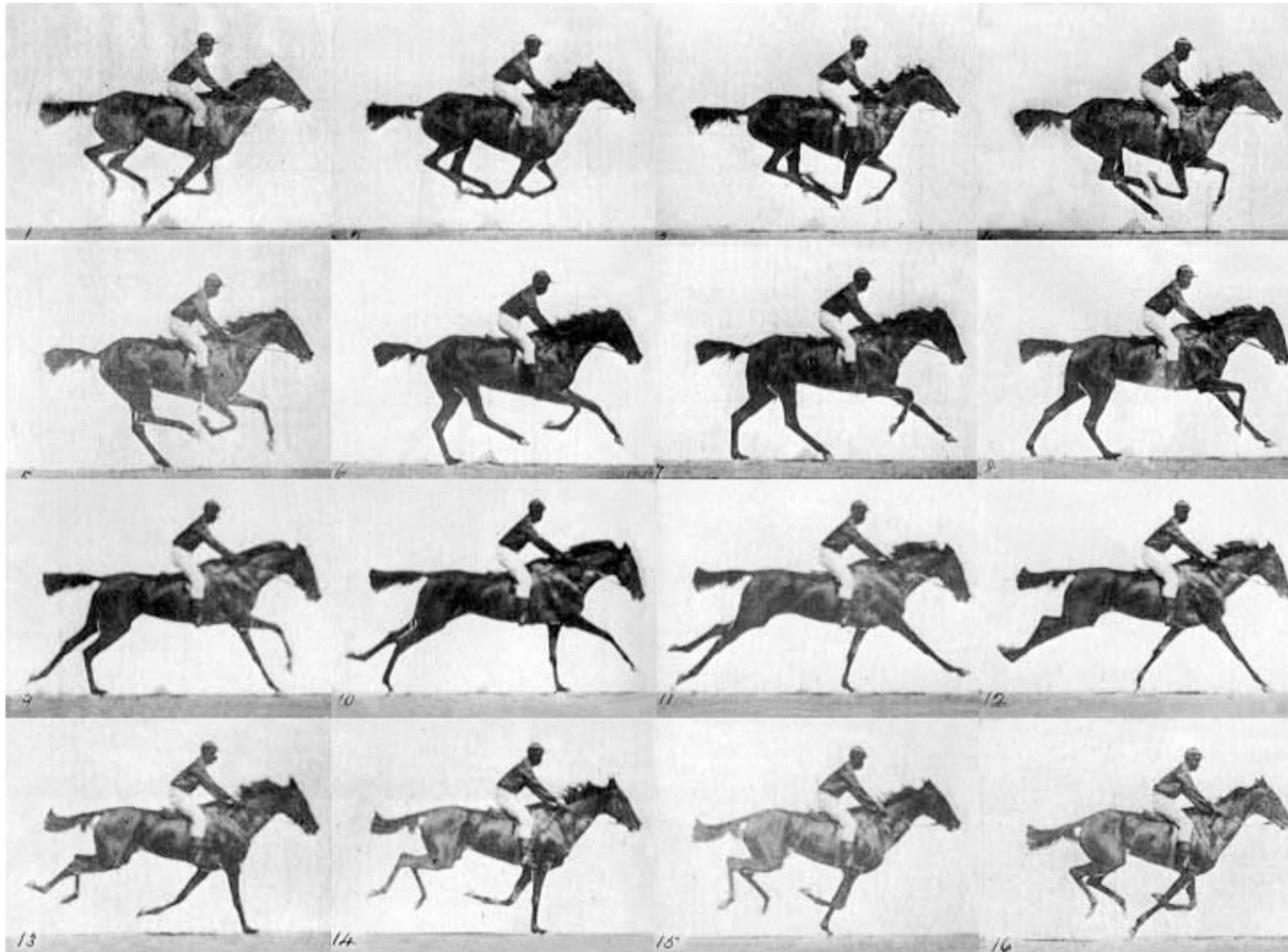


William Curtis (1746-1799)

The History of Visual Communication

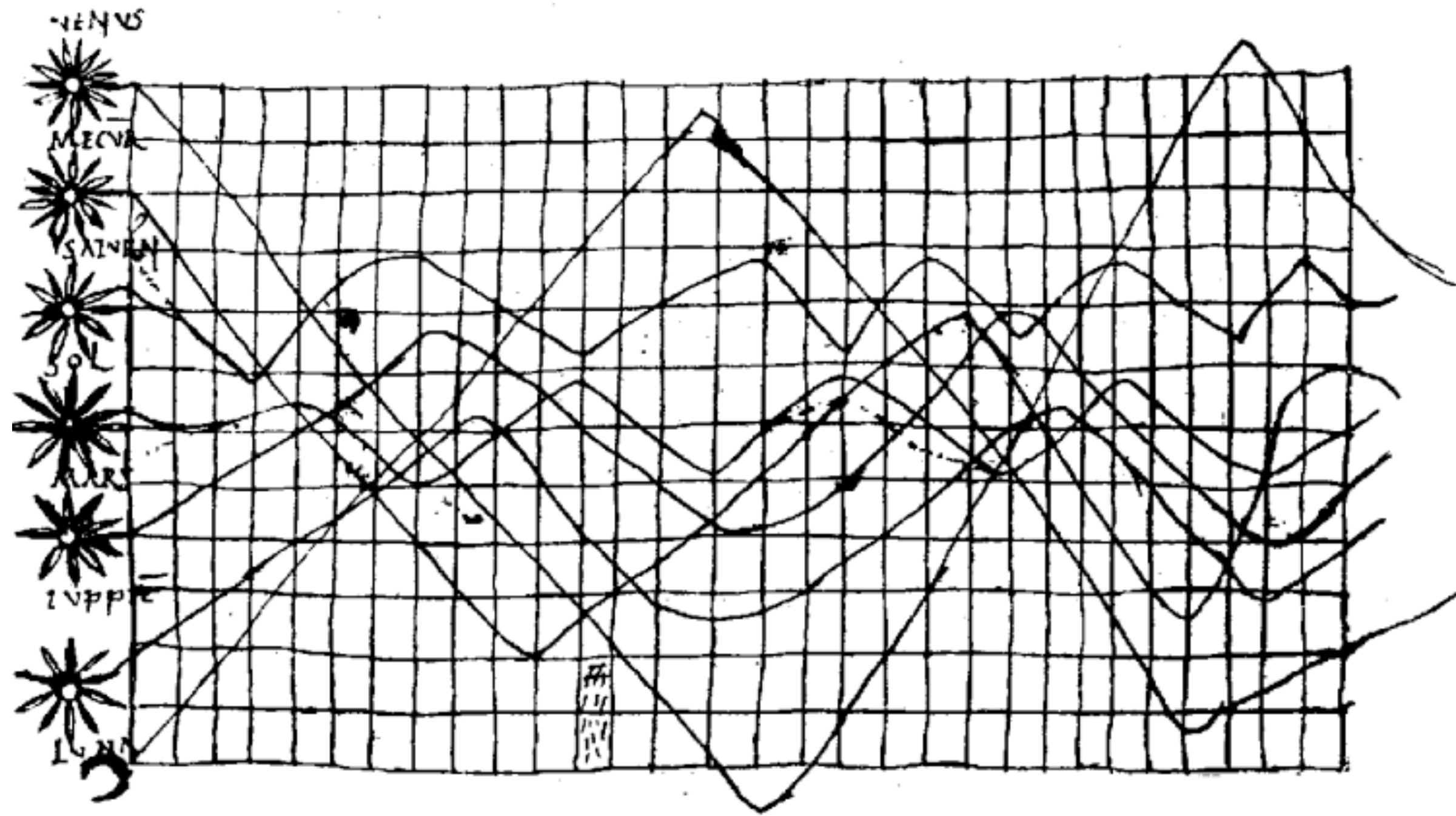
The Galileo Project, Rice University

Record

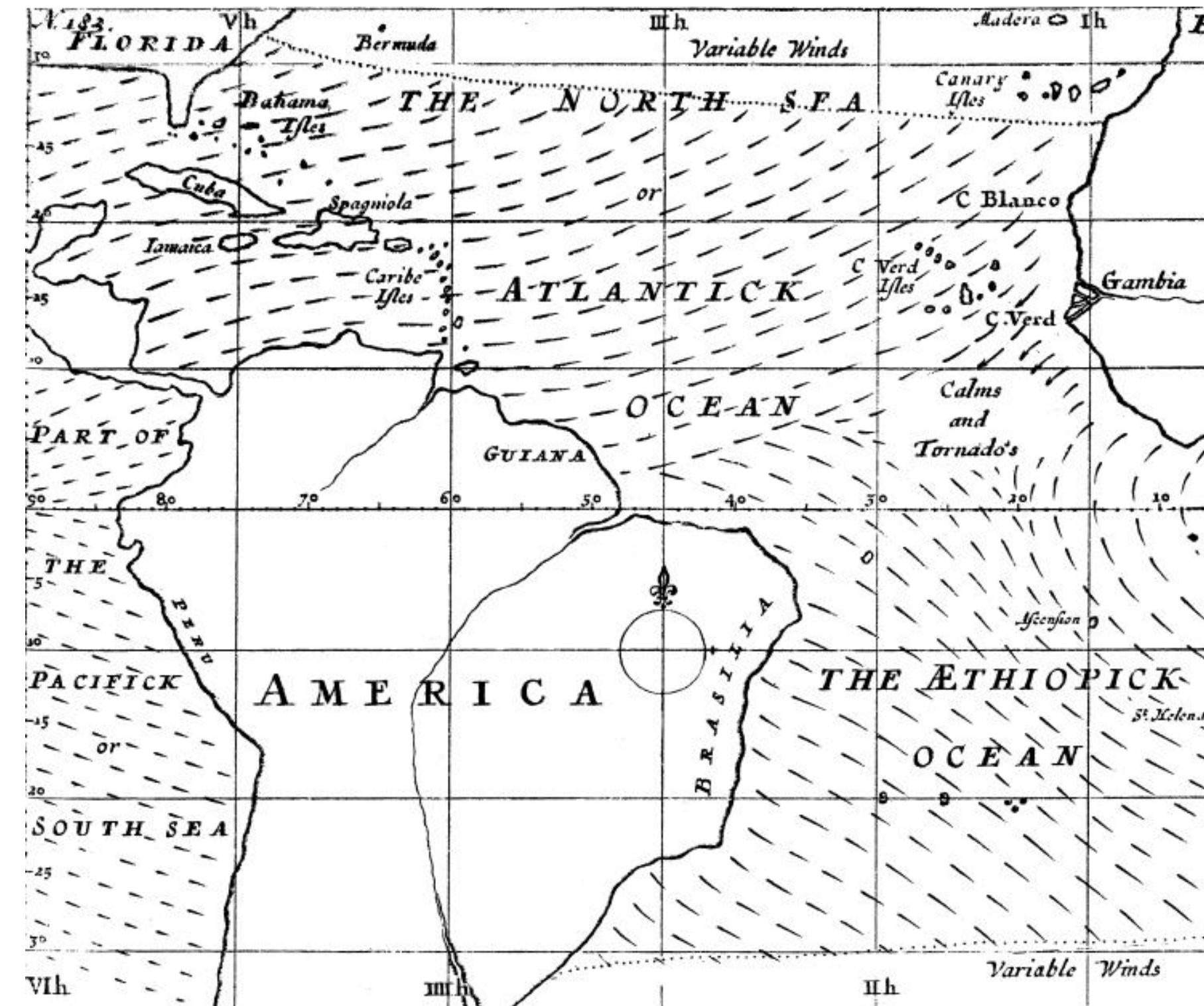


Eadweard J. Muybridge, 1878

Analyze



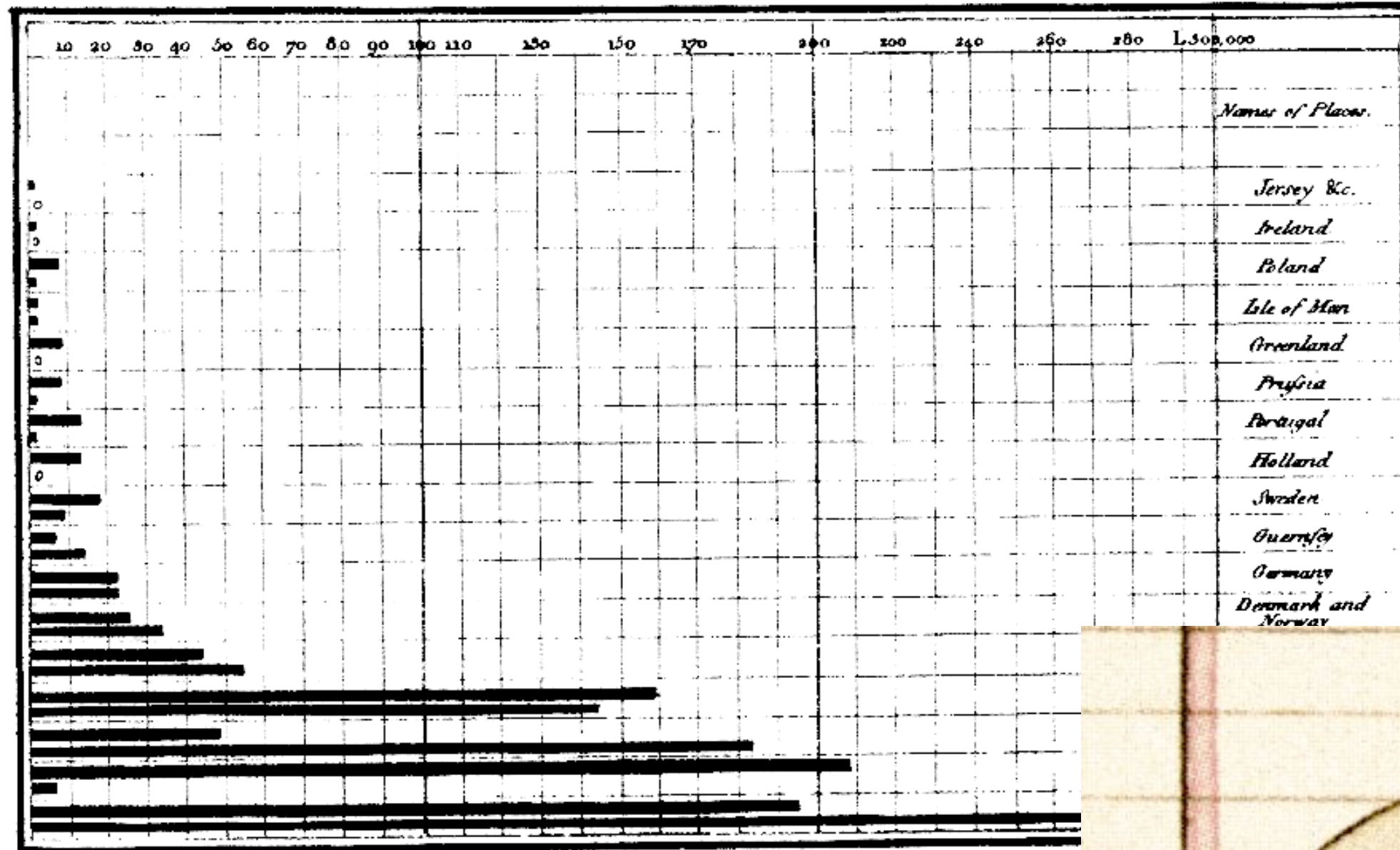
Planetary Movement Diagram, c. 950



Halley's Wind Map, 1686

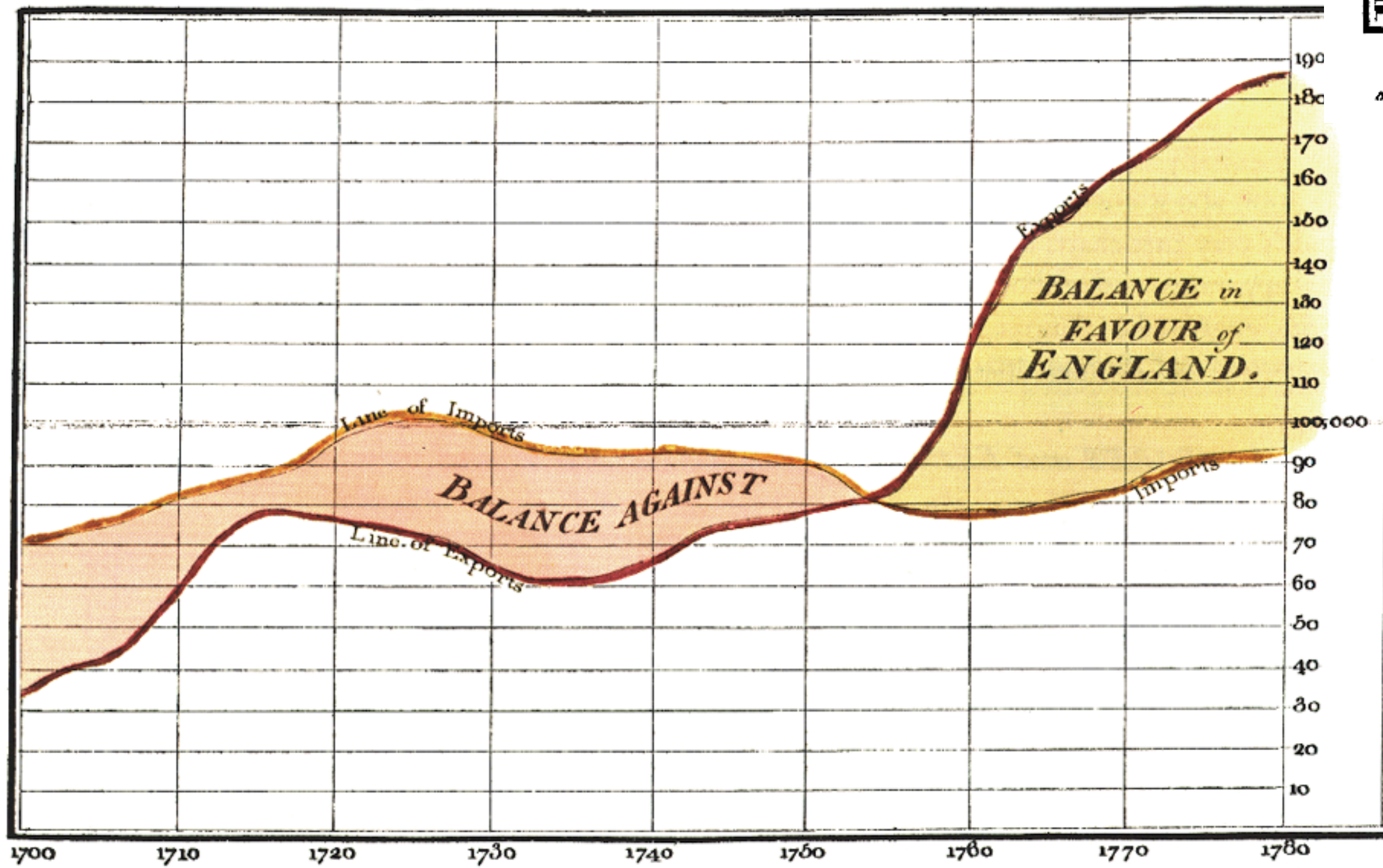
Analyze

Exports and Imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781.

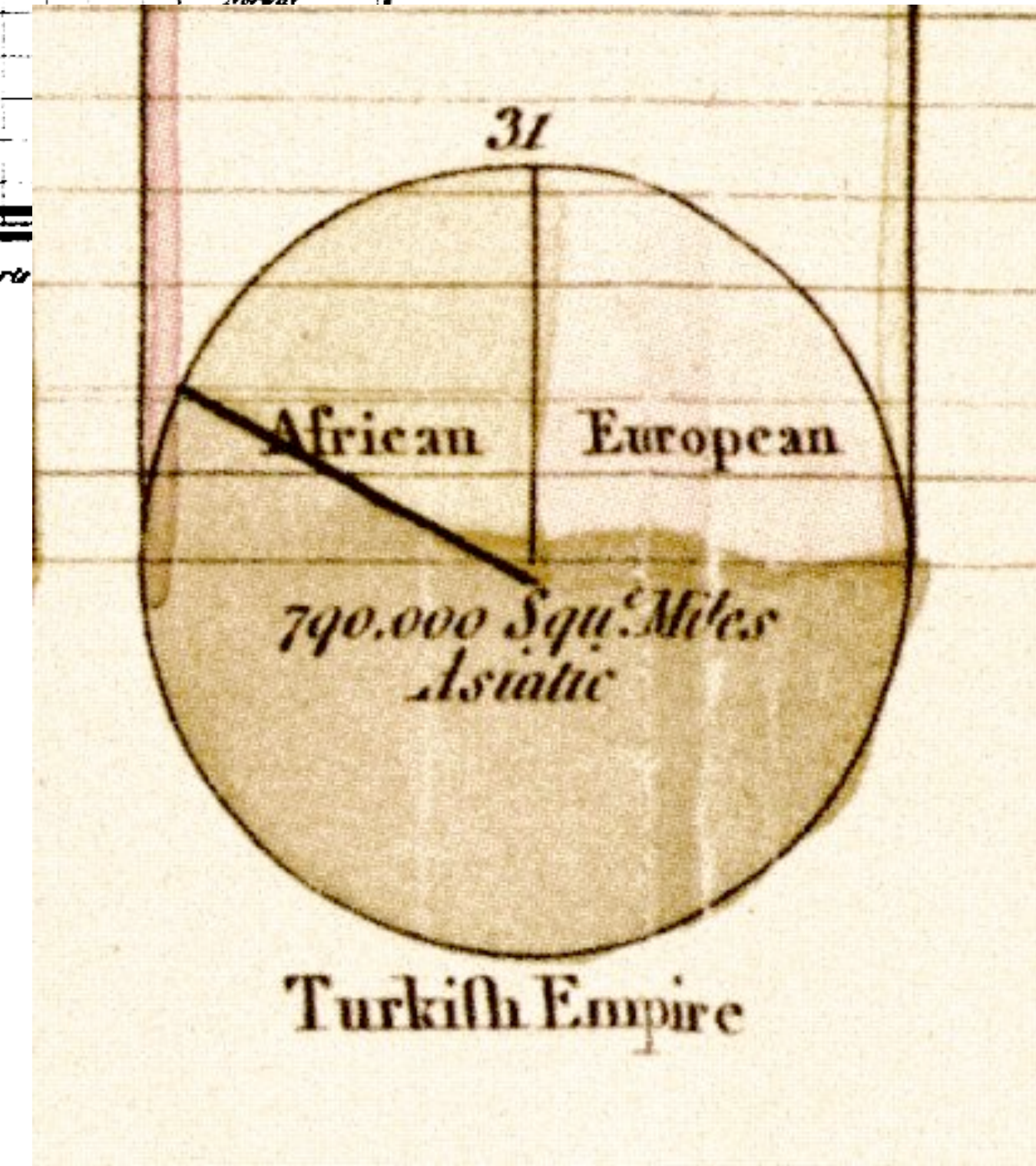


The upright divisions are Ten Thousand Pounds each. The Black Lines are Exports published in the Advertiser June 9th 1781 by W^m Playfair

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



W. Playfair, 1786



proportions of the Turkish Empire located in Asia, Europe and Africa before 1789

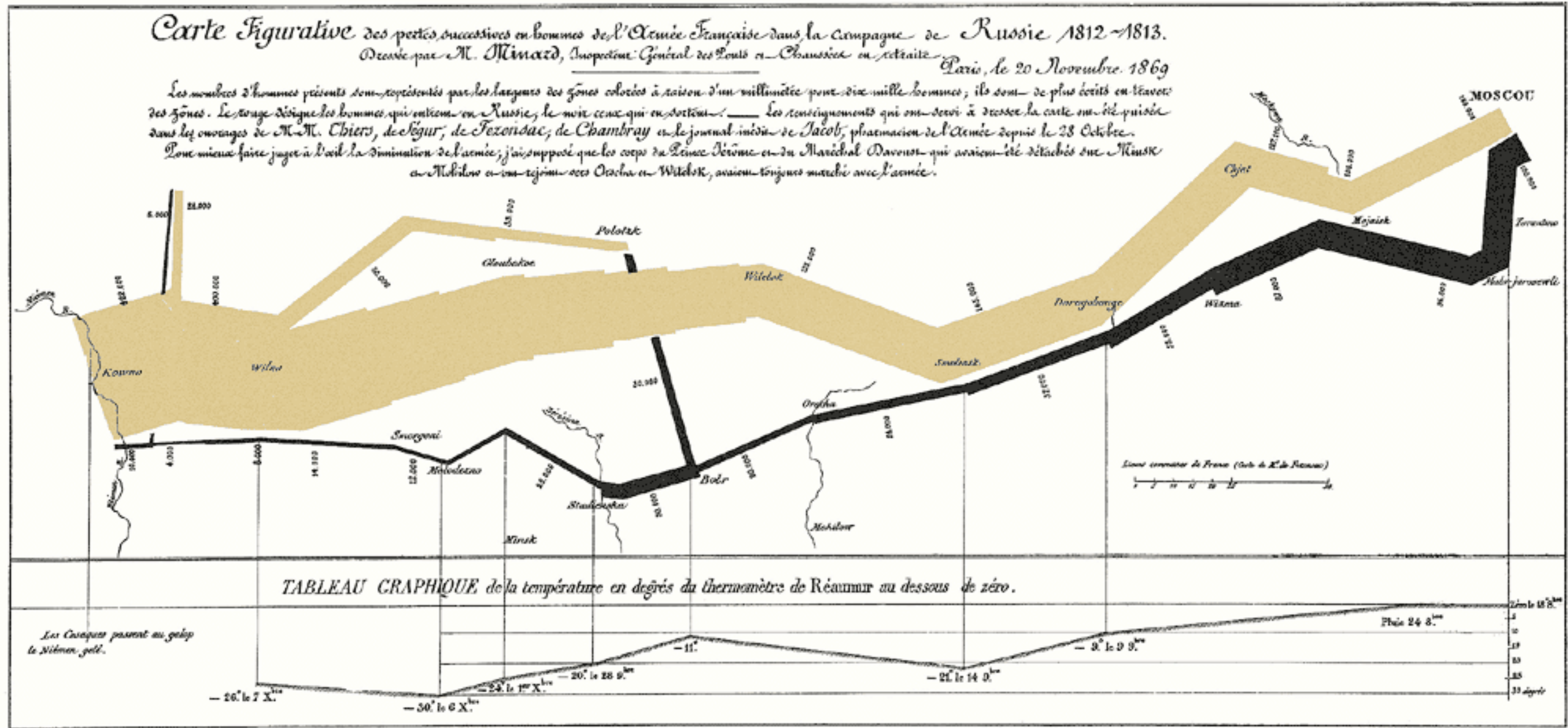
W. Playfair, 1801

Find Patterns



John Snow, 1854

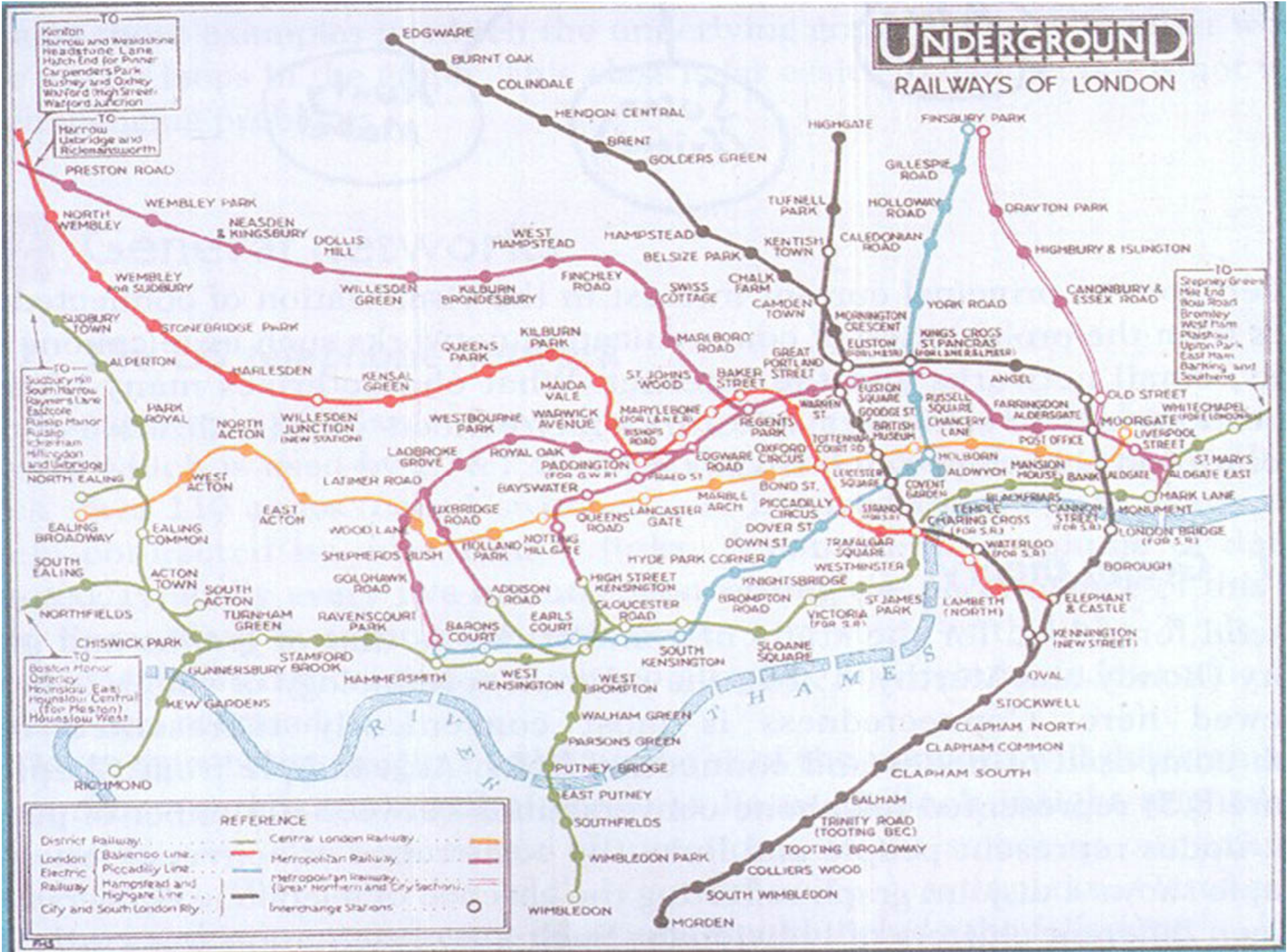
Communicate



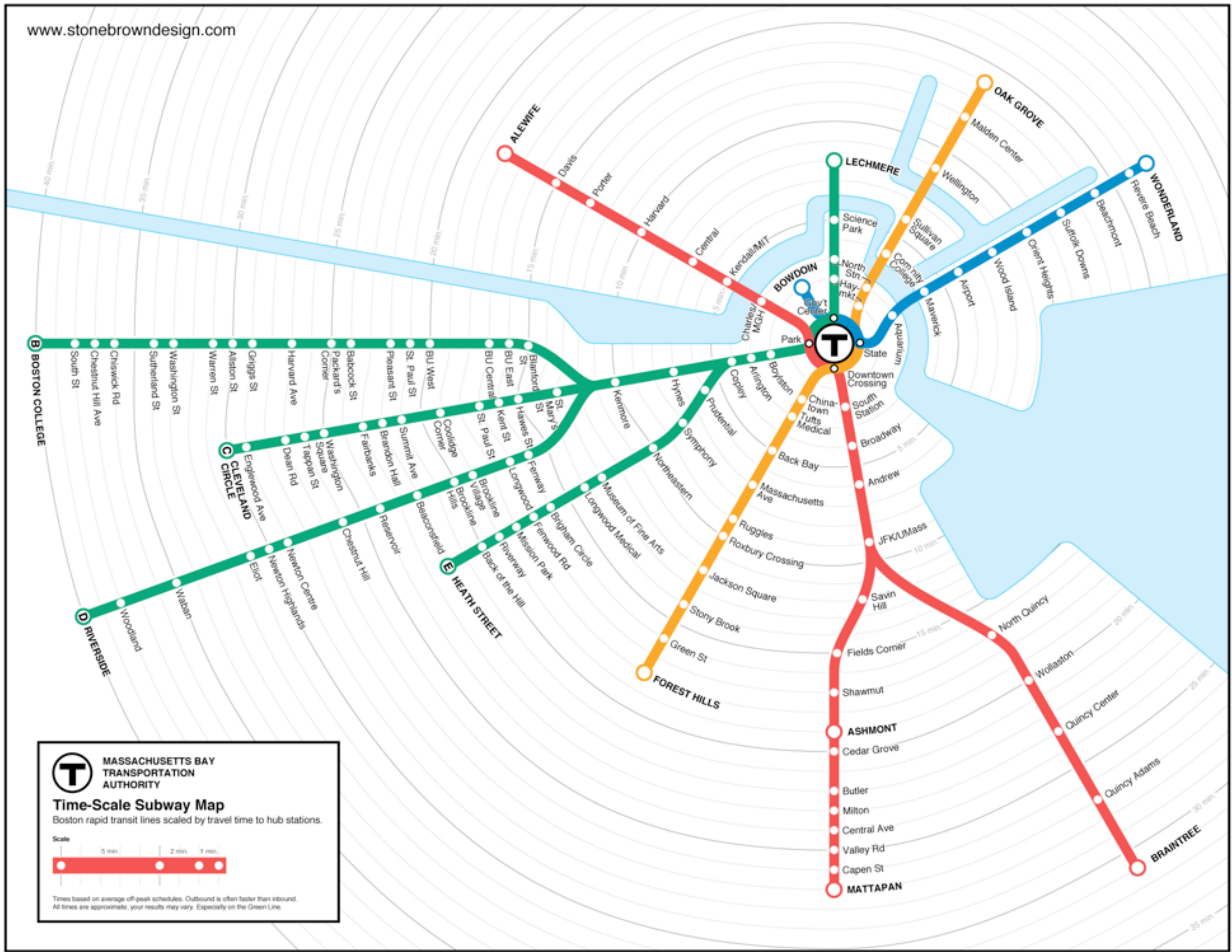


<http://infowetrust.com/scroll/>

Communicate



London Subway Map, 1927



ALEWIFE

OAK GROVE

WONDERLAND

B BOSTON COLLEGE

C CLEVELAND CIRCLE

D RIVERSIDE

E HEATH STREET

FOREST HILLS

ASHMONT

BRAINTREE

MATTAPAN

T

BOWDOIN

LECHMERE

Malden Center

Revere Beach

South St

Chestnut Hill Ave

Chiswick Rd

Sutherland St

Washington St

Warren St

Alston St

Griggs St

Harvard Ave

Packard's Corner

Babcock St

Pleasant St

St. Paul St

BU West

BU Central

BU East

Blanford St

St. Mary's

Harvard St

St. Paul St

Kenmore

Hynes

Prudential

Capley

Arlington

Boylston

China-town

Tufts Medical

South Station

Broadway

Andrew

JFK/UMass

Savin Hill

Fields Corner

Shawmut

Cedar Grove

Butler

Milton

Central Ave

Valley Rd

Capen St

Englewood Ave

Dean Rd

Washington Square

Tapscott St

Fairbanks

Brandon Hill

Summit Ave

Concord

Brookline Village

Brookline Hills

Beaconfield

Reservoir

Chestnut Hill

Newton Centre

Eliot

Newton Highlands

Woodland

Waban

Back of the Hill

Riverway

Mission Park

Fenwood Rd

Brighton Circle

Longwood Medical

Museum of Fine Arts

Northeastern

Symphony

Ruggles

Roxbury Crossing

Jackson Square

Stony Brook

Green St

North Quincy

Wollaston

Quincy Center

Quincy Adams

Wollaston

Quincy Center

Quincy Adams

Quincy Adams

Quincy Adams

40 min

35 min

30 min

25 min

20 min

15 min

10 min

5 min

5 min

10 min

15 min

20 min

25 min

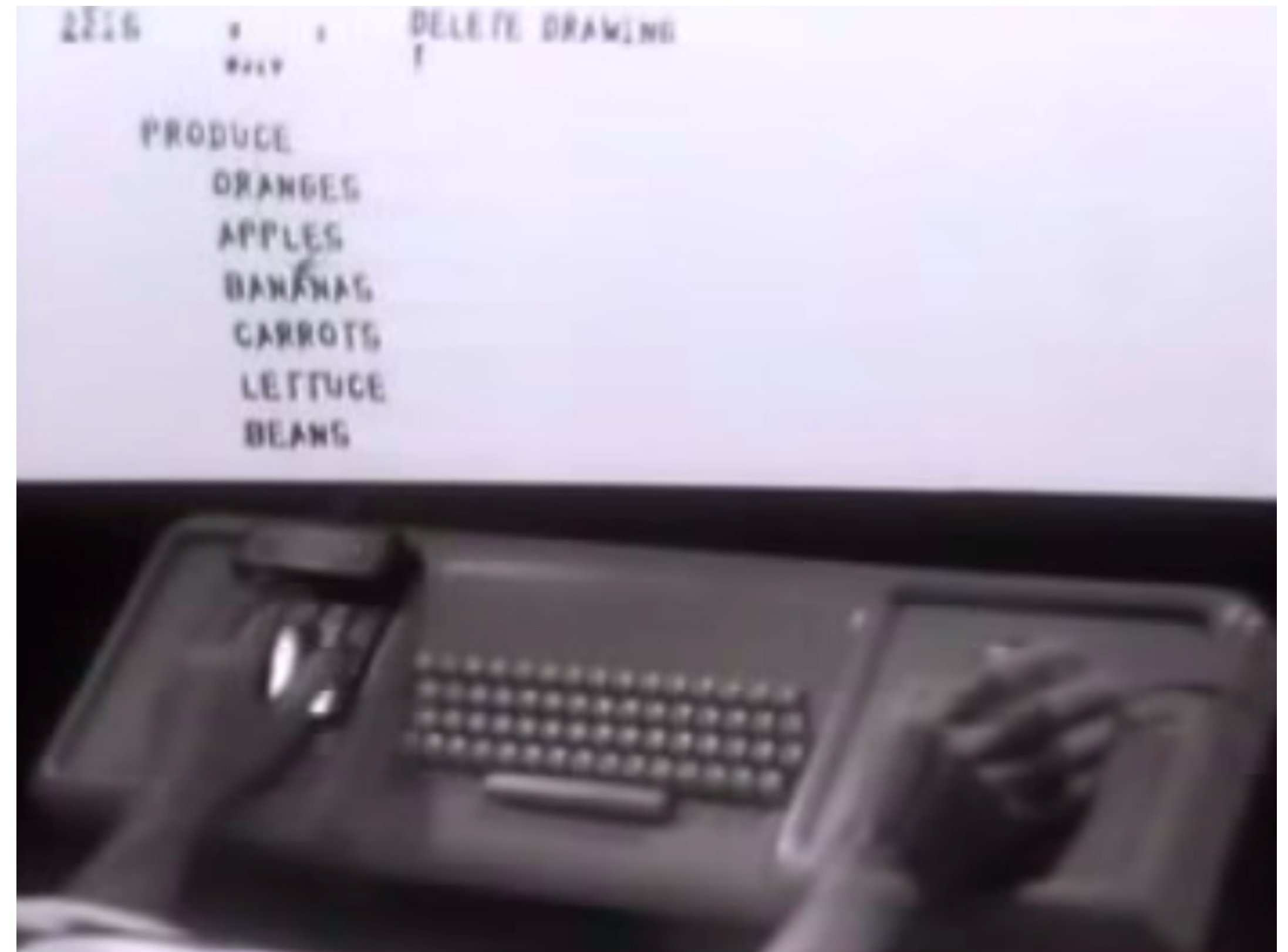
30 min

35 min

Interact



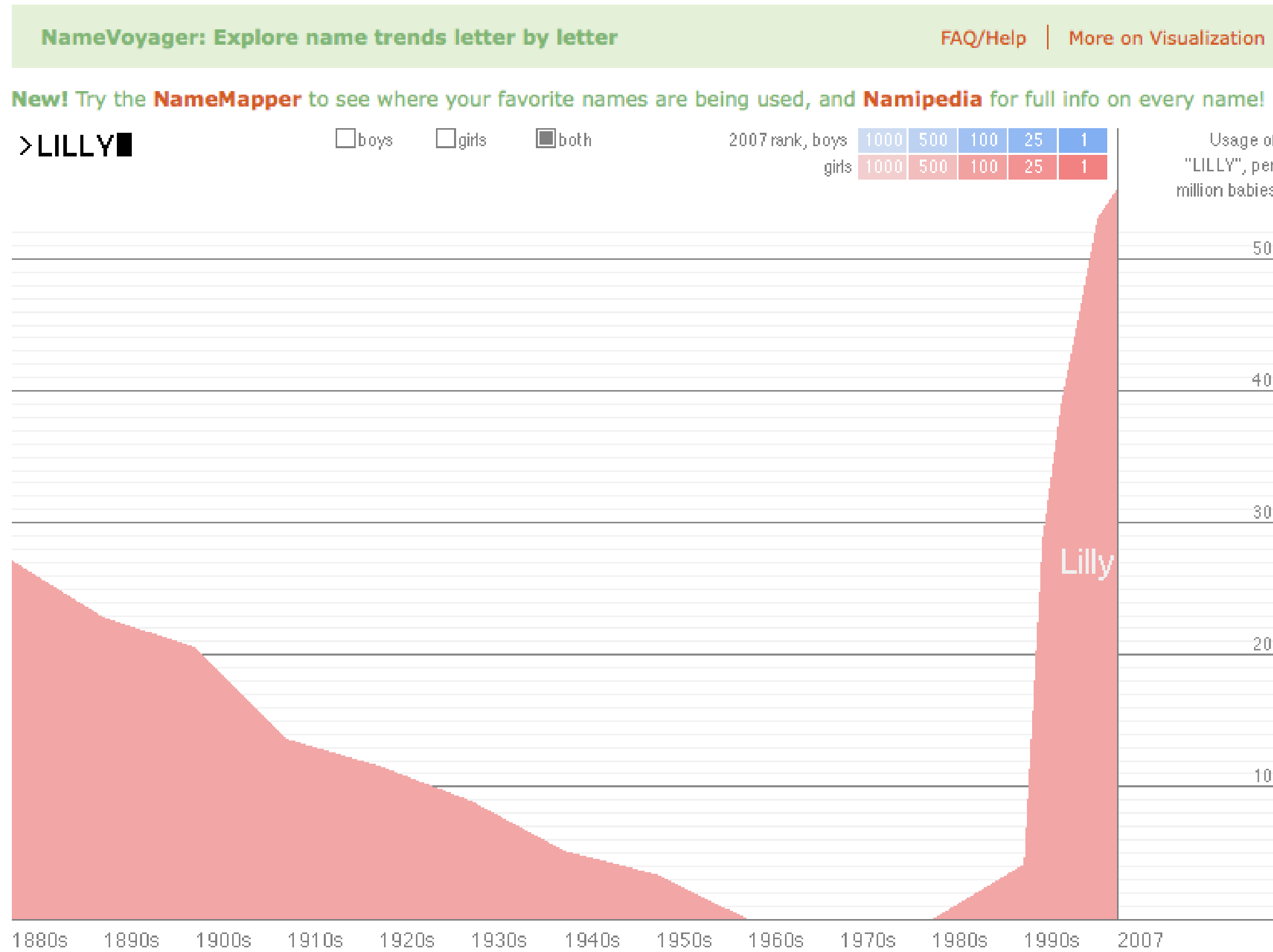
Ivan Sutherland, Sketchpad, 1963



Doug Engelbart, 1968

Modern Examples

Analyze



M. Wattenberg, 2005

Communicate



Hans Rosling, TED 2006

Who is CS-5630 / CS-6630?

Course Staff



Jen Rogers
Teaching Mentee



Kiran Gadhave
Teaching Mentee



Ilkin Safarli
Teaching Mentee

Alexander Lex

[@alexander_lex](#)
<http://alexander-lex.net>



Assistant Professor, Computer Science

Before that: Lecturer, Postdoctoral Fellow, Harvard

PhD in Computer Science, Graz University of Technology





visualization design lab

<http://vdl.sci.utah.edu/>



Miriah Meyer



Ilkin Safarli

Carolina Nobre

Sam Quinan

Jimmy Moore

Kiran Gadhave

Alexander Lex

Alex Bigelow

Pascal Goffin

Nina McCurdy

Ethan Kerzner

Jennifer Rogers

Haihan Lin

SCI Institute

Scientific Computing and Imaging Institute

Scientific Computing

Biomedical Computing

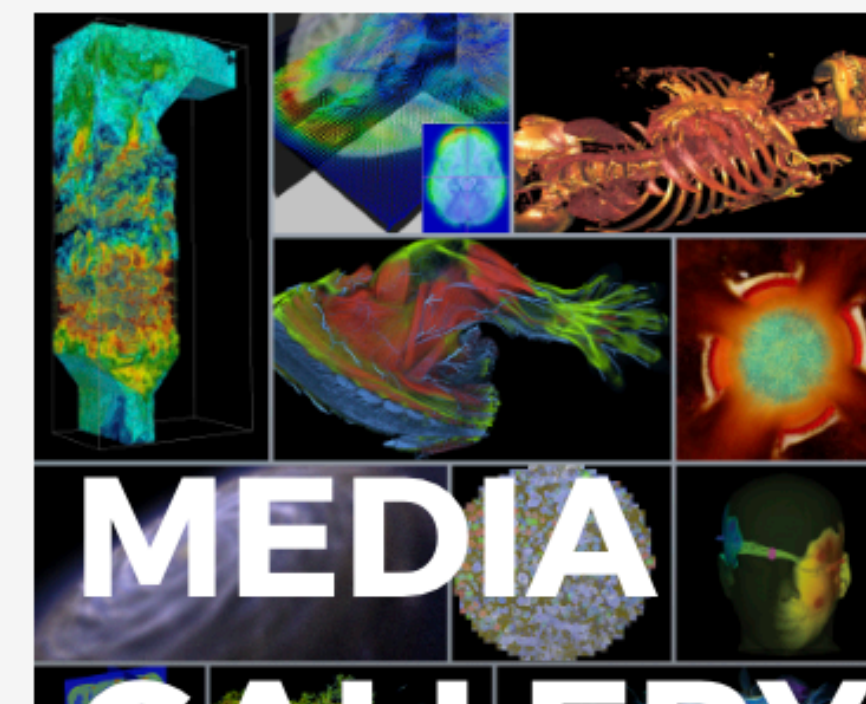
Scientific Visualization

Information Visualization

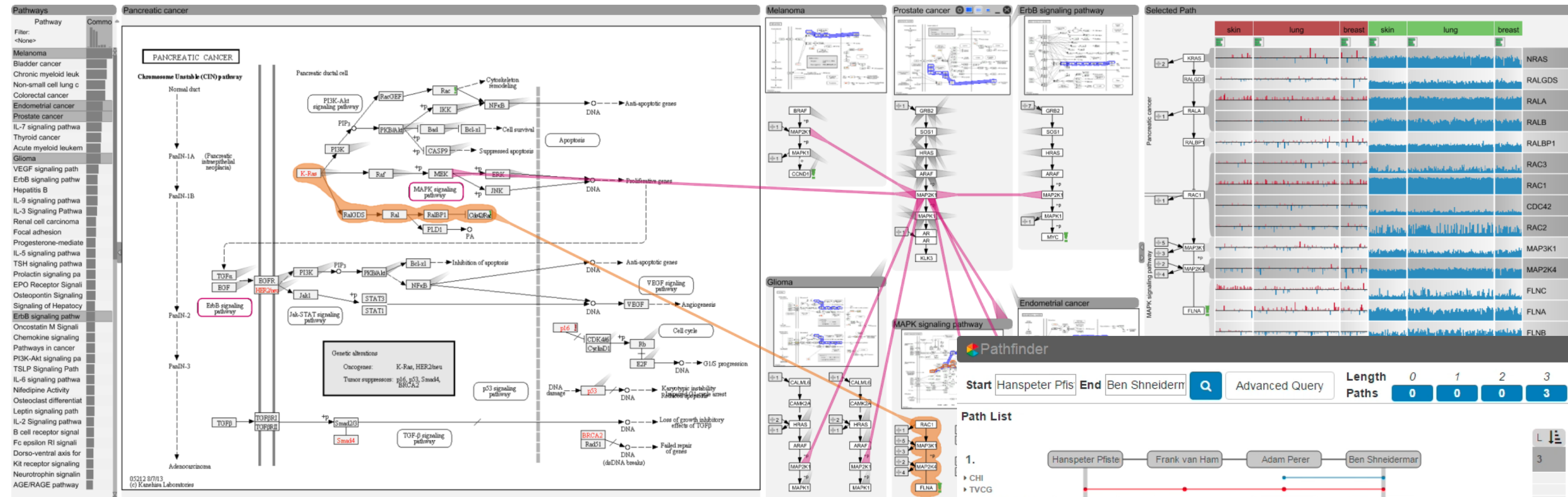
Image Analysis



<http://sci.utah.edu>



Large, Multivariate (Biological) Networks



MATCH p = (s)-[x:FLIGHT]->()-[y:FLIGHT]->(t) WHERE s.state in ['CA', 'OR', 'WA'] AND t.state in ['CT', 'ME', 'MA', 'RI', 'NH', 'VT', 'WA', 'TX', 'GA'] AND x.carrier = y.carrier AND x.arr_time < y.dep_time

Submit
 Query statistics
 Length 2
 Paths 18389

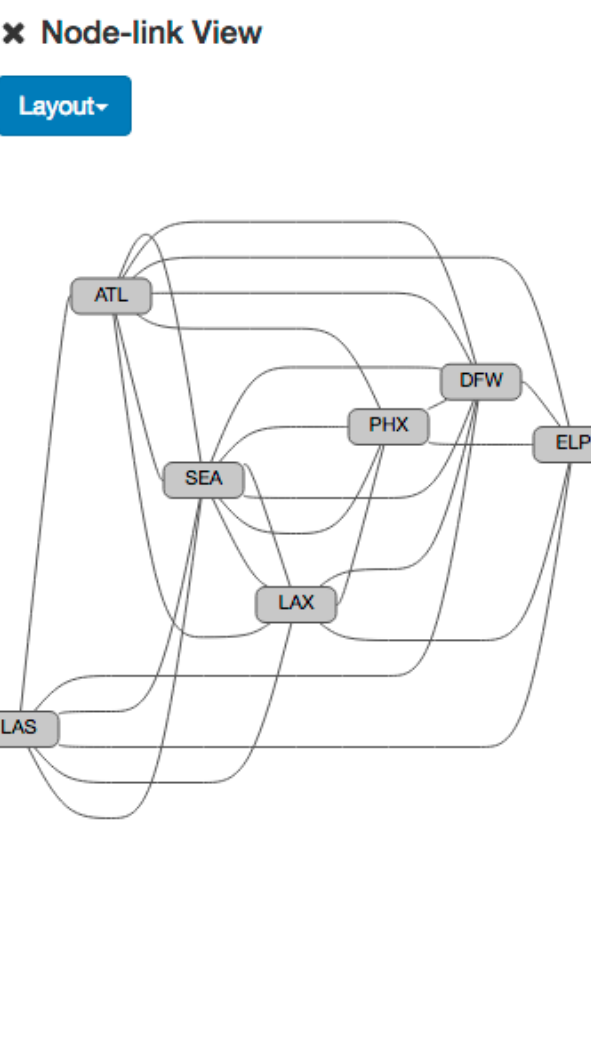
Collapse rows: state
 Collapse cols: none
 Sort order: Select
 Cell encoding: colormap
 Path metric: path count

Legend: 280-1955 (color scale)

city	state	market	id	count
Oakland	CA	32457	OAK	105
Los Angeles	CA	32575	LAX	209
San Diego	CA	33570	SAN	313
Monterey	CA	34922	MRY	417
Santa Barbara	CA	34689	SBA	521
San Francisco	CA	32457	SFO	625
Ontario	CA	32575	ONT	729
San Luis Obispo	CA	34236	SBP	
San Jose	CA	32457	SJC	
Sacramento	CA	33192	SMF	
Fresno	CA	31638	FAT	
Crescent City	CA	30930	CEC	
Burbank	CA	32575	BUR	
Carlsbad	CA	31041	CLD	
Palm Springs	CA	34282	PSP	
Santa Ana	CA	32575	SNA	
Arcata/Eureka	CA	30157	ACV	
Bakersfield	CA	30561	BFL	
Mammoth Lakes	CA	33388	MMH	
Long Beach	CA	32575	LGB	
Redding	CA	33792	RDD	
Santa Maria	CA	34905	SMX	
Portland	OR	34057	PDX	
Medford	OR	33264	MFR	
Bend/Redmond	OR	34489	RDM	
Eugene	OR	31603	EUG	
Pasco/Kennewick/Richland	WA	34252	PSC	

Intermediate Nodes

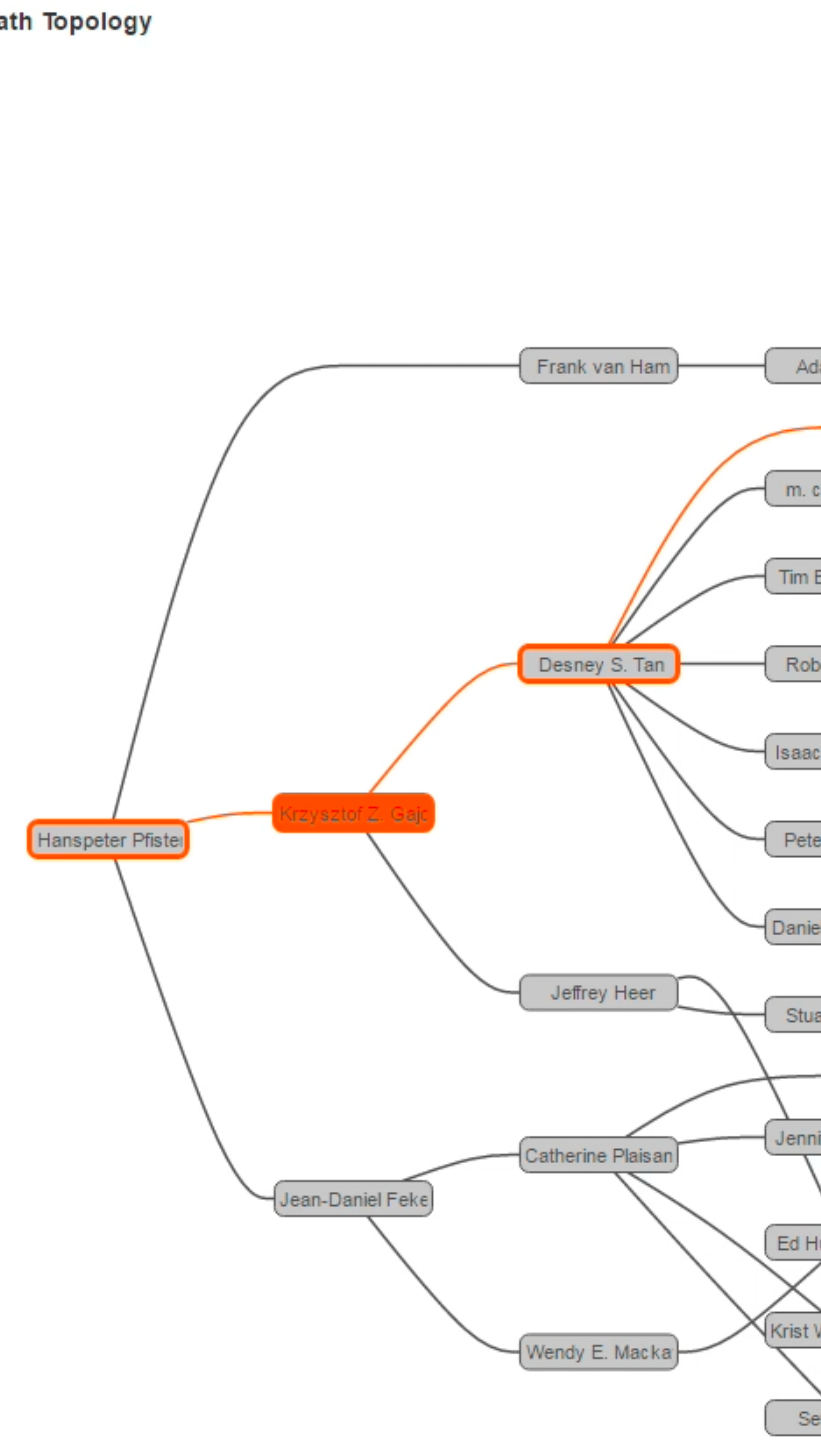
id	count
LAS	105
SEA	209
ORD	313
SLC	417
PHX	521
LAX	625
DFW	729
ELP	
ATL	



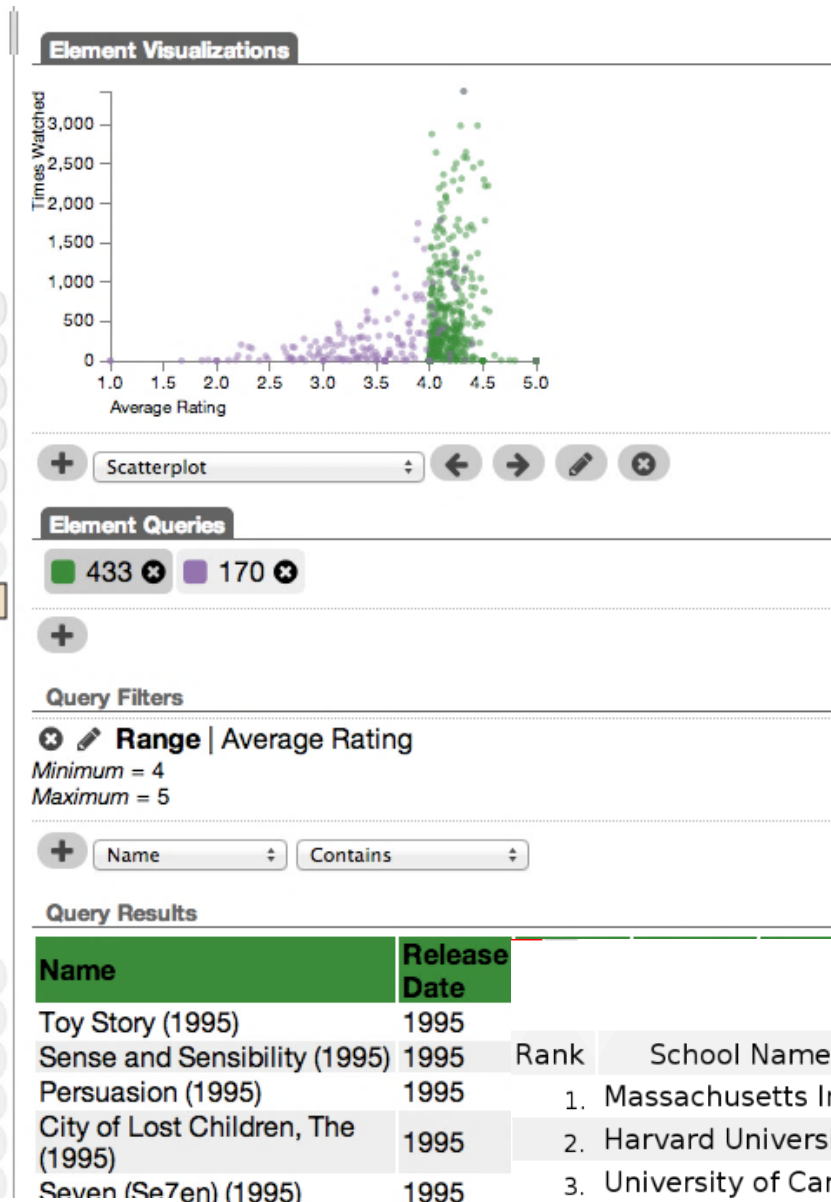
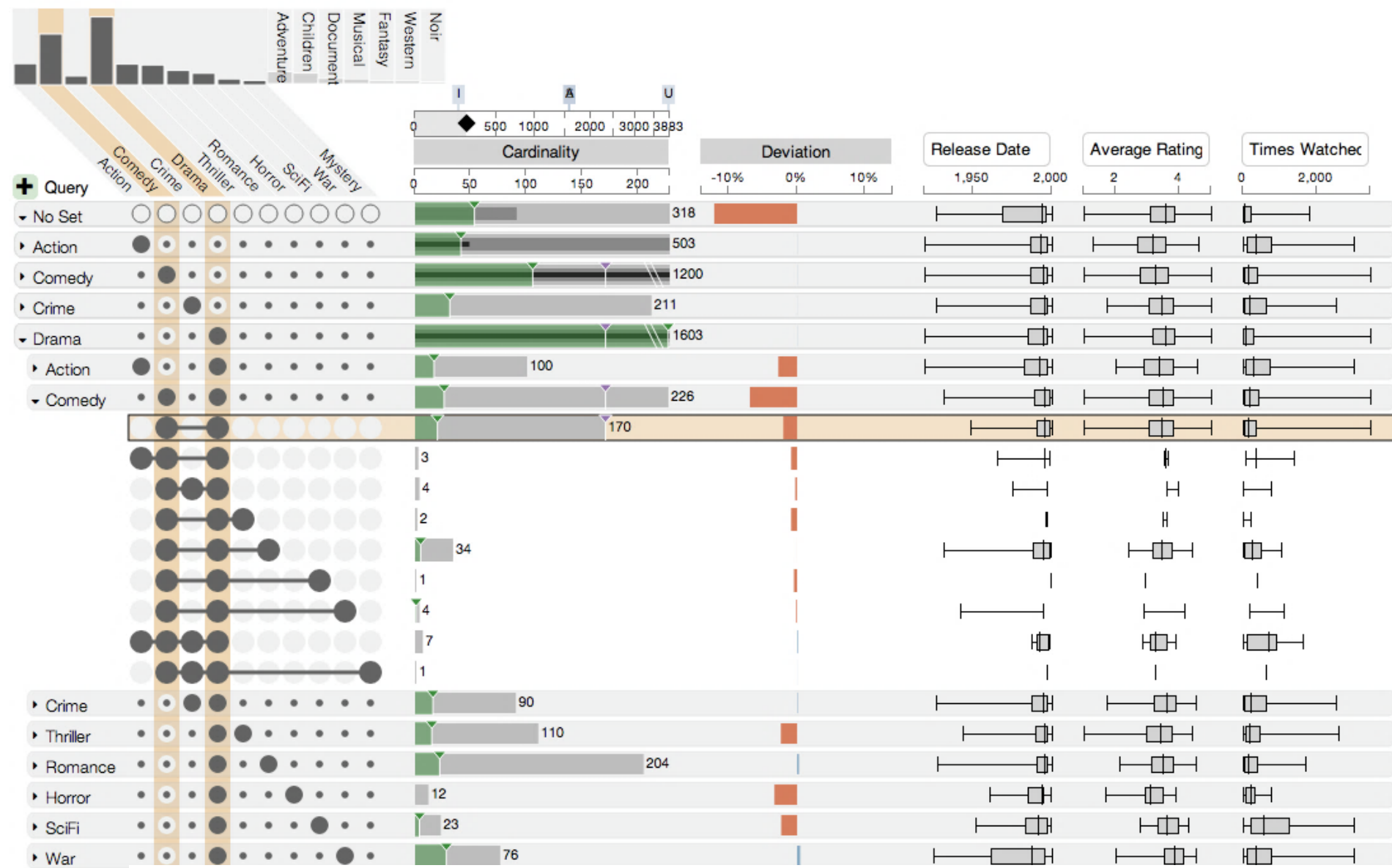
Path List

Start: Hanspeter Pfirs, End: Ben Shneiderm, Length: 3, Paths: 105

Path	Nodes	Length	Paths
1.	Hanspeter Pfirs - Frank van Ham - Adam Perer - Ben Shneiderm	3	105
1.	Hanspeter Pfirs - Kristof Z. Gaj - Desney S. Tan - Ben Shneiderm	3	105
1.	Hanspeter Pfirs - Jean-Daniel Feké - Catherine Plaisan - Ben Shneiderm	3	105
4.	Hanspeter Pfirs - Jean-Daniel Feké - Catherine Plaisan - Jennifer Golbeck - Ben Shneiderm	4	105
4.	Hanspeter Pfirs - Jean-Daniel Feké - Wendy E. Macka - Ed Hwai-hsin Ch - Ben Shneiderm	4	105
4.	Hanspeter Pfirs - Kristof Z. Gaj - Jeffrey Heer - Ed Hwai-hsin Ch - Ben Shneiderm	4	105

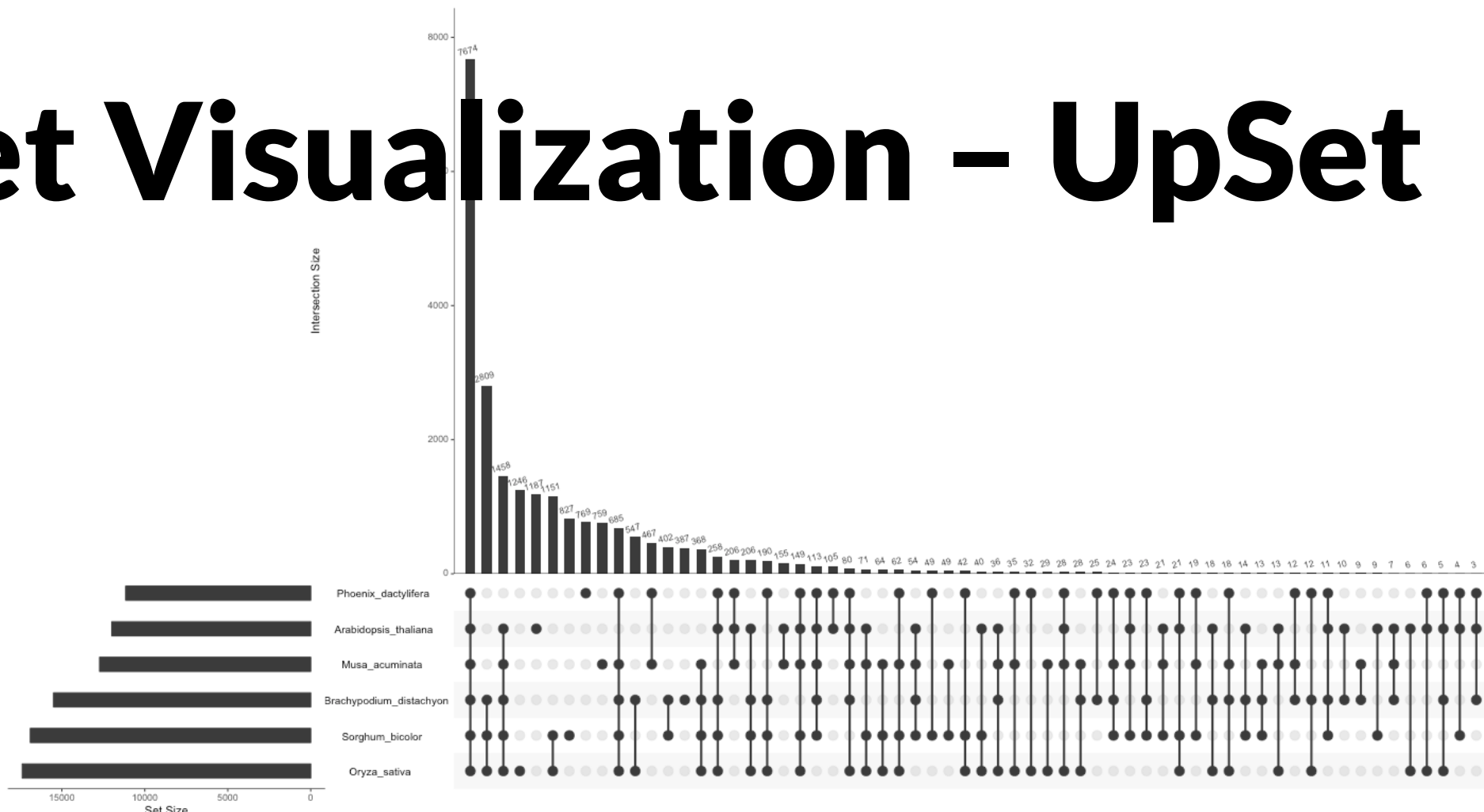


Multidimensional Data



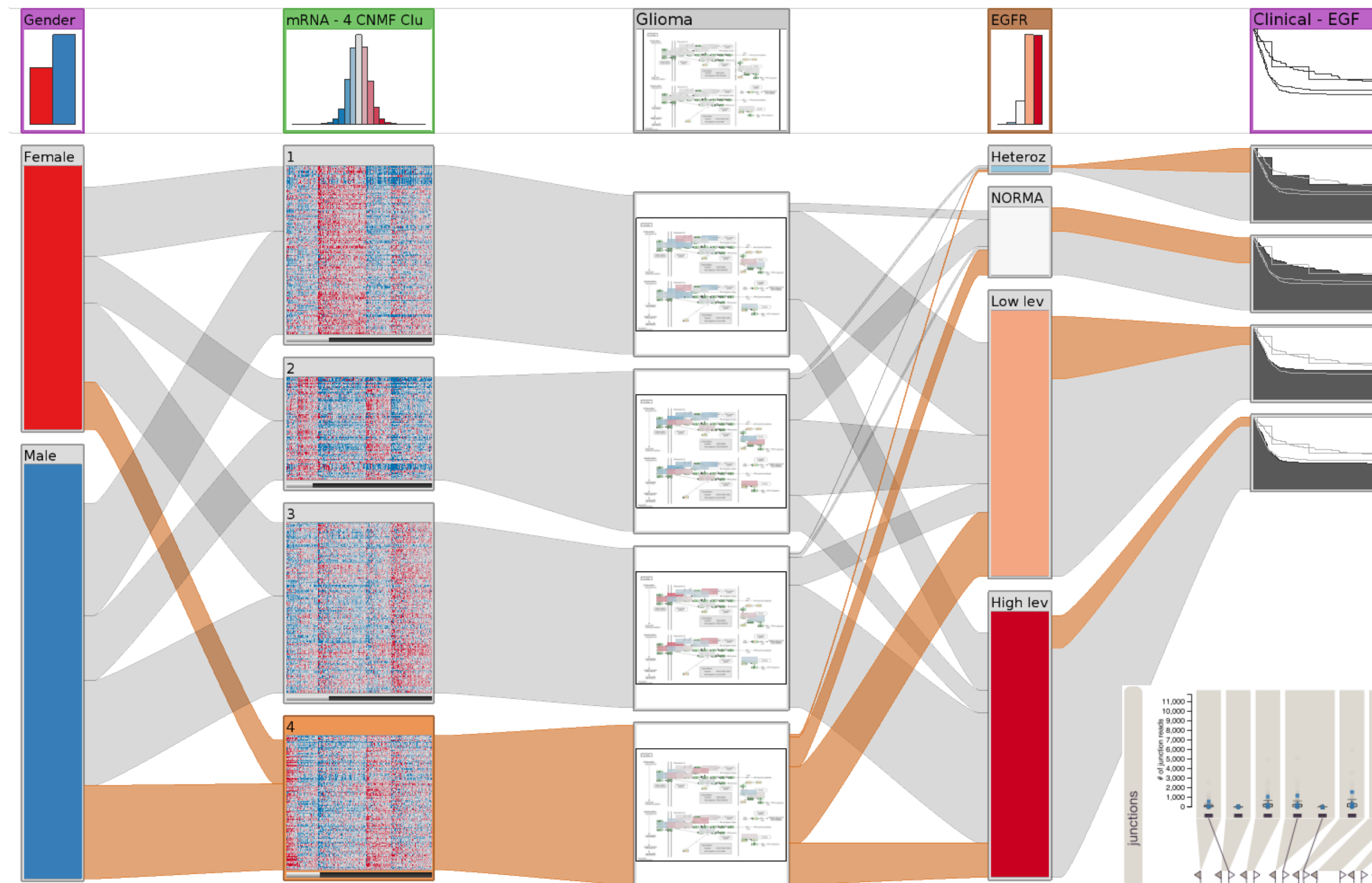
Multivariate Rankings - Lineup

Set Visualization - UpSet

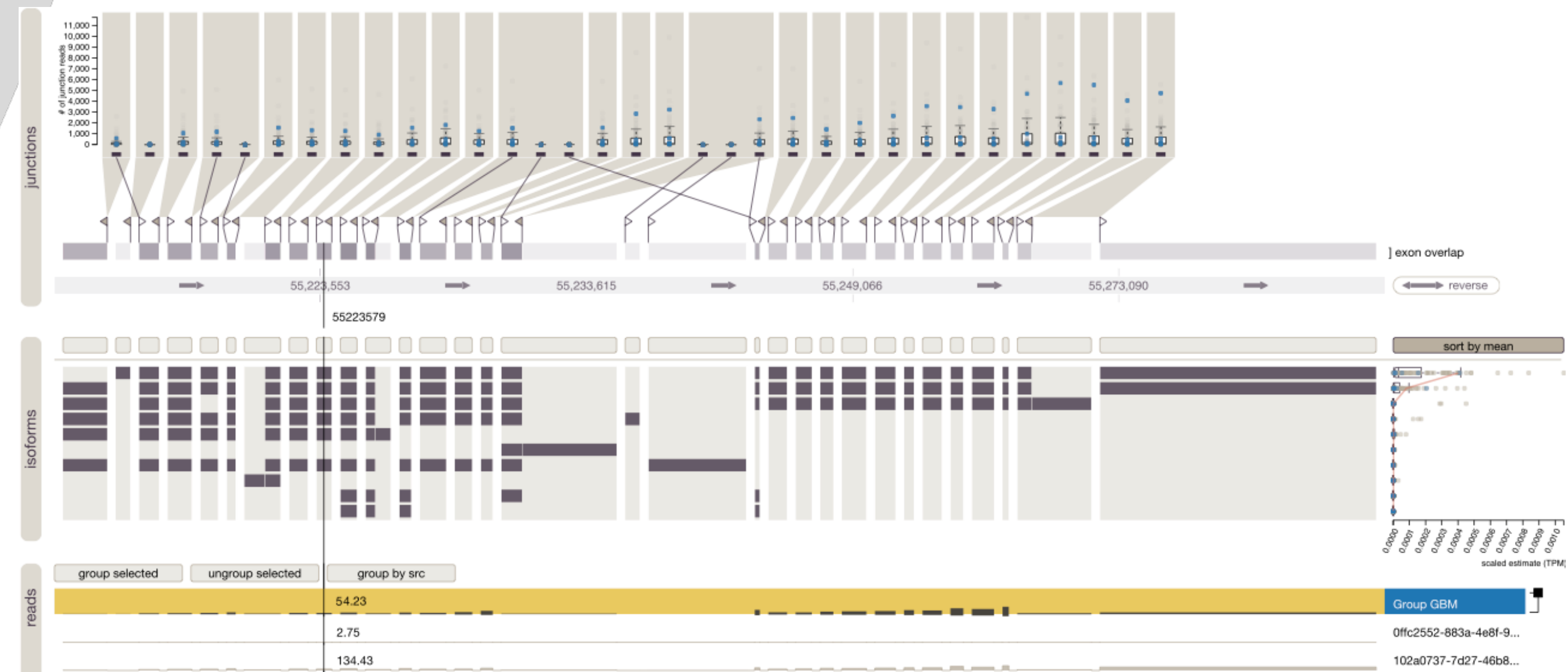


World University Ranking					World University Ranking										
Name	Release Date	Rank	School Name	Country	Acade	Employer repu	Faculty/ Citation	Separator	Rank	Academic reputation	Empl	Faculty/stud	Citations pe	In In	
Toy Story (1995)	1995				17.99%	32.94%	19.63%	19.63%	4.4		40.00%	10.00%	20.00%	20.00%	5.0 5.0
Sense and Sensibility (1995)	1995	1.	Massachusetts Insti	United States					1.						
Persuasion (1995)	1995	2.	Harvard University	United States					2.						
City of Lost Children, The (1995)	1995	3.	University of Camb	United Kingdom					3.						
Seven (Se7en) (1995)	1995	4.	Imperial College L	United Kingdom					4.						
		5.	University of Oxfor	United Kingdom					5.						
		6.	UCL (University Col	United Kingdom					6.						
		7.	Stanford University	United States					7.						
		8.	Yale University	United States					8.						
		9.	Princeton Universit	United States					9.						
		10.	University of Chica	United States					10.						
		11.	ETH Zurich (Swiss F	Switzerland					11.						
		12.	Columbia Universit	United States					12.						
		13.	University of Penns	United States					13.						
		14.	Cornell University	United States					14.						
		15.	University of Edinb	United Kingdom					15.						
		16.	Ecole Polytechniqu	Switzerland					16.						
		17.	King's College Lond	United Kingdom		93.7 (0.94)			17.						
		18.	University of Toron	Canada					18.						
		19.	McGill University	Canada					19.	94.6 (0.95)	89.9 (0.9)	79.3 (0.79)			
		20.	National University	Singapore					20.						
		21.	University of Michi	United States					21.						
		22.	University of Califo	United States					22.						
		23.	California Institute	United States					23.						
		24.	University of Bristol	United Kingdom					24.						
		25.	Duke University	United States					25.						

Genomic Data

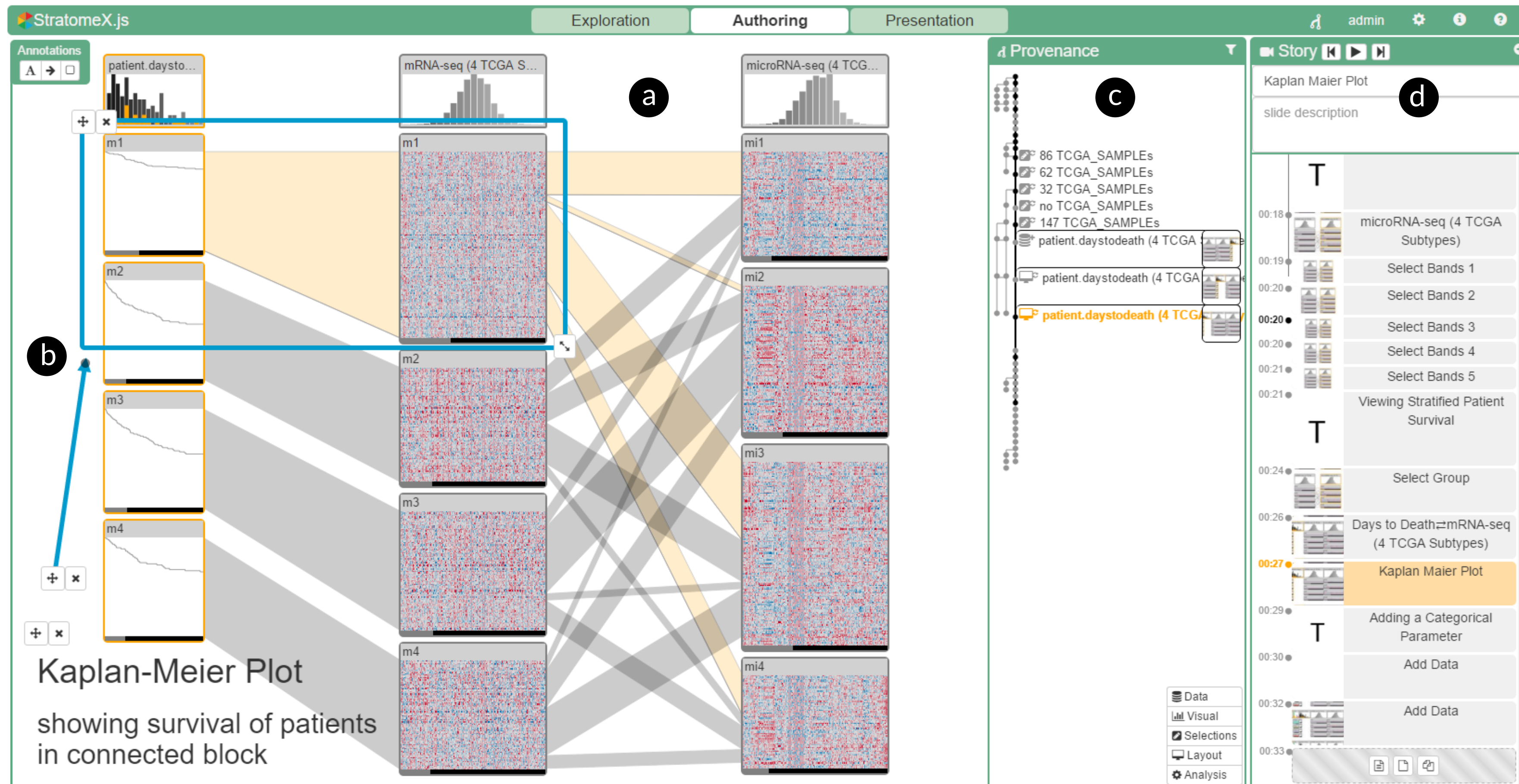


Alternative Splicing / mRNA-seq

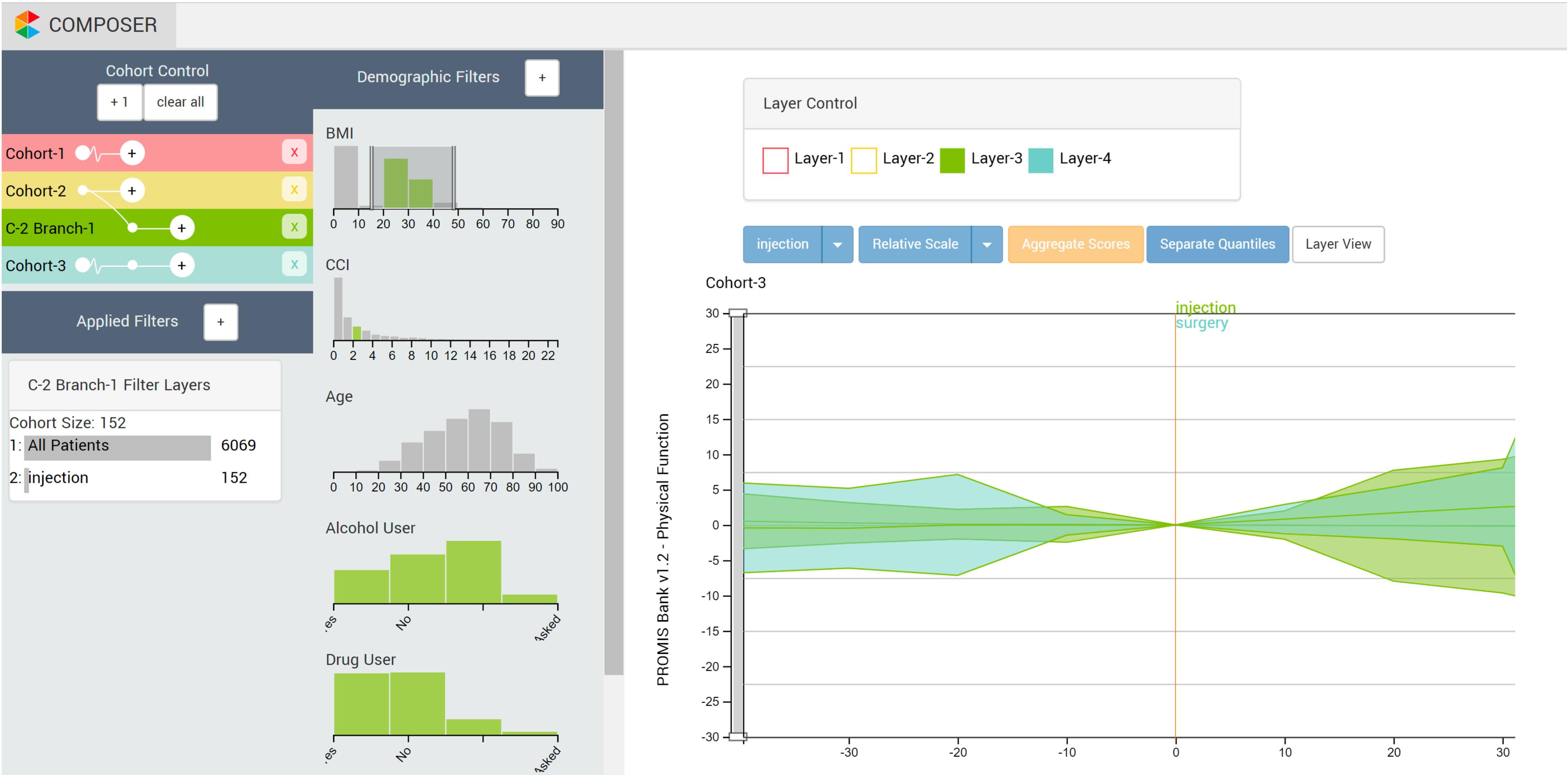


Cancer Subtypes / Omics Clustering and Stratification

Reproducibility, Storytelling, Annotation, and Integration in Computational Workflows



EHRs



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

Theory

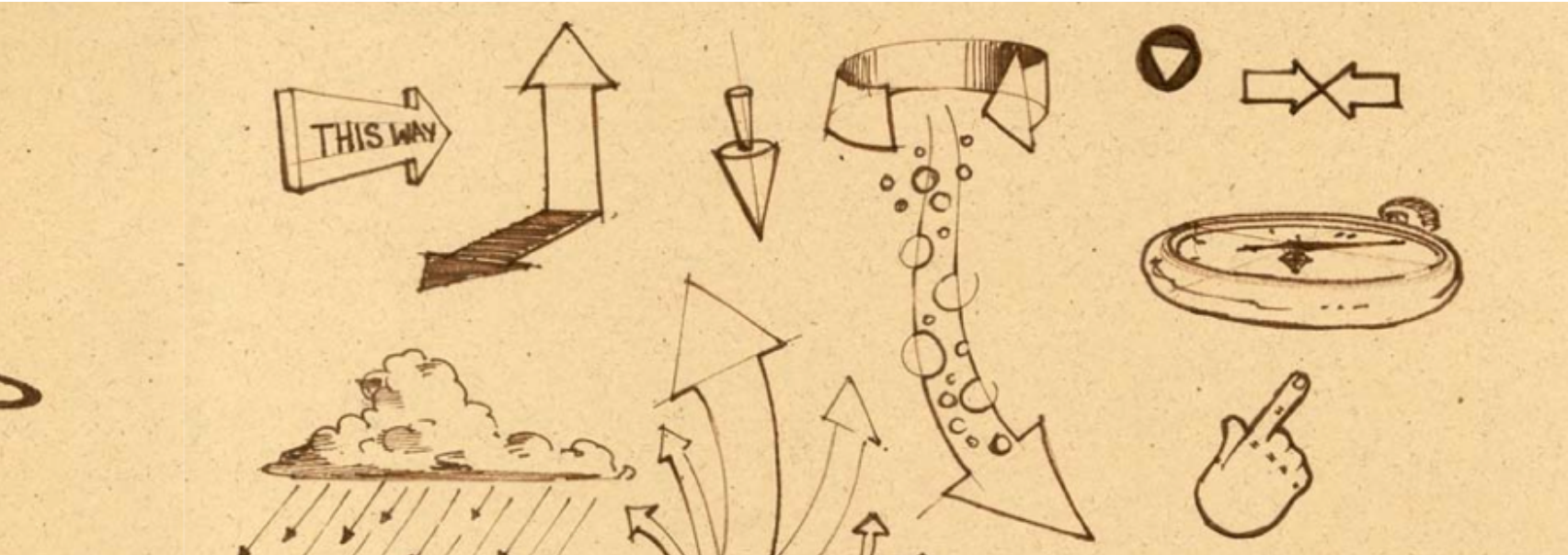
Lecture
Reading
Discussion

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: Tuesday and Thursday 12:25-1:45 pm, L103

WEB

Labs: Wednesday, 6:00-7:00 pm, L110 (scheduled on demand)

Review Lectures:
[YouTube Channel](#)

Three Parts:

I. Technical Foundations

HTML, Javascript, D3

II. Visualization Fundamentals

Perception, Visual encodings, Design Guidelines, Tasks..

III. Abstract Data Visualization

Tables, Graphs, Maps

Schedule

CS 5630/6630

Today September 2019

Print Week Month Agenda

Mon	Tue	Wed	Thu	Fri	Sat	Sun
26	27 12:25 Vis Lecture 14:00 Alex Lex Office	28 16:30 Ilkin's Office H 18:00 HW 1 Introduc	29 12:25 Vis Lecture 17:00 Kiran's Office	30 HW1 Due	31	Sep 1
2	3 12:25 Vis Lecture 14:00 Alex Lex Office	4 16:30 Ilkin's Office H	5 12:25 Vis Lecture 17:00 Kiran's Office	6 HW2 Due	7	8
9	10 12:25 Vis Lecture 14:00 Alex Lex Office	11 16:30 Ilkin's Office H	12 12:25 Vis Lecture 17:00 Kiran's Office	13 HW3 Due	14	15
16	17 12:25 Vis Lecture 14:00 Alex Lex Office	18 16:30 Ilkin's Office H	19 12:25 Vis Lecture 17:00 Kiran's Office	20 HW4 Due	21	22
23	24 12:25 Vis Lecture 14:00 Alex Lex Office	25 16:30 Ilkin's Office H	26 12:25 Vis Lecture 17:00 Kiran's Office	27	28	29
30	Oct 1 12:25 Vis Lecture 14:00 Alex Lex Office	2 16:30 Ilkin's Office H	3 Exam 1 12:25 Vis Lecture 17:00 Kiran's Office	4	5	6 Fall break

Events shown in time zone: Mountain Time - Denver

Subject to change

Week 1

Lecture 1: Introduction

Tuesday, August 20

What is visualization? Why is it important? Who are we? Course overview.

Recommended reading

- [A Tour through the Visualization Zoo](#). Jeffrey Heer, Michael Bostock, Vadim Ogievetsky. Communications of the ACM, 53(6), pp. 59-67, Jun 2010.
- [The Value of Visualization](#). Jarke van Wijk. Proceedings of the IEEE Visualization Conference, pp. 79-86, 2005.

Information <http://dataviscourse.net>

Visualization
for Data Science
CS-5630 / CS-6630



[Home](#) [Syllabus](#) [Schedule](#) [Project](#) [Resources](#) [Fame](#)



UpSet visualizing intersecting sets | 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. Visualization for data discovery and communication is an important part of the data science pipeline. 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 about the fundamentals of perception, the theory of visualization, good design practices for visualization, and how to develop your own web-based visualizations using HTML5, CSS, JavaScript, SVG, and **D3**.

The course begins by bootstrapping your web development skills, moves on to fundamentals of perception, introduces data types you will encounter, and then focuses on visualization techniques and methods for a broad range of data types. An integral component of the course are regular design critiques and redesigns that will hone your skills in understanding, critiquing and developing visualization techniques.

The course is offered in the fall term 2017 at the University of Utah in two variants: **CS-5630** for undergraduates and **CS-6630** for graduate students, with a special section of CS-6630 (002) designated for data certificate students. Classes start on Tuesday,

Communicate

Slack

<http://dataviscourse2019.slack.com/>

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

Please don't use personal messages to me or TAs

Only use e-mail for personal inquiries

Canvas

<https://utah.instructure.com/courses/574005>

Homework submissions, Grades

Office Hours

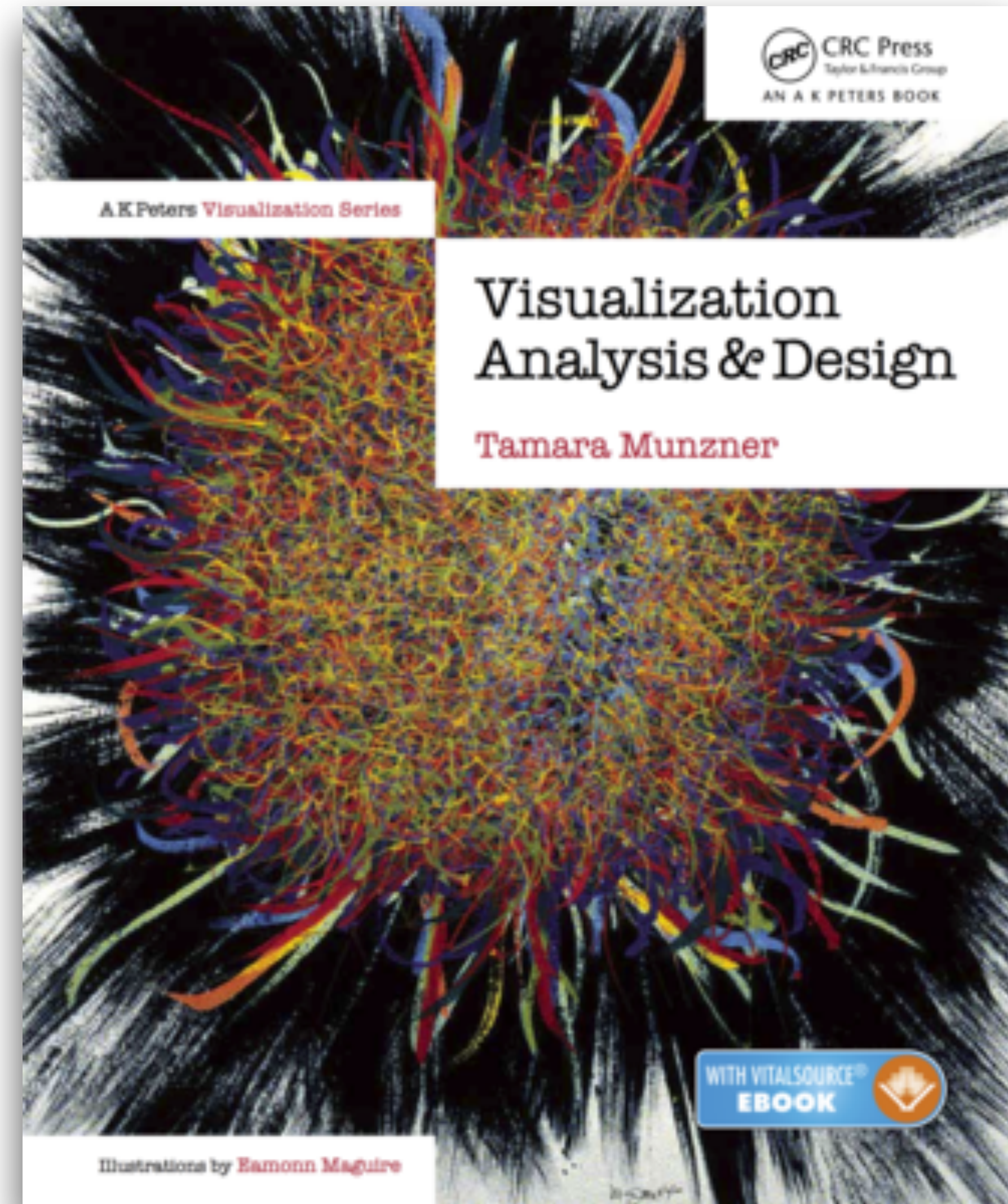
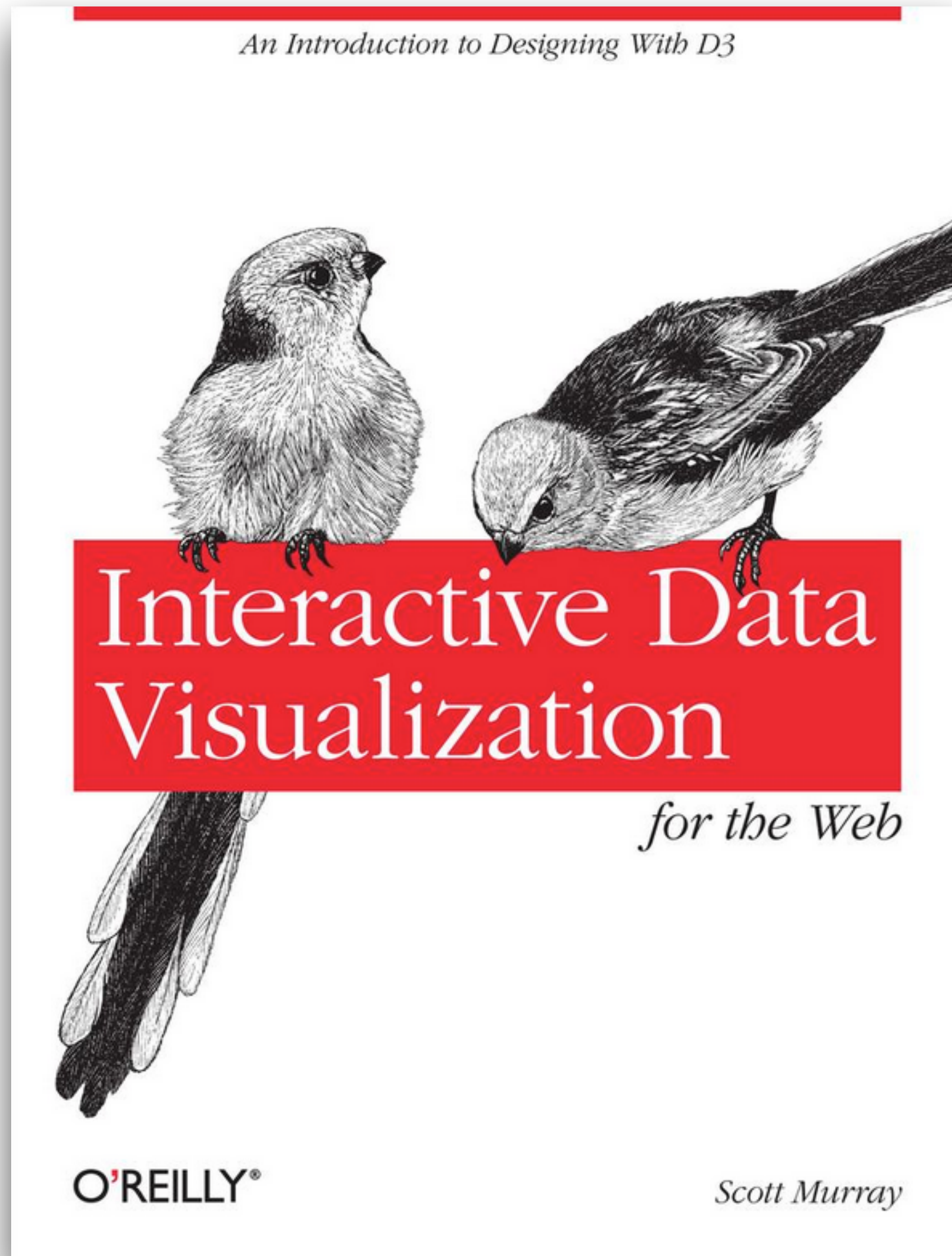
Alex: Tuesdays after Class, WEB 3887

TAs: starting next week

E-Mail

alex@sci.utah.edu

Required Books



Programming

HTML



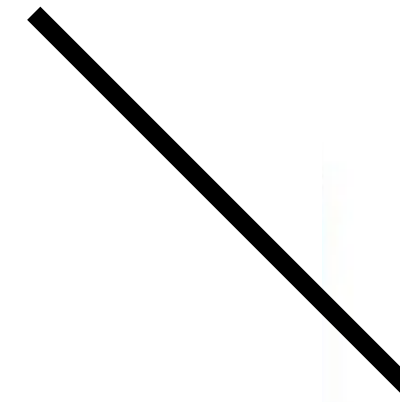
JS



Data-Driven Documents



Is this course for me ???



Prerequisites

Programming experience

C, C++, Java, Python, etc.

Willingness to think about user-centered design

This is not your average CS course! We care about the human in the loop!

Willingness to learn new software & tools

This can be time consuming

You will need to build skills by yourself!

Formalities

How are you graded?

6 Homework Assignments: 40%

Varying value, 2%-10%, depending on length/difficult

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

Due on Fridays, late days: -10% per day, up to two days.

Final Project: 40%

Teams, proposal and two milestones

Exams: 20%

Two exams: last class before fall break and end of term

Code of Conduct

- We are committed to providing an inclusive and harassment-free environment in all interactions regardless of gender, sexual orientation, disability, physical appearance, race, or religion.
- We do not tolerate harassment in any form.
- Please report any harassment to me or the appropriate university office, which you can find at <https://safeu.utah.edu/>
- Please review the syllabus on these issues and the student code of conduct at <https://regulations.utah.edu/academics/6-400.php>

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.

See also the SoC Academic Misconduct Policy:

http://www.cs.utah.edu/wp-content/uploads/2014/12/cheating_policy.pdf

You will fail the class if you cheat.

A "strike" will be recorded.

We will **automatically check for plagiarism** in all your submissions.

No Device Policy

No Computers, Tablets, Phones in lecture hall

except when used for exercises

Switch off, mute, flight mode

Why?

It's better to take notes by hand

Notifications are designed to grab your attention

*Applies to **theory lectures**, coding along in technical lectures encouraged*

This Week

HW0, including course survey

Lecture on Perception

Readings

D3 Book, Chapters 1-3

VDA Book, Chapter 1

Table of Contents	
Preface.....	ix
1. Introduction.....	1
Why Data Visualization?	1
Why Write Code?	2
Why Interactive?	2
Why on the Web?	3
What This Book Is	3
Who You Are	4
What This Book Is Not	5
Using Sample Code	5
Thank You	6
2. Introducing D3.....	7
What It Does	7
What It Doesn't Do	8
Origins and Context	9
Alternatives	10
Easy Charts	10
Graph Visualizations	12
Geomapping	12
Almost from Scratch	13
Three-Dimensional	13
Tools Built with D3	14
3. Technology Fundamentals.....	15
The Web	15
HTML	17
Content Plus Structure	18

Next Week

HW1 due

Introduction to Git, HTML, CSS

Office hours start!

<https://github.com/dataviscourse/2019-dataviscourse-homework>

The screenshot shows the GitHub repository page for 'dataviscourse / 2019-dataviscourse-homework'. At the top, there are navigation links for 'Code', 'Issues 0', 'Pull requests 0', 'Projects 0', 'Wiki', 'Security', 'Insights', and 'Settings'. On the right, there are buttons for 'Watch 0', 'Star 0', and 'Fork 0'. Below the navigation is a section for repository statistics: '2 commits', '1 branch', '0 releases', '1 contributor', and 'MIT' license. There are also buttons for 'Branch: master', 'New pull request', 'Create new file', 'Upload files', 'Find File', and 'Clone or download'. A commit history table shows a commit by 'alexsb' updating the README. Below this is a file list with 'hw0', 'LICENSE', and 'README.md'. The main content area shows the 'README.md' file with the following text:

Homeworks for Utah's Vis for Data Science Course

In subfolders in this directory you will find the homeworks for CS 6630 / CS 5630 - Visualization for Datascience.

More information can be found on the [course website](#)

We suggest you clone this repository:

```
git clone https://github.com/dataviscourse/2018-dataviscourse-homework
```

To receive updates and add newly released homeworks update your repository by cd-ing into the `2019-dataviscourse-homework` directory and running:

```
git pull
```

Submitting Homeworks

Please submit homeworks using the appropriate canvas dropbox.

Newish Track: Human Centered Computing

REQUIRED COURSES

CS 6540 - HCI (humans + interfaces)

CS 6963 - Advanced HCI (humans + things)

CS 6630 - Visualization for Data Science (humans + data)

ED PS 6010: Introduction to Stats and Research Design (methods)

ELECTIVES

Pre-approved course list from within CS and across campus

Up to 3 electives can be taken from outside CS

NON-CS COURSES

Design

DES 5320 - Typographic Communication

DES 5370 - Digital Fabrication

DES 5710 - Product Design and Development

Ed Psych

ED PSY 6030 - Introduction to Research Design

Psych

PSY 6120 - Advanced Human Cognition

PSY 6140 - Cognitive Neuroscience Approaches to Research

PSY 6420 - Methods in Social Psychology

PSY 6700 - Neuropsychology

Anthropology

ANTH 6169 - Ethnographic Methods

Sociology

SOC 6110 - Methods of Social Research

EAE

EAE 6900 - Games User Research

EAE 6900 - A.I. For Games

Companion Course: Visualization for Scientific Data



CS 5635 / CS 6635

Valerio Pascucci

Spring 2020

