

Lecture 1: Introduction

DS 4200

FALL 2020

Prof. Cody Dunne

NORTHEASTERN UNIVERSITY

Slides and inspiration from Michelle Borkin, Krzysztof Gajos, Hanspeter Pfister, Miriah Meyer, Jonathan Schwabish, and David Sprague

STAFF INTRODUCTIONS



Cody Dunne
Assistant Professor
Instructor



Megha Rao
TA



David Saffo, CS PhD
TA (Primary)

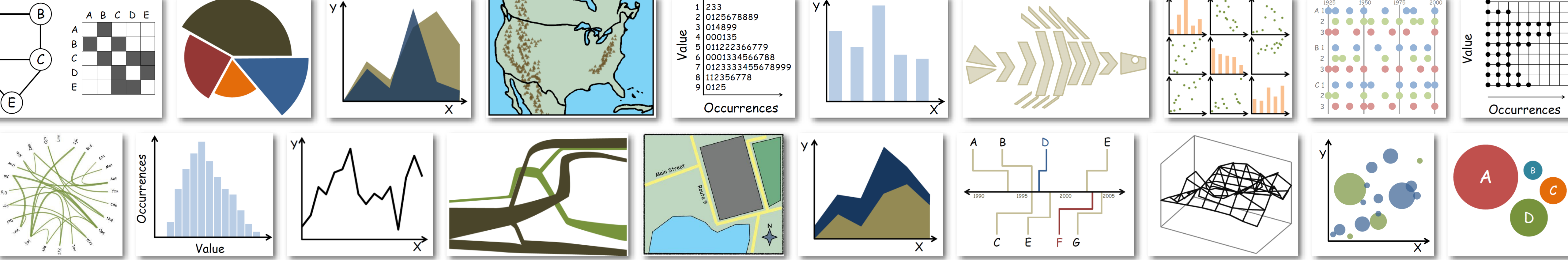


Eduardo Massud
IA

Course Homepage on Canvas

<https://c.dunne.dev/ds4200f20>

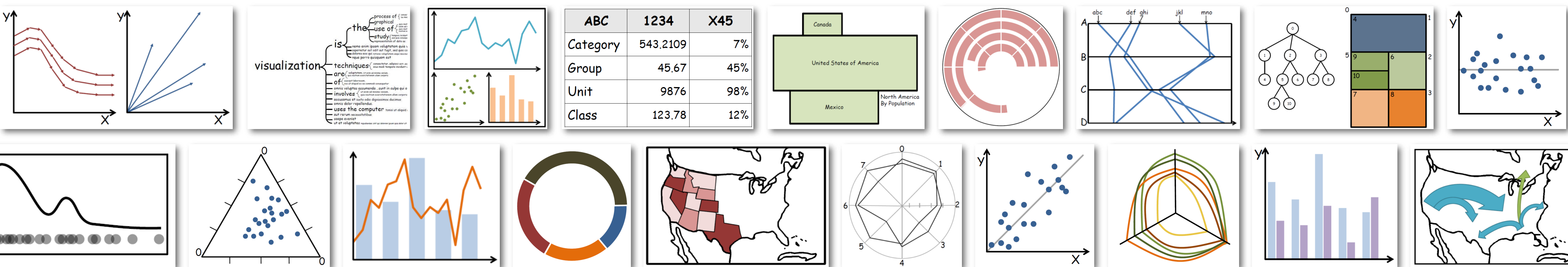
What is visualization
anyway?



(static or interactive)

(abstract or spatial)

visualization: the visual representation of data to reinforce human cognition

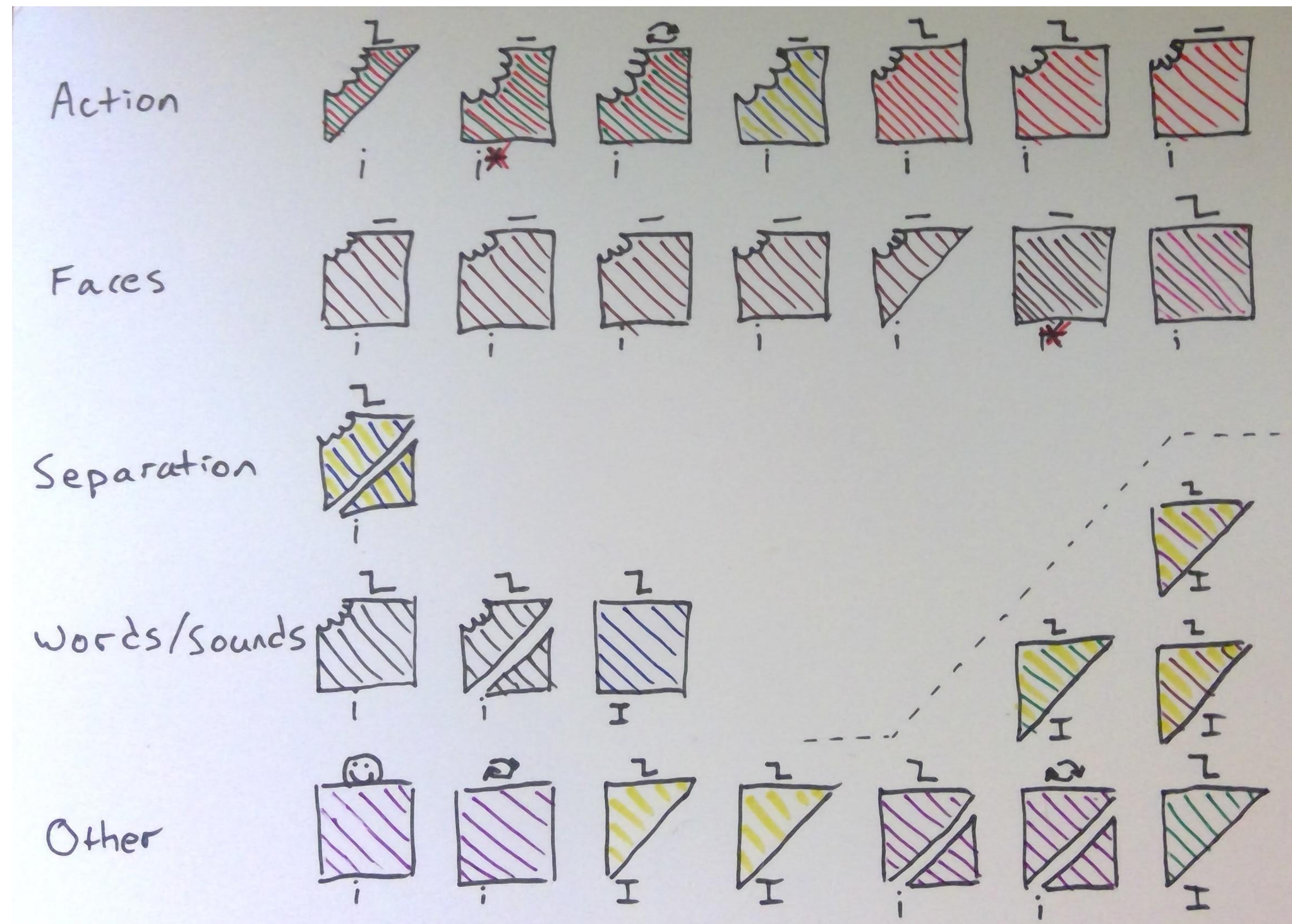


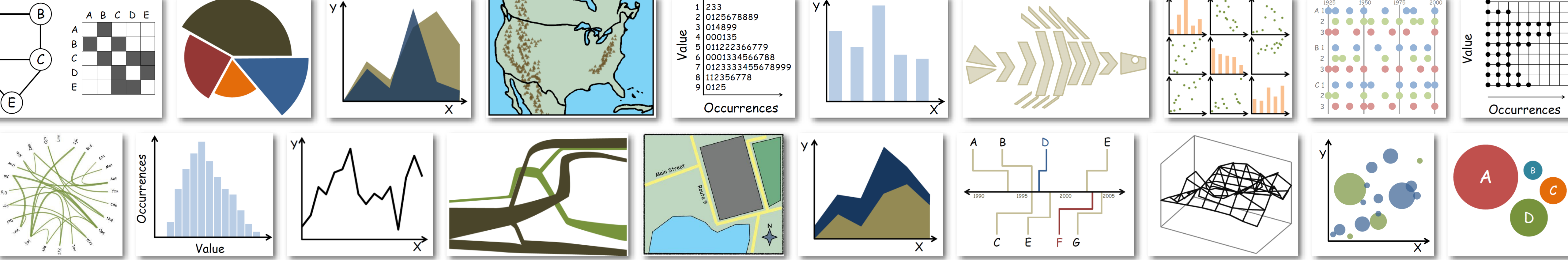
IN-CLASS EXERCISE

In-Class Sketching — Visualize Me

15 min

<https://northeastern.instructure.com/courses/18721/assignments/607608>

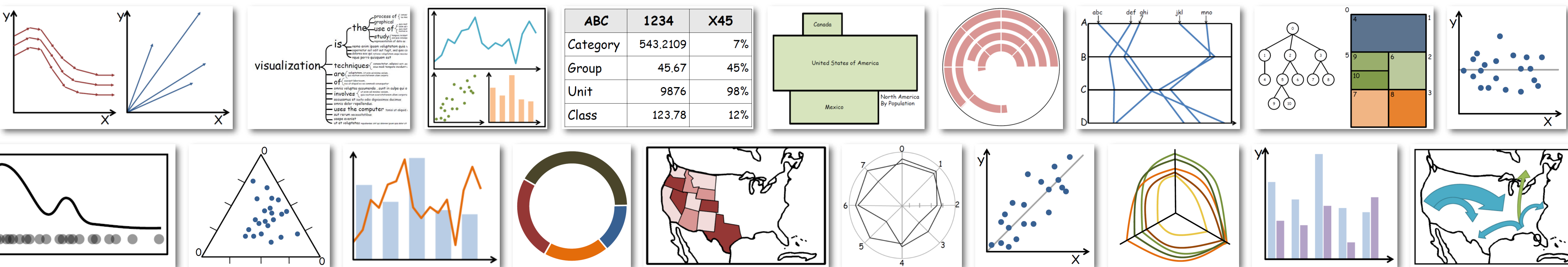




(static or interactive)

(abstract or spatial)

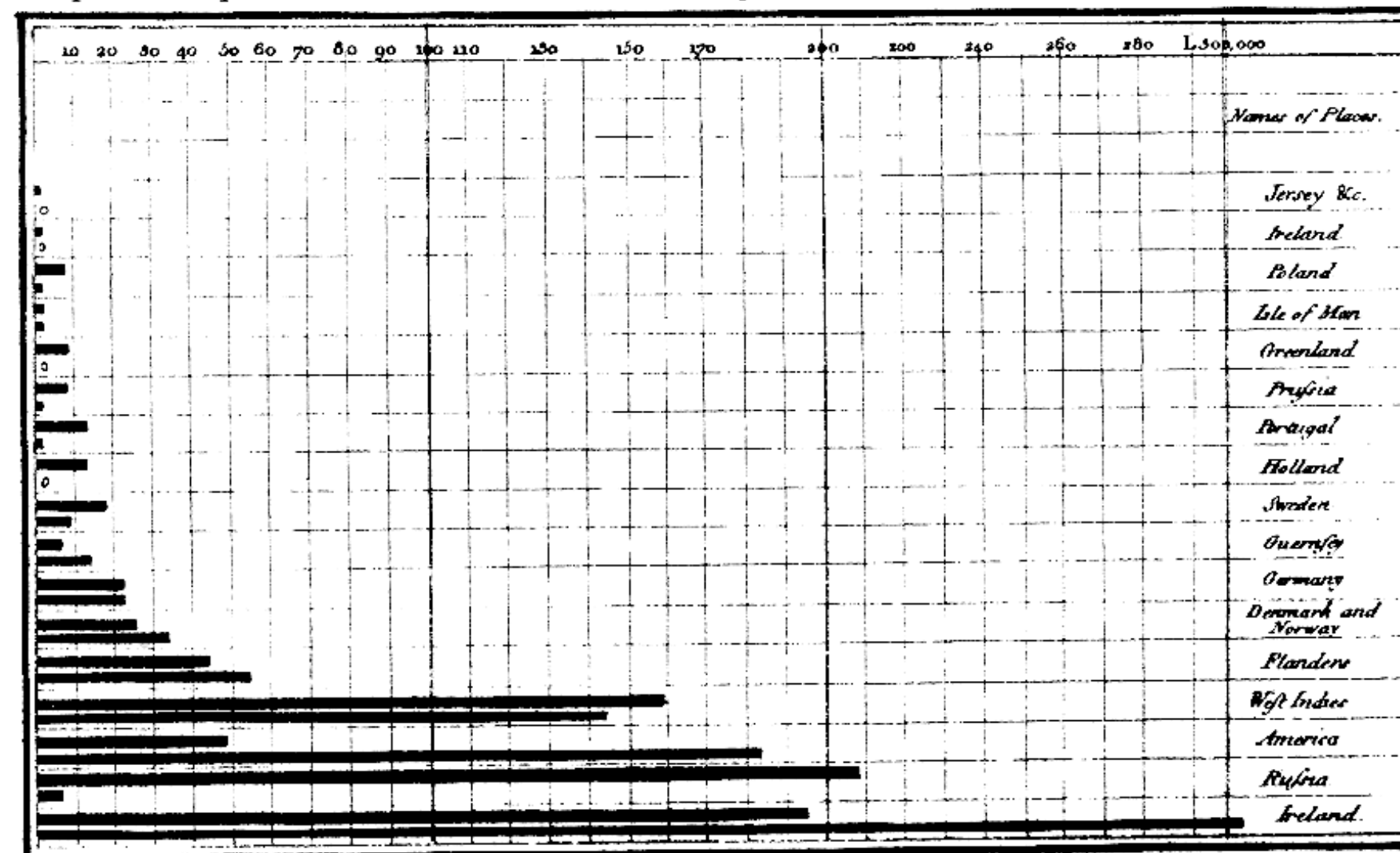
visualization: the visual representation of data to reinforce human cognition





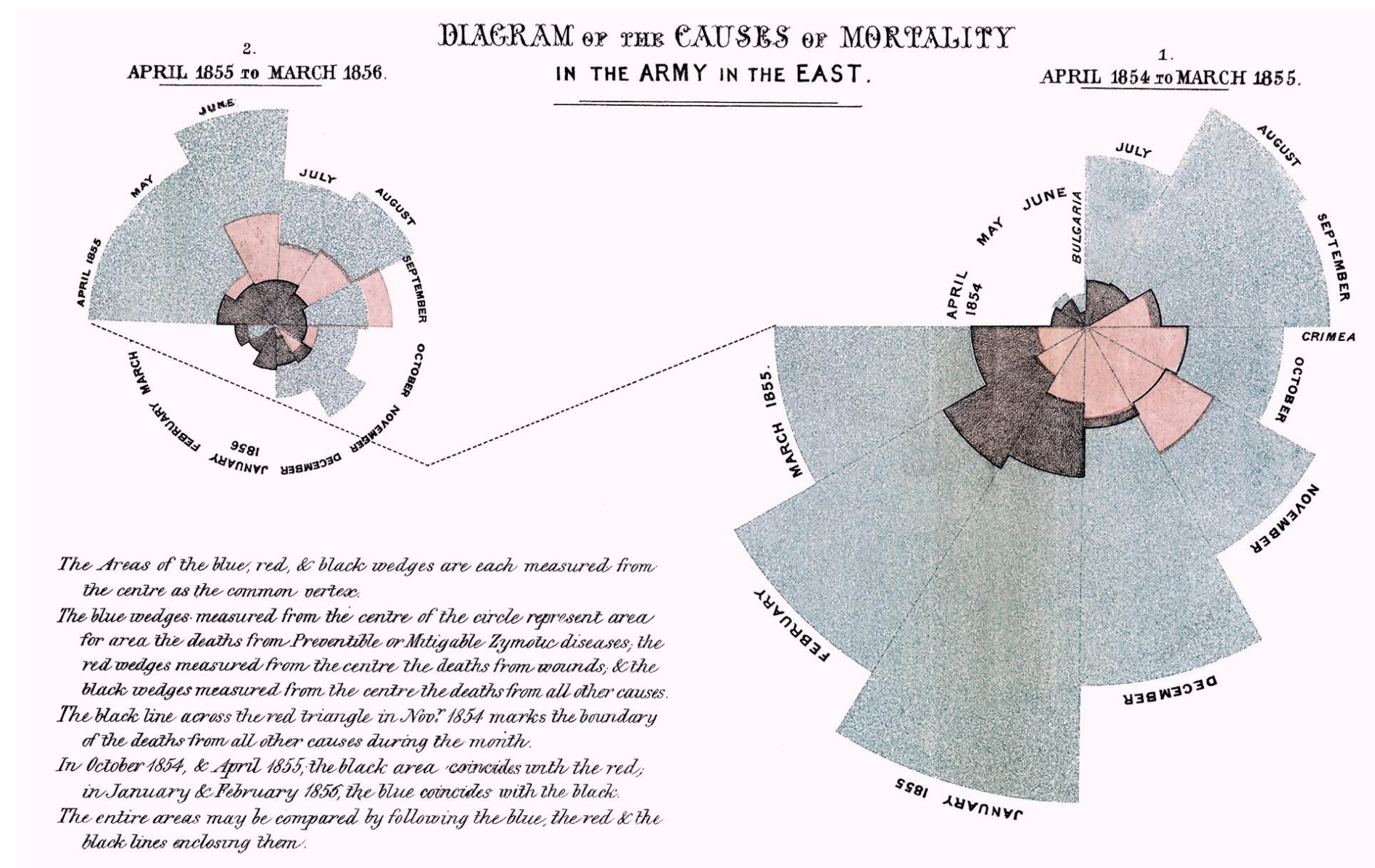
PTOLEMY (c. 150)

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

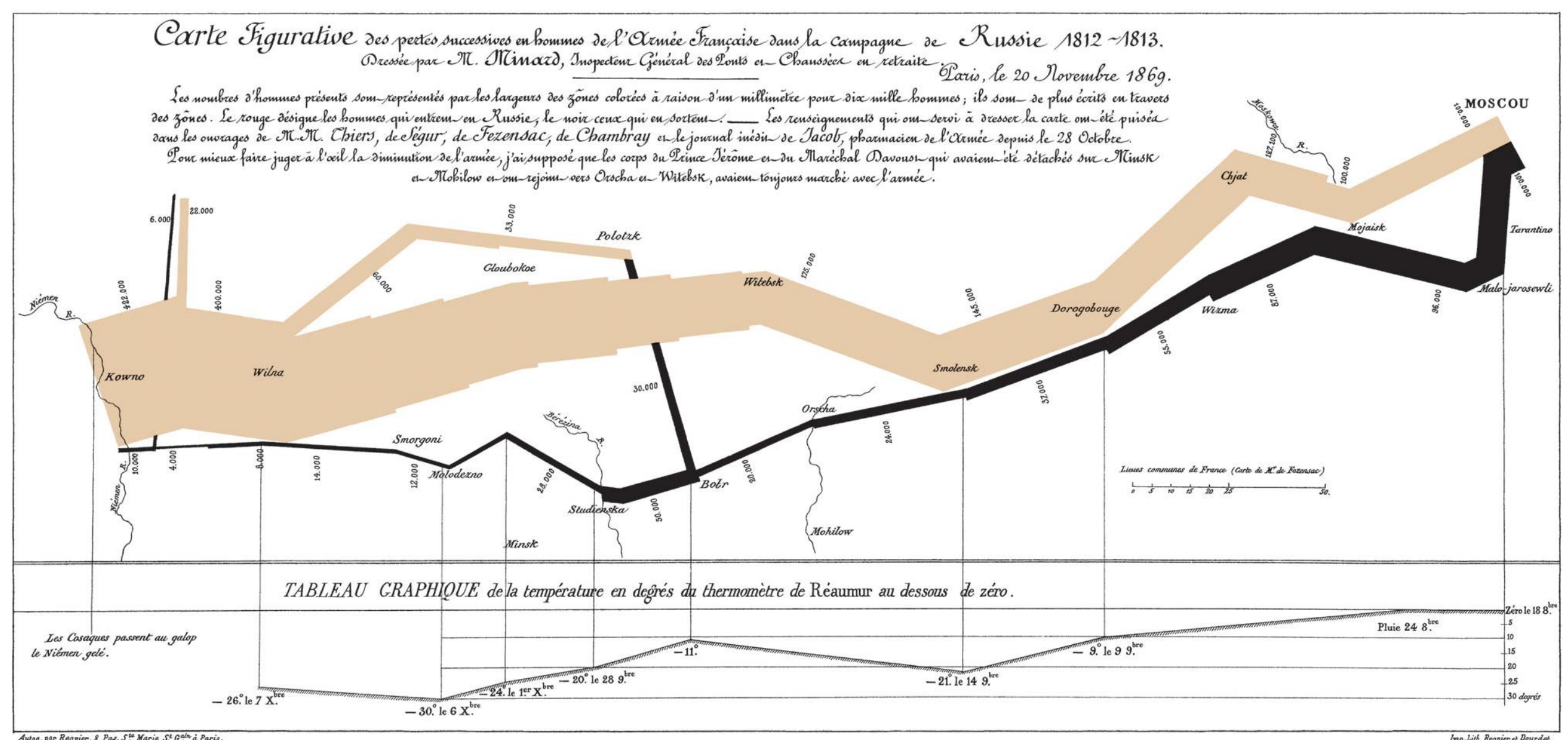


The Upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed Lines Imports.
 Published as the Act above June 7th 1788 by W^m Playfair
 Made comp^d 352. Strand. London.

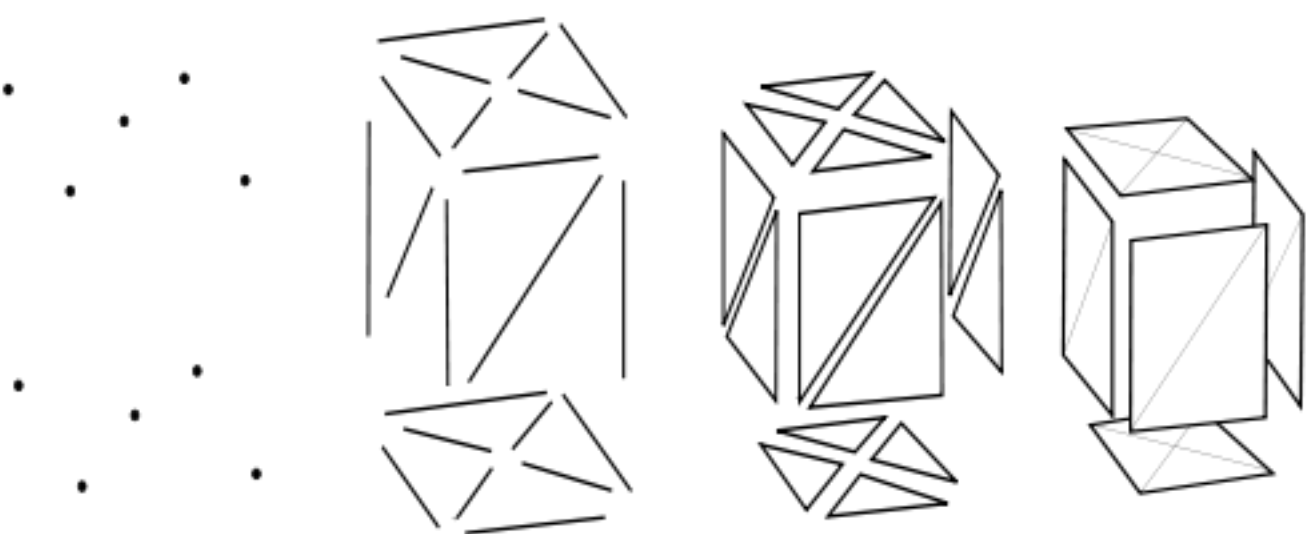
WILLIAM PLAYFAIR (c. 1786)



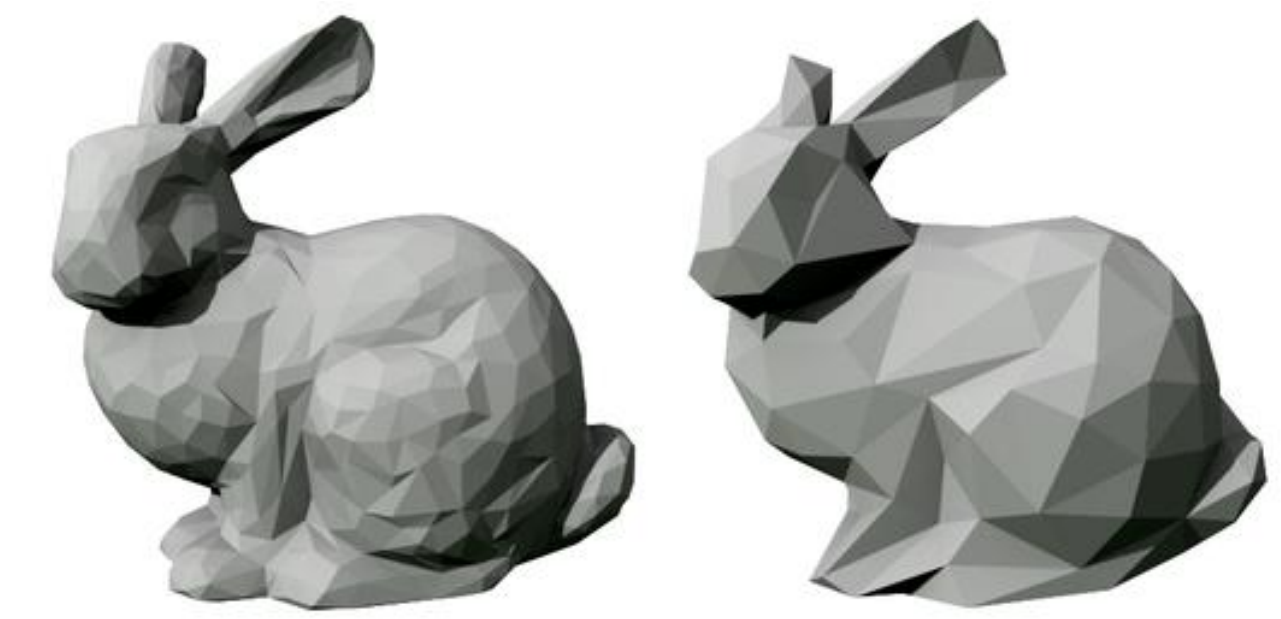
FLORENCE NIGHTINGALE (c. 1858)



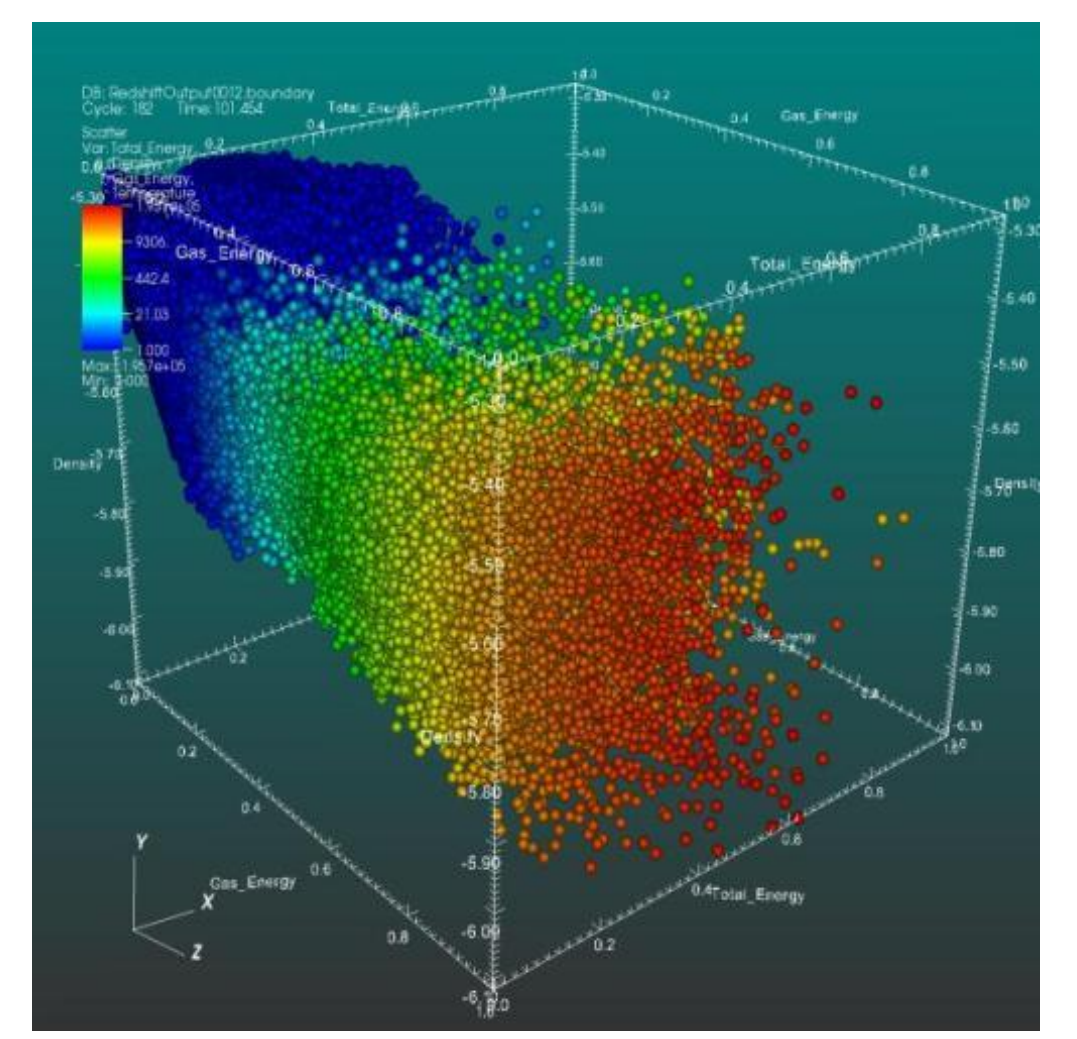
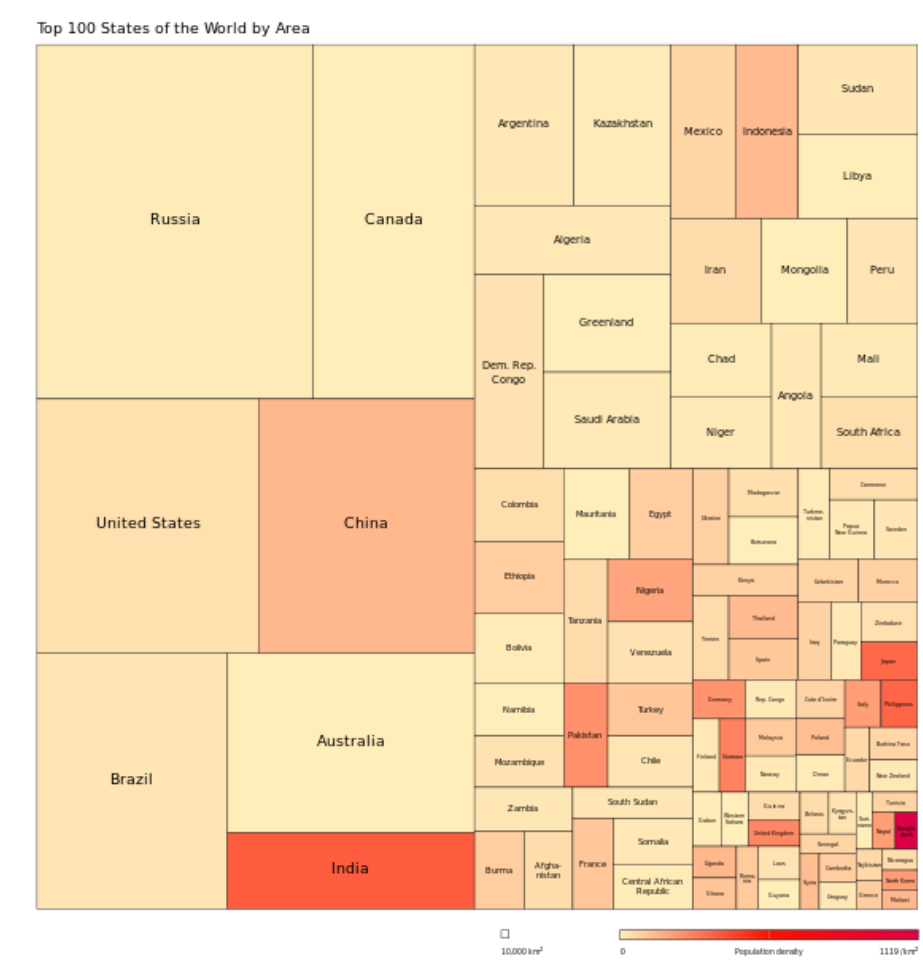
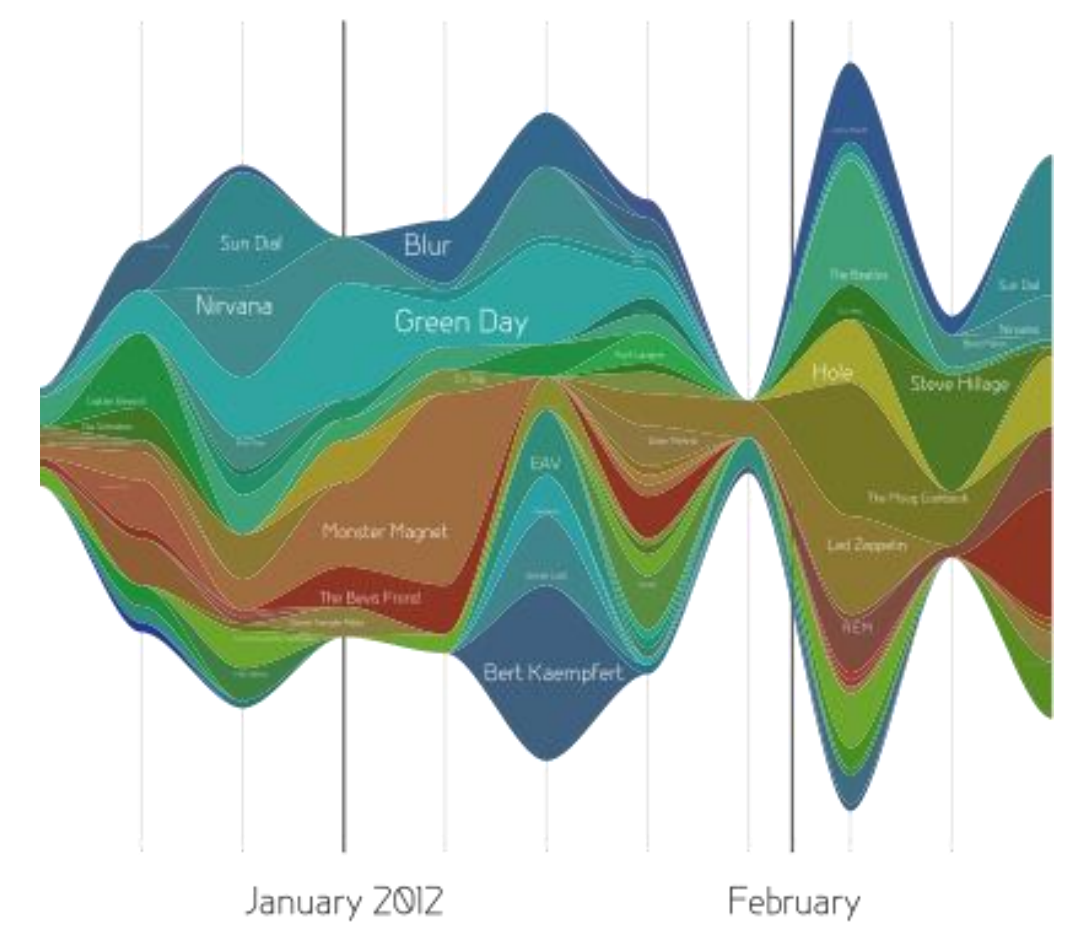
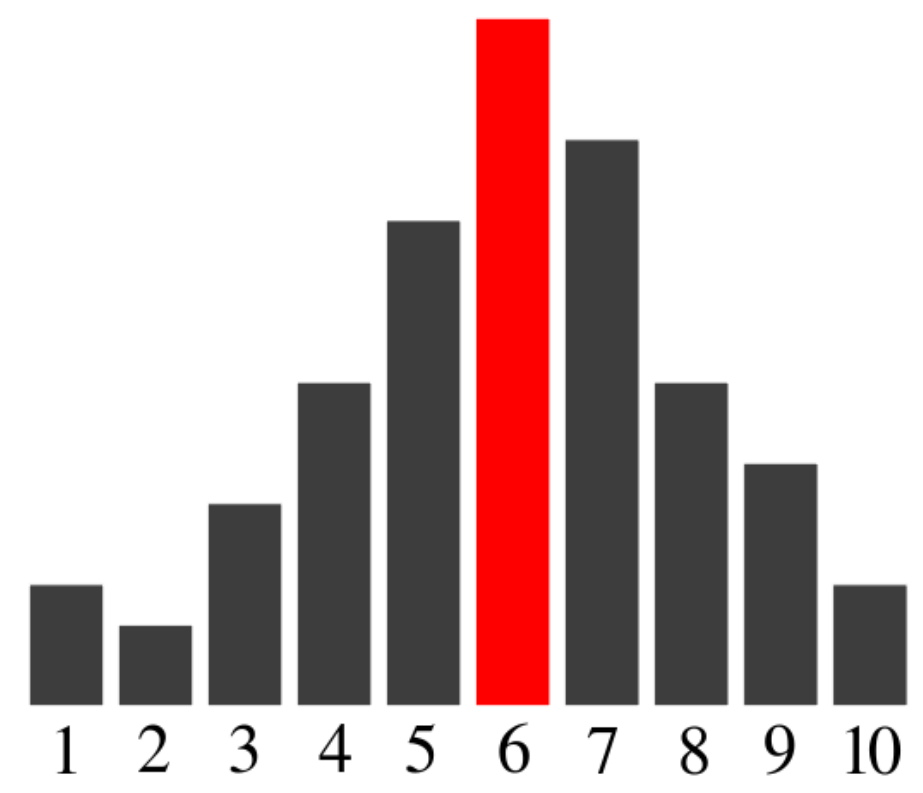
CHARLES MINARD (c. 1869)

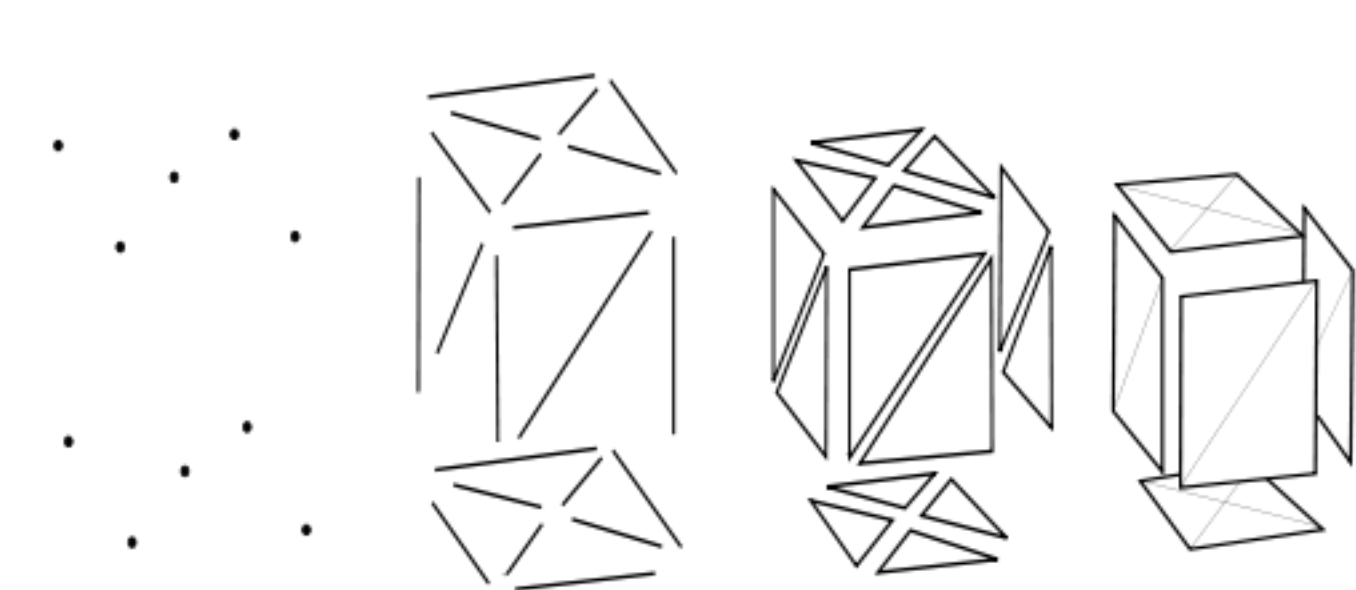


computer graphics

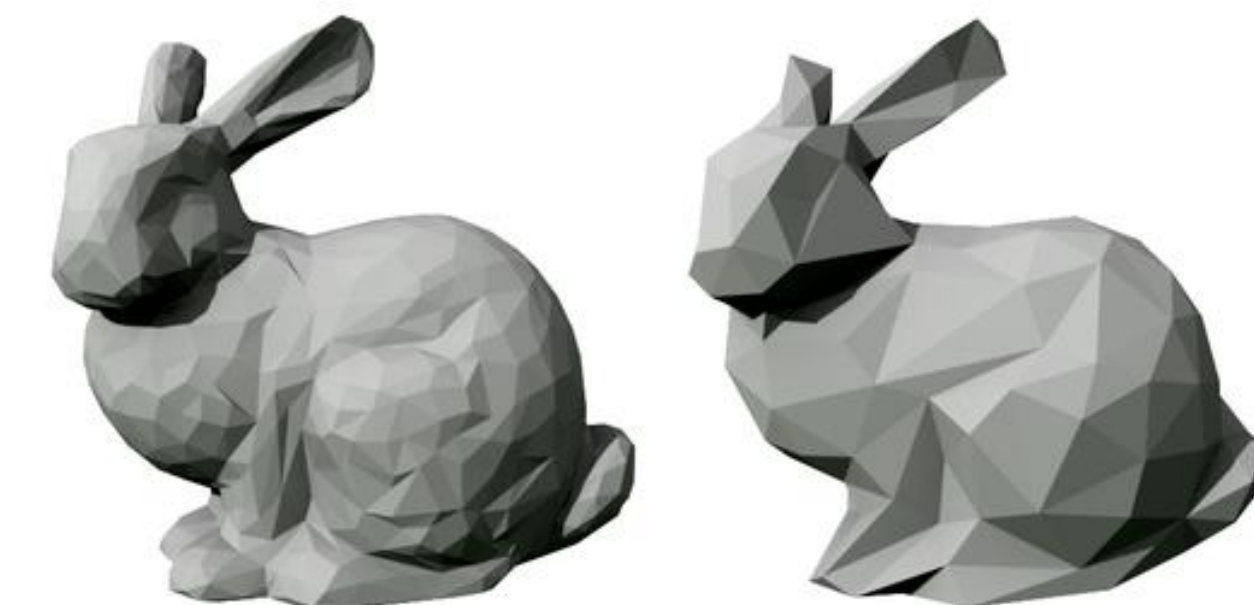


visualization

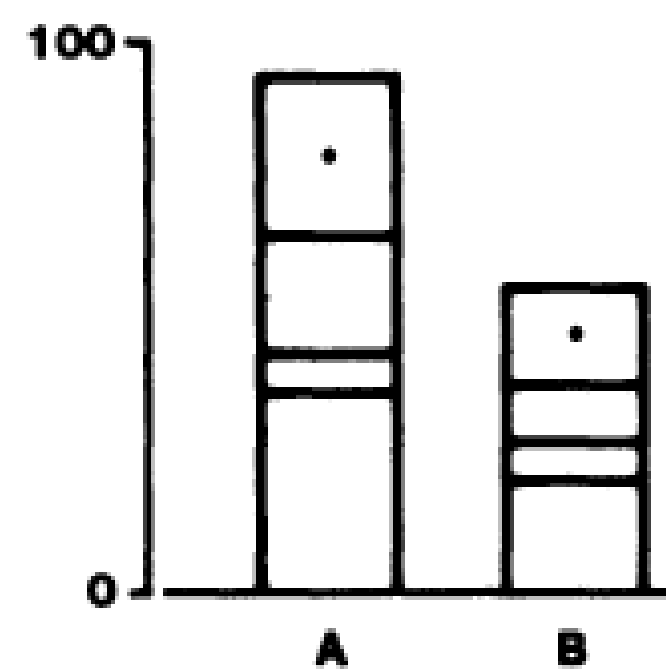
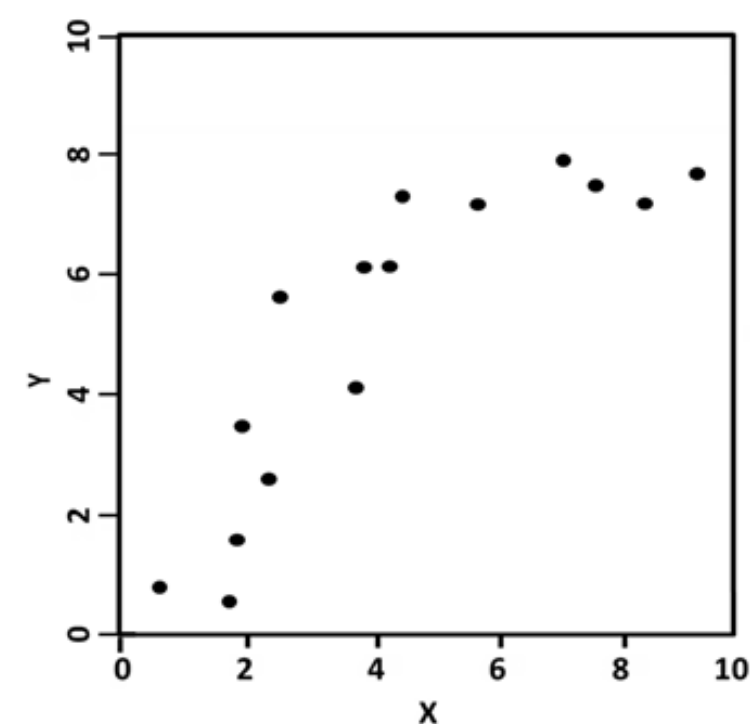




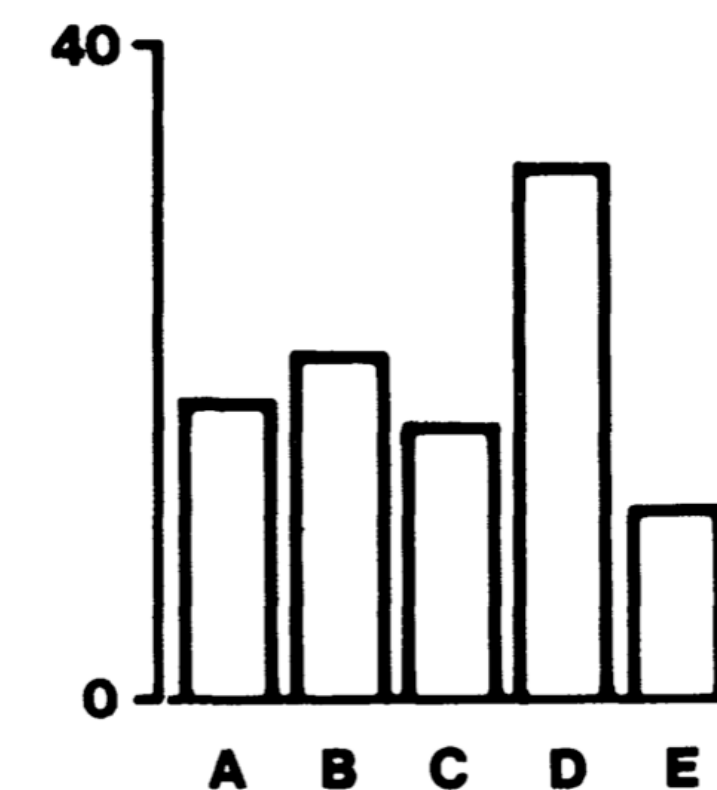
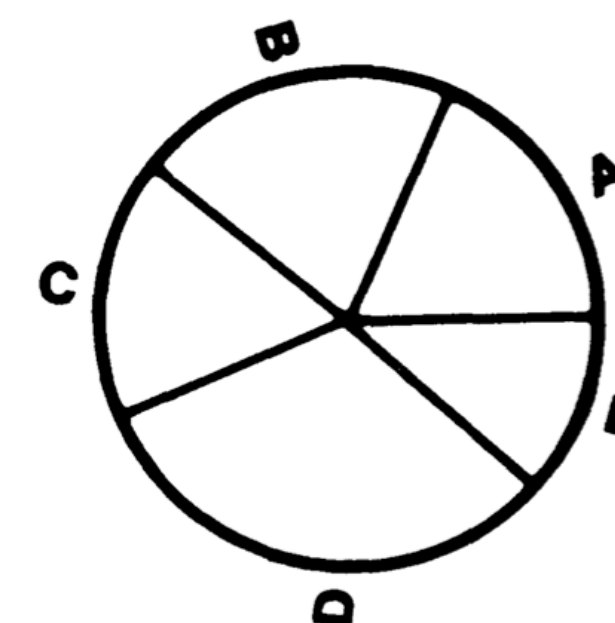
computer graphics

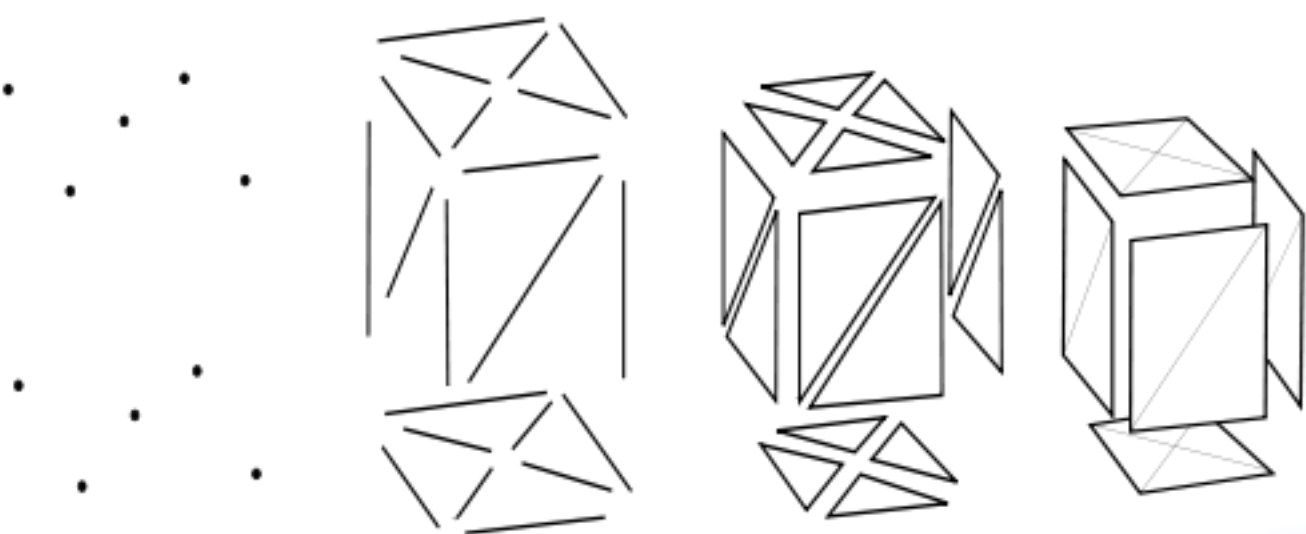


visualization

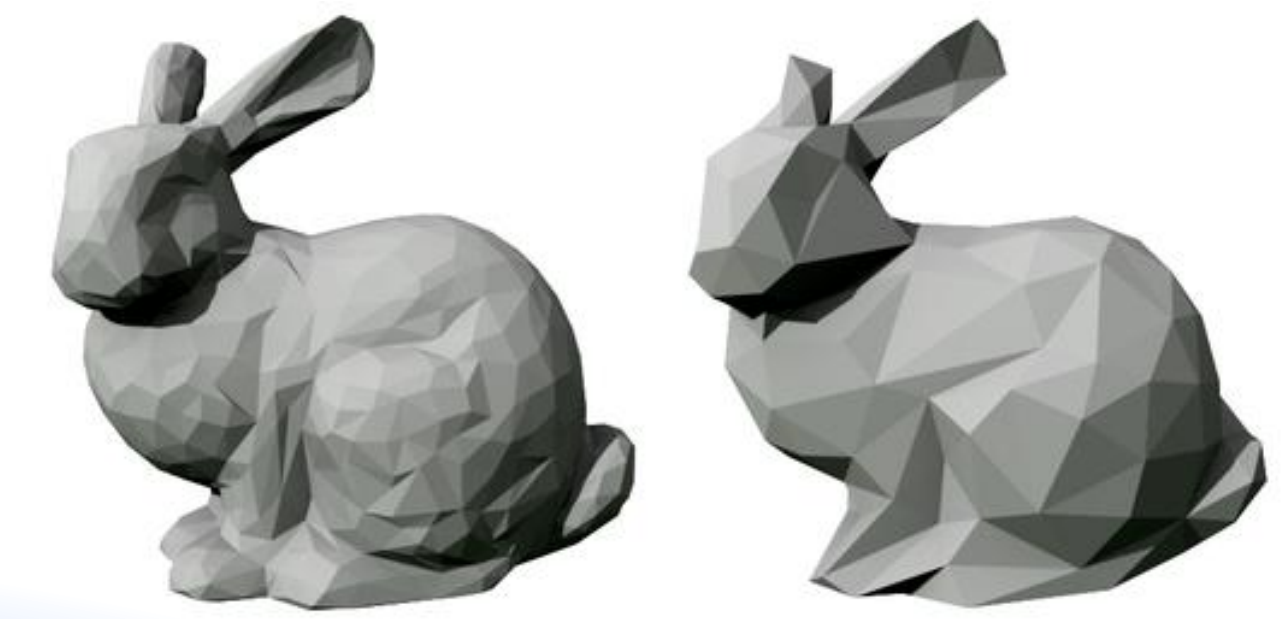


statistics





computer graphics



HCI

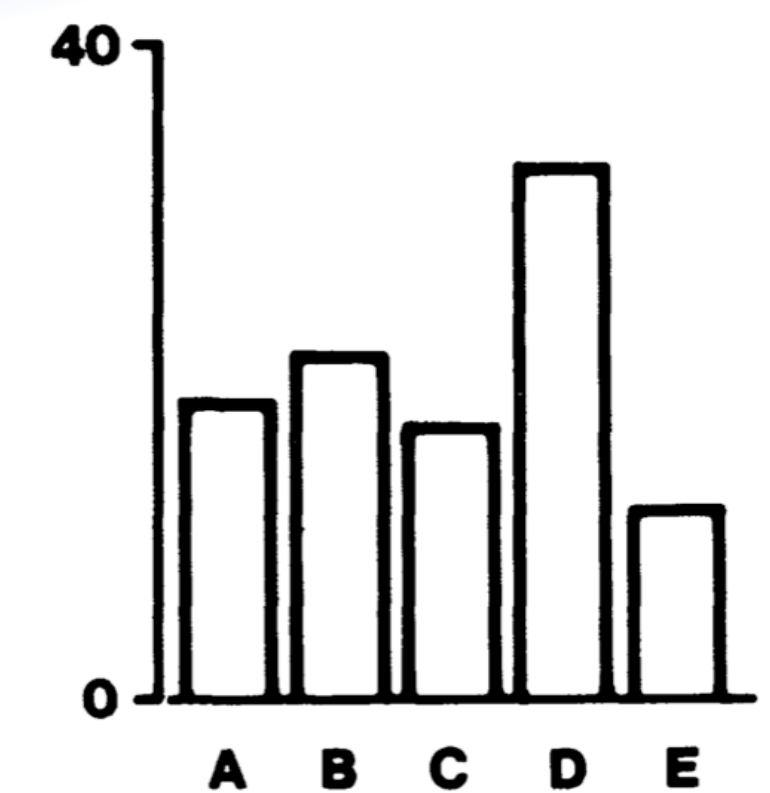
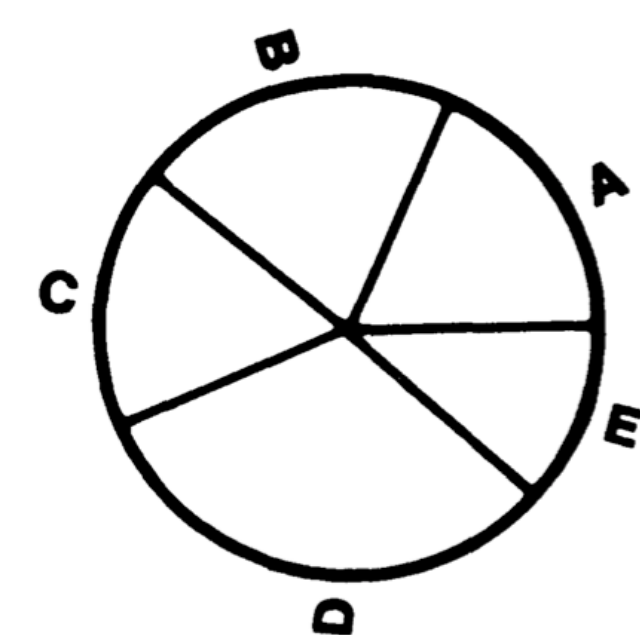
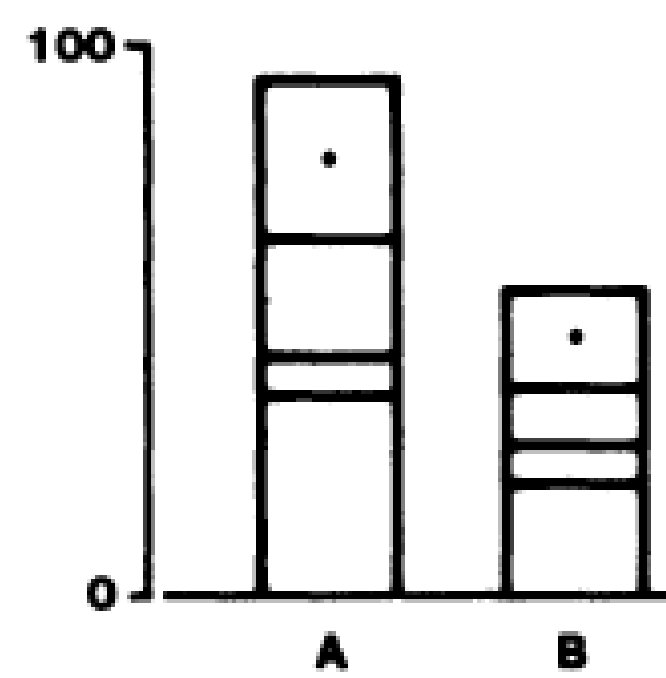
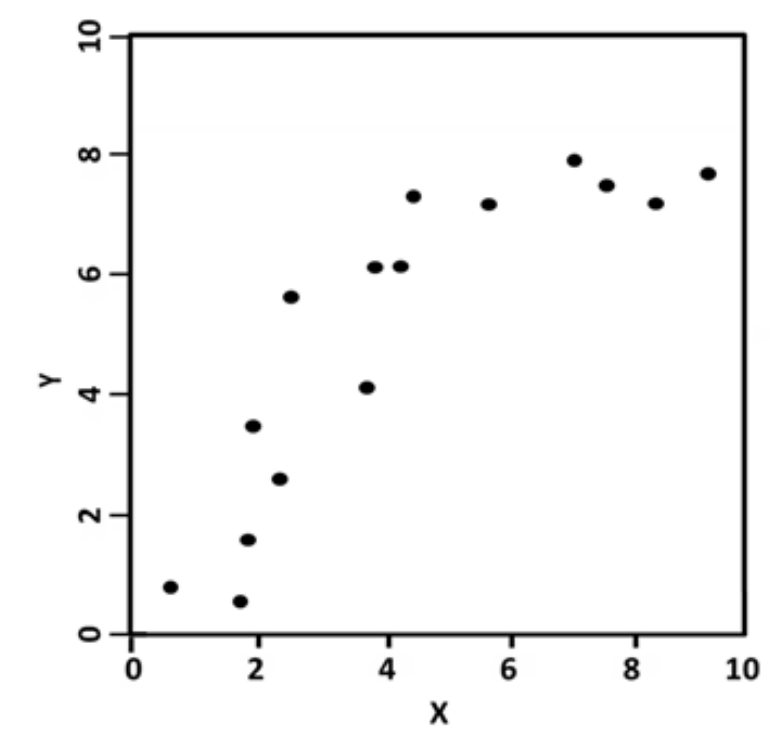
design

visualization

psychology

art

statistics



Ok, but why do we need
visualization?

LSST



Financial Markets



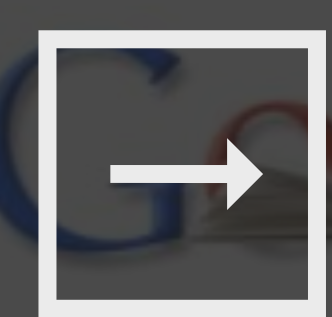
CHALLENGES:

- Scalability
- Complexity reduction
- Humans in-the-loop



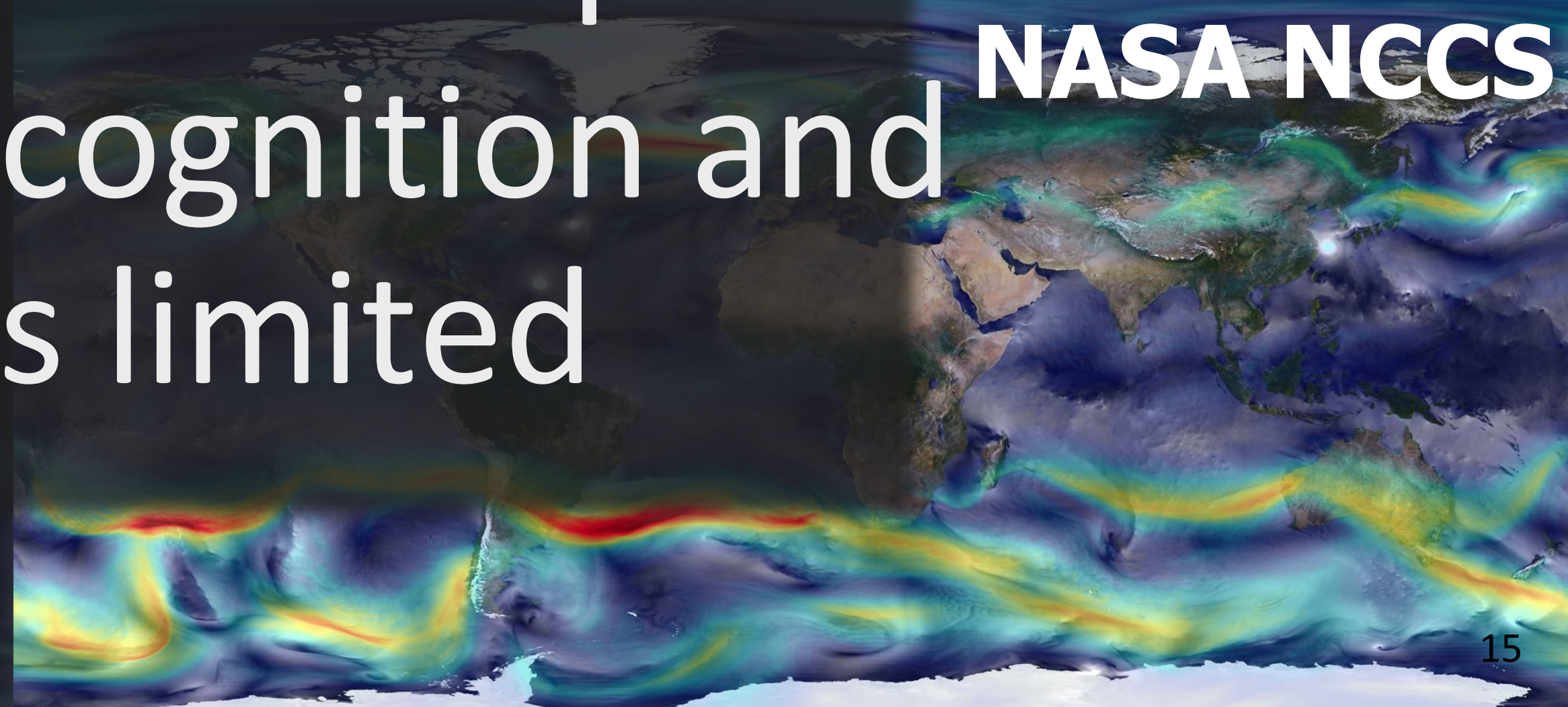
LHC

GenBank



human cognition and memory is limited

NASA NCCS



ngrams



q t f j n i x i g j u n a s b b t g r
k c l b v t x j x z x m x r g k l x
q h m z y w t e y j w n o

q t f j n i x i g j u n a s b b t g r
k c l b v t x j x z x m x r g k l x
q h m z y w t e y j w n o

“change blindness”



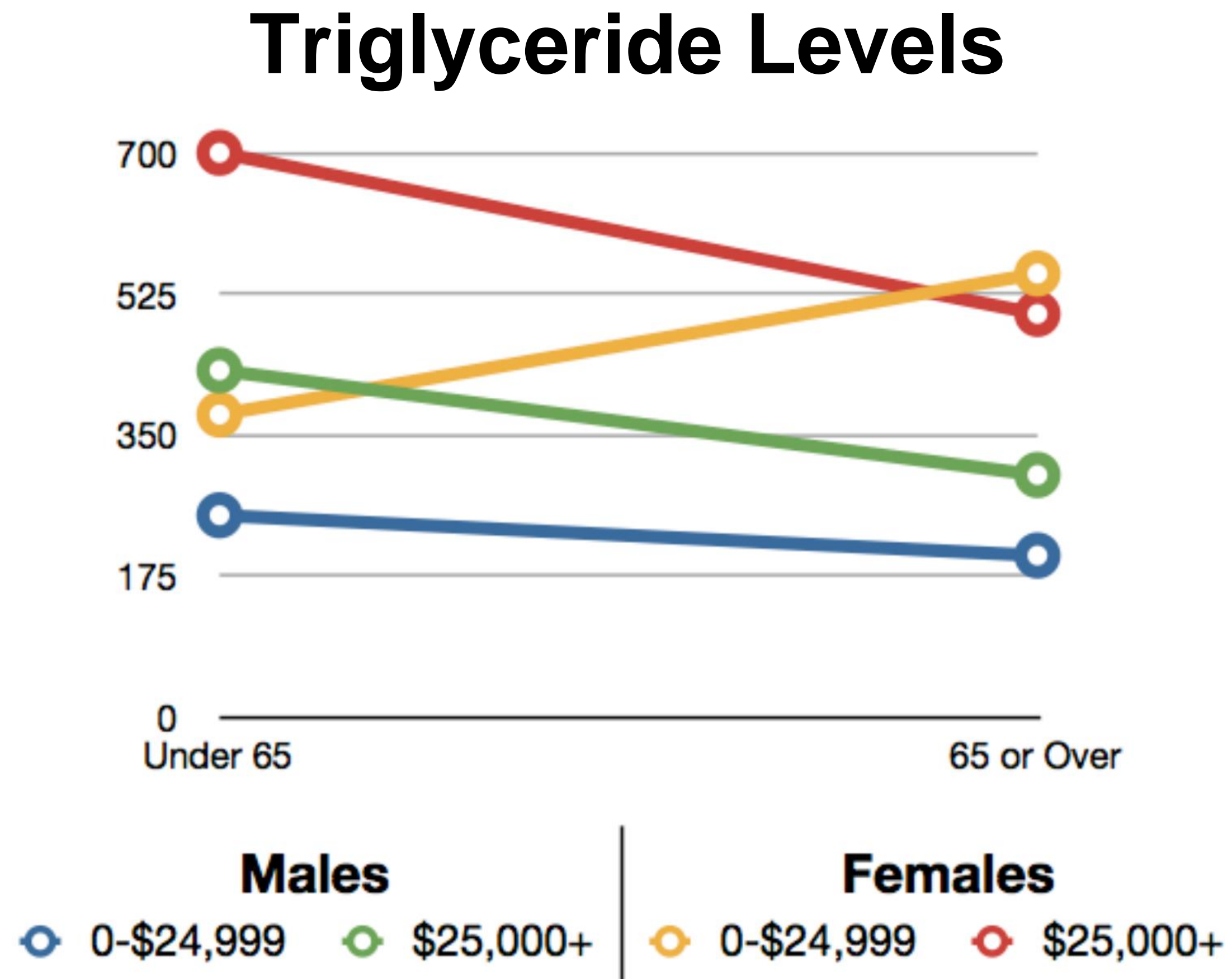
<https://www.youtube.com/watch?v=FW5xSQsspiQ>

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

Triglyceride Levels

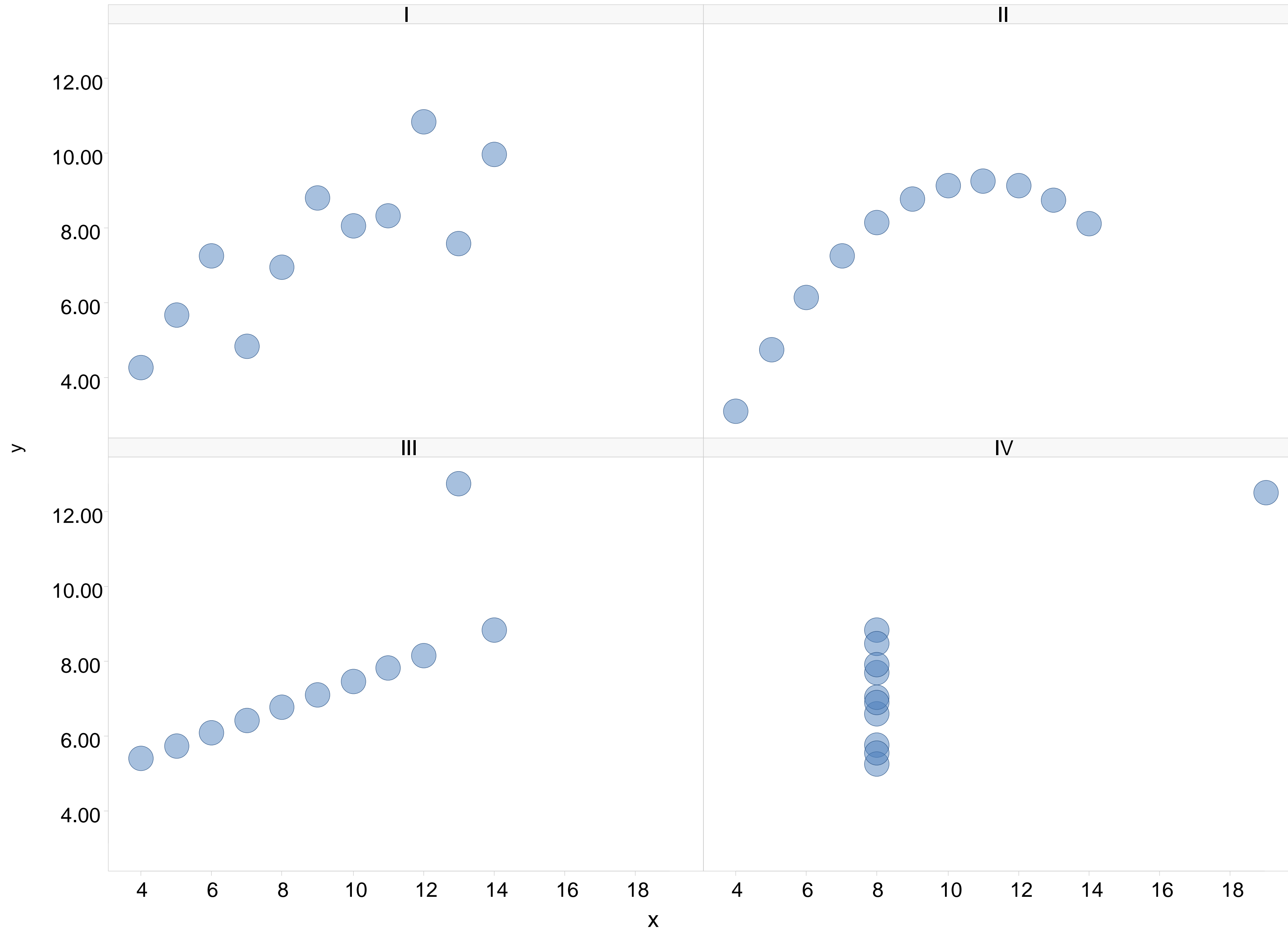
	Males		Females	
Income Group	Under 65	65 or Over	Under 65	65 or Over
0-\$24,999	250	200	375	550
\$25,000+	430	300	700	500

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



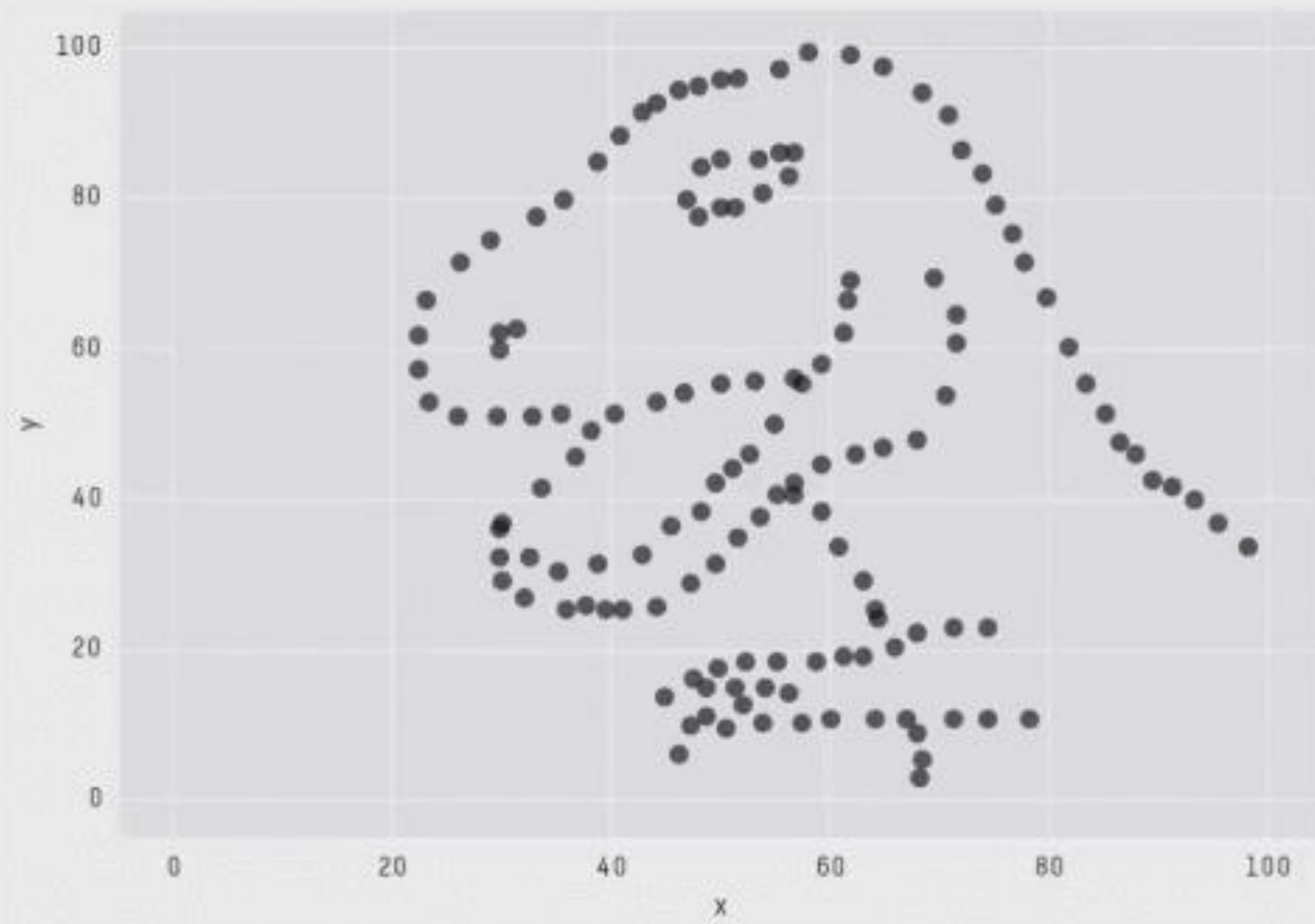
I		II		III		IV	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

	Value	Equality
X Mean	9	=
Y Mean	7.50	.00
X Variance	11	=
Y Variance	4.12	.00
Correlation	0.816	.000
Linear regression line	$y = 3.00 + 0.500x$.00 and .000



There are three types of lies: lies,
damned lies, and statistics

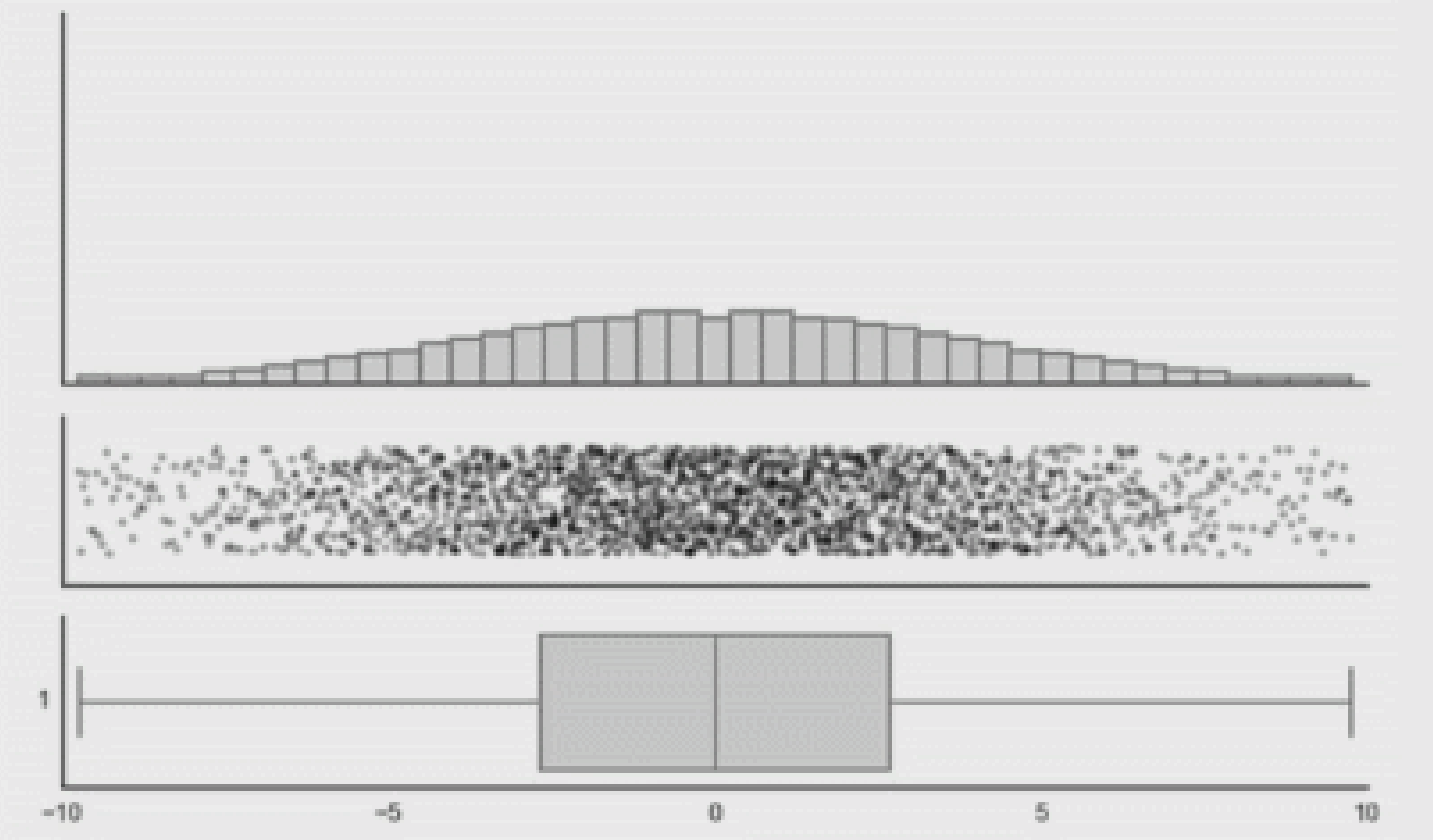
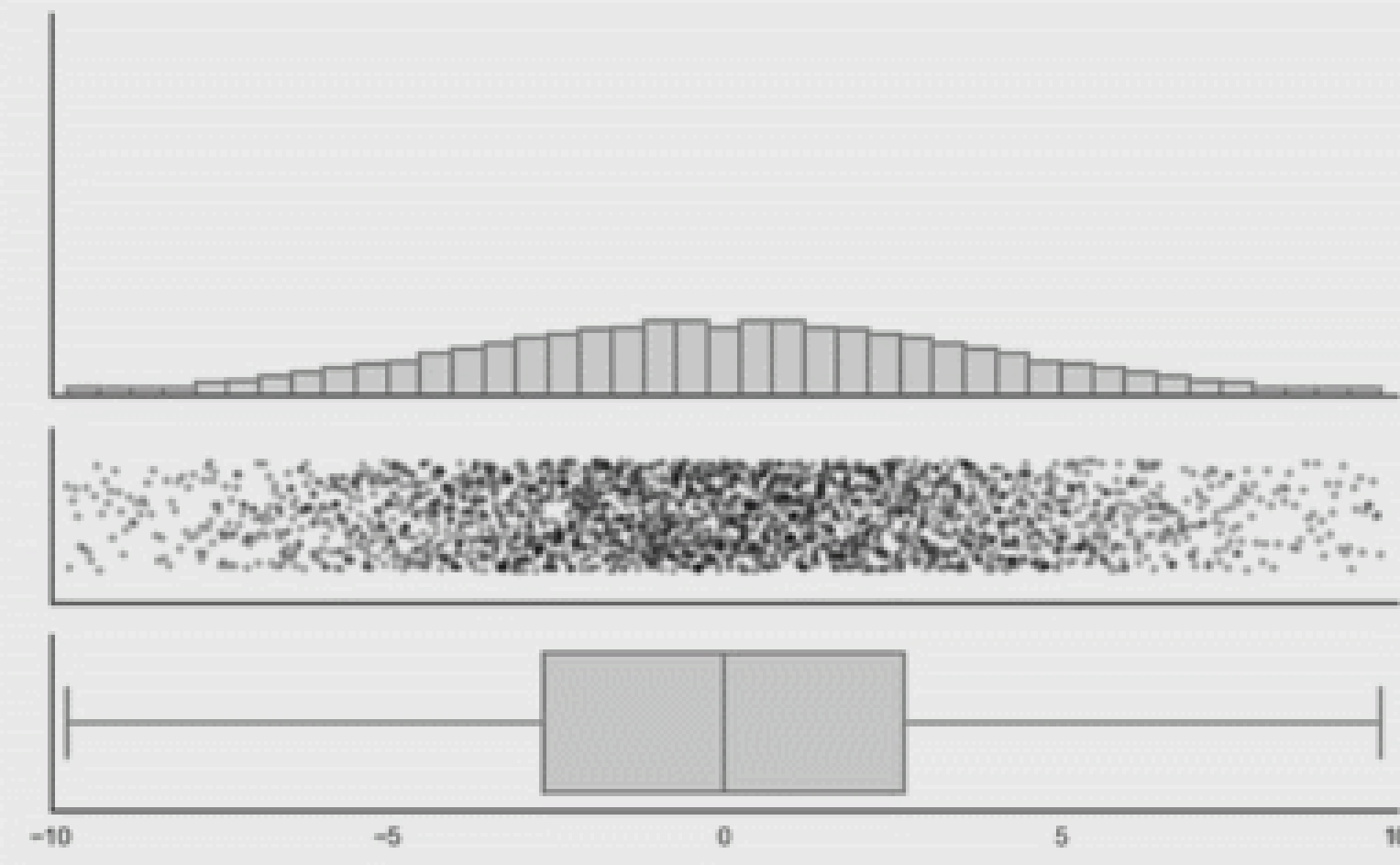
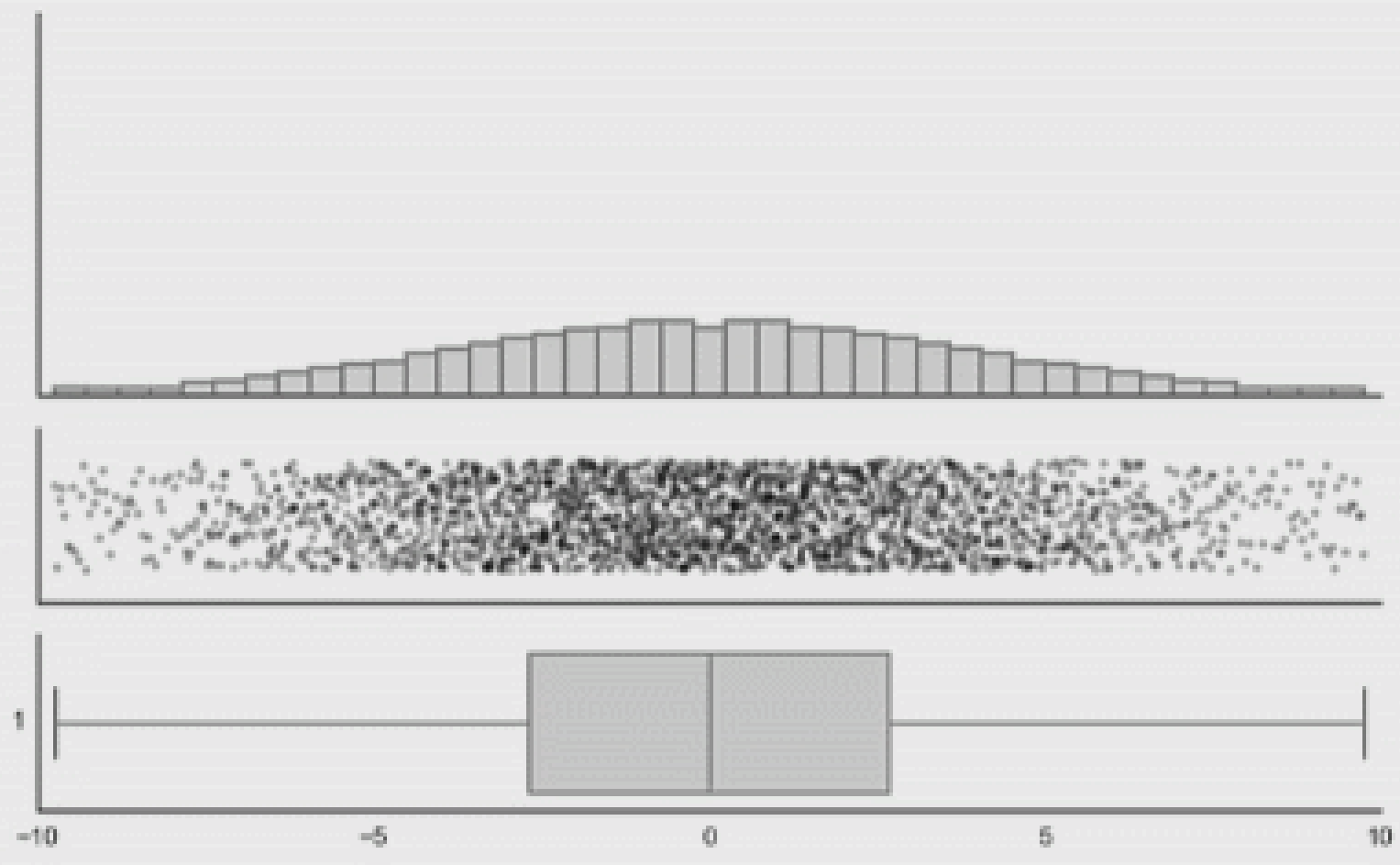
Mark Twain (maybe)

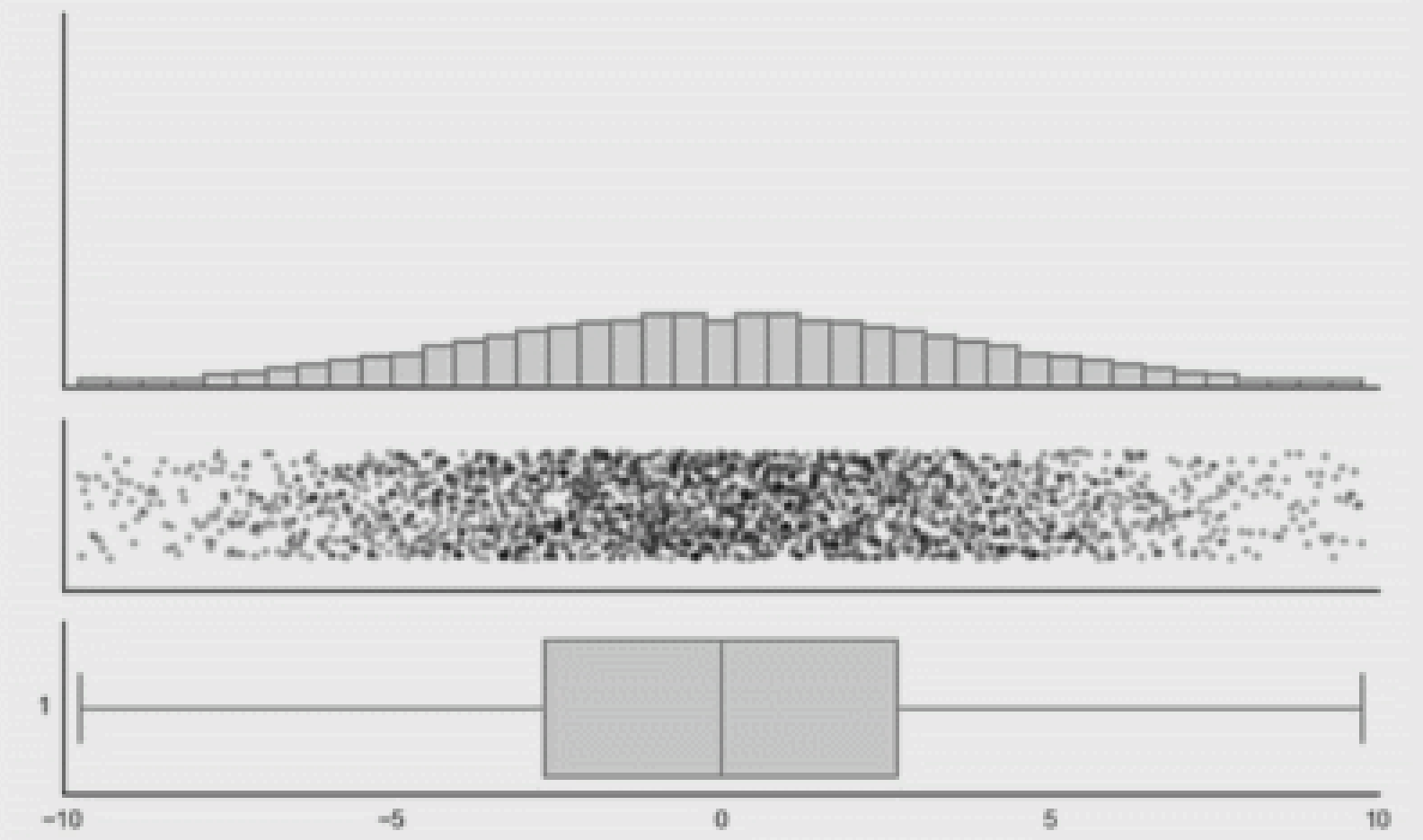
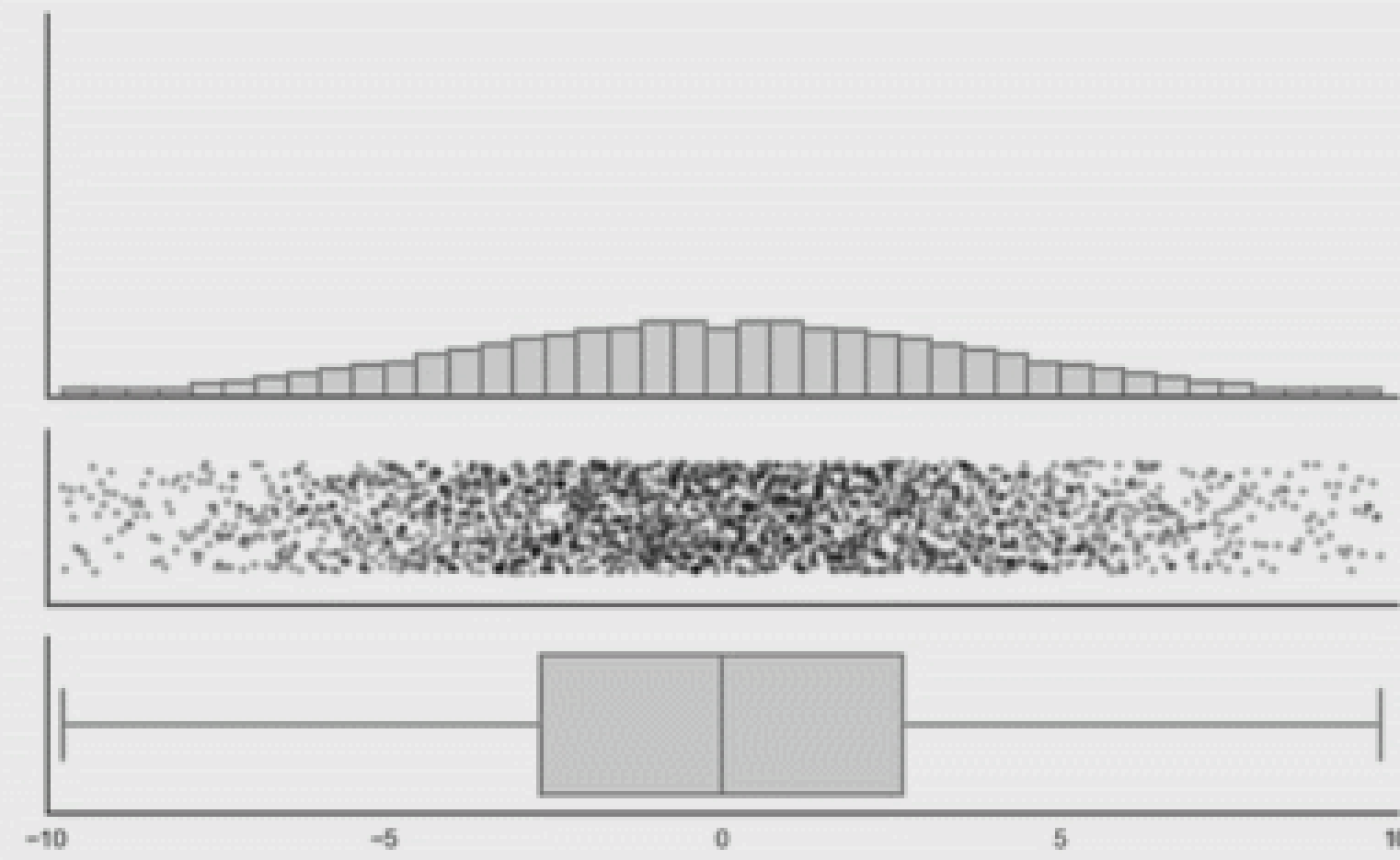
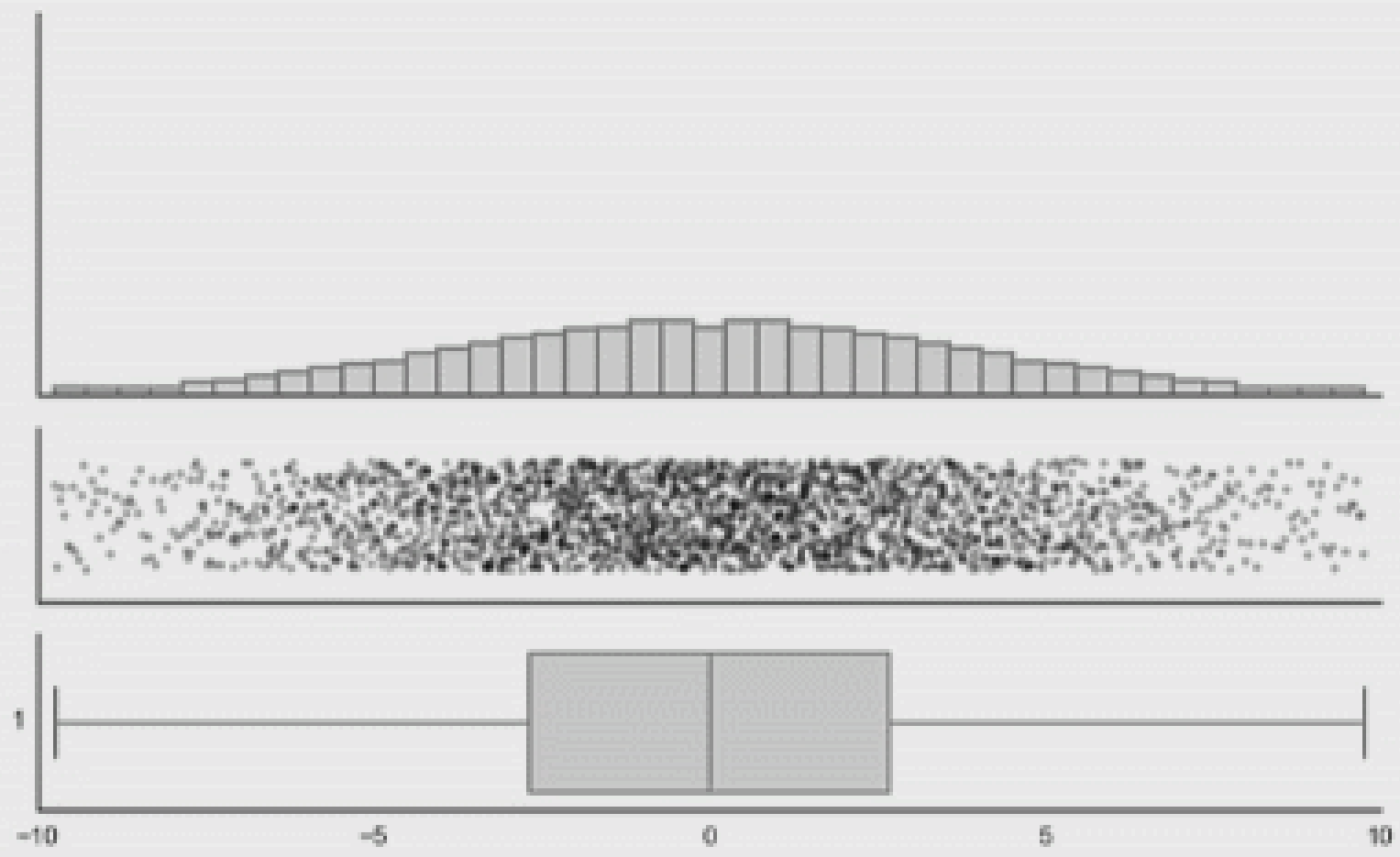


X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

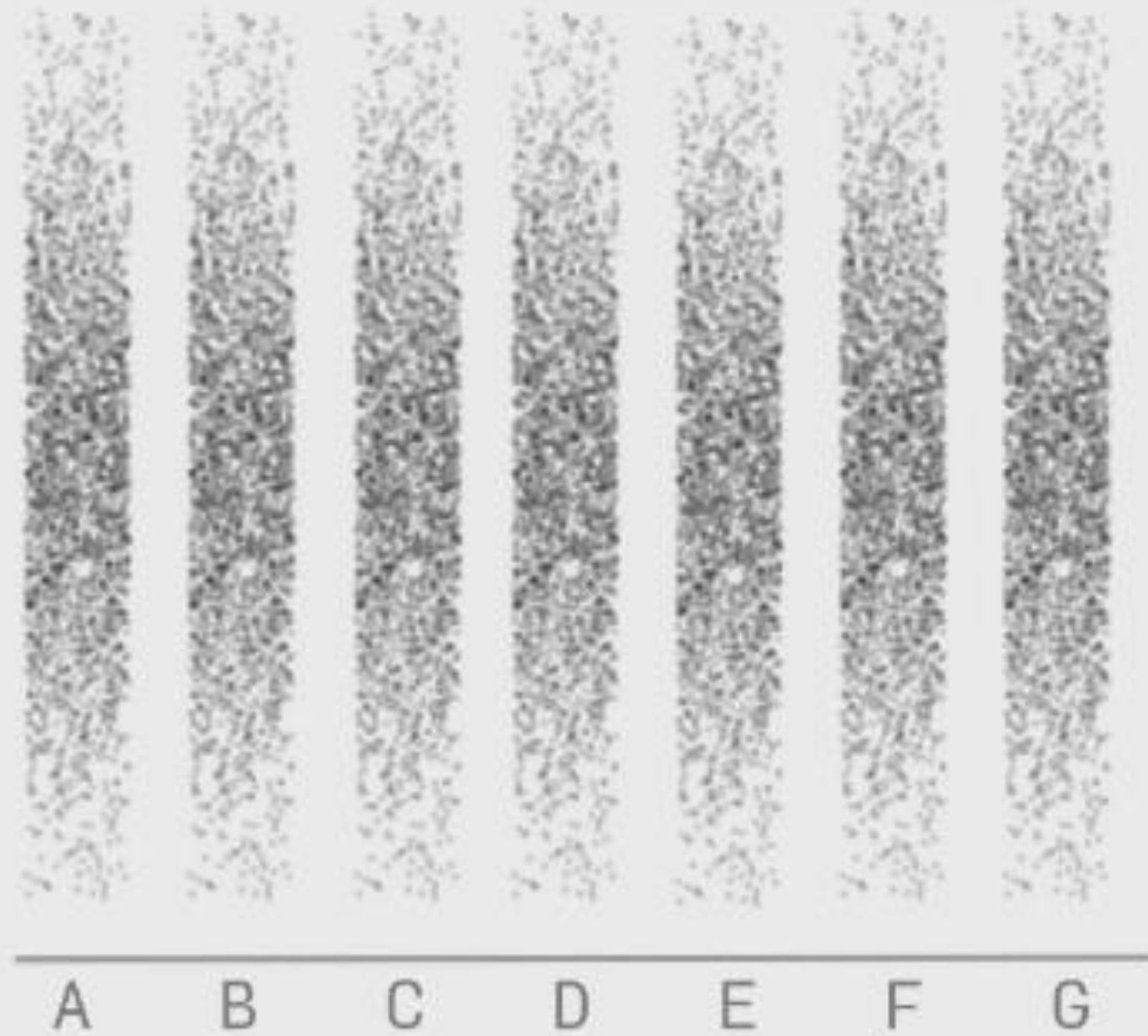
No catalogue of techniques can convey a willingness to look for what can be seen, whether or not anticipated. Yet this is at the heart of exploratory data analysis. ... the picture-examining eye is the best finder we have of the wholly unanticipated.

– Tukey, 1980

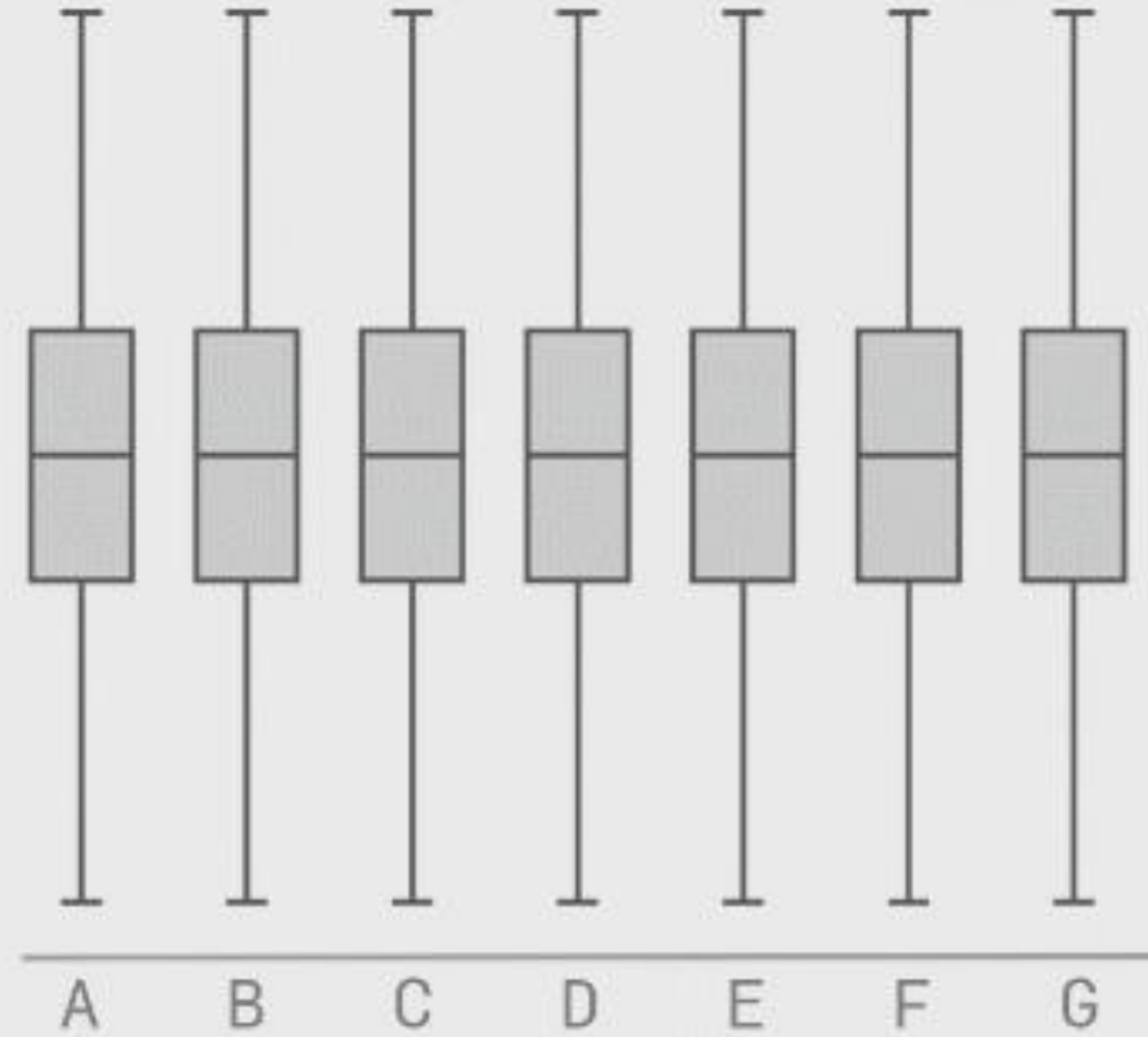




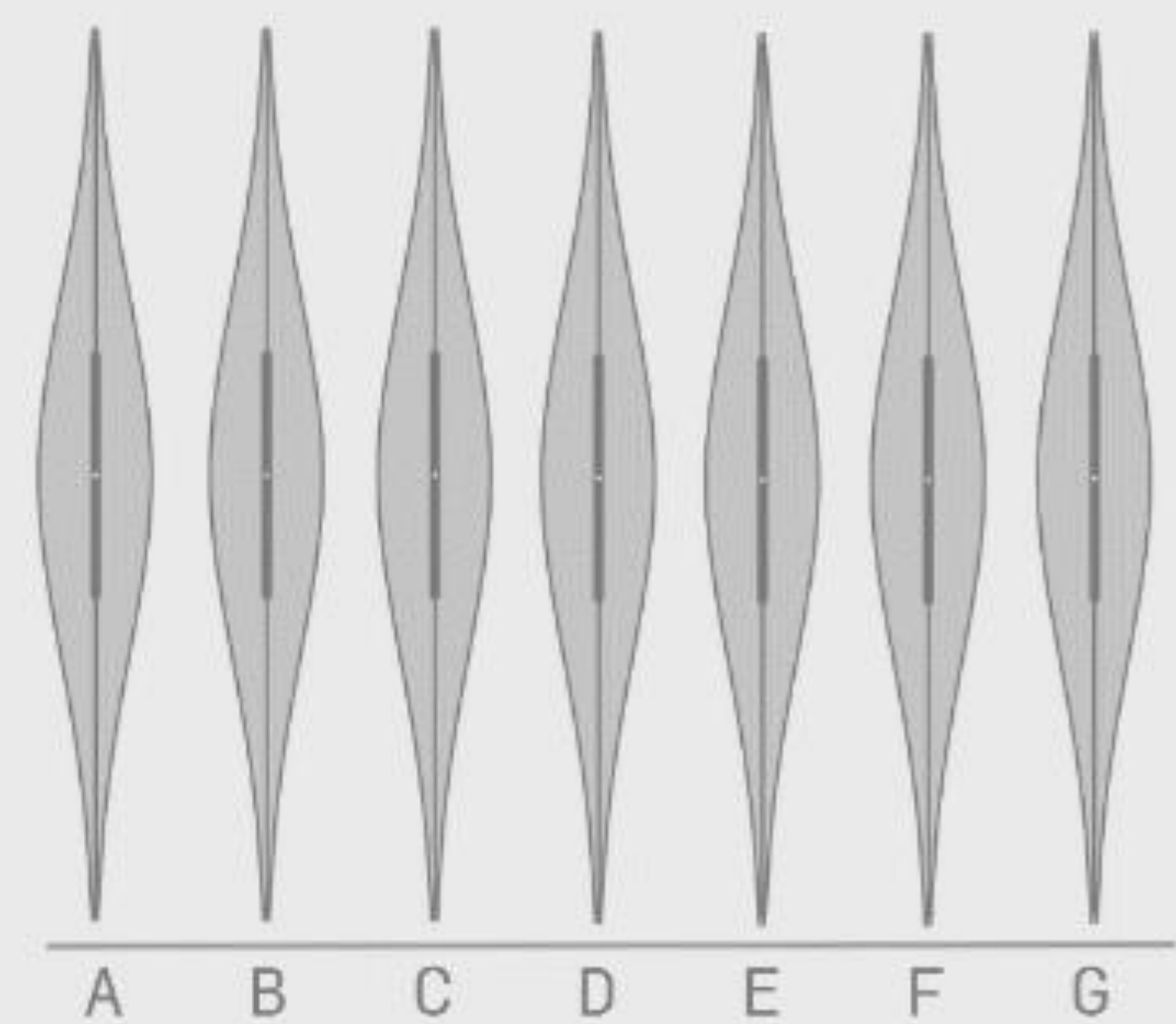
Raw Data



Box-plot of the Data



Violin-plot of the Data



Ok, but why do we need
visualization?

Why visualize your data?

- Help cognition
- Expand memory
- Generate hypotheses
- Answer questions
- Make decisions
- Find patterns
- Record
- Clarify
- Communicate
- Inspire

Why visualize your data?

Hans Rosling:

The best stats you've ever seen

TED2006 · 19:50 · Filmed Feb 2006

48 subtitle languages ?

View interactive transcript



Why visualize your data?



Generate hypotheses, Answer questions, Find patterns,
Record, Clarify, Communicate, Inspire

Hall of Fame or Hall of Shame



Prof. Krzysztof Gajos

wind map

past patterns

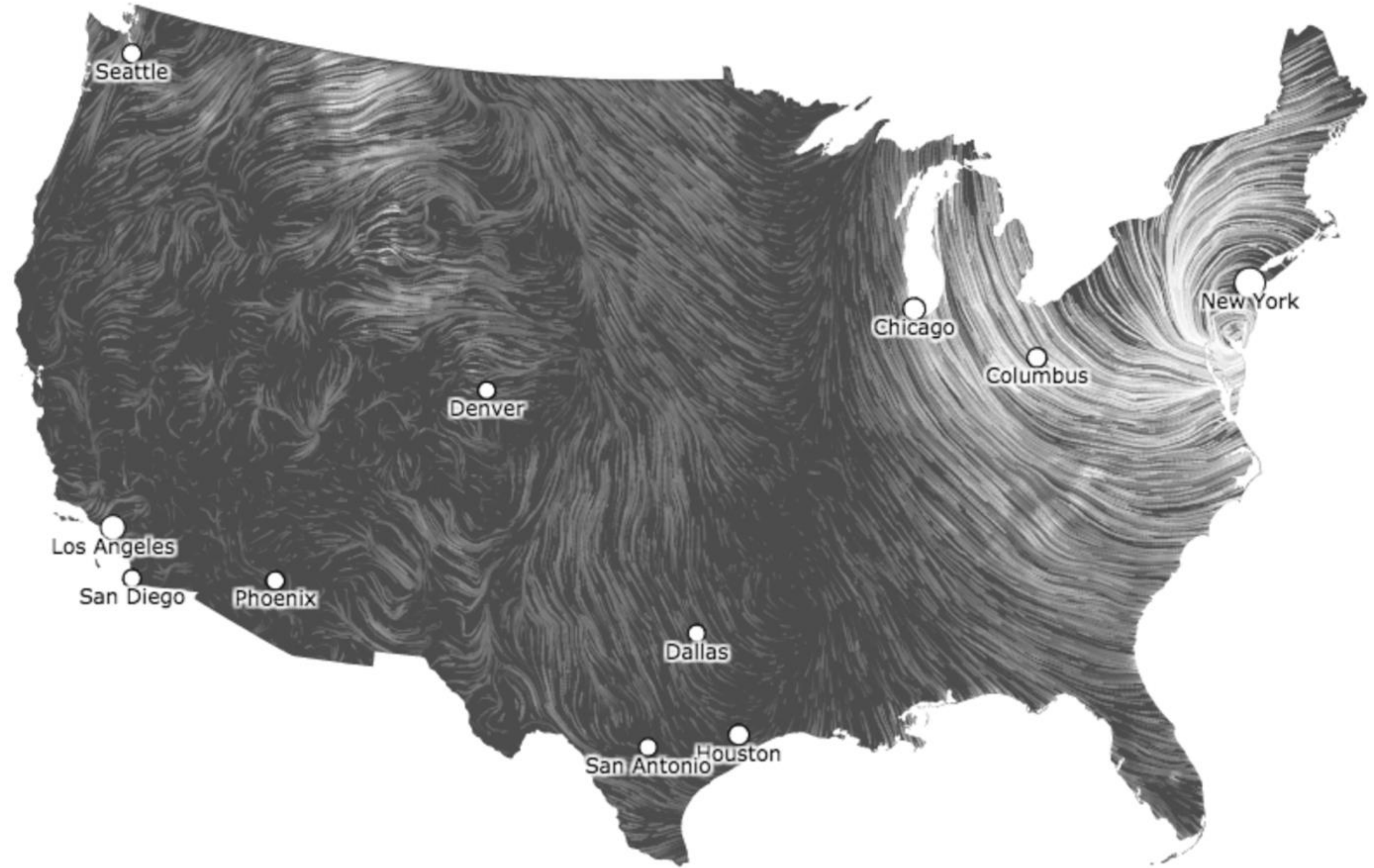
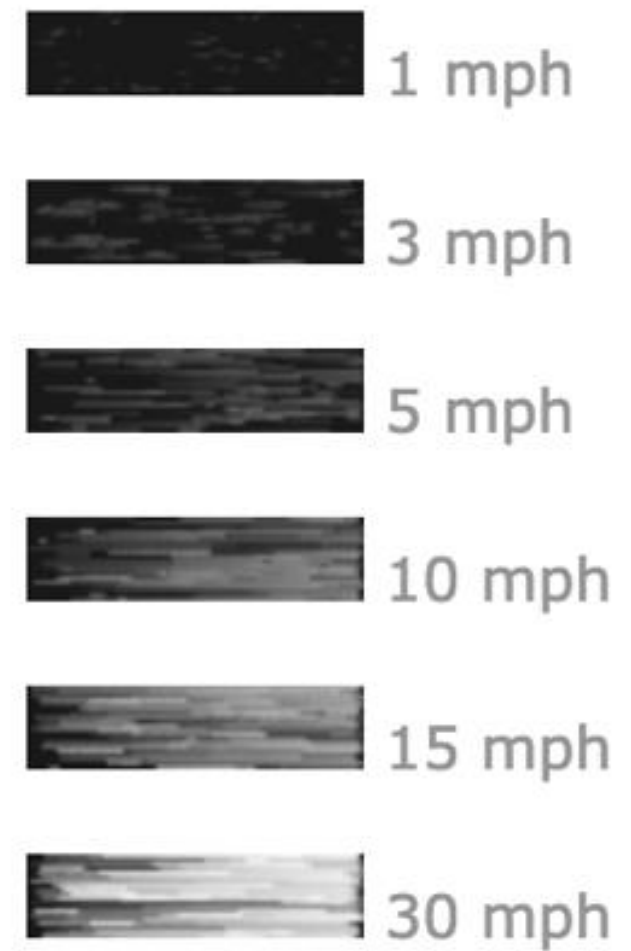
<previous next>

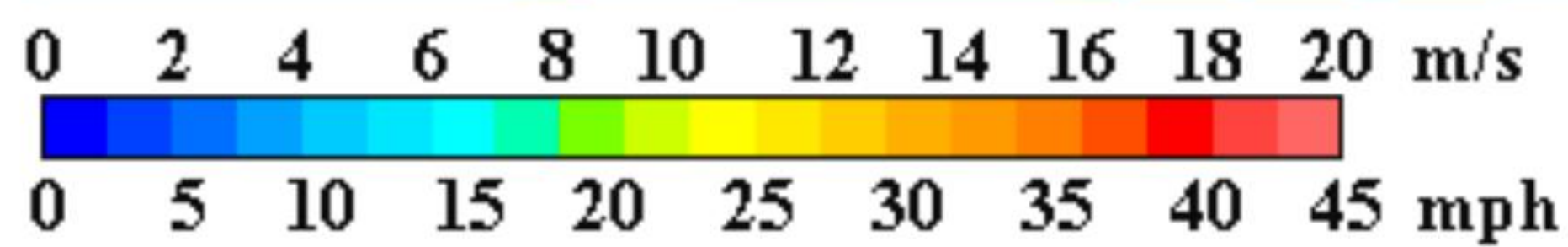
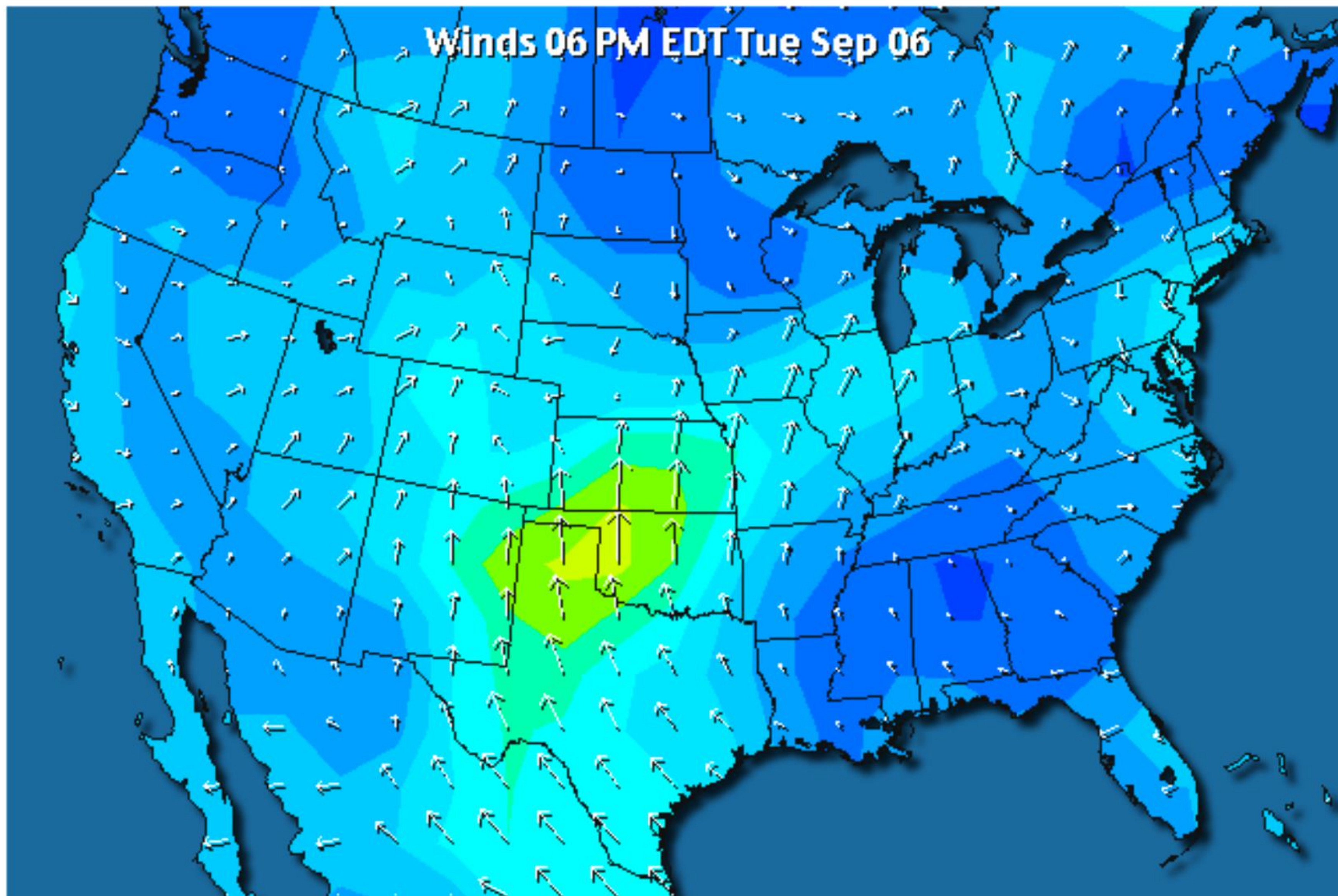
October 29, 2012

8:59 pm EST

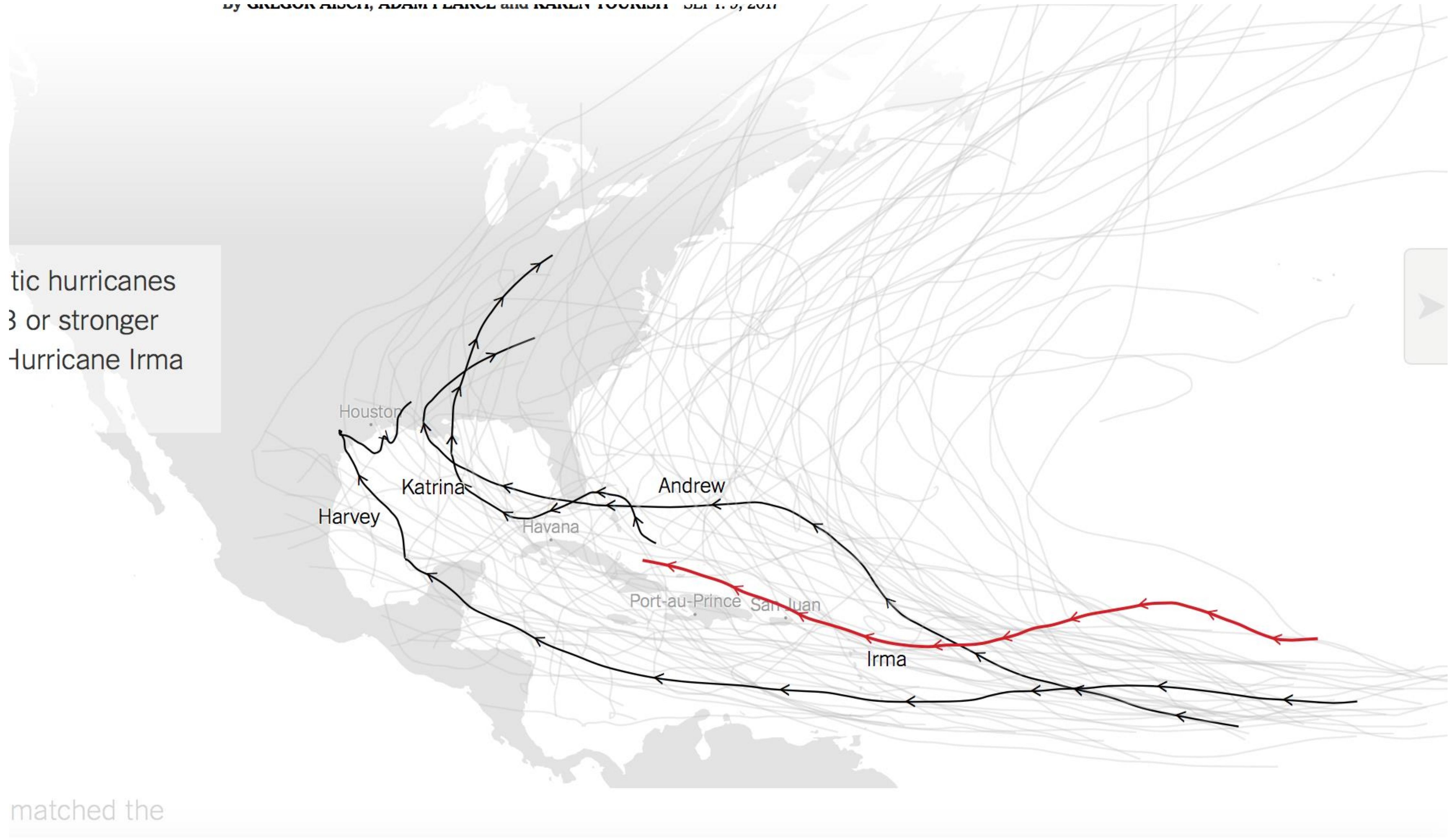
(time of forecast download)

top speed: **45.1 mph**
average: **9.4 mph**





tic hurricanes
3 or stronger
Hurricane Irma



matched the

Upcoming Assignments & Communication

<https://c.dunne.dev/ds4200f20>

Look at the upcoming assignments and deadlines (**12:01am**)!

- Textbook, Readings & Reading Quizzes
- Assignment 1a,b,c
- Projects

Everyday Required Supplies:

- 5+ colors of pen/pencil
- White paper
- Laptop and charger

Use Canvas Discussions for general questions, email the instructor/TAs for questions specific to you.