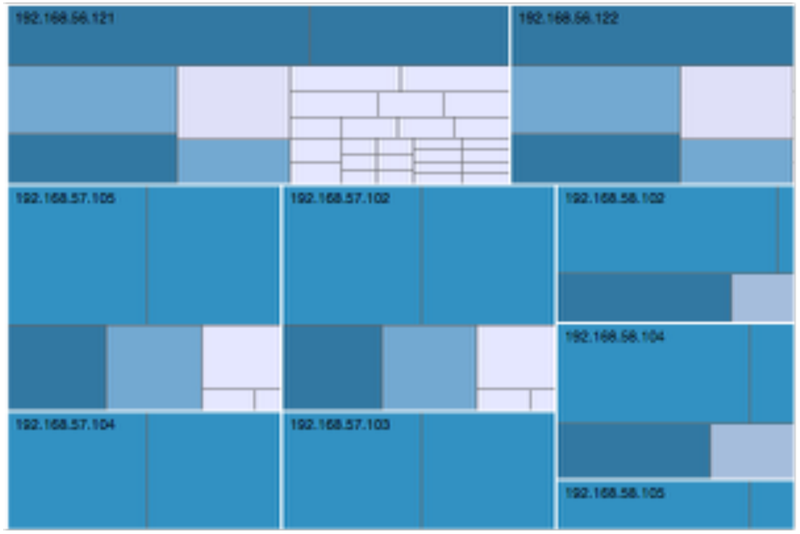
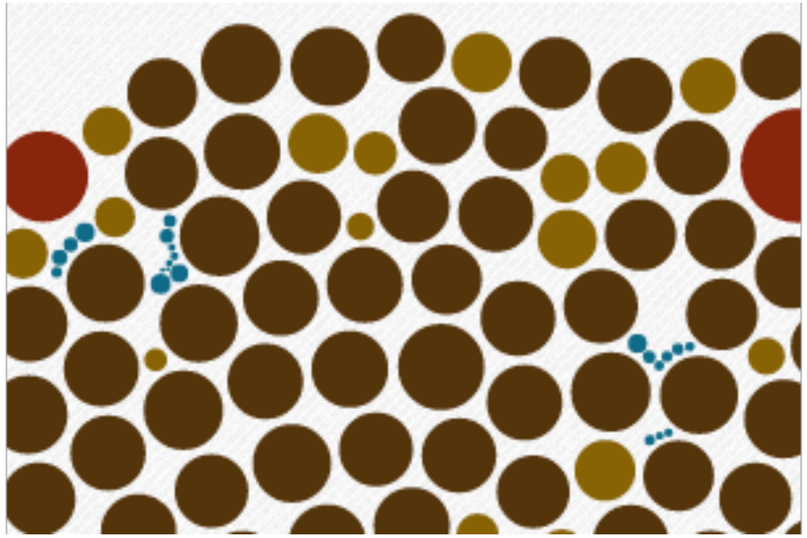
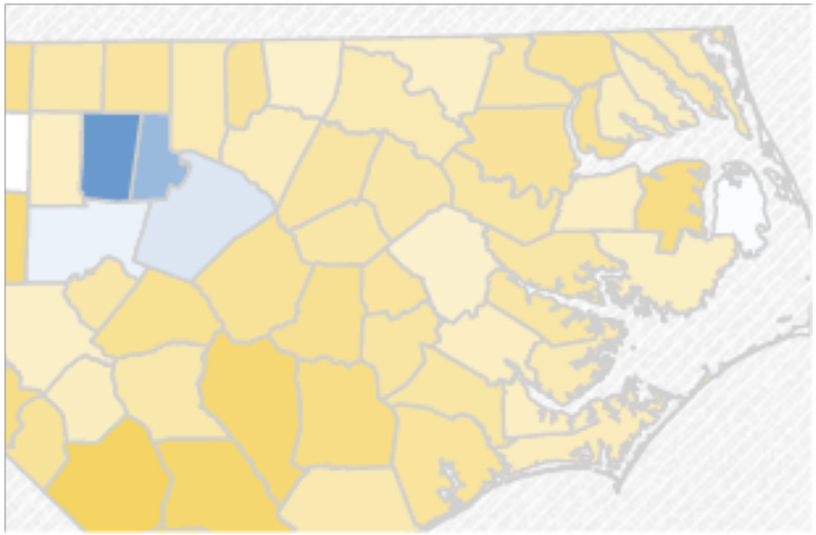
Nothing –
not the careful logic of mathematics,
not statistical models and theories,
not the awesome arithmetic power of modern computers –
nothing can substitute here for the flexibility
of the informed human mind...

Accordingly, both [analysis] approaches and techniques need to be structured so as to facilitate human involvement and intervention.

– John W. Tukey & Martin B. Wilk, Data Analysis & Statistics, 1966







Data Visualization

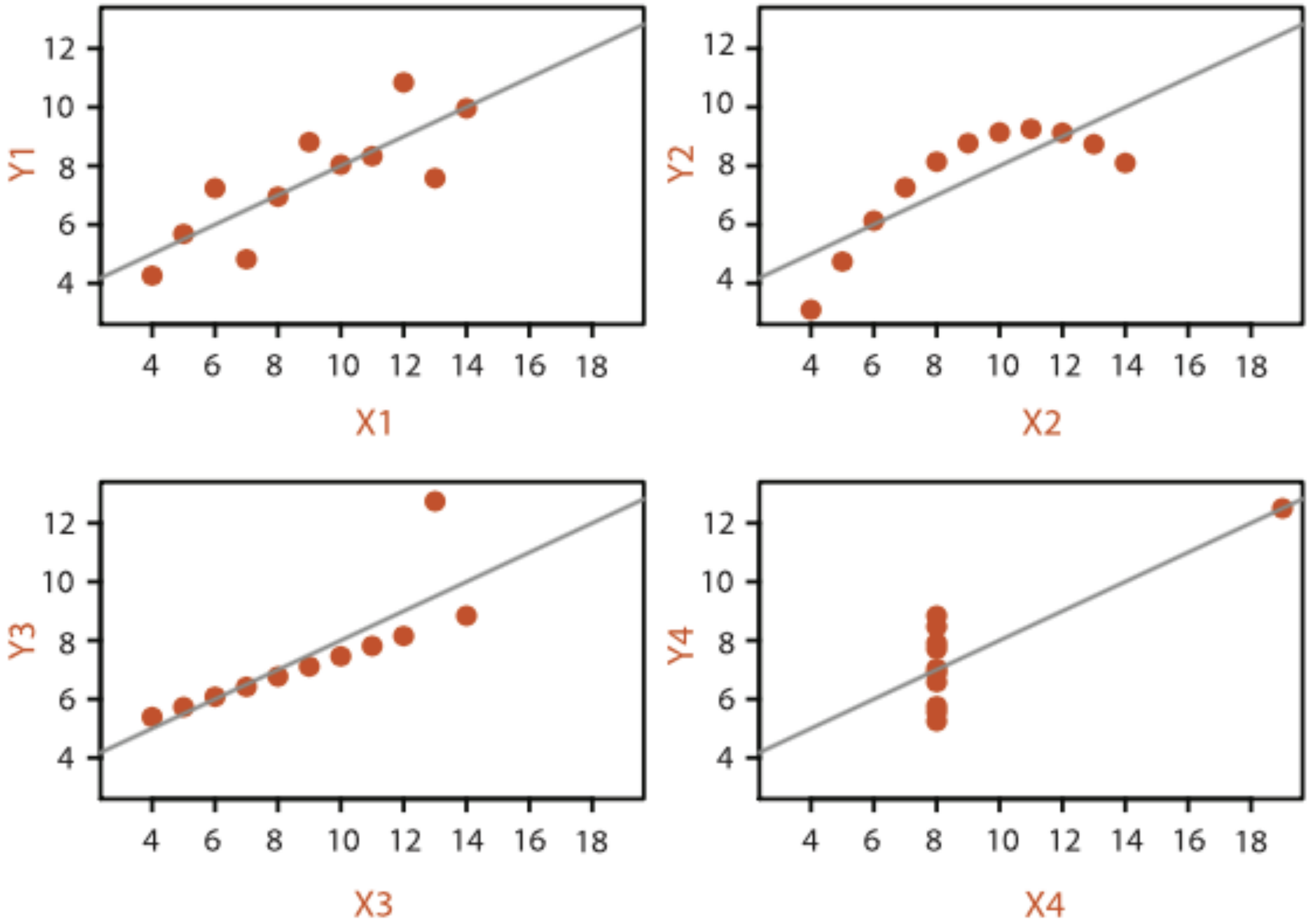
Visualization

Visualization

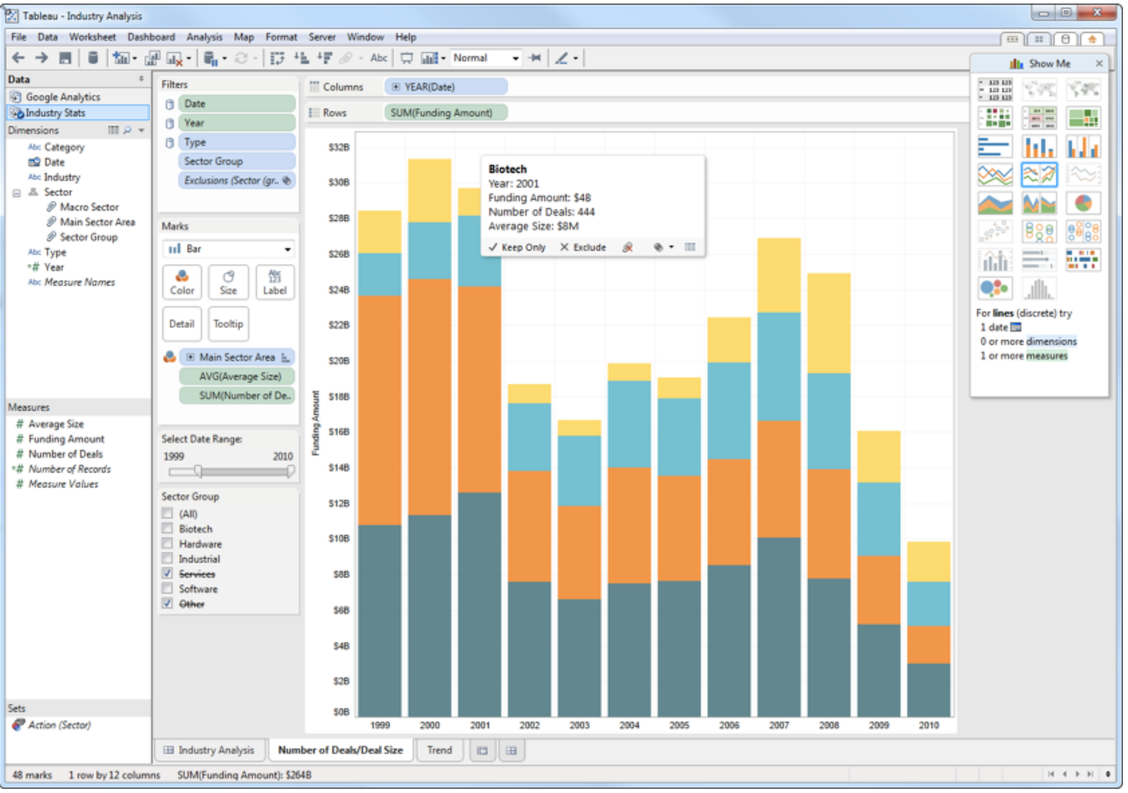
…is an indispensable tool for analysis and understanding.

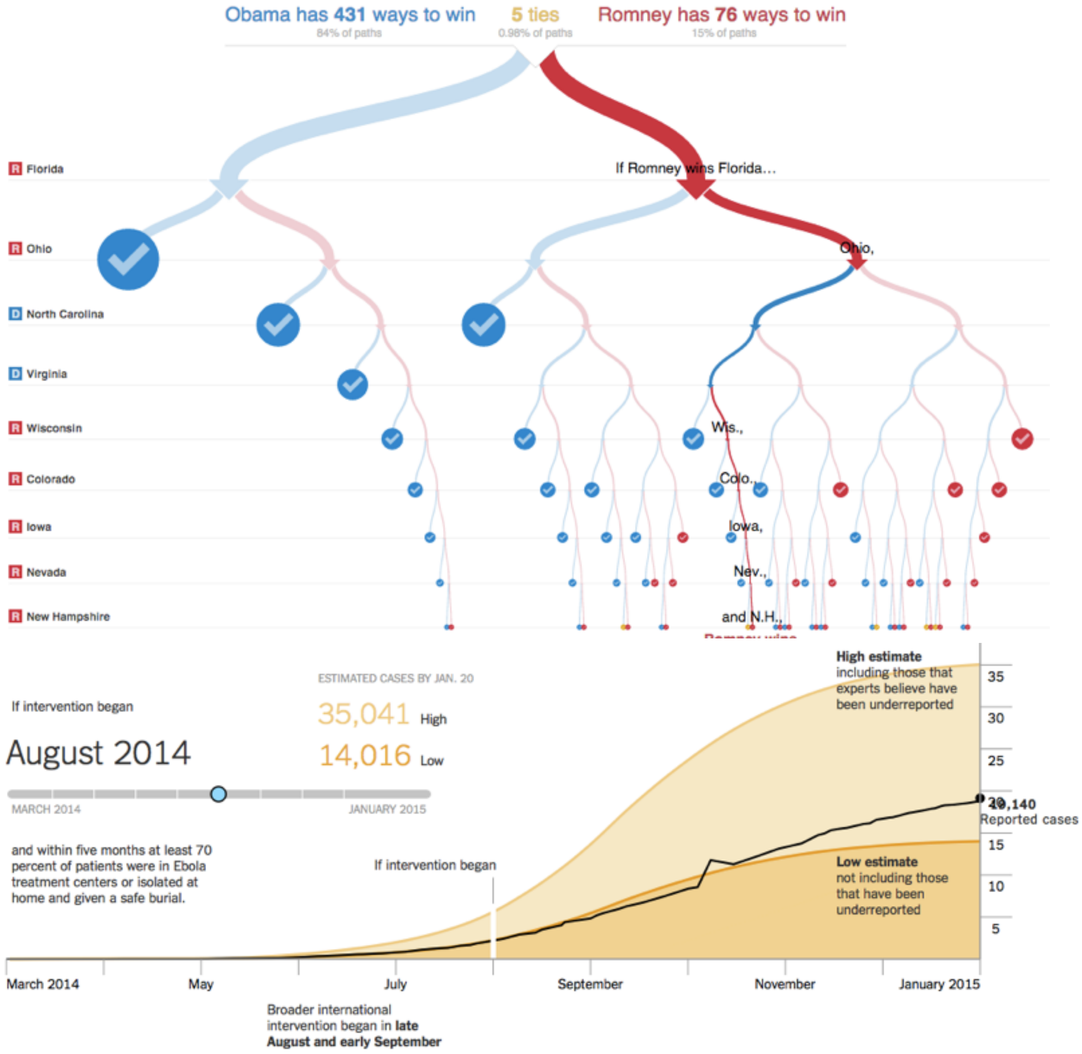
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	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

	1		2		3		4	
	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
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



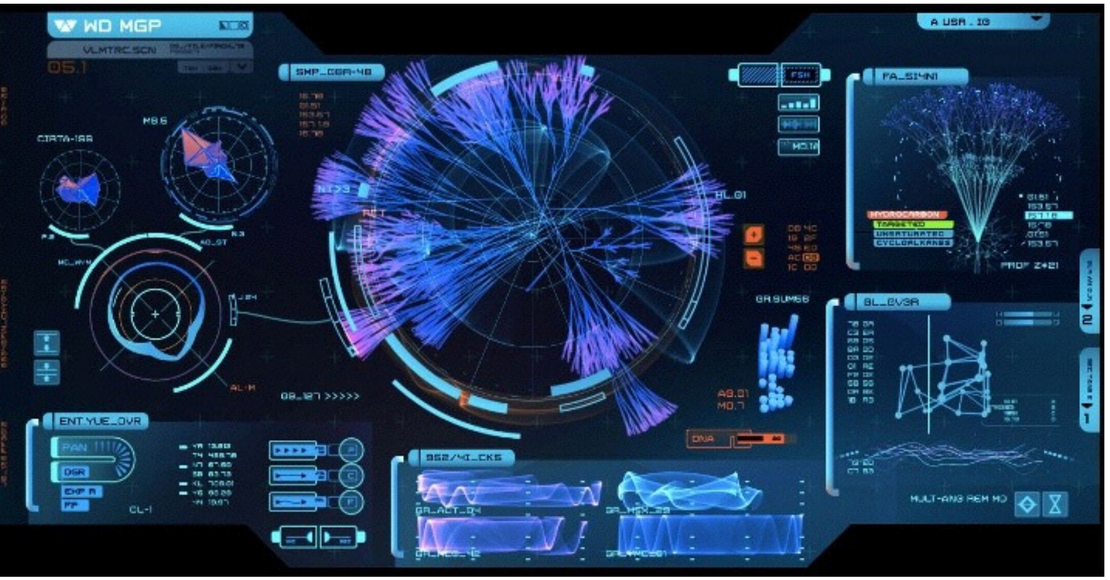
Francis Anscombe, Graphs in Statistical Analysis, 1973.





Exploratory

Expository



Exploratory

Expository

I’m Lane

Administrative

Course Description

“In this course we will study the theory and practice of data visualization.

Topics include the fundamental principles, concepts, and techniques of visualization and how visualization can be used to uncover and communicate data-driven insights.”

Learning Goals

1. Critically evaluate and deconstruct data visualizations.

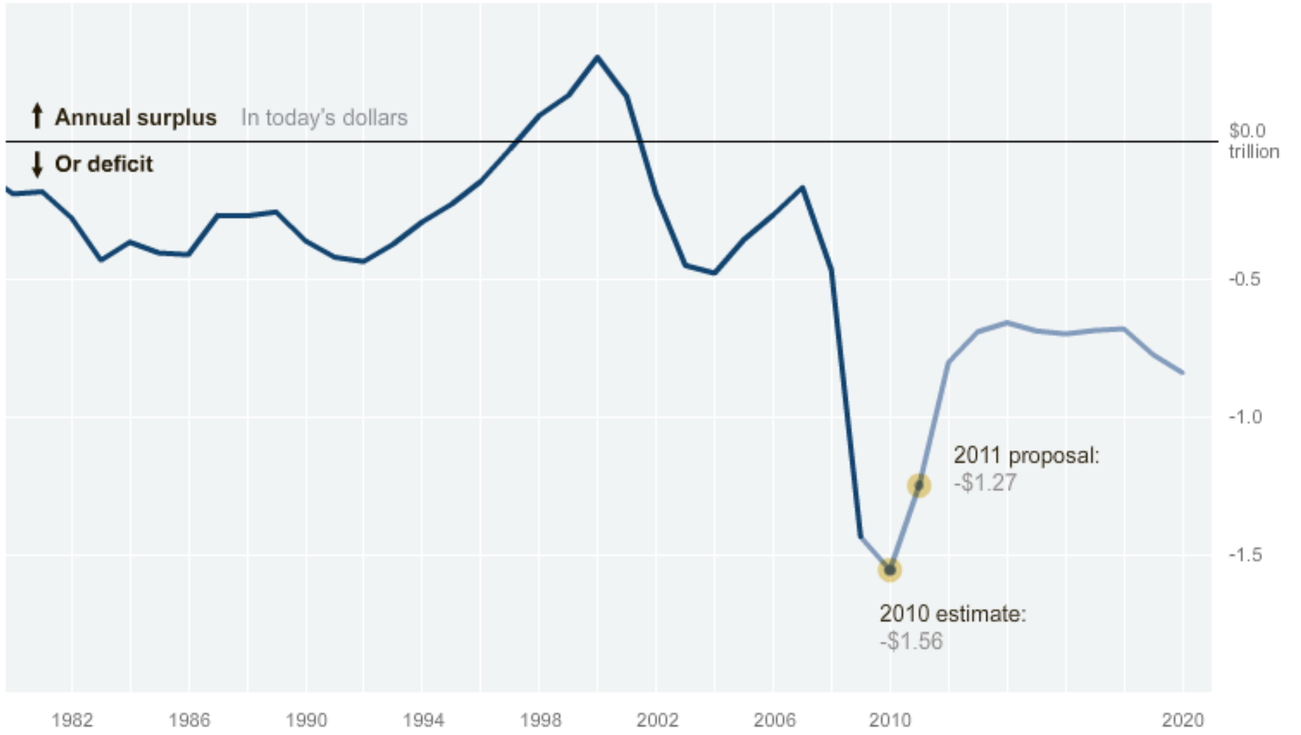
Published: February 2, 2010

Budget Forecasts, Compared With Reality

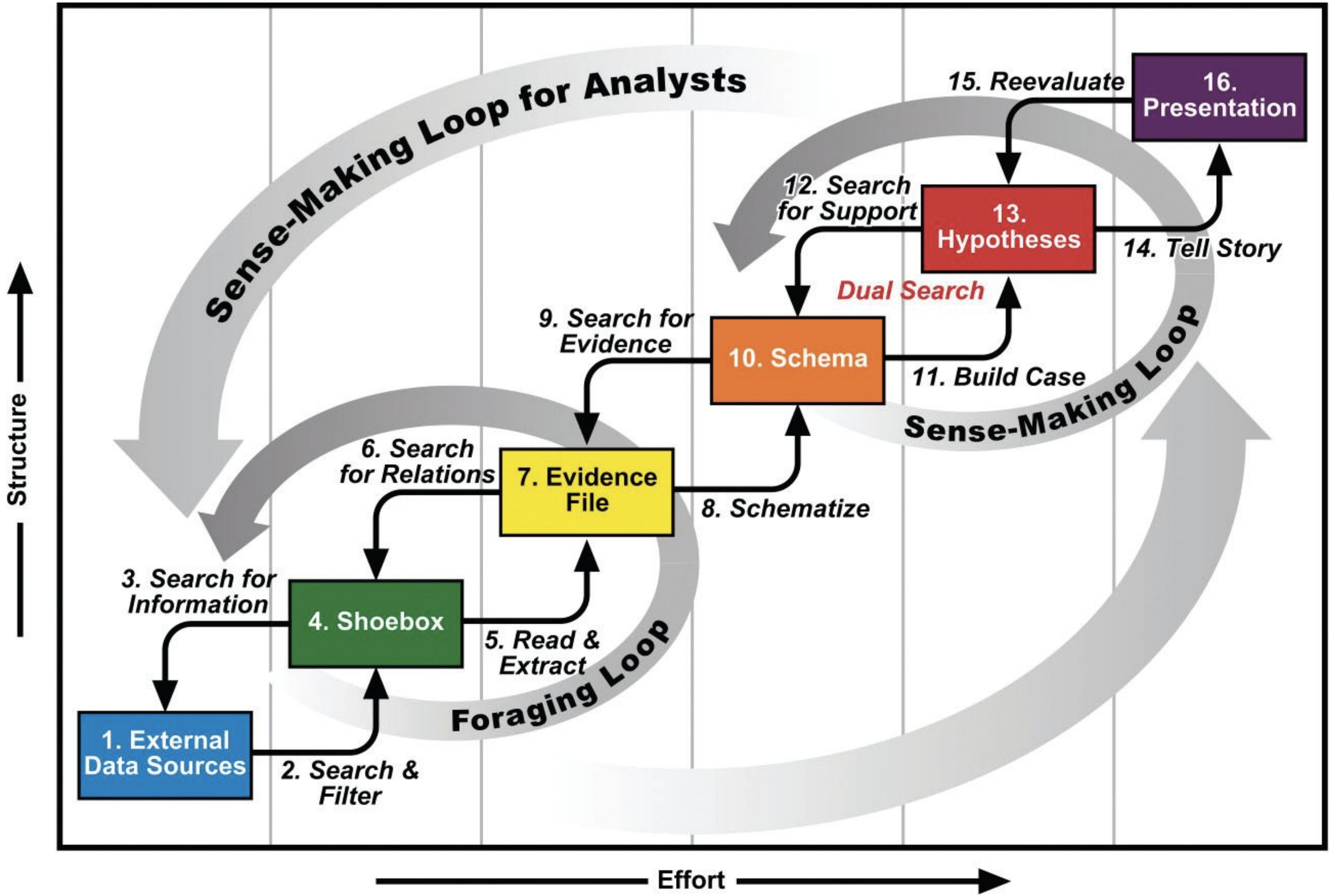
Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

Falling short

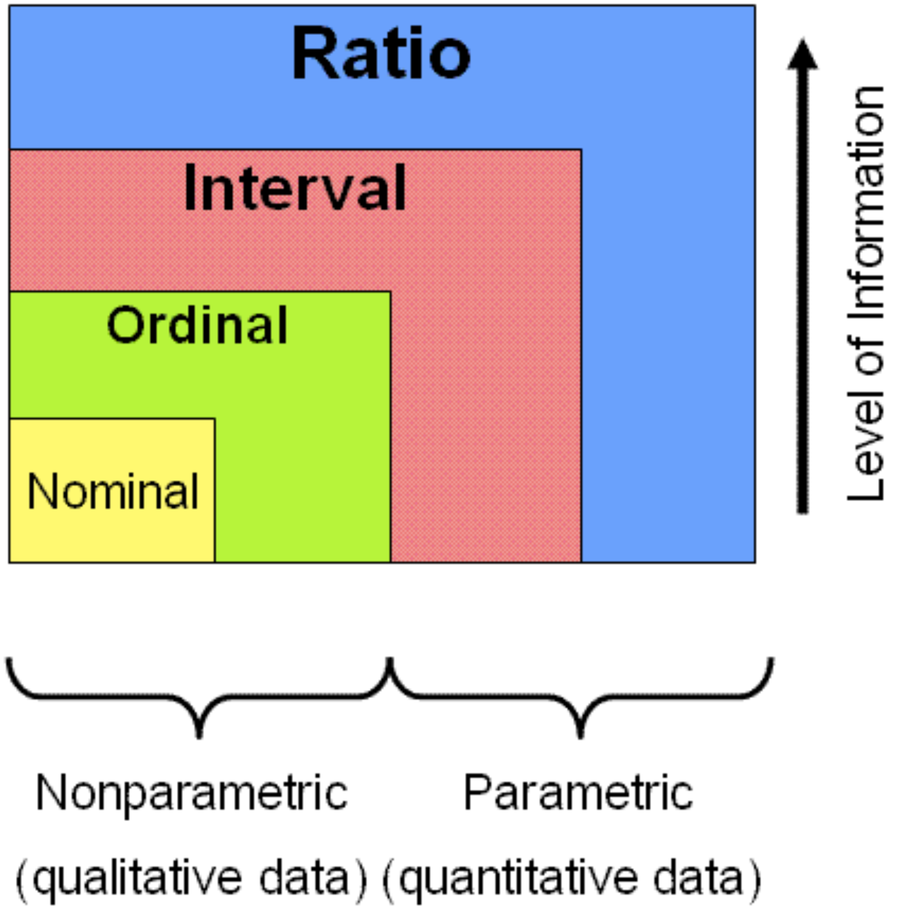
President Obama's budget proposal estimates a deficit of $1.6 trillion for the current fiscal year and $1.3 trillion in 2011.



2. Identify application areas for visualization in analysis workflows.


Sense-Making Loop for Analysts
15. Reevaluate
16. Presentation
12. Search for Support
13. Hypotheses
14. Tell Story
Dual Search
9. Search for Evidence
10. Schema
11. Build Case
Sense-Making Loop
6. Search for Relations
7. Evidence File
8. Schematize
3. Search for Information
4. Shoebox
5. Read & Extract
Foraging Loop
1. External Data Sources
2. Search & Filter
Structure
Effort

3. Evaluate the characteristics and structure of data you encounter to refine design options.



DATA

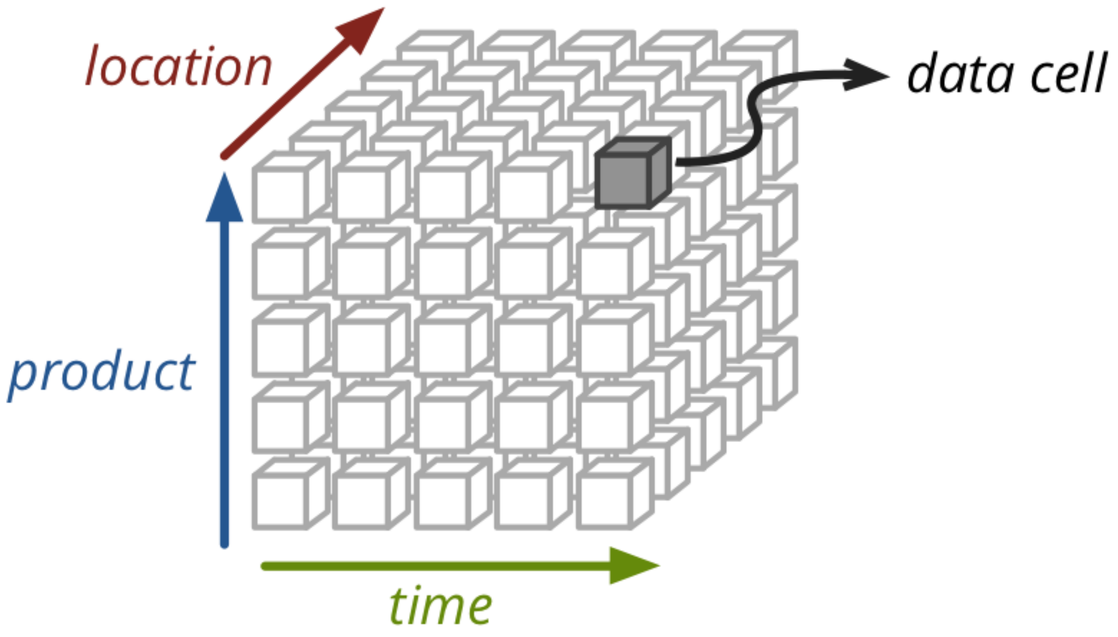
D

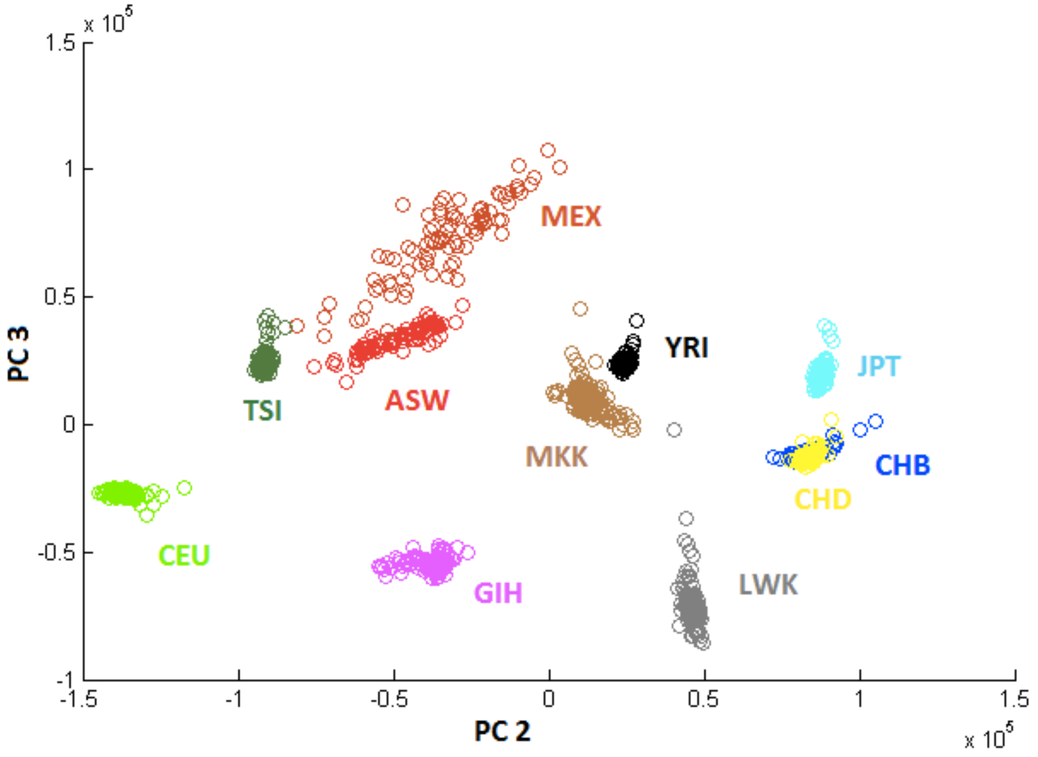
A T

A

DATA

4. Use algorithms, aggregation, sampling, and similar techniques to refine and manipulate data.

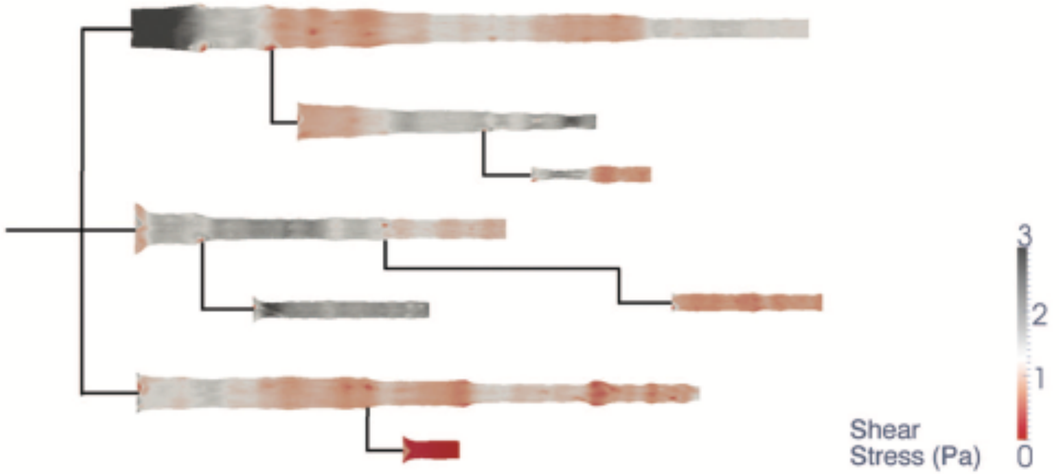

location
data cell
product
time

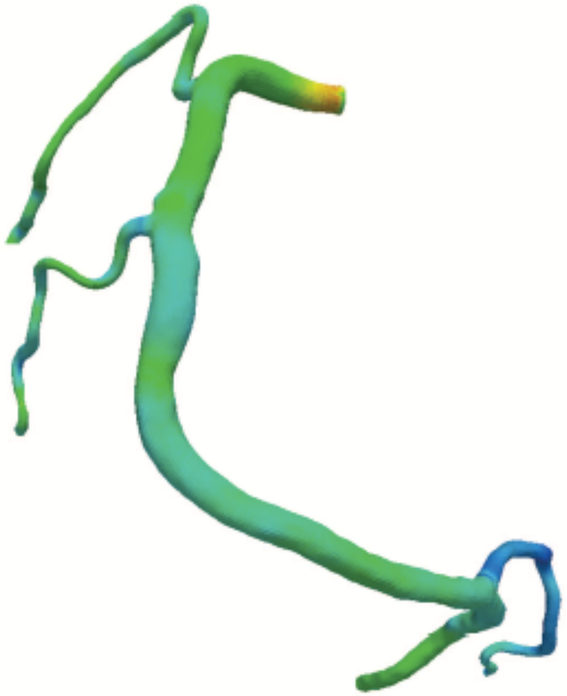

x 10^5
1.5
1
0.5
0
-0.5
-1
PC 3
-1.5
-1
-0.5
0
0.5
1
1.5
PC 2
x 10^5
MEX
YRI
JPT
TSI
ASW
MKK
CHB
CHD
CEU
GIH
LWK

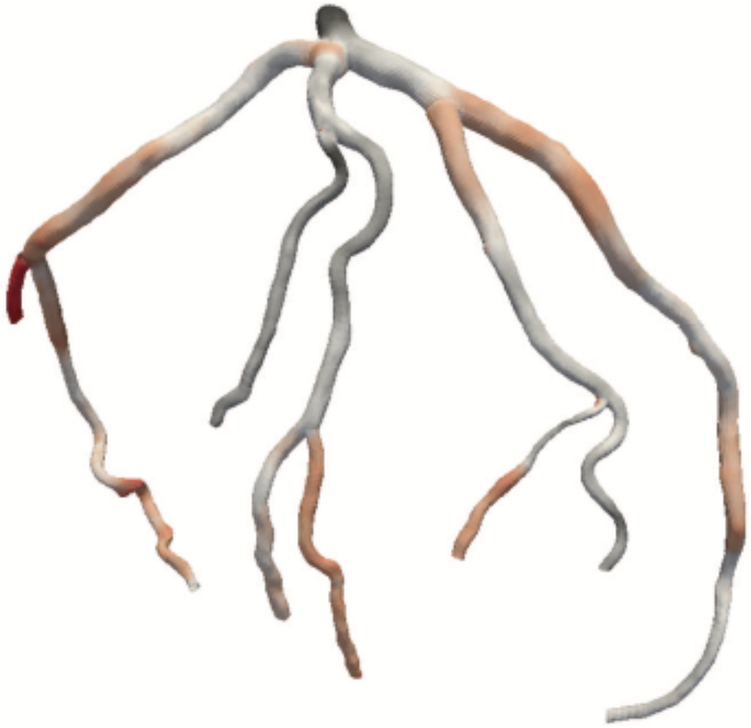
5. Apply knowledge of how people perceive and reason with visualizations in your designs.

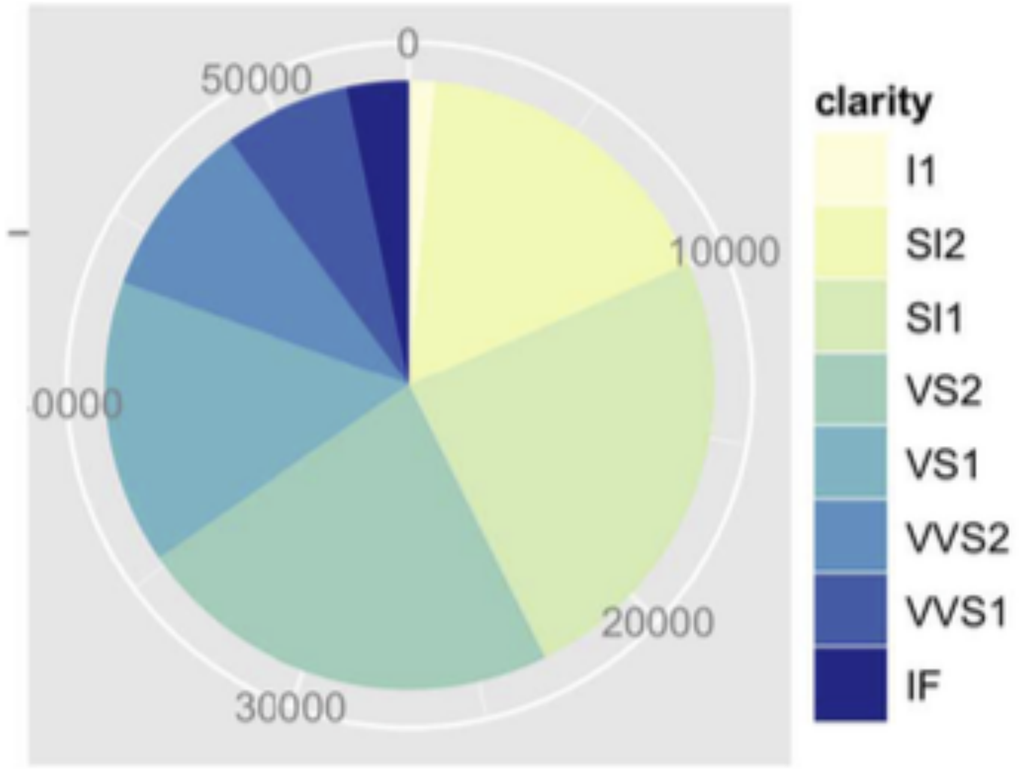
3
2
1
0
Shear
Stress (Pa)

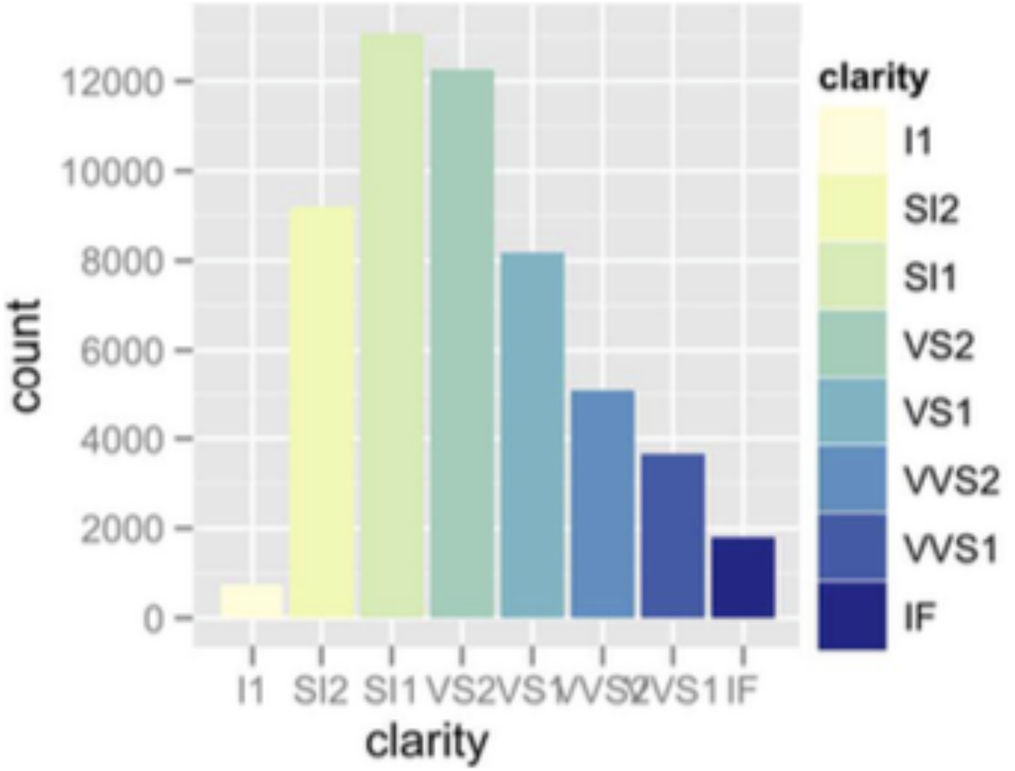
3
2
1
0
Shear
Stress (Pa)


3
2
1
0
Shear
Stress (Pa)






0
50000
10000
20000
30000
clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF


12000
10000
8000
6000
4000
2000
0
count
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF
clarity
clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF

6. Design and develop interactive data visualizations.

Domain situation
Observe target users using existing tools

Data/task abstraction

Visual encoding/interaction idiom
Justify design with respect to alternatives

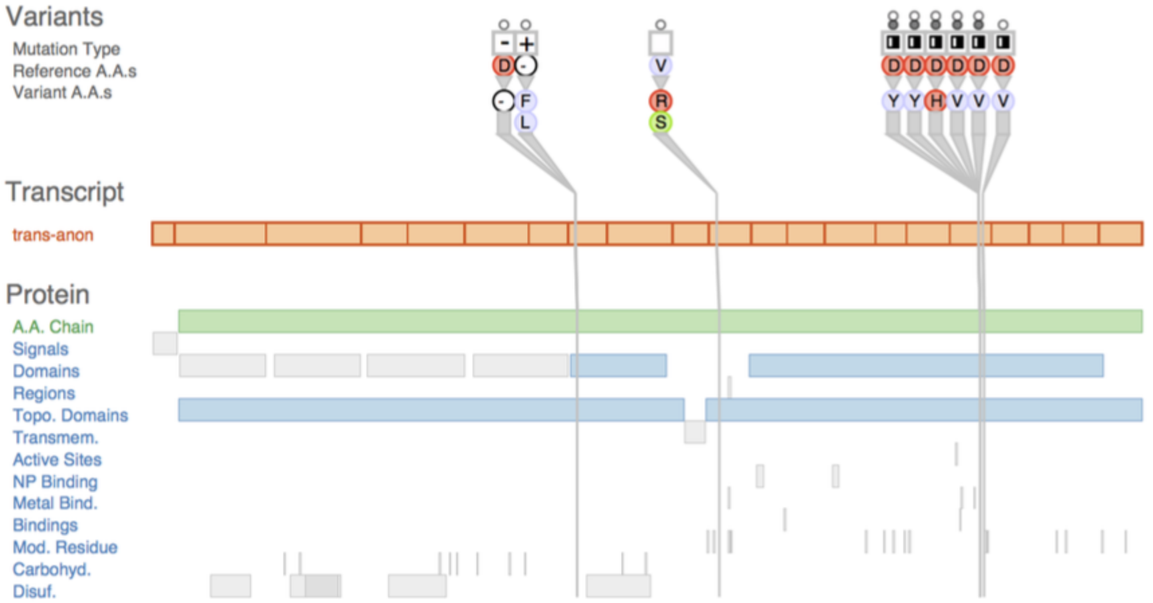
Algorithm
Measure system time/memory
Analyze computational complexity

Analyze results qualitatively

Measure human time with lab experiment (*user study*)

Observe target users after deployment (*field study*)

Measure adoption



Course Structure

Reading:

1-2 Chapters / wk

Lectures 1/wk

https://
cs573-16f.github.io/

Lab 1/wk

5 (ish)

Assignments

2

Exams

1

Final Project

Grading

40% assignments

20% labs

30% final project

10% exams

Attendance

Assignments

A0: Course Survey

A1: Hello World: GitHub and d3

A2

A3

A4

A5

TBA

A2 (visualization, ten ways)

A3 (interaction/animated transitions)

A4 (multiple views/server comm)

A5 (experiment)

100/120 for minimum requirements

writeup with every assignment:
+ 10 for Design justification
+ 10 for Technical achievement

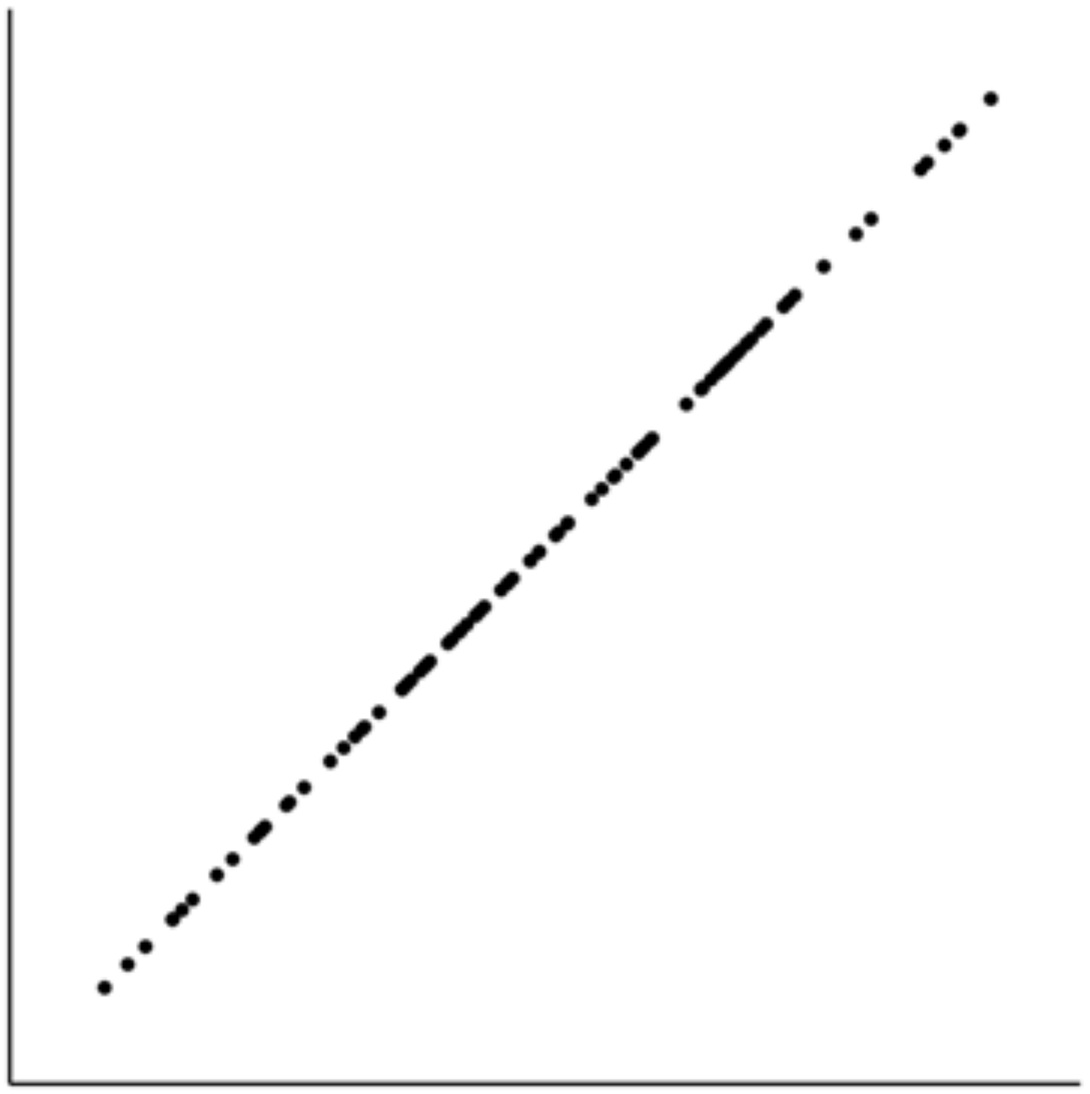
Labs

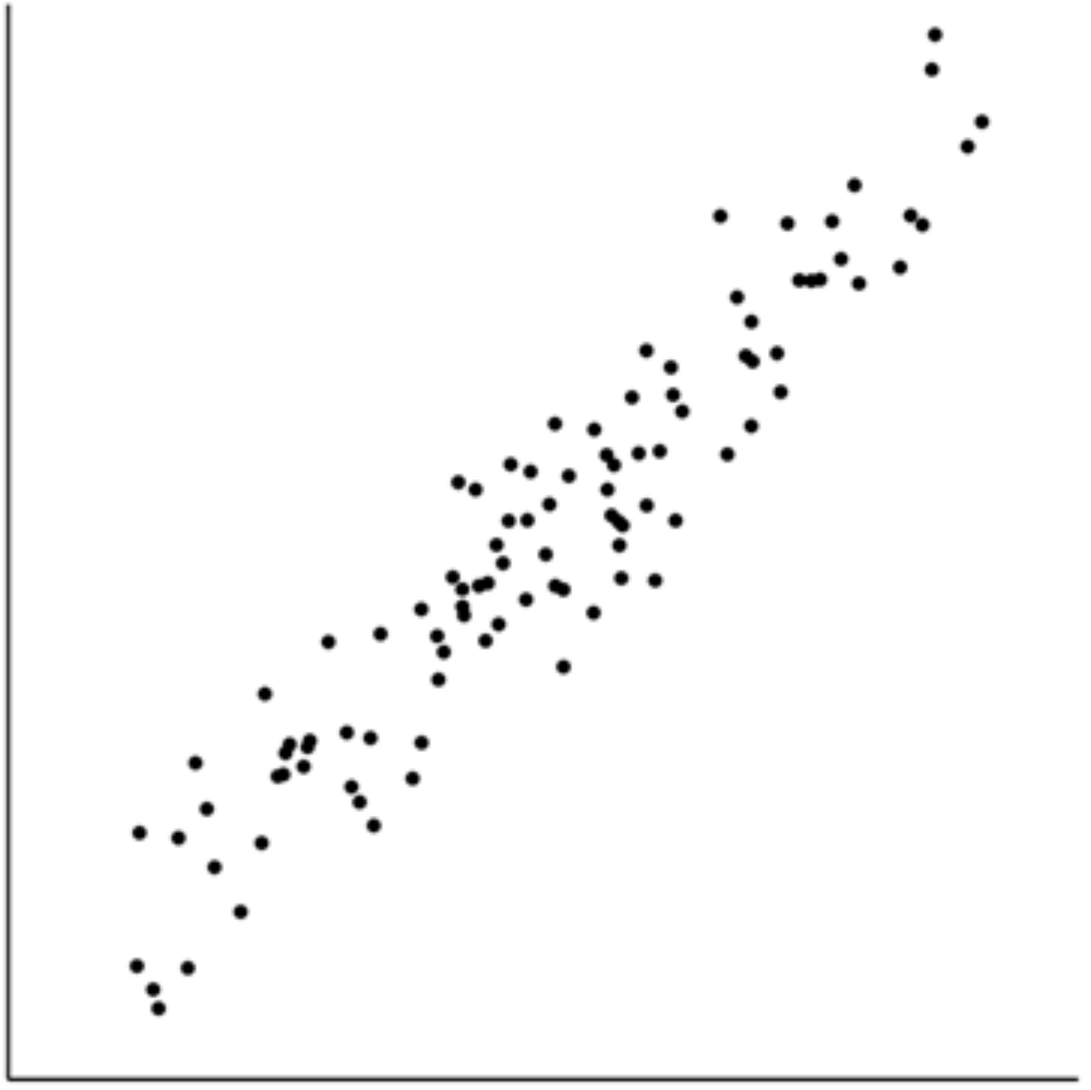
Final Project

Vis@WPI

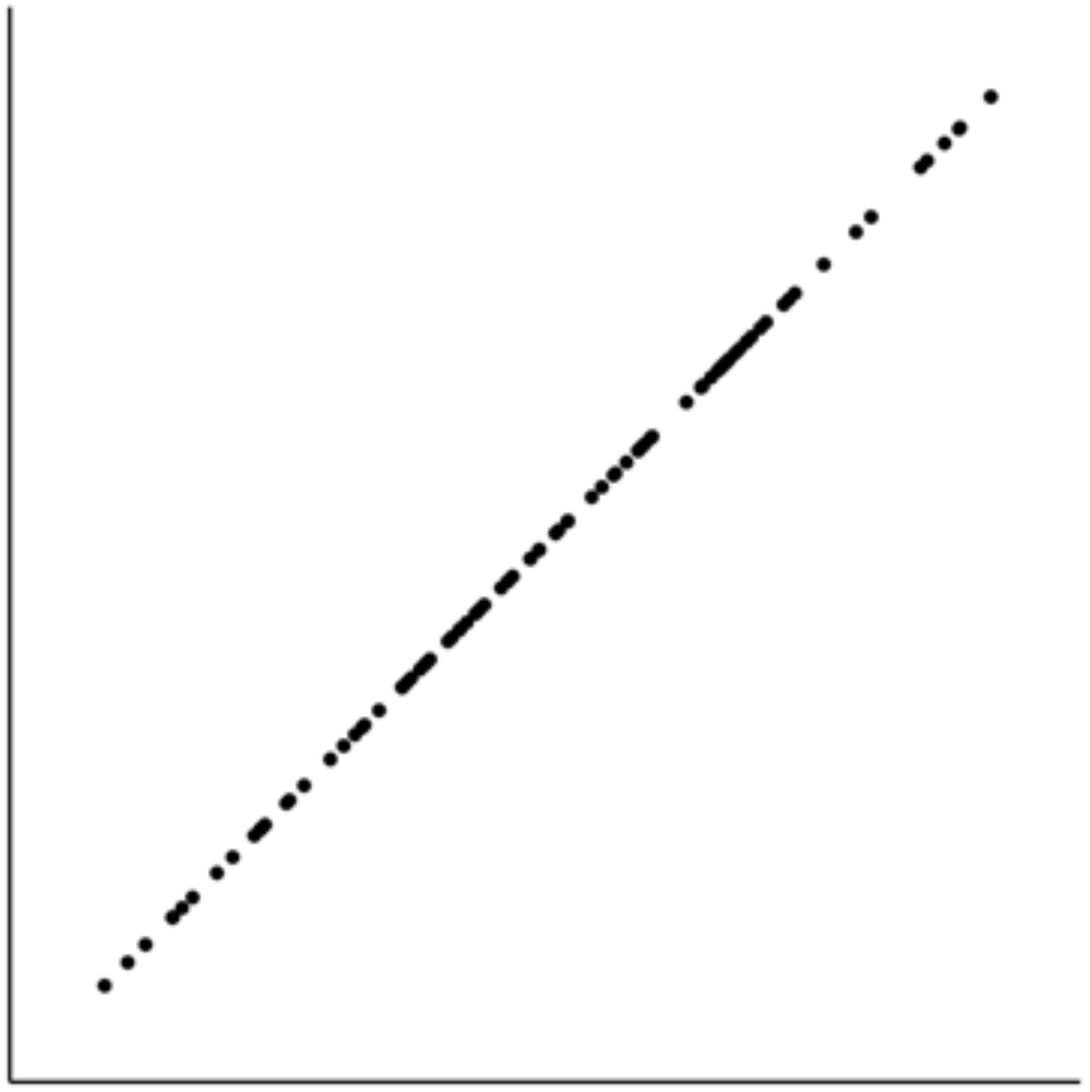
Research

Psychophysiological methodology: Which is more correlated?

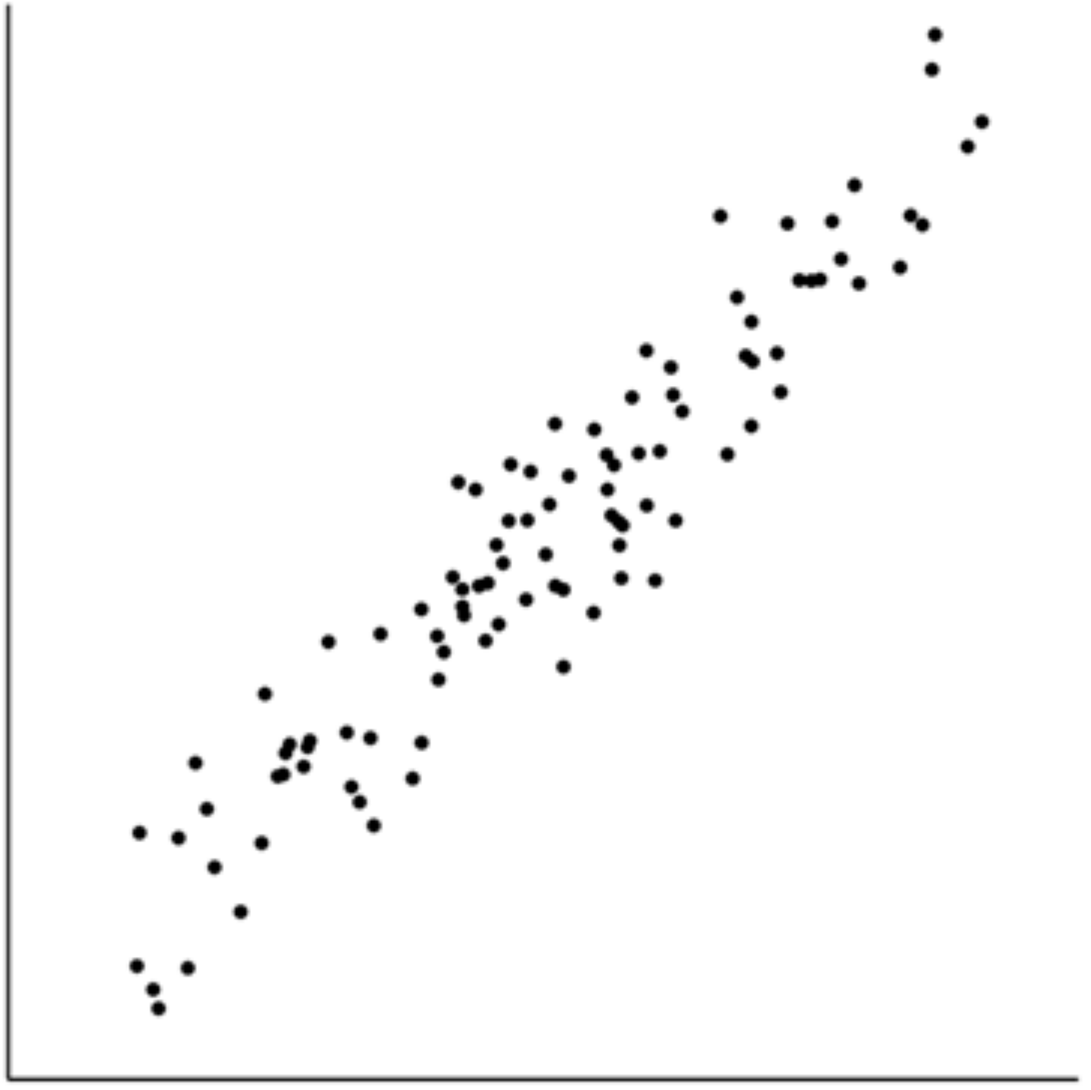


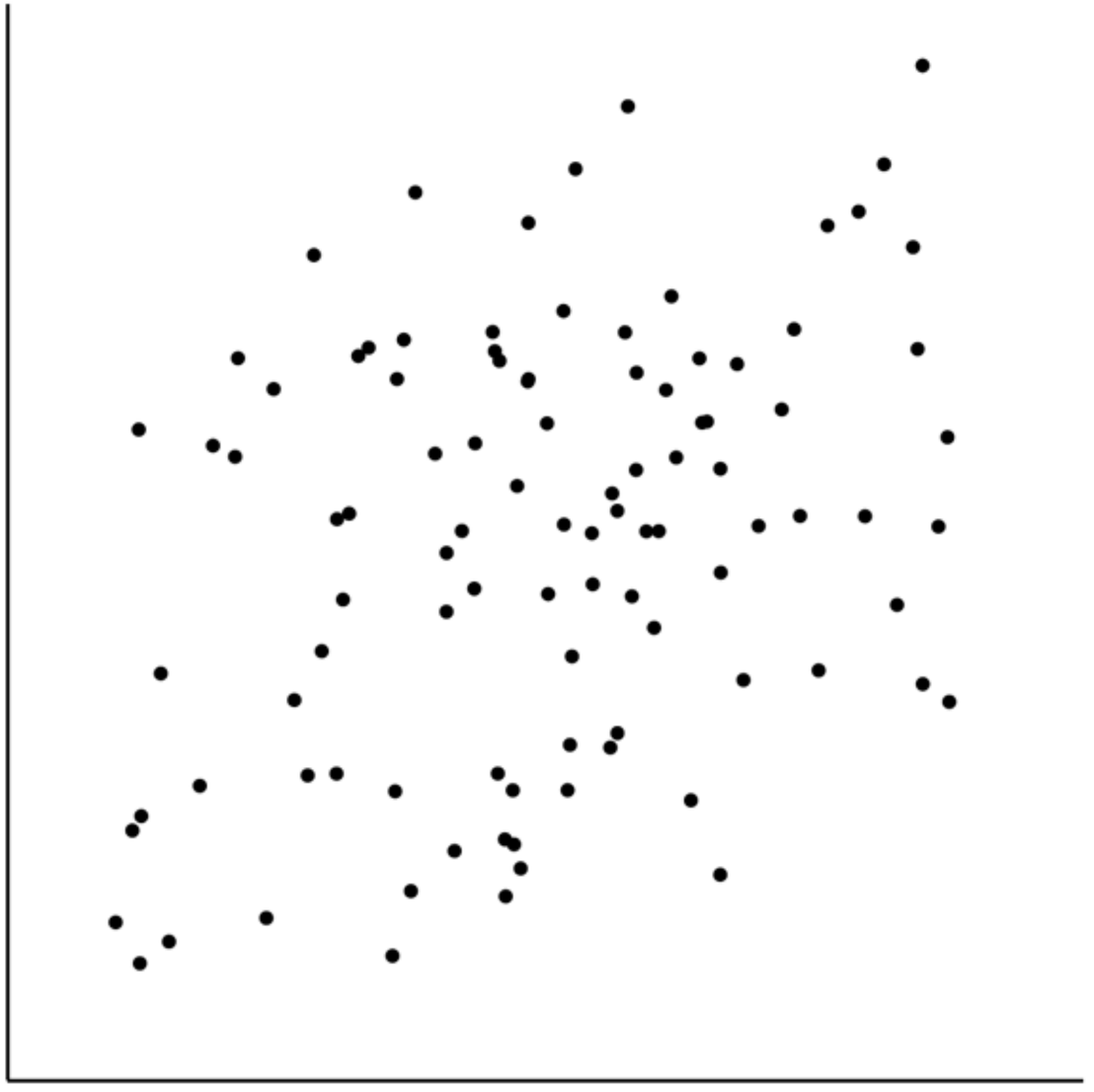


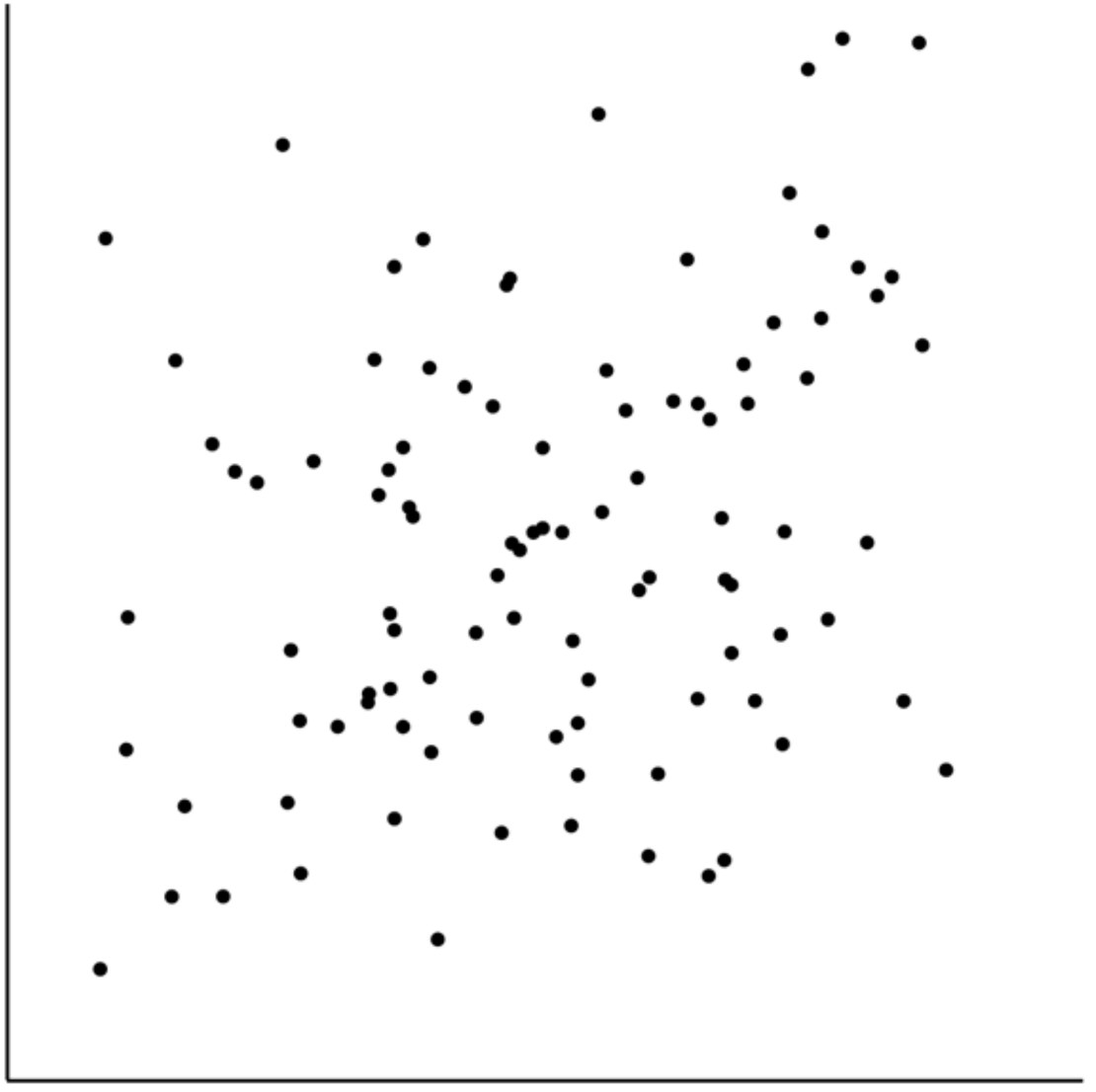
r = 1



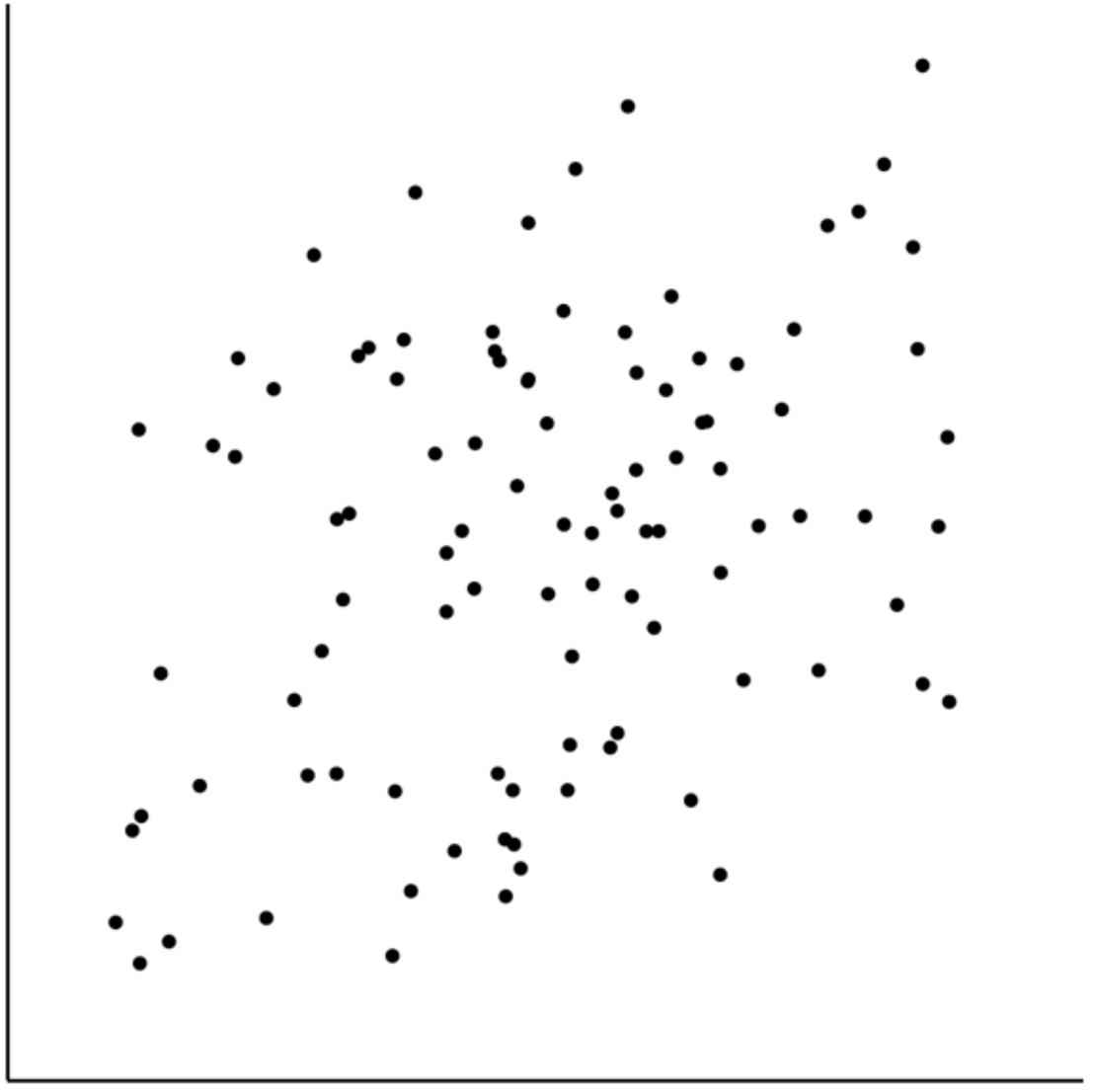
r = 0.95



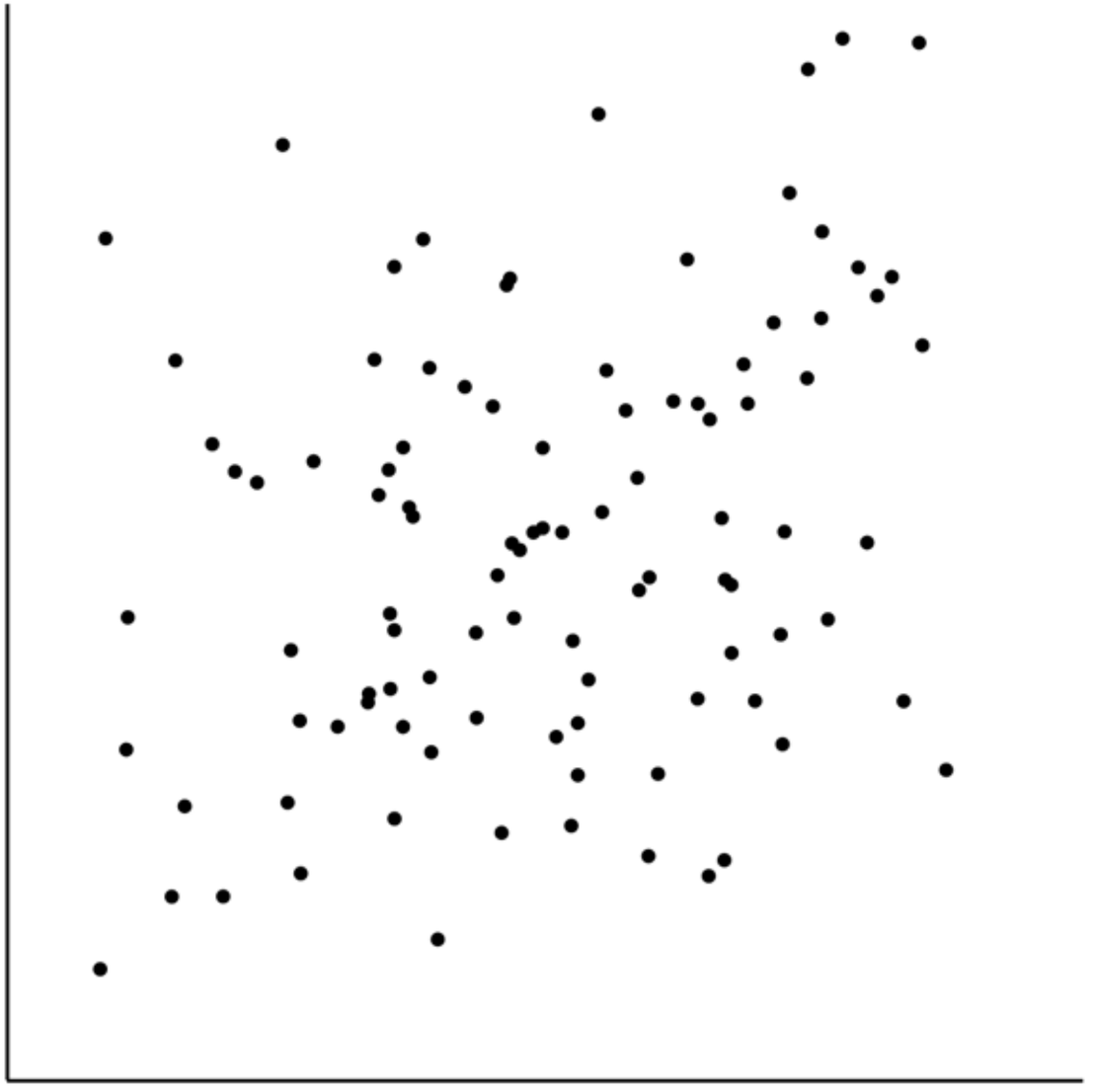




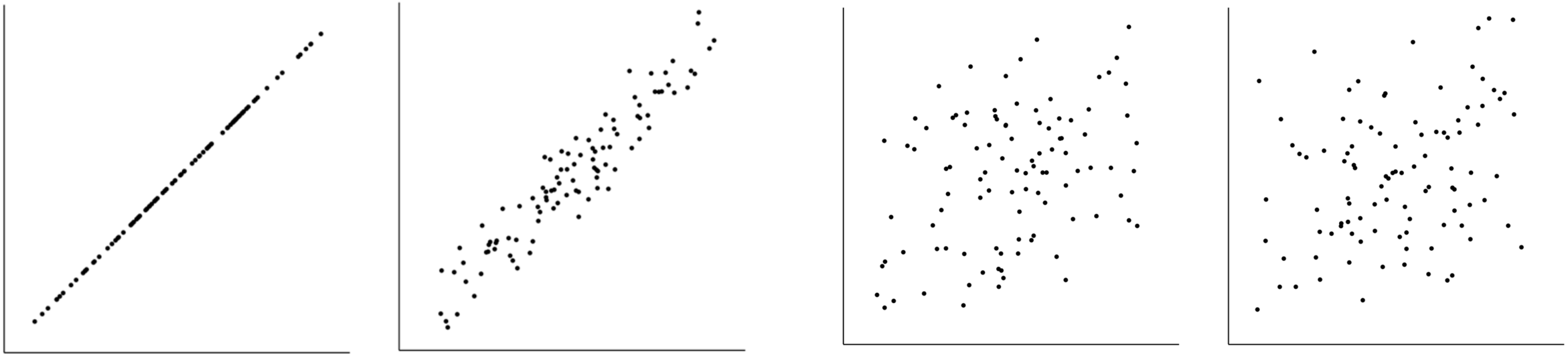
r = 0.35



r = 0.3



Same difference but harder, why?



1.0 *0.95* *0.35* *0.3*

d3 v3 -> v4

word of caution

Who should drop this course?

Let's do A1