INTRODUCTION TO DATA SCIENCE

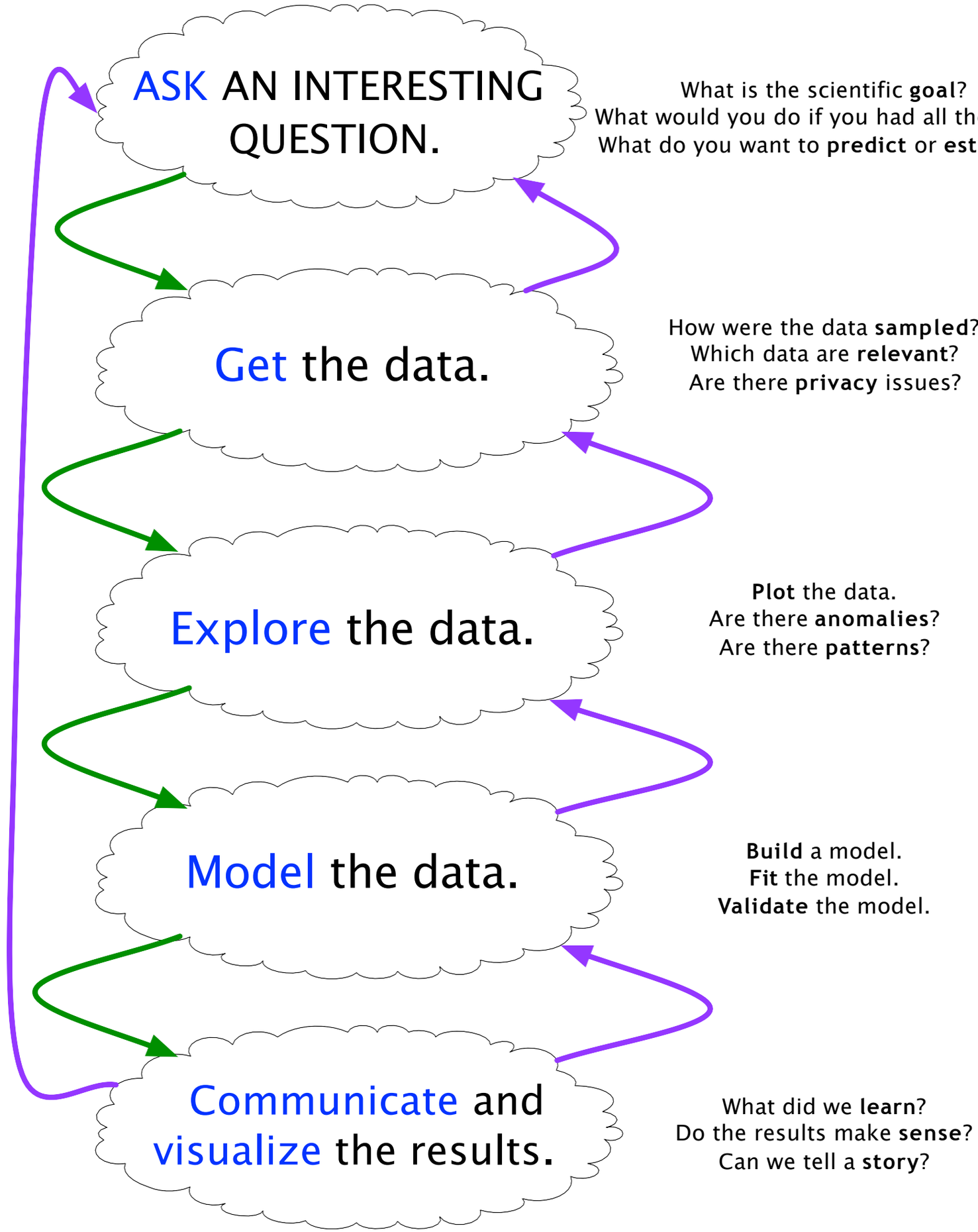
Lecture 2

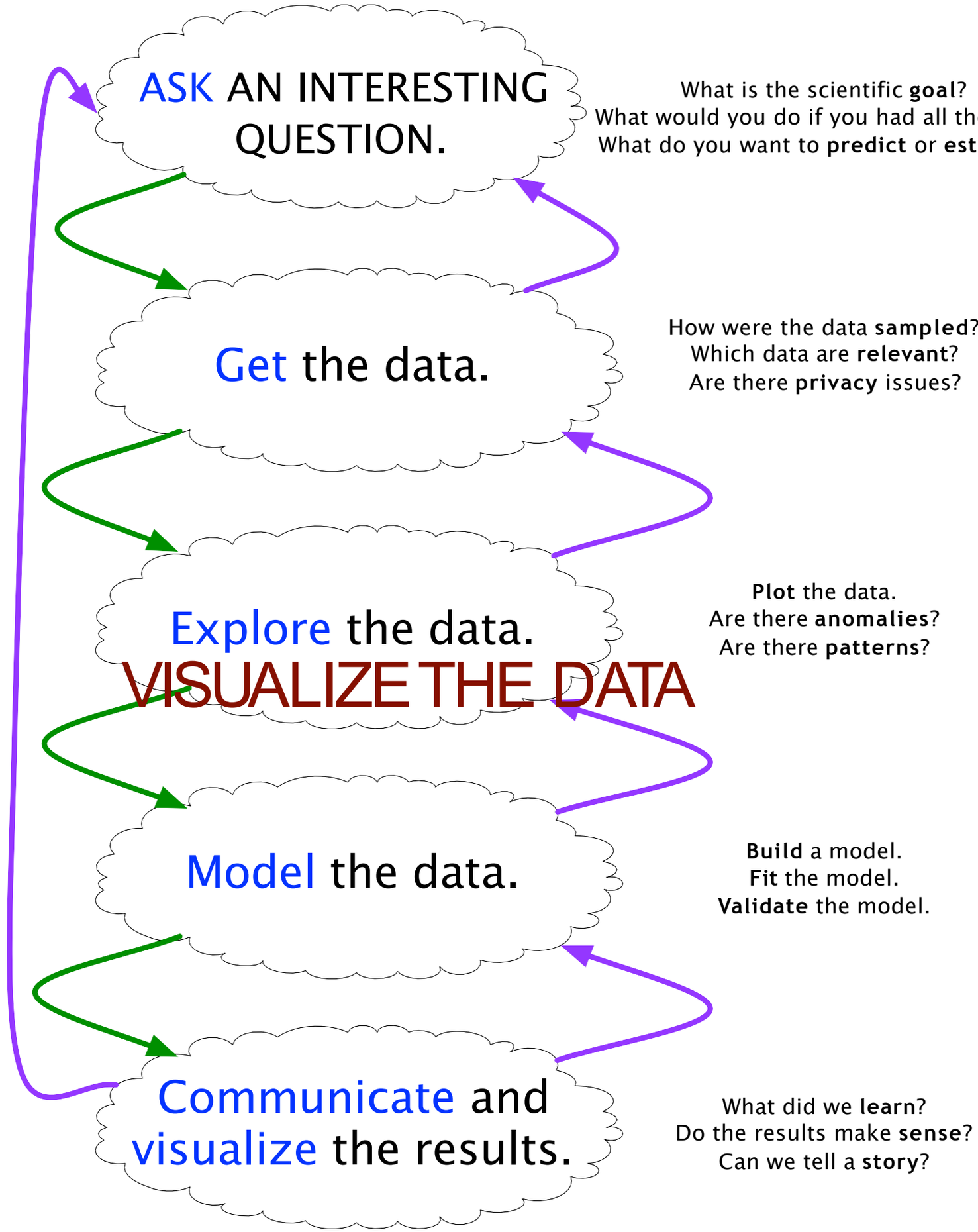
Effective Exploratory Data Analysis and Visualization

Data Science and Engineering Department
Faculty of Informatics
ELTE University

Zakarya Farou zakaryafarou@inf.elte.hu

THE DATA SCIENCE PROCESS


ASK AN INTERESTING QUESTION.
What is the scientific **goal**?
What would you do if you had all the **data**?
What do you want to **predict** or **estimate**?
Get the data.
How were the data **sampled**?
Which data are **relevant**?
Are there **privacy** issues?
Explore the data.
Plot the data.
Are there **anomalies**?
Are there **patterns**?
Model the data.
Build a model.
Fit the model.
Validate the model.
Communicate and visualize the results.
What did we **learn**?
Do the results make **sense**?
Can we tell a **story**?


ASK AN INTERESTING QUESTION.
What is the scientific **goal**?
What would you do if you had all the **data**?
What do you want to **predict** or **estimate**?
Get the data.
How were the data **sampled**?
Which data are **relevant**?
Are there **privacy** issues?
Explore the data.
VISUALIZE THE DATA
Plot the data.
Are there **anomalies**?
Are there **patterns**?
Model the data.
Build a model.
Fit the model.
Validate the model.
Communicate and visualize the results.
What did we **learn**?
Do the results make **sense**?
Can we tell a **story**?

Anscombe’s quartet

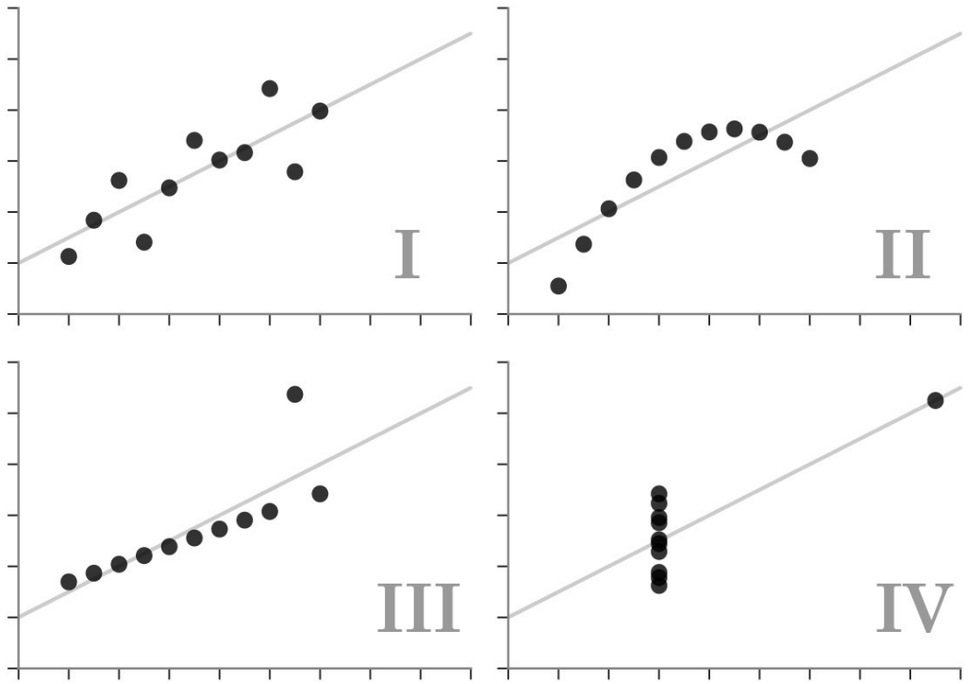
Anscombe's Data										
Observation	x1	y1		x2	y2		x3	y3		x4	y4
1	10	8.04		10	9.14		10	7.46		8	6.58
2	8	6.95		8	8.14		8	6.77		8	5.76
3	13	7.58		13	8.74		13	12.74		8	7.71
4	9	8.81		9	8.77		9	7.11		8	8.84
5	11	8.33		11	9.26		11	7.81		8	8.47
6	14	9.96		14	8.1		14	8.84		8	7.04
7	6	7.24		6	6.13		6	6.08		8	5.25
8	4	4.26		4	3.1		4	5.39		19	12.5
9	12	10.84		12	9.13		12	8.15		8	5.56
10	7	4.82		7	7.26		7	6.42		8	7.91
11	5	5.68		5	4.74		5	5.73		8	6.89
				Summary Statistics							
N	11	11		11	11		11	11		11	11
mean	9.00	7.50		9.00	7.500909		9.00	7.50		9.00	7.50
SD	3.16	1.94		3.16	1.94		3.16	1.94		3.16	1.94
r	0.82			0.82			0.82			0.82	

The statistical information for these four data sets are approximately similar!

Same Stats, Different Graphs

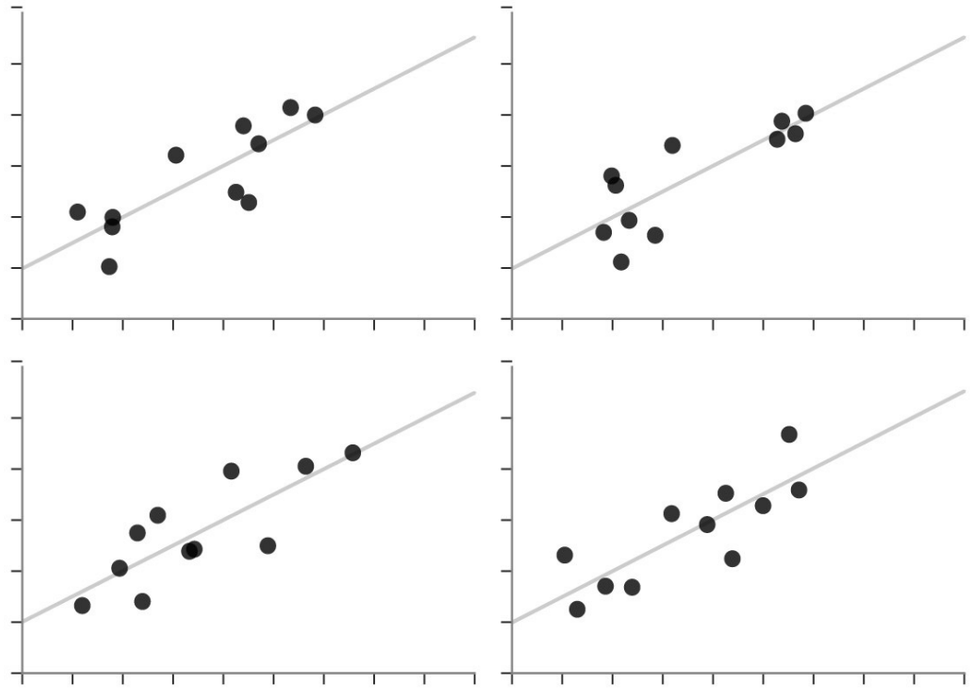
✔ Anscombe's Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.



✖ Unstructured Quartet

Each dataset here also has the same summary statistics. However, they are not *clearly different* or *visually distinct*.



ANSCOMBE'S QUARTET FOUR DATASETS

- **DS1:** fits the linear regression model pretty well.
- **DS2:** cannot fit the linear regression model because the data is non-linear.
- **DS3 – DS4:** shows the outliers involved in the data set, which the linear regression model cannot handle.
- Anscombe's quartet helps us to understand the importance of data visualization and how easy it is to fool a regression algorithm.
- Before attempting to interpret and model the data or implement any machine learning algorithm, we first need to visualize the data set to help build a well-fit model.

https://www.autodeskresearch.com/publications/samestats

EXAMPLE: ANTIBIOTICS

WILL BURTIN, 1951

DATA

Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

DATA

Genus, *Species*

Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

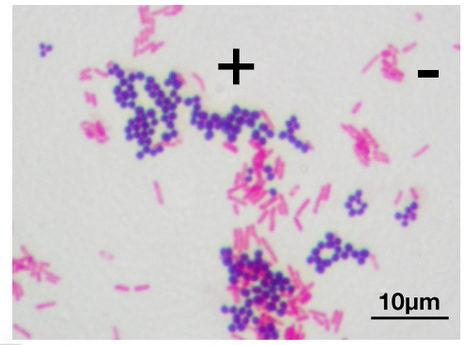
DATA

Genus, *Species*

Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

DATA

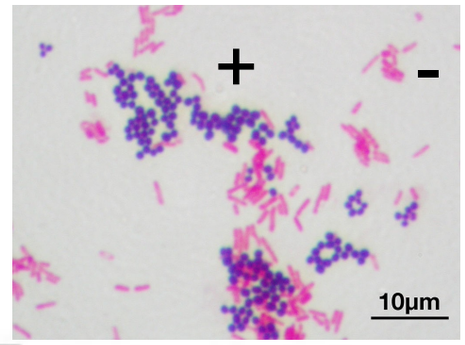


Genus, *Species*

Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

DATA



Genus, *Species*

Min. Inhibitory Concentration [ml/g]

Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

WHAT QUESTIONS ?

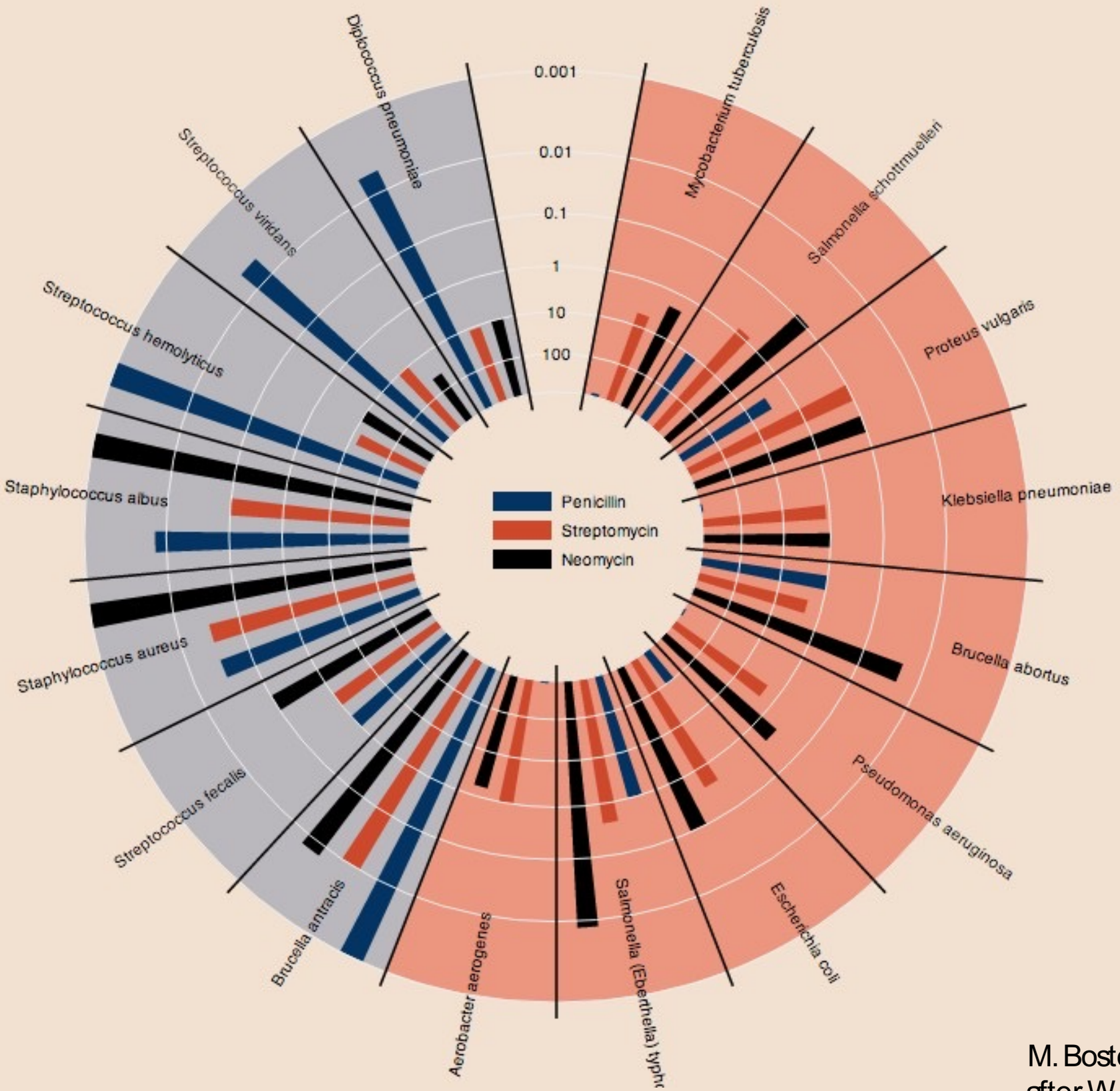
Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

HOW EFFECTIVE ARE THE DRUGS?

Gram Positive

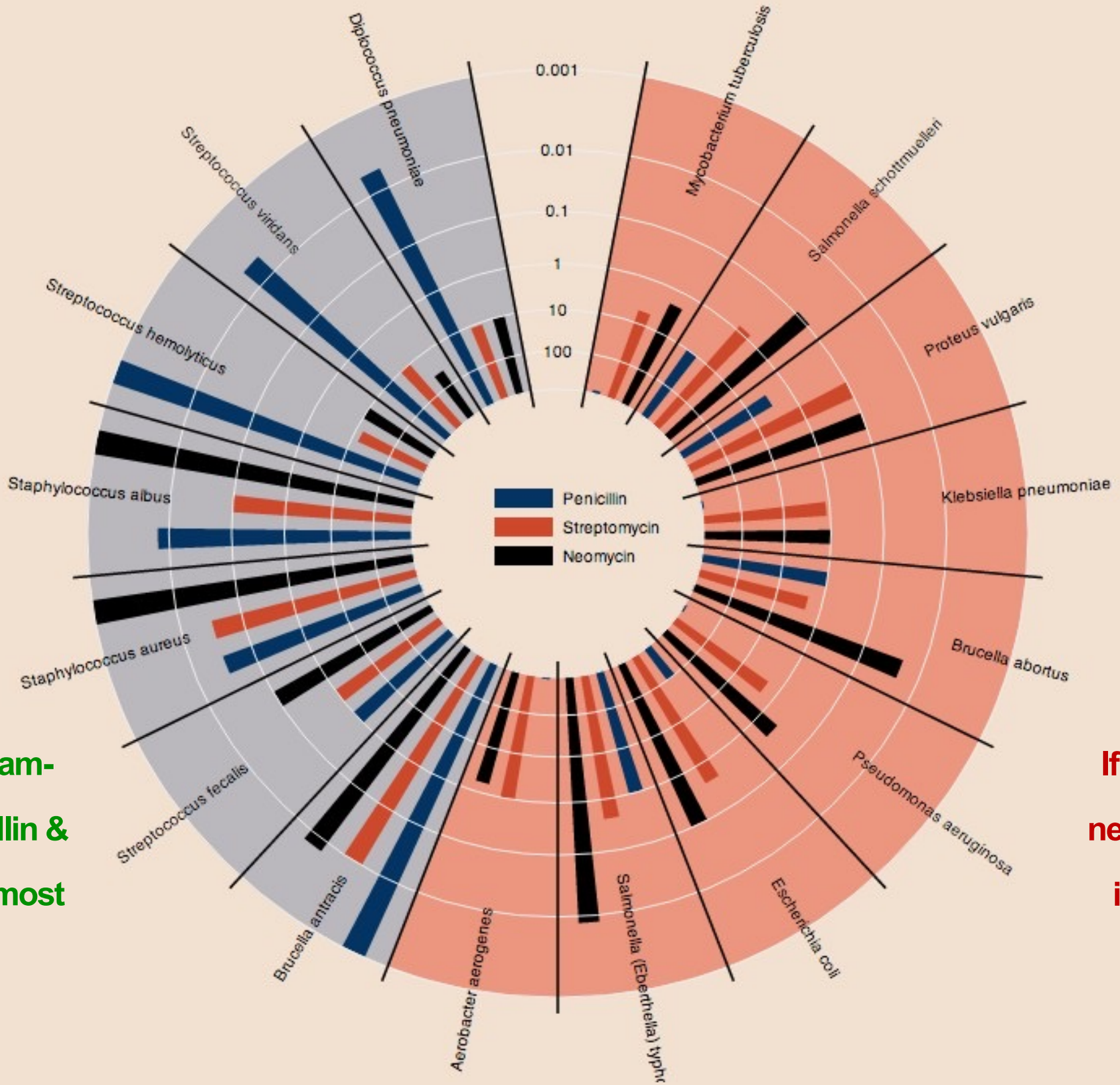
Gram Negative



M. Bostock, Protovis
after W. Burtin, 1951

HOW EFFECTIVE ARE THE DRUGS?

Gram Positive



Gram Negative

If bacteria is gram-positive, Penicillin & Neomycin are most effective.

If bacteria is gram-negative, Neomycin is most effective.

EXPLORATORY DATA ANALYSIS

"The greatest value of a picture is when it forces us to notice what we never expected to see."

John Tukey

VISUALIZATION GOALS

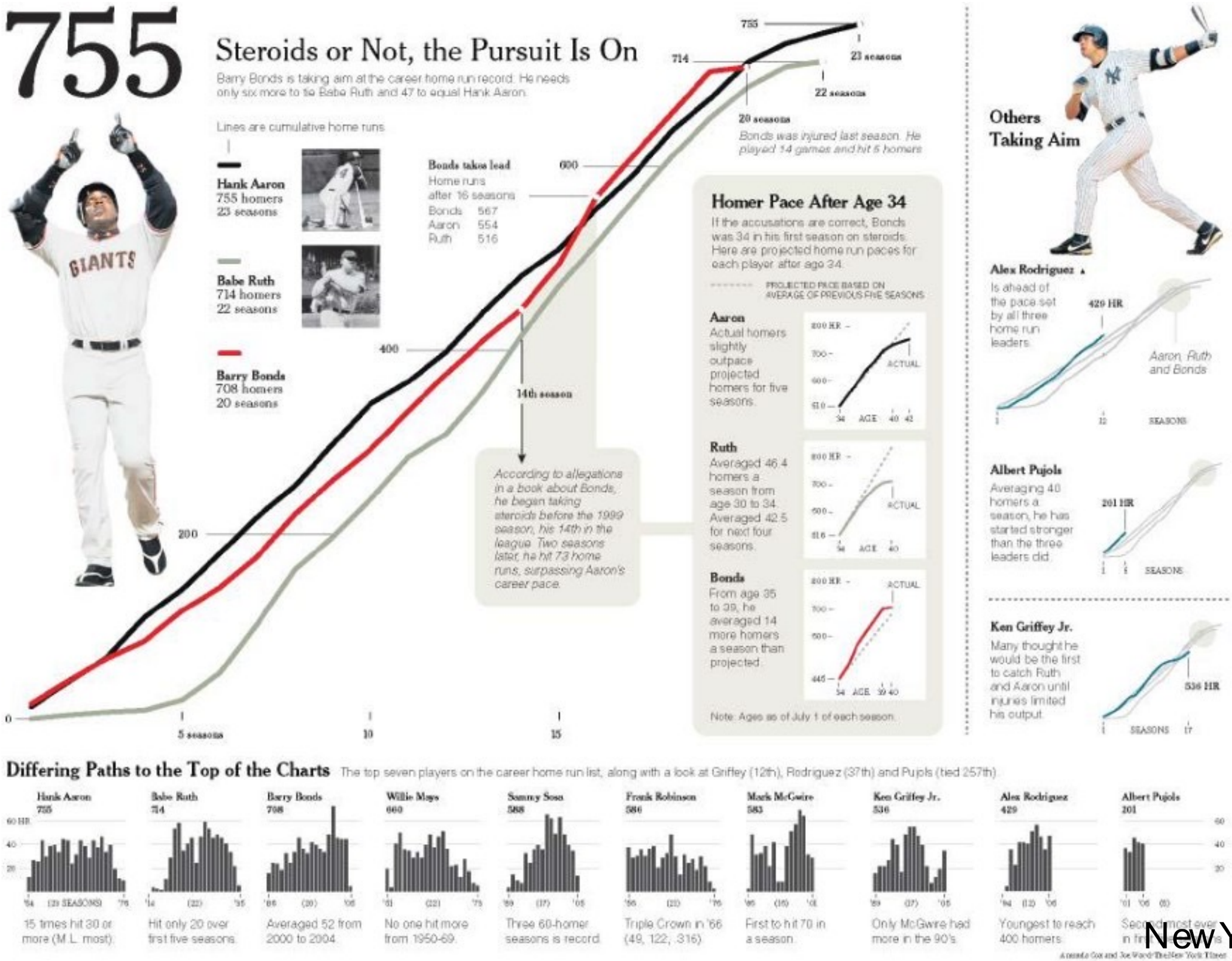
Communicate (Explanatory)

- Present data and ideas
- Explain and inform
- Provide evidence and support
- Influence and persuade

Analyze (Exploratory)

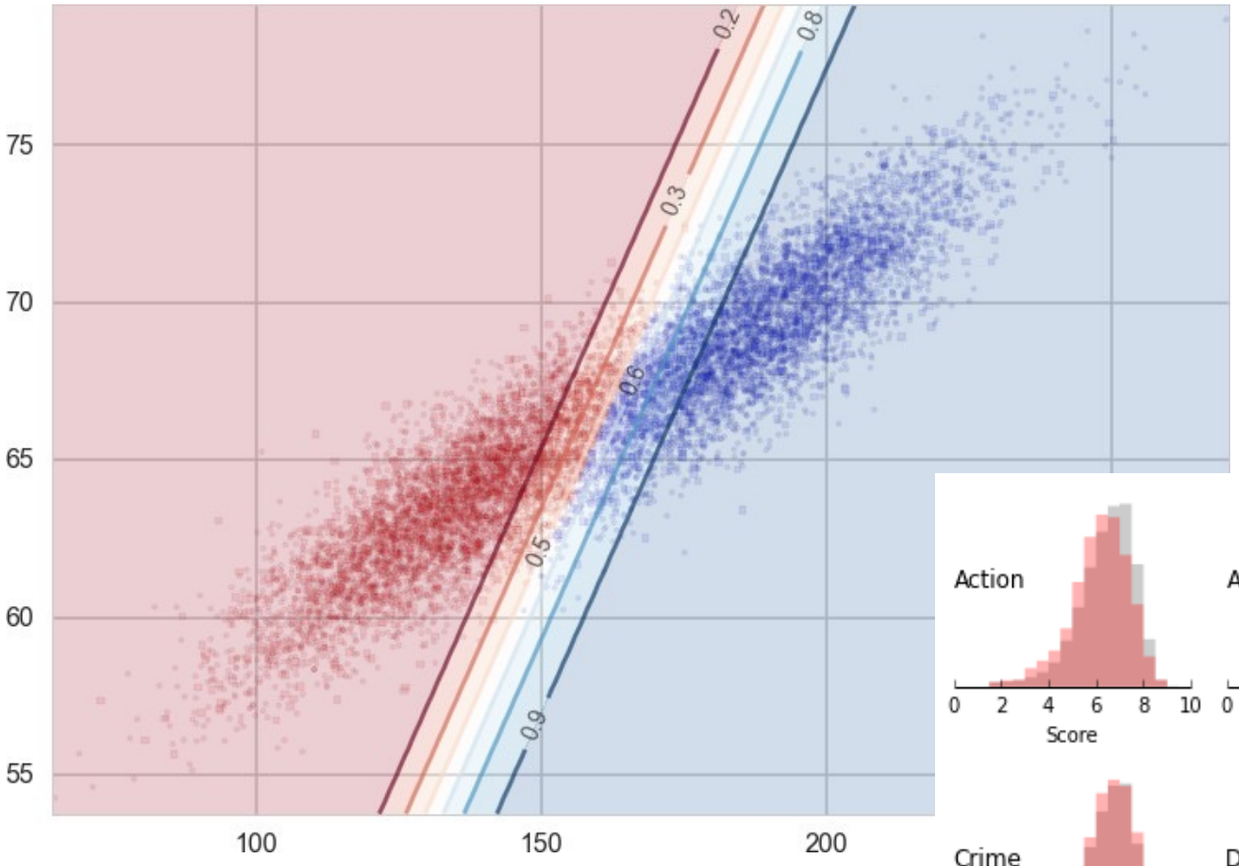
- Explore the data
- Assess a situation
- Determine how to proceed
- Decide what to do

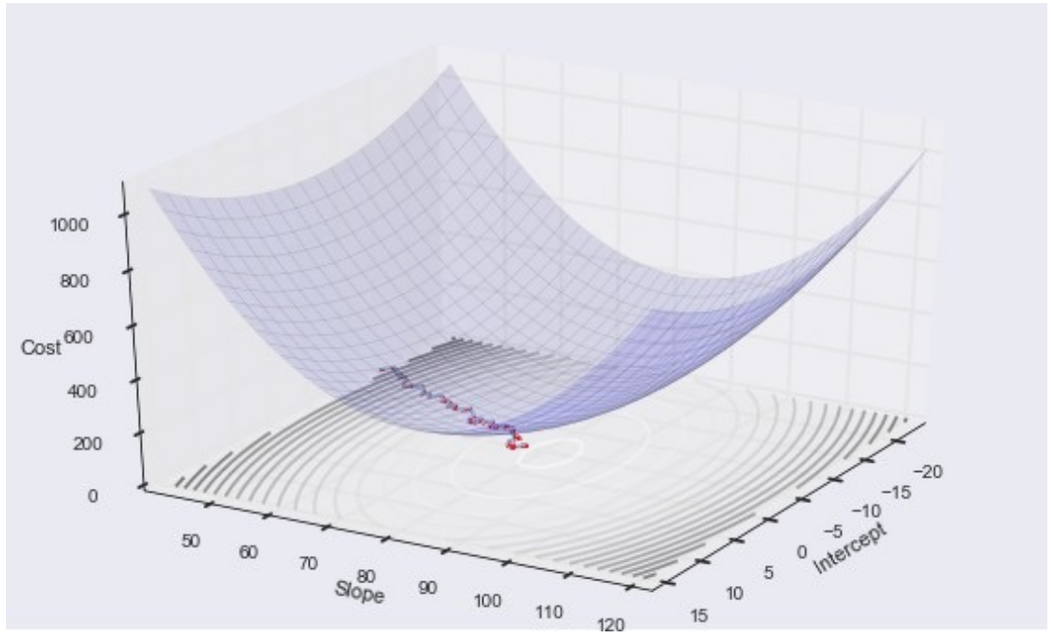
COMMUNICATE

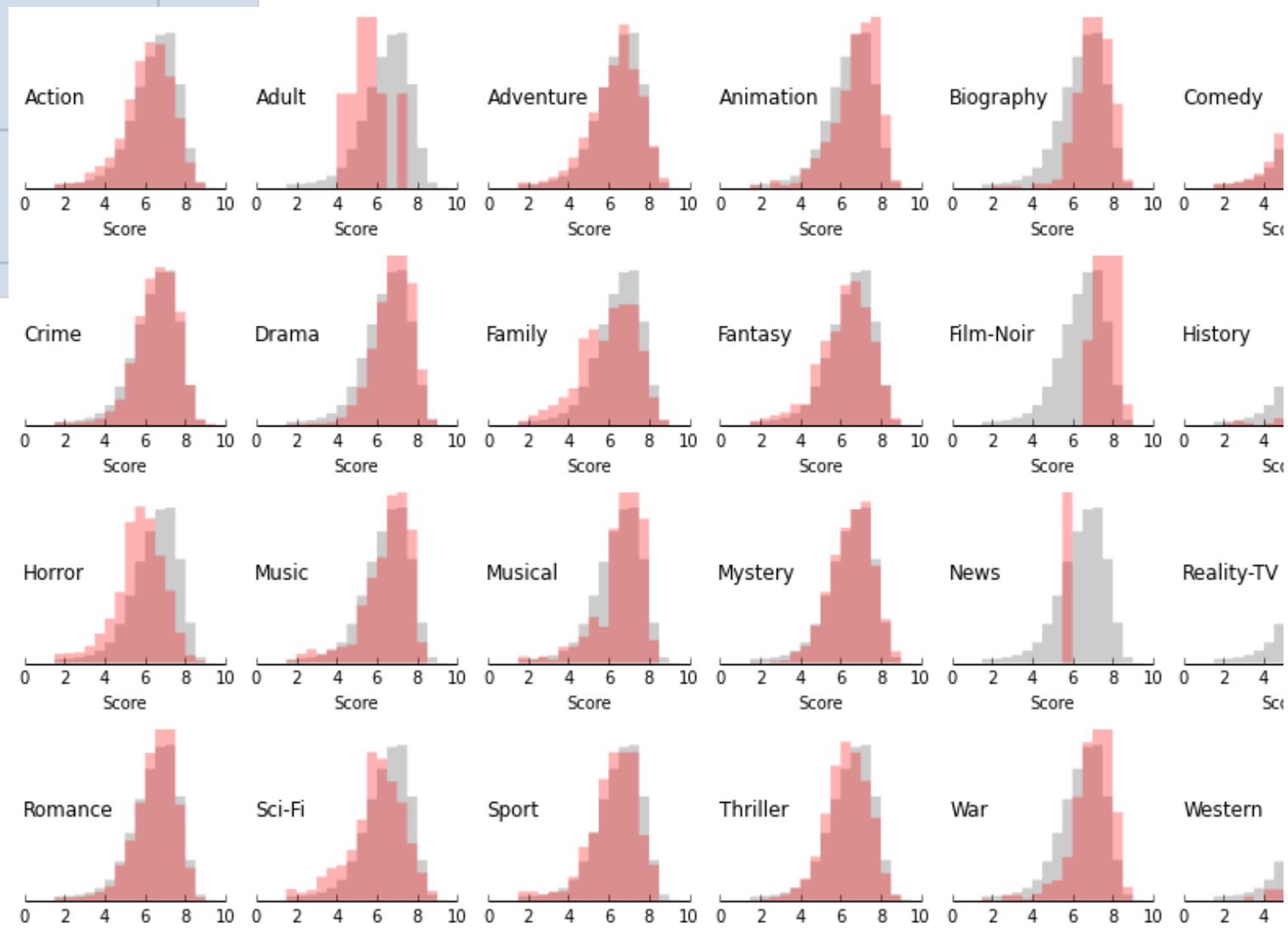


New York Times

EXPLORE


0.2
0.3
0.5
0.6
0.8
0.9
75
70
65
60
55
100
150
200


Cost
1000
800
600
400
200
0
50
60
70
80
90
100
110
120
Slope
15
10
5
0
−5
−10
−15
−20
Intercept


Action
Adult
Adventure
Animation
Biography
Comedy
Crime
Drama
Family
Fantasy
Film-Noir
History
Horror
Music
Musical
Mystery
News
Reality-TV
Romance
Sci-Fi
Sport
Thriller
War
Western
0 2 4 6 8 10
Score

EDA WORKFLOW

1. **Build** a DataFrame from the data (ideally, put all data in this object)
2. **Clean** the DataFrame. It should have the following properties
    - Each row describes a single object
    - Each column describes a property of that object
    - Columns are numeric whenever appropriate
    - Columns contain atomic properties that cannot be further decomposed
3. Explore **global properties**. Use histograms, scatter plots, and aggregation functions to summarize the data.
4. Explore **group properties**. Use groupby and small multiples to compare subsets of the data.

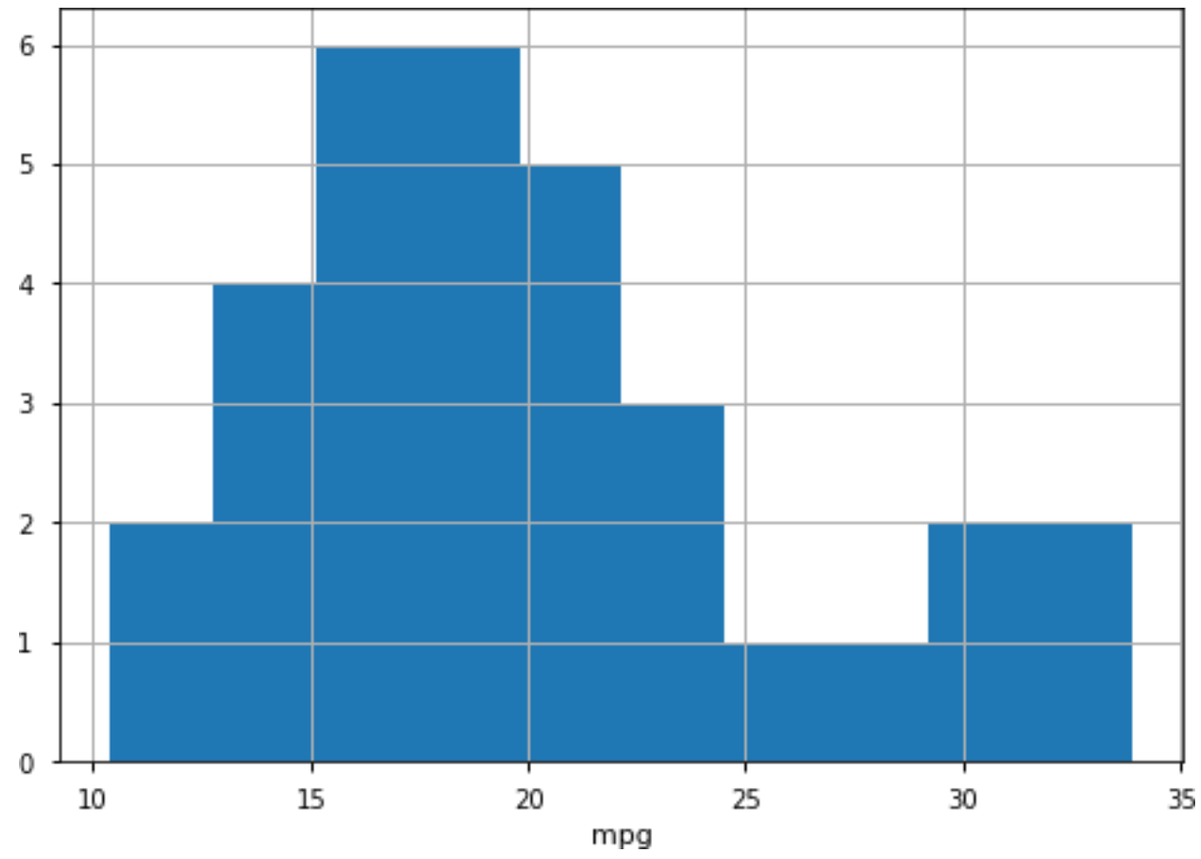
VISUALIZATION OPTIONS

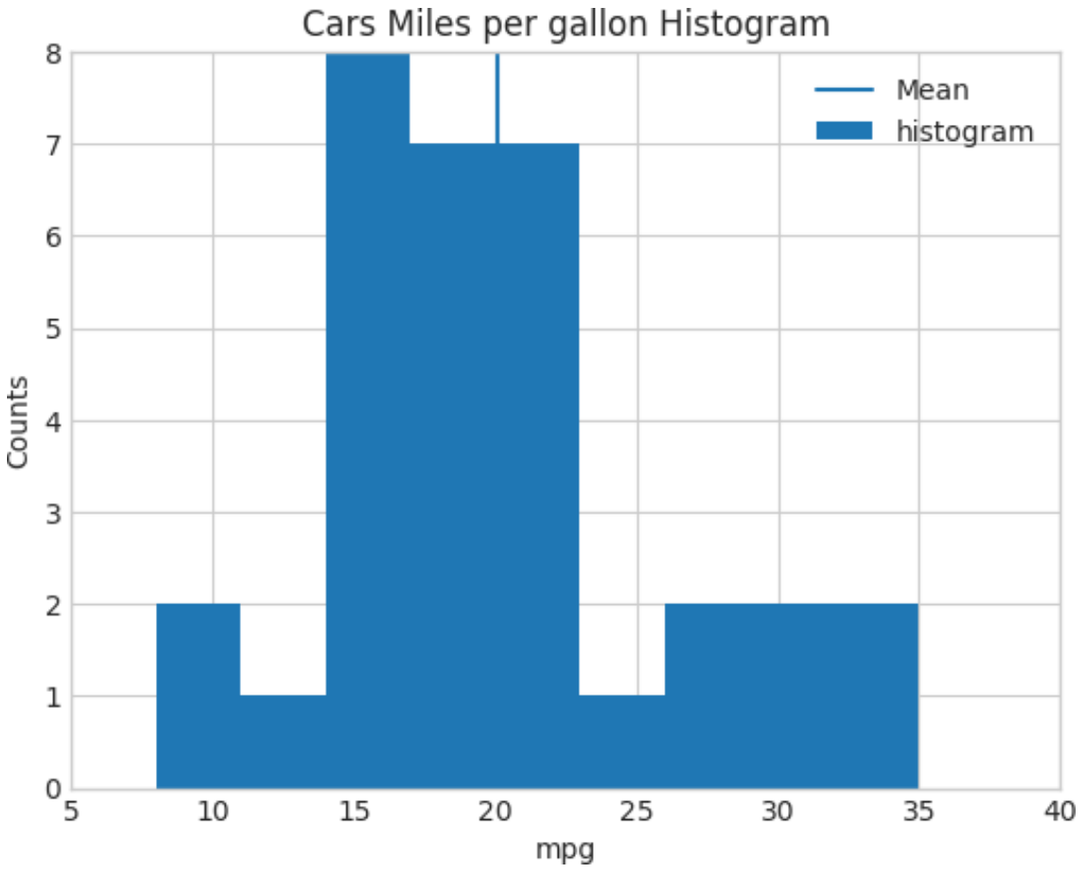
- Pandas Visualization module
- Matplotlib
- Seaborn
- The above three are inter-mixable
- Other options: Bokeh, Vega, Vincent, Altair

CARS DATASET

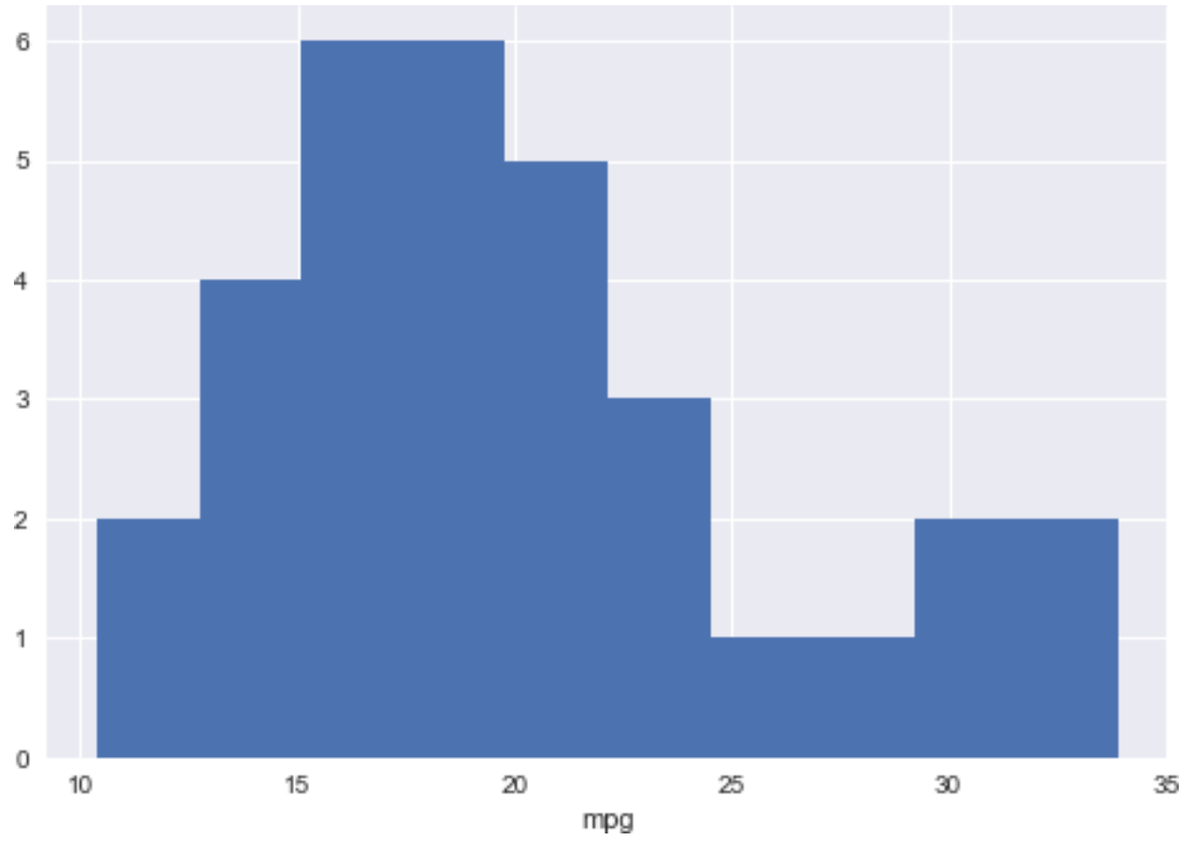
	name	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb	maker
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4	Mazda
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4	Mazda
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1	Datsun
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1	Hornet
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2	Hornet

Basic Pandas/matplotlib

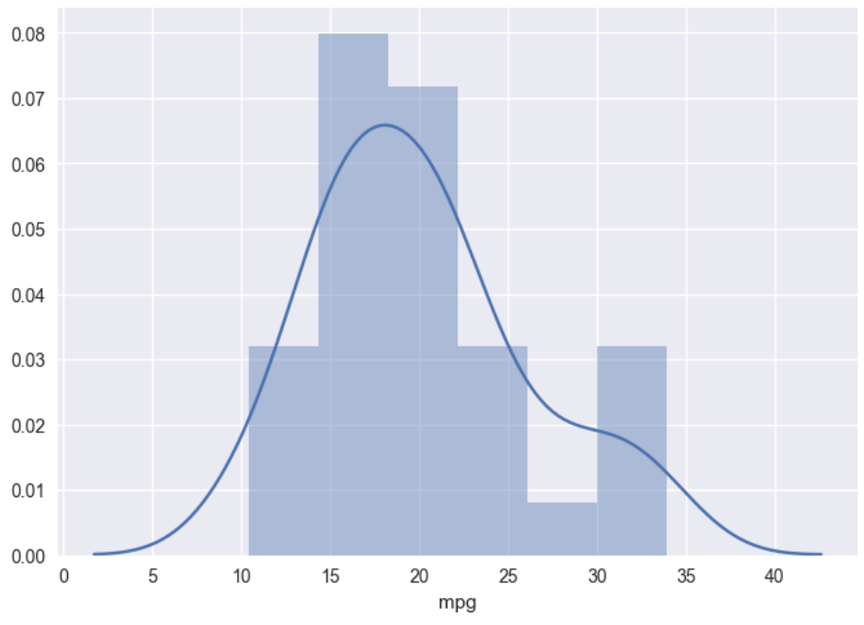




Can set limits, tick styles, scales, add lines, annotations, titles, legends

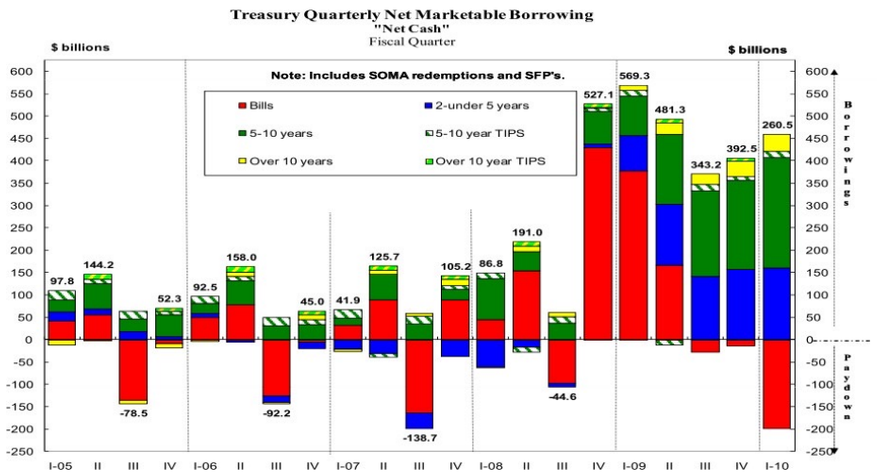


Seaborn provides a different visual style and lots of canned plots.



EFFECTIVE VISUALIZATIONS

NOT EFFECTIVE...



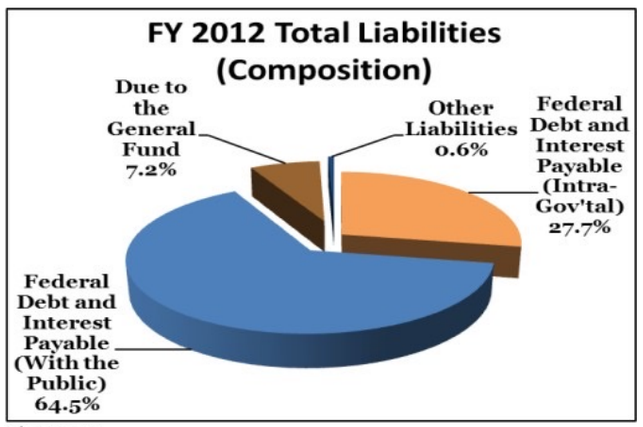
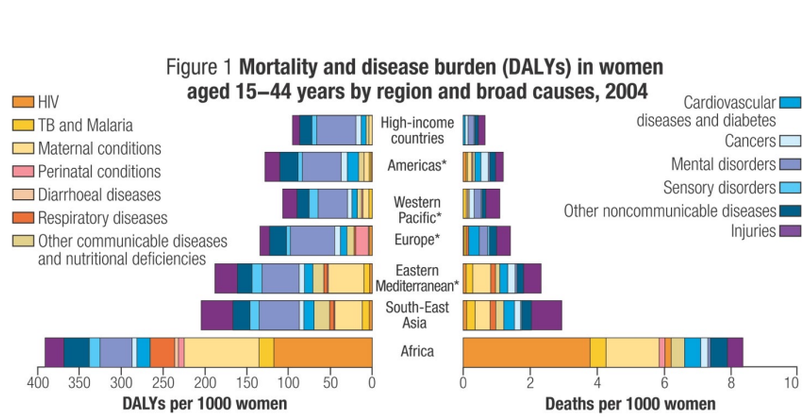
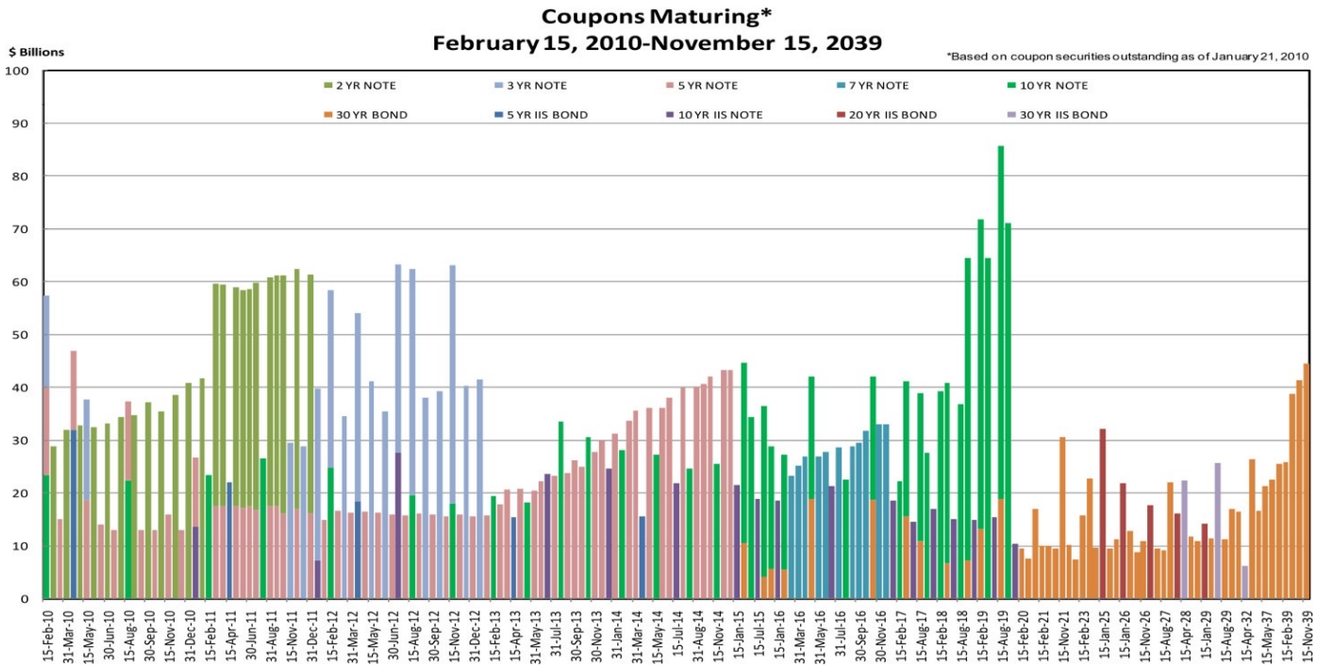


Figure 10





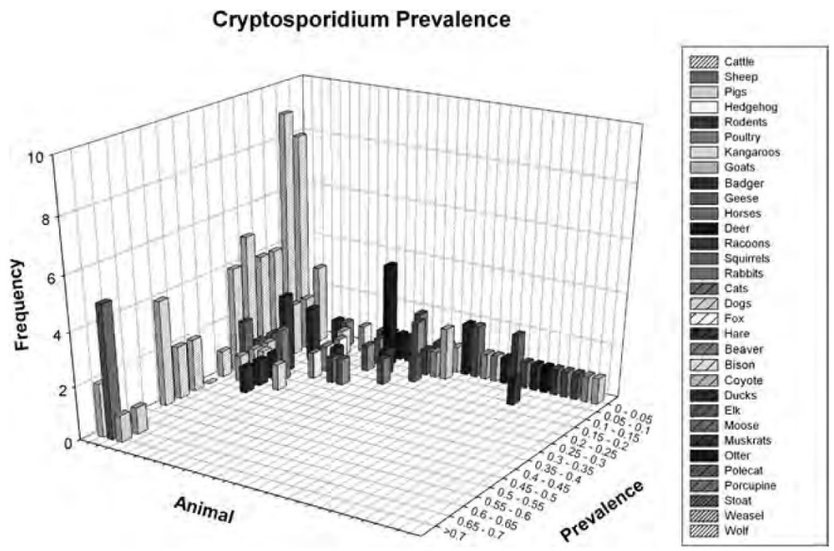


Figure 5.2 Mean prevalence rates of *Cryptosporidium* oocysts by animal species.

Sources: US Treasury and WHO reports

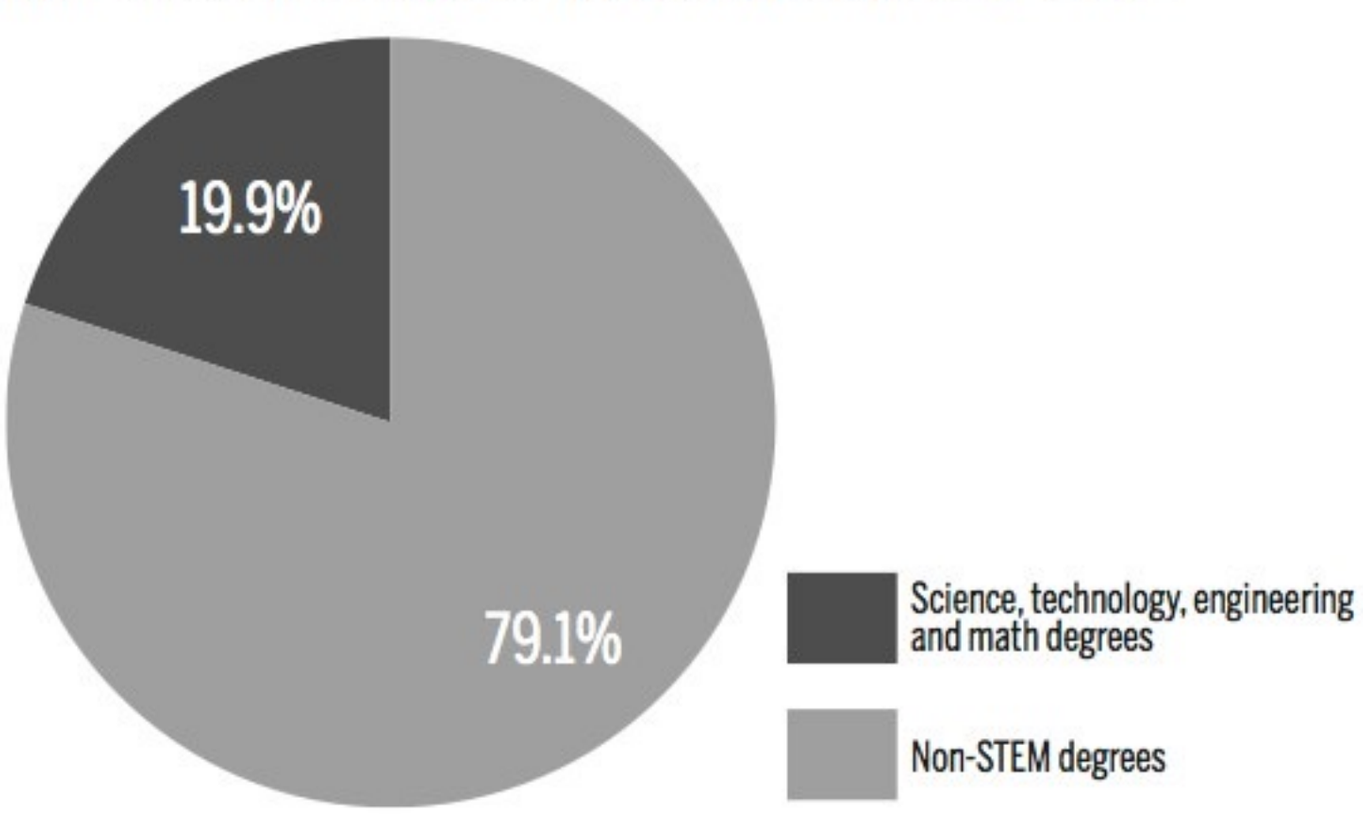
EFFECTIVE EDA VISUALIZATION

1. Have graphical integrity
2. Keep it simple
3. Use the right display
4. Use color sensibly

1. GRAPHICAL INTEGRITY

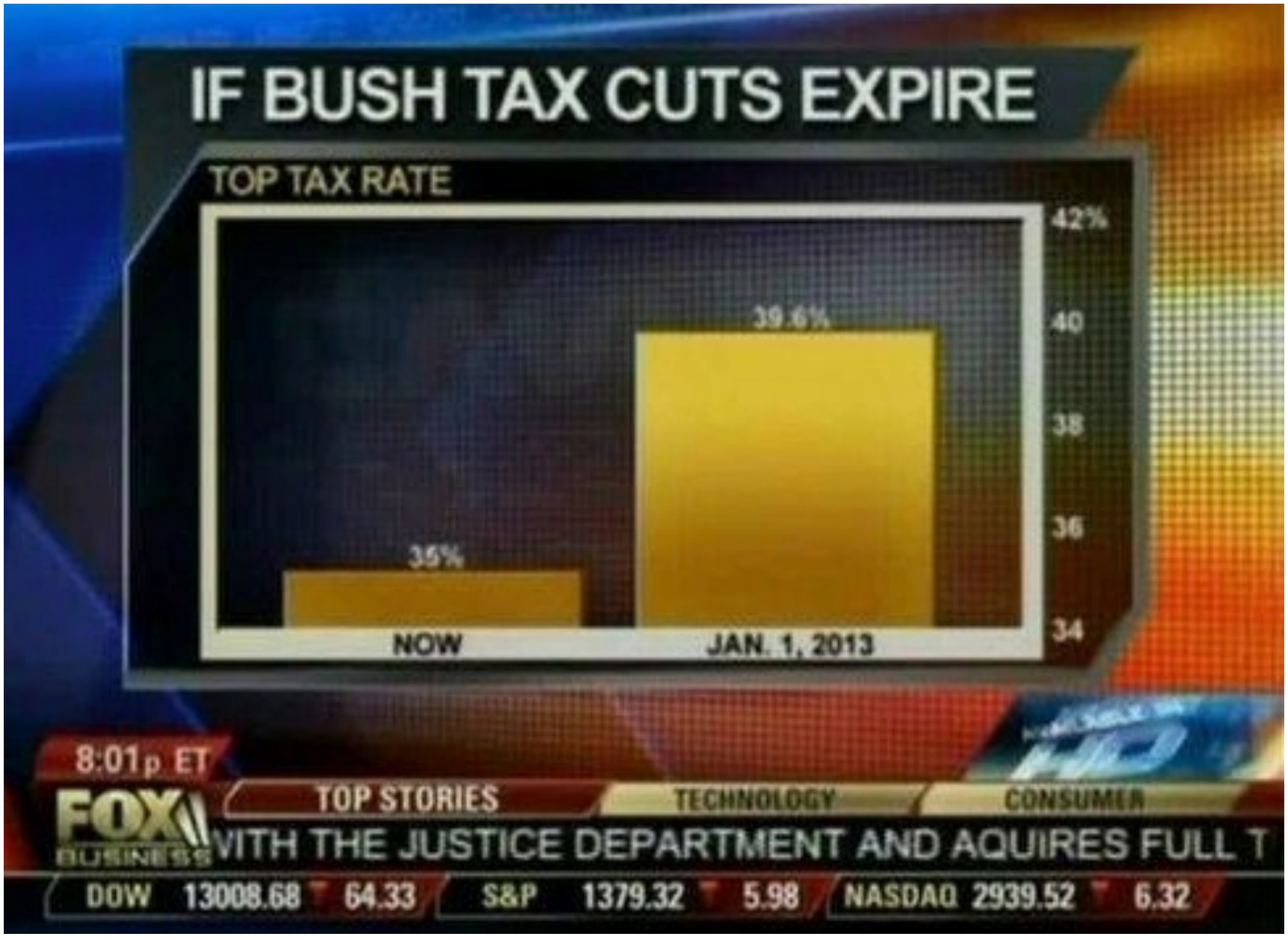
Yale Daily News | THE OLDEST COLLEGE DAILY ESTABLISHED 1878

CHART YALE GRADUATES' MAJORS, CLASS OF 2011

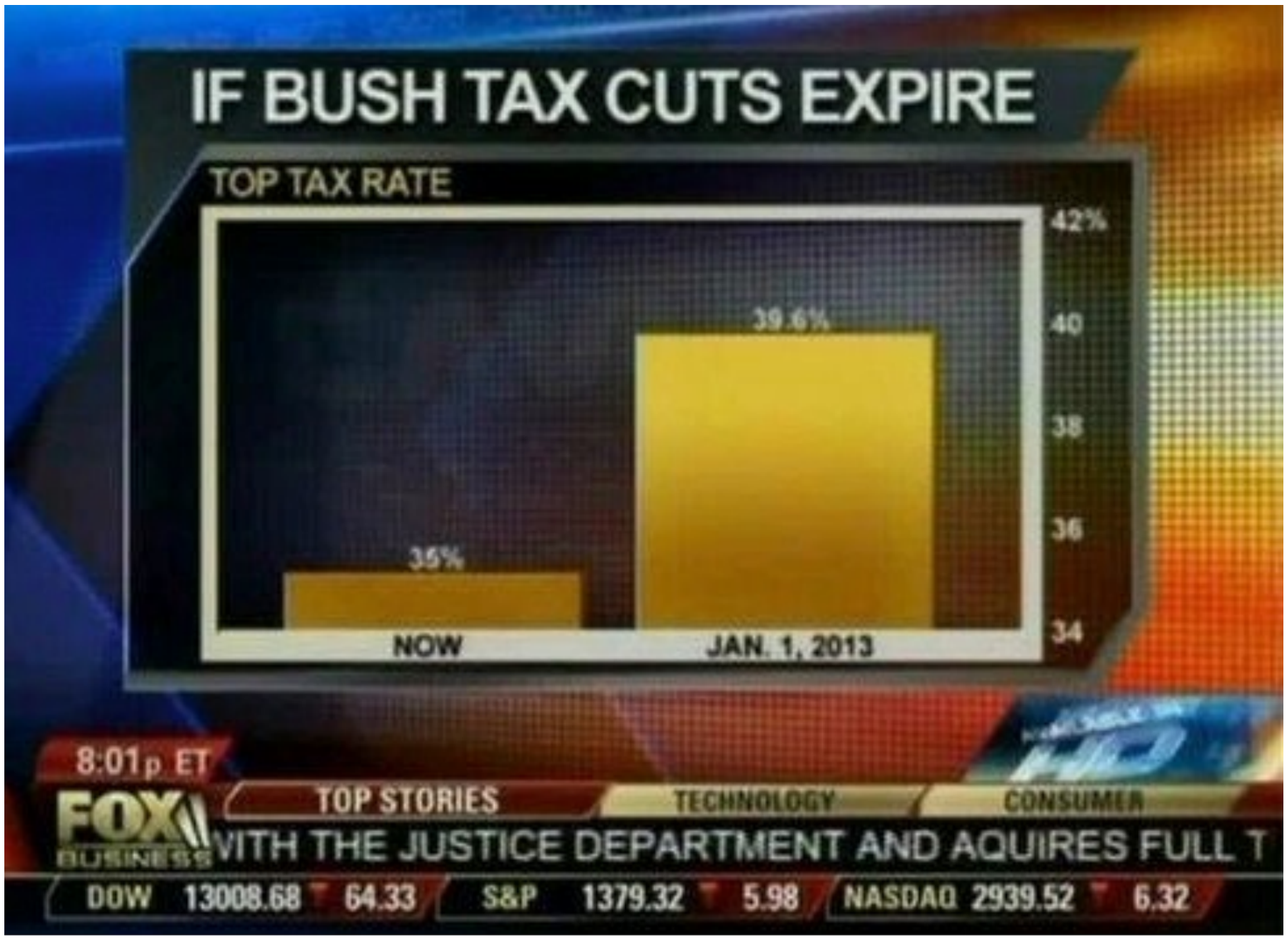


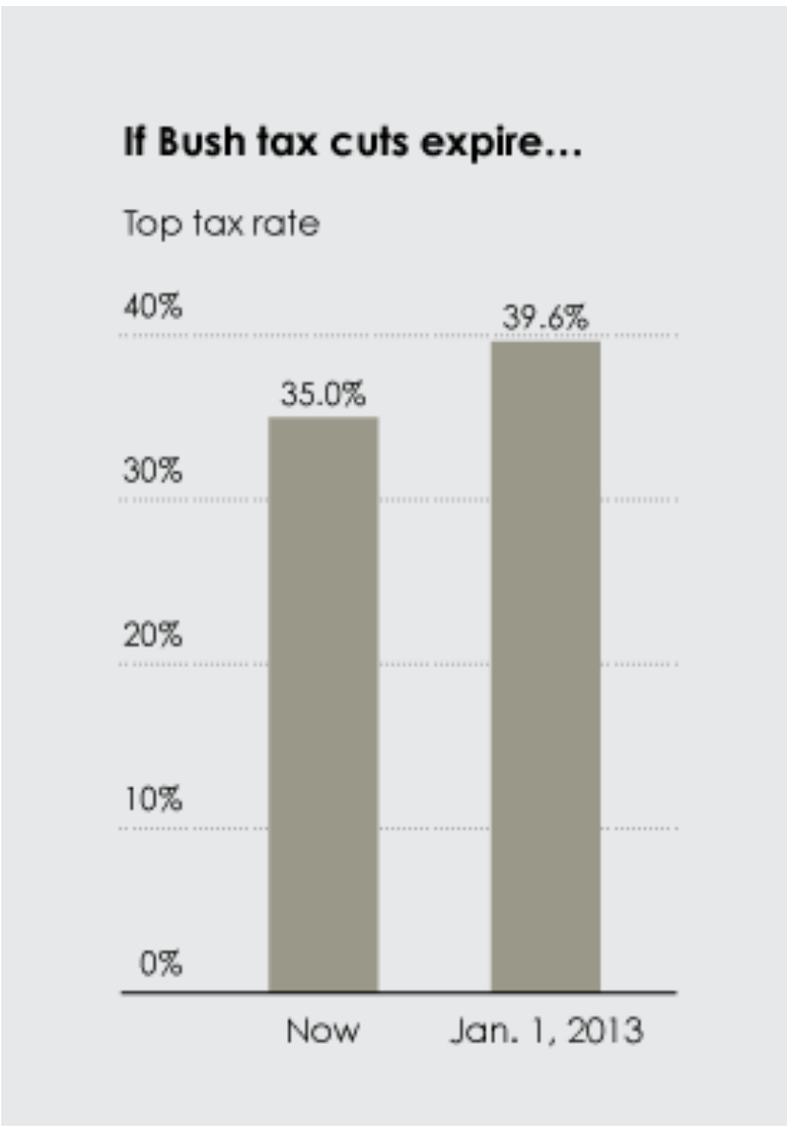
September 13, 2011

GRAPHICAL INTEGRITY

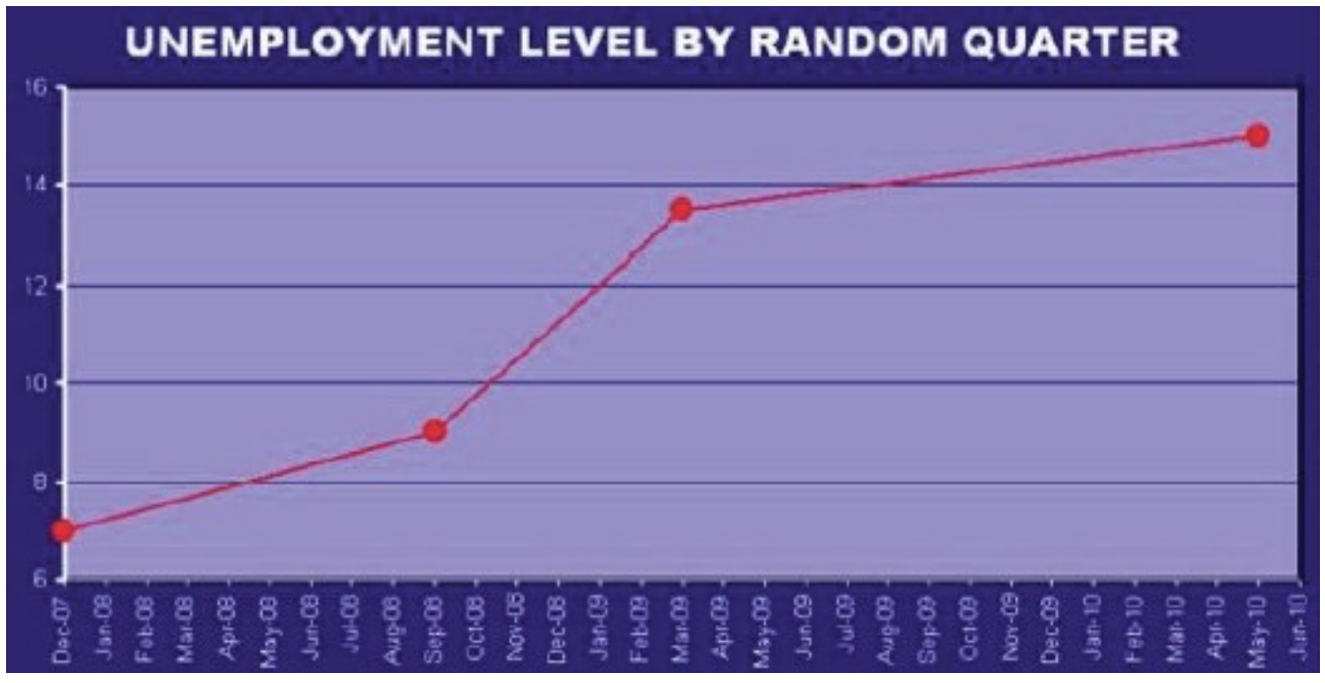


SCALE DISTORTIONS

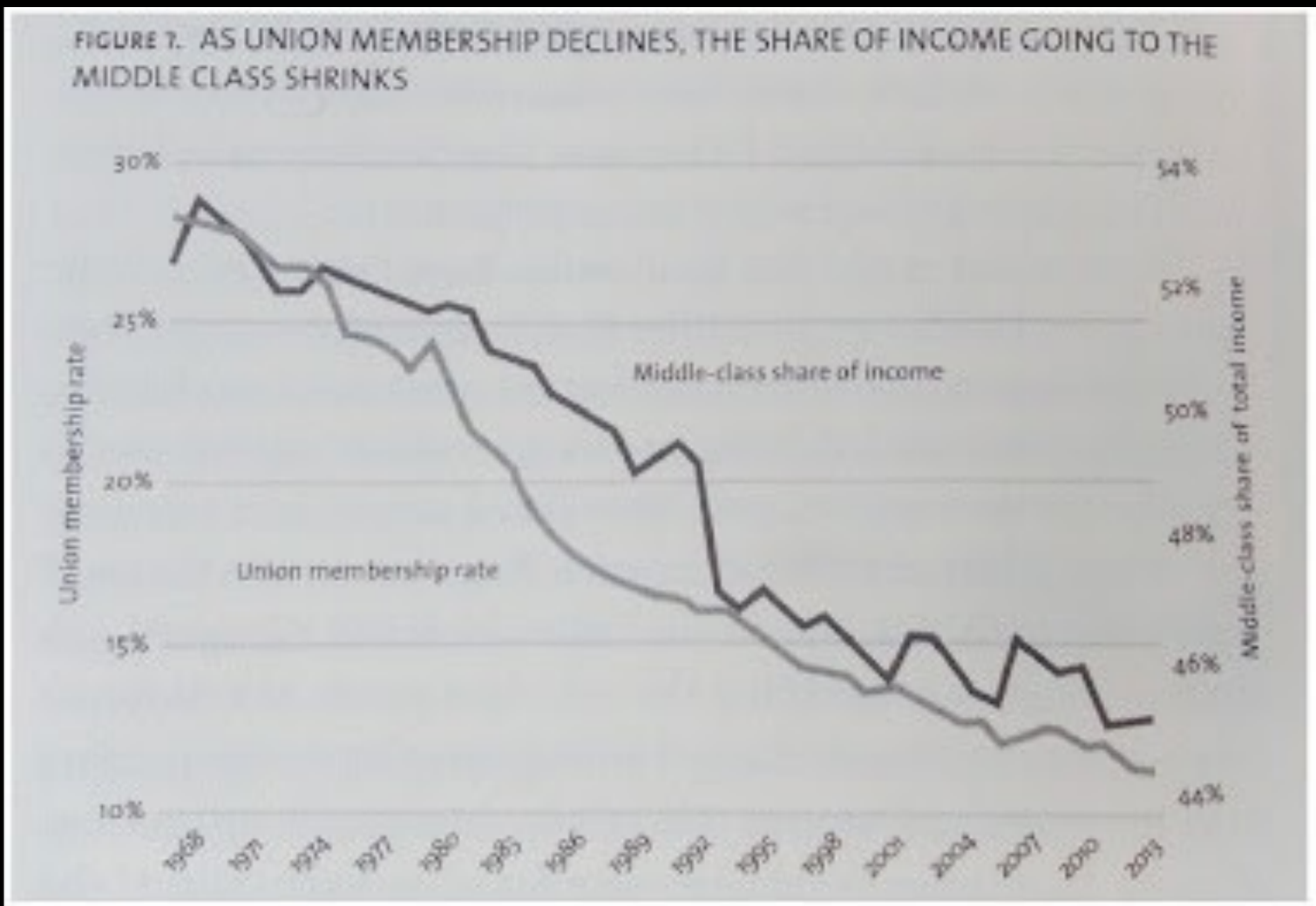


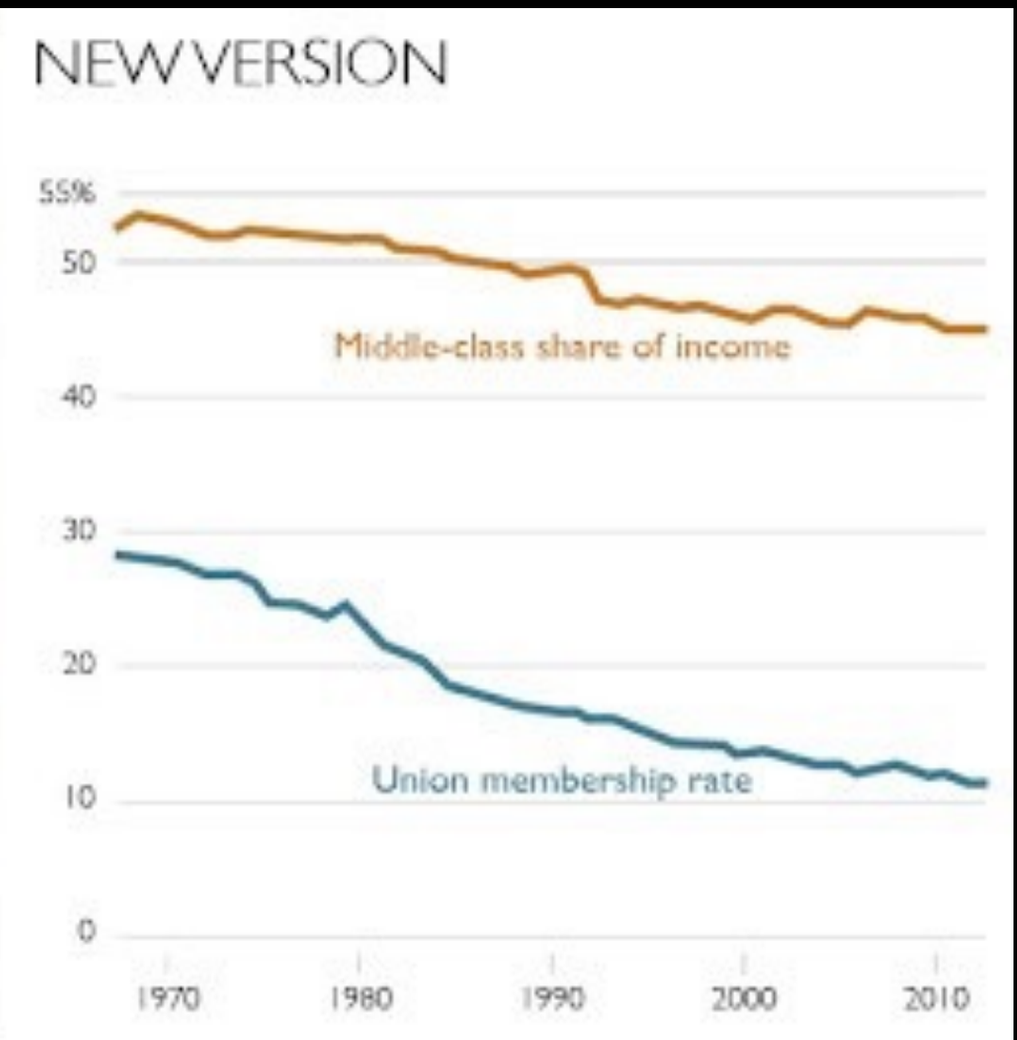


SCALE DISTORTIONS



"DOUBLE THE AXES, DOUBLE THE MISCHIEF"
(QUOTE FROM GARY SMITH'S STANDARD DEVIATIONS)





Graphic from Robert Reich's Saving Capitalism

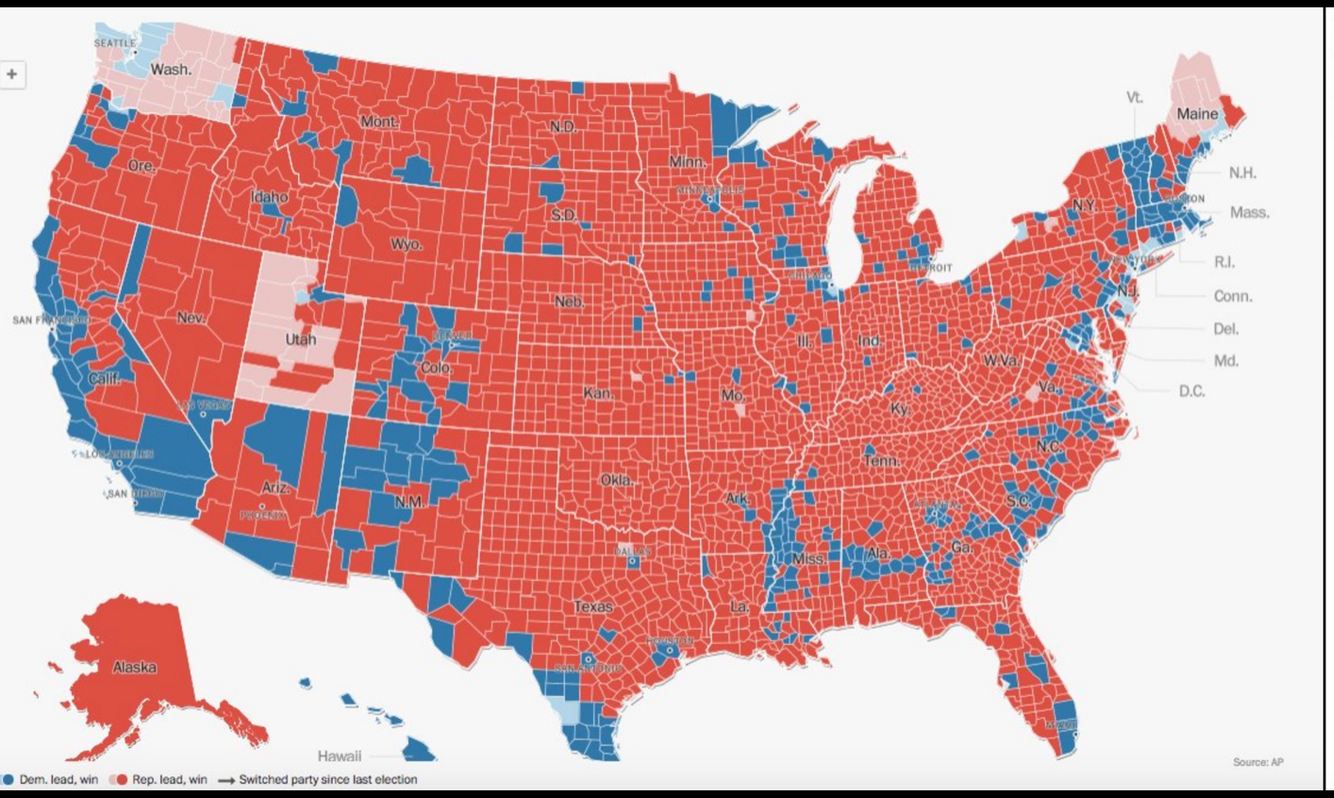
http://www.thefunctionalart.com/2015/10/double-axes-double-mischief.html

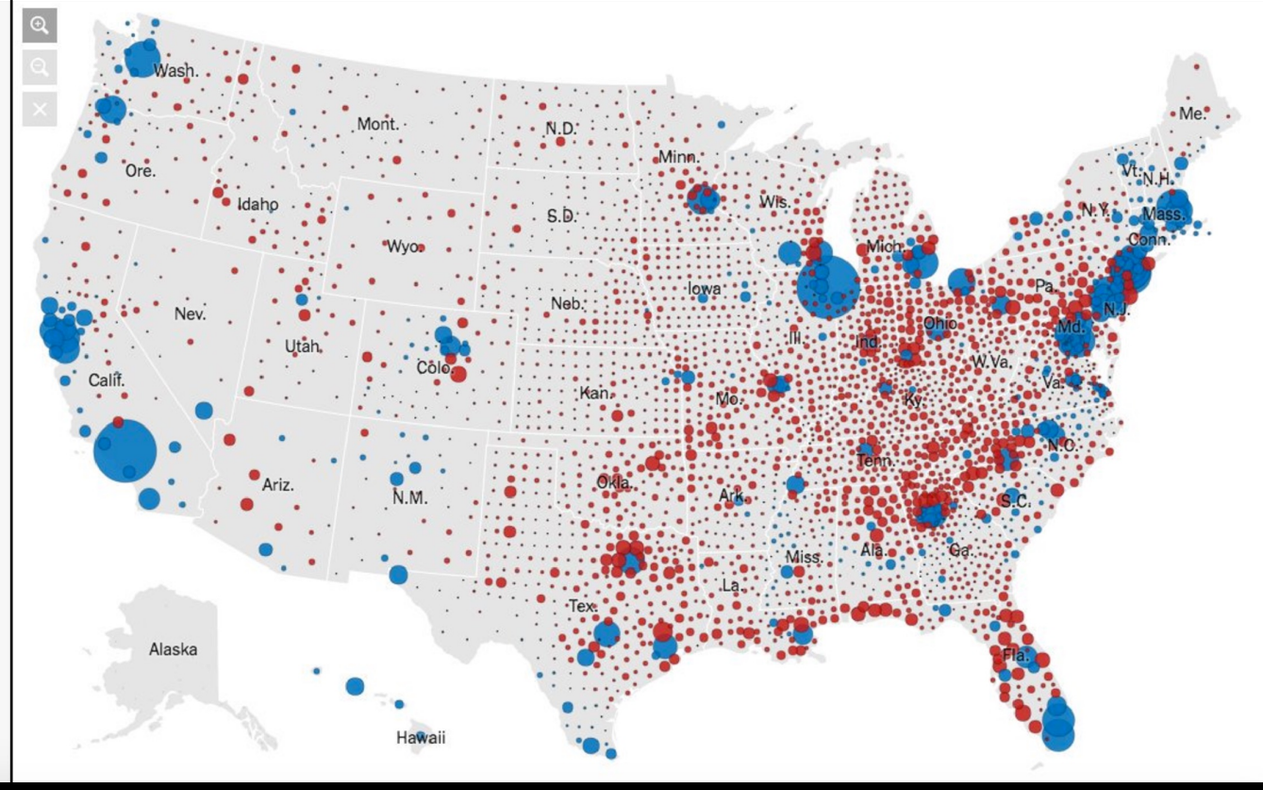
Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

BE PROPORTIONAL

Trey Yingst @TreyYingst · May 11

Spotted: A map to be hung somewhere in the West Wing





US Presidential Election 2016

Results mapped at county level showing the candidate with the largest vote share in each area

Overall result:

Trump
62,979,636 votes (46.1%)
306 electoral votes
(received 304 in the Electoral College)

Clinton
65,844,610 votes (48.2%)
232 electoral votes
(received 227 in the Electoral College)

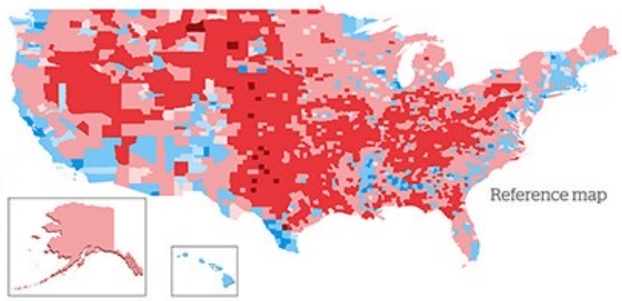
Other candidates
7,804,213 votes (5.7%)

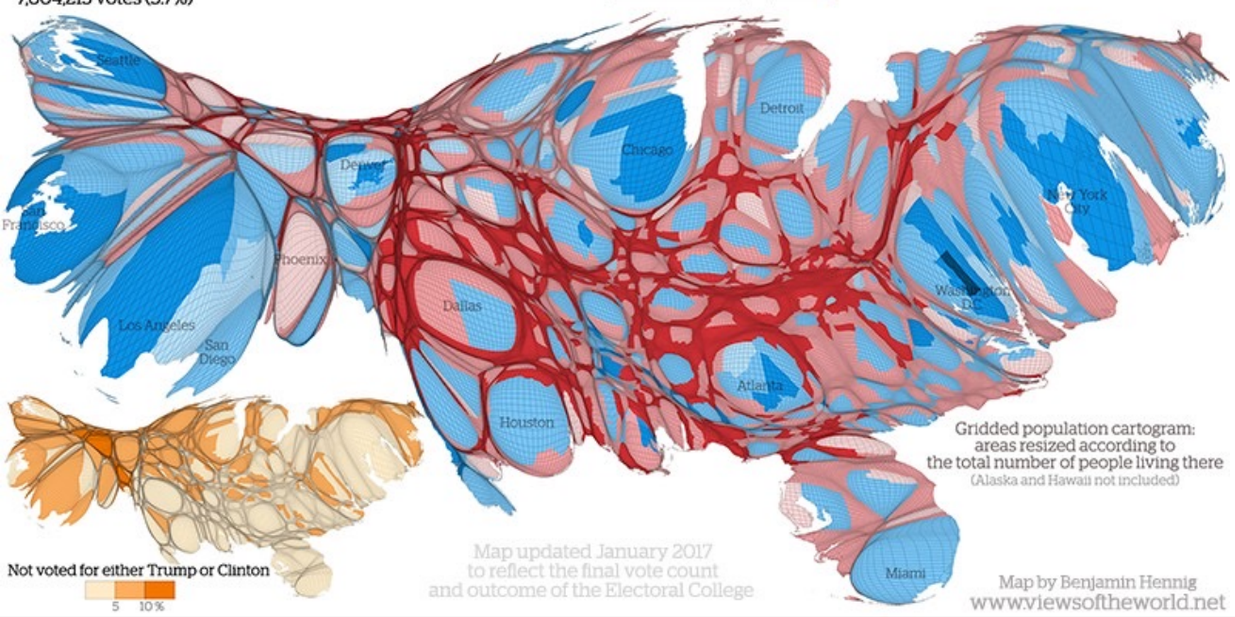
Vote share
of candidate with most votes

0 50 70 90%

Trump

Clinton



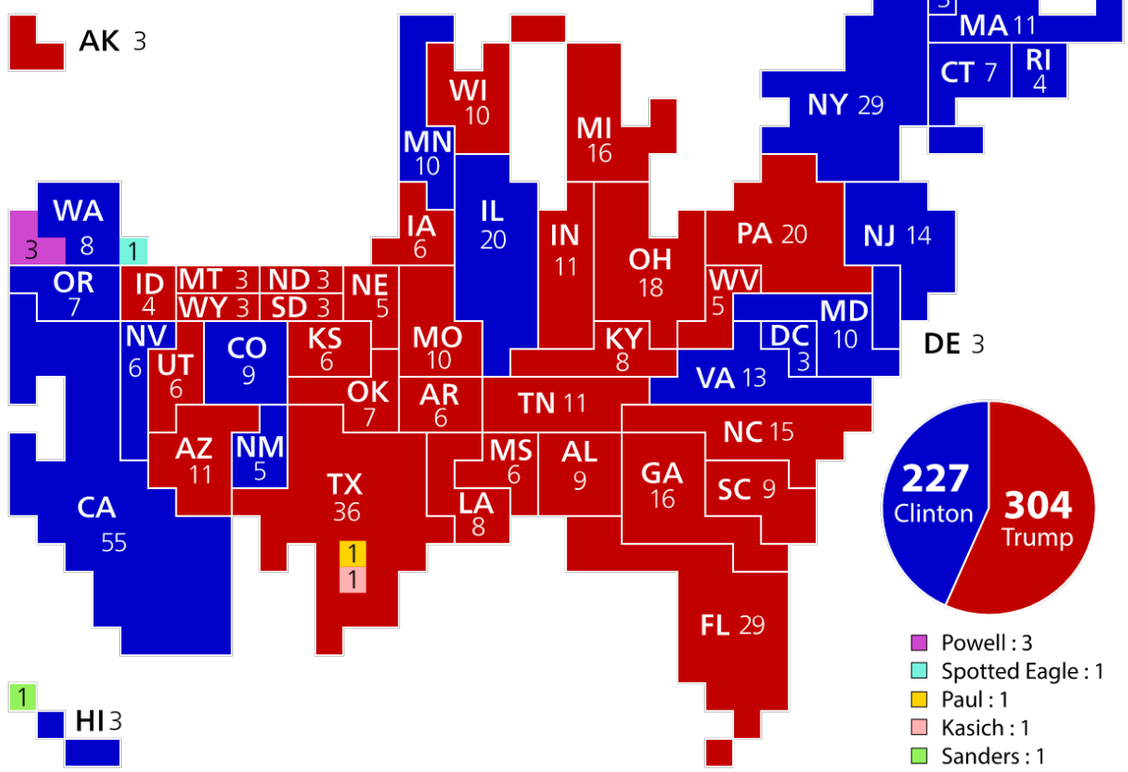


Gridded population cartogram:
areas resized according to
the total number of people living there
(Alaska and Hawaii not included)

Map updated January 2017
to reflect the final vote count
and outcome of the Electoral College

Map by Benjamin Hennig
www.viewsoftheworld.net

2016 Electoral Vote

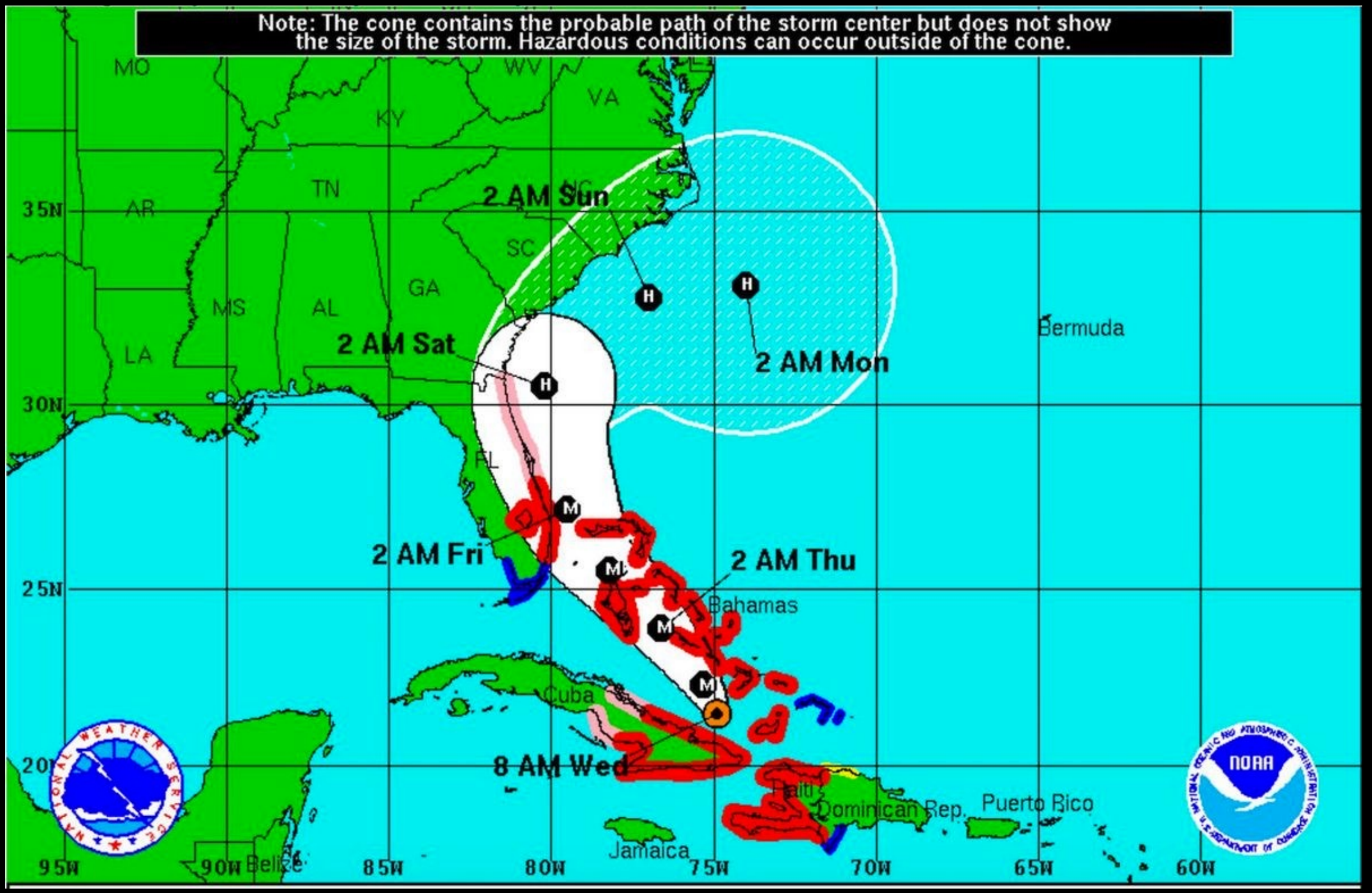


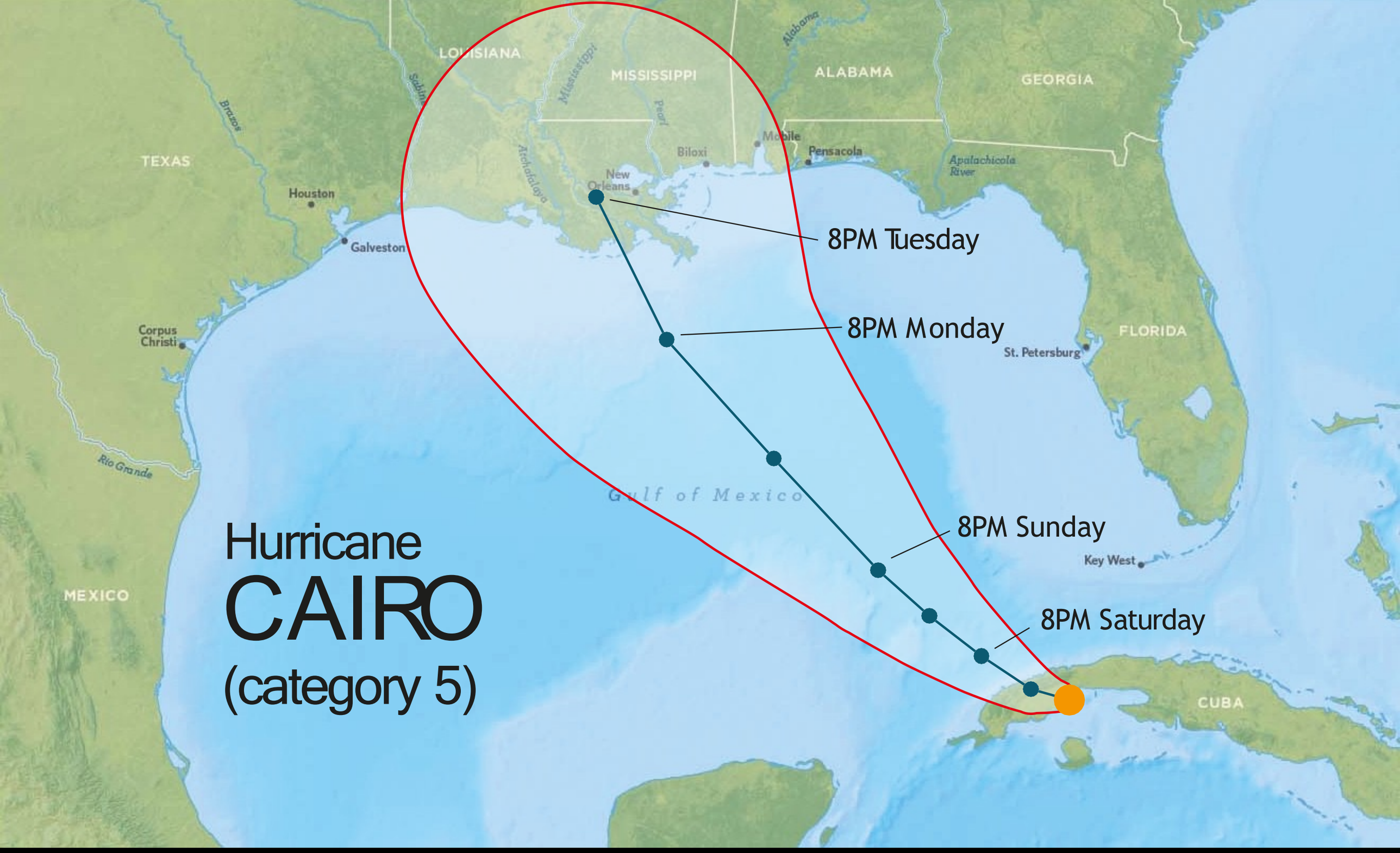
227 Clinton

304 Trump

- Powell : 3
- Spotted Eagle : 1
- Paul : 1
- Kasich : 1
- Sanders : 1

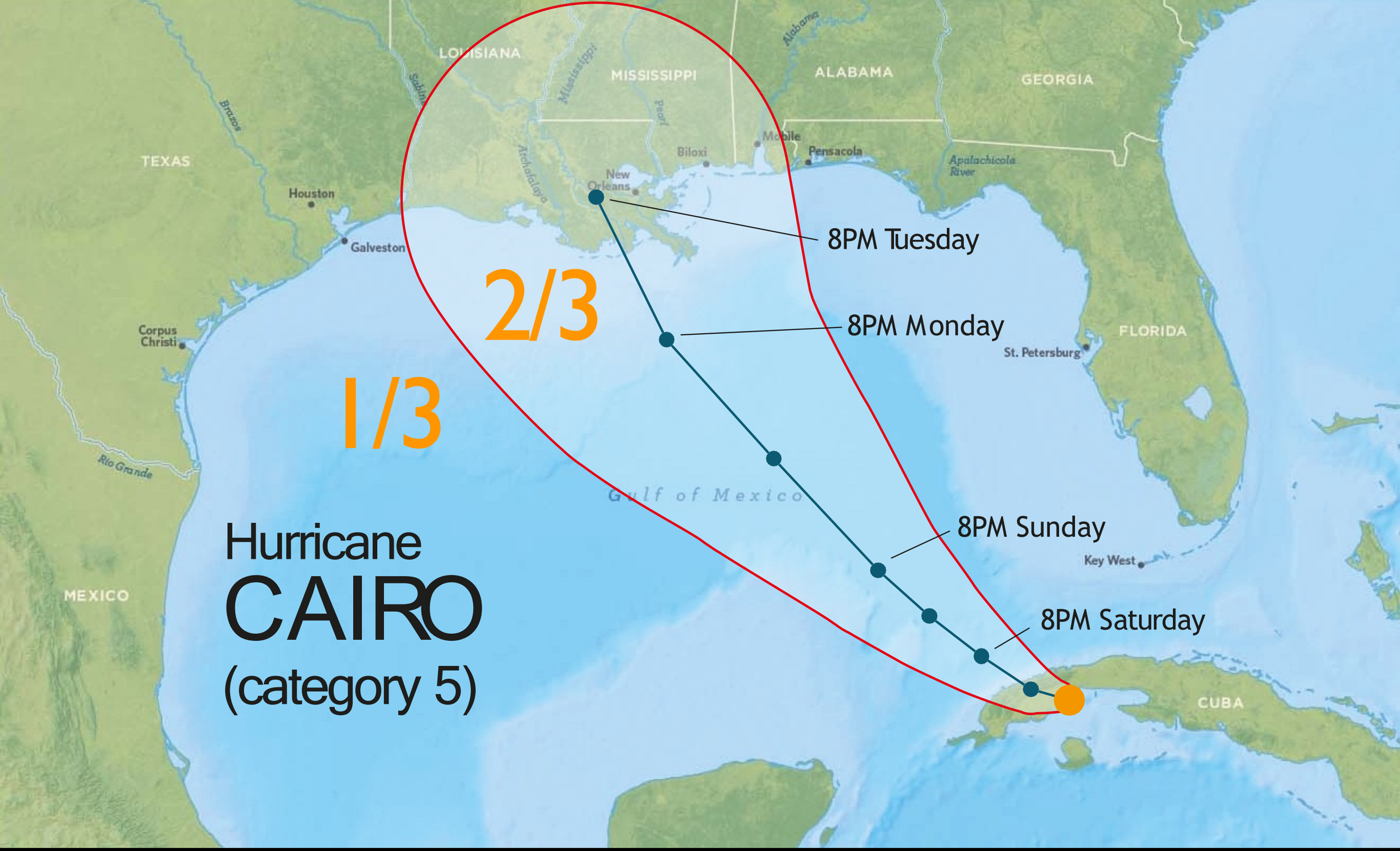
INCLUDE UNCERTAINTY


Note: The cone contains the probable path of the storm center but does not show the size of the storm. Hazardous conditions can occur outside of the cone.
MO
KY
WV
VA
TN
NC
AR
SC
GA
MS
AL
LA
FL
2 AM Sun
2 AM Sat
2 AM Mon
2 AM Fri
2 AM Thu
8 AM Wed
Bermuda
Bahamas
Cuba
Haiti
Dominican Rep.
Puerto Rico
Jamaica
Belize
35N
30N
25N
20N
95W
90W
85W
80W
75W
70W
65W
60W
NATIONAL WEATHER SERVICE
NOAA



What you show

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo



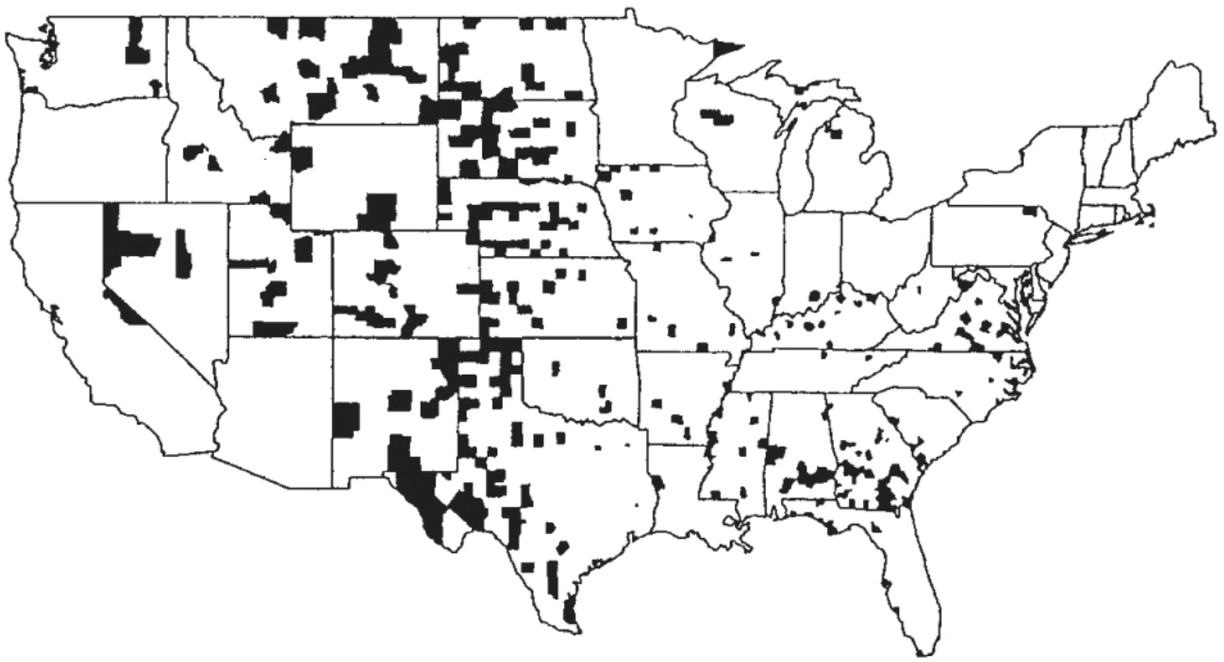
What non-scientists are not aware of (cone is just 66% probability)

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

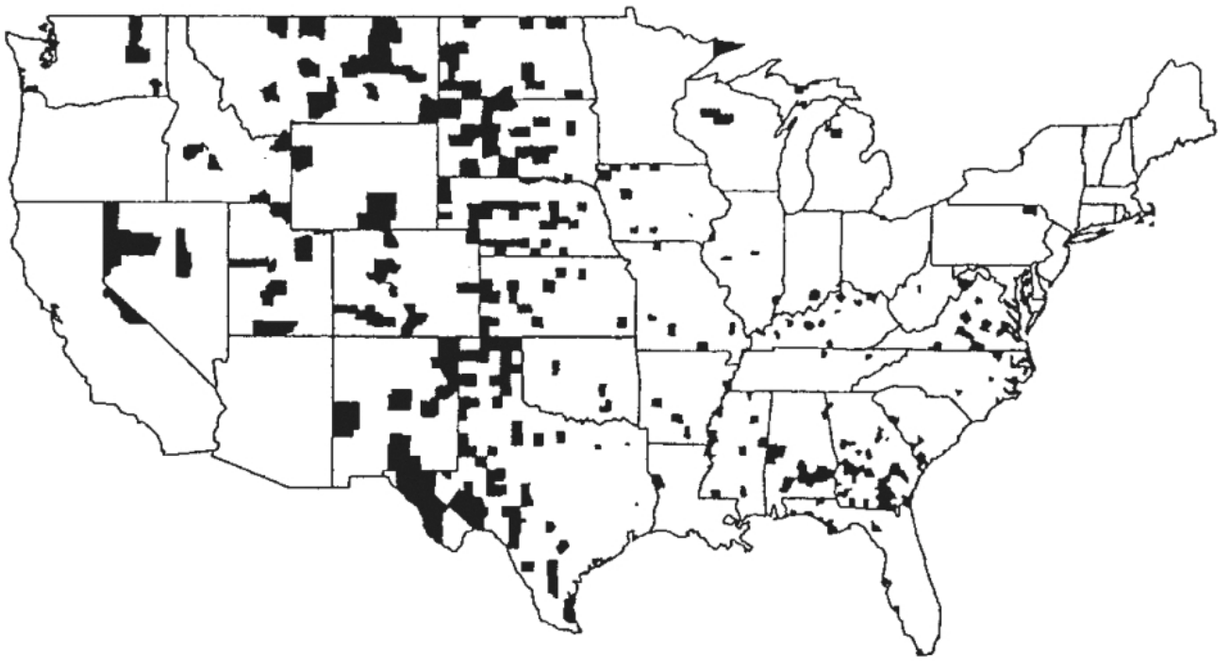
What we could be showing instead

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

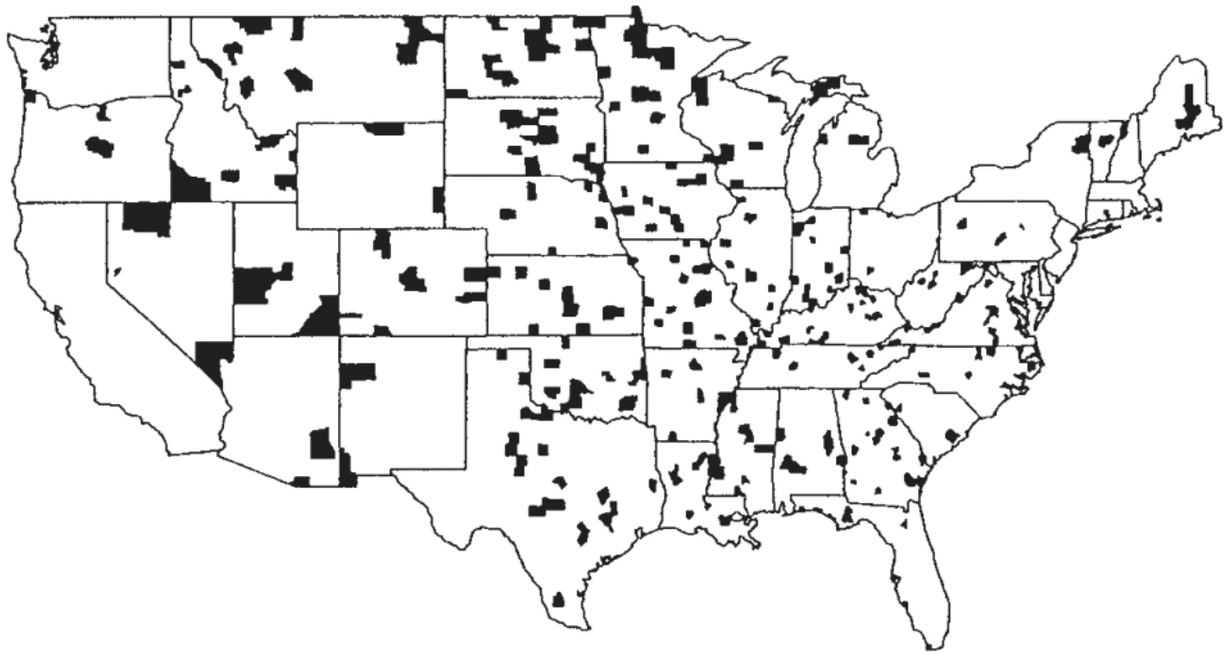
PLOT ALL YOUR DATA



Counties with the LOWEST kidney cancer death rates (1980-1989)



Counties with the LOWEST kidney cancer death rates (1980-1989)

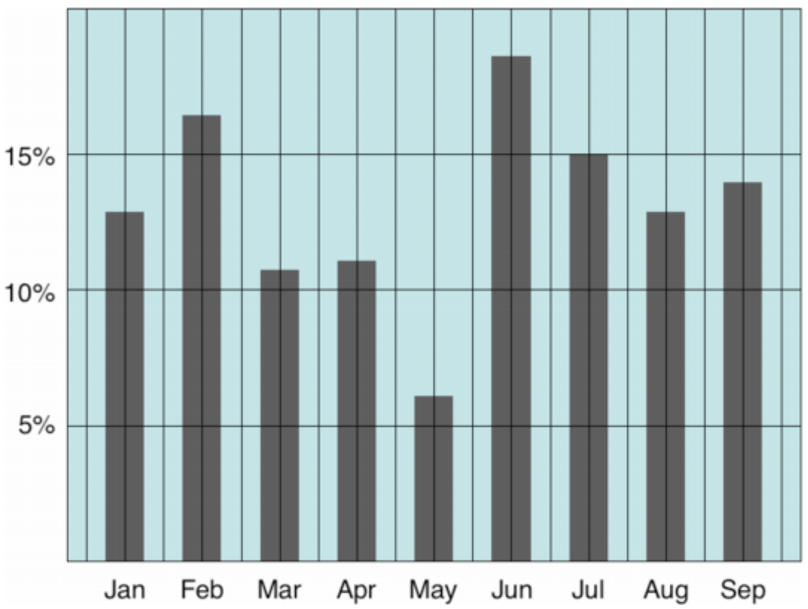


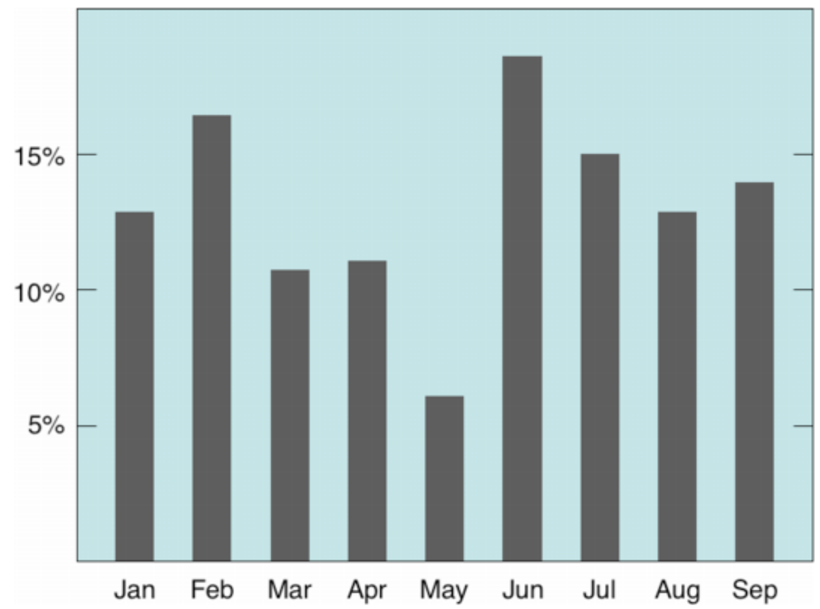
Counties with the HIGHEST kidney cancer death rates (1980-1989)

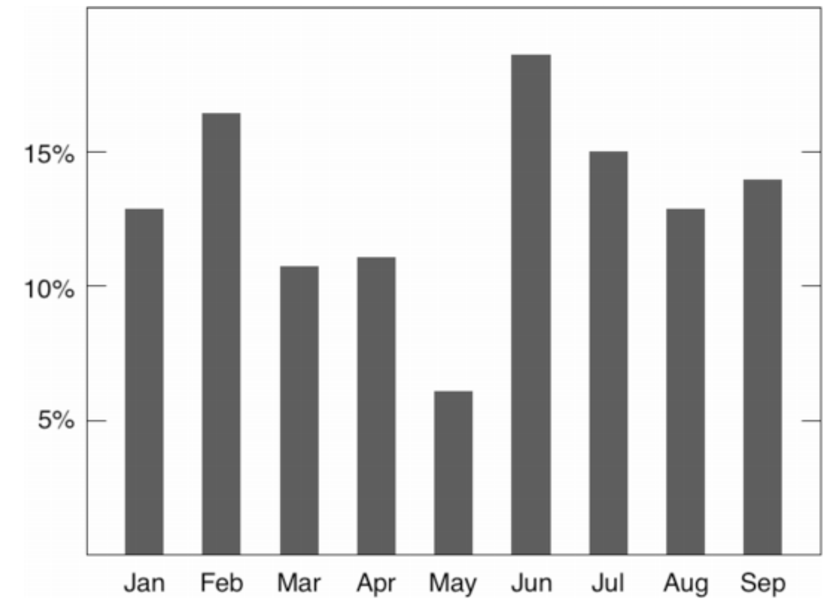
2. KEEP IT SIMPLE

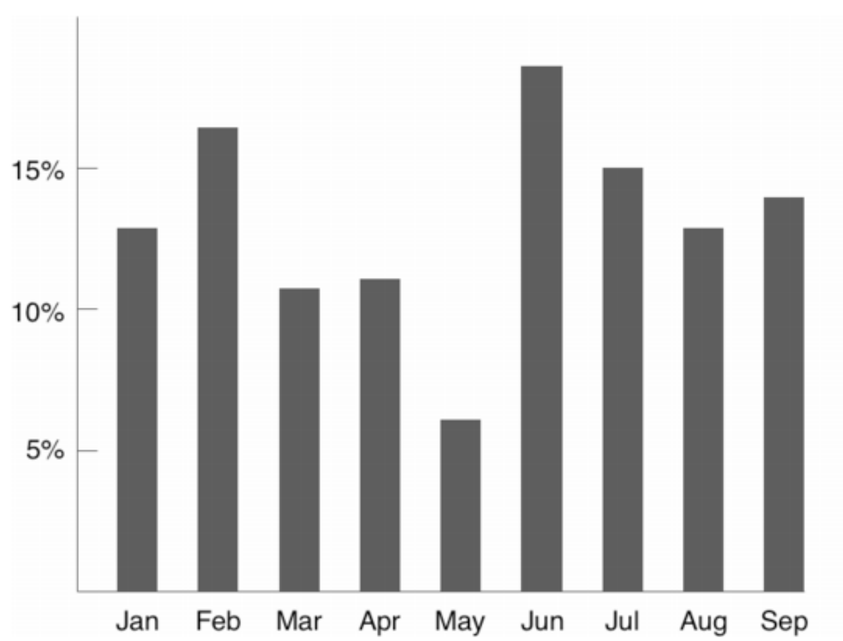
AVOID CHARTJUNK

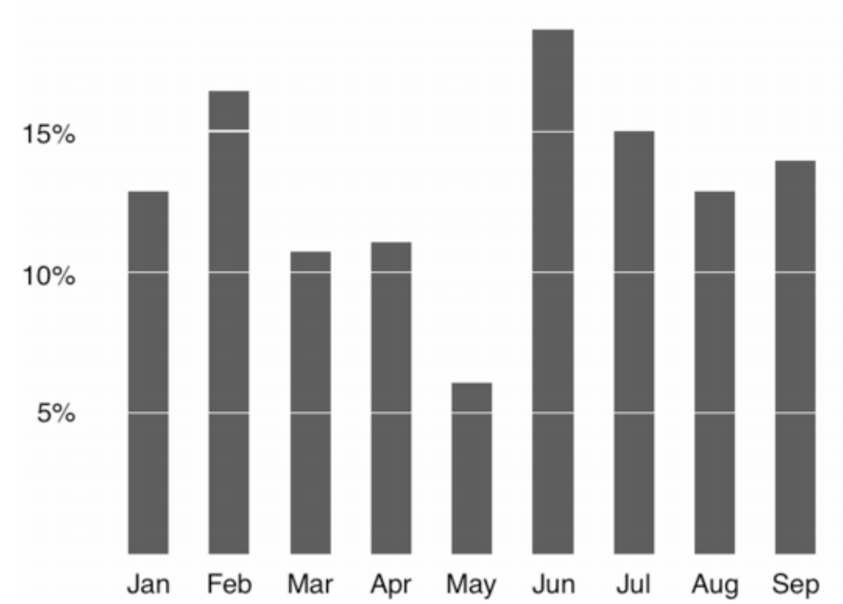
EXTRANEOUS VISUAL ELEMENTS THAT DISTRACT FROM THE MESSAGE



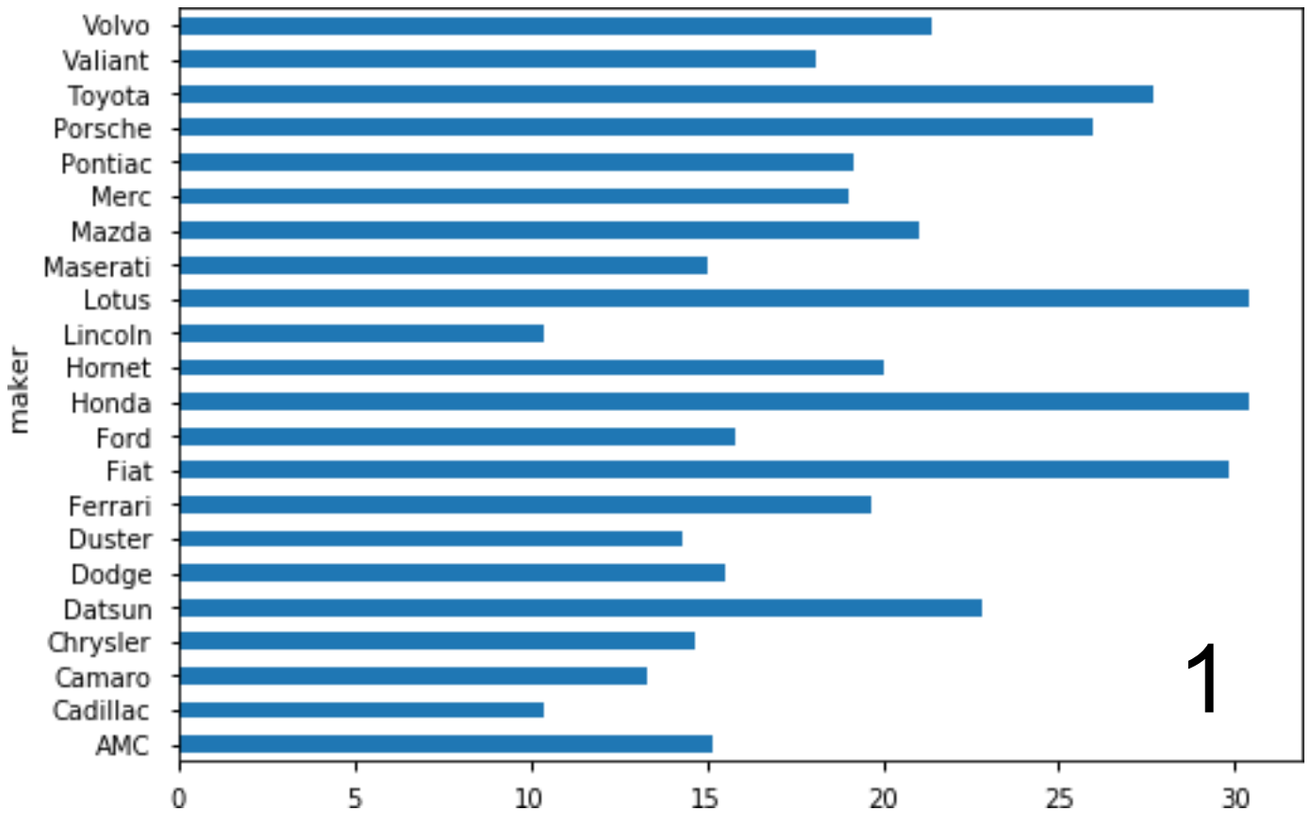


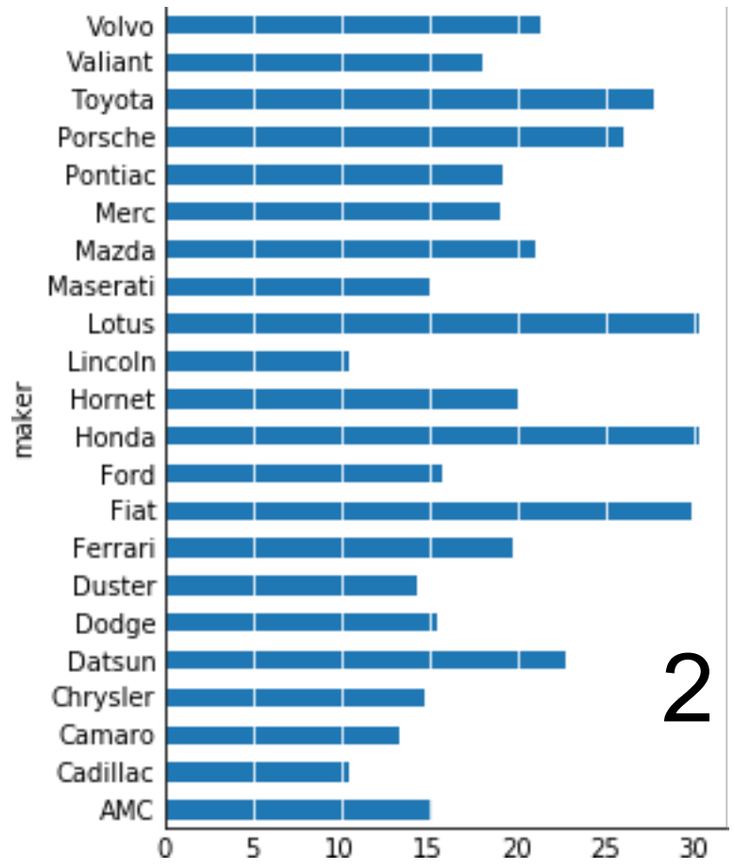


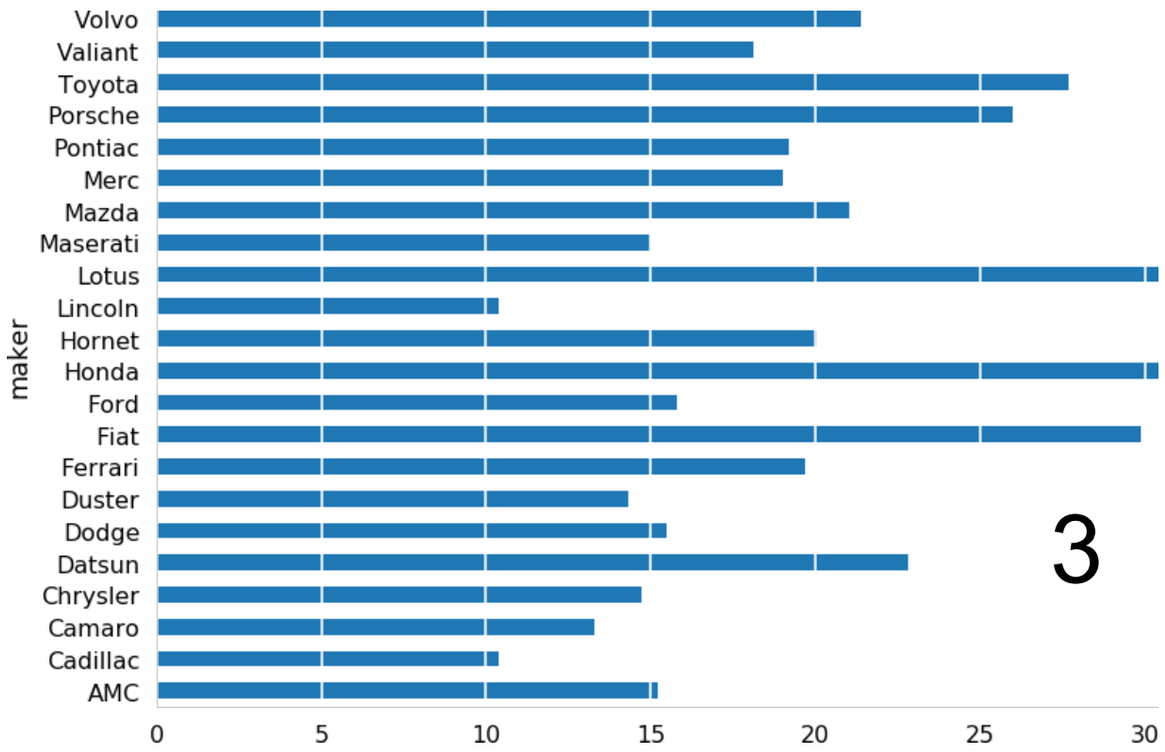


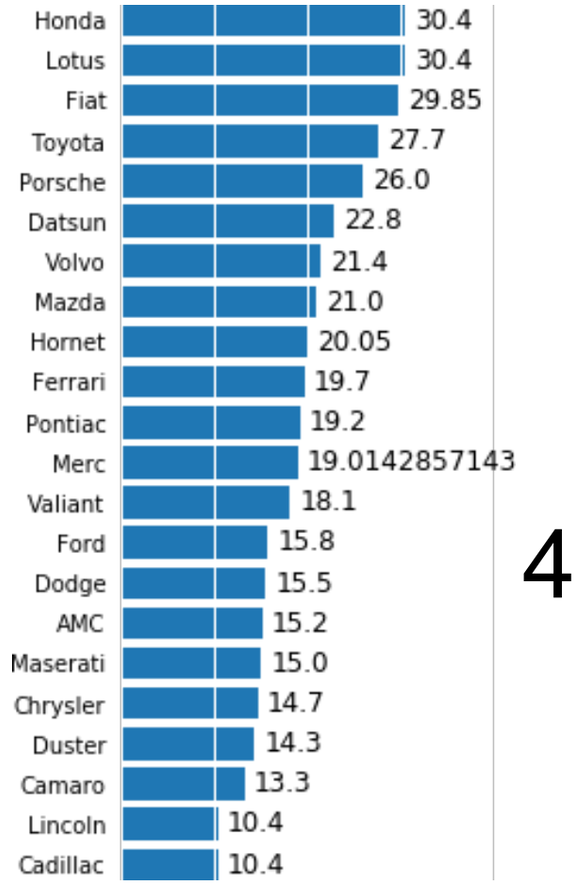


ongoing, Tim Brey

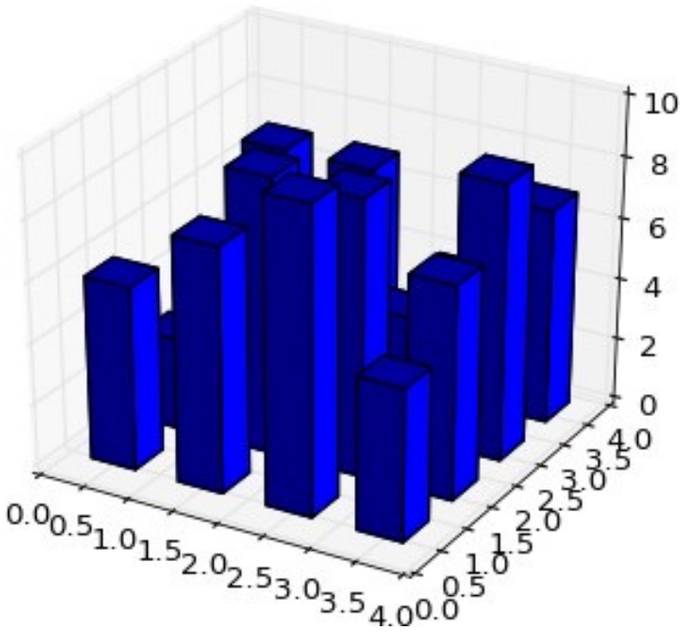

maker
Volvo
Valiant
Toyota
Porsche
Pontiac
Merc
Mazda
Maserati
Lotus
Lincoln
Hornet
Honda
Ford
Fiat
Ferrari
Duster
Dodge
Datsun
Chrysler
Camaro
Cadillac
AMC
0
5
10
15
20
25
30
1

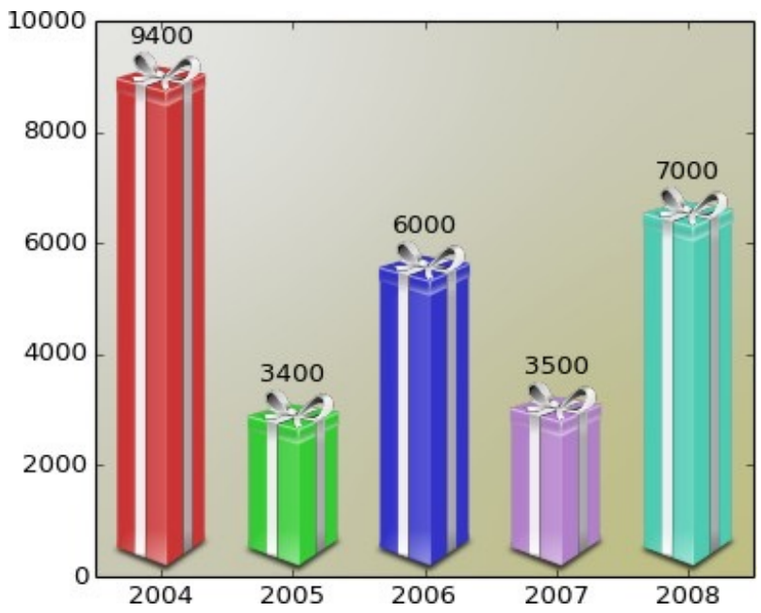

maker
Volvo
Valiant
Toyota
Porsche
Pontiac
Merc
Mazda
Maserati
Lotus
Lincoln
Hornet
Honda
Ford
Fiat
Ferrari
Duster
Dodge
Datsun
Chrysler
Camaro
Cadillac
AMC
0
5
10
15
20
25
30
2


maker
Volvo
Valiant
Toyota
Porsche
Pontiac
Merc
Mazda
Maserati
Lotus
Lincoln
Hornet
Honda
Ford
Fiat
Ferrari
Duster
Dodge
Datsun
Chrysler
Camaro
Cadillac
AMC
0
5
10
15
20
25
30
3

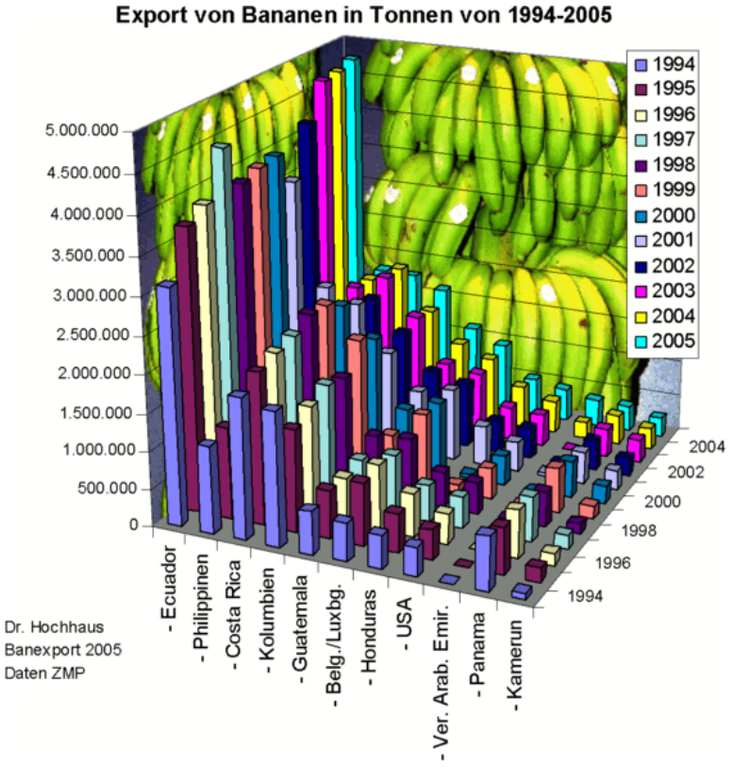

Honda 30.4
Lotus 30.4
Fiat 29.85
Toyota 27.7
Porsche 26.0
Datsun 22.8
Volvo 21.4
Mazda 21.0
Hornet 20.05
Ferrari 19.7
Pontiac 19.2
Merc 19.0142857143
Valiant 18.1
Ford 15.8
Dodge 15.5
AMC 15.2
Maserati 15.0
Chrysler 14.7
Duster 14.3
Camaro 13.3
Lincoln 10.4
Cadillac 10.4
4

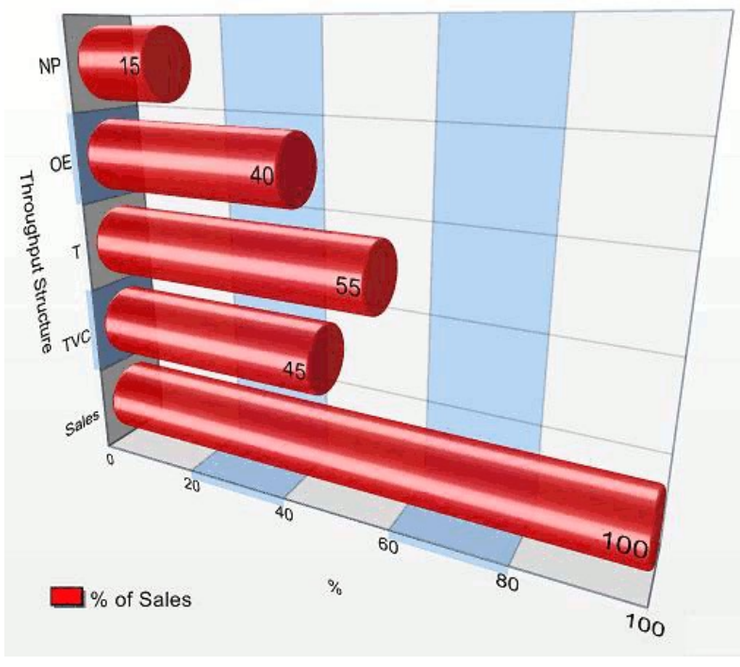
DON'T!





matplotlib gallery

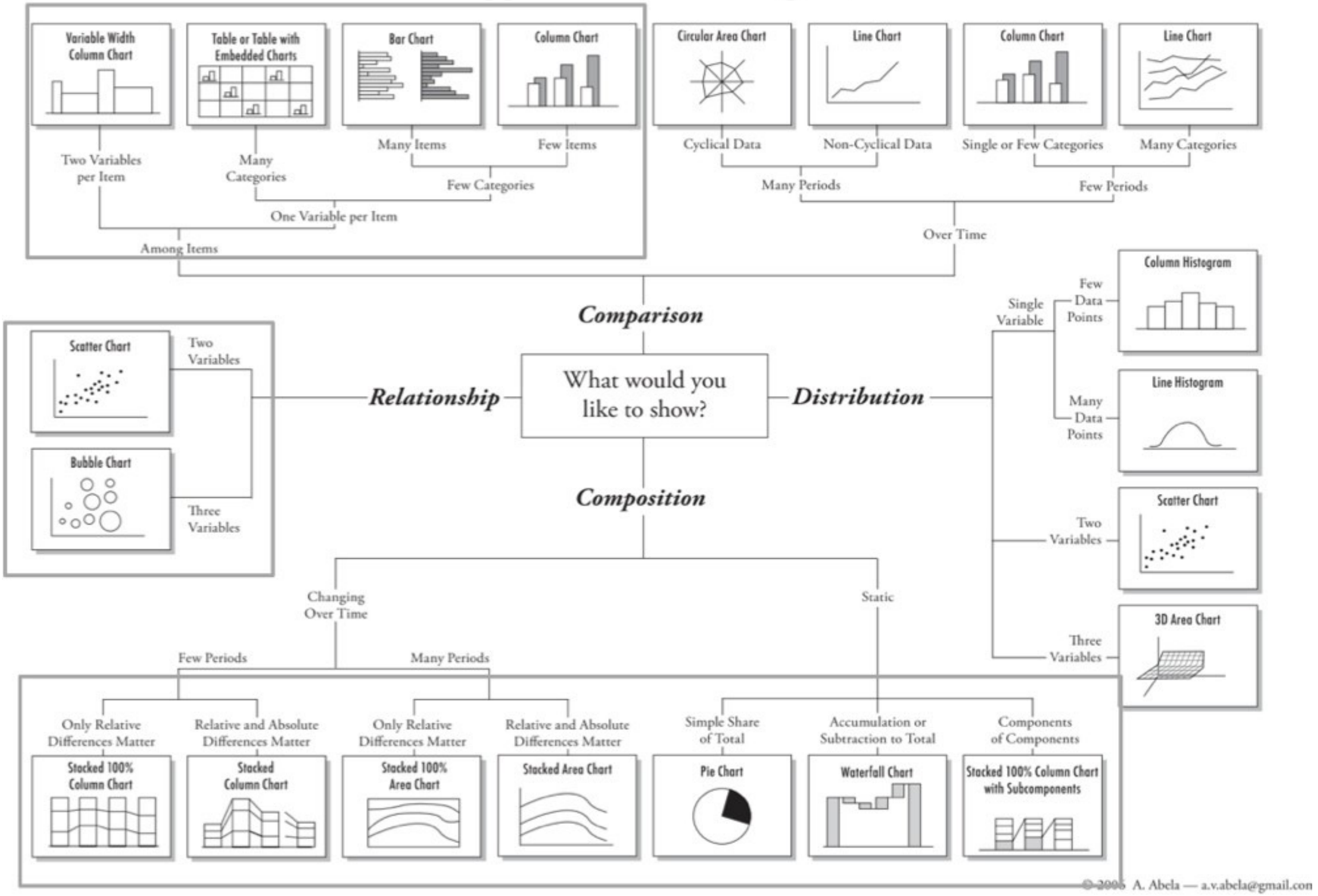




Excel Charts Blog

3. USE THE RIGHT DISPLAY

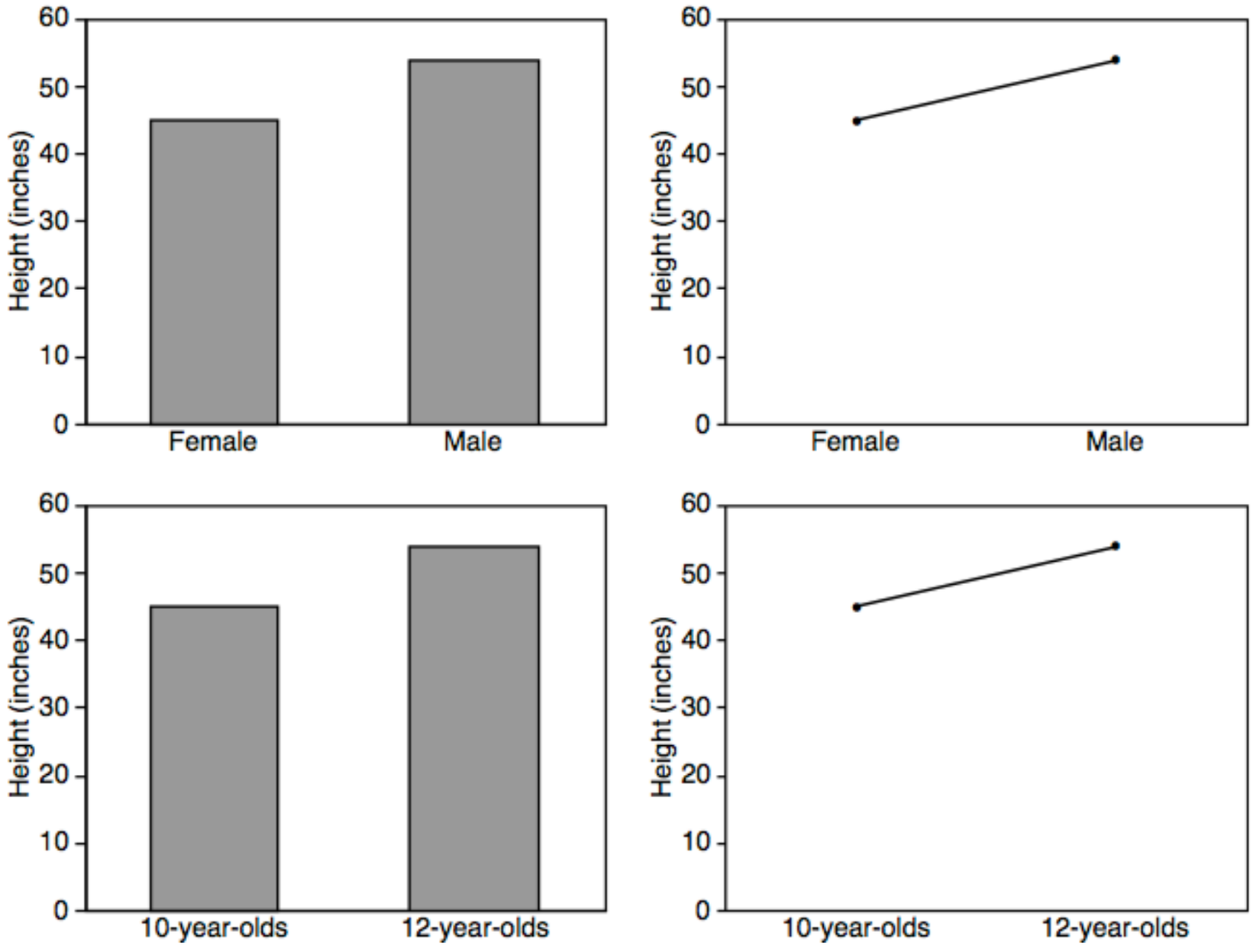
Chart Suggestions—A Thought-Starter



http://extremepresentation.typepad.com/blog/files/choosing_a_good_chart.pdf

COMPARISONS

BARS vs LINES

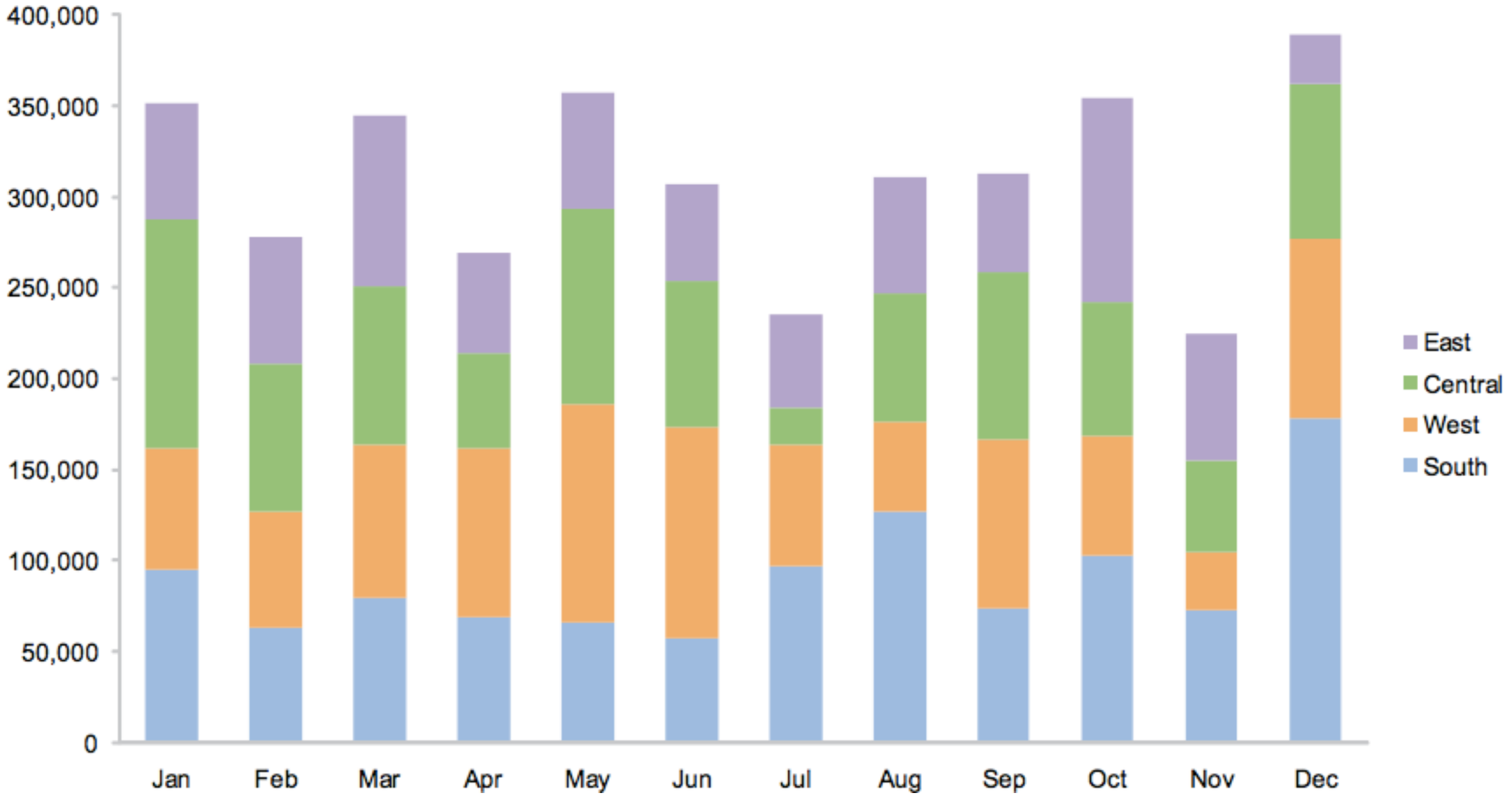


Zacks 1999

PROPORTIONS

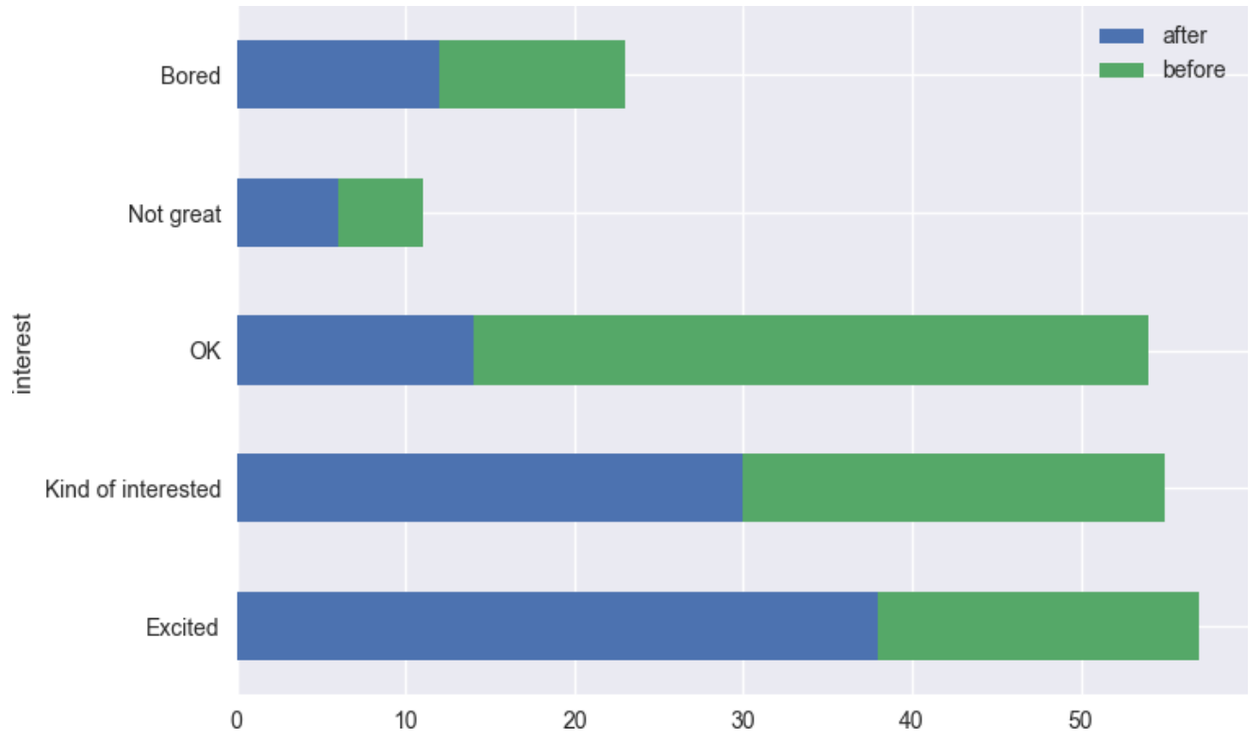
PIE CHARTS

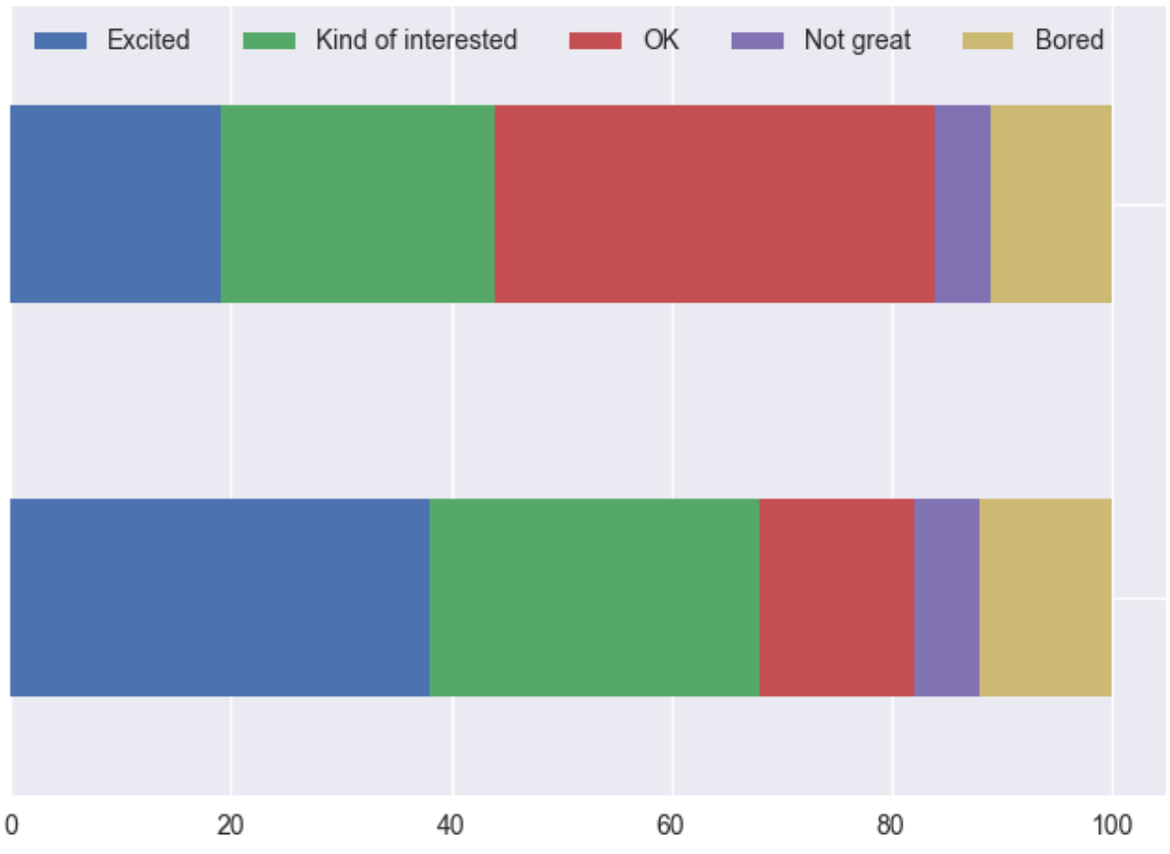
STACKED BAR CHART



S. Few

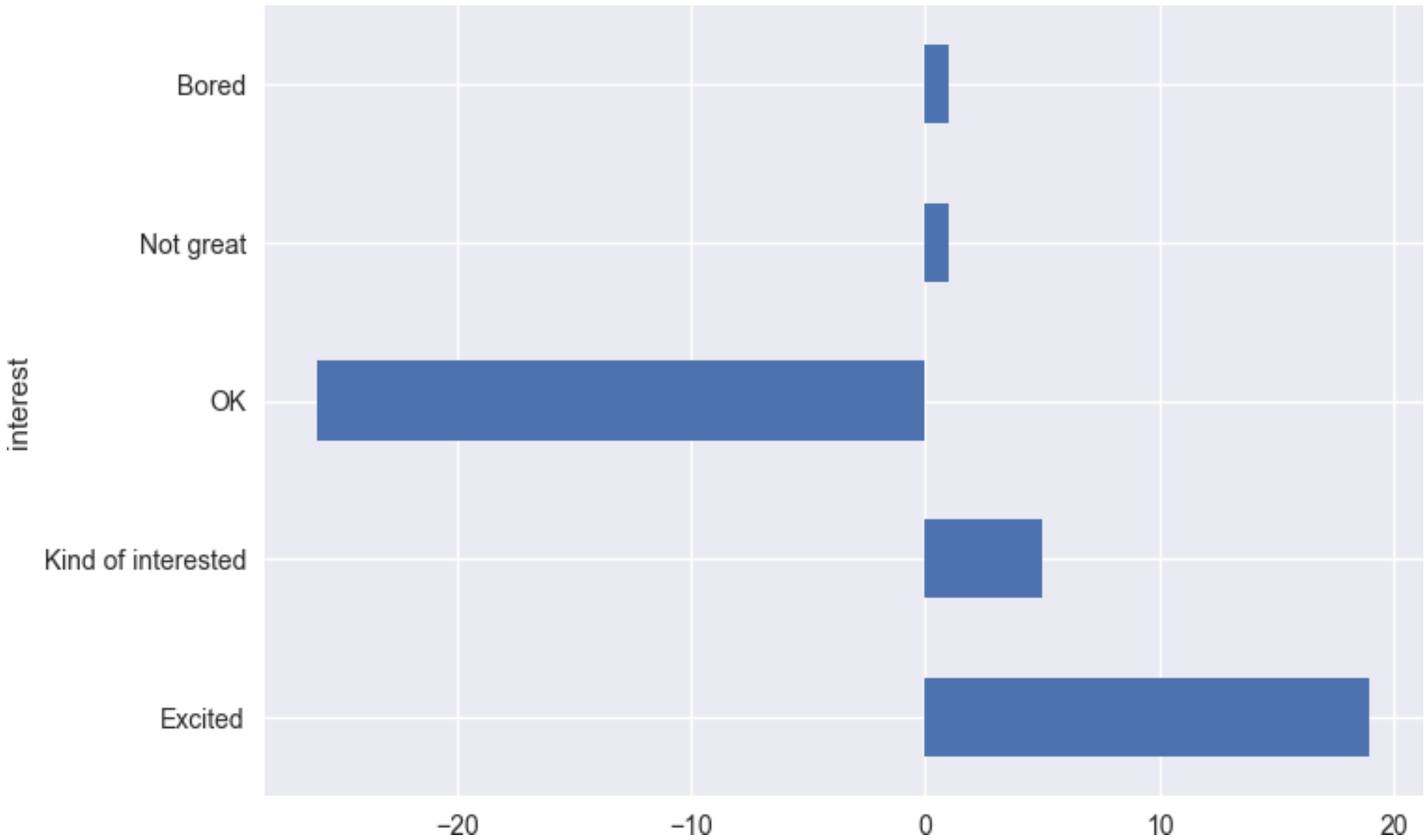
Stacked bar, not very useful

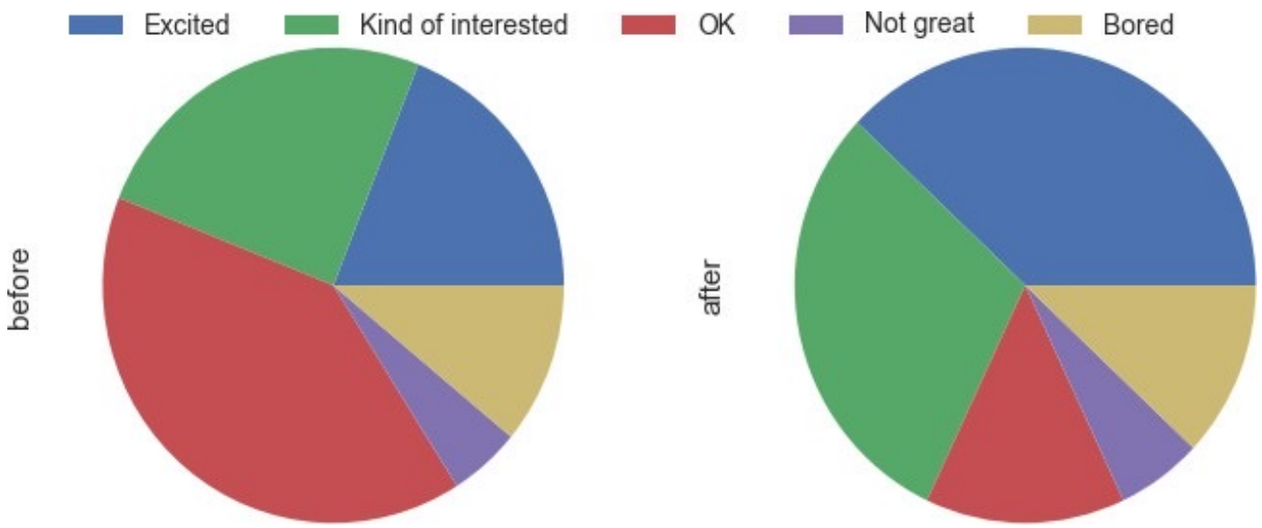




Data Transposed Bar Chart

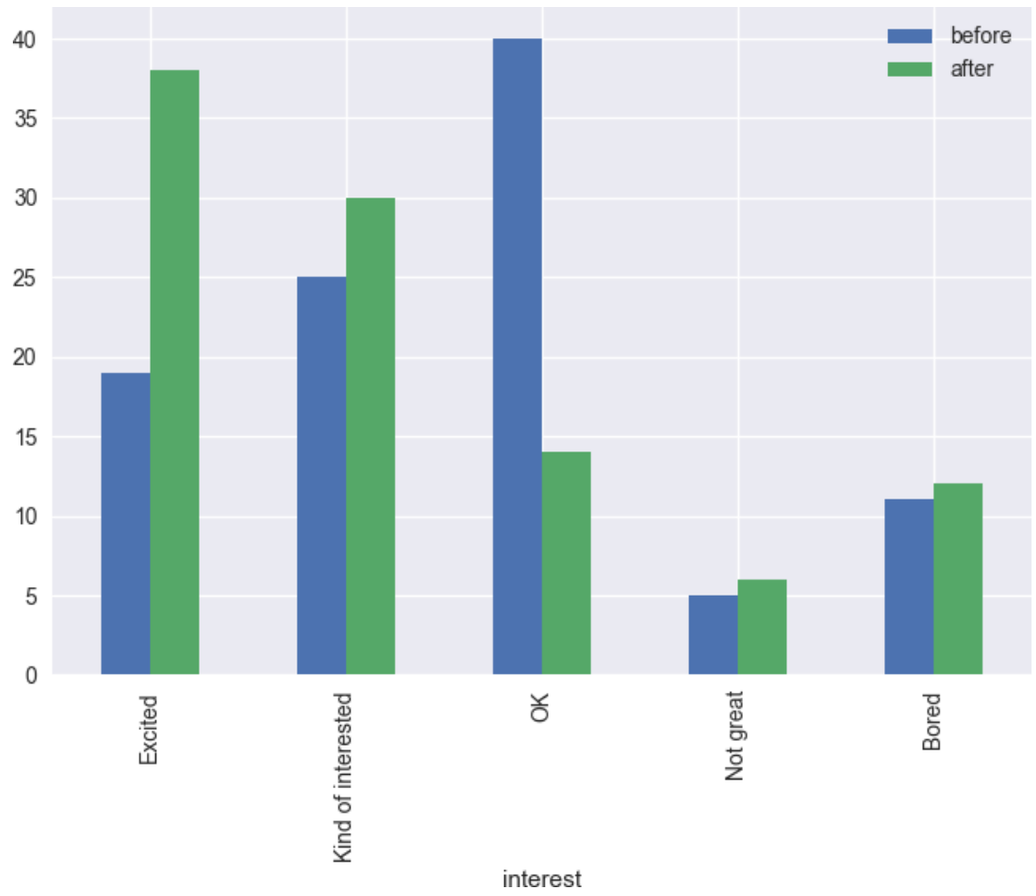
DIFFERENCE BAR CHART



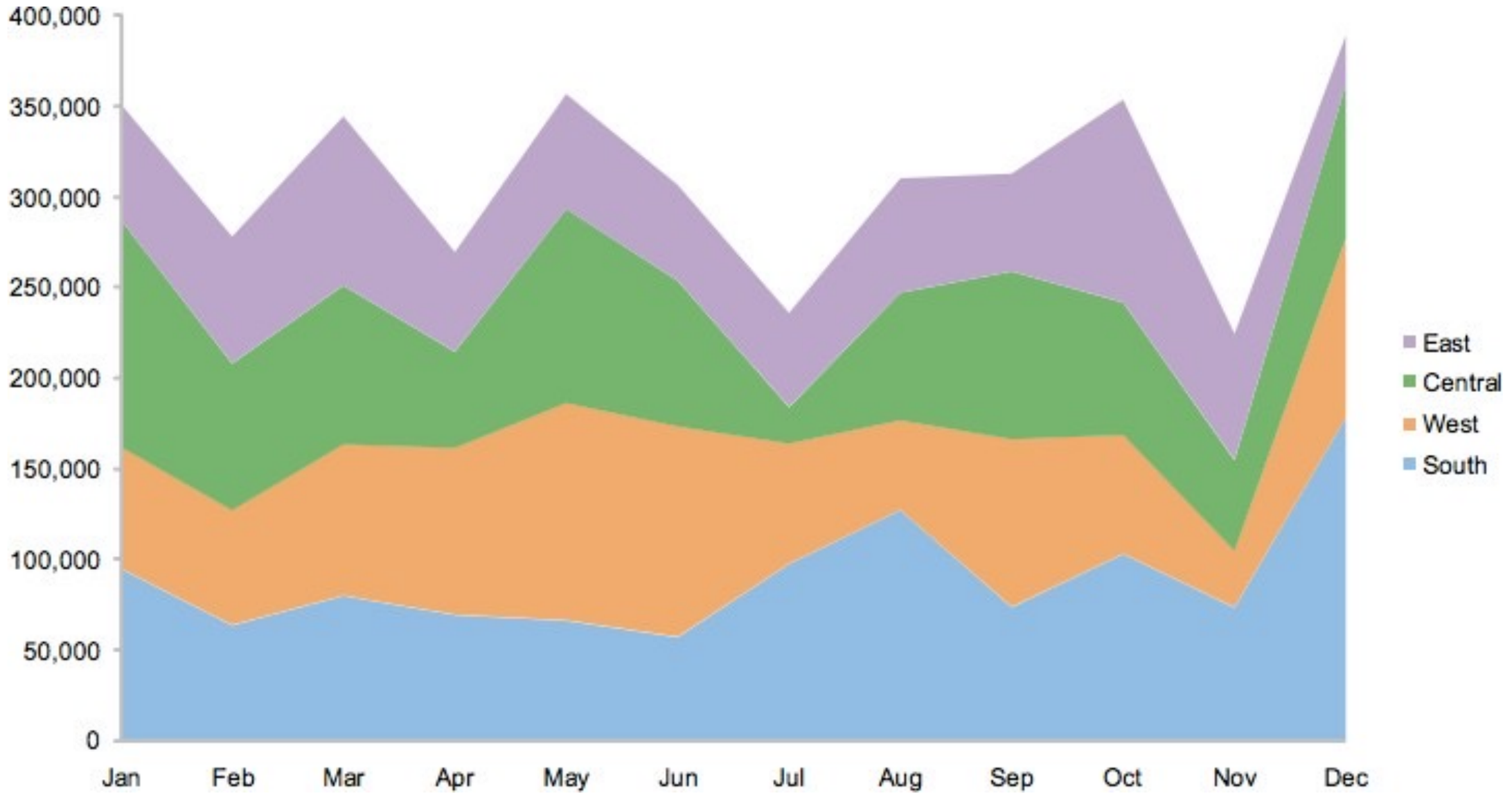


Pie

Side by side bar



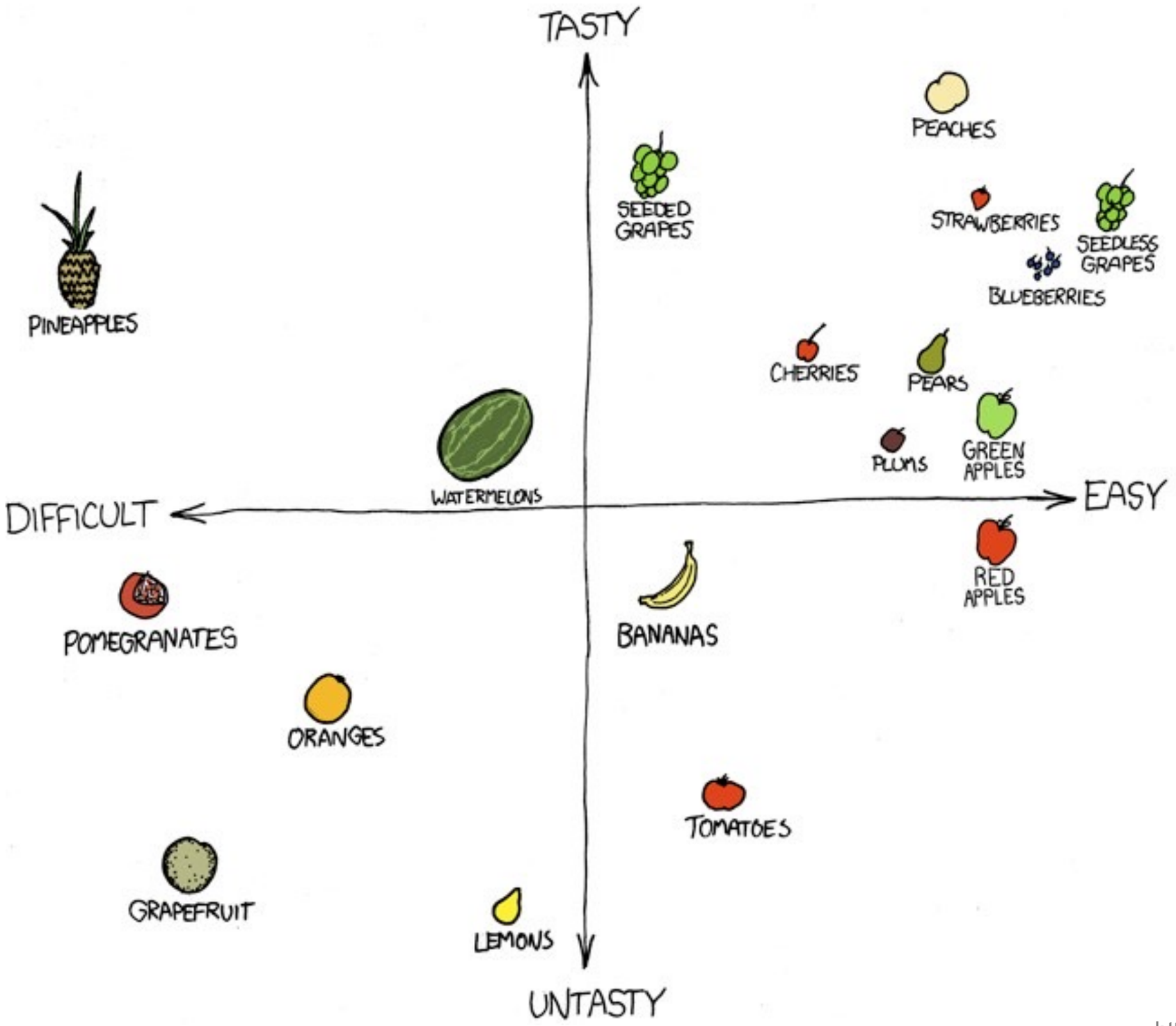
STACKED AREA CHART



S. Few

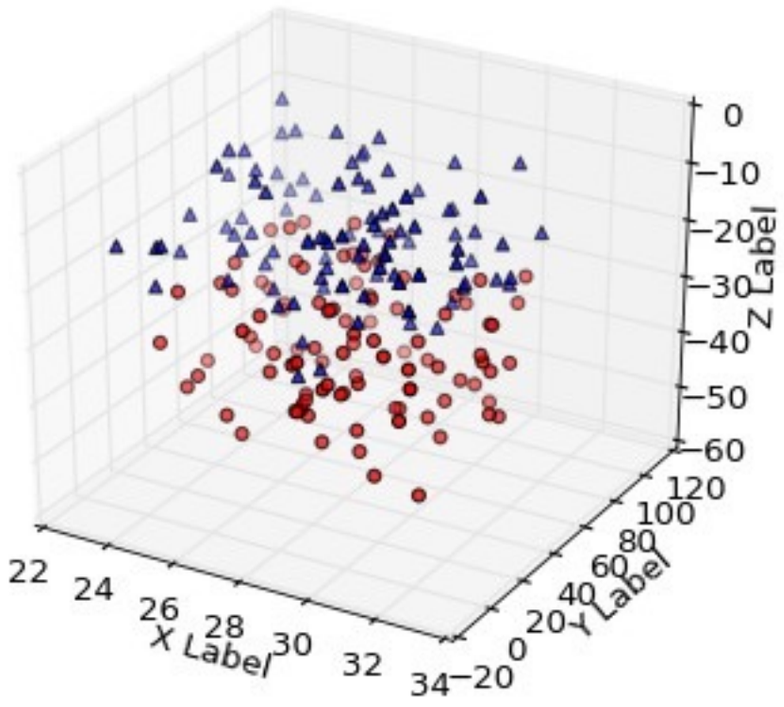
CORRELATIONS

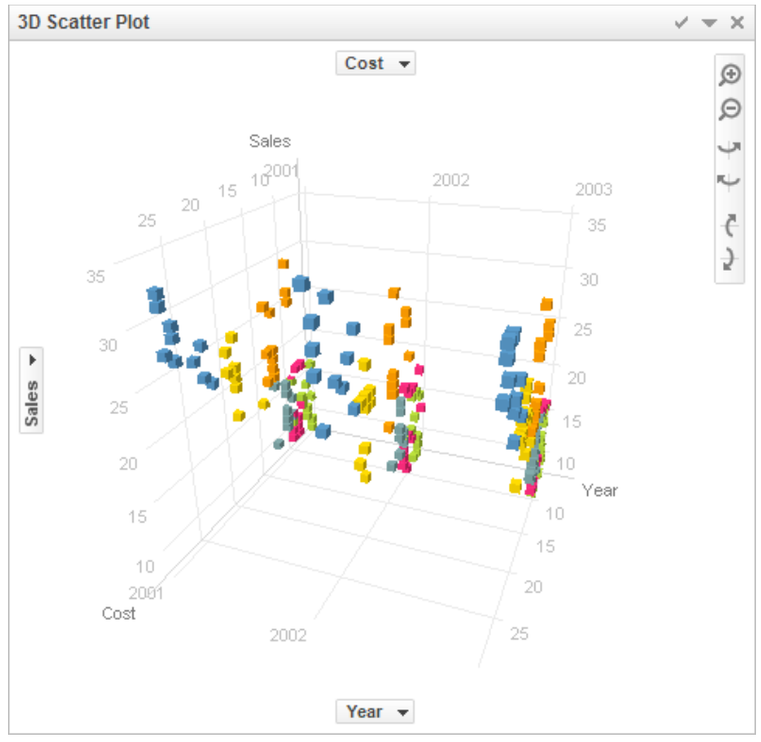
SCATTERPLOTS

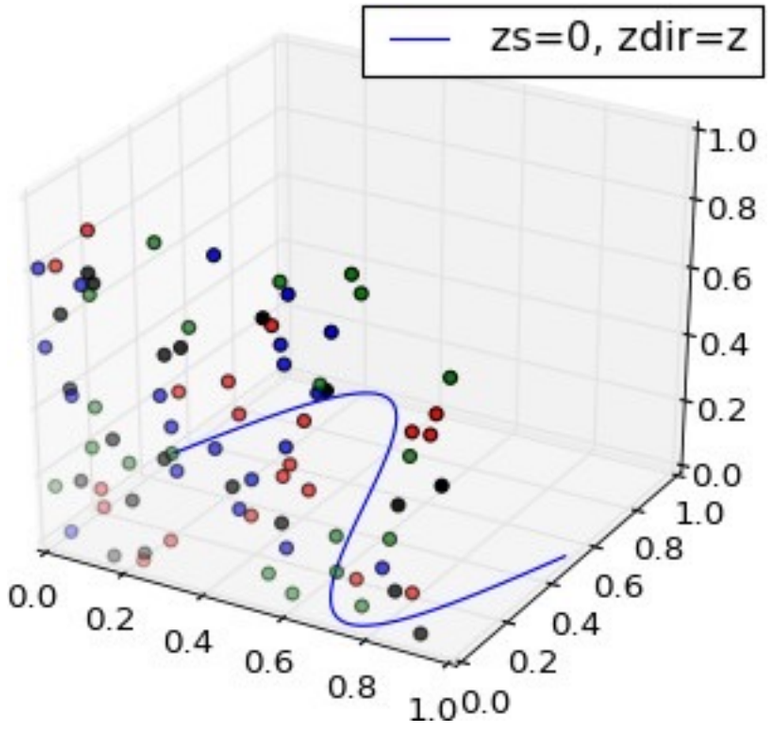


http://xkcd.com/388/

DON'T!



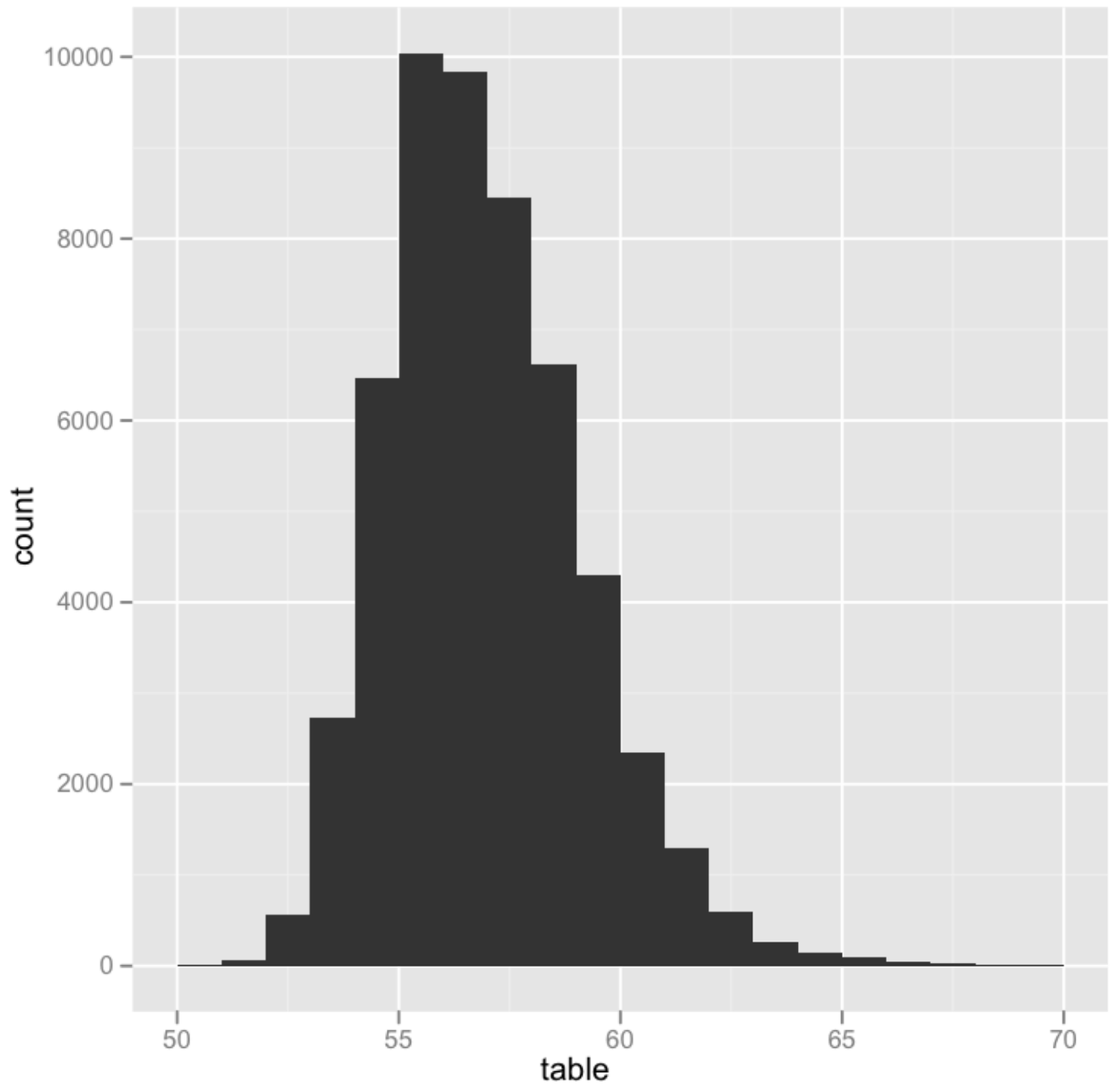




matplot3d tutorial

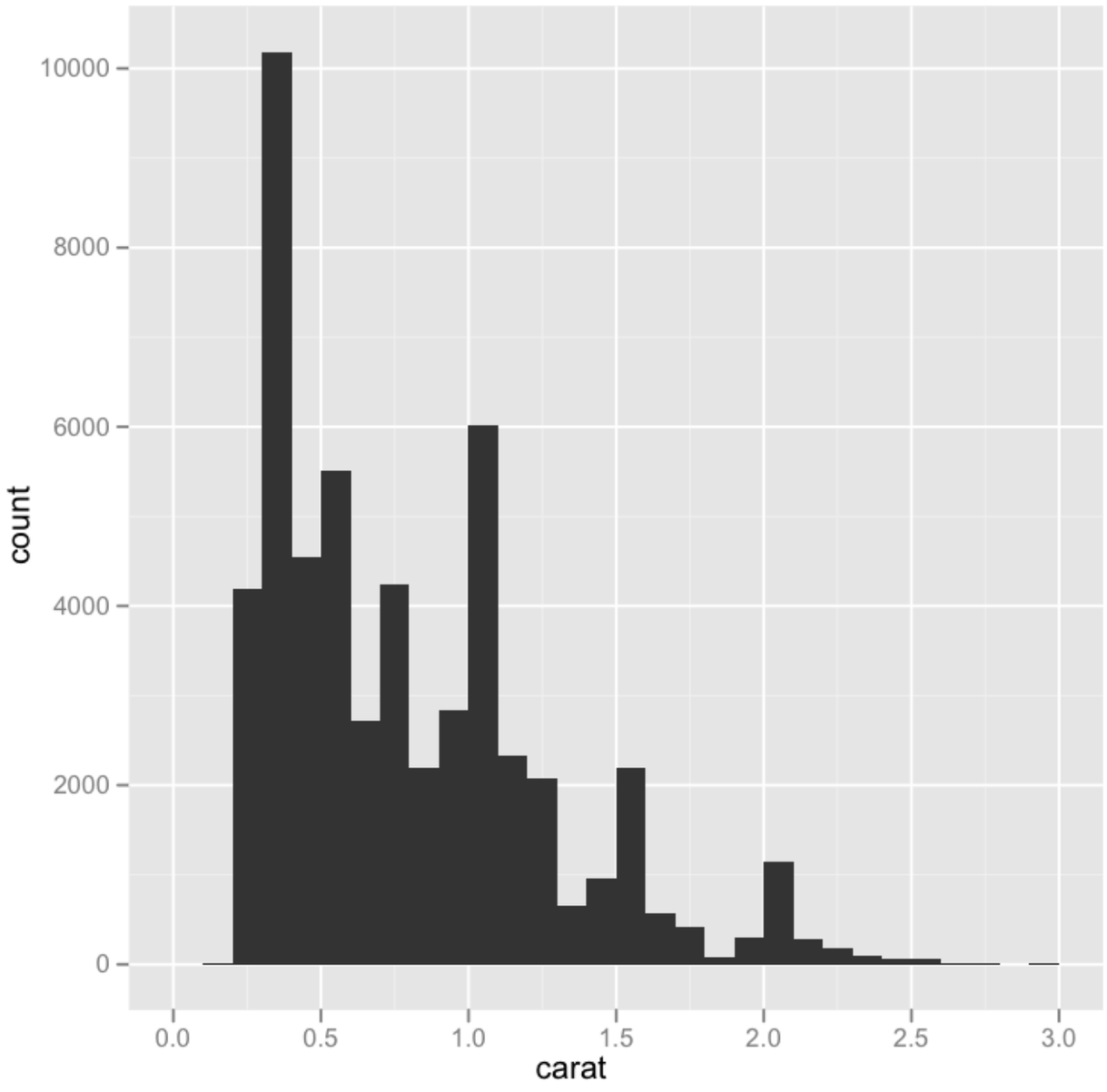
DISTRIBUTIONS

HISTOGRAM

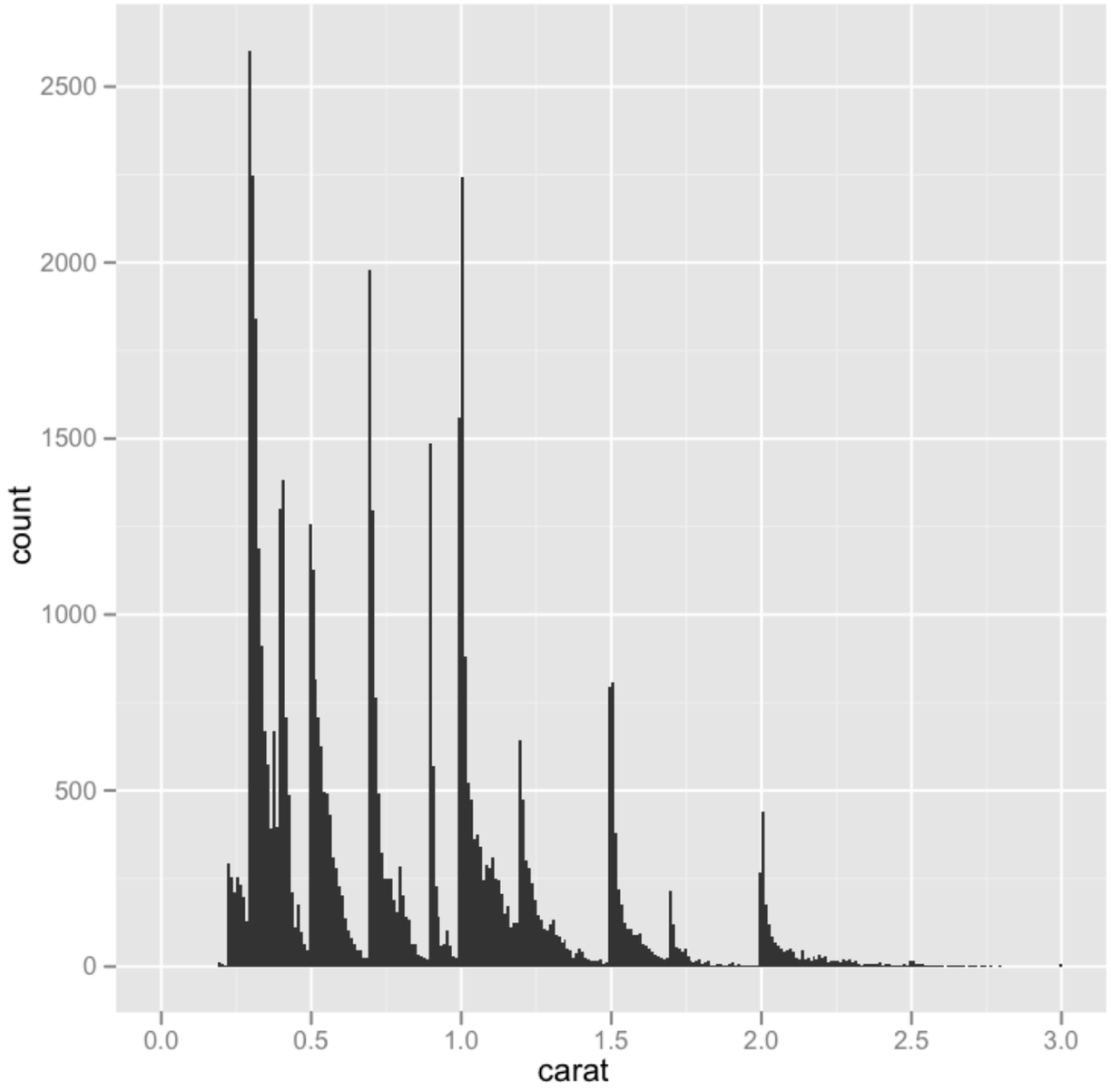


ggplot2

BIN WIDTH

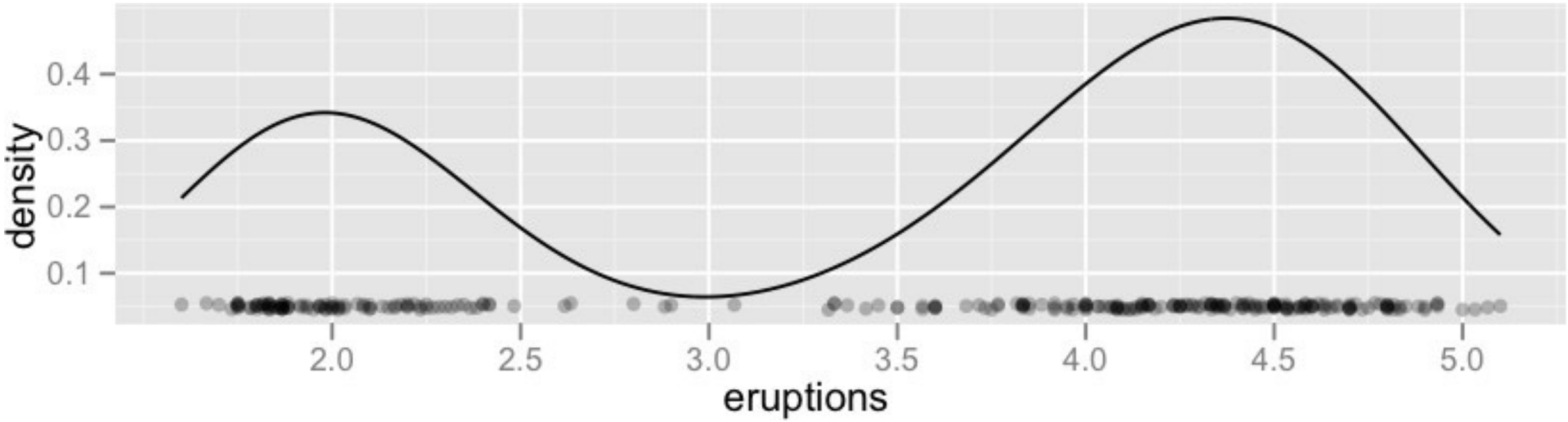


binwidth = 0.1

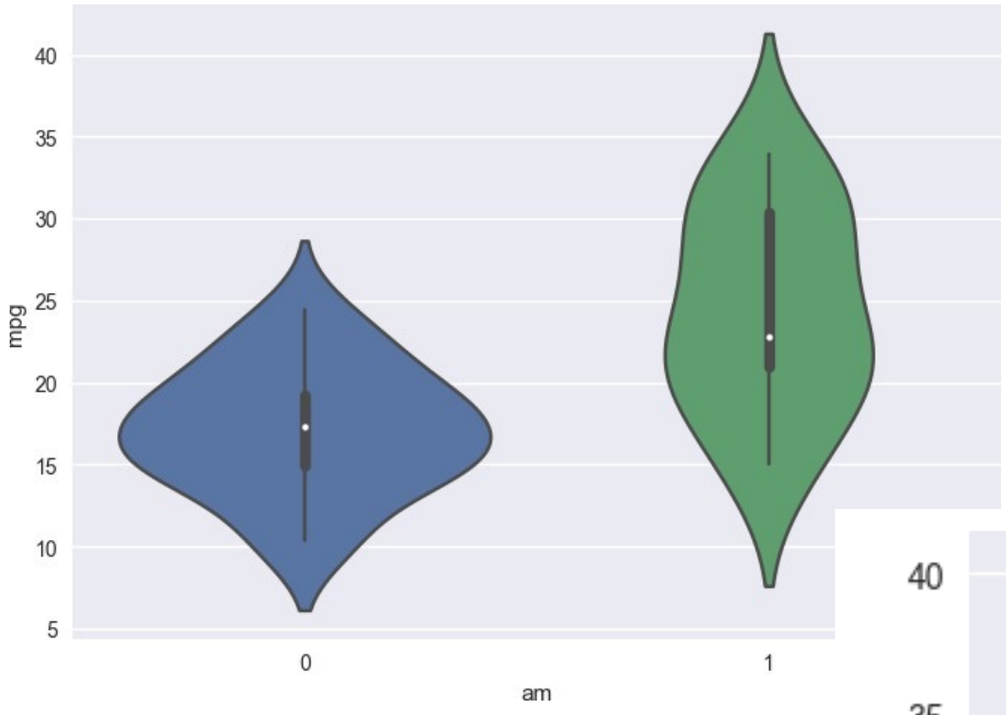


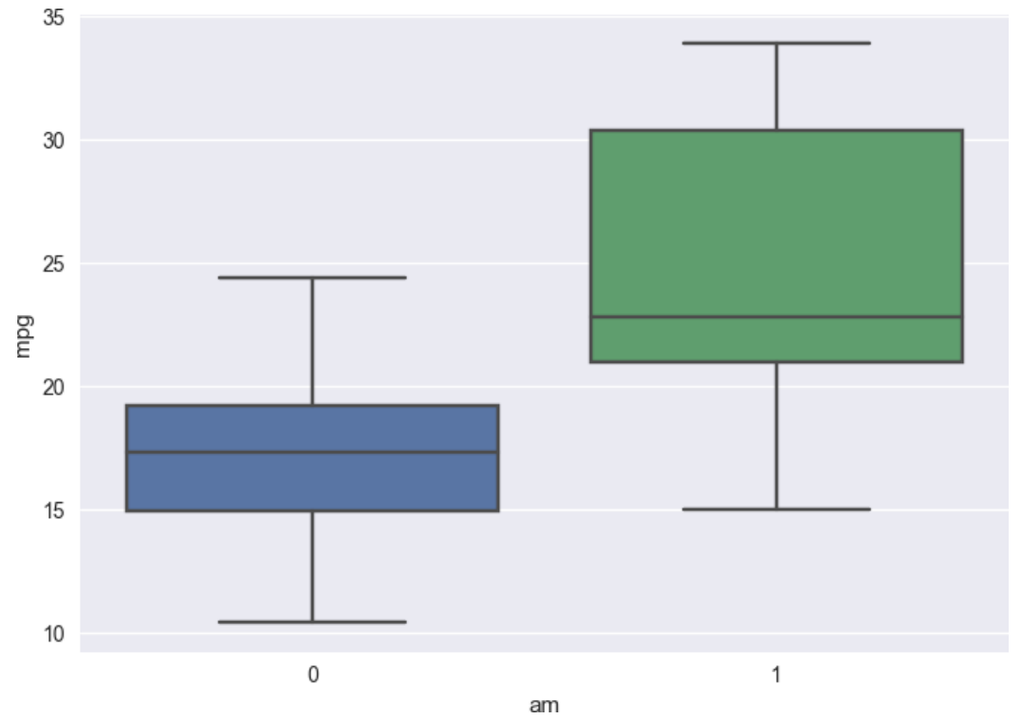
binwidth = 0.01

DENSITY PLOTS

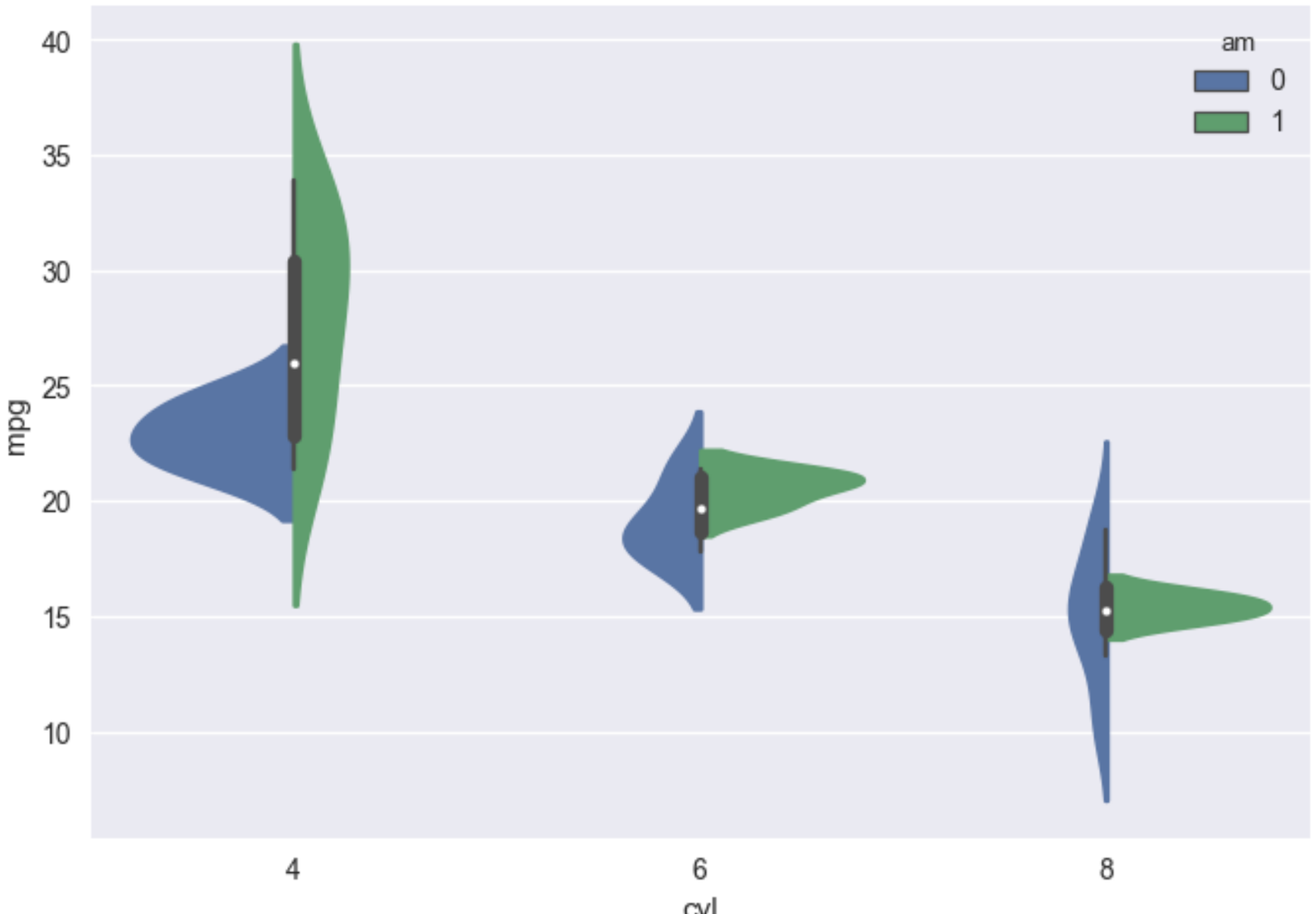


GROUP



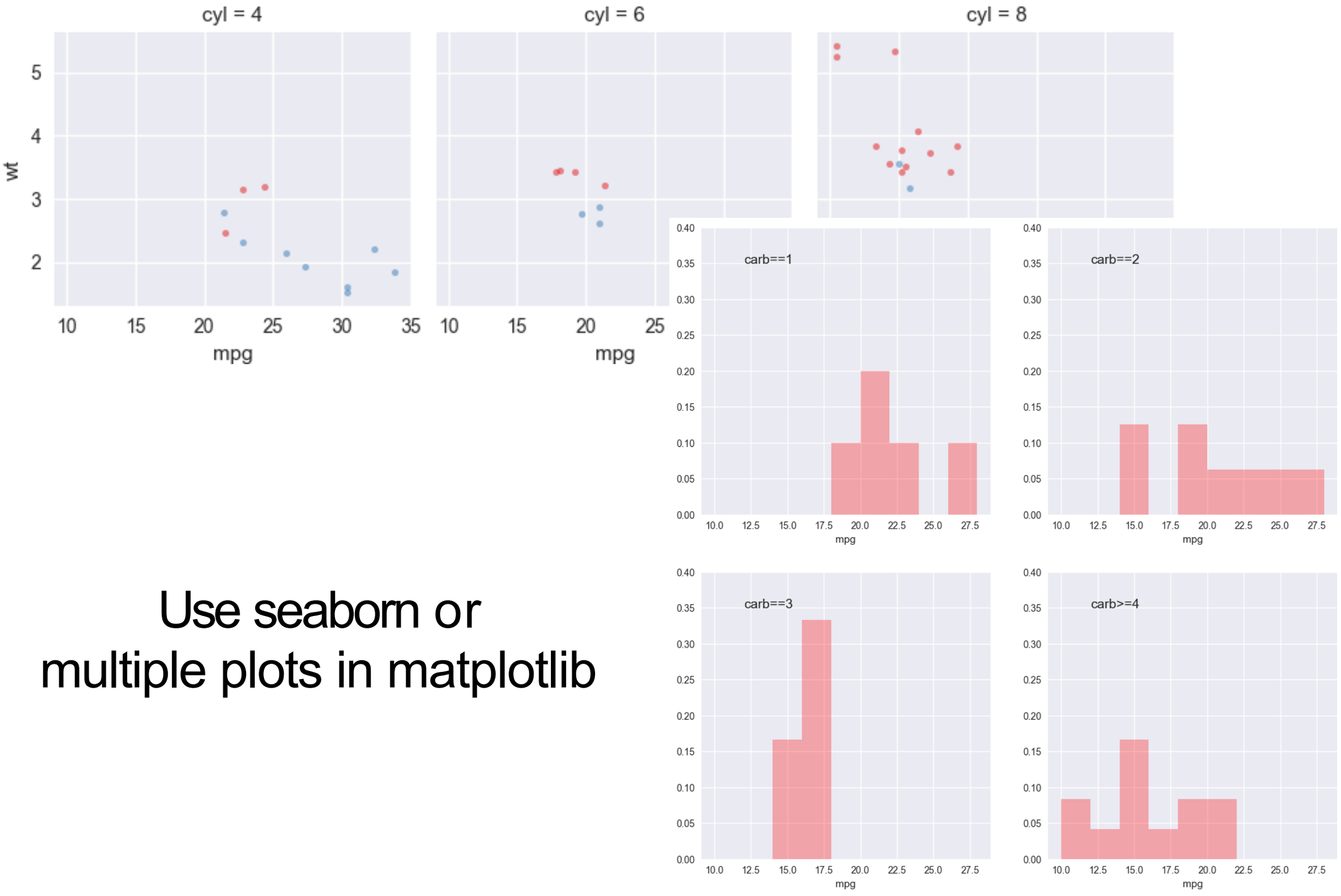


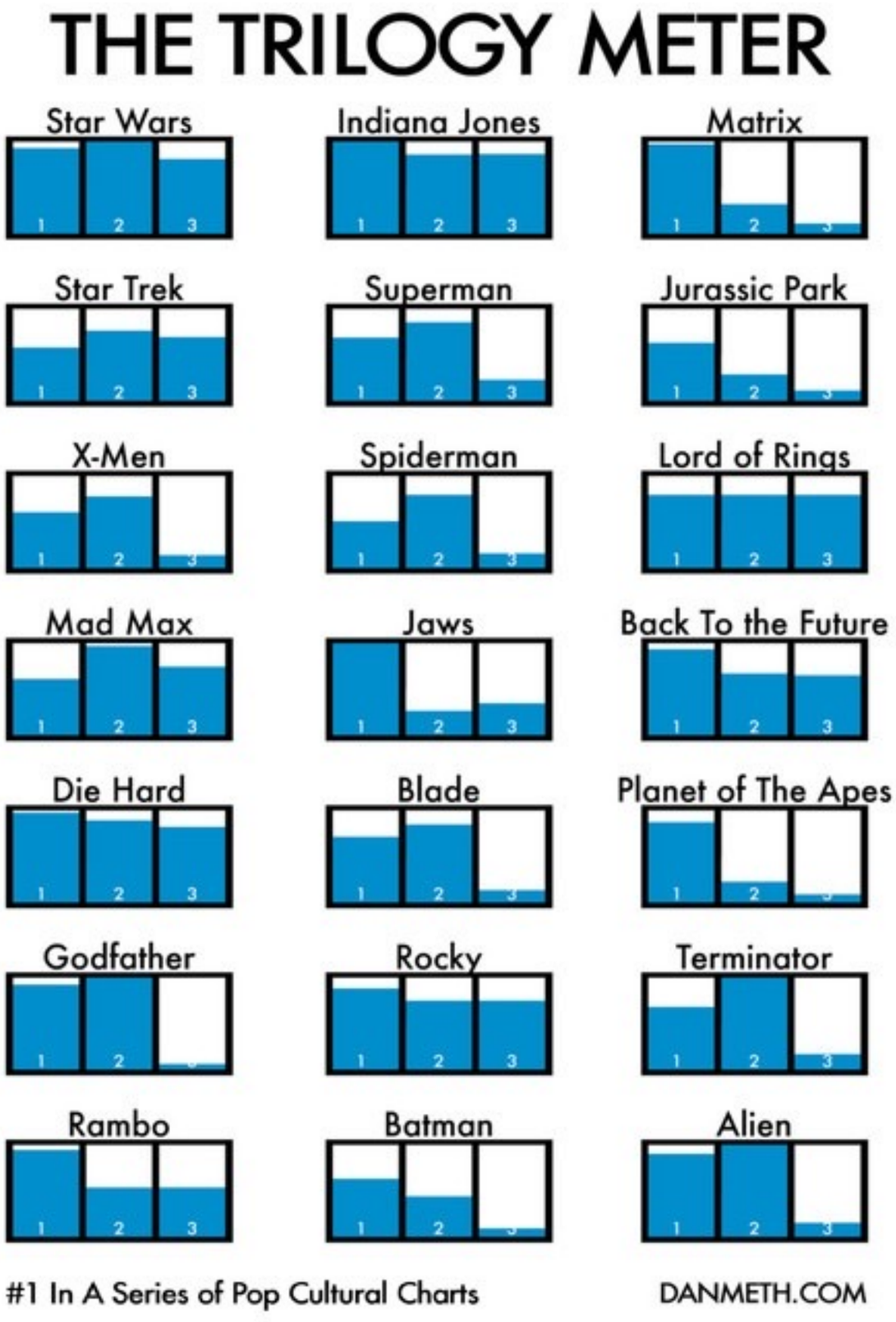
getting complex…



FACETING AND SMALL MULTIPLES

Use seaborn or
multiple plots in matplotlib

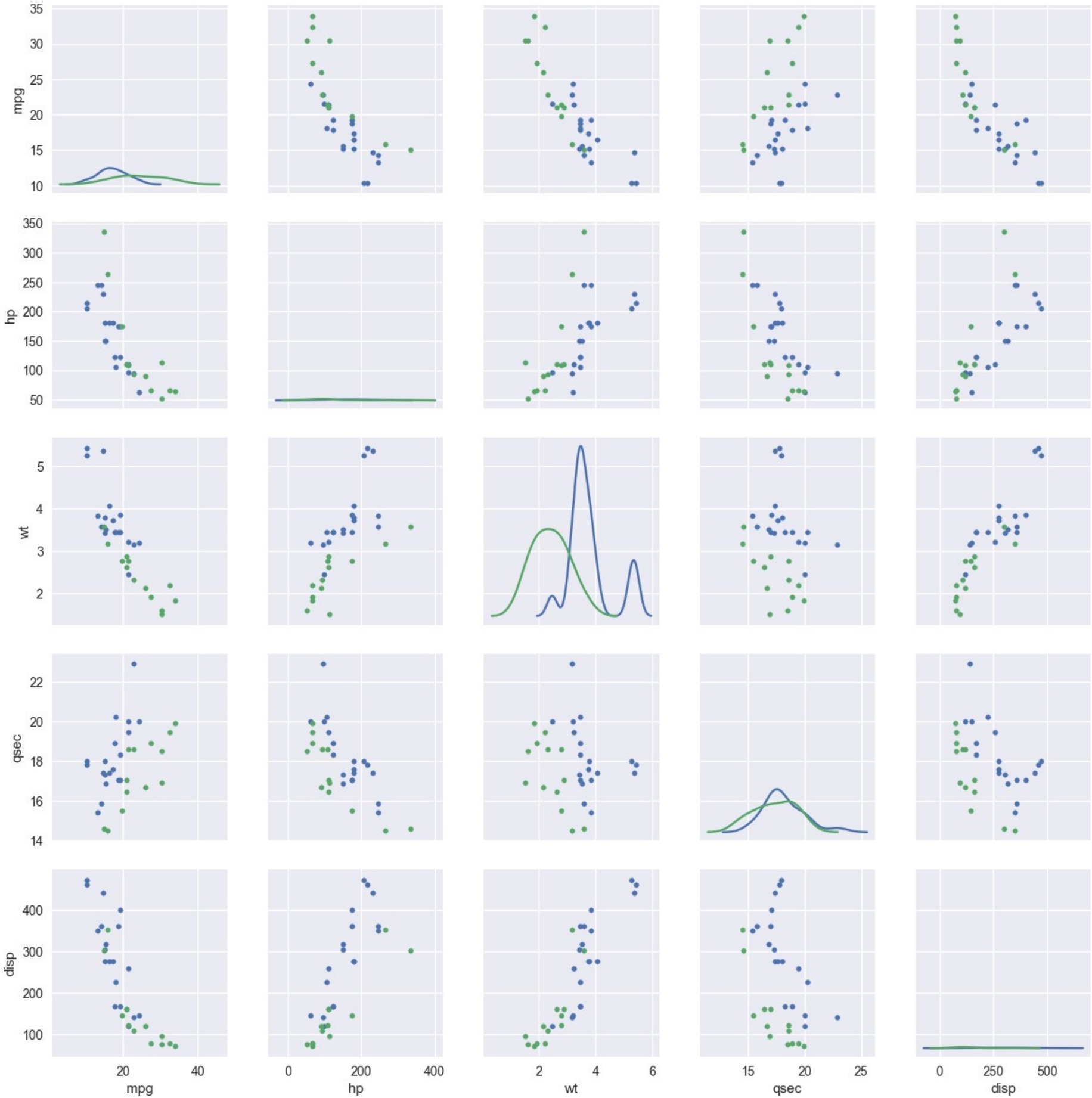




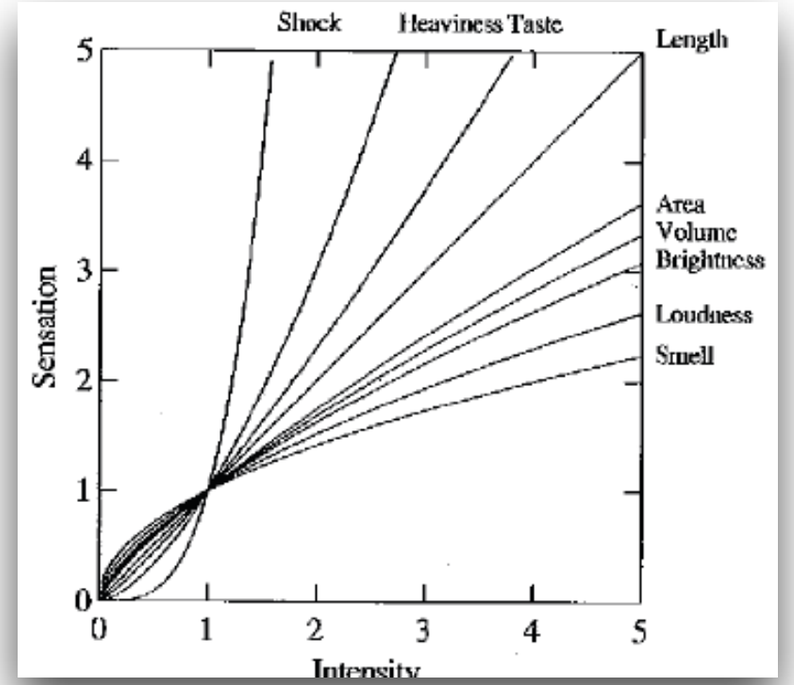
SMALL MULTIPLES

SPLOM

The scatterplot matrix



PERCEPTUAL EFFECTIVENESS



Stephen's Power Law, 1961

	Nominal	Ordinal	Quantitative
Position	✔	✔	✔
Size	✔	✔	~
(Grey)Value	✔	✔	~
Texture	✔	~	✖
Color	✔	✖	✖
Orientation	✔	✖	✖
Shape	✔	✖	✖

✔ = Good
~ = OK
✖ = Bad

J. Bertin, 1967

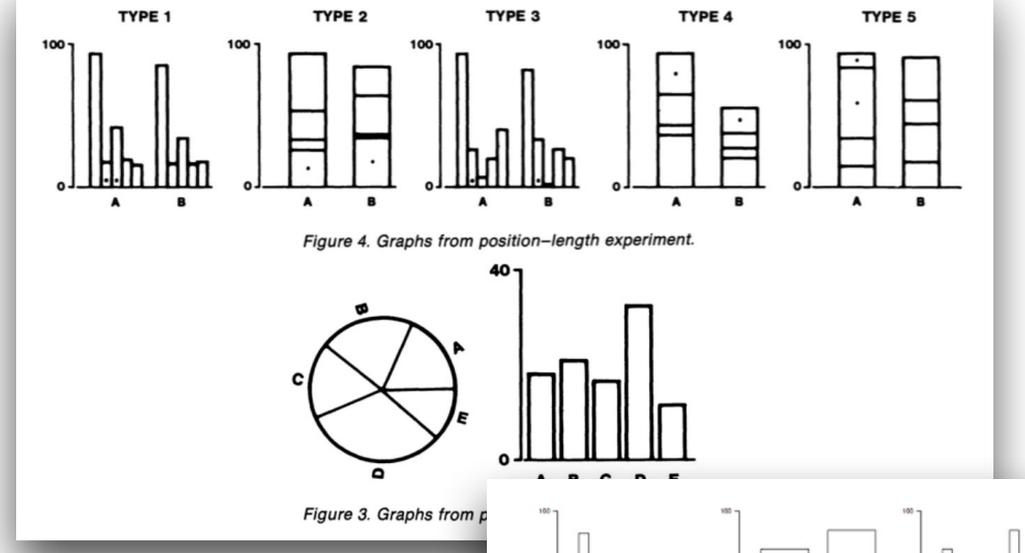
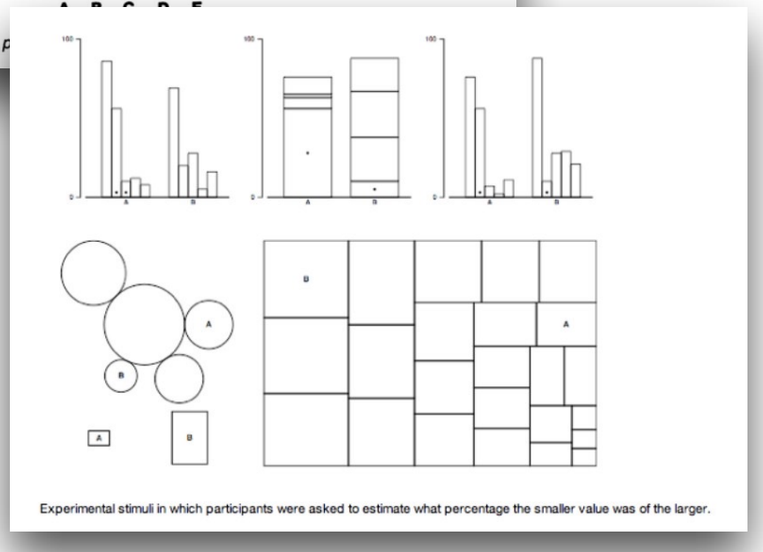


Figure 4. Graphs from position–length experiment.

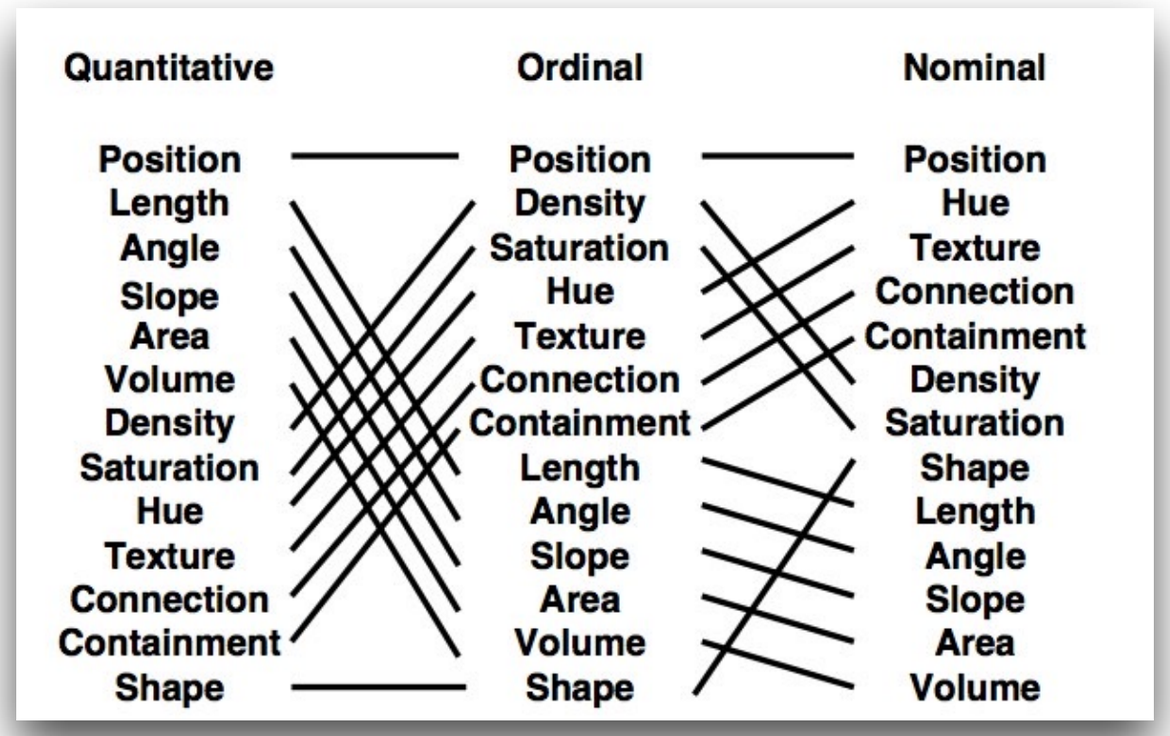
Figure 3. Graphs from p

Cleveland / McGill, 1984



Experimental stimuli in which participants were asked to estimate what percentage the smaller value was of the larger.

Heer / Bostock, 2010



J. Mackinlay, 1986

HOW MUCH LONGER?

A

B

HOW MUCH LONGER?

A

B

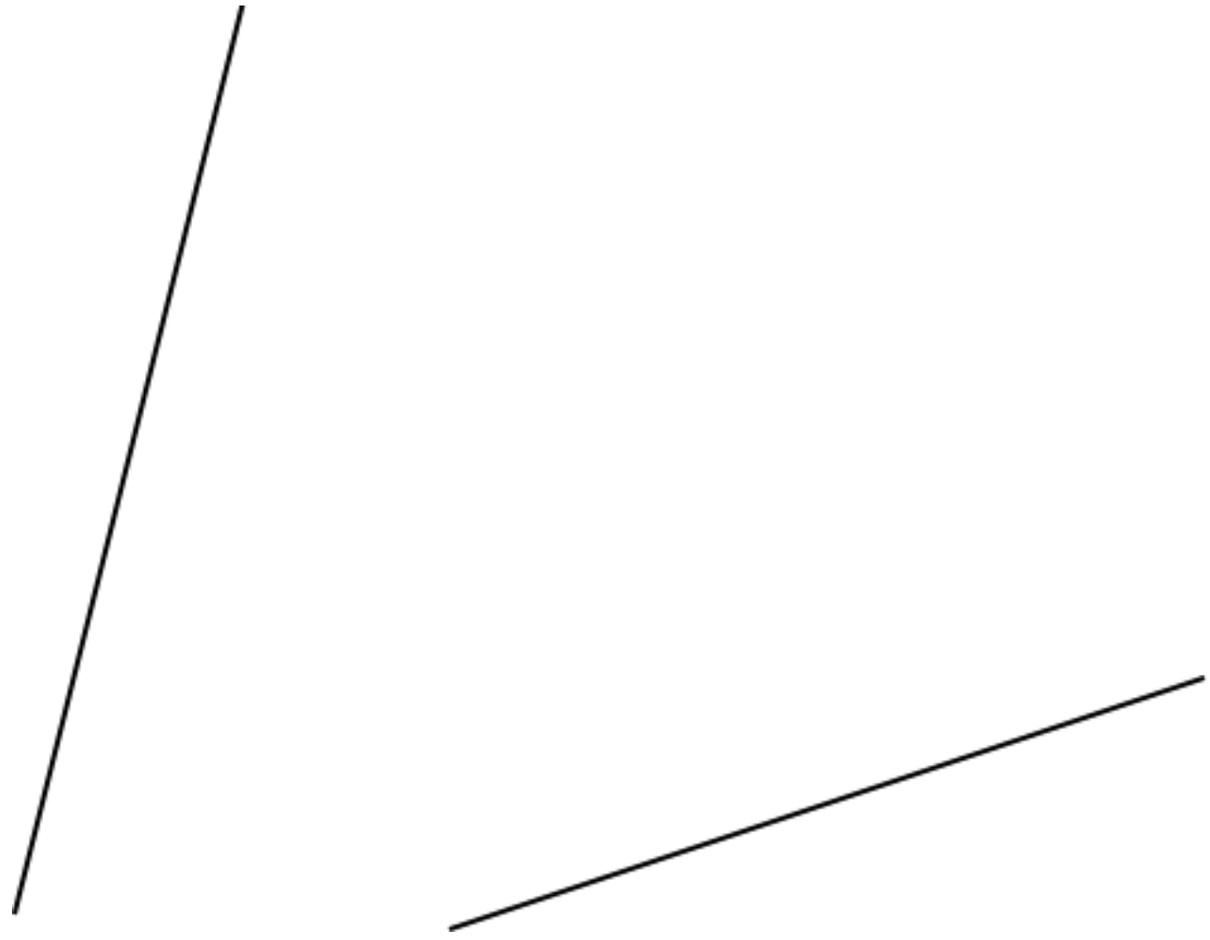
4x

HOW MUCH STEEPER SLOPE?

A

B

HOW MUCH STEEPER SLOPE?

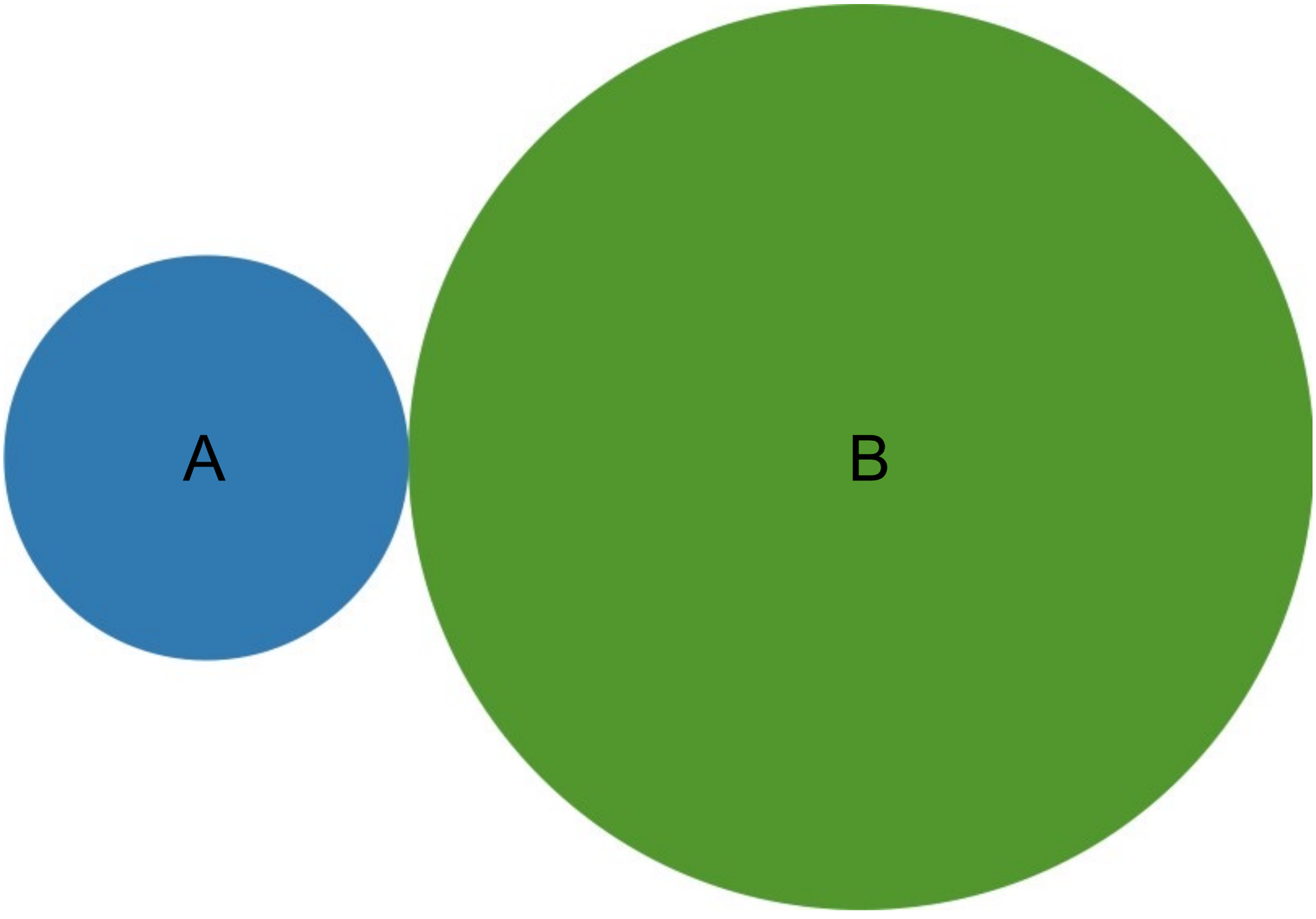


A

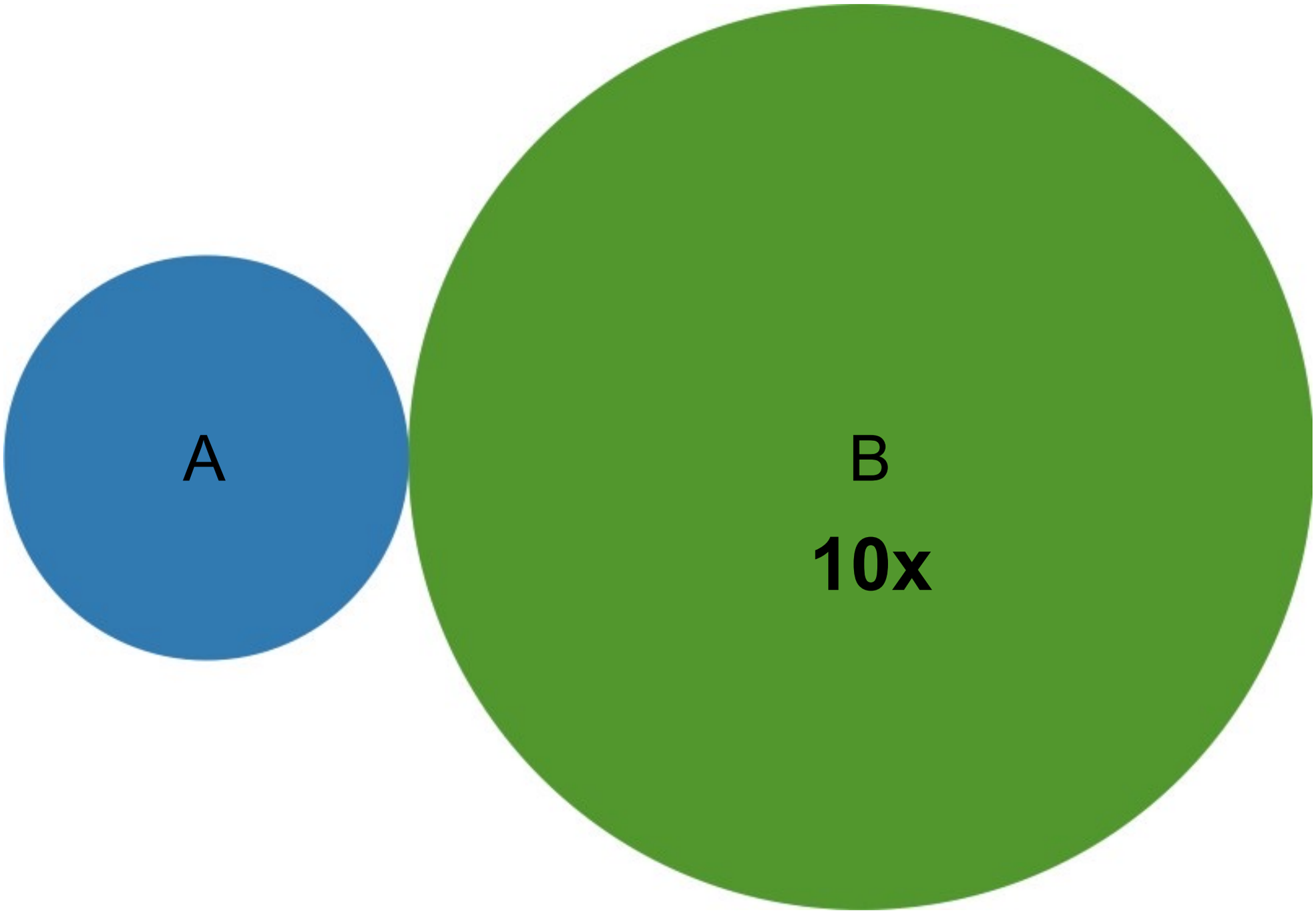
B

4x

HOW MUCH LARGER AREA?



HOW MUCH LARGER AREA?



HOW MUCH DARKER?

A

B

HOW MUCH DARKER?

A

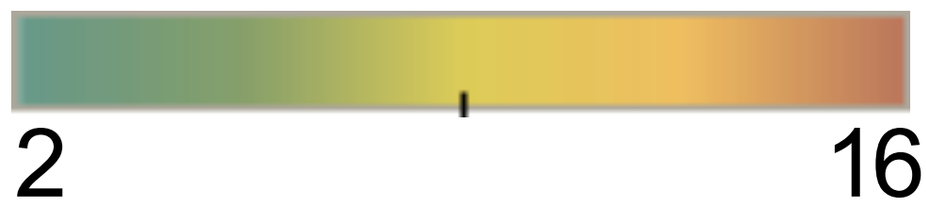
B

2x

HOW MUCH BIGGER VALUE?

A

B



HOW MUCH BIGGER VALUE?

A

B

4x

Most Efficient

Position

Length

Slope

Angle

Area

Intensity

Color

Shape

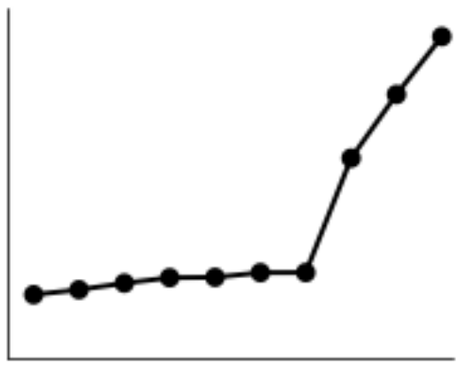
Least Efficient

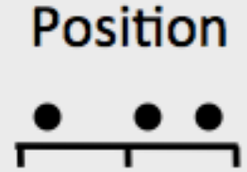
Quantitative

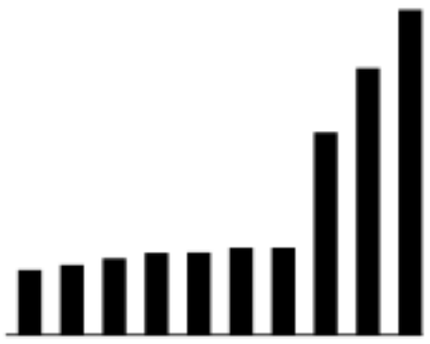
Ordered

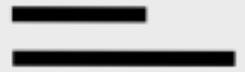
Categories

MOST EFFECTIVE

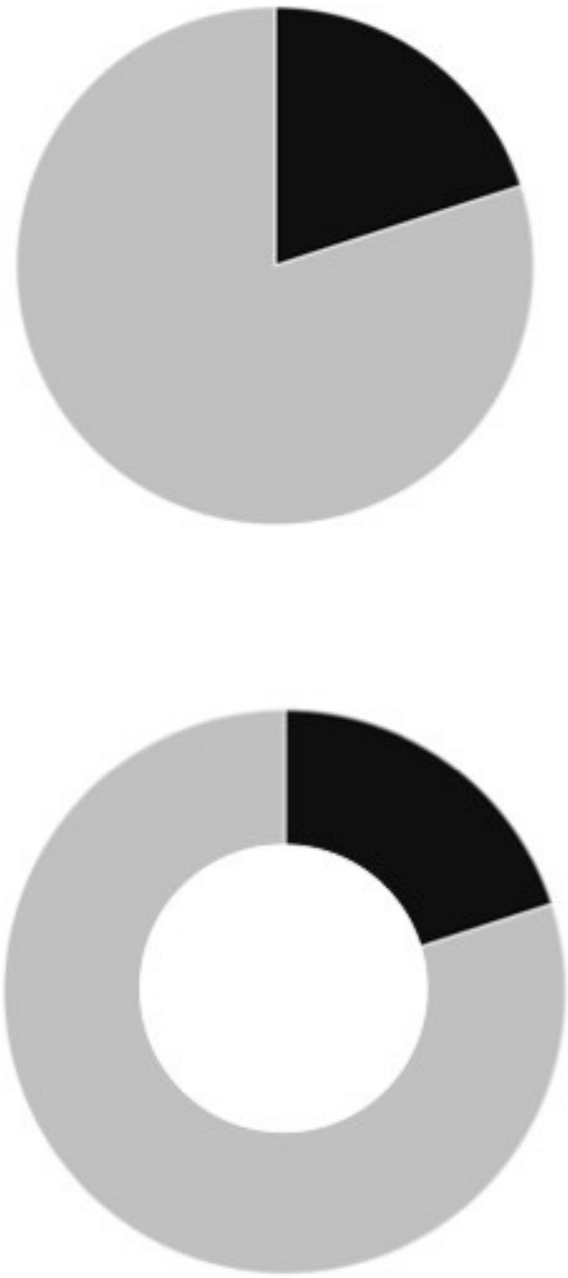


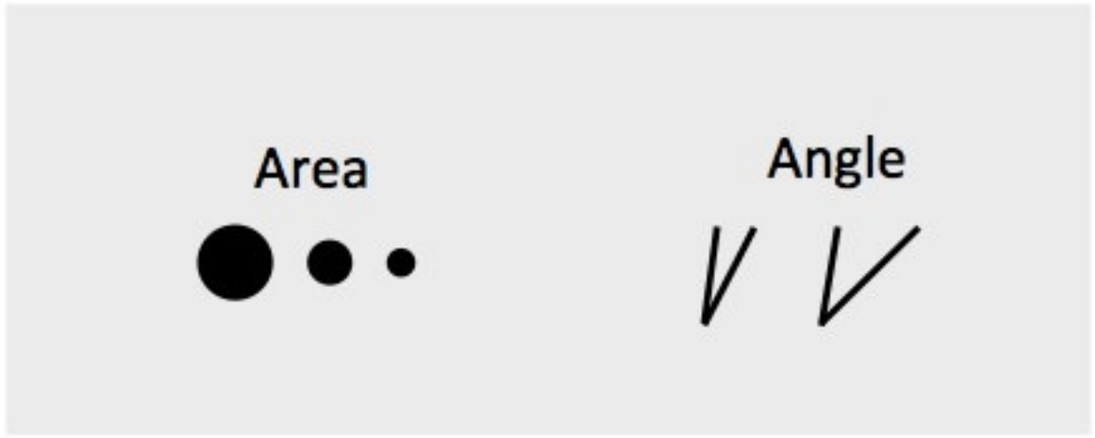


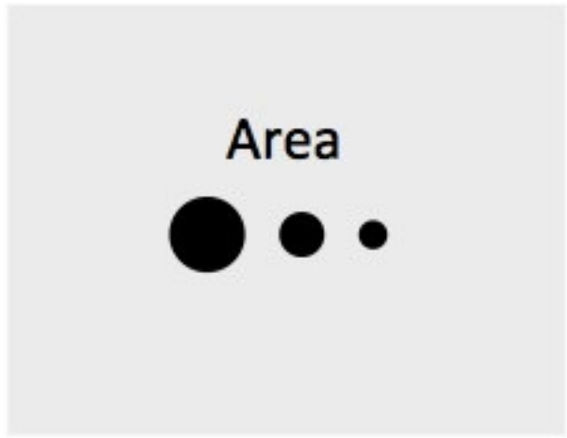




LESS EFFECTIVE

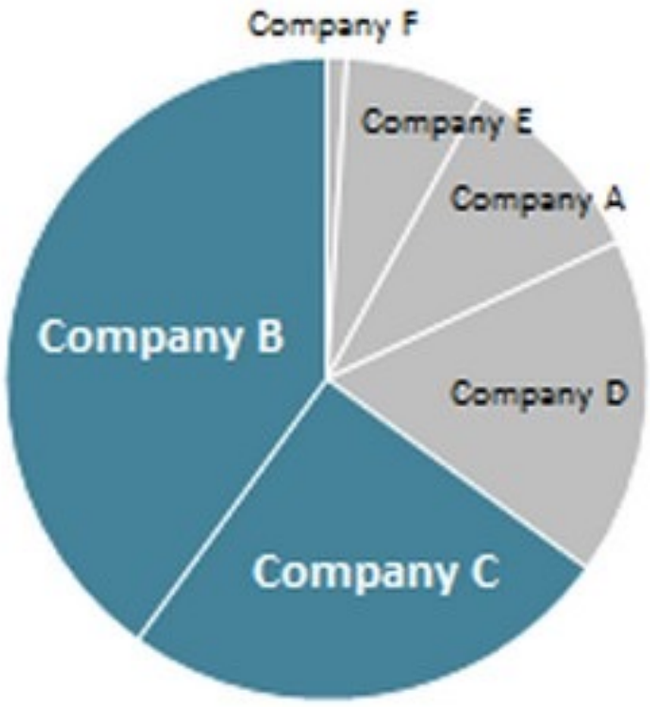


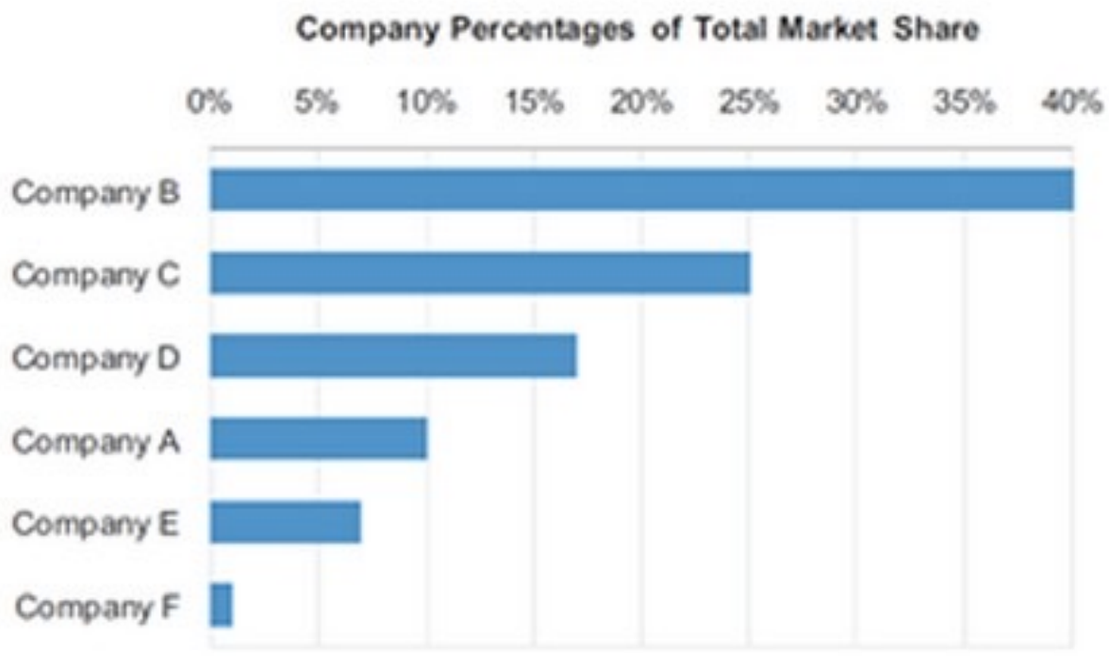




PIE vs BAR CHARTS

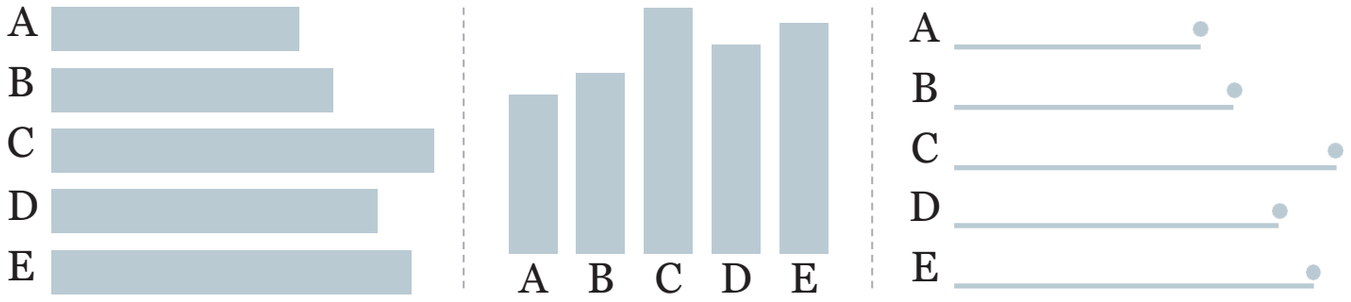
65% of the market is controlled by companies B and C



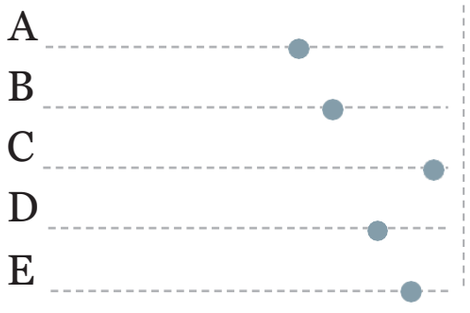


Data visualization and visual encoding

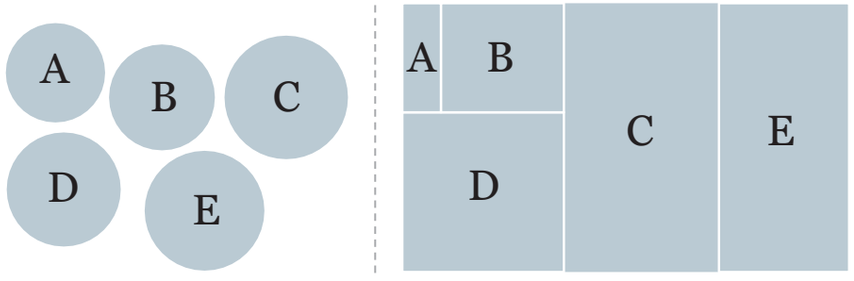
LENGTH OR HEIGHT



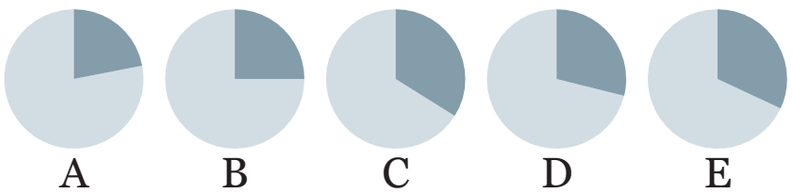
Position



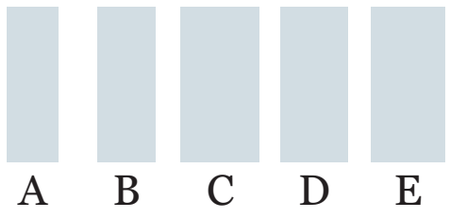
Area



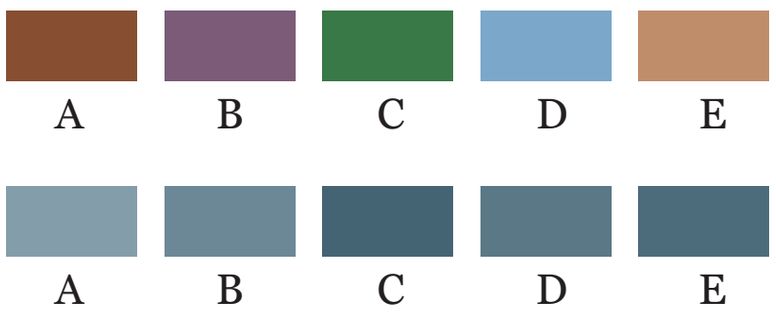
Angle/area



Line weight



Hue and shade



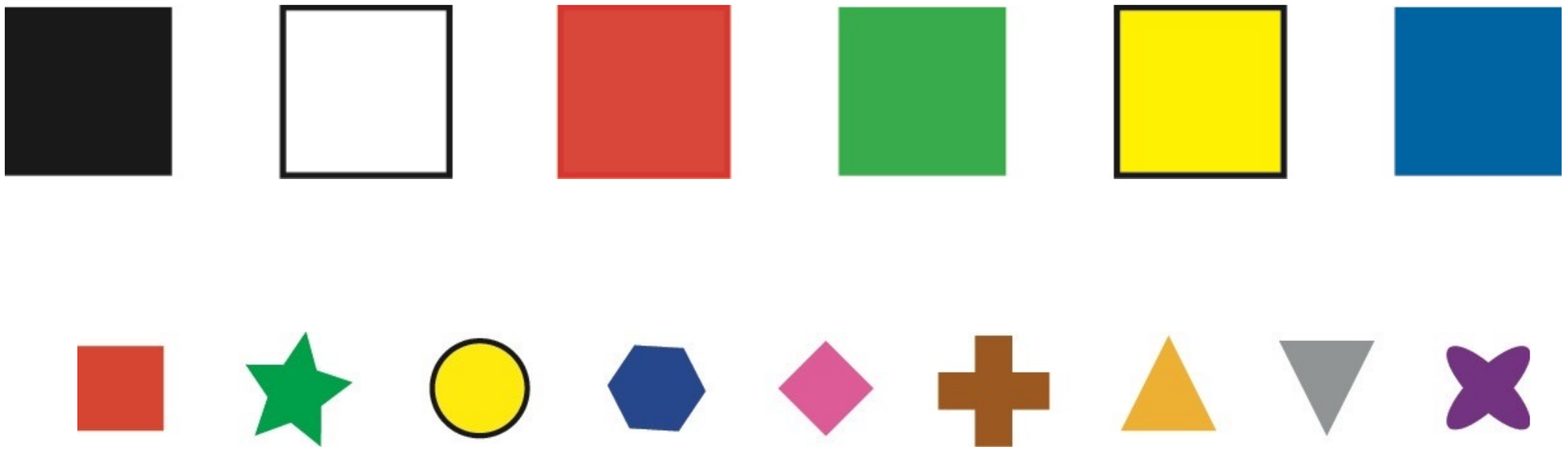
Figures represented in all these graphics: 22%, 25%, 34%, 29%, 32%

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

4. USE COLOR STRATEGICALLY

Colors for Categories

Do not use more than 5-8 colors at once!



Ware, "Information Visualization"

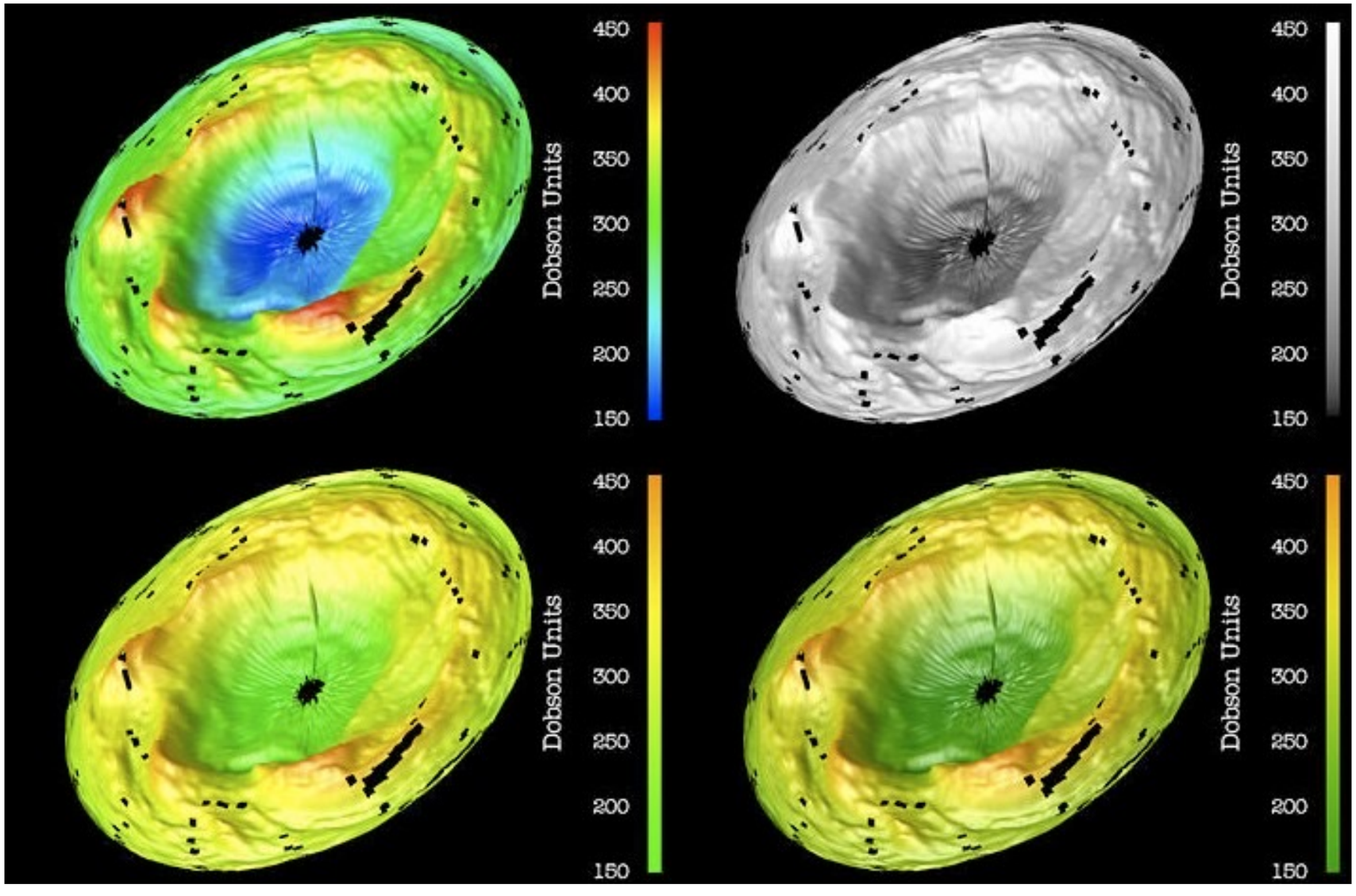
Colors for Ordinal Data

Vary luminance and saturation

Zeilis et al, 2009, "Escaping RGBland: Selecting Colors for Statistical Graphics"

COLORS FOR QUANTITATIVE DATA

Hue
(Rainbow)

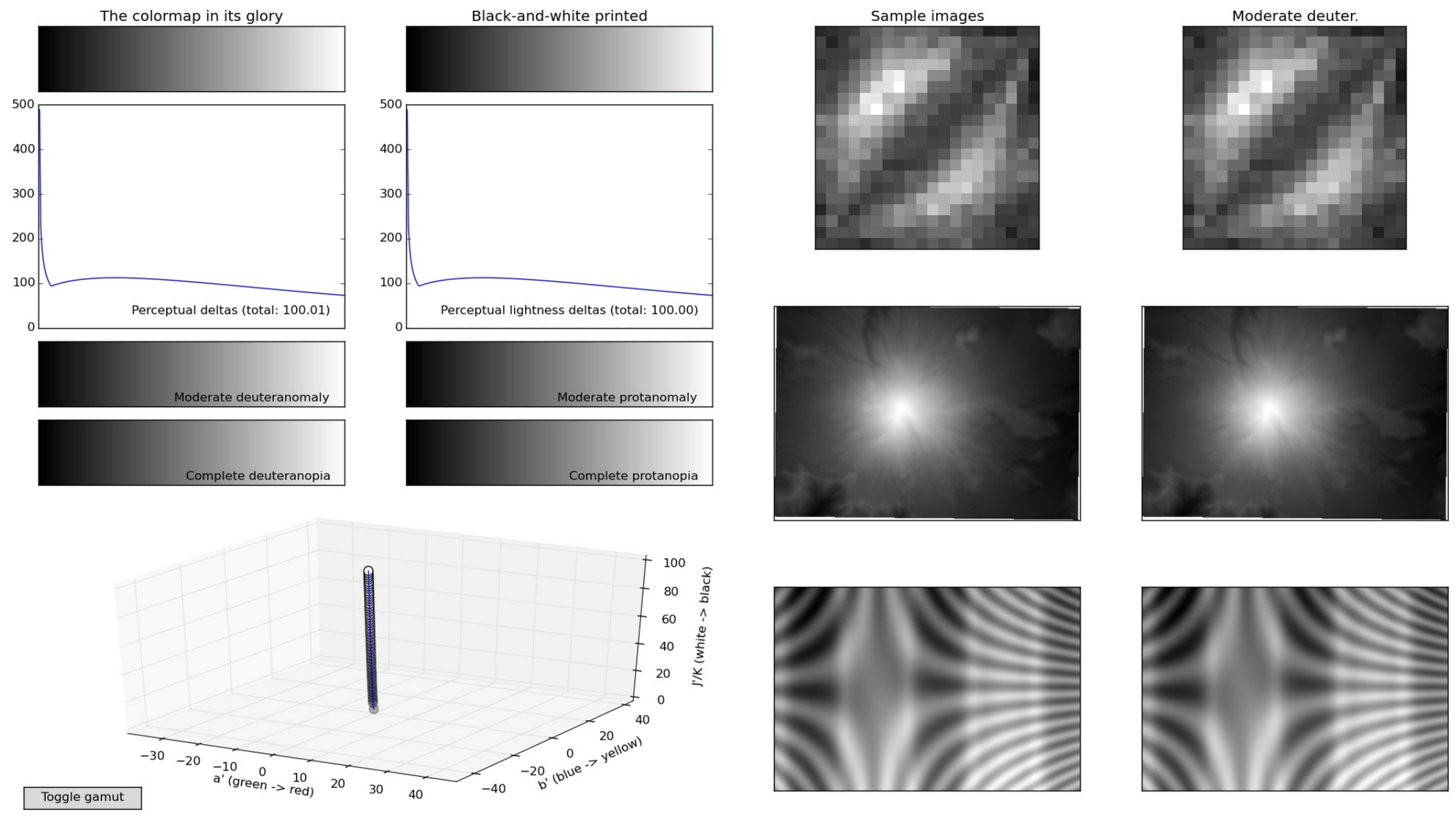


Luminance

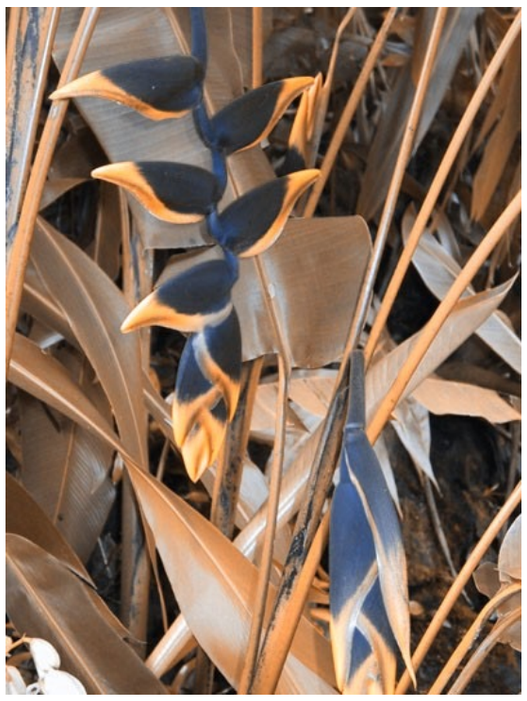
Luminance
& Hue

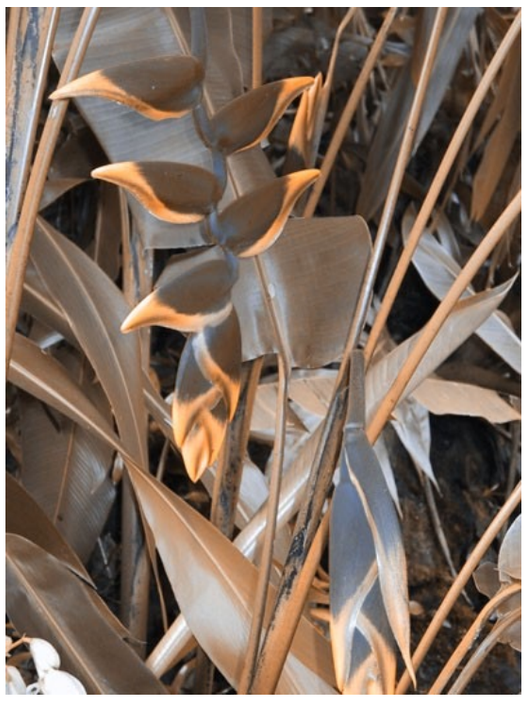
Rogowitz and Treinish, Why should engineers and scientists be worried about color?

GRAY


Colormap evaluation: gray
The colormap in its glory
Black-and-white printed
Sample images
Moderate deuter.
500
400
300
200
100
0
Perceptual deltas (total: 100.01)
Perceptual lightness deltas (total: 100.00)
Moderate deuteranomaly
Moderate protanomaly
Complete deuteranopia
Complete protanopia
100
80
60
40
20
0
J'/K (white -> black)
−30
−20
−10
0
10
20
30
40
a' (green -> red)
−40
−20
0
20
40
b' (blue -> yellow)
Toggle gamut

COLOR BLINDNESS





Protanope

Deuteranope

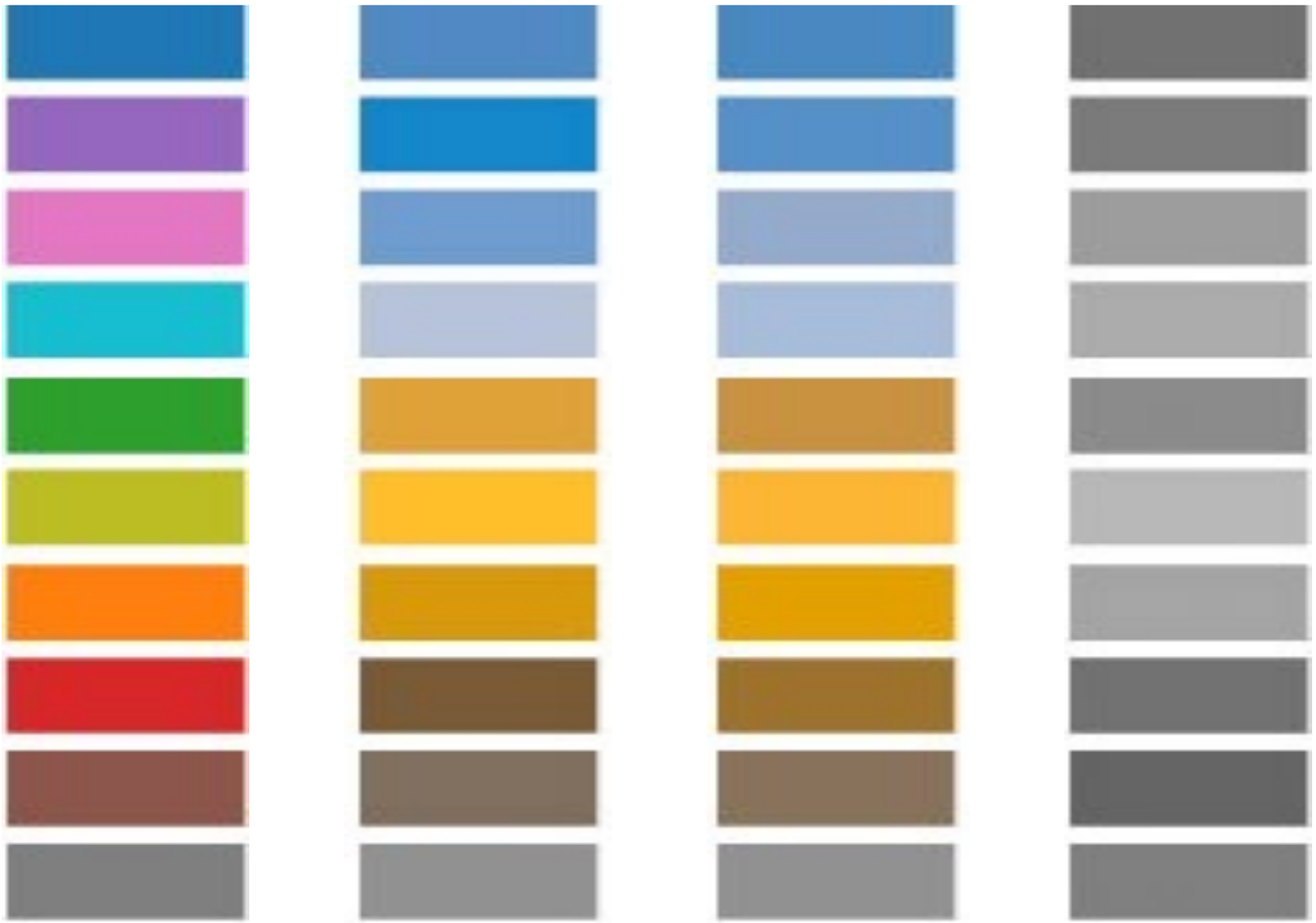
Tritanope

Red /green deficiencies

Blue / Yellow deficiency

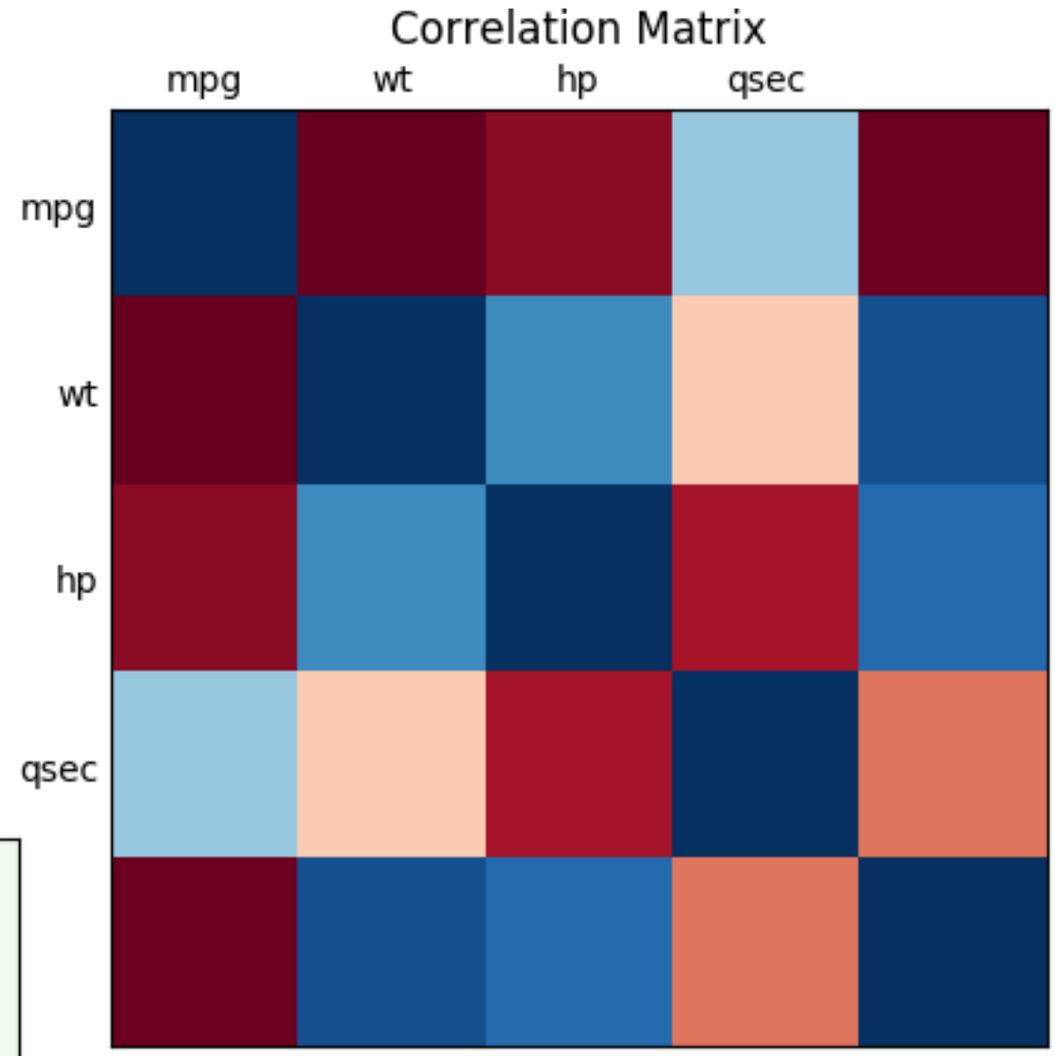
Based on slide from Stone

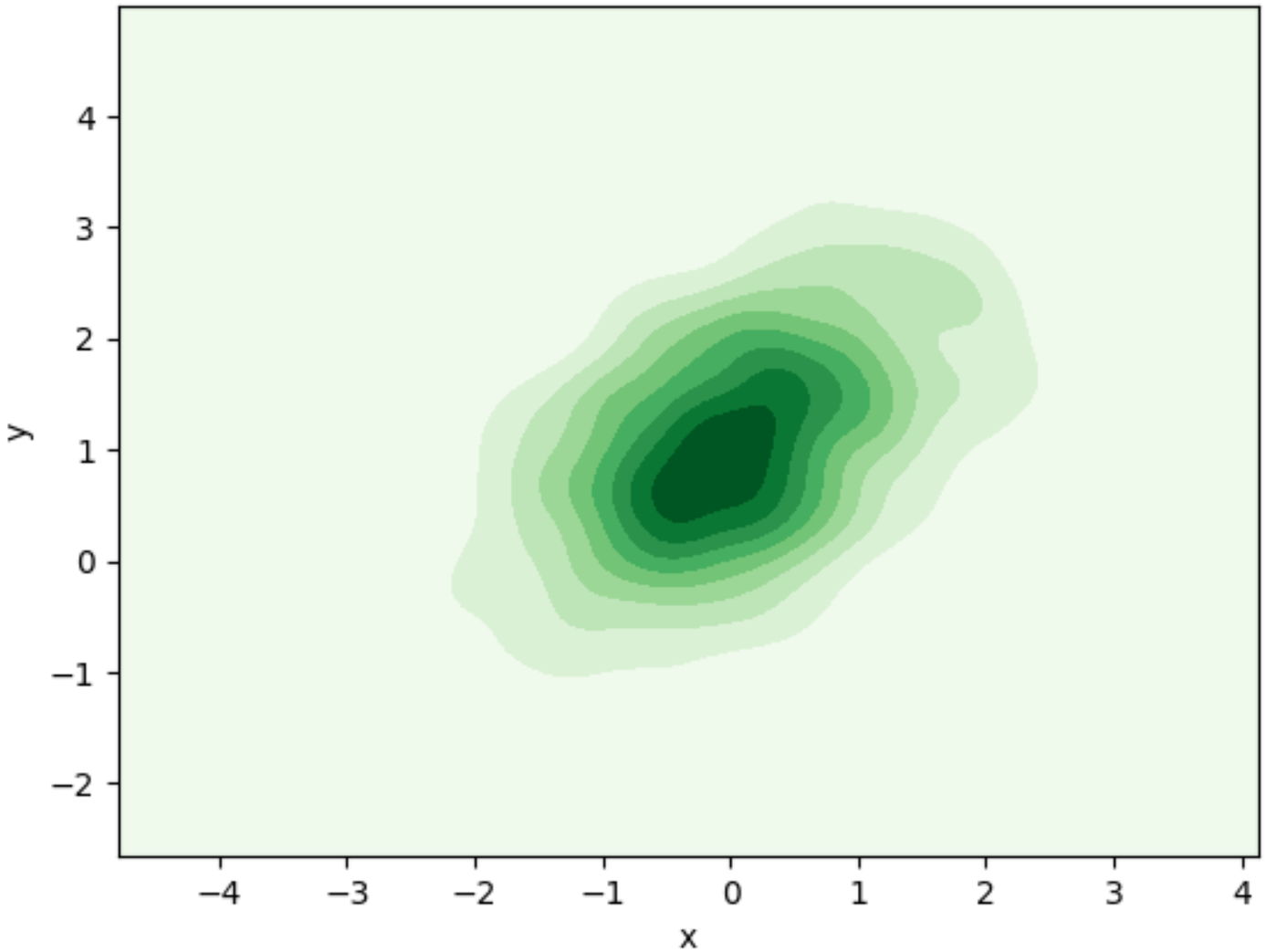
COLOR BLINDNESS



Based on slide from Stone

Diverging Palette for Quantitative or Ordinal





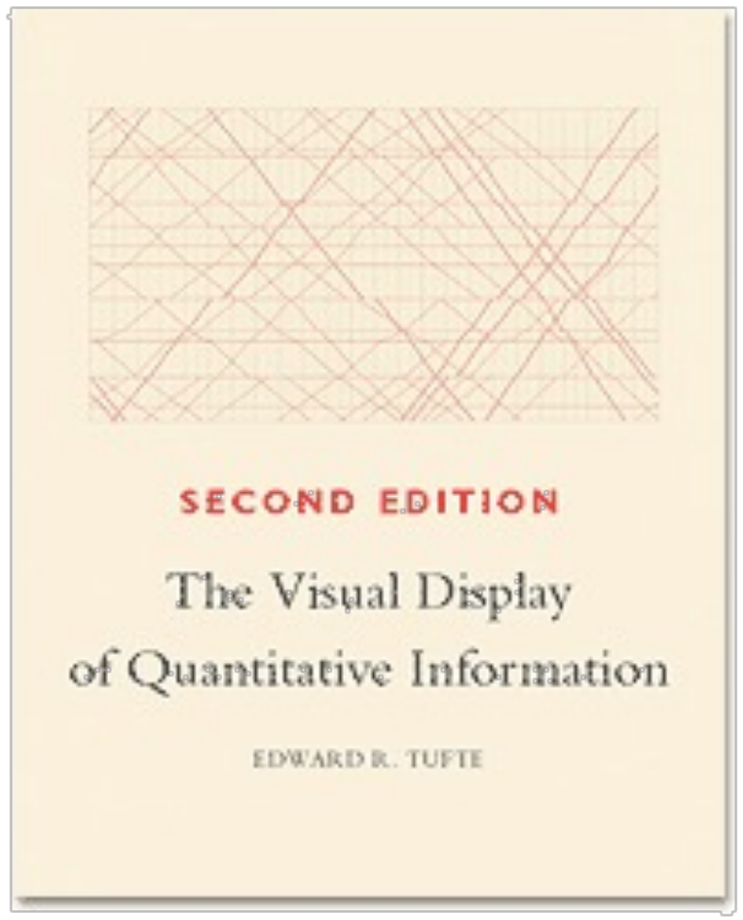
Sequential Palette for Densities

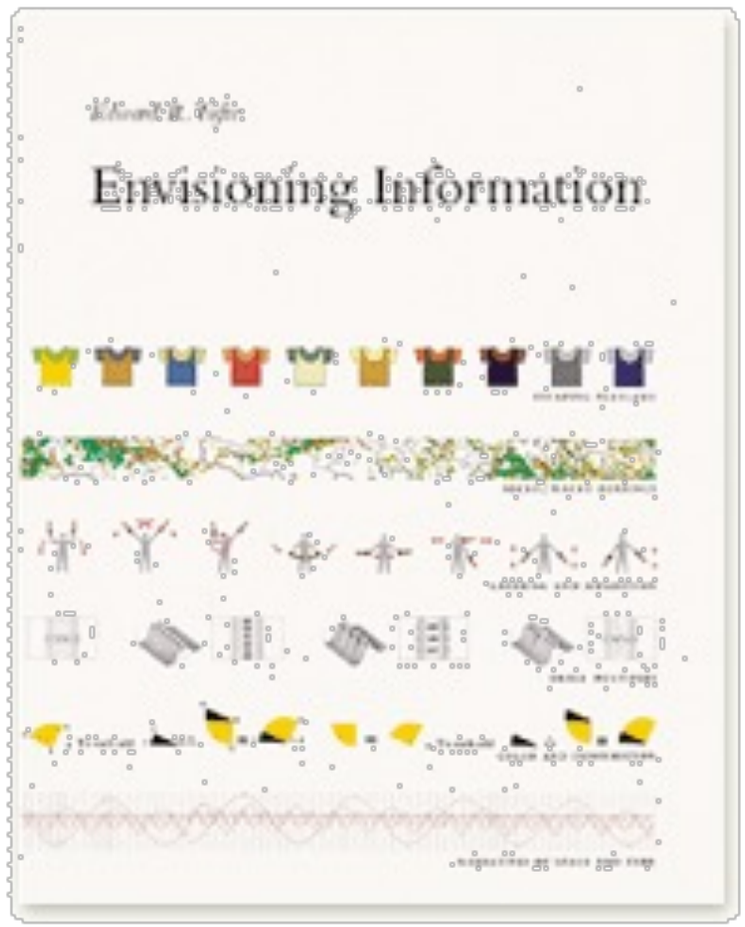
SUMMARY: EFFECTIVE VISUALIZATIONS

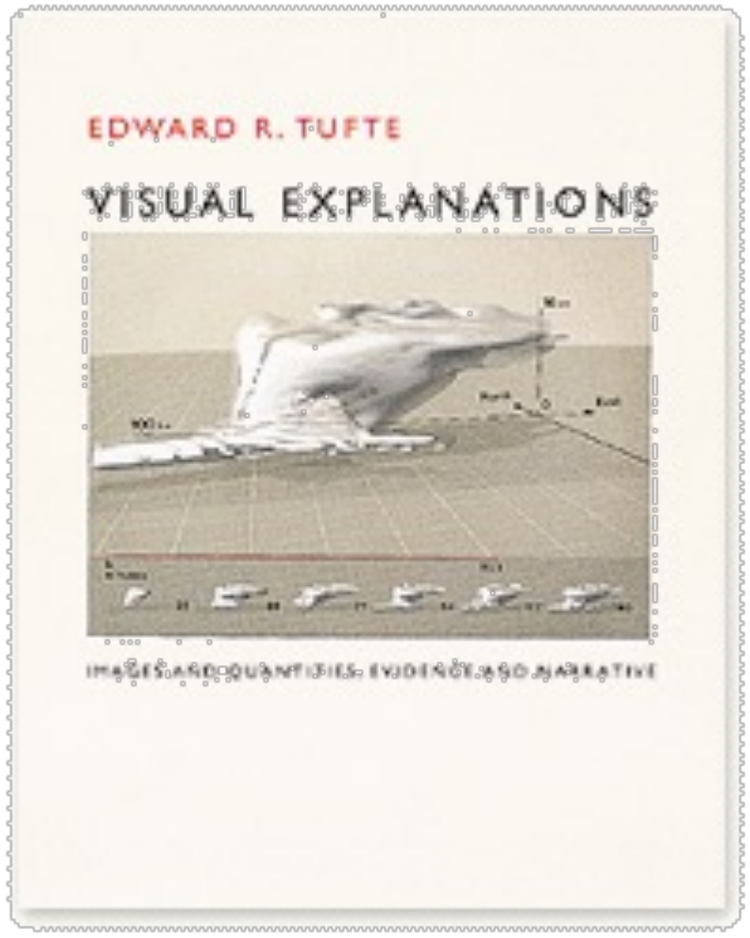
1. Have graphical integrity
2. Keep it simple
3. Use the right display
4. Use color strategically

FURTHER READING

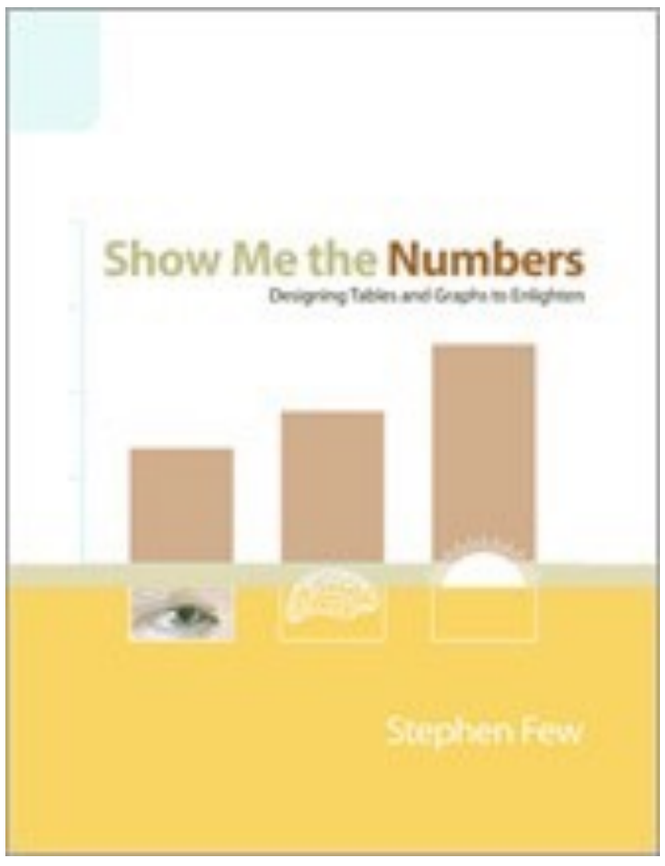
EDWARD TUFTE

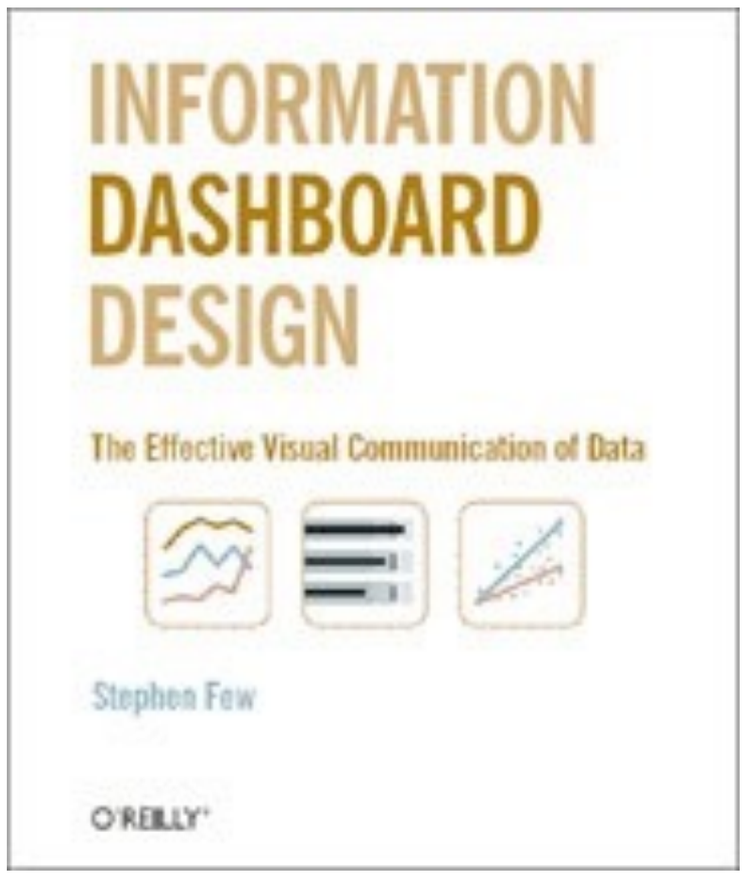






STEPHEN FEW


Show Me the Numbers
Designing Tables and Graphs to Enlighten
Stephen Few


INFORMATION
DASHBOARD
DESIGN
The Effective Visual Communication of Data
Stephen Few
O'REILLY


Now you see it
Stephen Few

Thank you for your attention

QUESTIONS?