Charts

DATA 201: Thinking With Data
Winter 2022

Jonathan Hudson, Ph.D.
Instructor
Department of Computer Science
University of Calgary

Tuesday, March 22, 2022

UNIVERSITY OF CALGARY

What and Why?

Tables
(are not charts)

Tables

A type of visualization

Mostly values

Colours and formatting

Most accurate way to indicate and compare data

Don’t scale well, and can be hard to draw conclusions from

	A	B	C	D	E	F	G	H	I	J	K	L
1	0	0	0.6931	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0.6931	0	1.3459	0	0	0	0	0	0	0.6931	0	0
4	0.6931	0	0	0	0	0	0	0	0	0	0	0
5	1.0986	0	0	0	1.0986	0	0	0	0	0	0	0.6931
6	0	1.3459	1.7918	0	0.6931	0	0.6931	0	0	0	0	0
7	1.0986	2.1972	2.0794	0	2.4849	1.0986	1.0986	1.0986	0	0.6931	0	0
8	3.5553	4.0254	4.7875	2.0794	3.1781	2.0794	2.4849	2.9444	0	1.3863	1.0986	1.0986
9	4.7362	5.0562	5.1818	2.4849	3.4965	2.7081	3.0445	3.912	1.6094	2.5649	2.4849	0
10	5.7366	5.2832	5.9428	2.5649	4.1431	3.989	3.8712	3.5553	2.4849	3.2958	1.3863	1.0986
11	5.7004	5.0173	6.2126	2.3979	4.4773	2.5649	3.8067	2.7726	2.8332	1.6094	1.0986	0.6931
12	4.1589	2.6391	4.2047	0.6931	2.3026	1.7918	2.1972	0	0.6931	0.6931	0	0
13	0	0.6931	2.0794	0	0	0.6931	1.0986	0	0	0	0	0
14	0	0	1.0986	0	0	0	0	0	0	0	0	0
15	0.6931	0	0.6931	0	0	0	0	0	0	0	0	0
16	0.6931	0	0	0	0	0	0	0	0	0	0	0
17	1.0986	0.6931	1.3863	0	0	0.6931	0	0	0	0.6931	0	0
18	0	1.3863	1.0986	1.6094	0	0.6931	0.6931	0	0	1.3863	0	0.6931
19	0	1.0986	1.3863	1.0986	0	0	0	0	0	0	0	0.6931
20	2.0794	0	1.6094	0	0.6931	0	0	0	0	0.6931	0	0
21	2.1972	1.7918	2.3979	0	0	0	1.0986	1.3863	1.3863	0.6931	0	0.6931

Tables

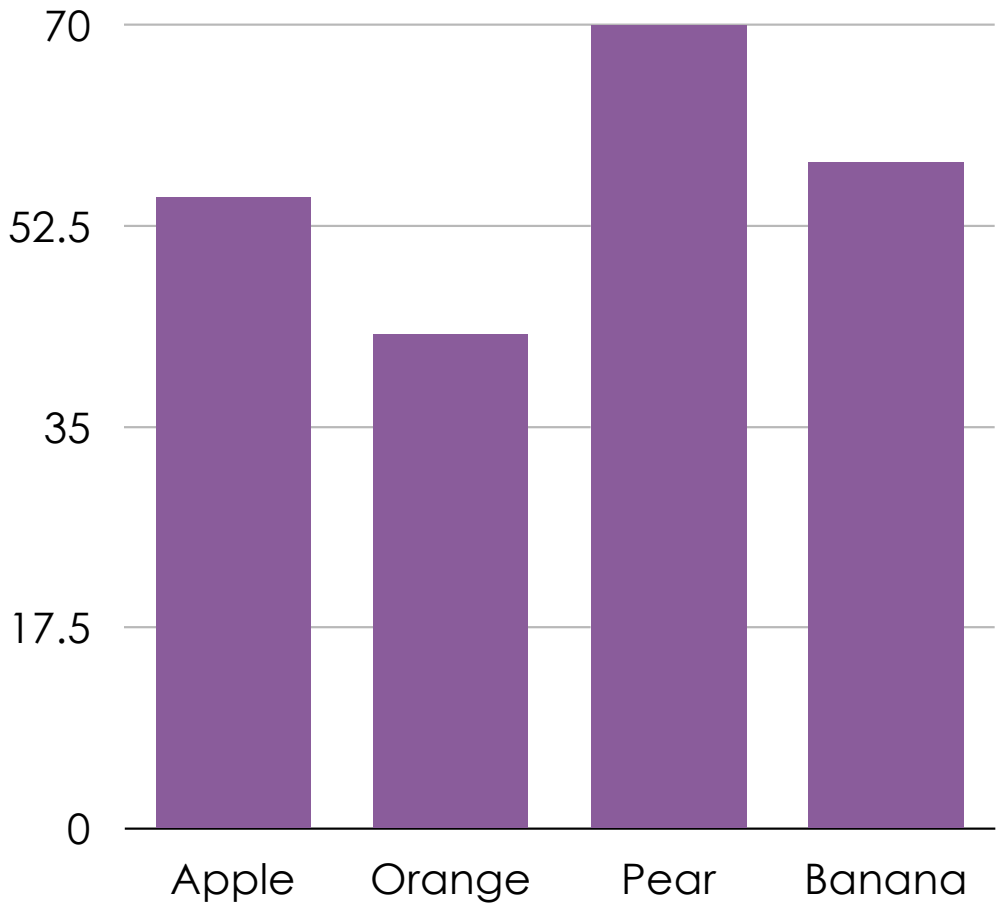
If numbers were all that matters it would be all we communicate in papers.

Good scientific papers have both, a table and a visualization like a chart.

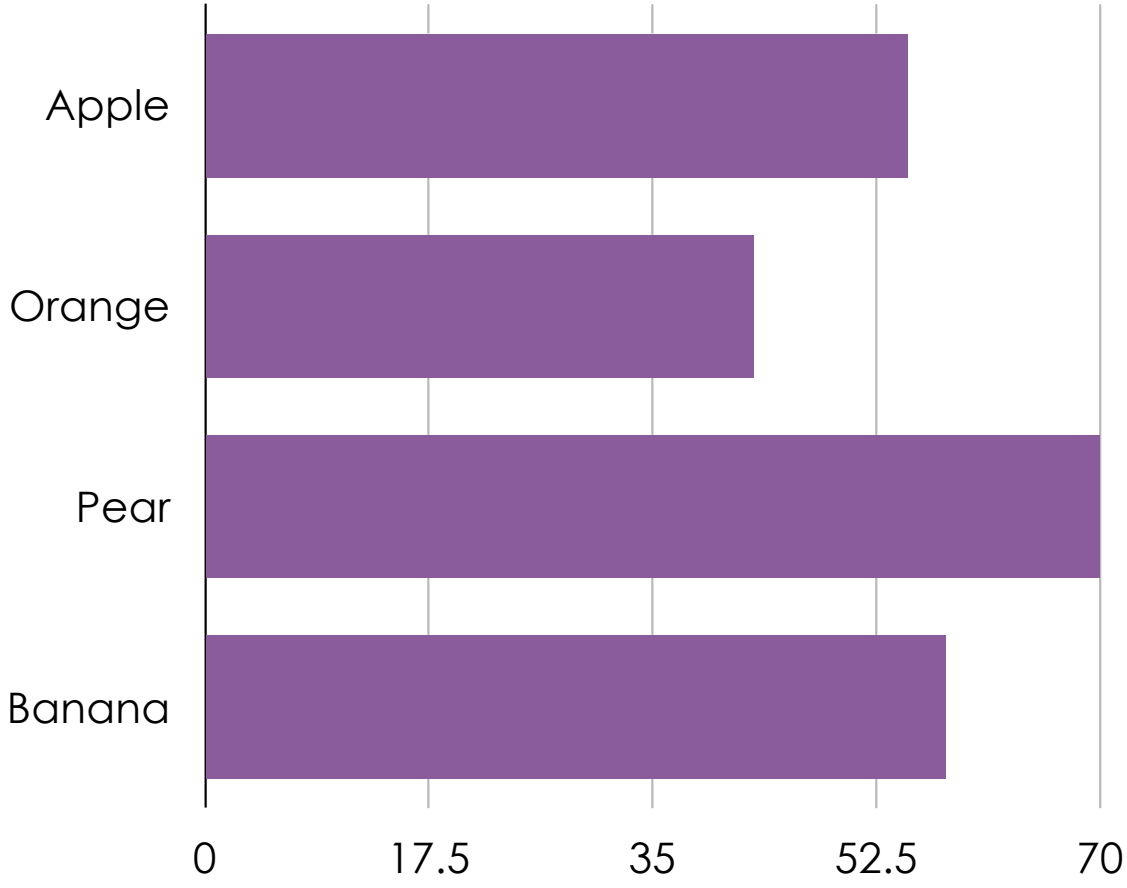
	A	B	C	D	E	F	G	H	I	J	K	L
1	0	0	0.6931	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0.6931	0	1.3459	0	0	0	0	0	0	0.6931	0	0
4	0.6931	0	0	0	0	0	0	0	0	0	0	0
5	1.0986	0	0	0	1.0986	0	0	0	0	0	0	0.6931
6	0	1.3459	1.7918	0	0.6931	0	0.6931	0	0	0	0	0
7	1.0986	2.1972	2.0794	0	2.4849	1.0986	1.0986	1.0986	0	0.6931	0	0
8	3.5553	4.0254	4.7875	2.0794	3.1781	2.0794	2.4849	2.9444	0	1.3863	1.0986	1.0986
9	4.7362	5.0562	5.1818	2.4849	3.4965	2.7081	3.0445	3.912	1.6094	2.5649	2.4849	0
10	5.7366	5.2832	5.9428	2.5649	4.1431	3.989	3.8712	3.5553	2.4849	3.2958	1.3863	1.0986
11	5.7004	5.0173	6.2126	2.3979	4.4773	2.5649	3.8067	2.7726	2.8332	1.6094	1.0986	0.6931
12	4.1589	2.6391	4.2047	0.6931	2.3026	1.7918	2.1972	0	0.6931	0.6931	0	0
13	0	0.6931	2.0794	0	0	0.6931	1.0986	0	0	0	0	0
14	0	0	1.0986	0	0	0	0	0	0	0	0	0
15	0.6931	0	0.6931	0	0	0	0	0	0	0	0	0
16	0.6931	0	0	0	0	0	0	0	0	0	0	0
17	1.0986	0.6931	1.3863	0	0	0.6931	0	0	0	0.6931	0	0
18	0	1.3863	1.0986	1.6094	0	0.6931	0.6931	0	0	1.3863	0	0.6931
19	0	1.0986	1.3863	1.0986	0	0	0	0	0	0	0	0.6931
20	2.0794	0	1.6094	0	0.6931	0	0	0	0	0.6931	0	0
21	2.1972	1.7918	2.3979	0	0	0	1.0986	1.3863	1.3863	0.6931	0	0.6931

Bar/Column Charts

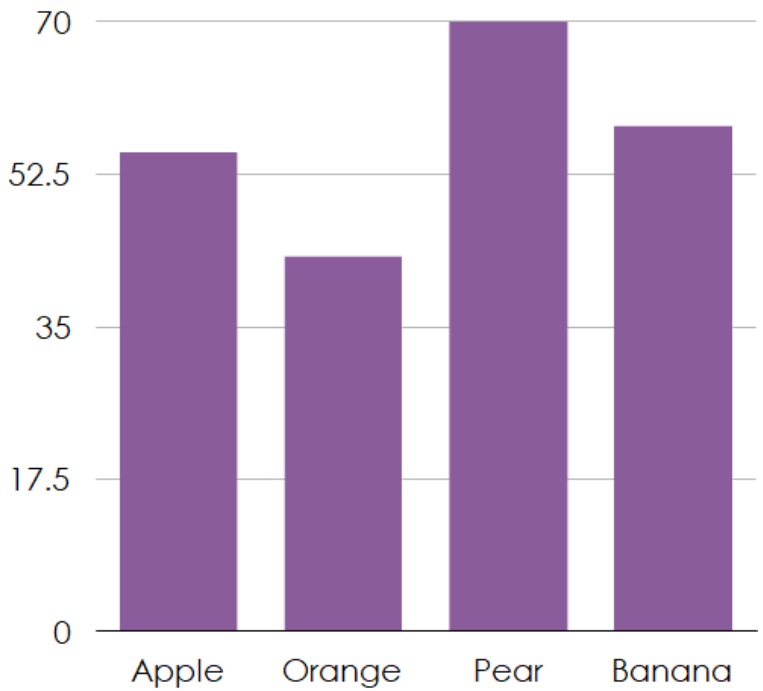
Column Chart / Vertical Bar Chart



Bar Chart / Horizontal Bar Chart



UNIVERSITY OF CALGARY

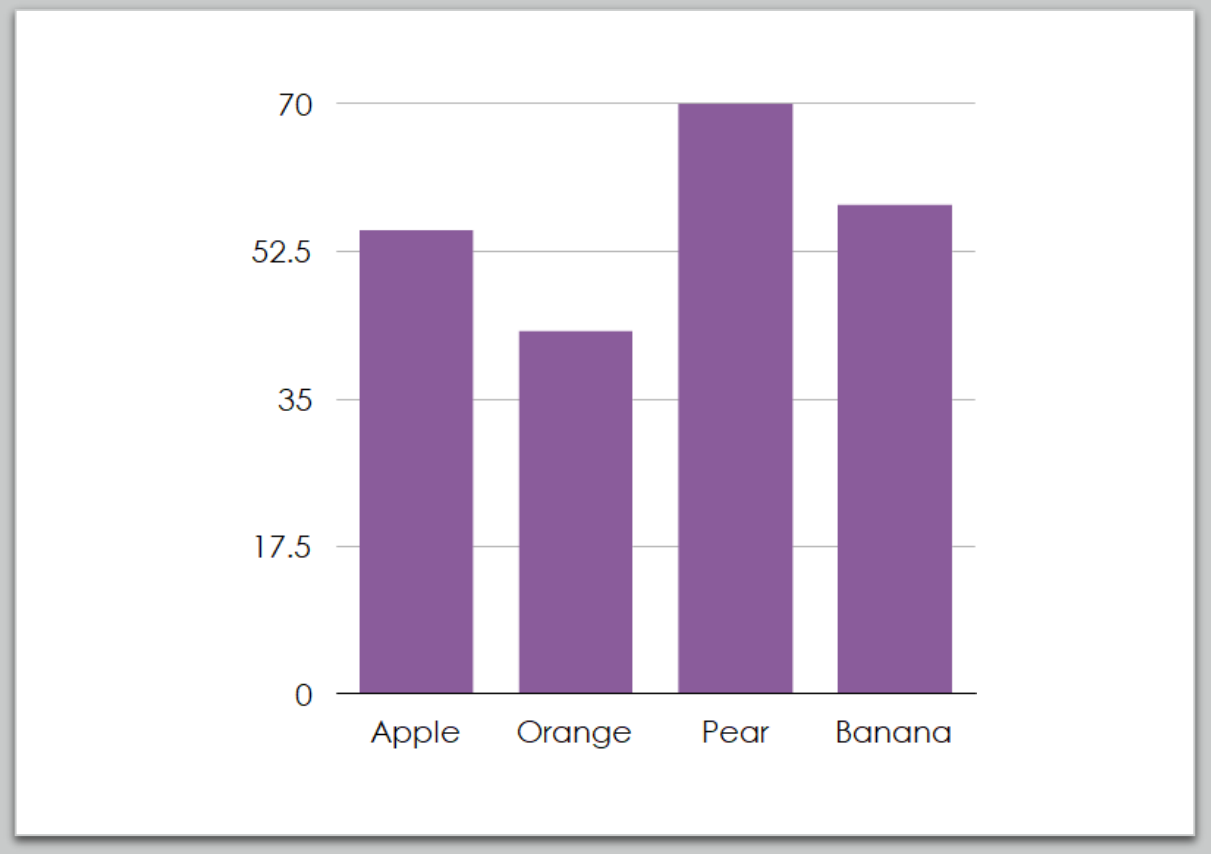


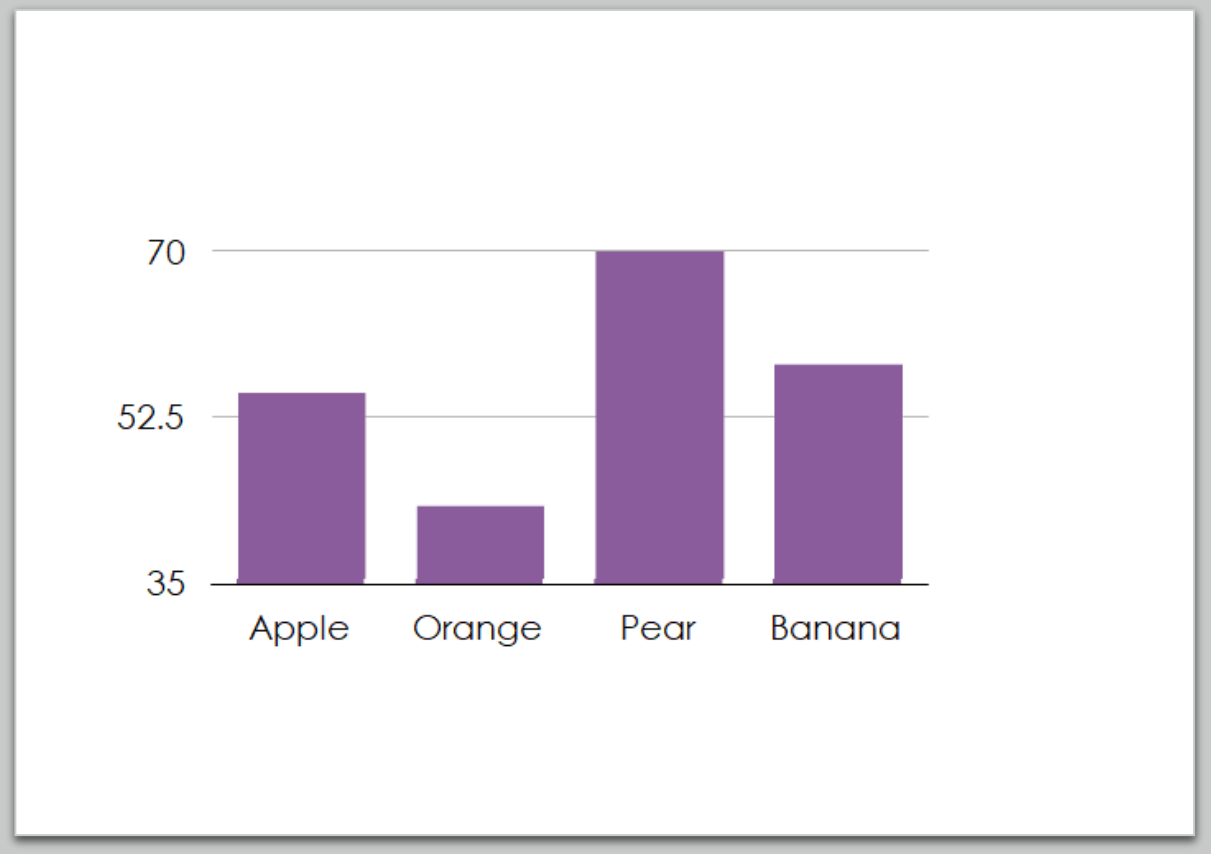
Column Charts

- Compare across categories
- Ex. What percentage of people like each type of fruit?
- Y is response (or count)
- X is category
- Column charts are best in terms of Y as a response variable
- Bar charts work well when labels are long or data is more of a natural horizontal idea (length!)

Column Charts

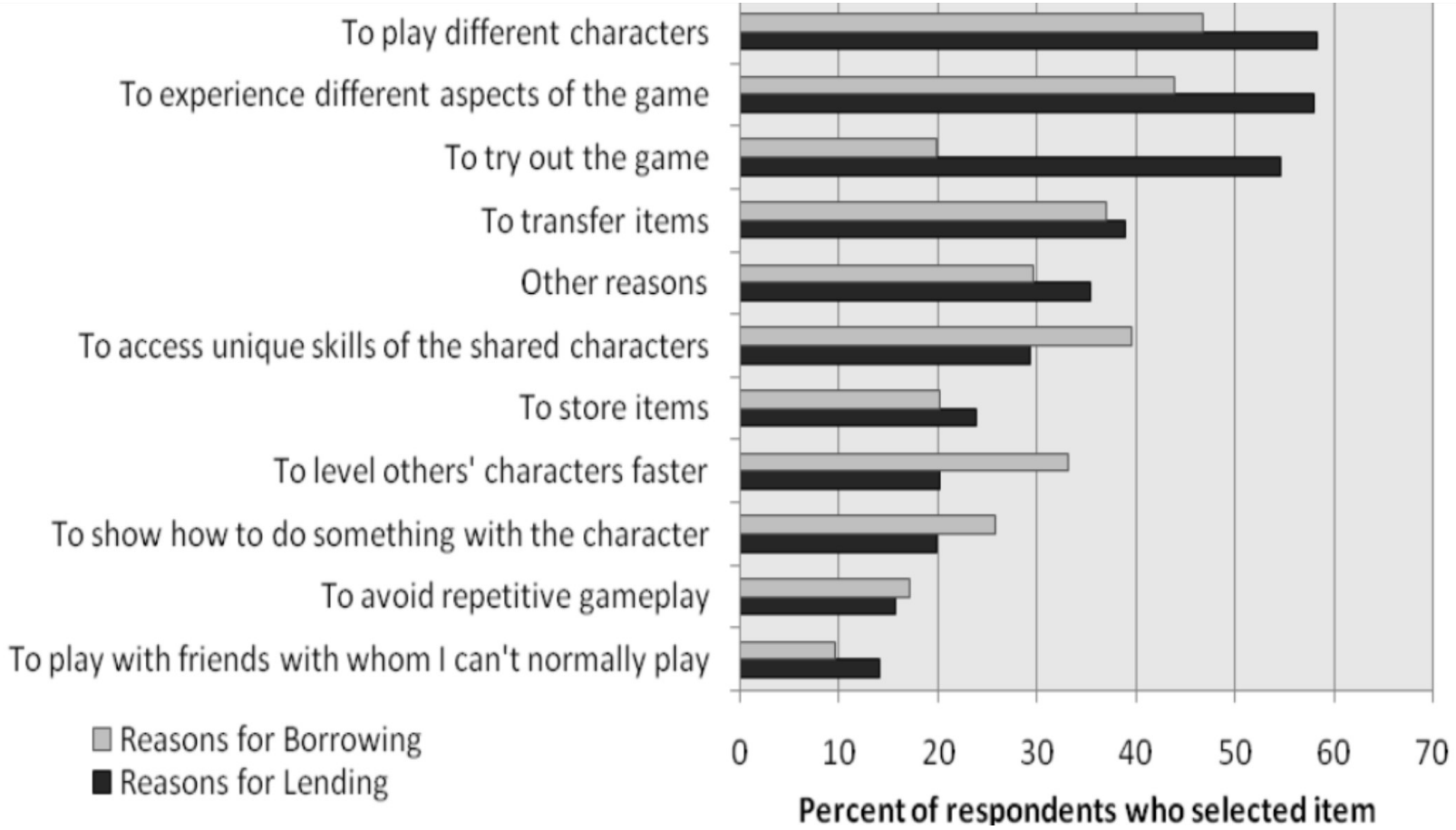
- Start your Y axis at 0!
- Alternate Y axis origins can be acceptable but when they are used it is because those reading chart understand context of non-zero origin.
- Many non-zero origin choices are deceptive.





Wong et al., 2009 (Multi-set Bar Chart)
Example of primary/secondary categories

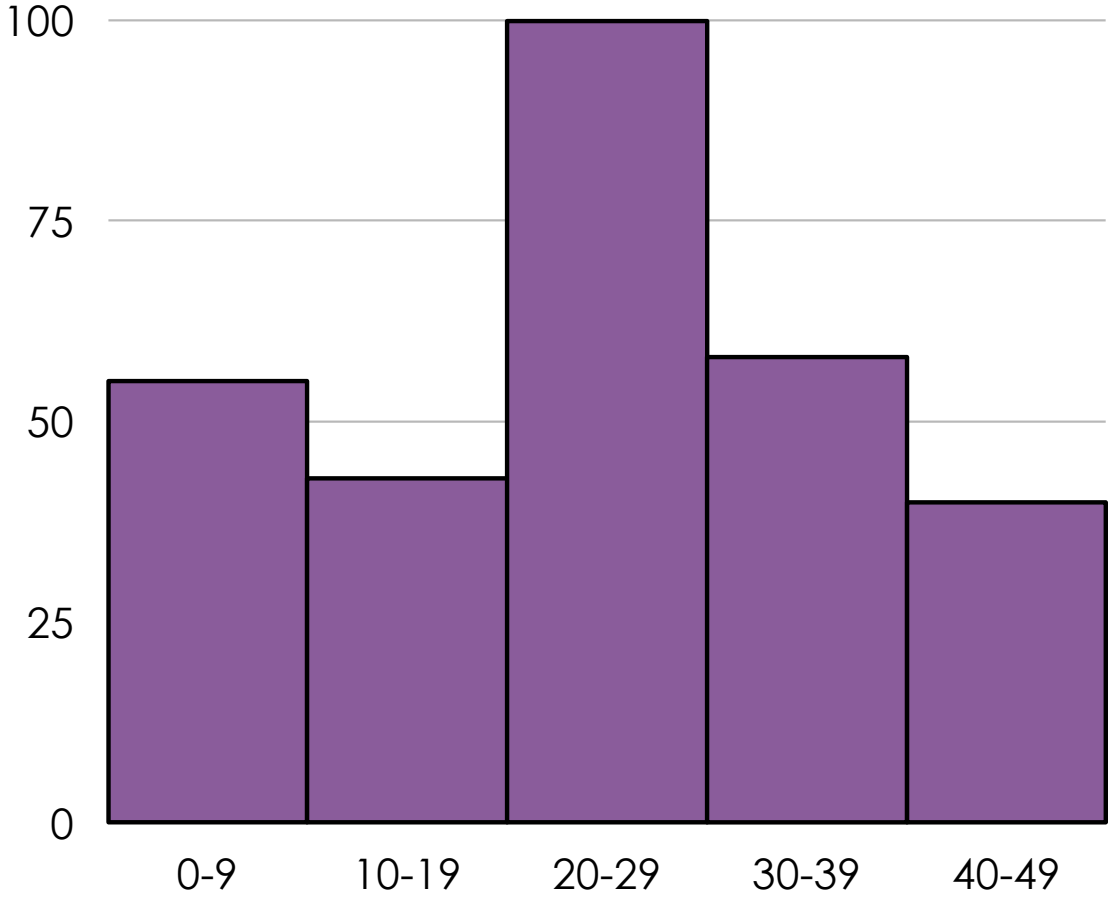
- Colour of bar (or texture, etc.) indicates another category for comparison

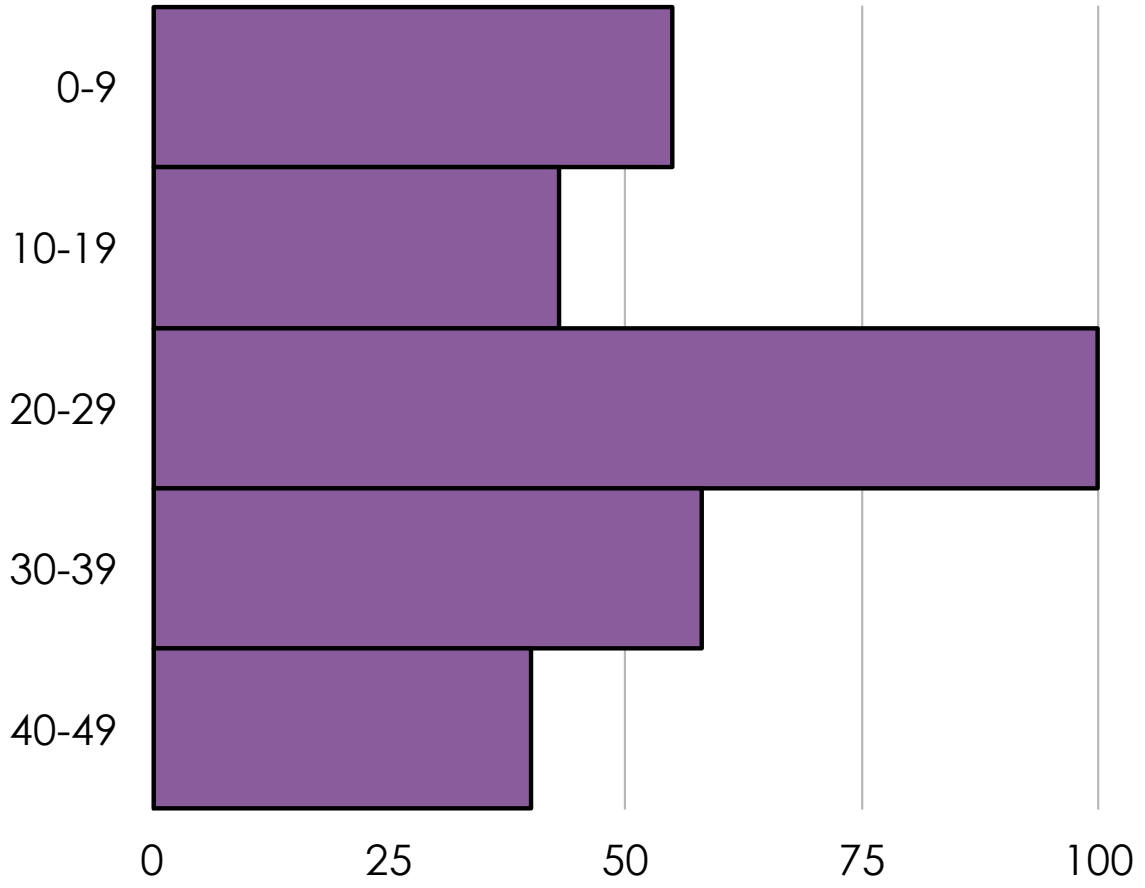


UNIVERSITY OF CALGARY

Histograms

Histograms

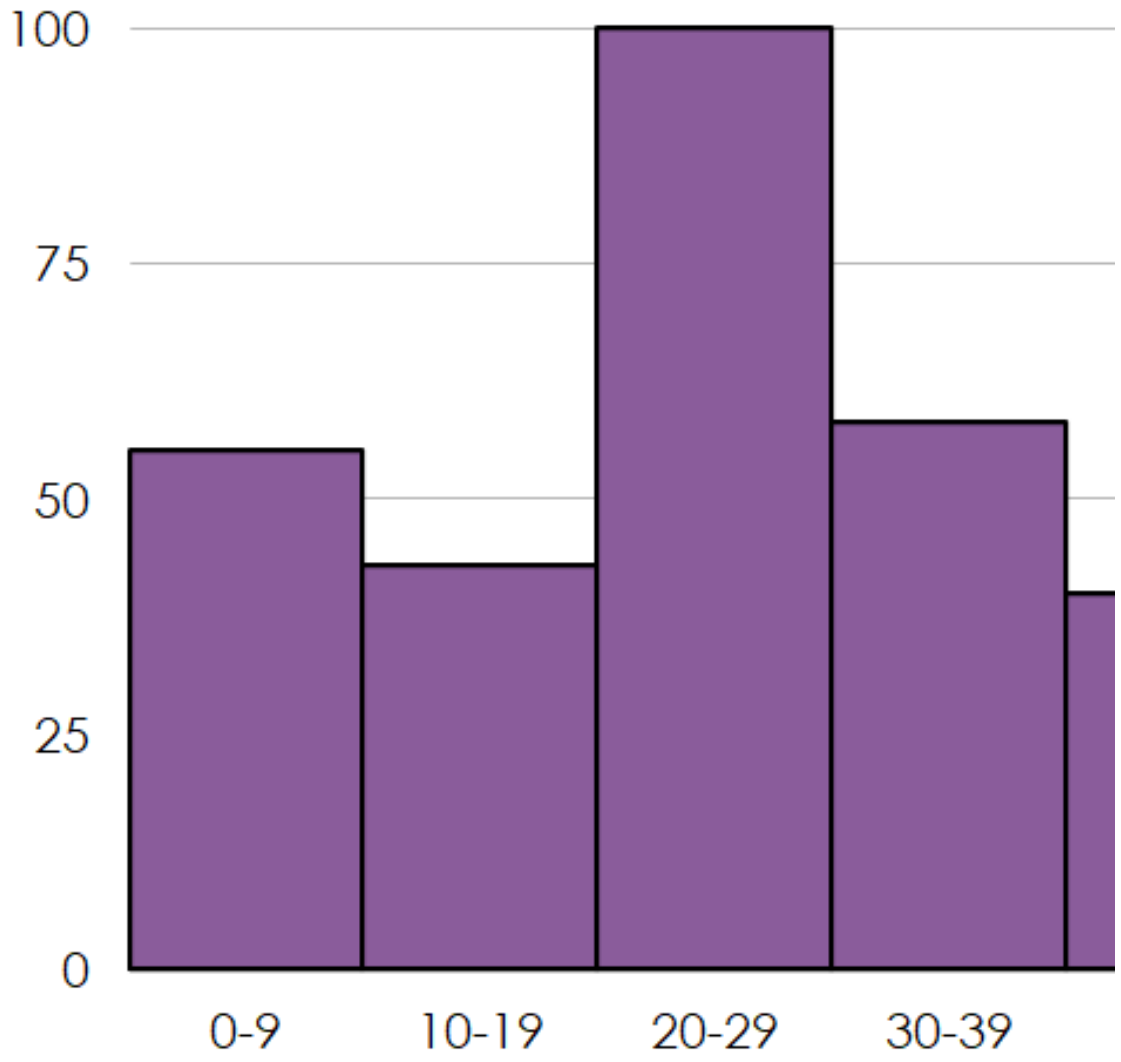




UNIVERSITY OF CALGARY

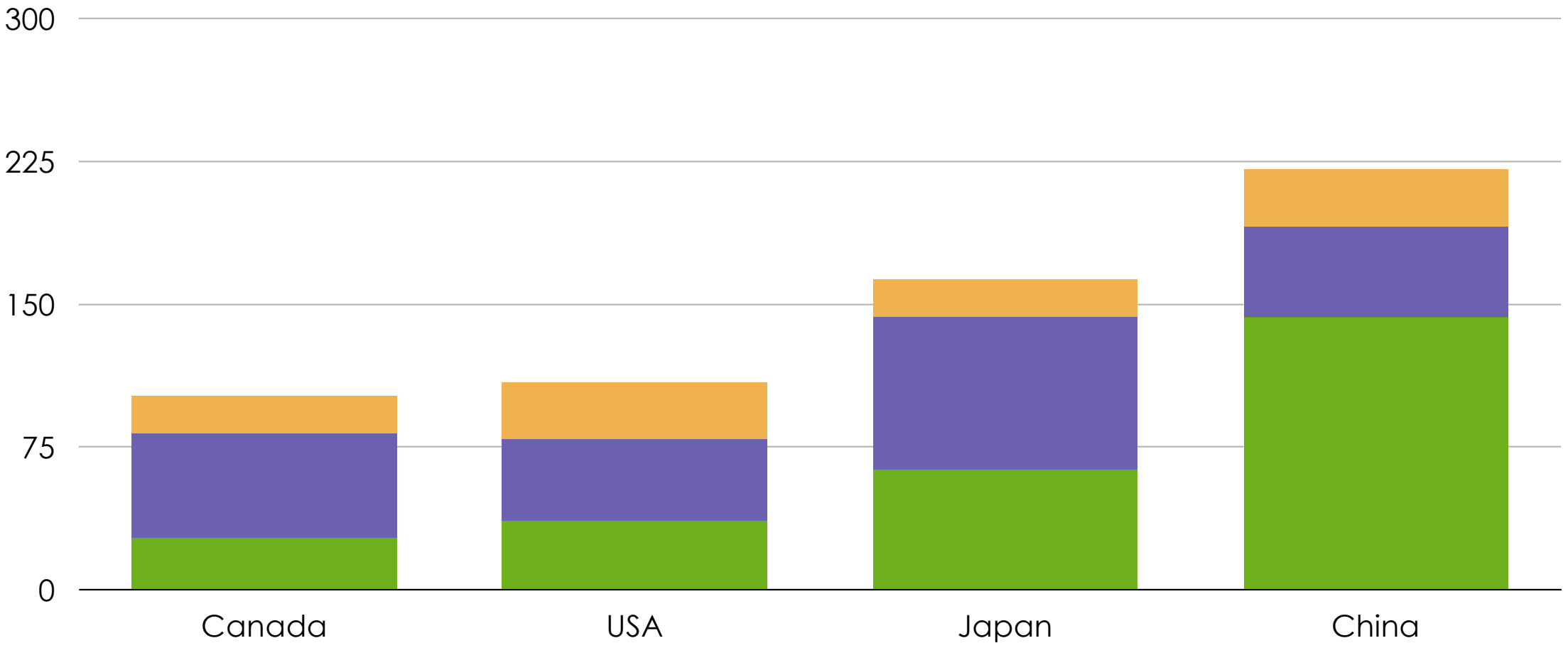
Histograms

- Distribution based data
- Ex. Student grades, data that might have categories or groupings
- Each x in X is a bucket (a smaller range portion of the total range)
- Y is response (count in category)
- Like bar chart but categories are related as being ranges which are part of a complete range
- Range choice can be used to manipulate data's appearance.



Stacked Charts / Histograms

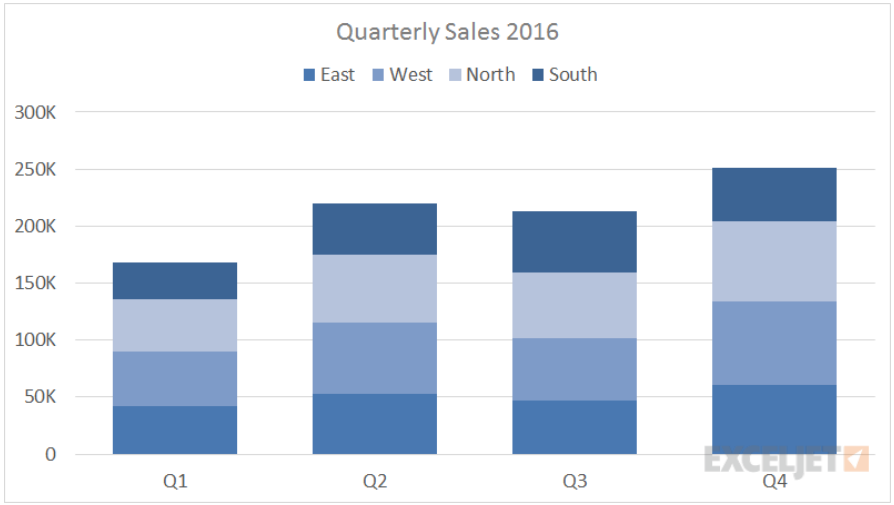
Stacked Charts



UNIVERSITY OF CALGARY

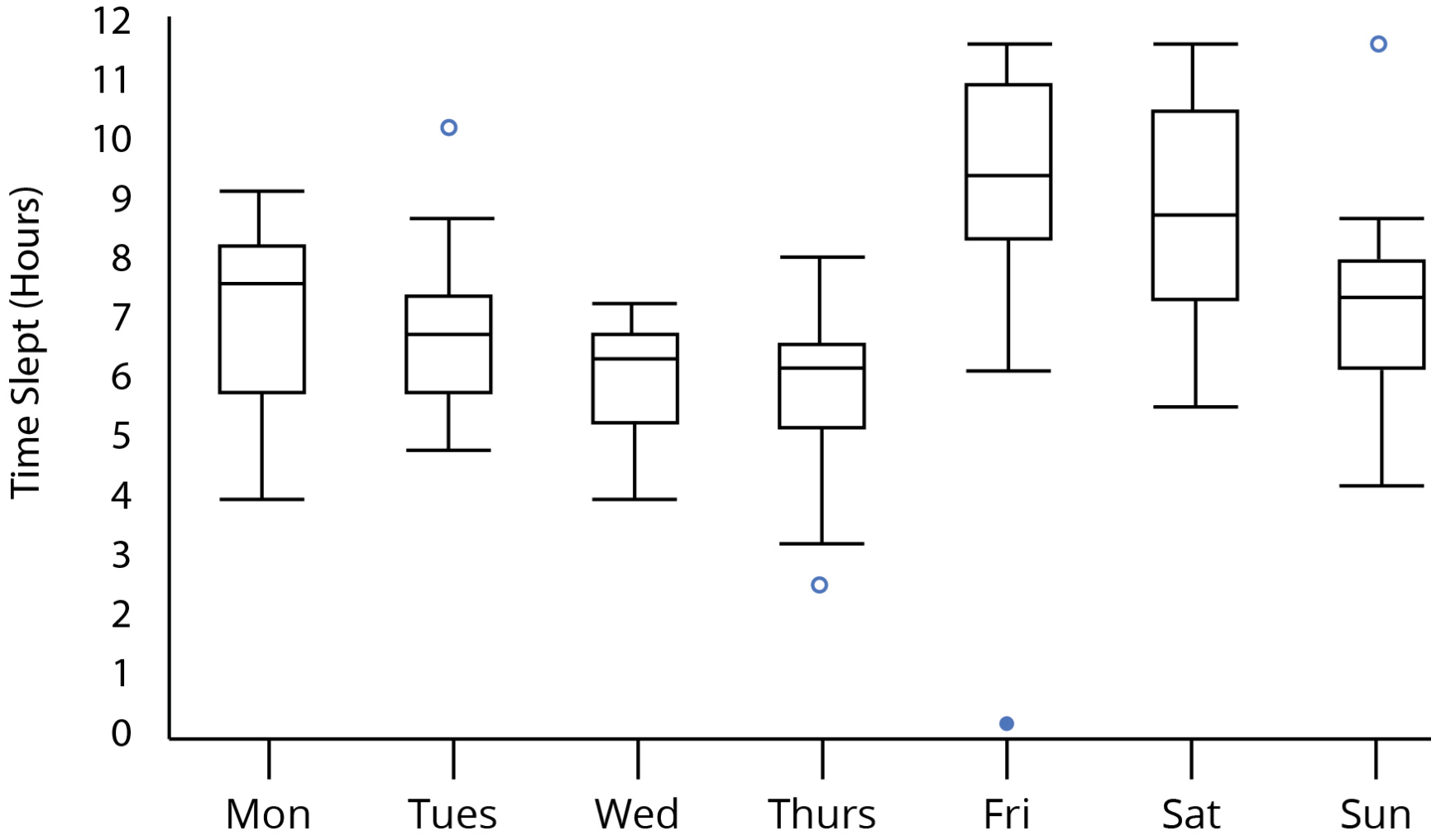
Stacked Charts

- Variant of bar type charts
- Part to whole comparisons
- Primary and secondary categories, but the secondary categories are all part of the whole for each primary category
- Ex. Sales per quarter for company, but also divided by area of country within that quarter.
- Note that secondary categories become hard to numerically compare.



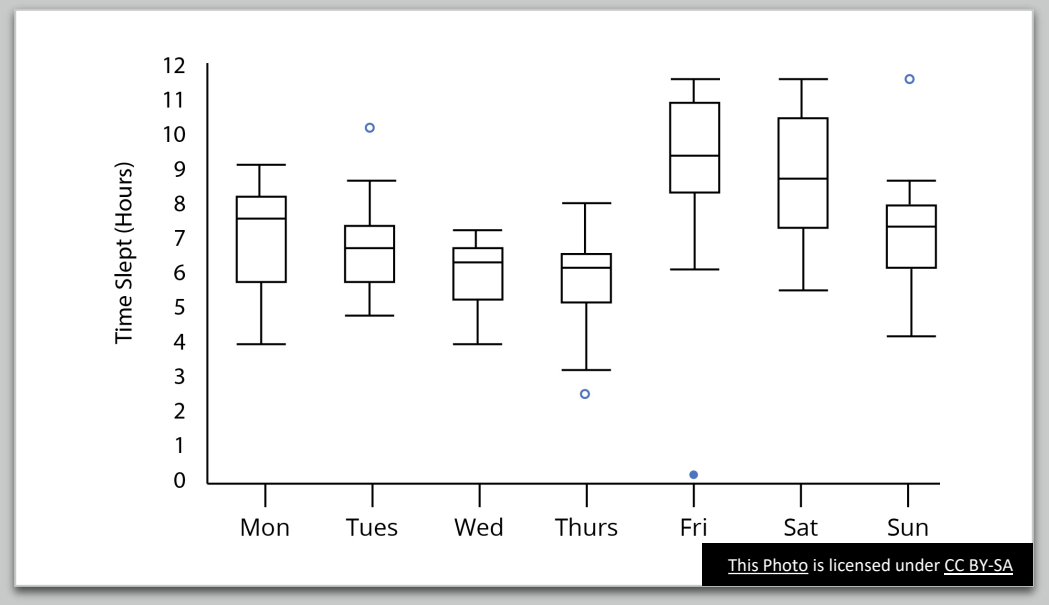
Box-And-Whisper Charts

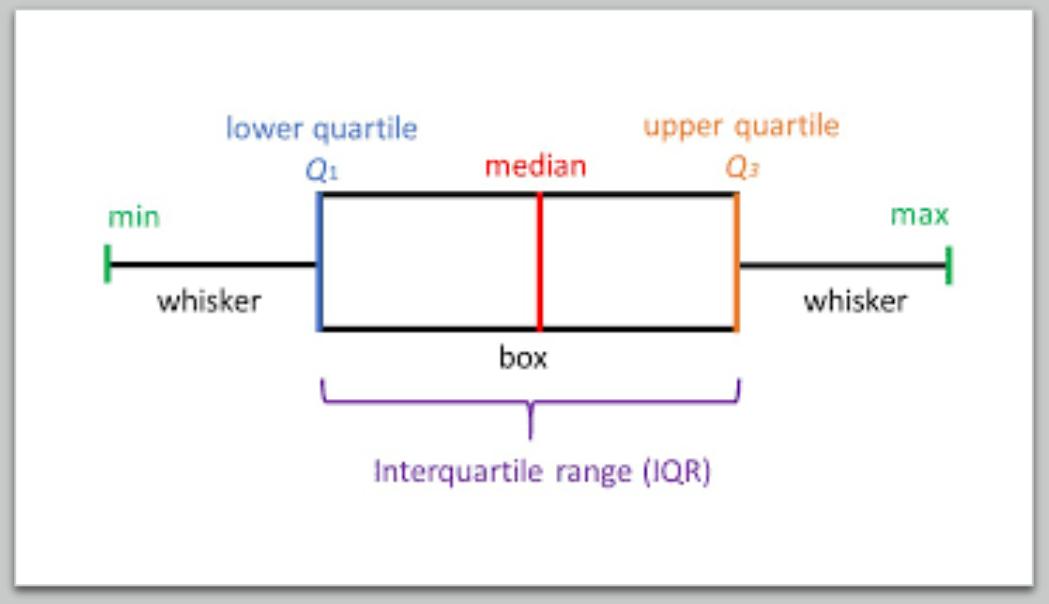
Box-And-Whisper



Box-And-Whisper

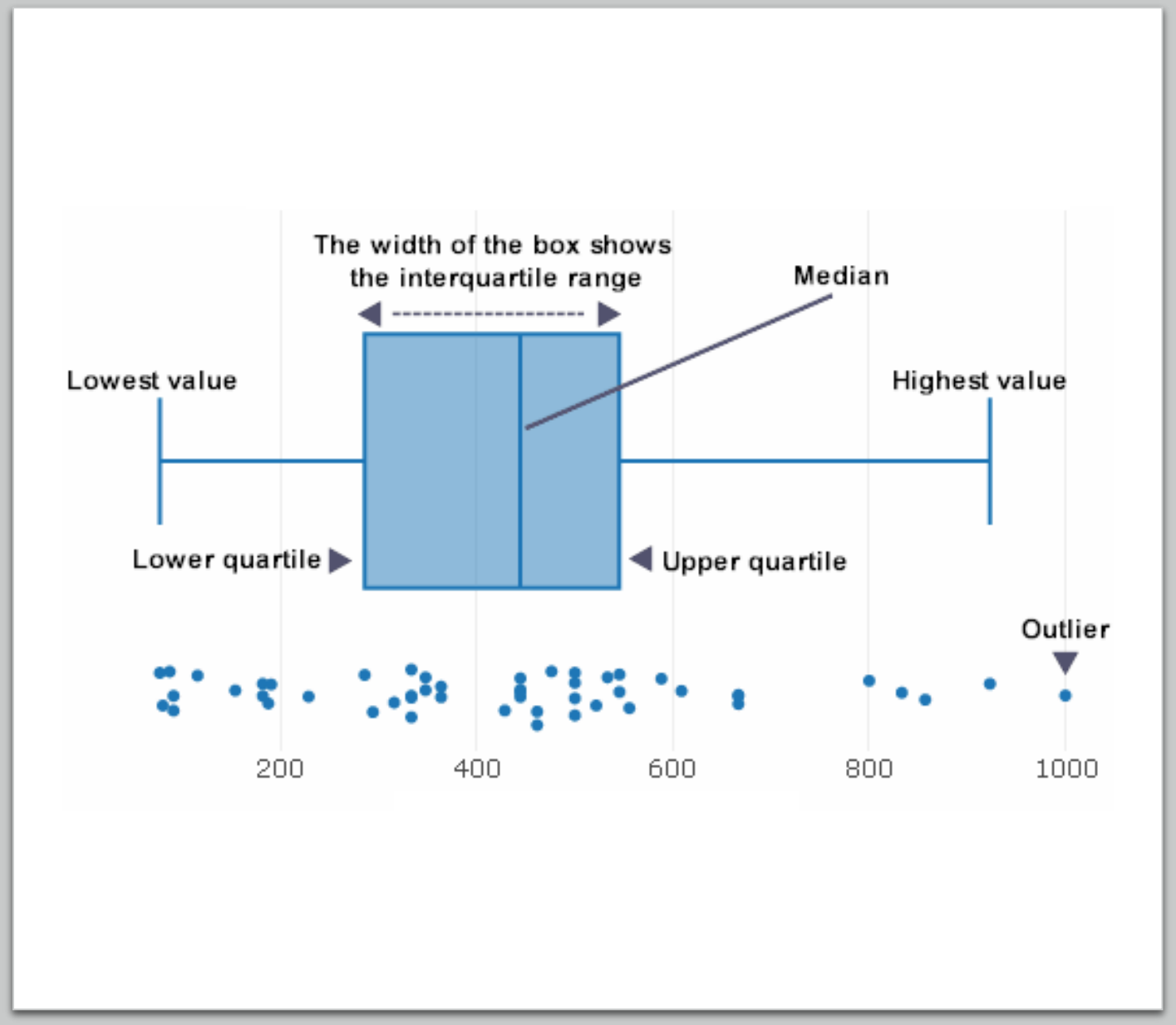
- Center line (sometimes an X) is the median of the collection of values in the category
- The top/bottom of box are 25% and 75% quartile values.
- The whiskers (top and bottom thin pieces) are the minimum and maximum values
- Dots are outliers that fall a certain extent away from the mean (technically they are the true mimima and maxima)
- Common in sciences to convey multiple data points in more statistical manner





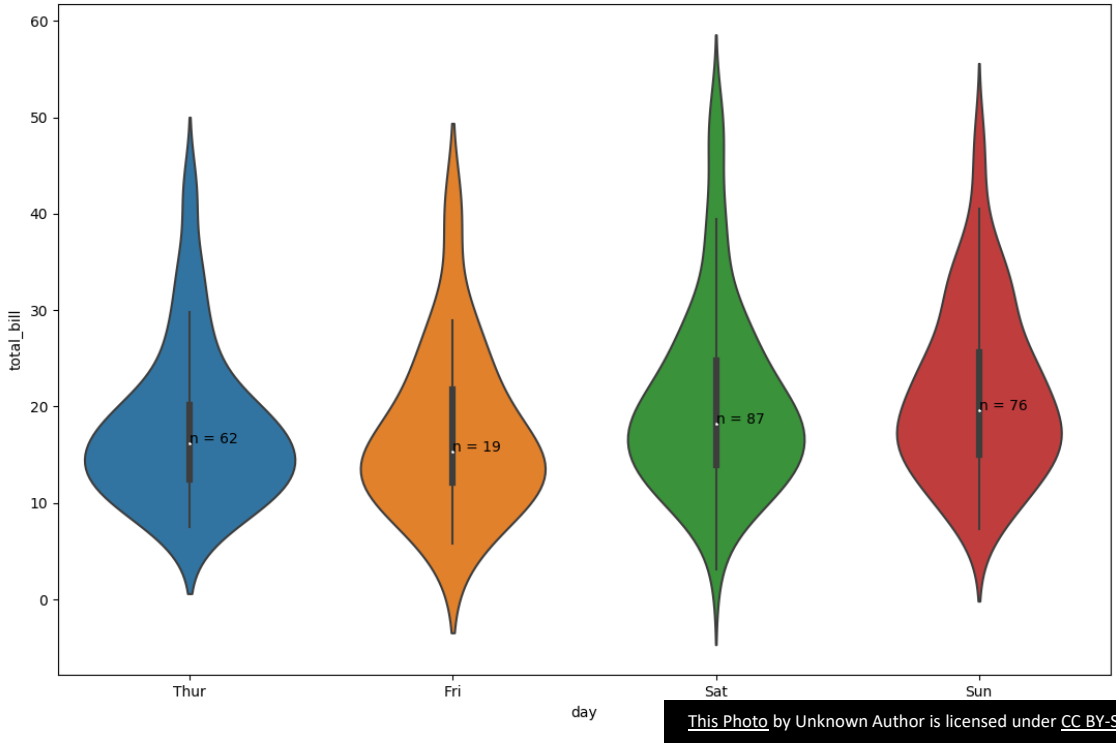
Box-And-Whisper and Bee Swarm Chart

- Same data can be shown as a ‘bee swarm’
- A plot of points but ‘jittered’ (randomly distributed when they overlap)
- Gives more exact data but is often not detail that is needed



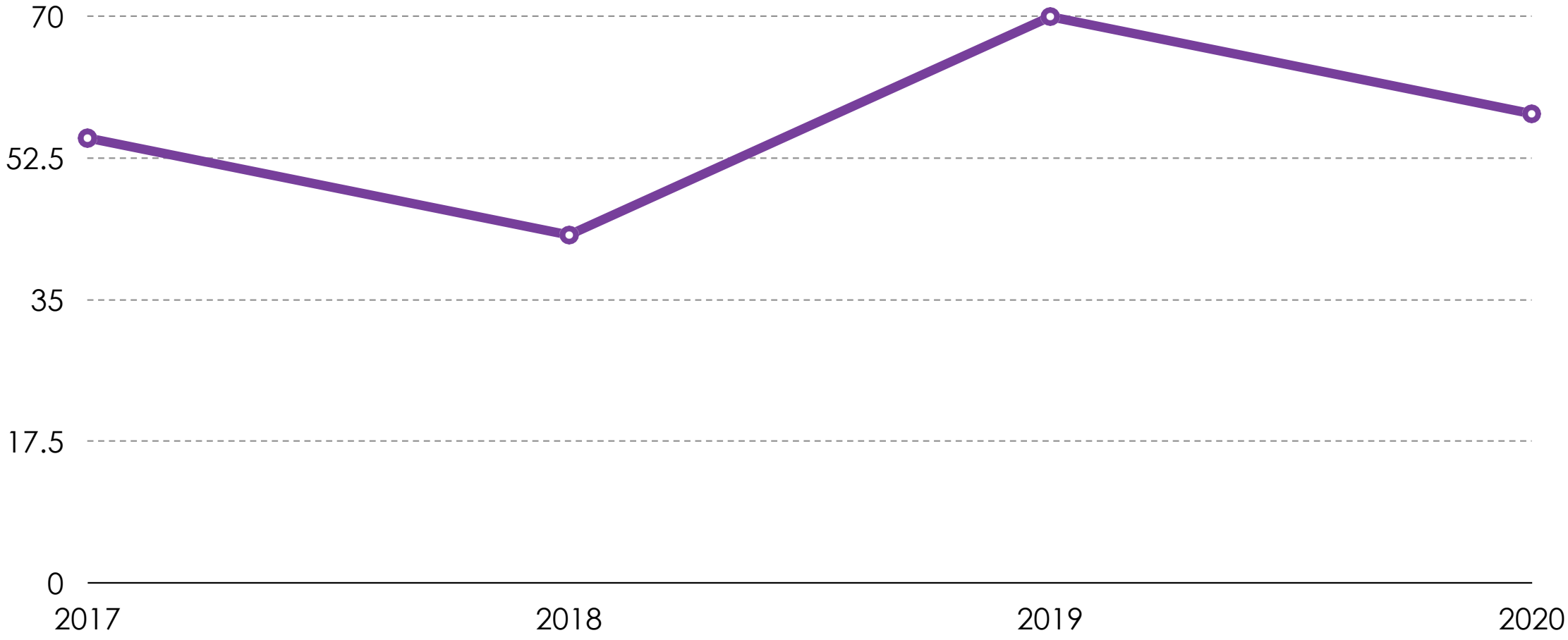
Box-And-Whisper and Violin Chart

- Same data can be shown as a ‘violin’
- A variant of ‘bee swarm’ where the points aren’t ‘jittered’ but instead of a box and whisper box, a variable width shape is used.
- Some degree of smoothing of data is done to width change to give smooth shape.



Line Charts

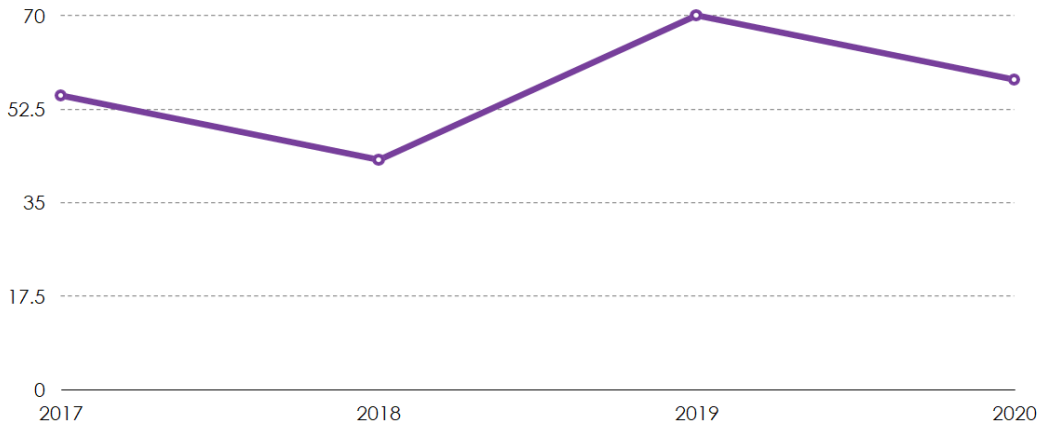
Line Charts



UNIVERSITY OF CALGARY

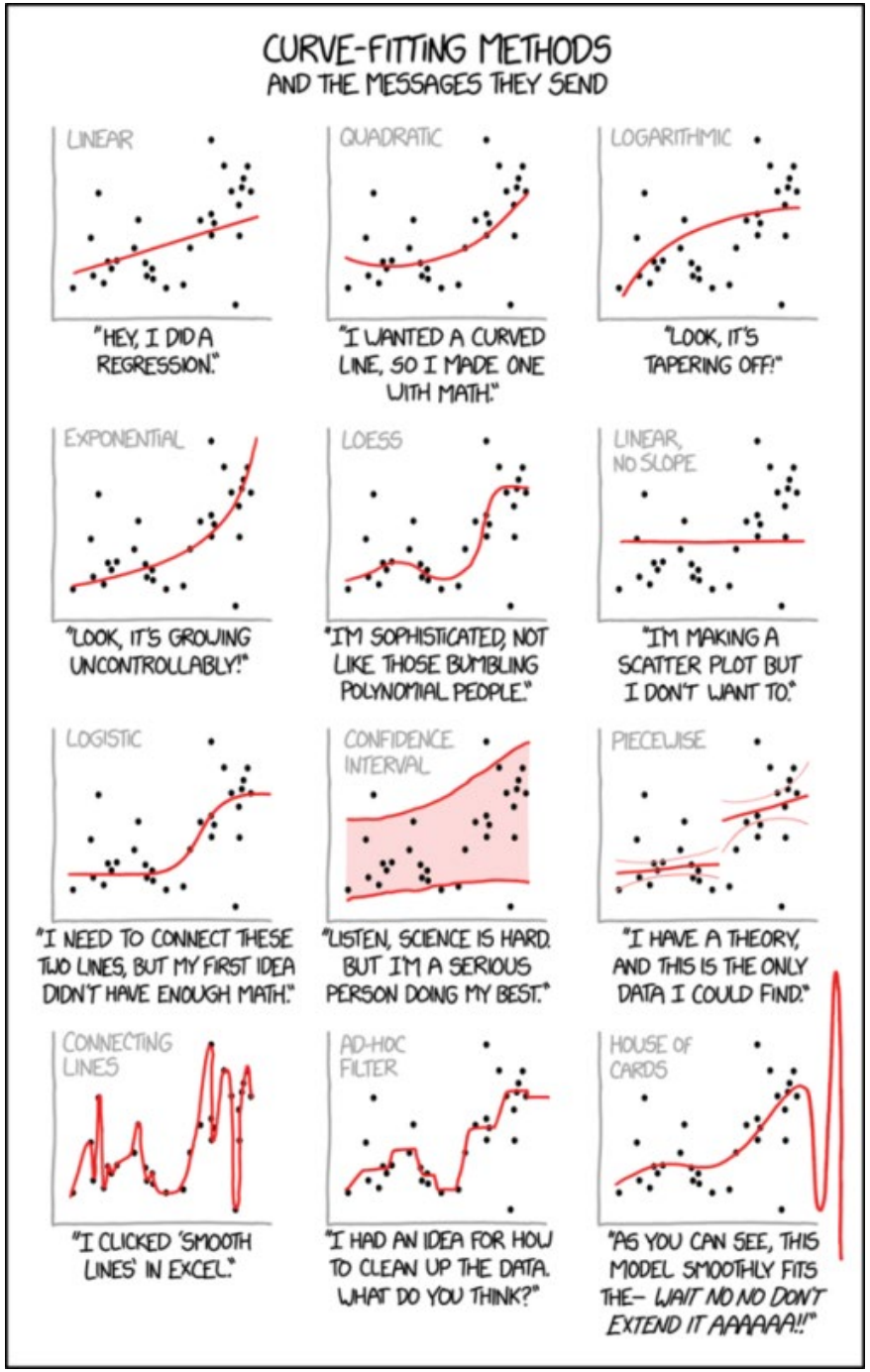
Line Charts

- Data changing over time
- Y (up-down is response)
- X is interval (it is clearest if this is consistent scale)
- Predict trends



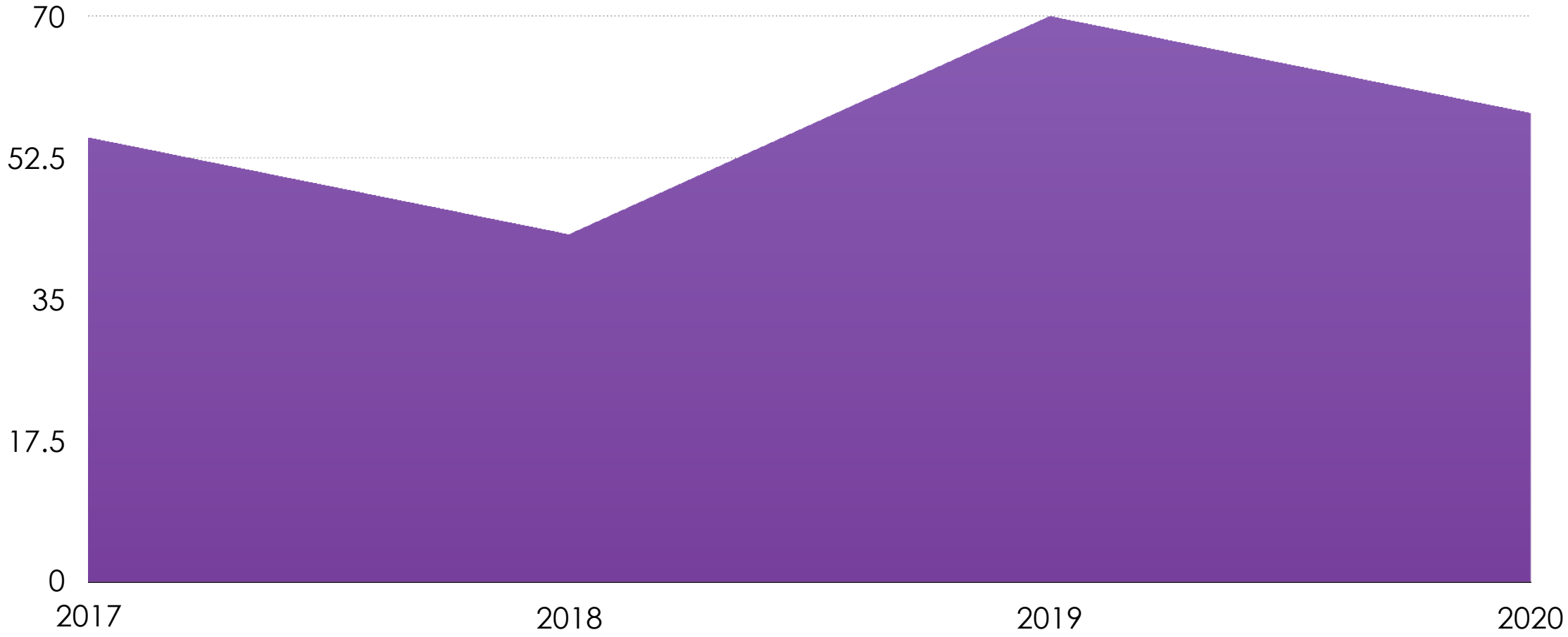
Line Charts

- Data changing over time
- Y (up-down is response)
- X is interval (it is clearest if this is consistent scale)
- Imply trends?



Area Charts

Area Charts



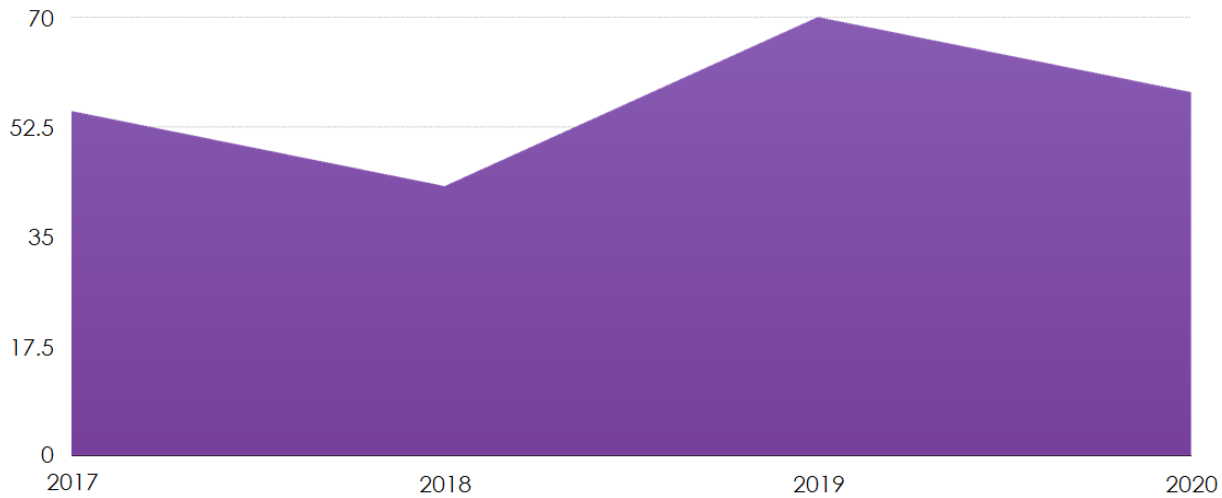
UNIVERSITY OF CALGARY

Area Charts

No immediate benefit over a line chart, but when only 1 line gives more emphasis to quantity

Also works better when 0 is meaningful quantity relative to Y data

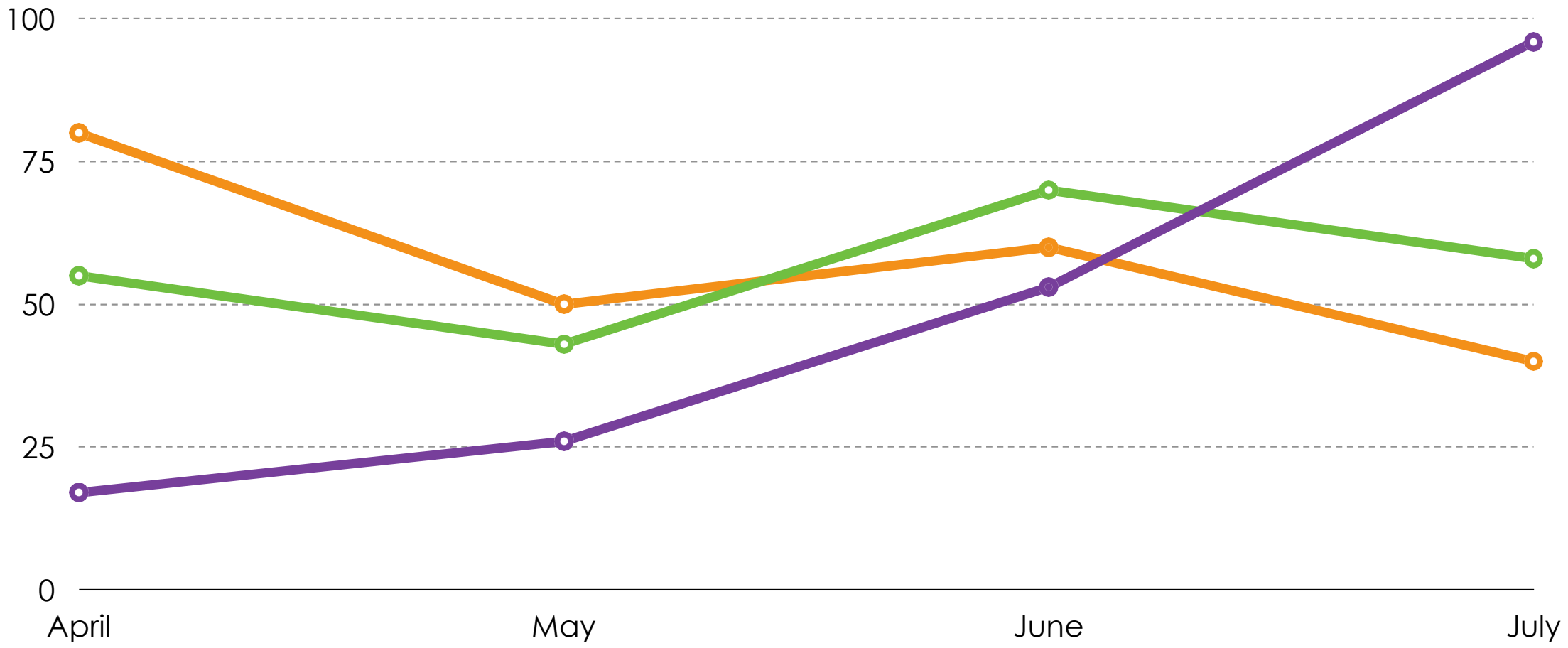
Most useful when combined with stacking



UNIVERSITY OF CALGARY

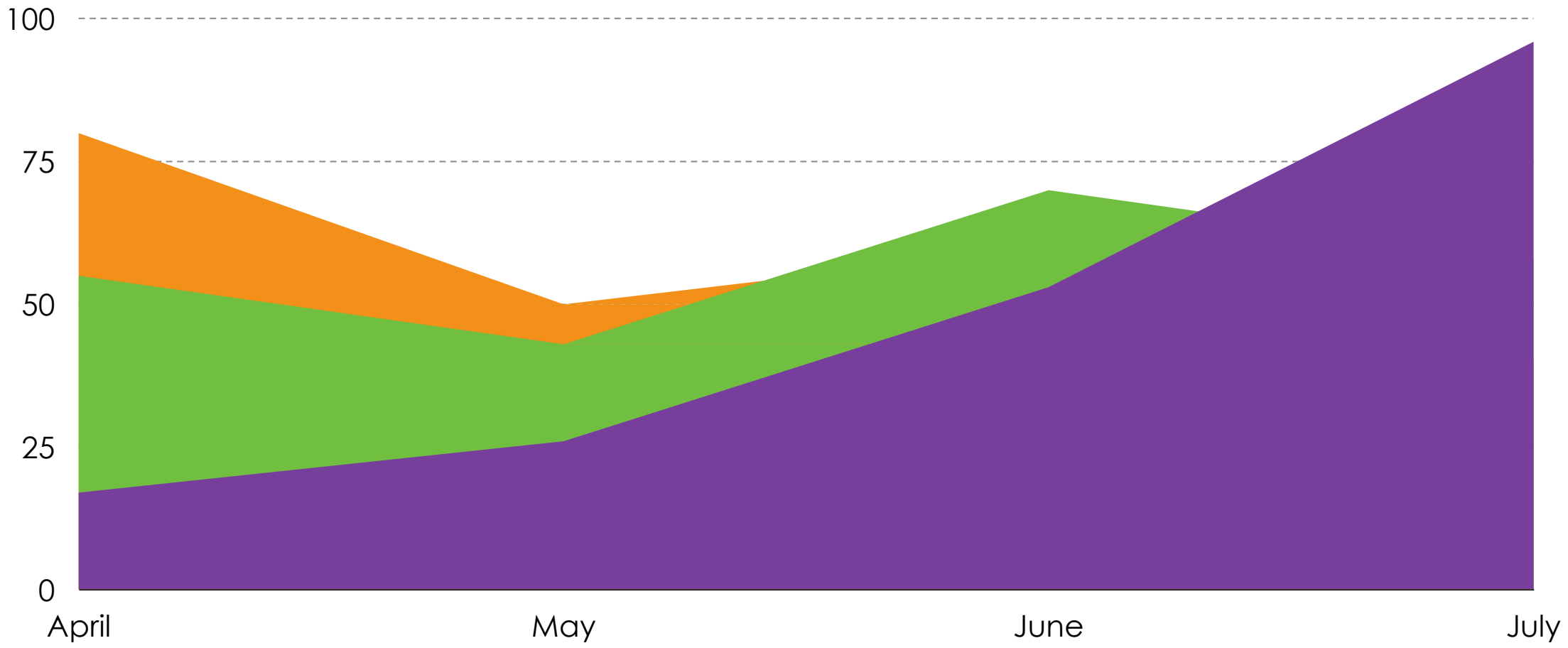
Line vs. Area

Line Chart



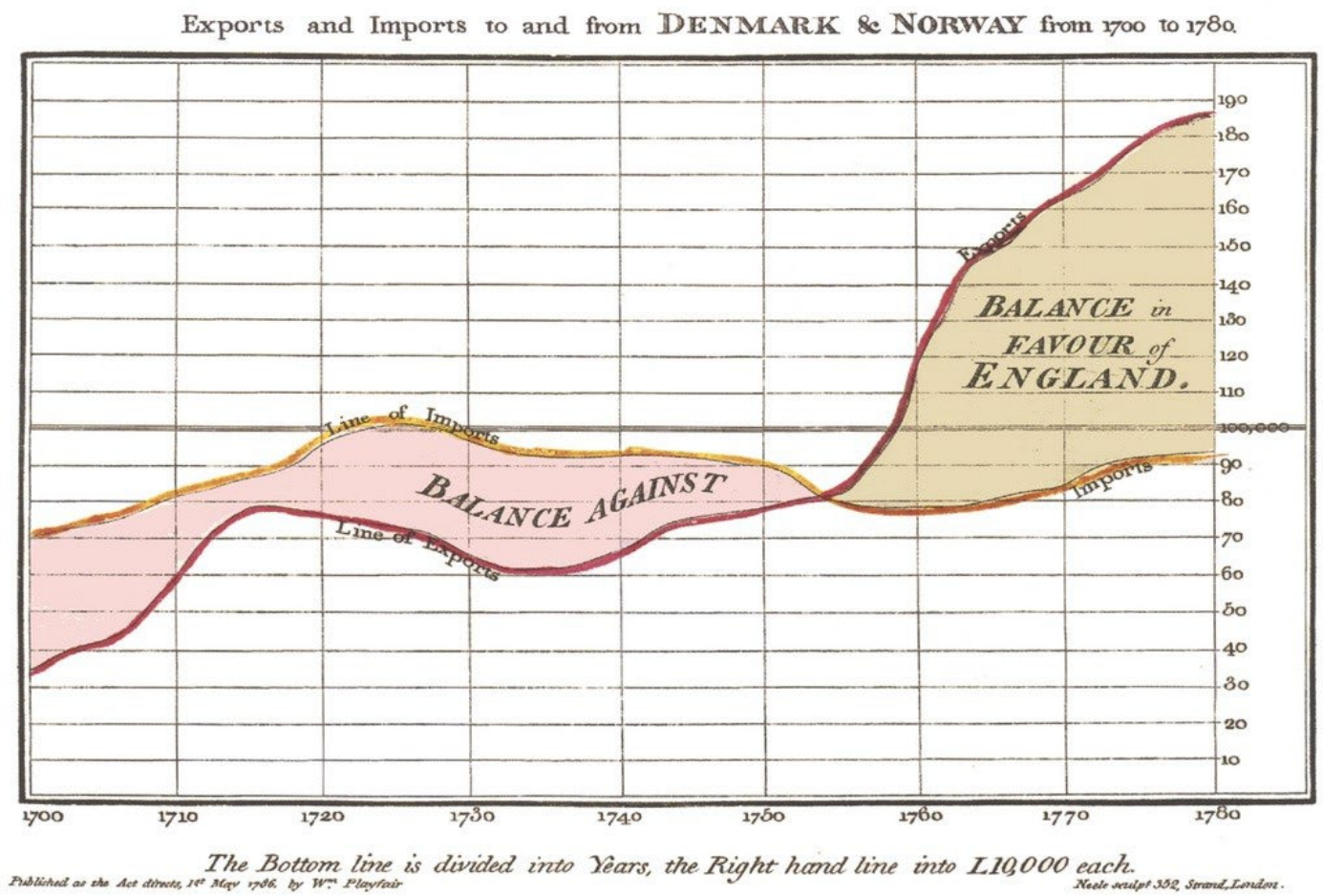
UNIVERSITY OF CALGARY

Area Chart



UNIVERSITY OF CALGARY

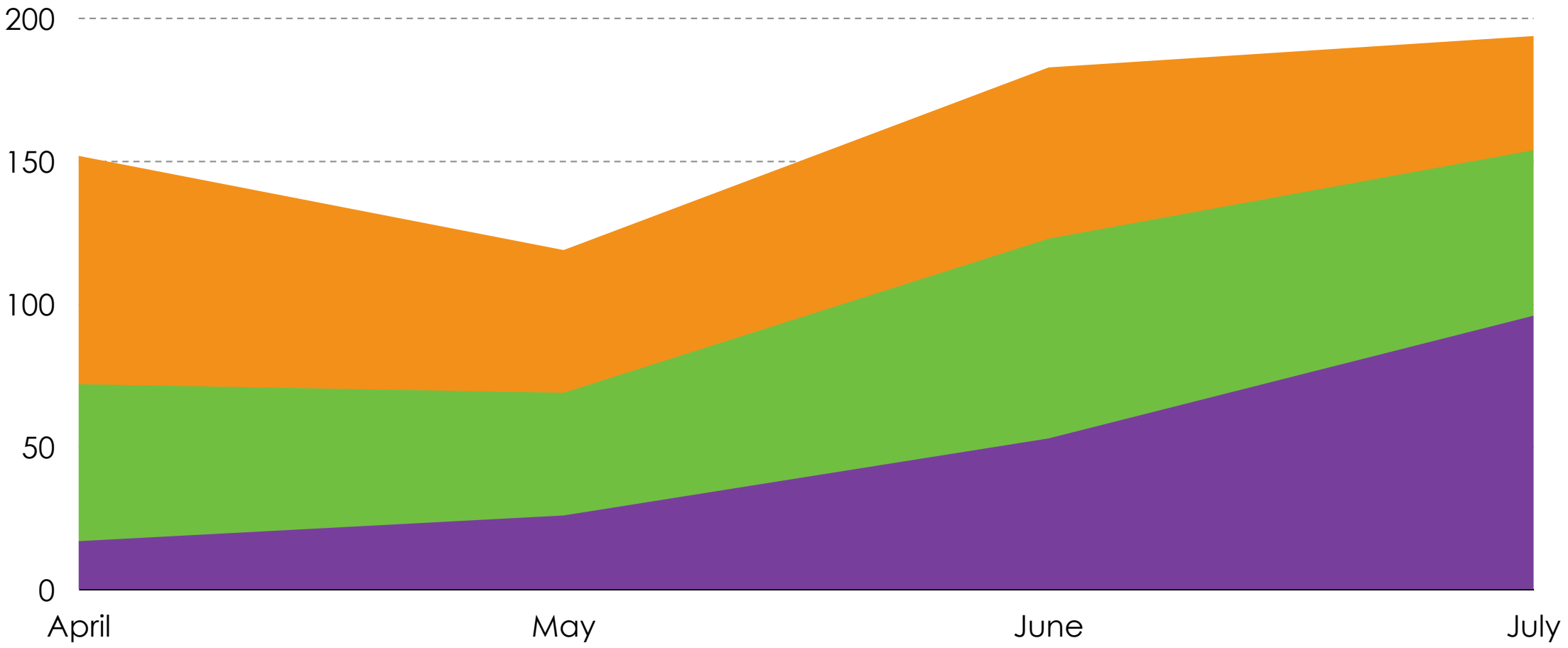
William Playfair - The Commercial and Political Atlas, 1786



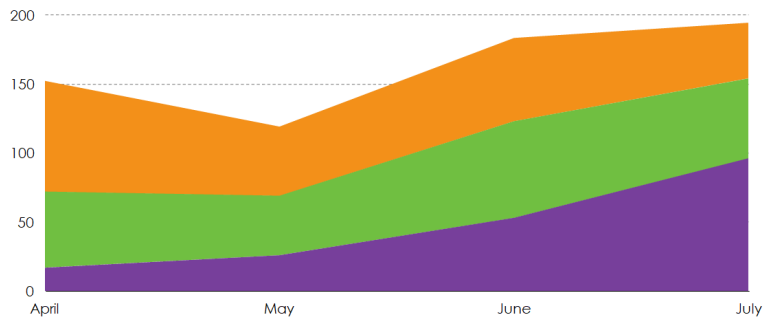
UNIVERSITY OF CALGARY

Stacked Area Charts

Stacked Area Charts



UNIVERSITY OF CALGARY



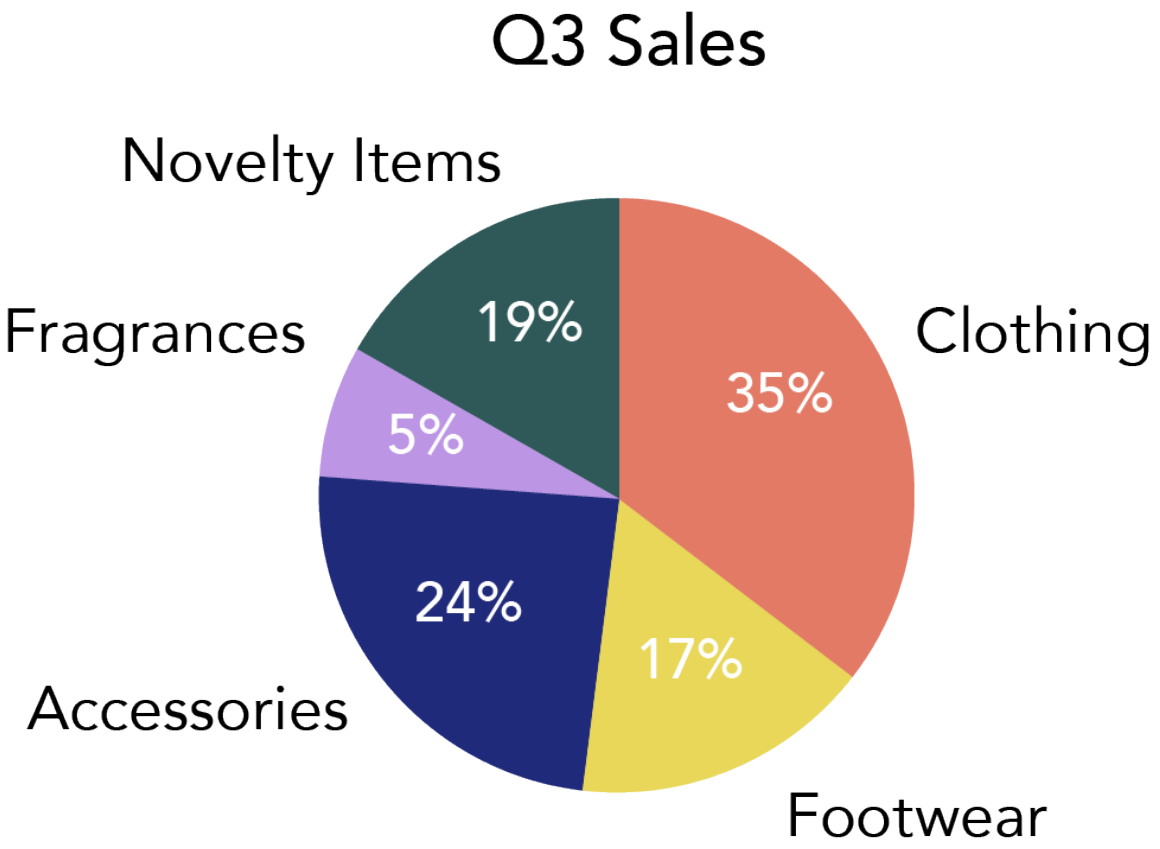
Stacked Area Charts

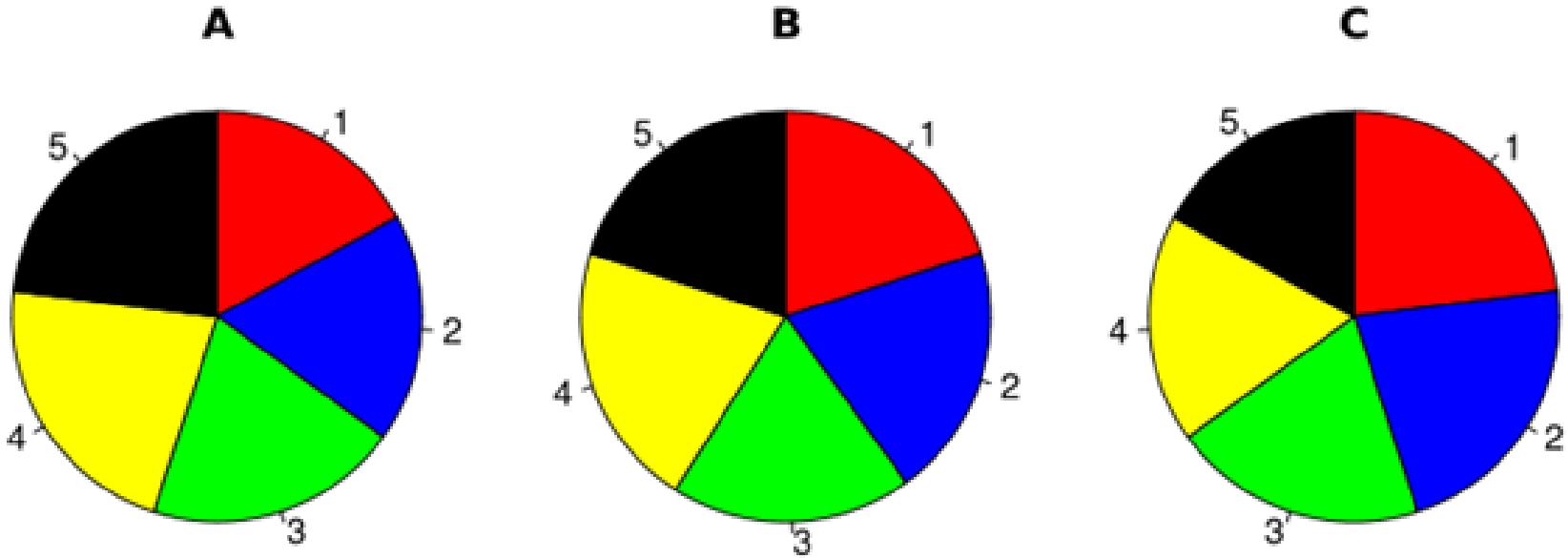
- A line chart form of stacked bar chart
- The top line becomes cumulative measure and partial measures are indicated by portions between each line
- Choice of stack order can be deceiving
- A danger with these is that visual area can be a distraction from being able to determine numerical data from chart

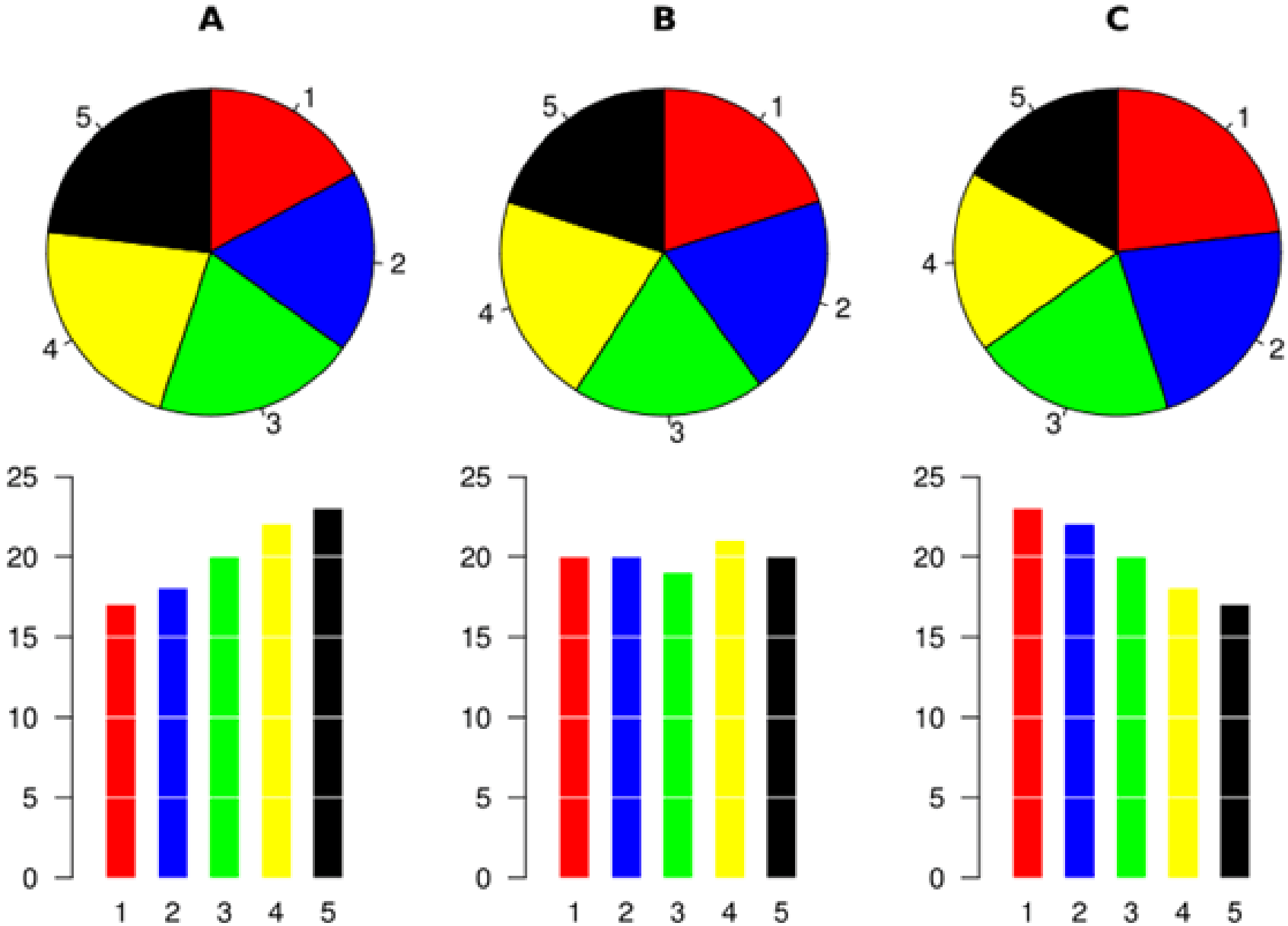
Pie Charts and Donut Charts

Pie/Donut Charts

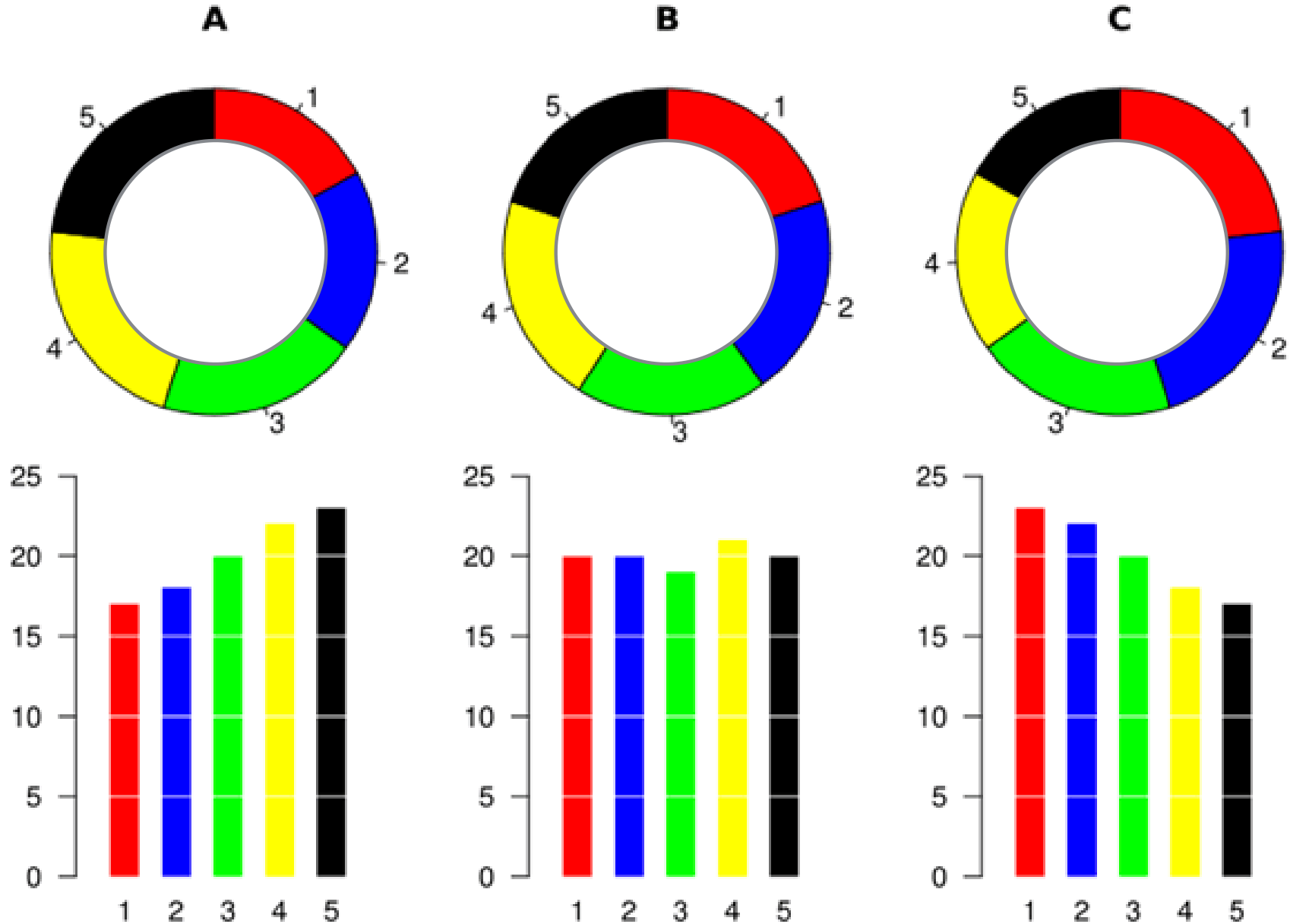
- Part of whole
- Natural idea of 100% being the complete circle (or donut)
- Each slice should have area proportional category value
- Hard to accurately draw by hand





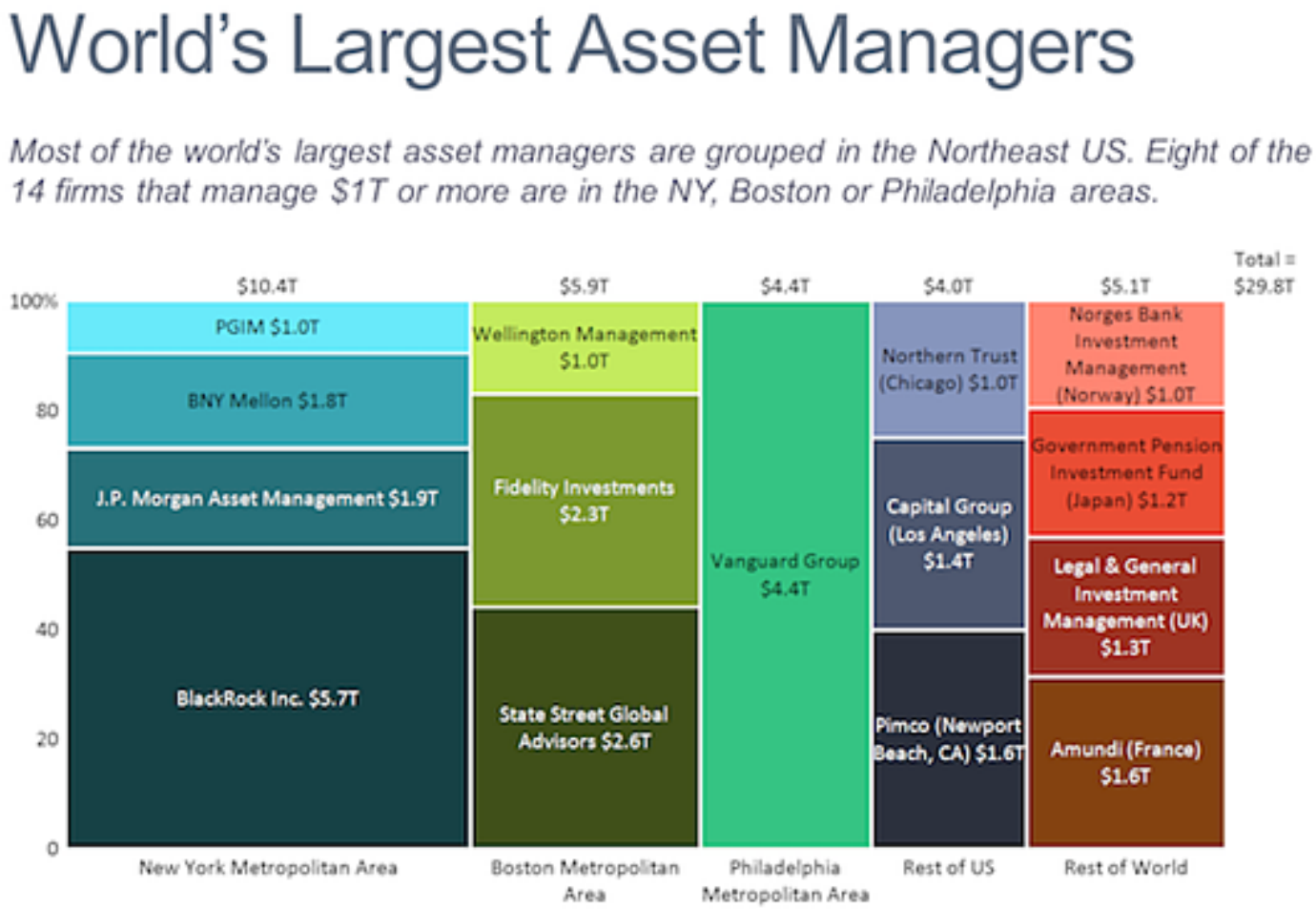


UNIVERSITY OF CALGARY



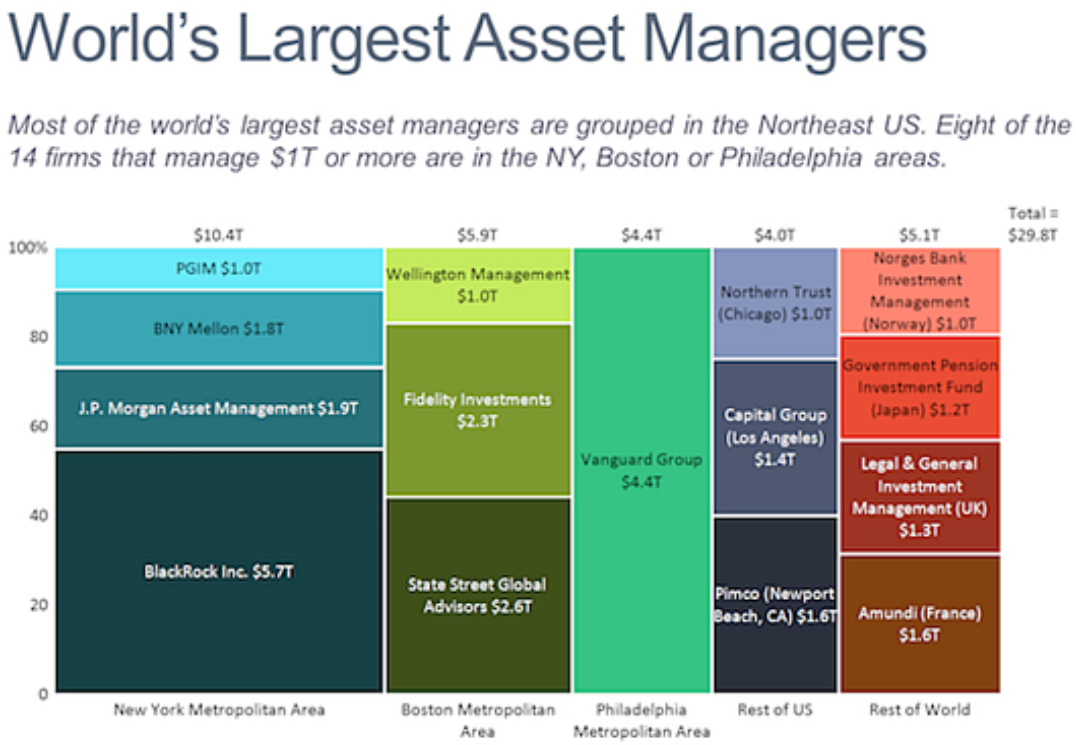
Mekko Charts (Colour Blocking Charts)

Mekko Chart



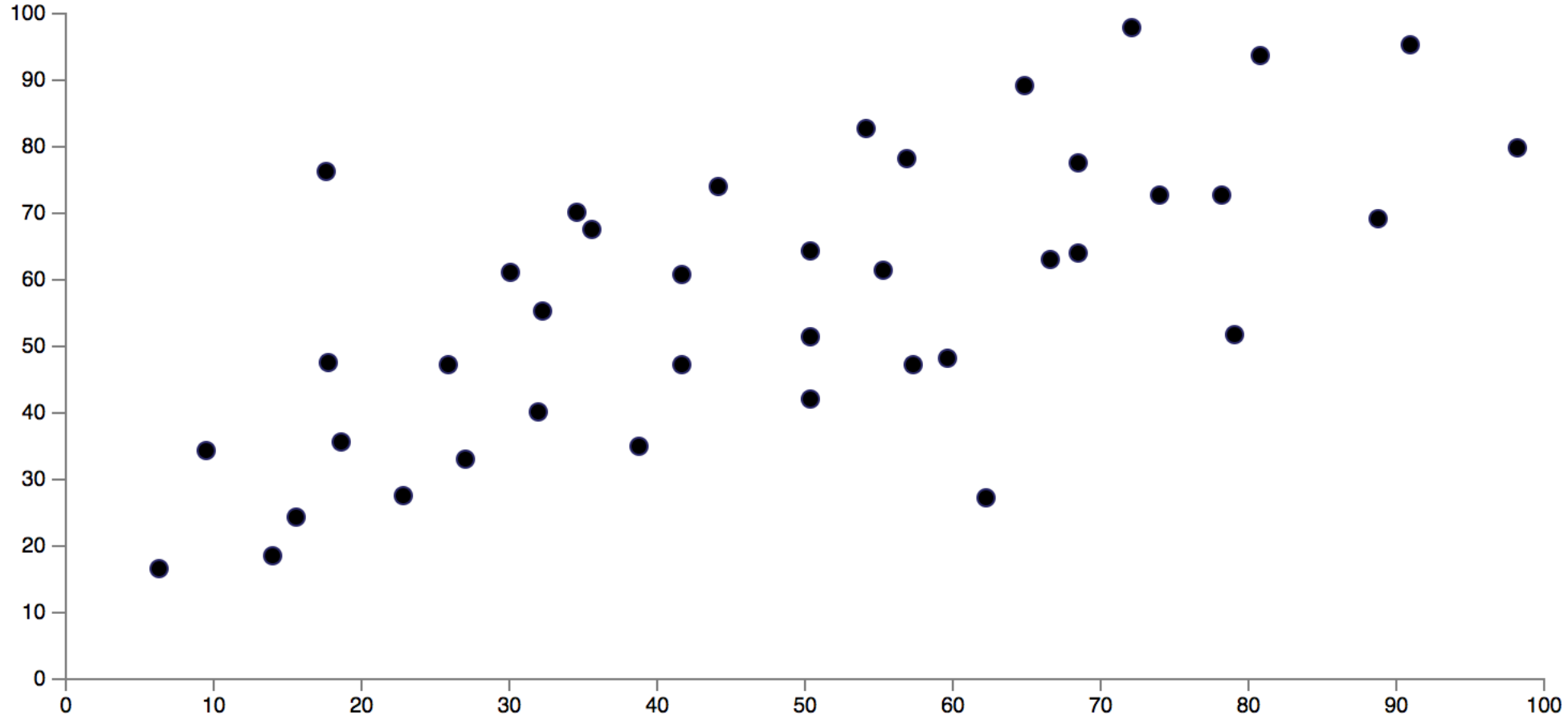
Mekko Chart

- Combination of column chart with pie chart functionality
- Parts of a whole idea
- Each sub area is relative part of whole area
- Also can compare data as a category and often stacked part of category as seen in this example
- Colour block charts don't include column chart properties (basically a rectangular pie chart)



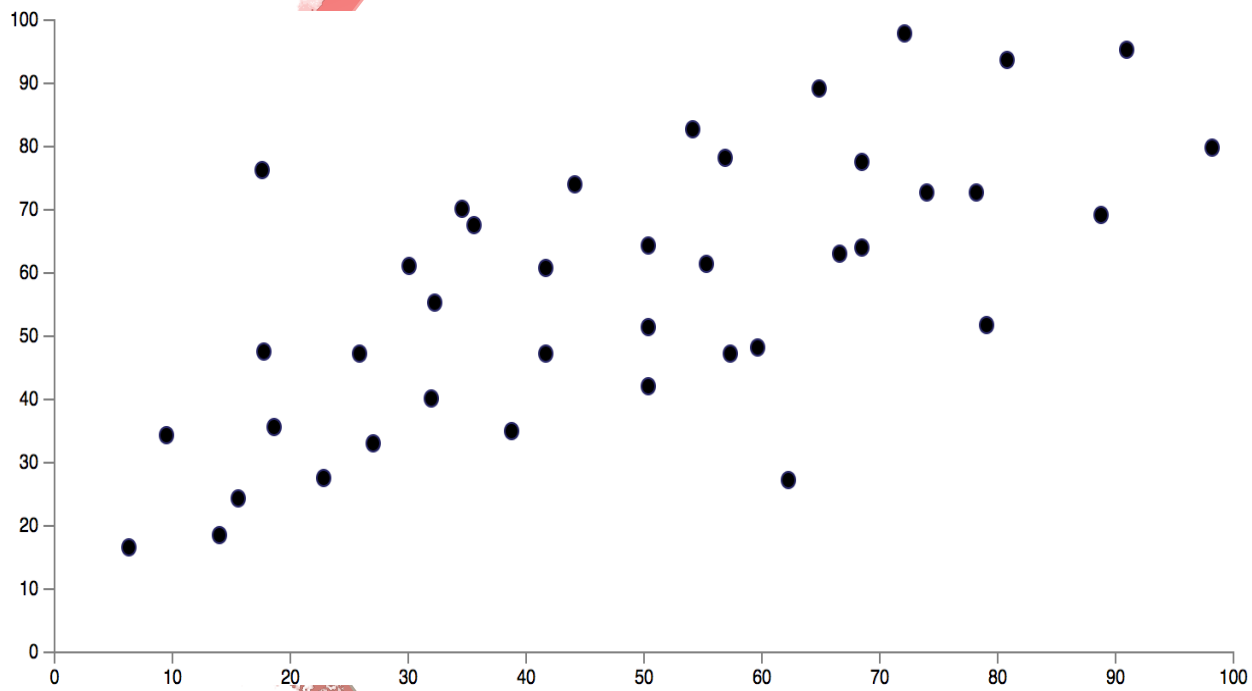
Scatter Charts (Plots)

Scatter Plots



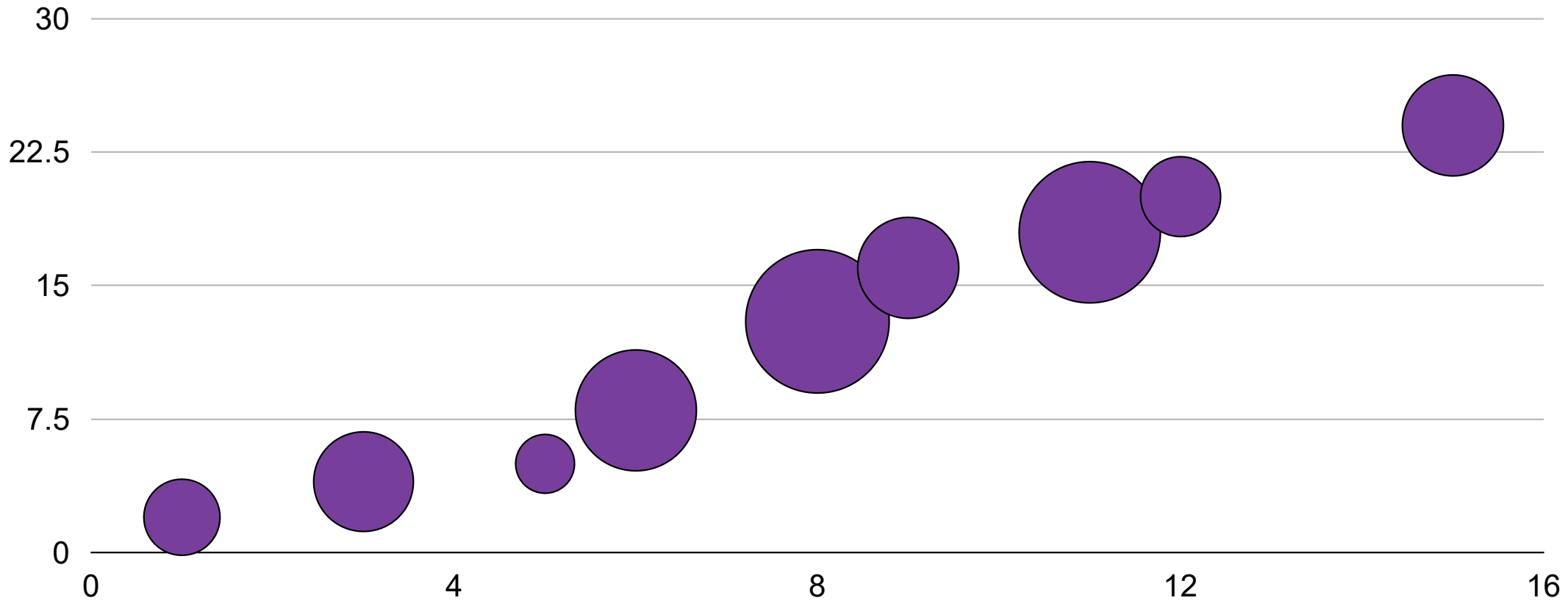
Scatter Plots

- Relationship between two variables
- Often from numerous experiments or measurements
- Reveals distributions (clusters of points, or pattern of points imply relationships or correlations)
- Can find outliers in data otherwise existing in a table of data



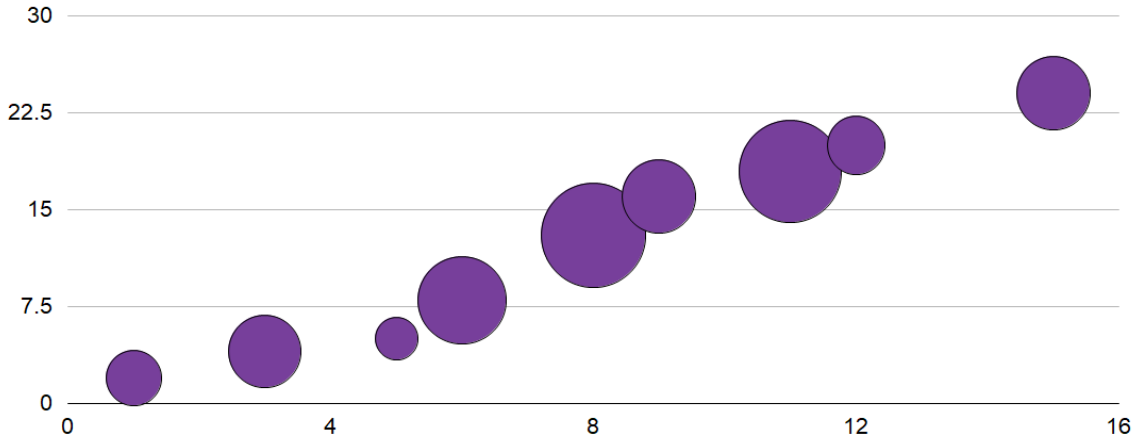
Bubble Charts

Bubble Chart



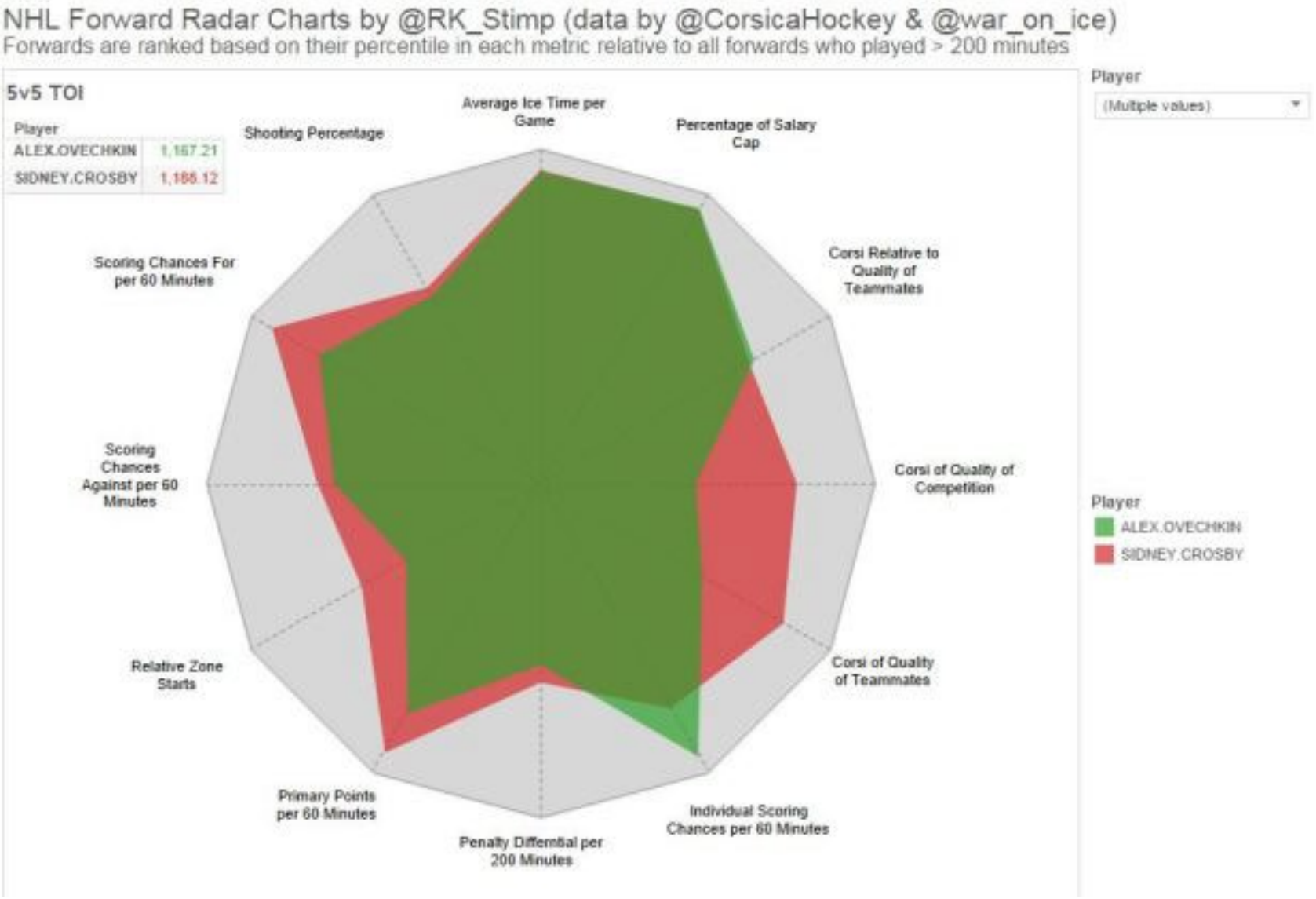
Bubble Chart

- Generally lets you expand data with and X and Y to have a third Z characteristic
- Sometimes X is category and Y is response, but sometimes both are inputs and Z (area) is response
- **Remember from visualization that size isn't great for quantitative data (I get gut reaction of ordering but not clear numerical number)**
- **Sometimes colour is a fourth variable and often a way to give a category when X and Y are both input variables**



Radar Charts

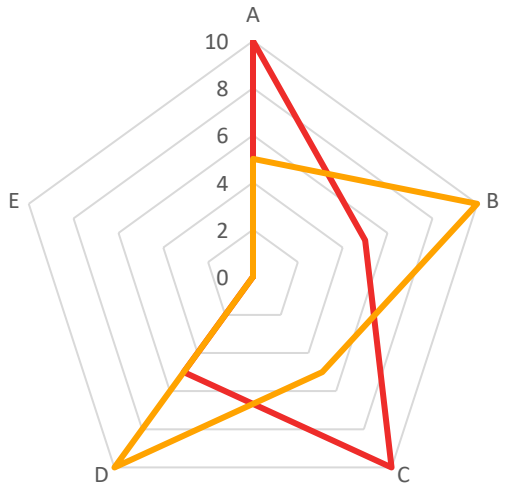
Radar Charts

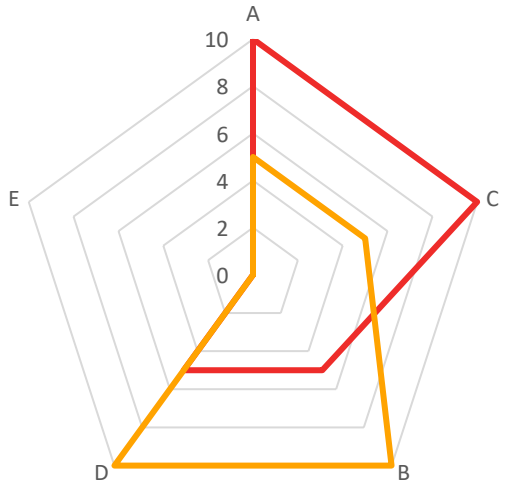


UNIVERSITY OF CALGARY

Radar Charts

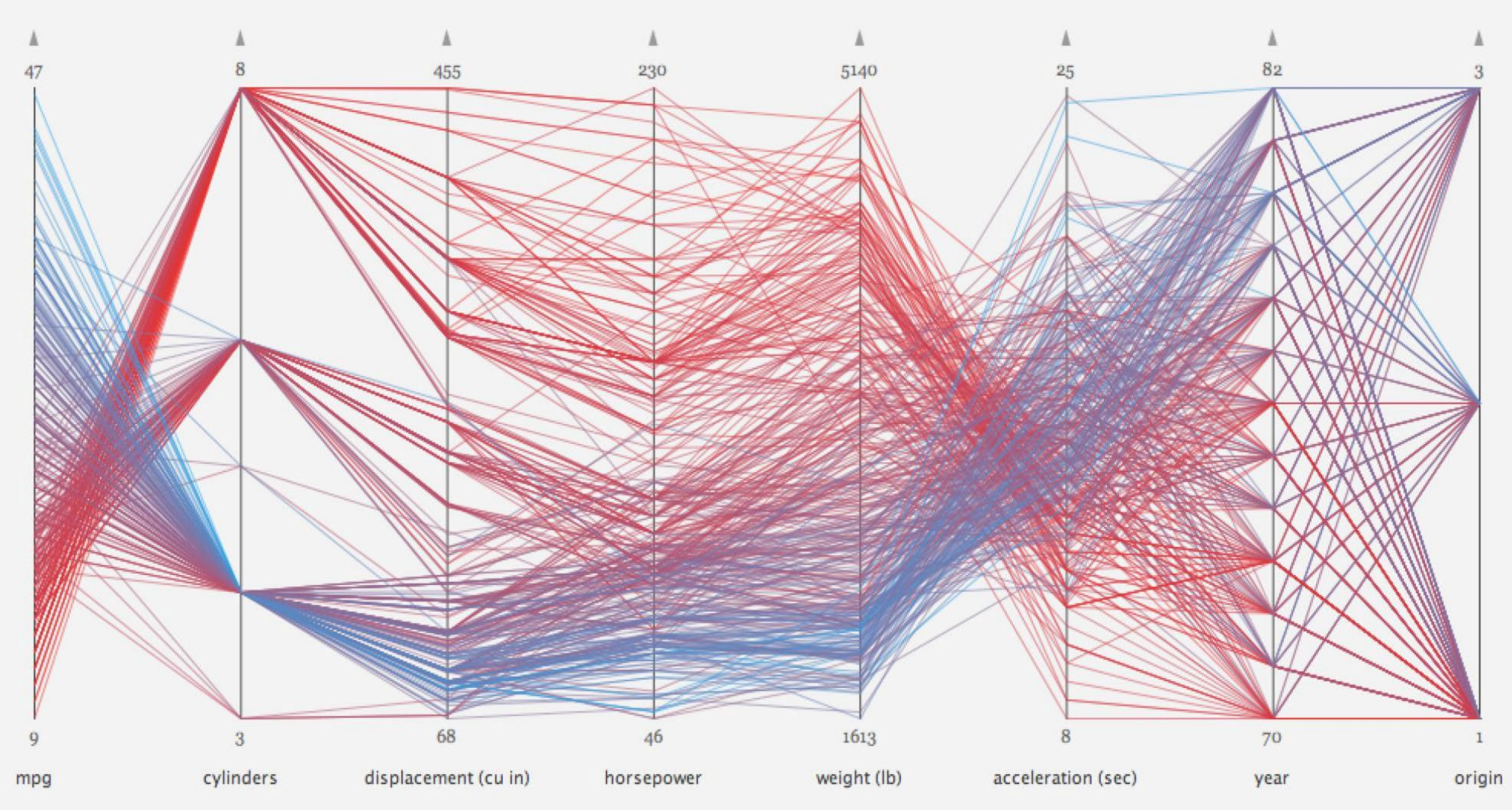
- Like a circular line chart of categories
- Ordering and area are deceiving visually
- Stacking becomes hard to perceive
- Quantity of categories also makes data less clear
- Is there a common scale for categories?



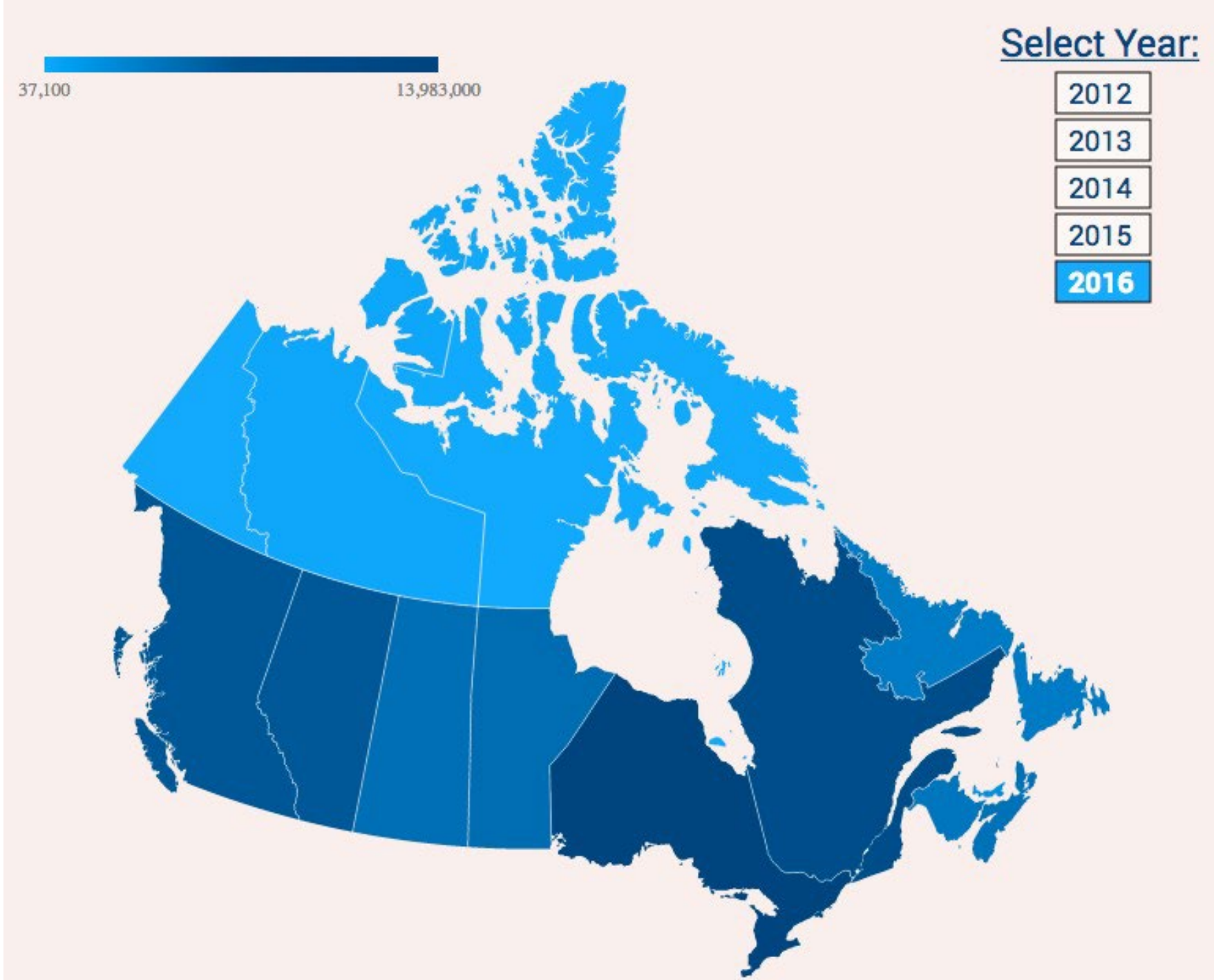


Parallel Coordinates

Parallel Coordinates



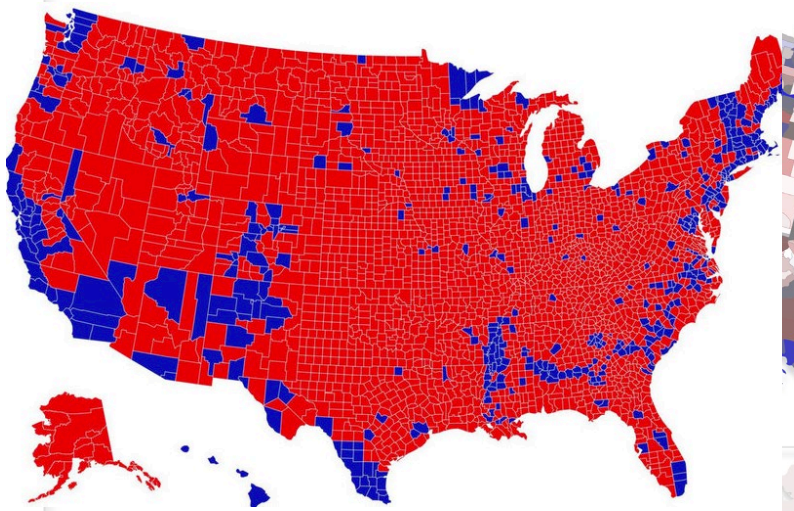
Map Charts

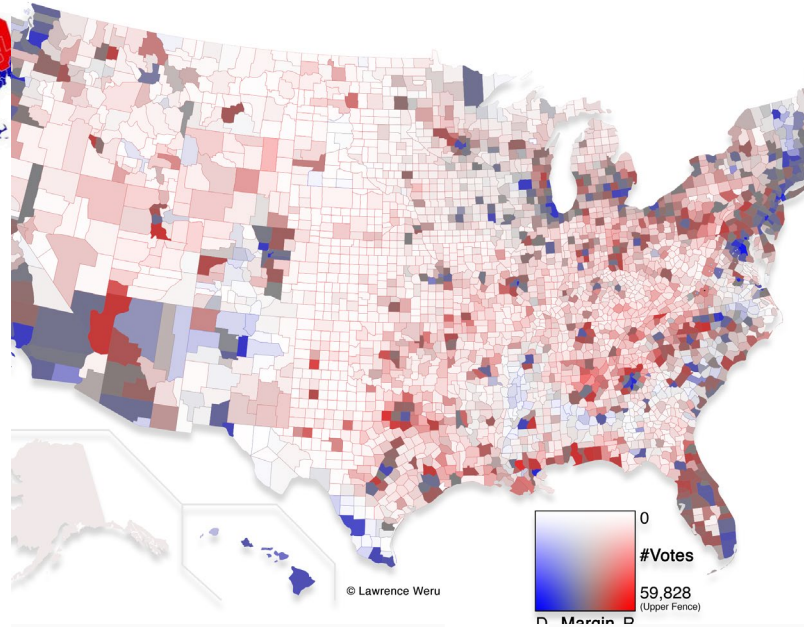


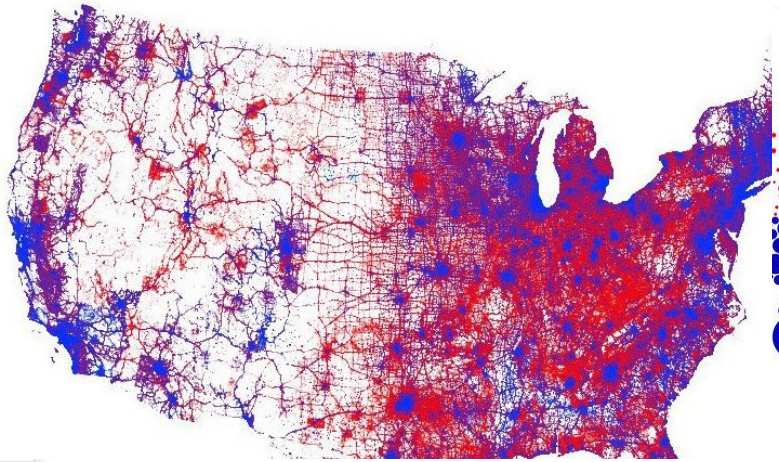
UNIVERSITY OF CALGARY

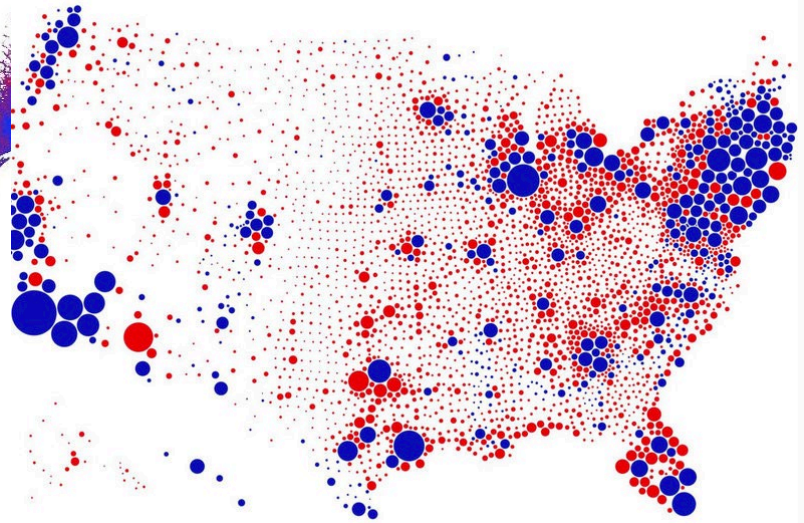
Area/Colour Can Be Deceptive

- Top left is county based voting in 2016 US Presidential election
- Bottom left one dot per vote, density map (https://nymag.com/intelligencer/2018/03/a-new-2016-election-voting-map-promotes-subtlety.html)
- Bottom right is one circle per county, sized per population (https://www.core77.com/posts/90771/A-Great-Example-of-Better-Data-Visualization-This-Voting-Map-GIF)
- Top right is counties colour by scale for both population and extremity of vote (https://nymag.com/intelligencer/2018/03/a-new-2016-election-voting-map-promotes-subtlety.html)



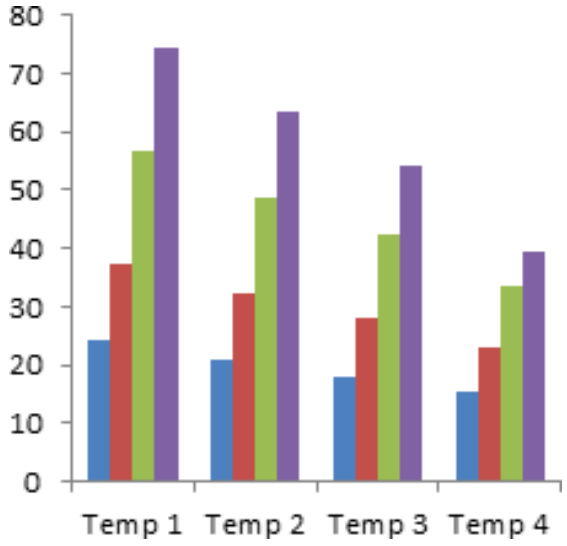


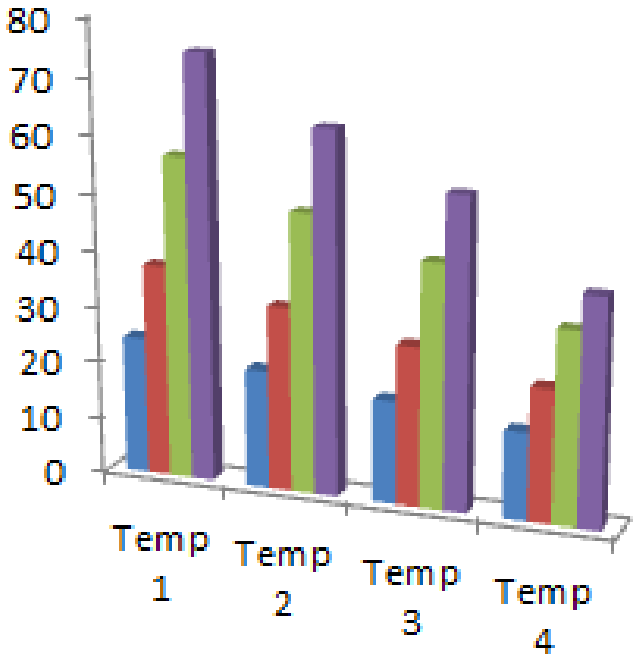


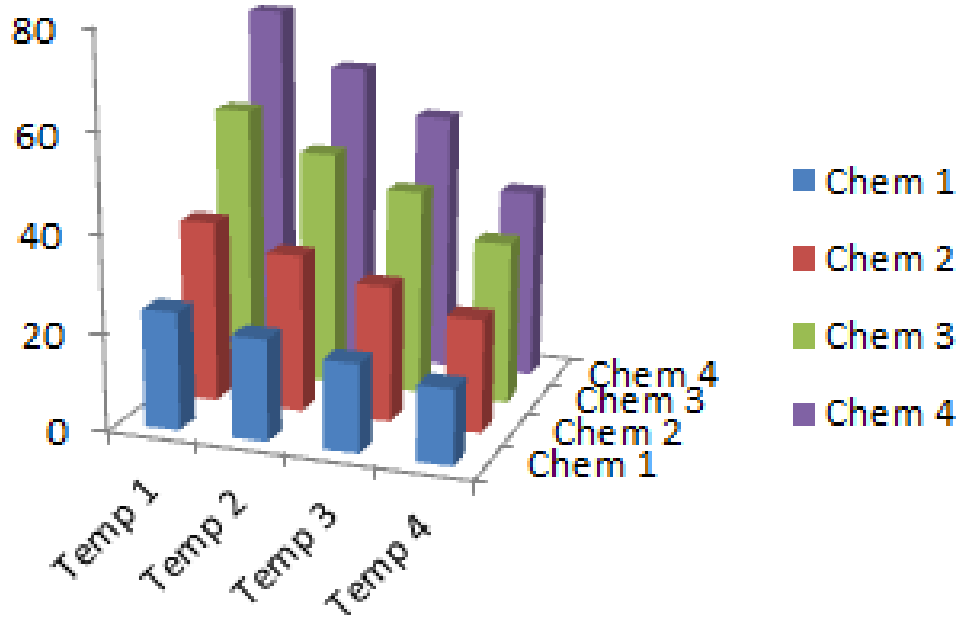


2D vs. 3D

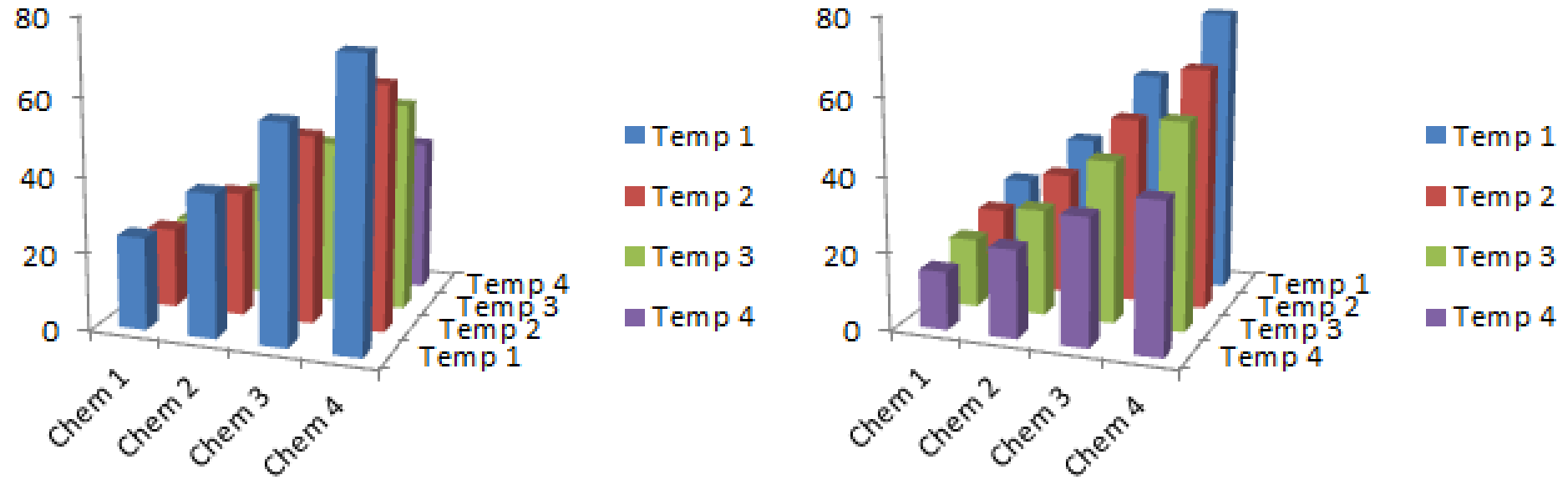
https://peltiertech.com/excel-3d-charts-charts-with-no-value







https://peltiertech.com/excel-3d-charts-charts-with-no-value



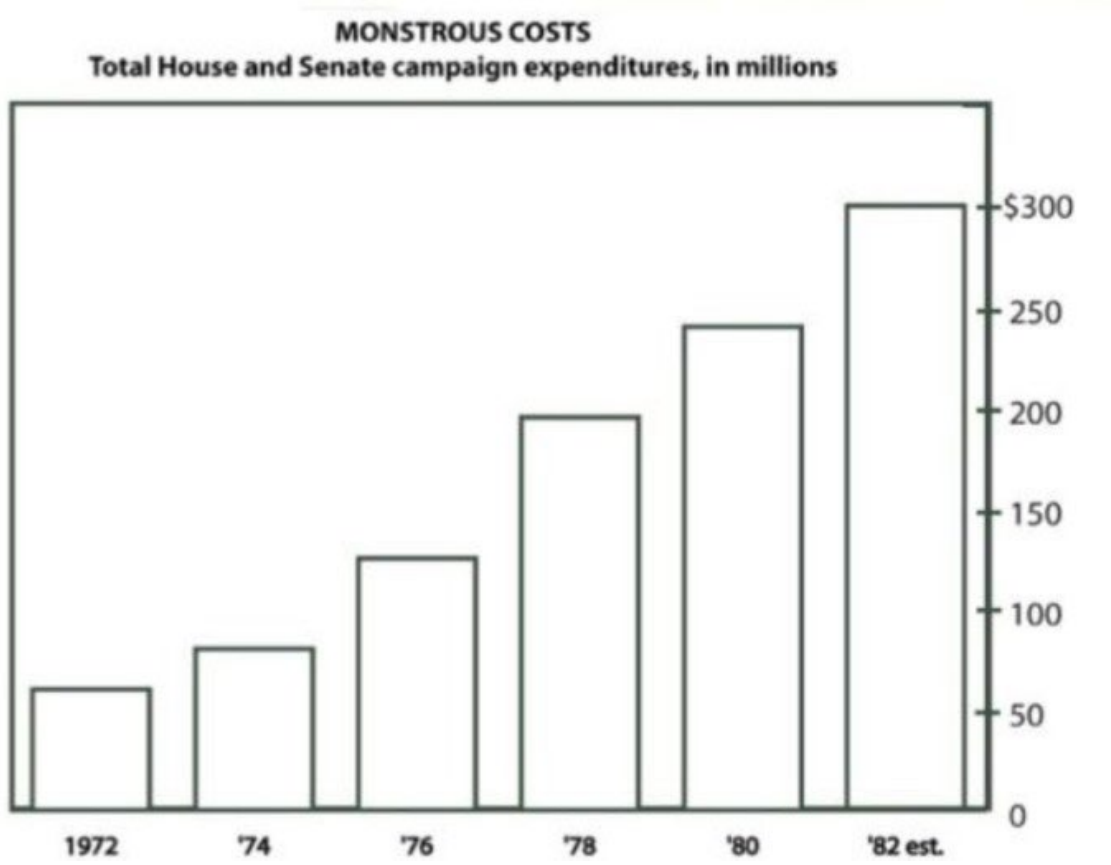
Two-Axis Charts?

Two-Axis Charts

- Used to imply correlation
- Generally considered to be bad practice and often deceptive
- Easy to change y-axis scale to manipulate ‘apparent correlation’

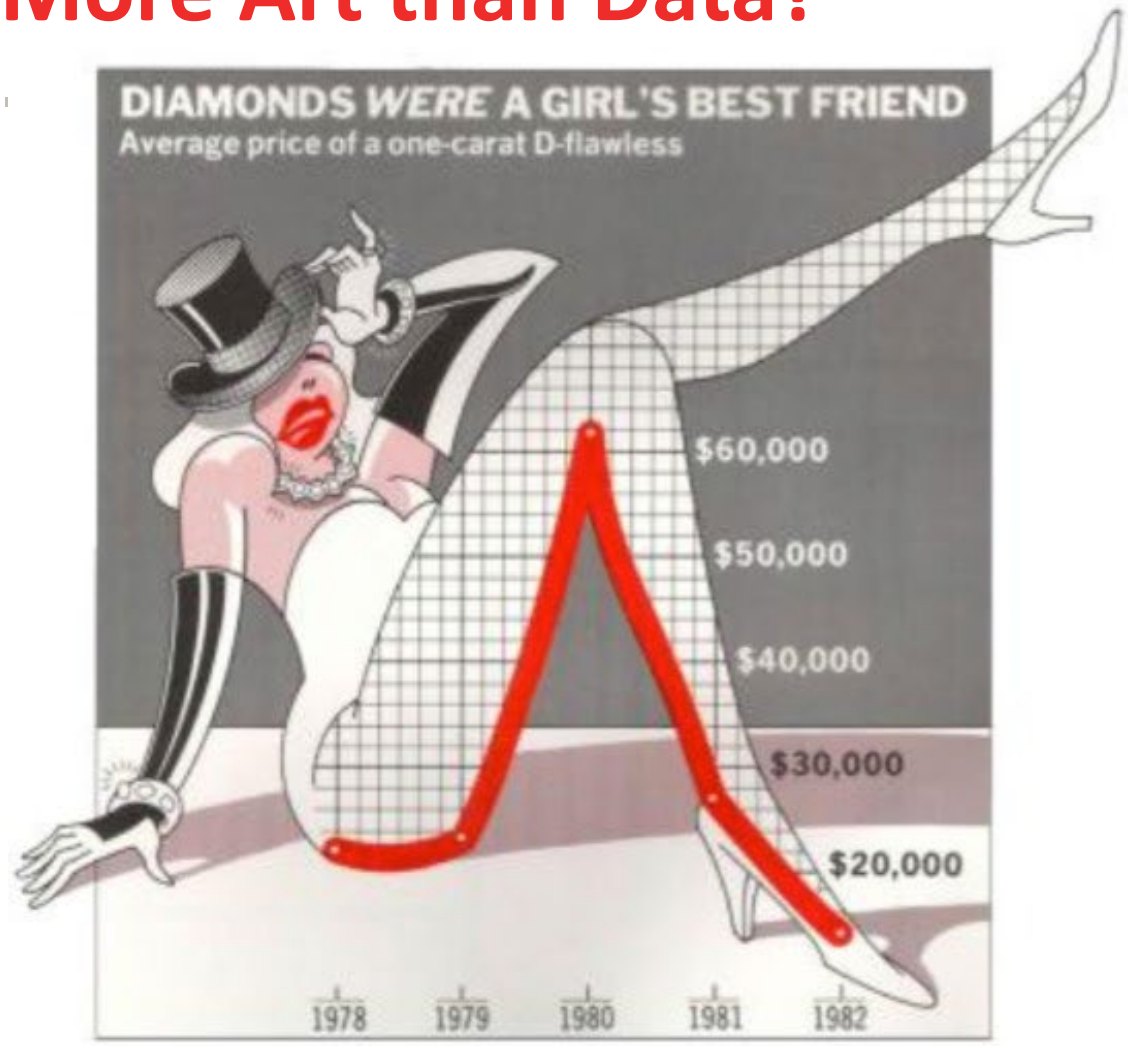
Chart Junk

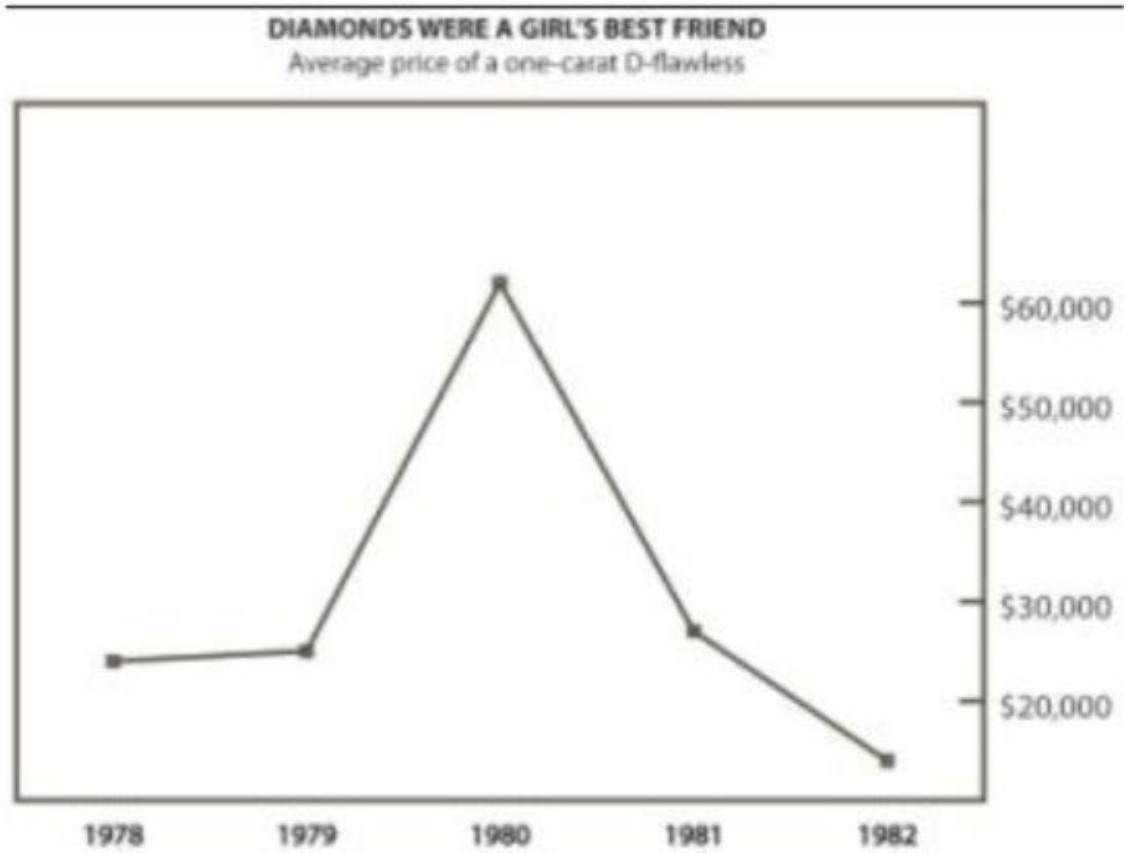
More Art than Data?



Bateman et al. 2010

More Art than Data?

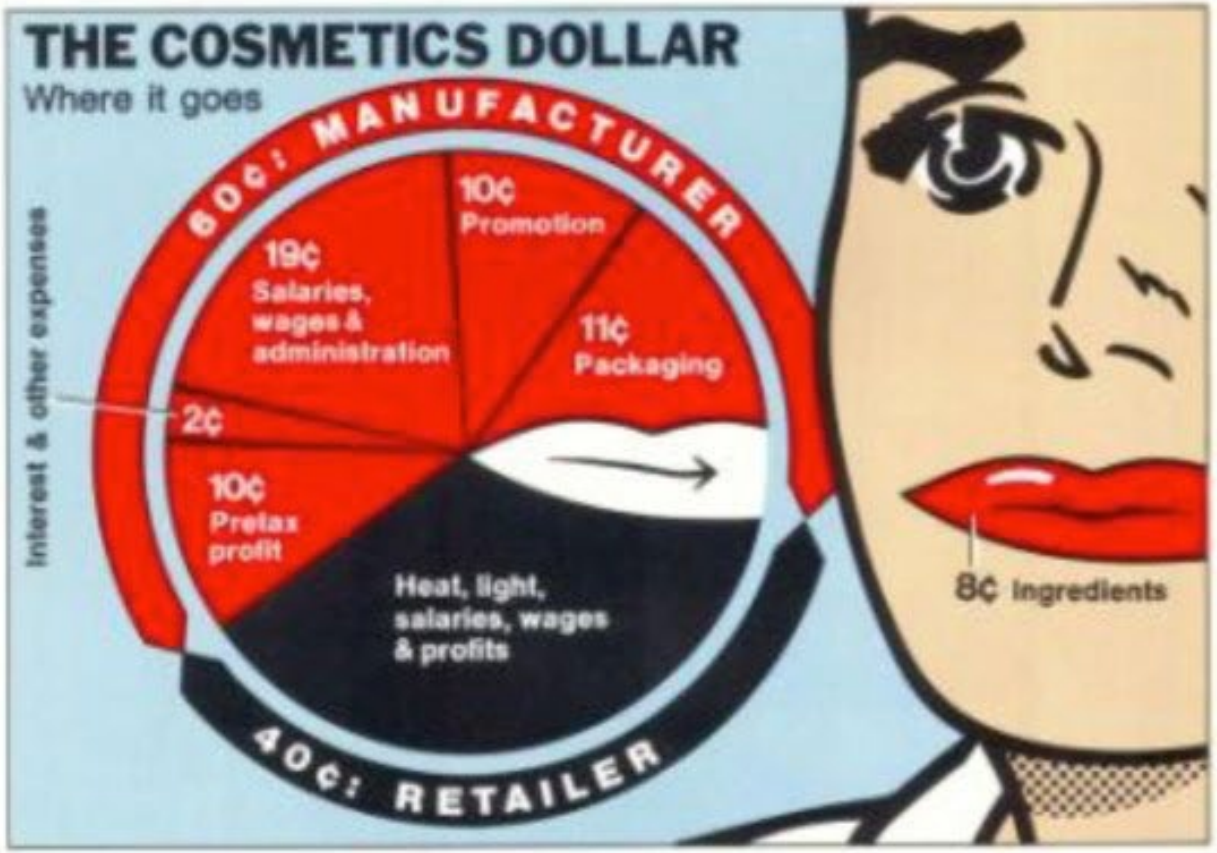


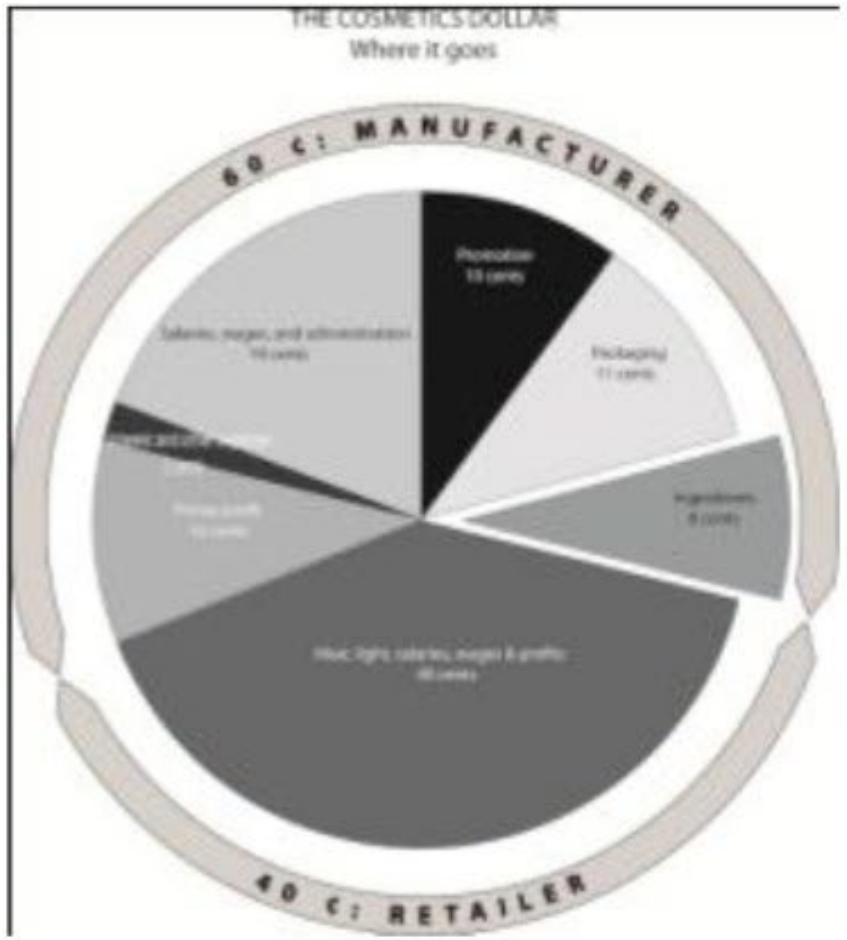


Bateman et al. 2010

UNIVERSITY OF CALGARY

More Art than Data?

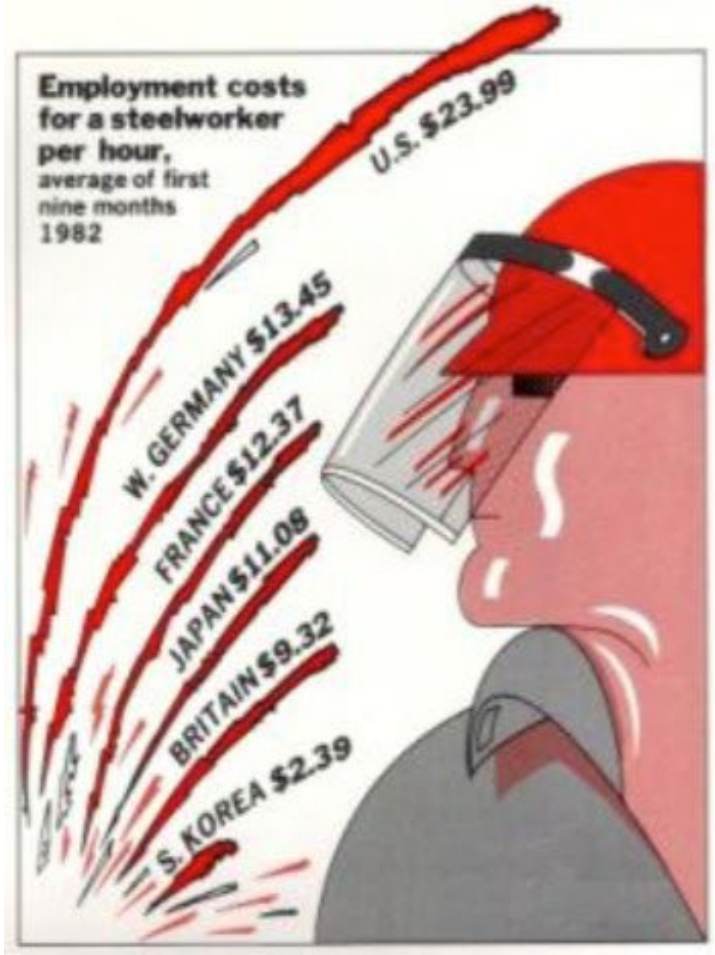




Bateman et al. 2010

UNIVERSITY OF CALGARY

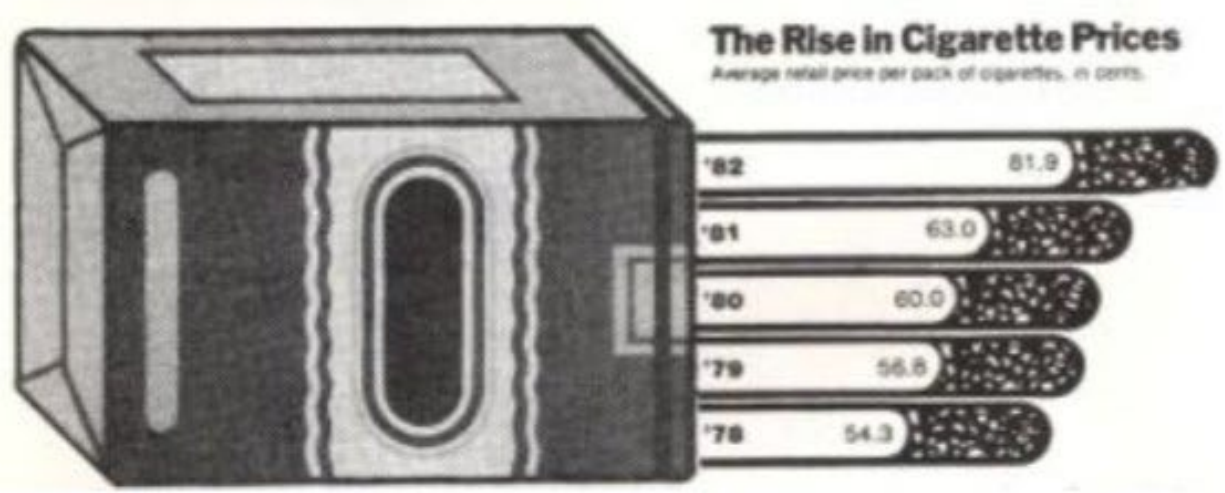
More Art than Data?

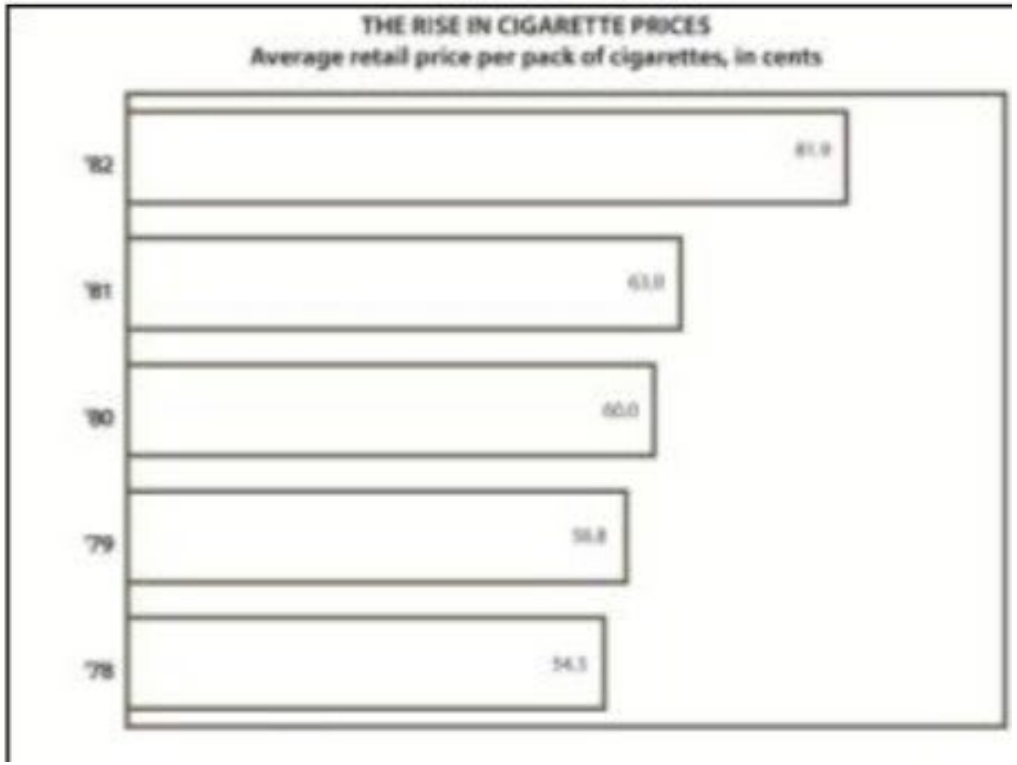


Bateman et al. 2010

UNIVERSITY OF CALGARY

More Art than Data?





Bateman et al. 2010

Onward to … Presentations

Jonathan Hudson
jwhudson@ucalgary.ca
https://pages.cpsc.ucalgary.ca/~jwhudson/

UNIVERSITY OF CALGARY