W3D1 – DataViz Lecture

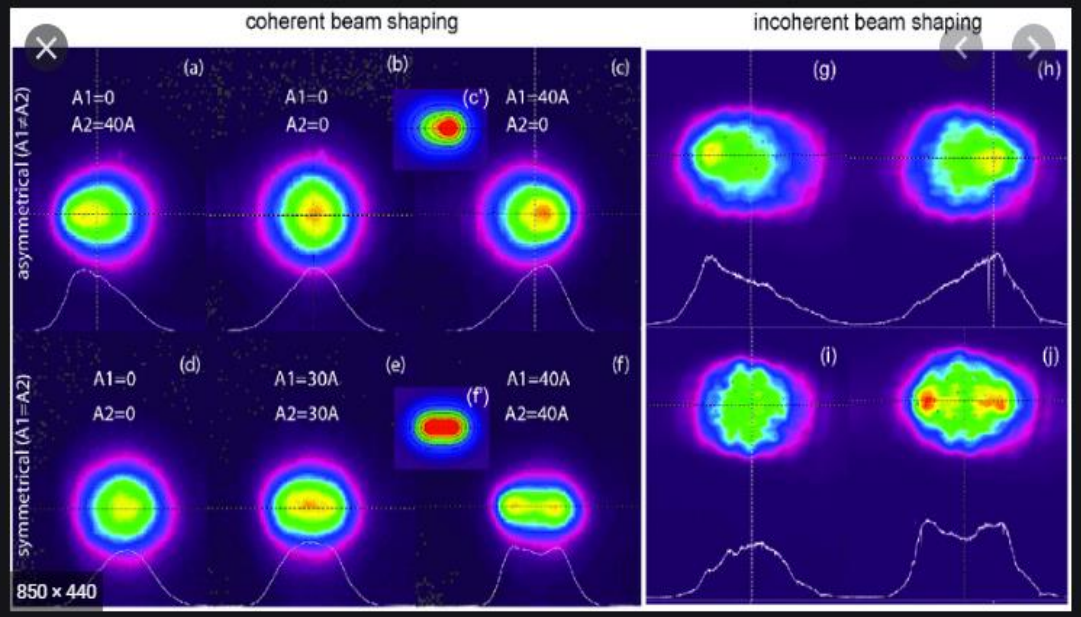
LHL DS Bootcamp July 2020

Arunabh Singh

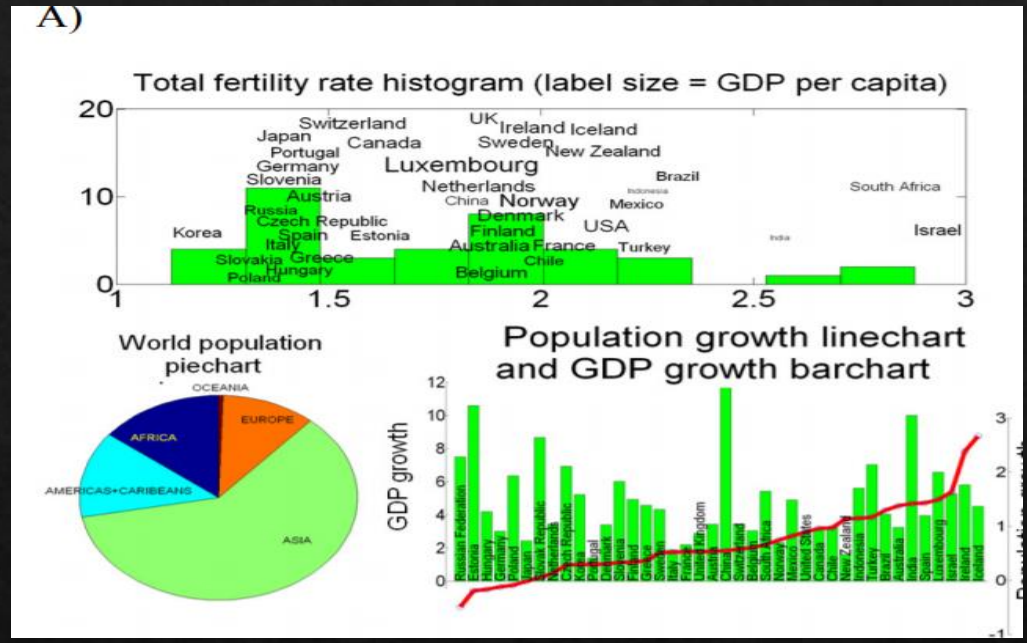
Why Data Visualization?

- A picture is worth a thousand words.
- A high level representation of data.
- Be able to spot mistakes easy to overlook
- Understanding the market audience (Marketing, Politics, and Sales).
- Predicting trend (Time Series)

Design Principles



Symmetrical



Asymmetrical

Radial

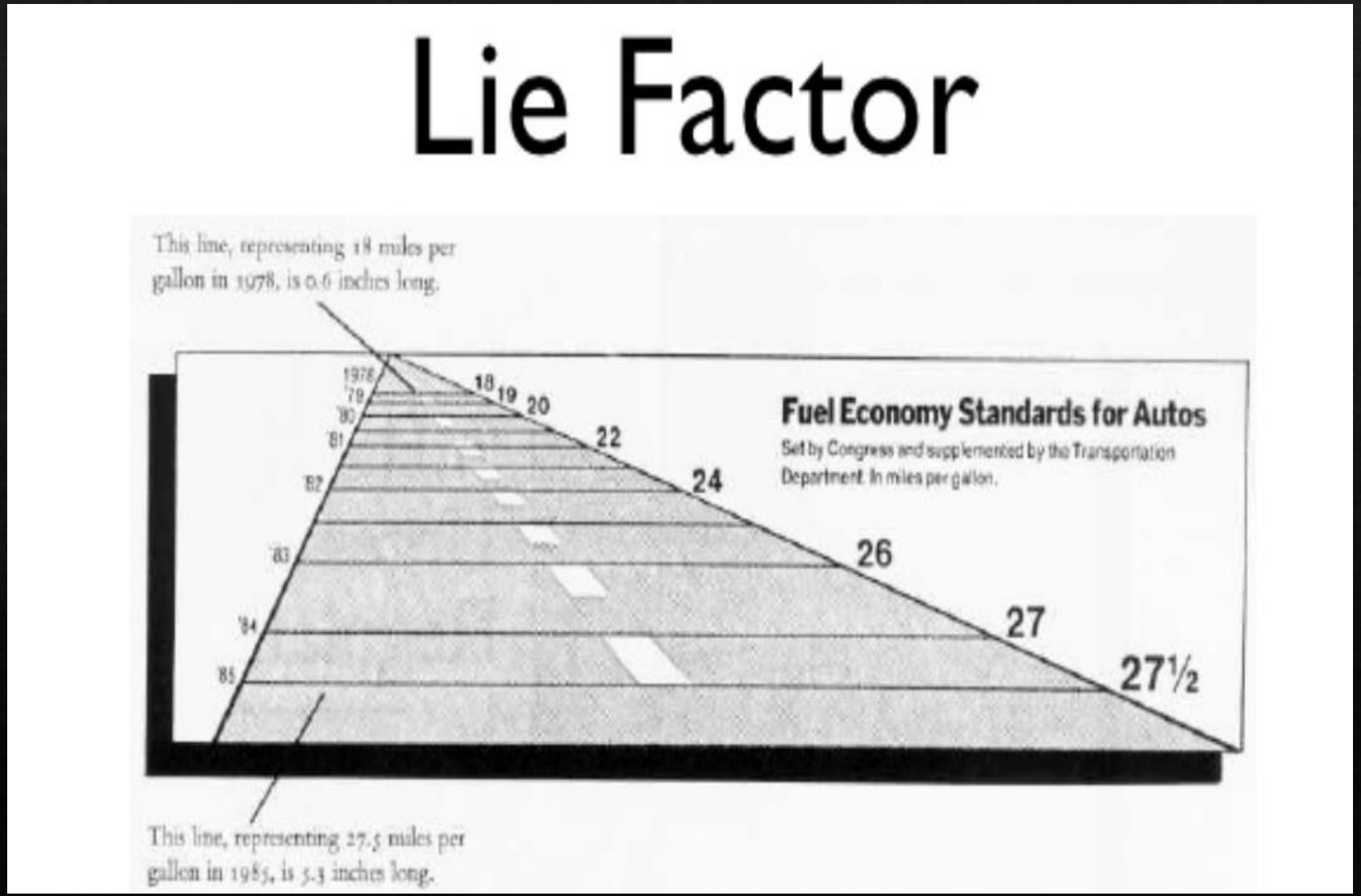
More Design Principles

- Emphasize – draw attention using colors, size, certain locations like top-left corner.
- Movement – Follow an ‘F ‘ pattern of reading for user and place topics in that fashion.
- Movement – Follow an ‘F ‘ pattern of reading for user and place topics in that fashion.
- Pattern - establish a “pattern” by using similar objects, colors, and chart types to display the information. Disrupting the pattern naturally draws curiosity.
- Repetition, Proportion and Rhythm

Tufte's Integrity Principles

1. Representations of numbers should match their true proportions.
2. Labeling should be clear and detailed.
3. Designs should not vary from some ulterior motive, but show only data variations.
4. Well known units are best when representing money.
5. The number of dimensions represented should be the same as the number of dimensions in the data.
6. Representations should not imply an unintended context.

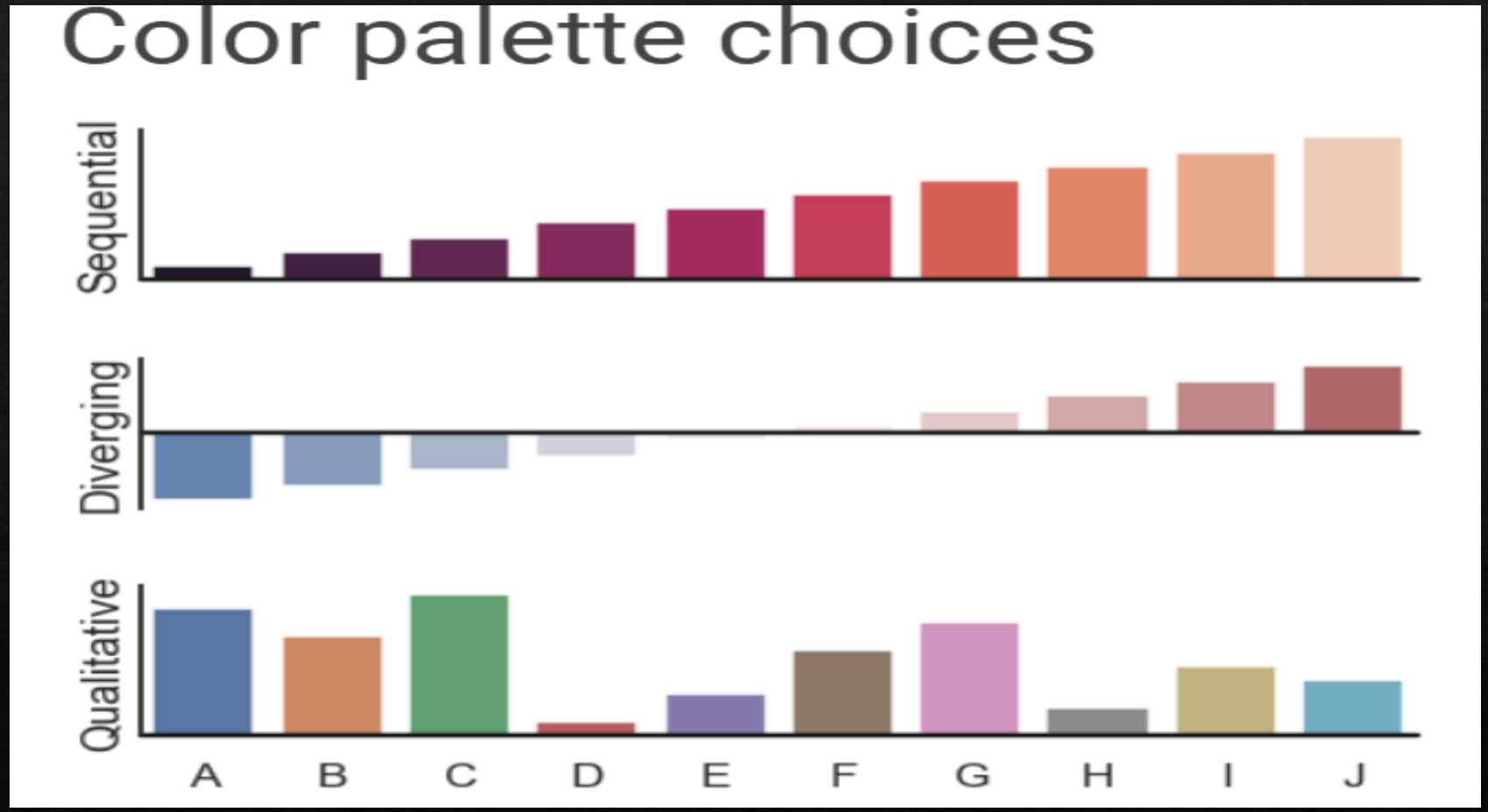
Lie Factor



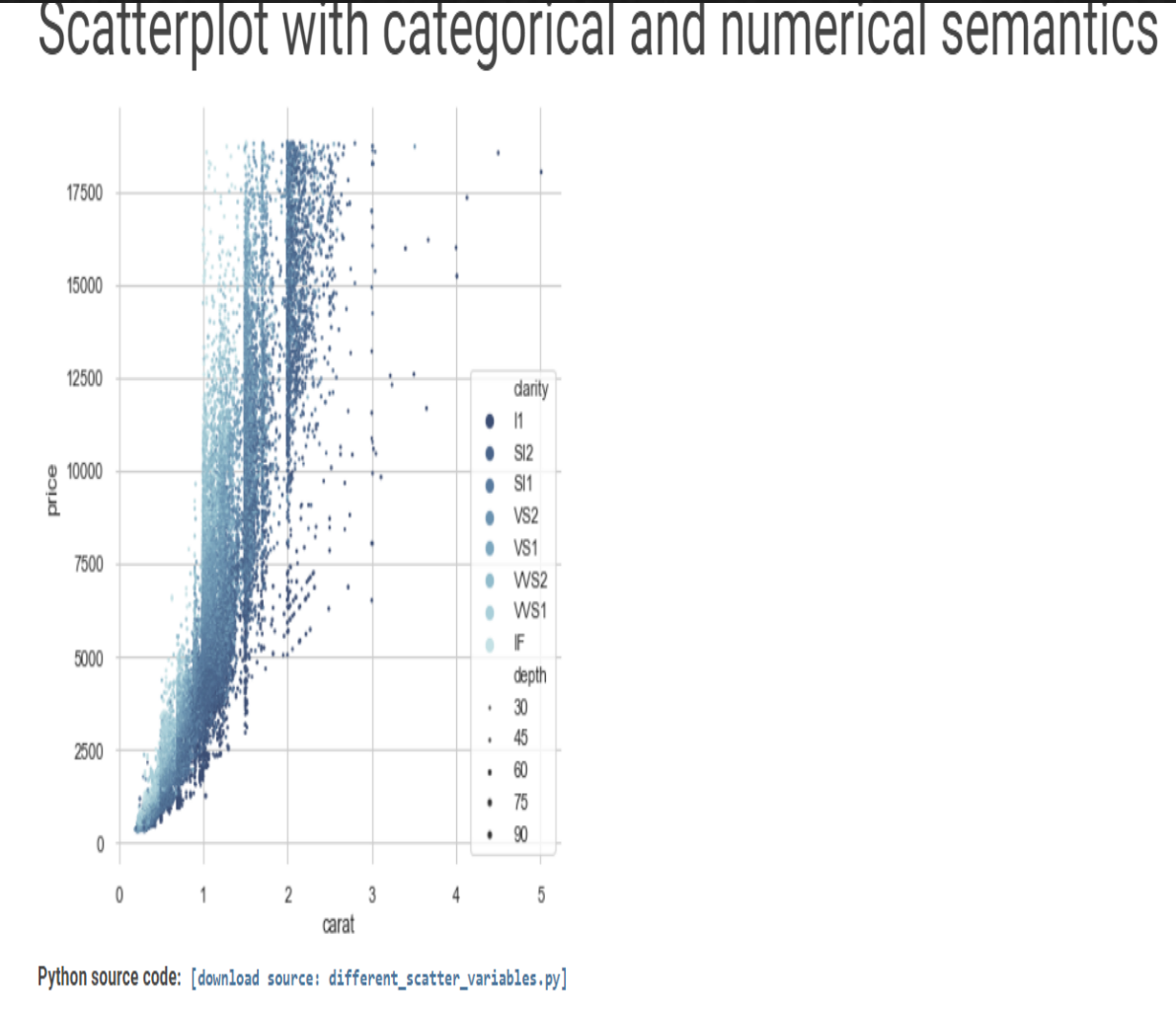
- Lie Factor = Size of Effect in graph/Size of effect in Data.
- Lie Factor > 1 means graph overstates the effect.

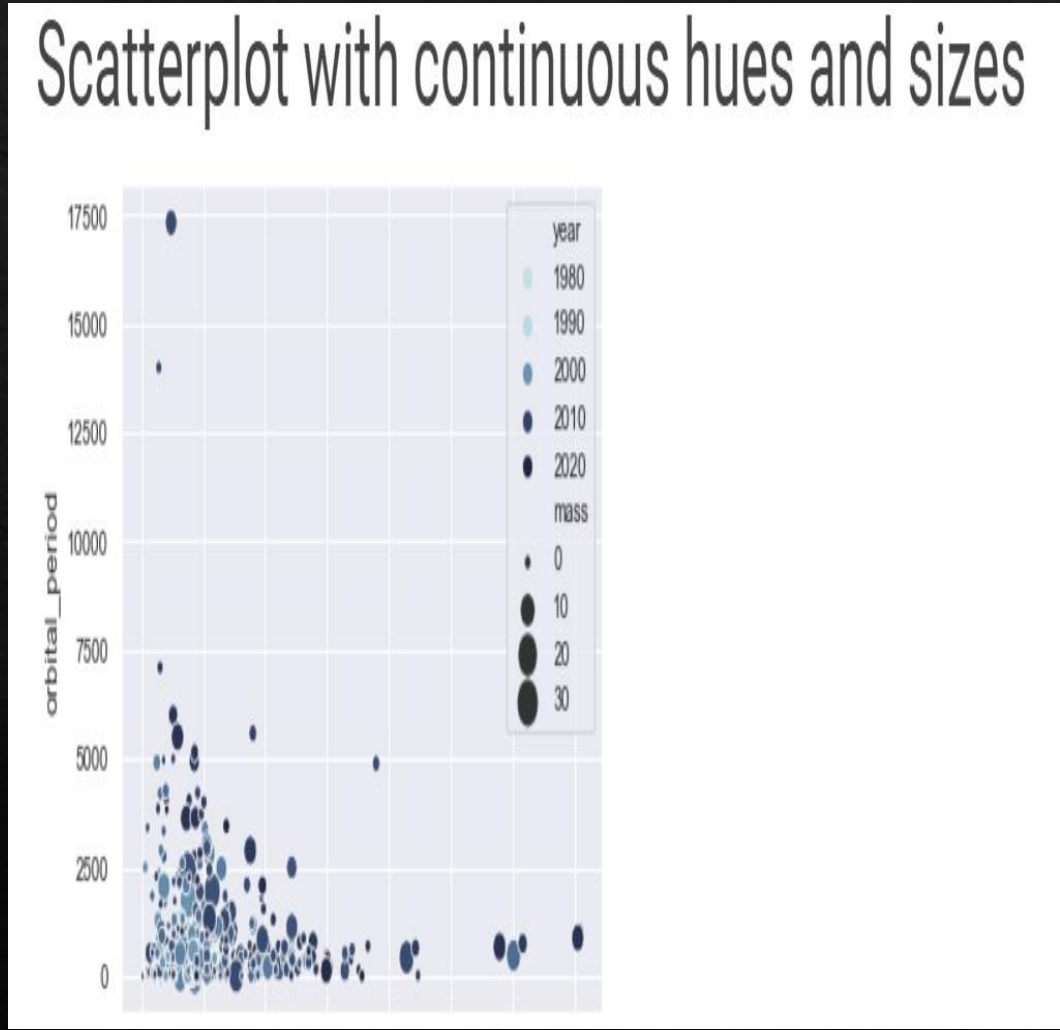
Basic Type of Charts

Bar Plot

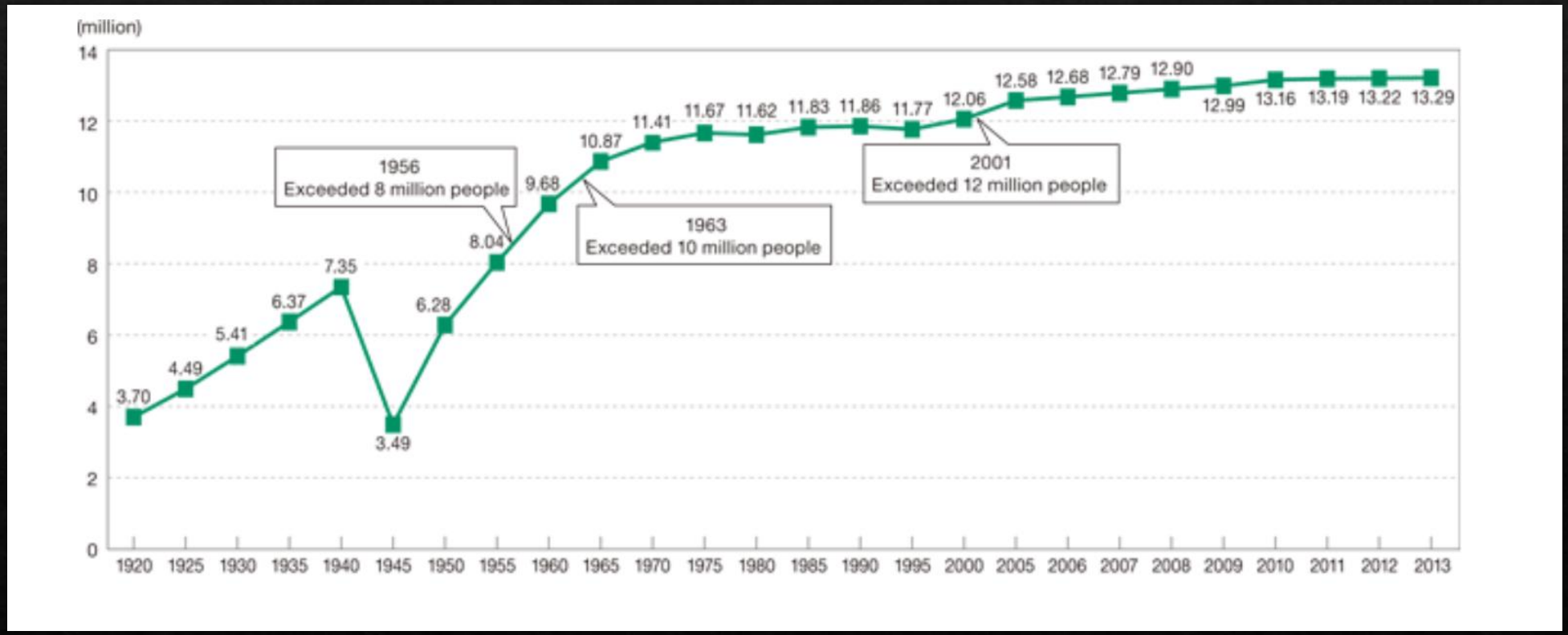

Color palette choices
Sequential
Diverging
Qualitative
A
B
C
D
E
F
G
H
I
J

Scatterplot

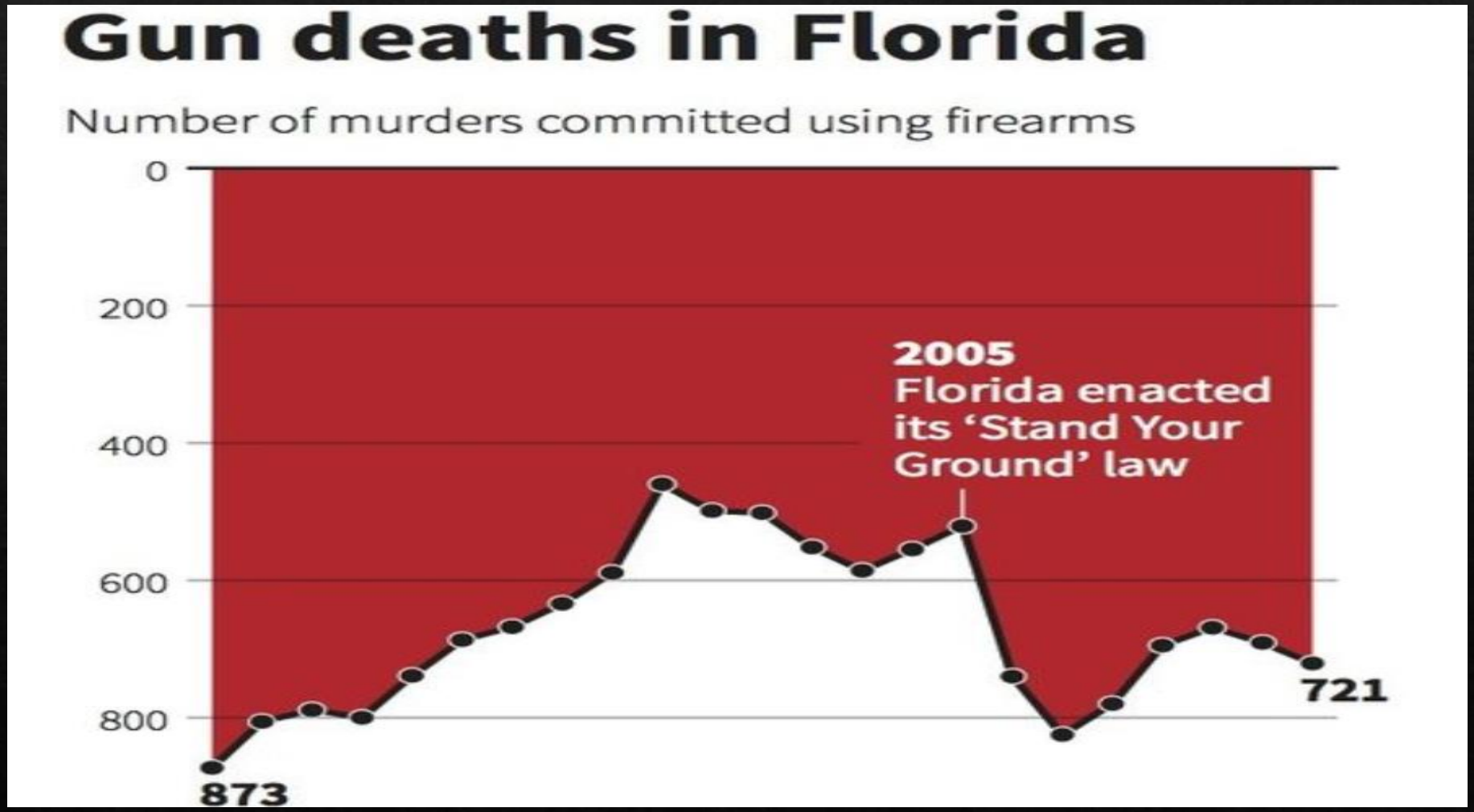




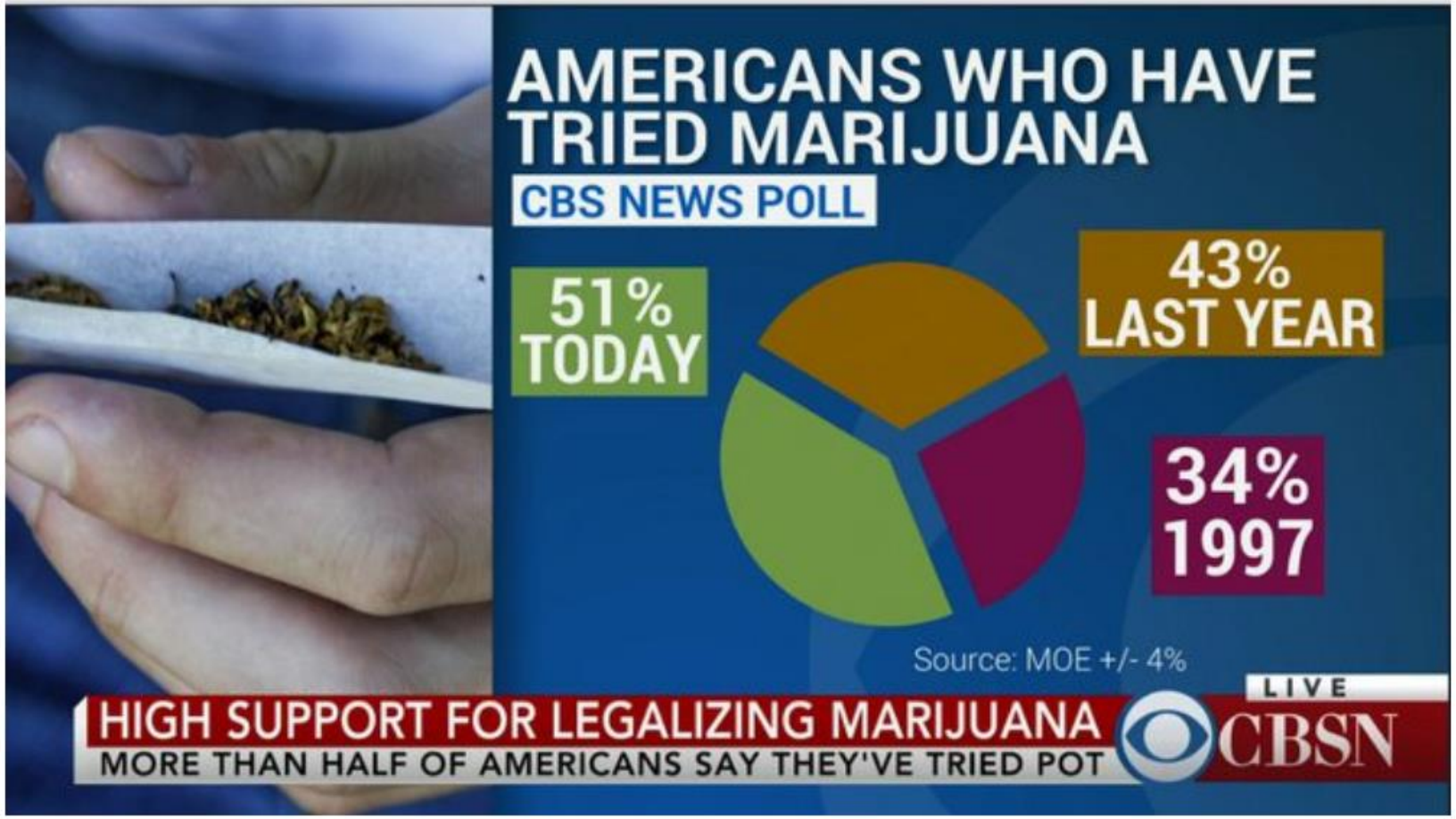
What's wrong ?


(million)
14
12
10
8
6
4
2
0
3.70
4.49
5.41
6.37
7.35
3.49
6.28
8.04
9.68
10.87
11.41
11.67
11.62
11.83
11.86
11.77
12.06
12.58
12.68
12.79
12.90
12.99
13.16
13.19
13.22
13.29
1956
Exceeded 8 million people
1963
Exceeded 10 million people
2001
Exceeded 12 million people
1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2006 2007 2008 2009 2010 2011 2012 2013

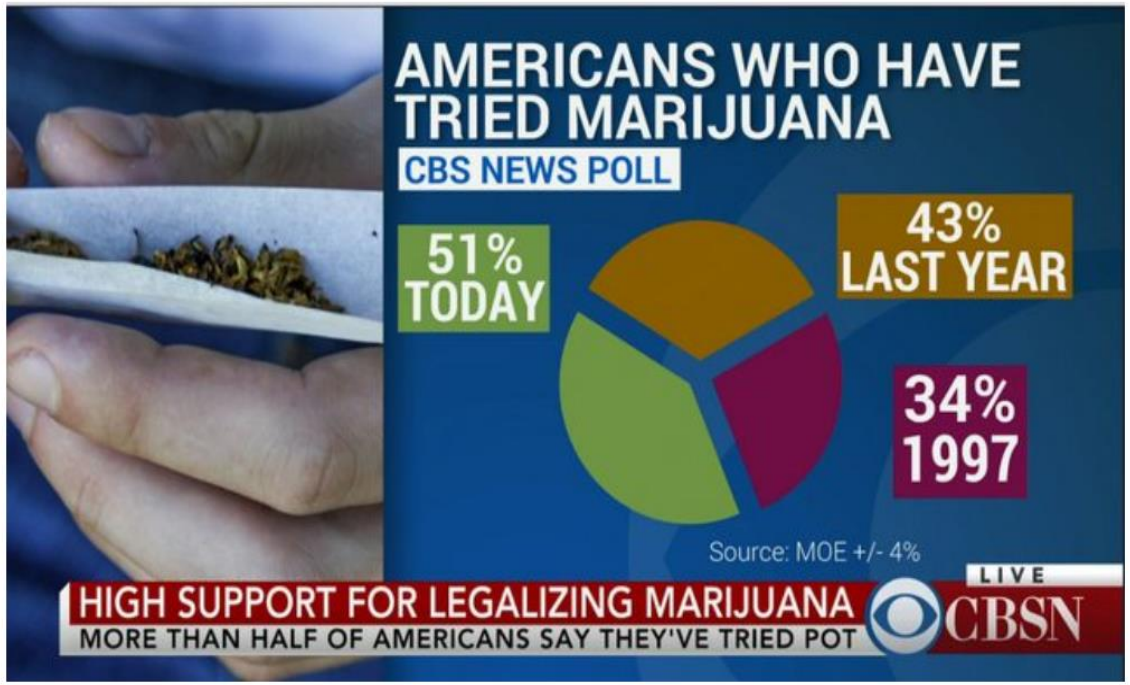
See anything incorrect?

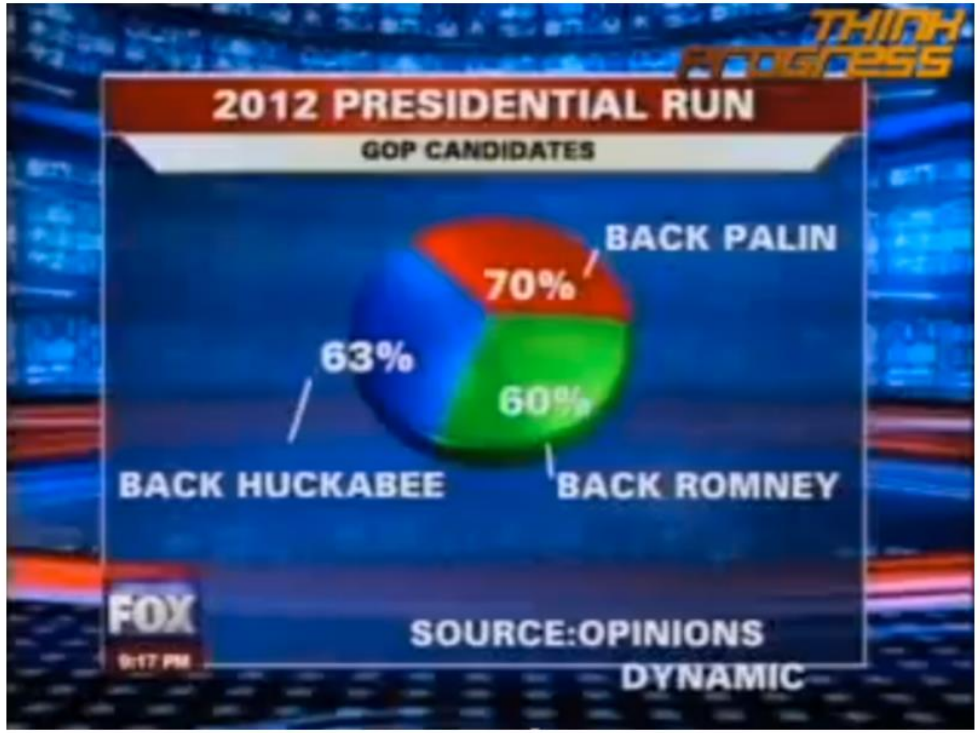


Bad at Math!!

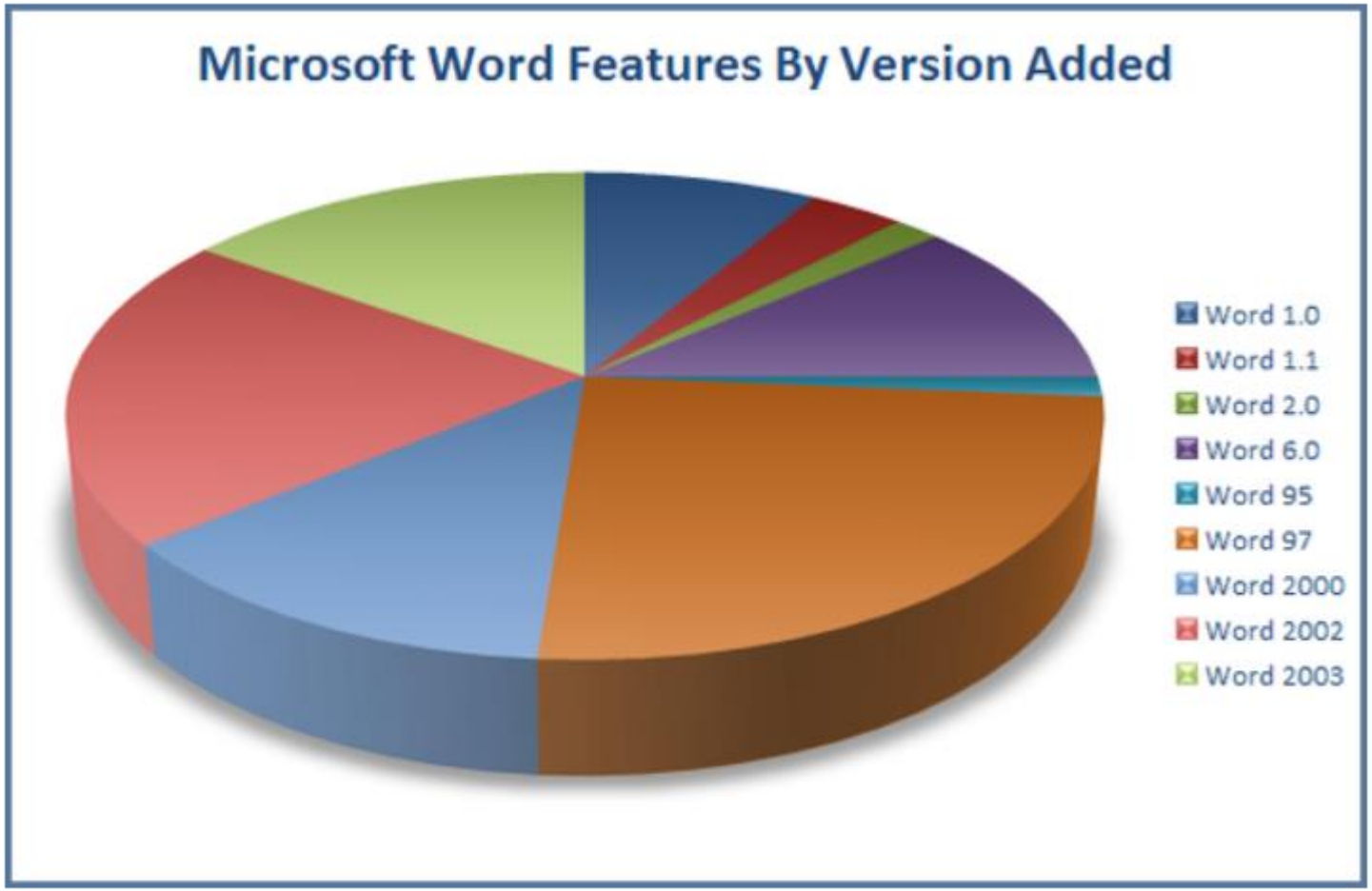

AMERICANS WHO HAVE TRIED MARIJUANA
CBS NEWS POLL
51% TODAY
43% LAST YEAR
34% 1997
Source: MOE +/- 4%
HIGH SUPPORT FOR LEGALIZING MARIJUANA
MORE THAN HALF OF AMERICANS SAY THEY'VE TRIED POT
LIVE
CBSN

Media and Maths!!

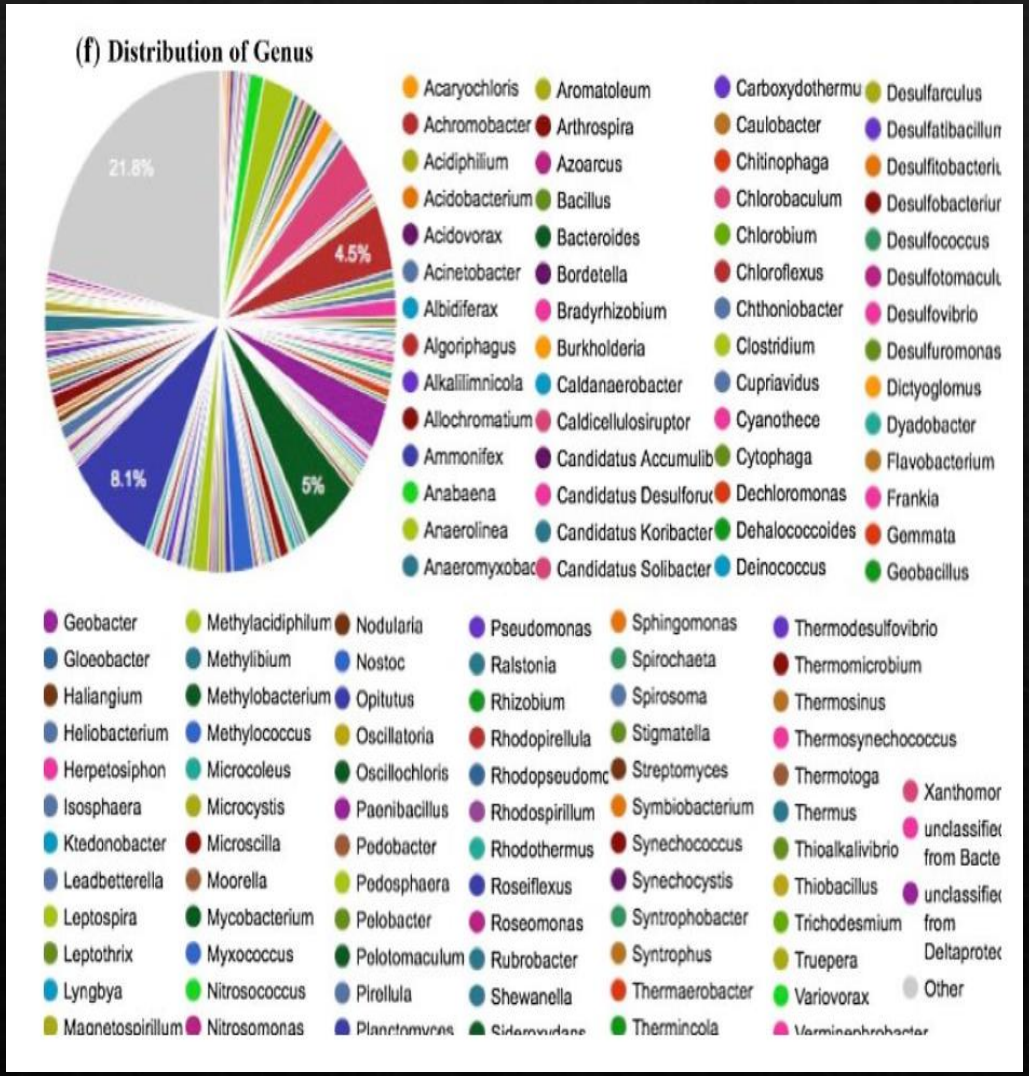


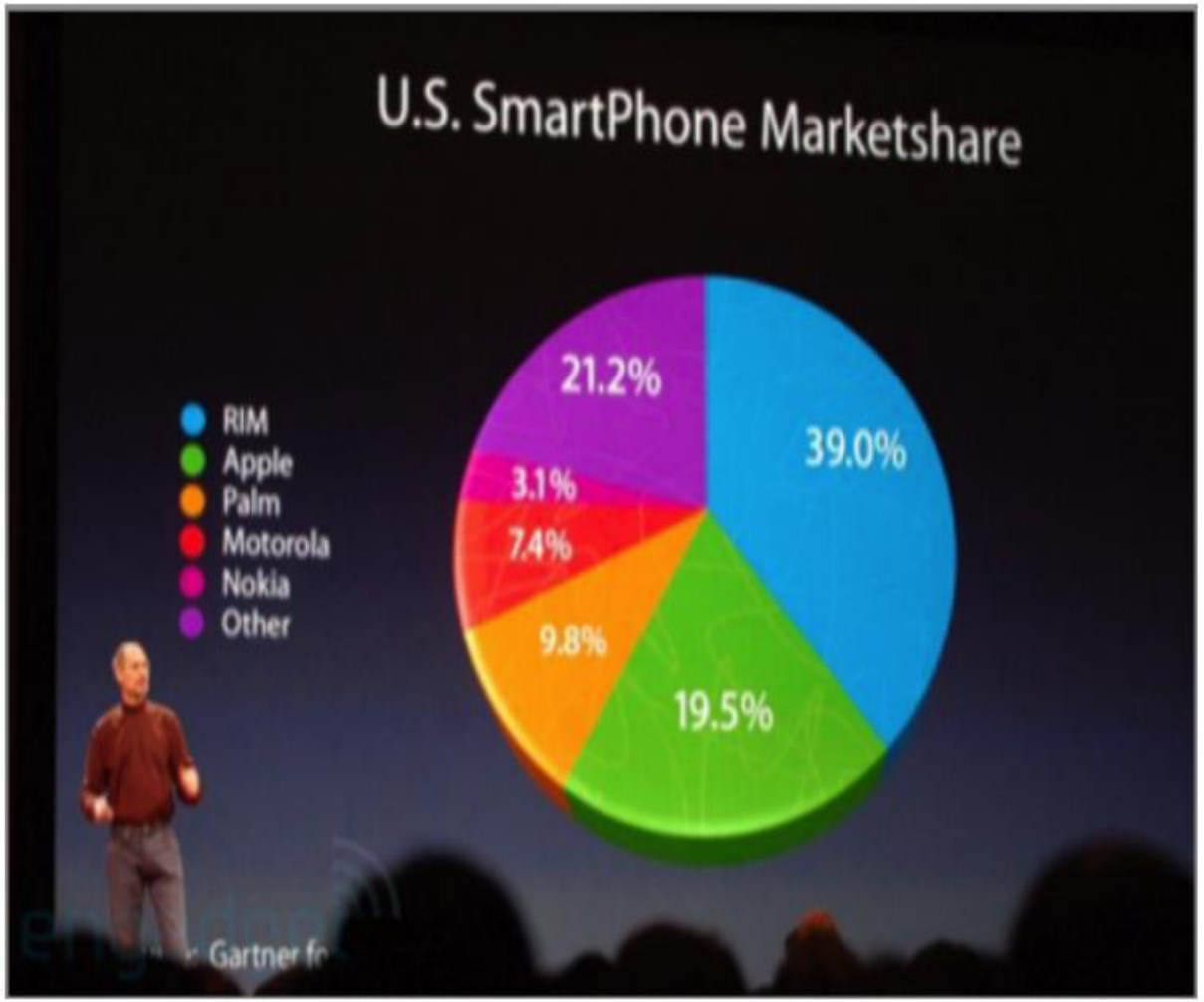


Nice Pie Chart eh!



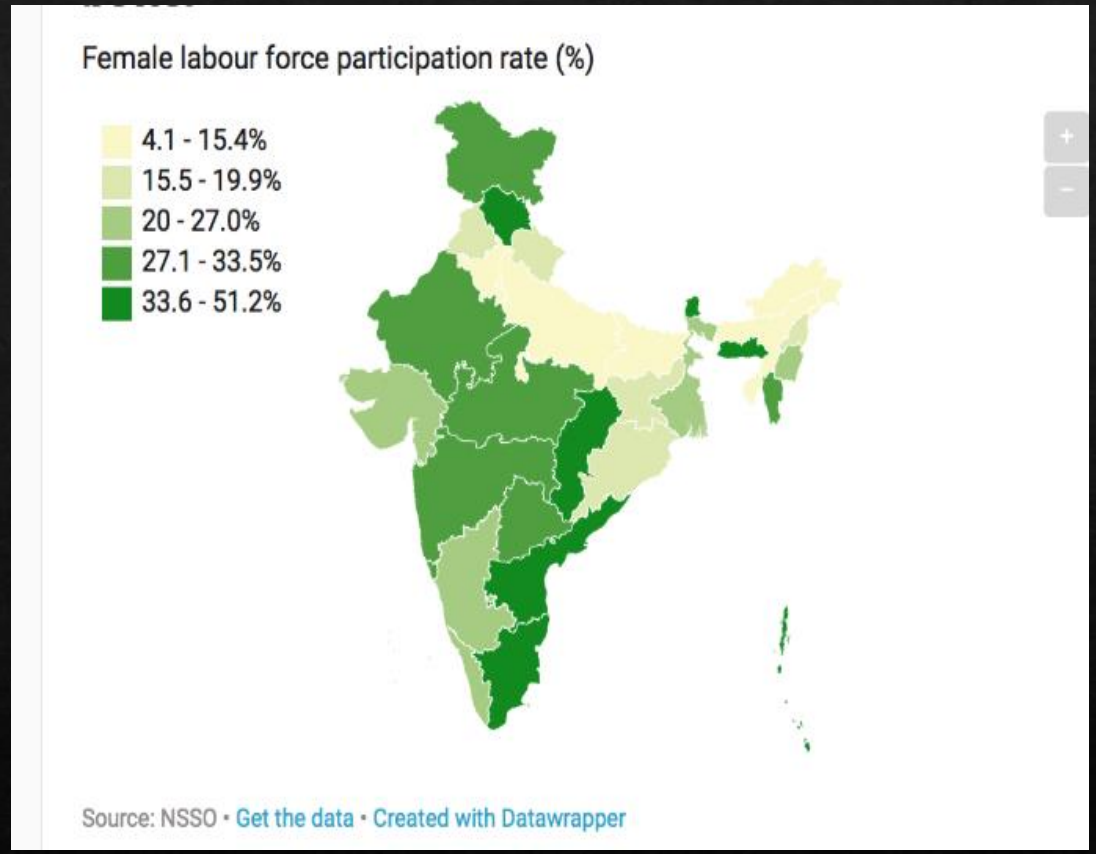
Pie Charts just get a bad rep



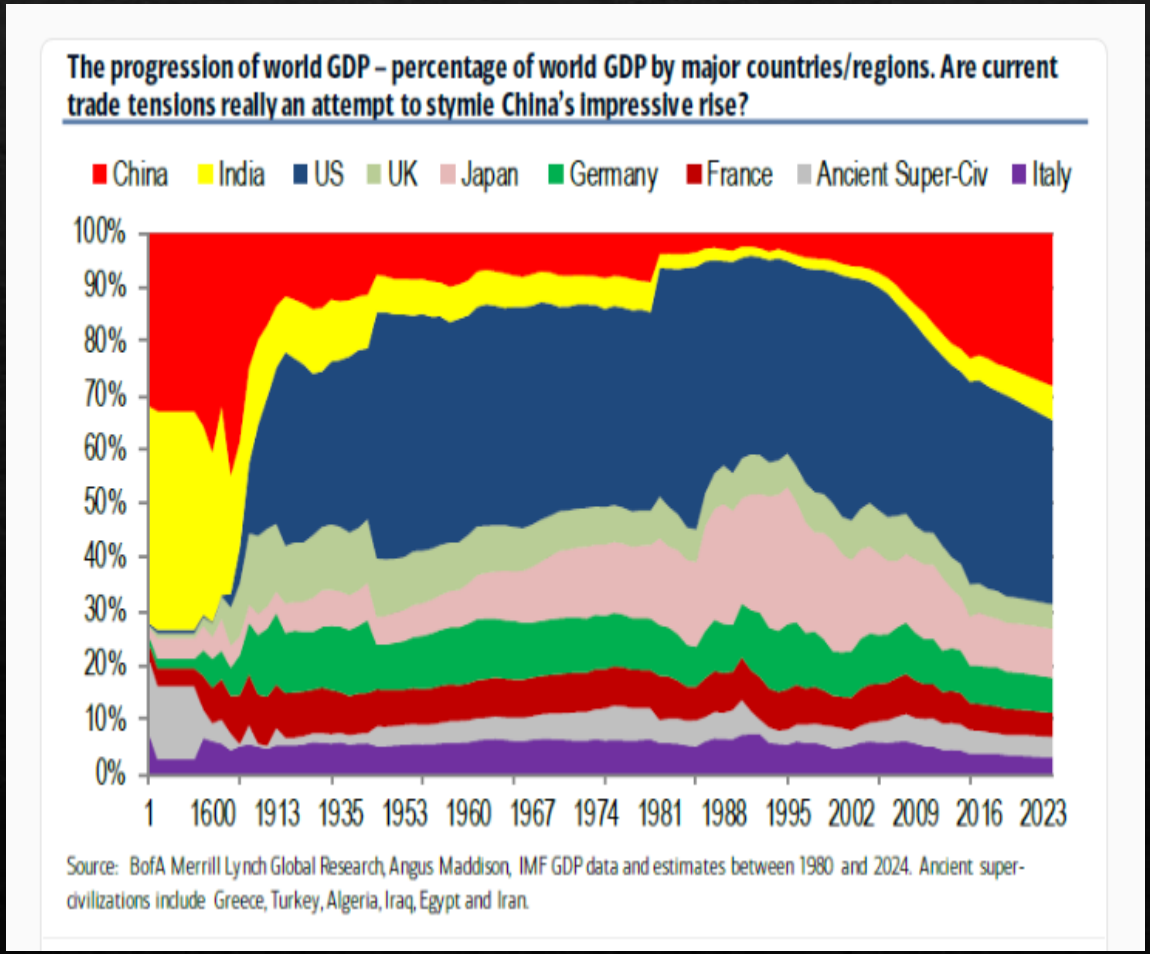


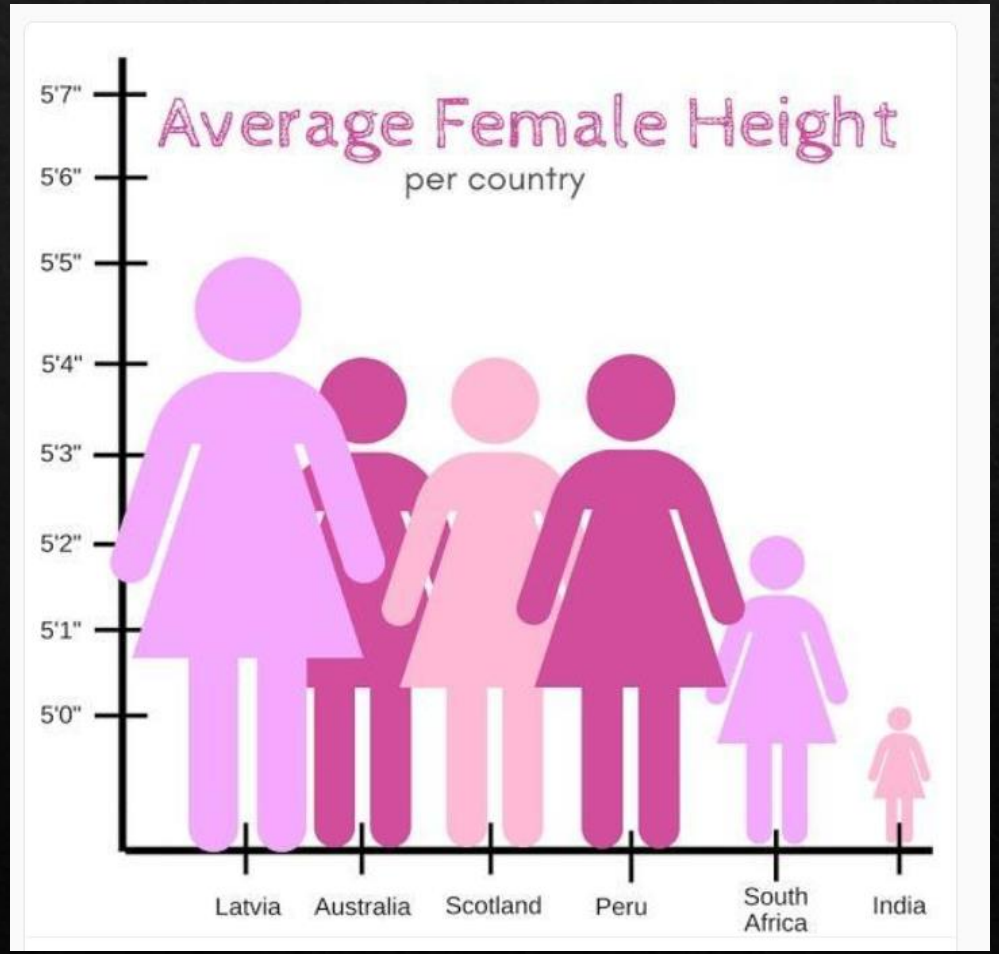
Some other Examples

Less obvious ones



Scale always matters!! Period!





Scale always matters!! Period!