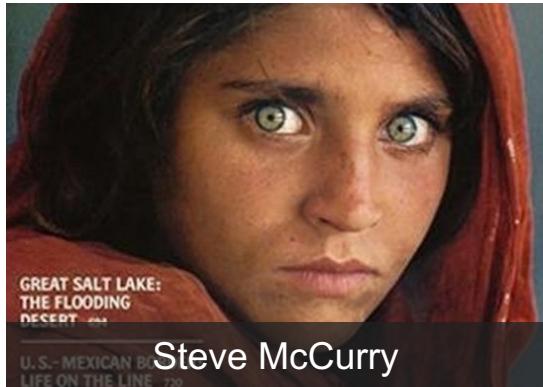


Introduction to Visualization

A picture is worth a thousand words

Visualization is the representation of **information** in a **graphical form**

Photographs



Street Art



Comics



A picture is worth a thousand words



A picture is worth a thousand ~~words~~ rows

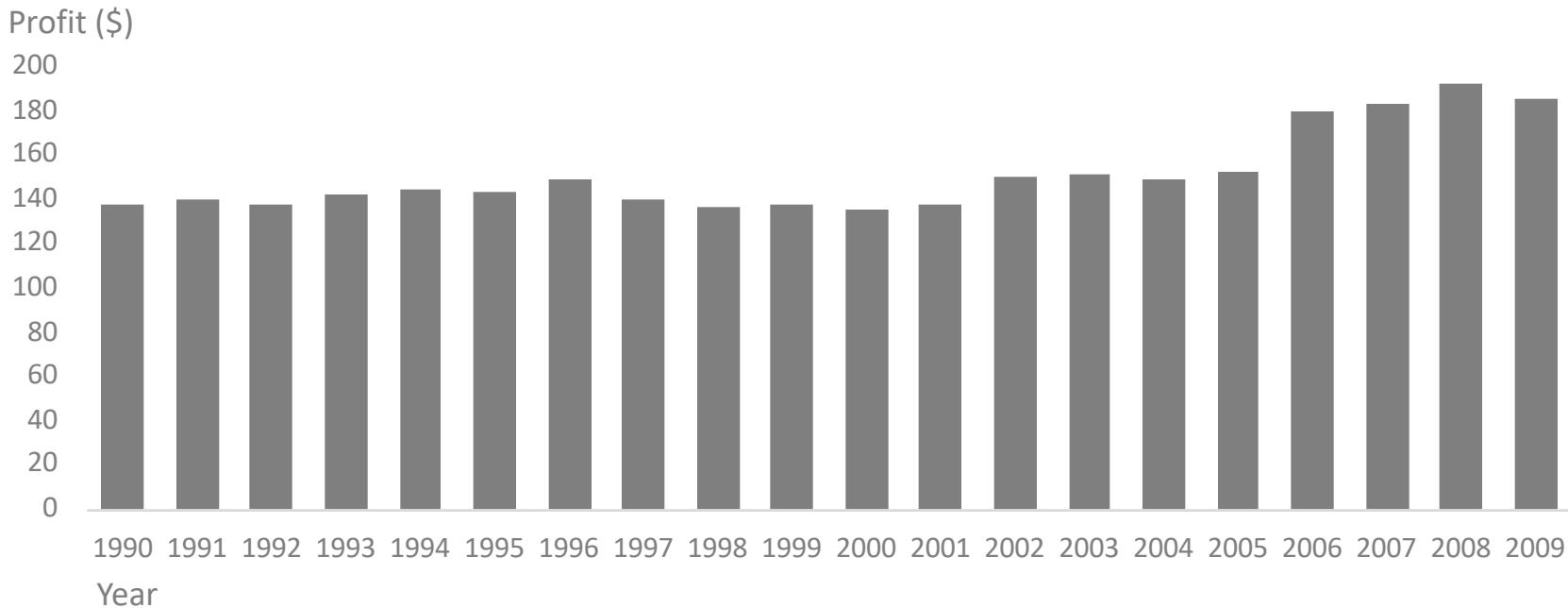
Which is the most profitable year?

Year	Profits (\$)
1990	138.1
1991	140.2
1992	137.8
1993	142.1
1994	144.5
1995	143.2
1996	148.9
1997	139.9
1998	136.5
1999	138.2
2000	135.5
2001	137.4
2002	150.4
2003	150.9
2004	149.3
2005	152.3
2006	180.1
2007	183.2
2008	192.5
2009	185.2



A picture is worth a thousand ~~words~~ rows

Which is the most profitable year?



“

The **purpose** of visualization
is **insight**, not pictures

”

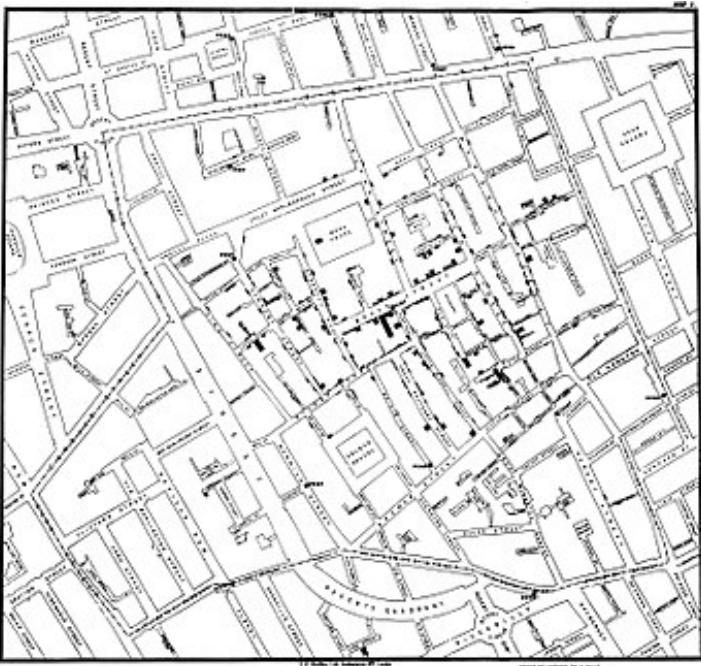
BEN SHNEIDERMAN



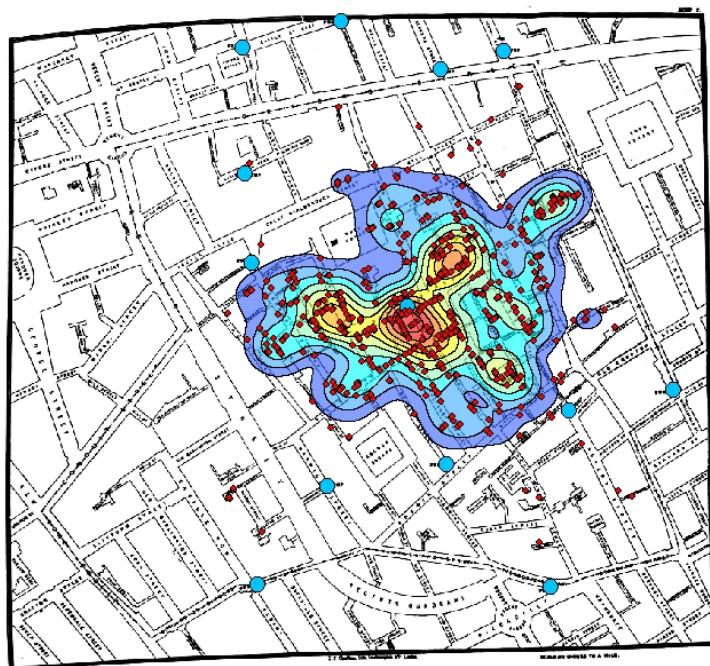


EXAMPLES

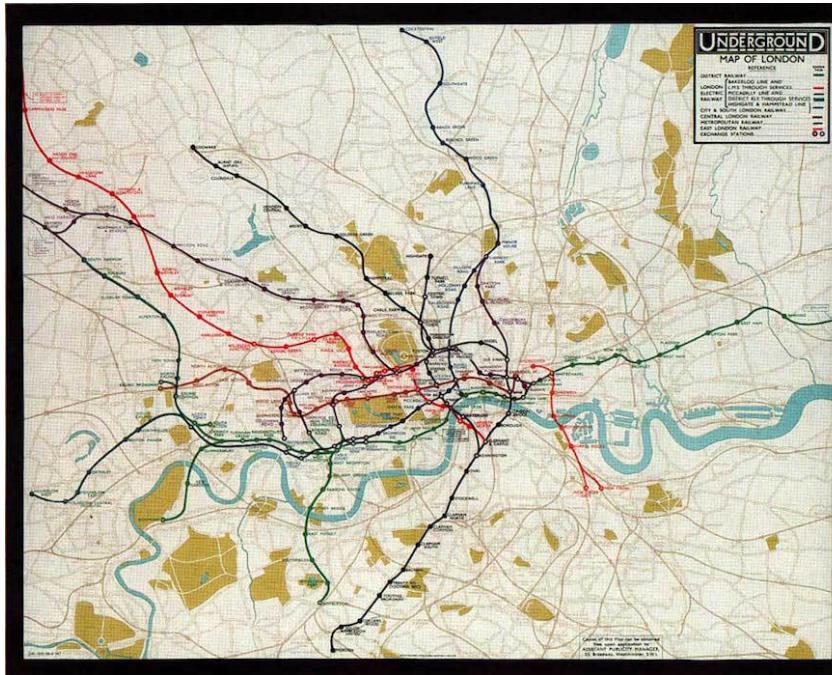
1854 Broad Street Cholera Outbreak



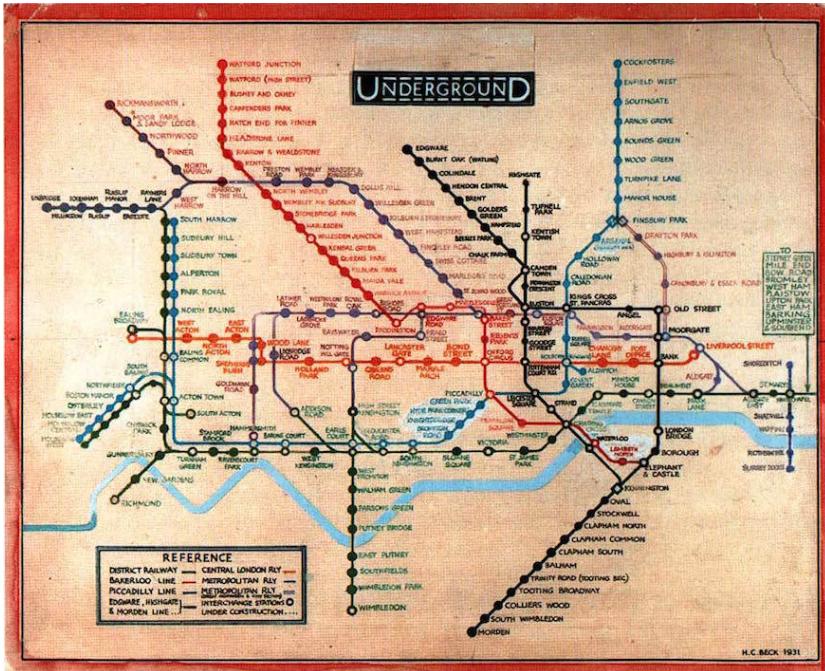
John Snow



London Tube Map - 1908



London Tube Map – Harry Beck 1931





TYPES OF DATA

Types of Data

- ▶ Qualitative
 - ▶ Ordinal
 - ▶ Nominal
- ▶ Quantitative
 - ▶ Continuous
 - ▶ Discrete



Qualitative Data

- ▶ Arise when the observations fall into distinct categories
 - Eye color: blue, green, brown
 - Exam results: A, B, C, D, F
 - Socio-economic status: low, middle, high
- ▶ Nominal: no natural order (e.g. eye color)
- ▶ Ordinal: order exists (e.g. exam results, socio-economic status)



Quantitative Data

- ▶ Arise when the observations are counts or measurements
 - Age
 - Weight
 - Temperature
- ▶ Discrete: measurements only take certain values (e.g. age)
- ▶ Continuous: measurements can take on any value within a range (e.g. weight, temperature)



Binning

- ▶ Continuous data can be converted to discrete or qualitative data
- “Low” $\text{Weight} < 100 \text{ lbs}$
- “Medium” $100 \text{ lbs} \leq \text{Weight} < 200 \text{ lbs}$
- “High” $200 \text{ lbs} \leq \text{Weight}$



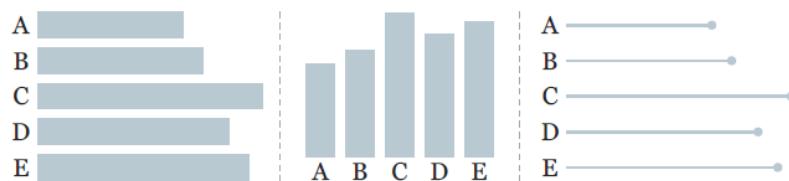


TYPES OF CHARTS

Visual Encodings

Transform numbers into a graphical representation

Length or height



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

Big Numbers

- ▶ Great to display simple information
- ▶ Tips:
 - ▶ It is good practice to present a relative metric

Page Views:

5,567

Daily Average: 3,625



Tables

- ▶ Conveys comparisons across categories
- ▶ Tips:
 - ▶ Too big is less effective

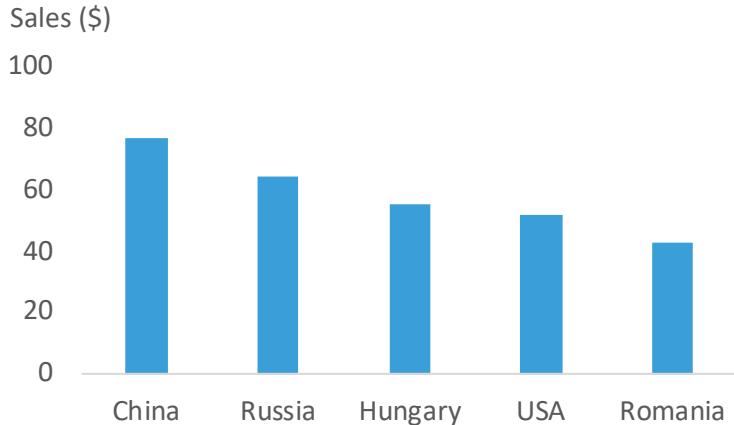
	Like chocolates	Do not like chocolates
Kids	23	5
Grown Ups	35	32



Bar Charts

- ▶ Uses length to represent a measure
- ▶ Compare different groups or categories
- ▶ Track changes over time

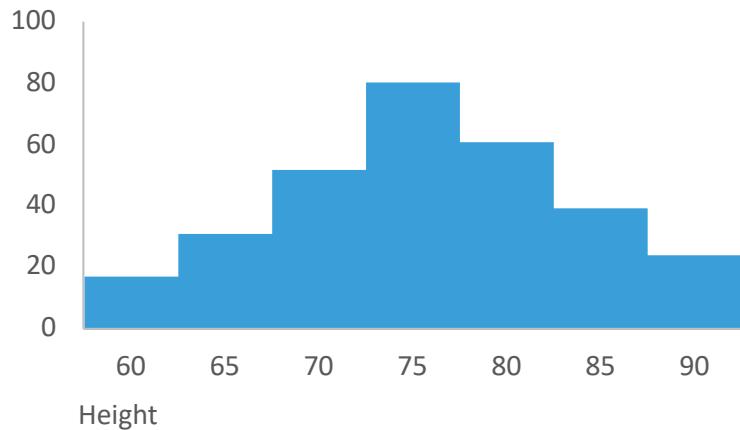
- ▶ Tips:
 - ▶ Sort based on something meaningful (e.g. length, alphabetical)
 - ▶ Start scale at 0
 - ▶ Horizontal, vertical and stacked bars are commonly used



Histograms

- ▶ Represents the distribution of data

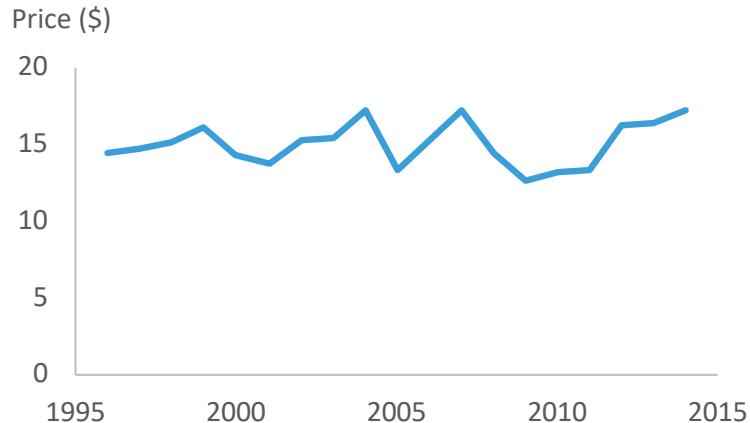
- ▶ Tips:
 - ▶ Range of values must be binned
 - ▶ Can be normalized (sum of all frequencies equals 1)
 - ▶ Estimates the probability functions
 - ▶ It is a variation of a bar chart



Line Charts

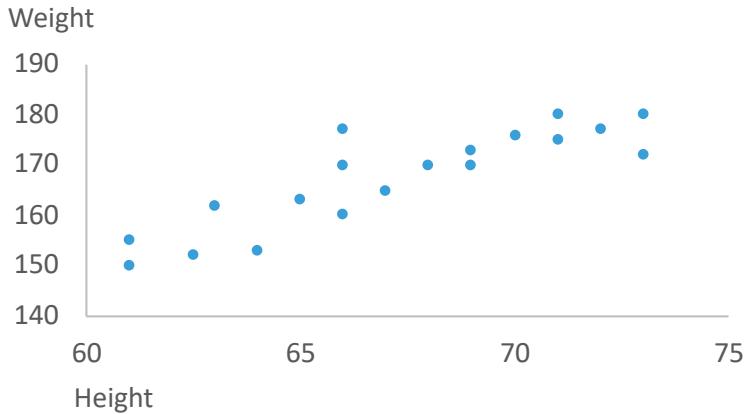
- ▶ Track changes over time
- ▶ Allows for potential forecasting
- ▶ Height and slope lets us see trends

- ▶ Tips:
 - ▶ Ideal for continuous data
 - ▶ Time is represented in the x-axis
 - ▶ Do not use if x-axis is non-numeric



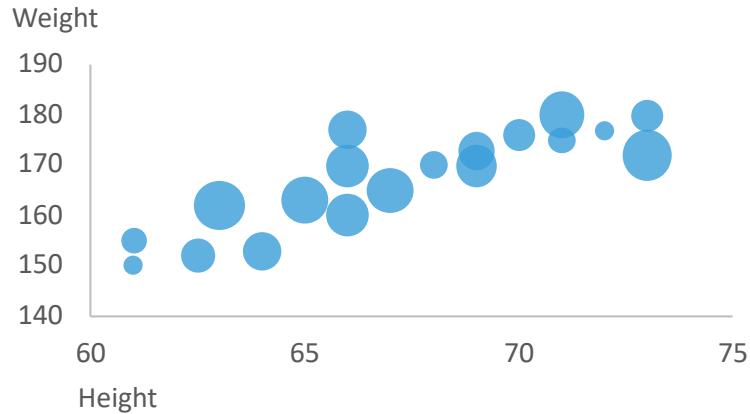
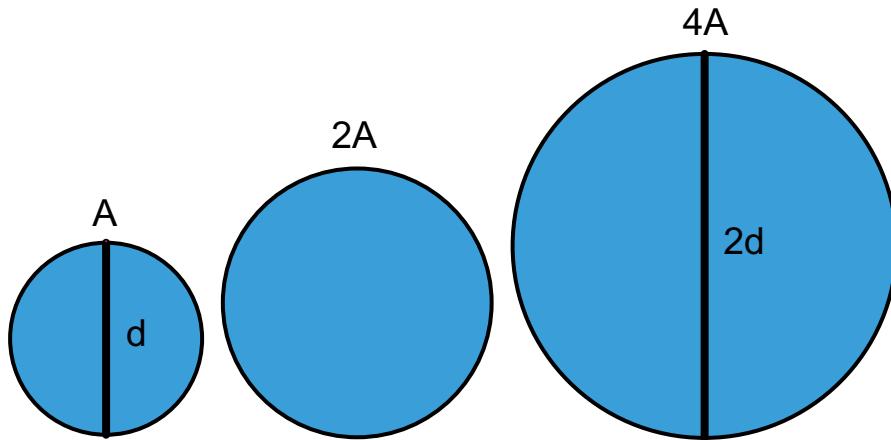
Scatter Plots

- ▶ Displays relationship between two measures
- ▶ Tips:
 - ▶ Best for continuous data
 - ▶ Do not use for qualitative data



Bubble Charts

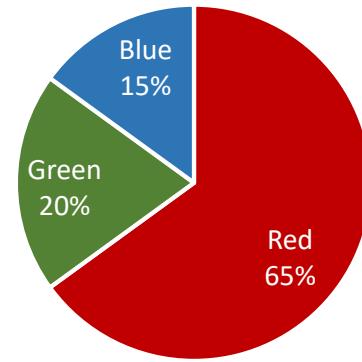
- ▶ Used for scatter plots with 3 variables
- ▶ Tips:
 - ▶ Size causes confusion



Pie Charts

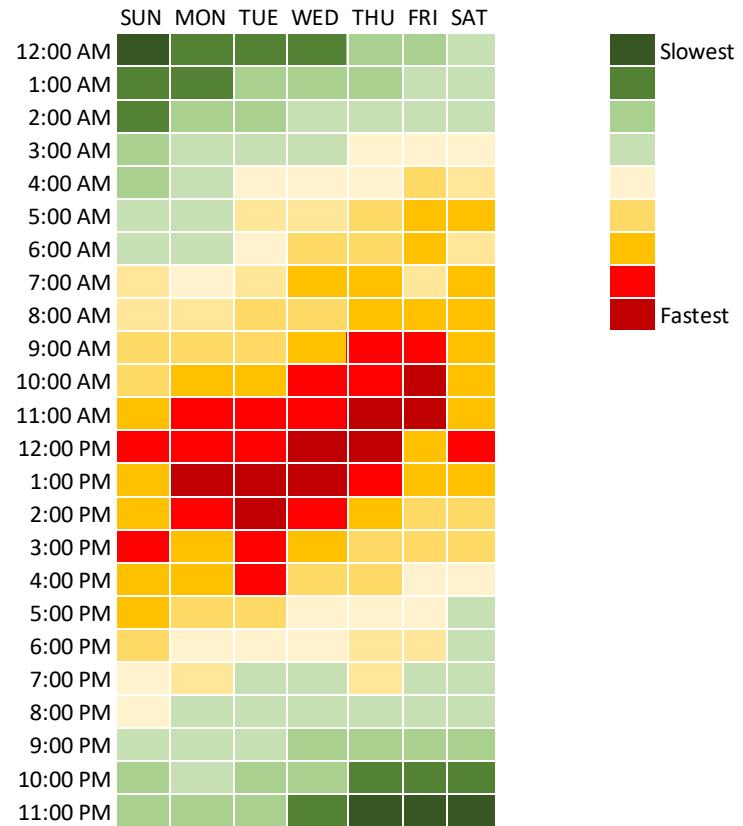
- ▶ Compare parts of a whole

- ▶ Tips:
 - ▶ All parts must add up to 100%
 - ▶ Best for categorical data
 - ▶ Order clockwise by size or any meaningful order
 - ▶ Hard to read when:
 - ▶ Components have similar size
 - ▶ More than 3 parts



Heat Maps

- ▶ Use color to display trends or emphasize information



Presentation Types

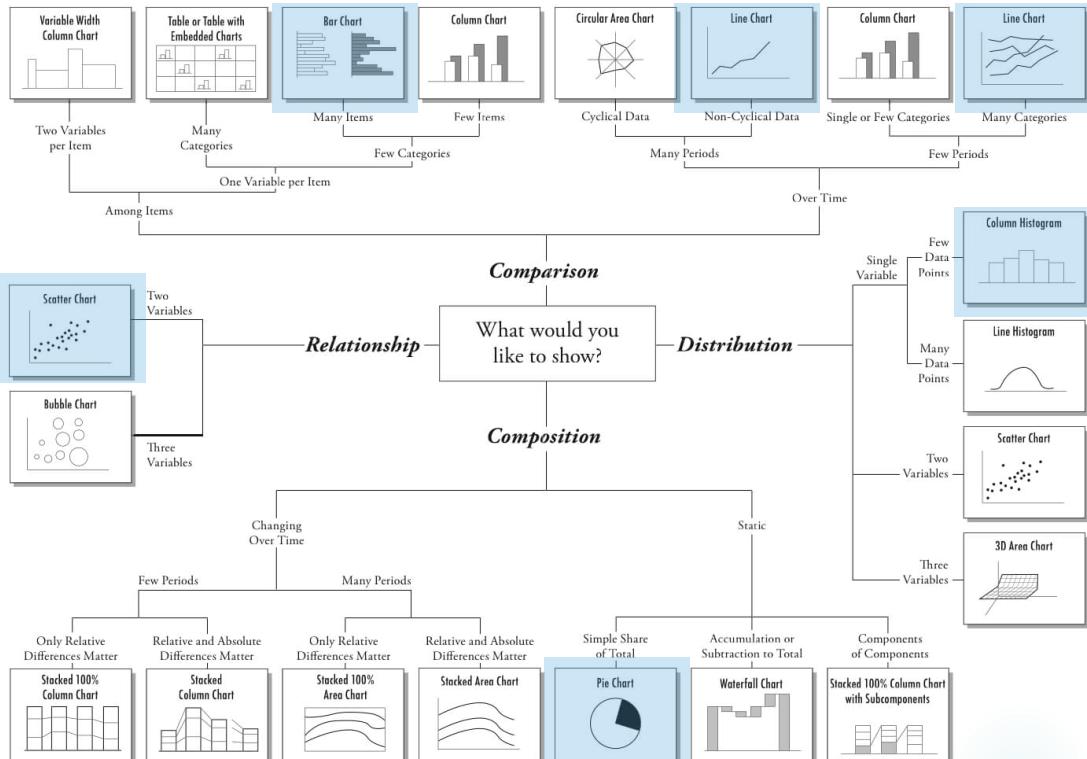
- ▶ Comparison
 - ▶ How do our sales compare to the last few years?
- ▶ Composition
 - ▶ What percent of our users like football?
- ▶ Relationship
 - ▶ How does the stock price change with the weather?
- ▶ Distribution
 - ▶ What is the typical age of our customers?



Presentation Types

Chart Suggestions—A Thought-Starter

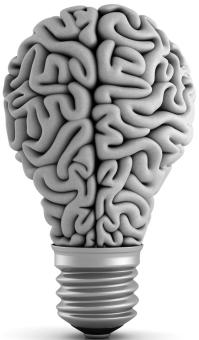
www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com



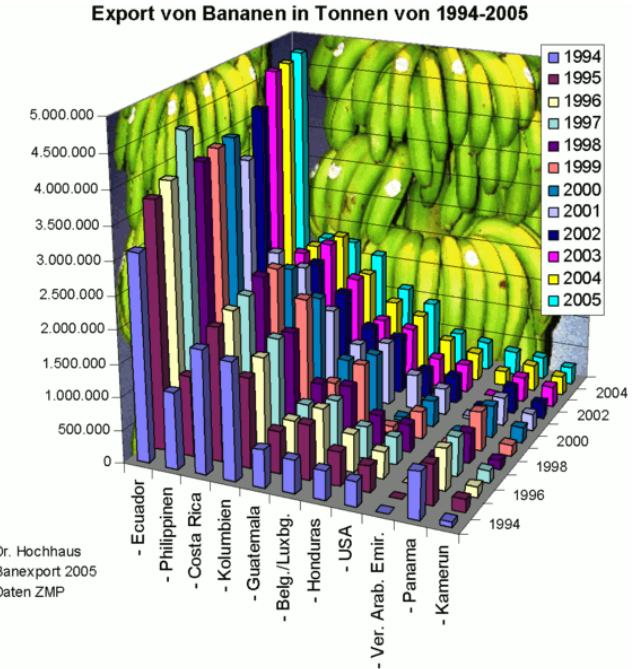


PROBLEM STATEMENT

Problem Statement



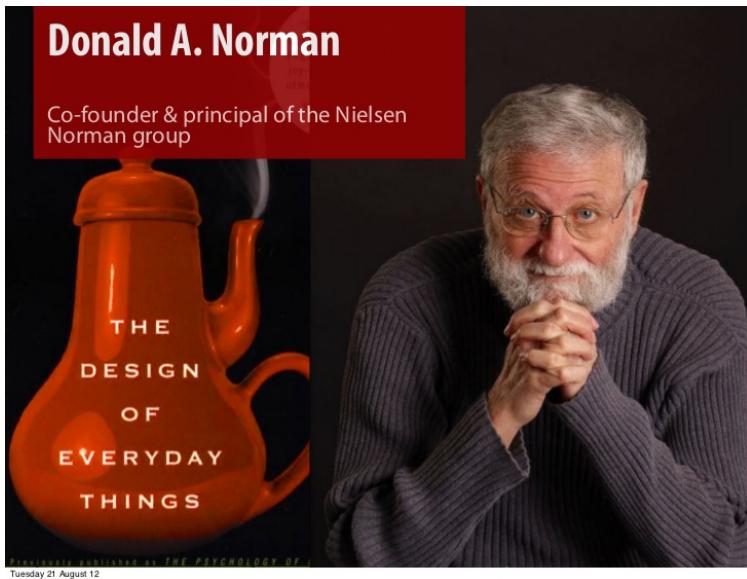
Problem Statement





DESIGN

Don Norman - 1988



Norman Door

- ▶ A door where the design tells you to do the opposite of what you're actually supposed to do
- ▶ A door that gives the wrong signal and needs a sign to correct it



Discoverability

- ▶ The ability to discover what operations one can do





TEXT

Text Orientation

Vertical text orientation requires
mental effort!



Text Orientation

Vertical text orientation requires
mental effort!

s
o
d
o
e
s
t
h
i
s
!

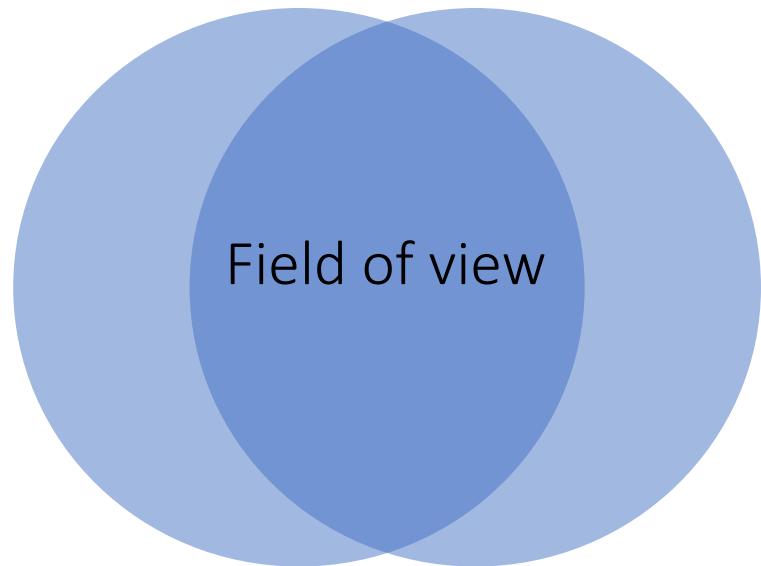


Text Orientation

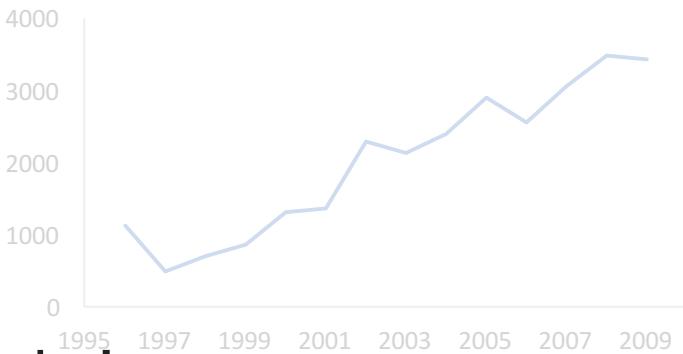
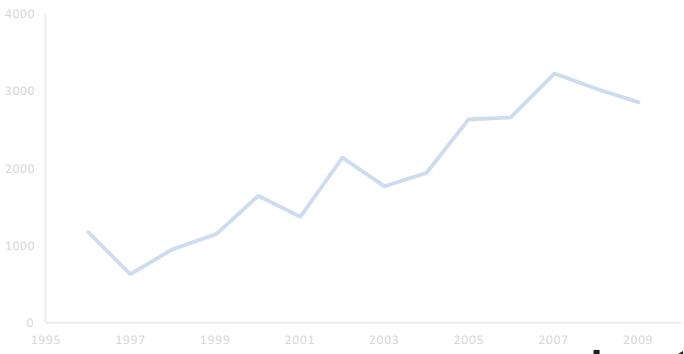
Vertical text orientation requires
mental effort!

s
o
d
o
e
s
t
h
i
s
!

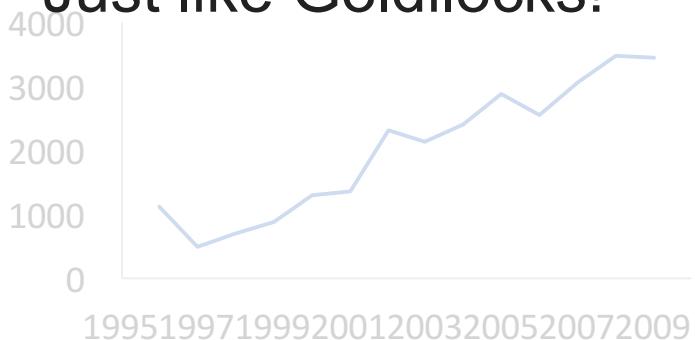
Horizontal text is easier to read!



Font Size



Just like Goldilocks!





COLOR

Color Palette

- ▶ Max 5 colors



<http://tools.medialab.sciences-po.fr/iwanthue/>

<http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3>

<http://www.color-hex.com/color-palettes/>

<https://coolors.co/>

<https://www.canva.com/color-palette/>

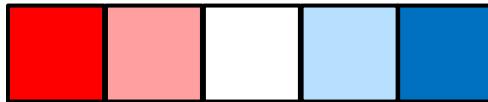


Color

- ▶ Sequential



- ▶ Diverging



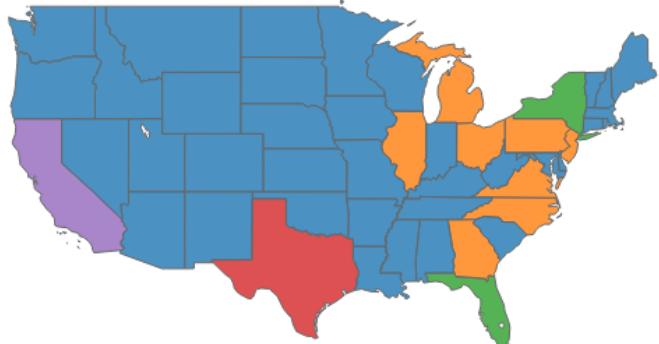
- ▶ Categorical



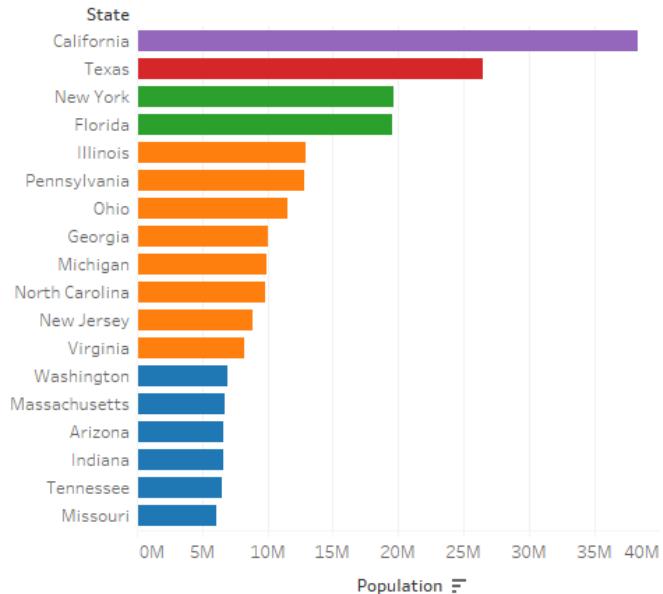
- ▶ Highlight



Color consistency across charts

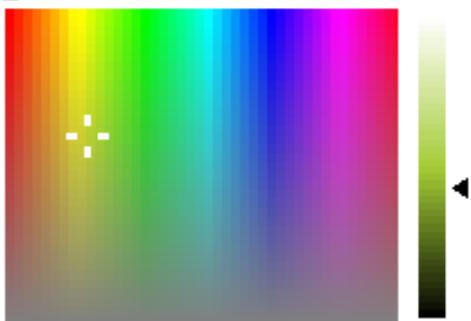


© OpenStreetMap contributors



Color theory

Colors:



R: 141
G: 175
B: 45

R: 141
G: 45
B: 175

R: 211
G: 115
B: 9

R: 0
G: 0
B: 0

R: 255
G: 255
B: 255

R: 110
G: 110
B: 110



RGB to grayscale

$$\text{Grayscale} = 0.21 \times R + 0.72 \times G + 0.07 \times B$$

RGB Image



Grayscale Image

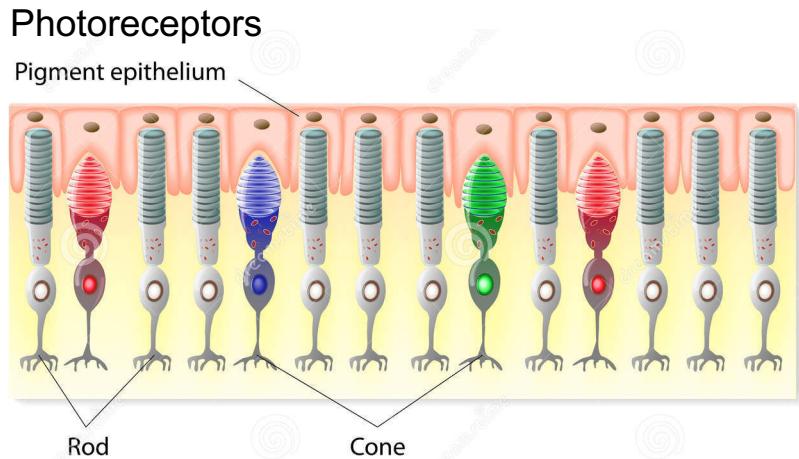
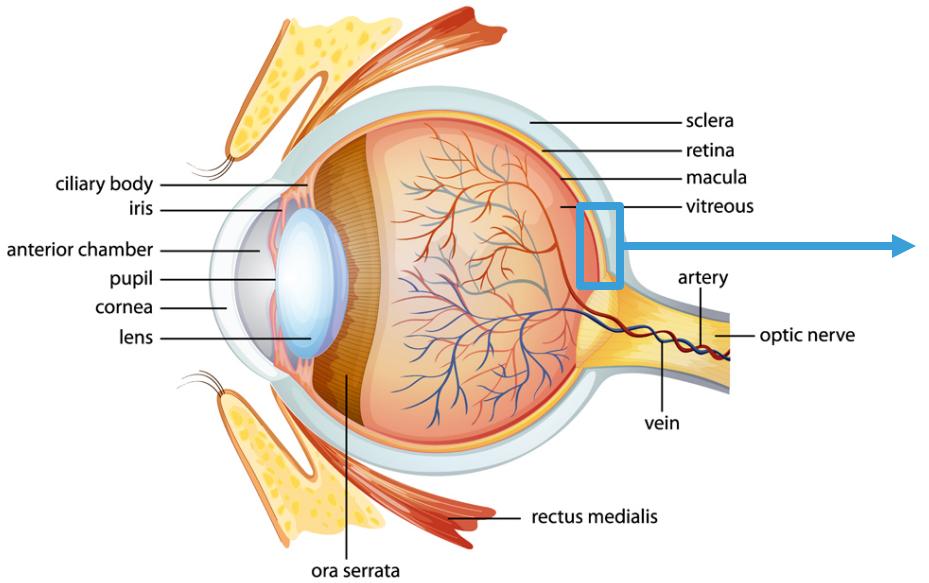


RGB to grayscale

$$\text{Grayscale} = 0.21 \times R + 0.72 \times G + 0.07 \times B$$

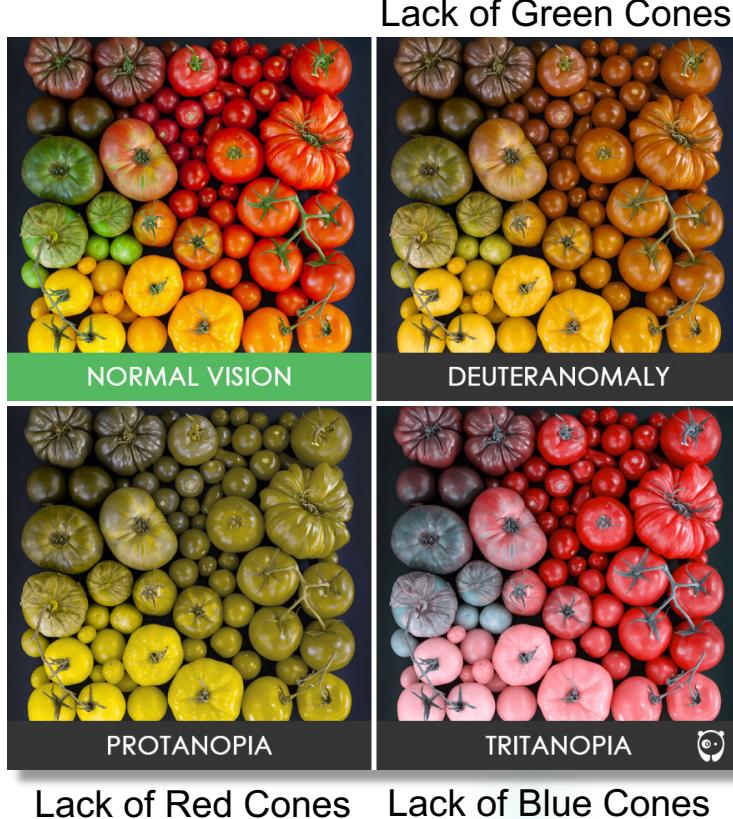


Anatomy of the human eye



Color blindness

0.5% 8.0%



Color blindness

Instead of **red** and **green**
use **blue** and **orange**

Resources:

<http://asada.tukusi.ne.jp/cvsimulator/e/>

<https://accessgarage.wordpress.com/2013/02/09/458/>





HIGHLIGHT

Discriminate background



Florida Bark Mantis





TEST: How many threes?

How many three's?

5498731840

4893128612

2634085106

1592059852

3854634876



How many three's?

54987**3**1840

489**3**128612

26**3**4085106

1592059852

38546**3**4876



Balance background and foreground

ACVLSIGBSLWUHKAJSLHV

ACVLSIGBSLWUHKAJSLHV

ACVLSIGBSLWUHKAJSLHV



Pre-attentive attributes

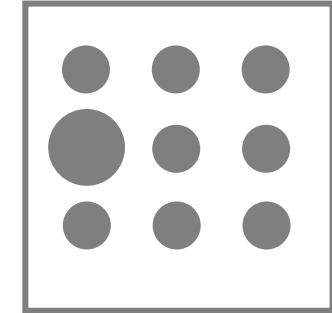
Orientation



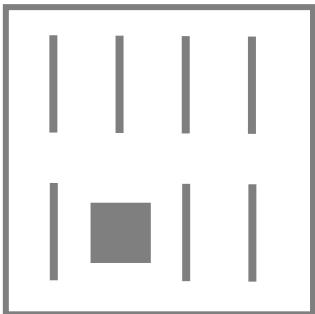
Width



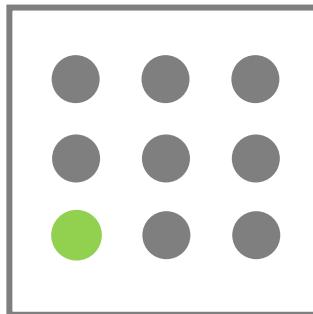
Size



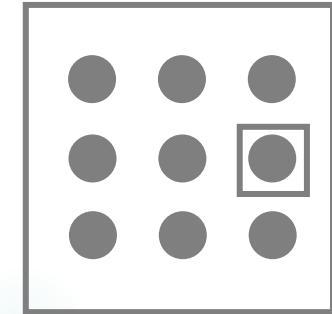
Shape



Color



Enclosure

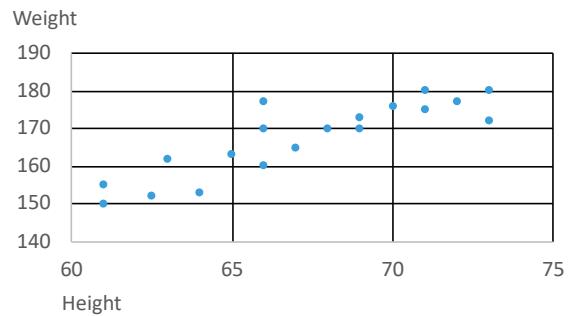


What is the main point?



Gridlines

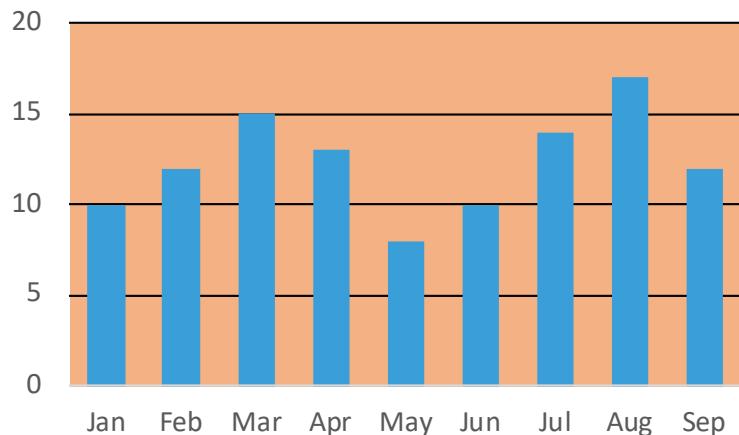
- ▶ Move to background or completely remove



Increase Data-Ink Ratio

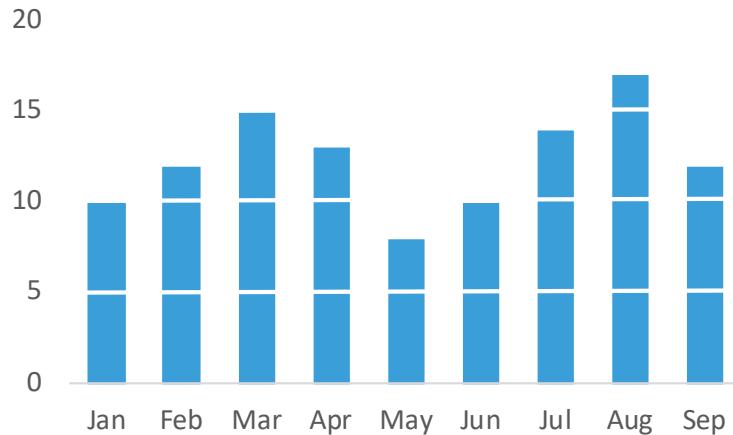
- ▶ Unnecessary:

- ▶ Background color
- ▶ Gridlines



“Above all else show the data”

Tufte - 1983



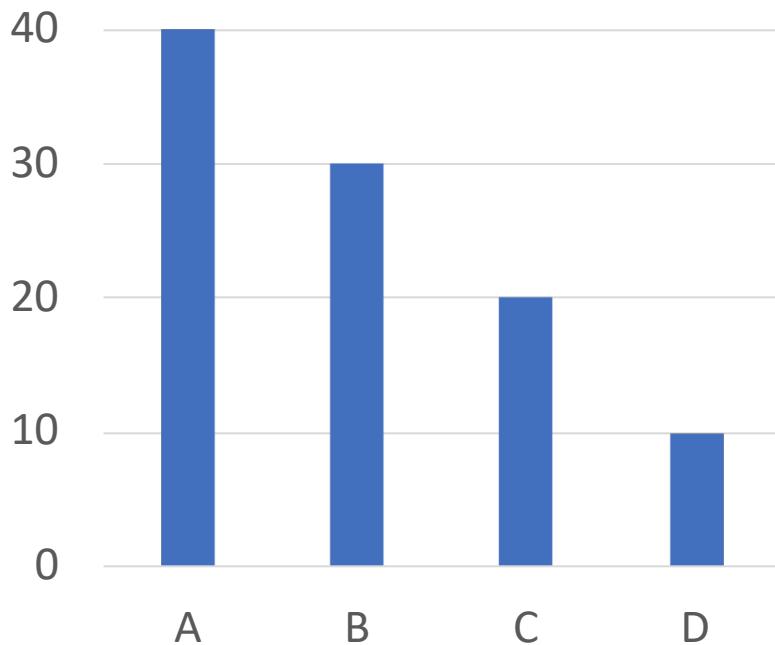
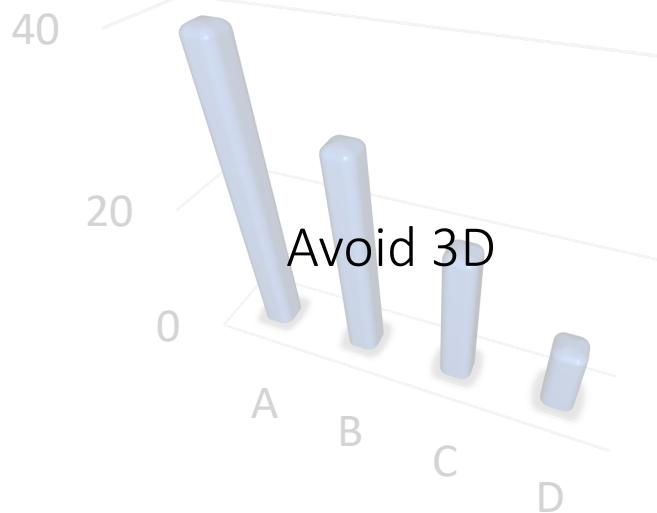
30 second rule



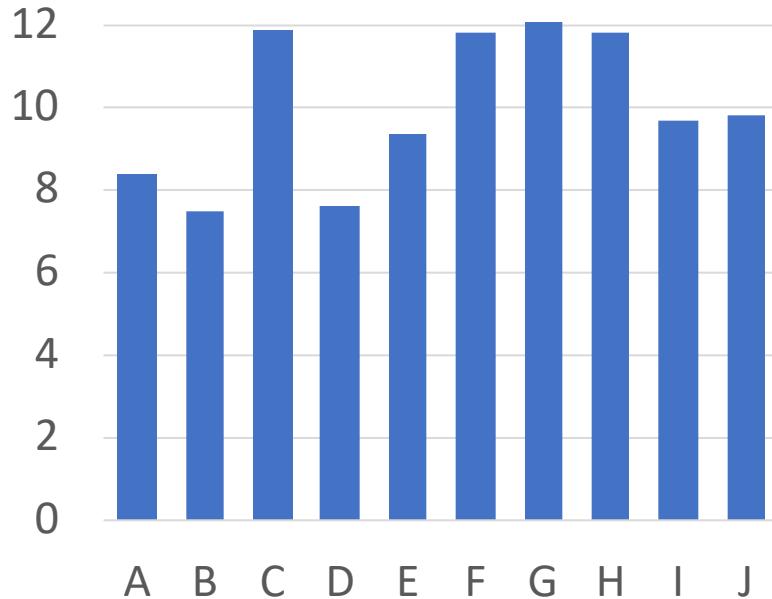


COMMON MISTAKES

What is the value of “C”?

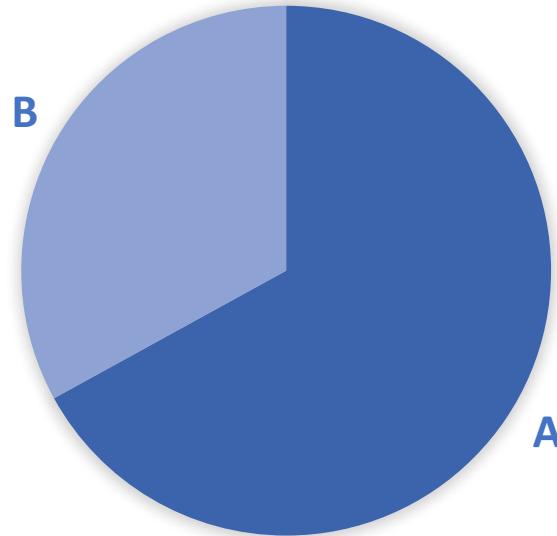


Which is the biggest slice?

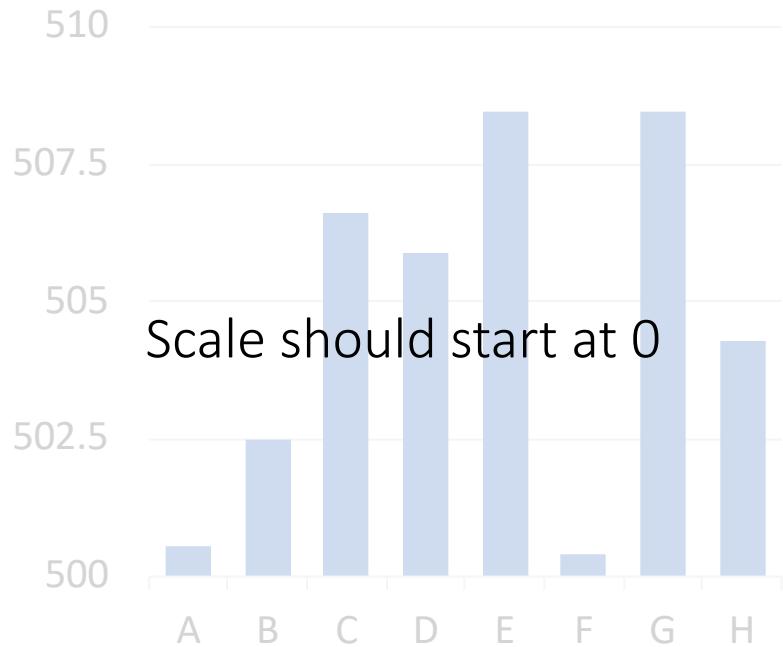


Sometimes pie charts work well

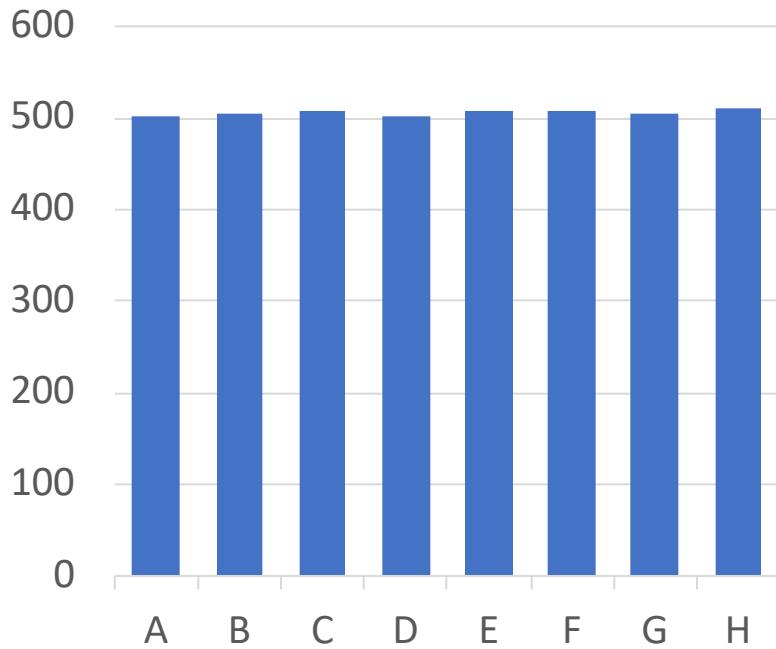
2 or 3 slices



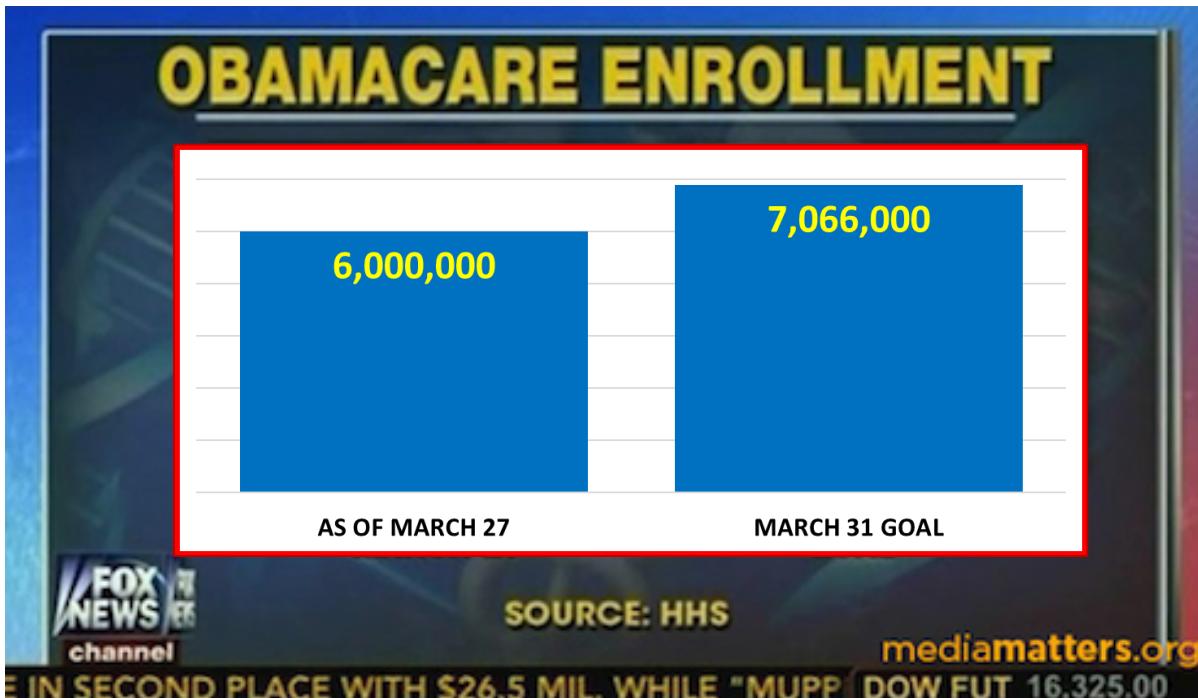
What is the conclusion?



Scale should start at 0



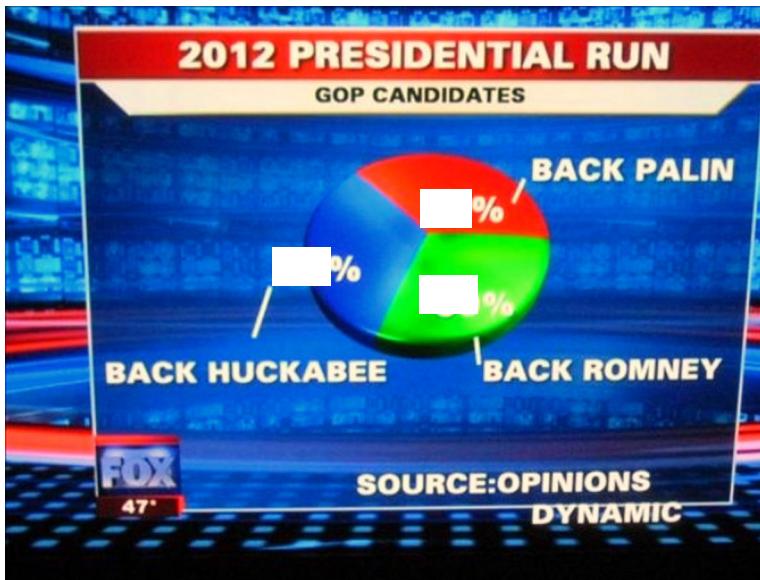
Example



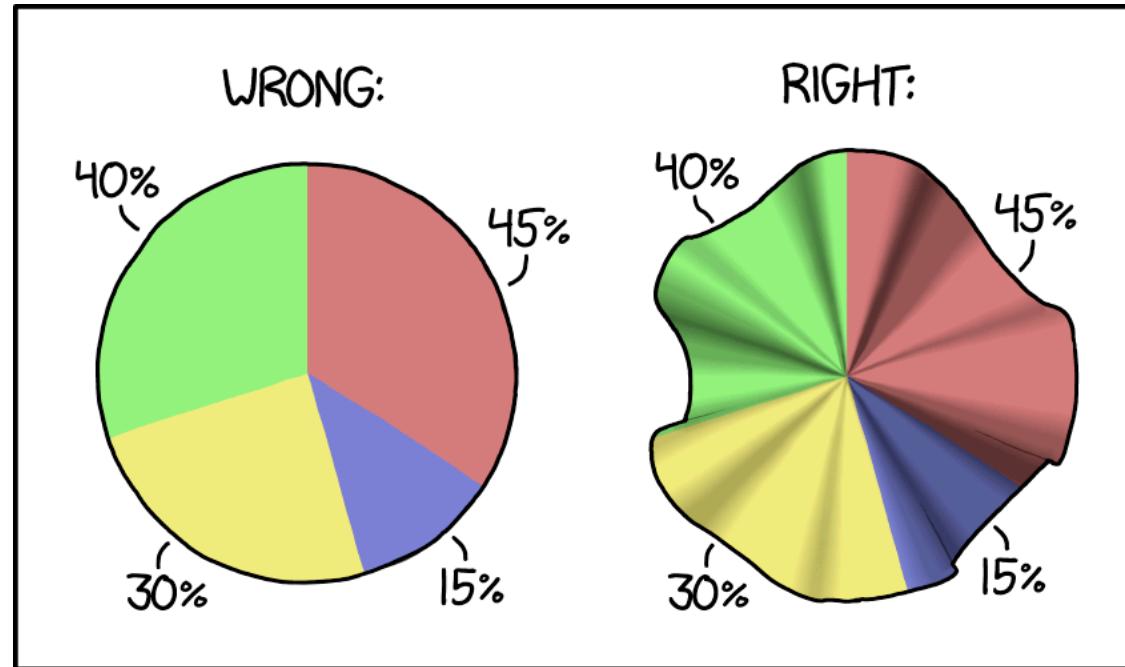
Example



Example



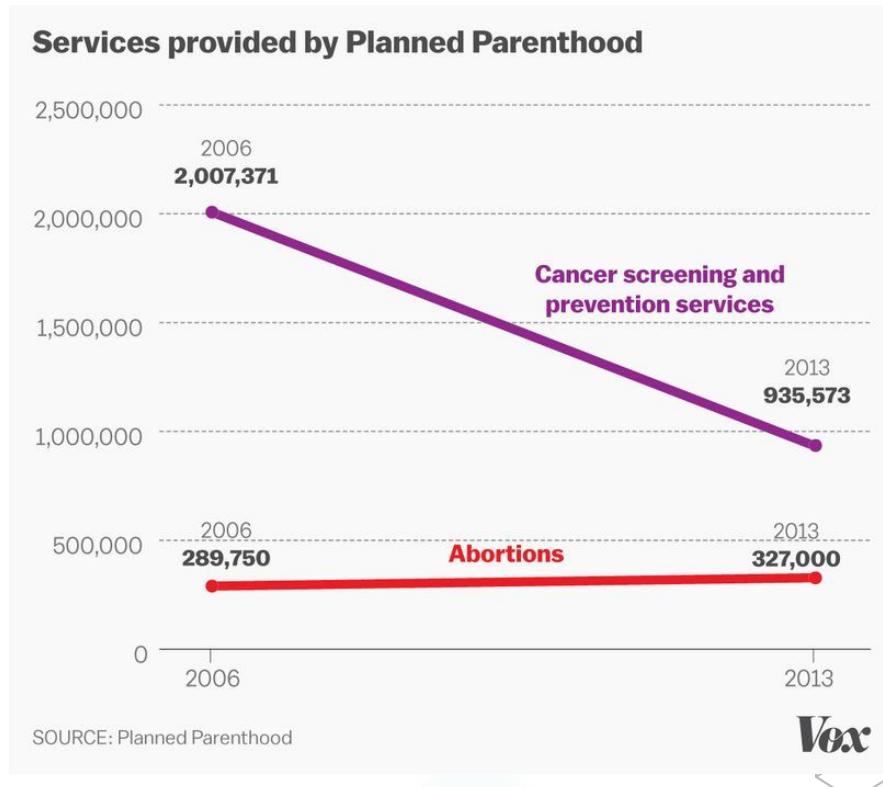
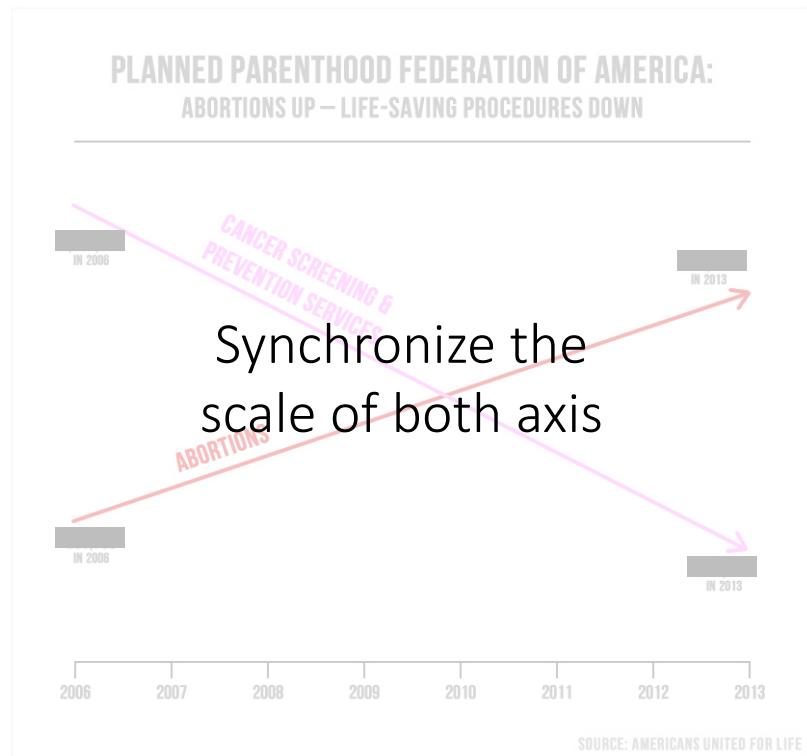
XKCD



HOW TO MAKE A PIE CHART IF YOUR
PERCENTAGES DON'T ADD UP TO 100



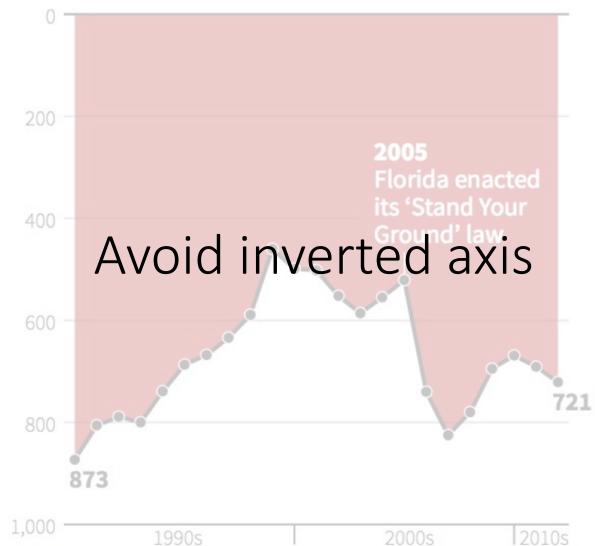
Example



Example

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS





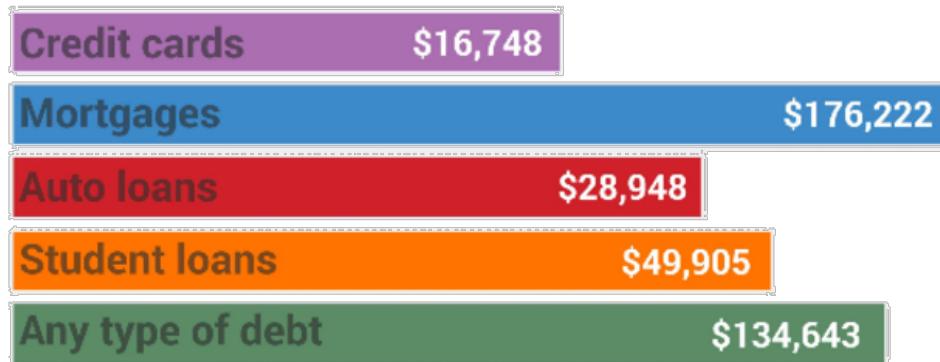
EXAMPLES OF BAD VISUALIZATIONS

What is wrong with this visual?

- ▶ Bar length
- ▶ Colors

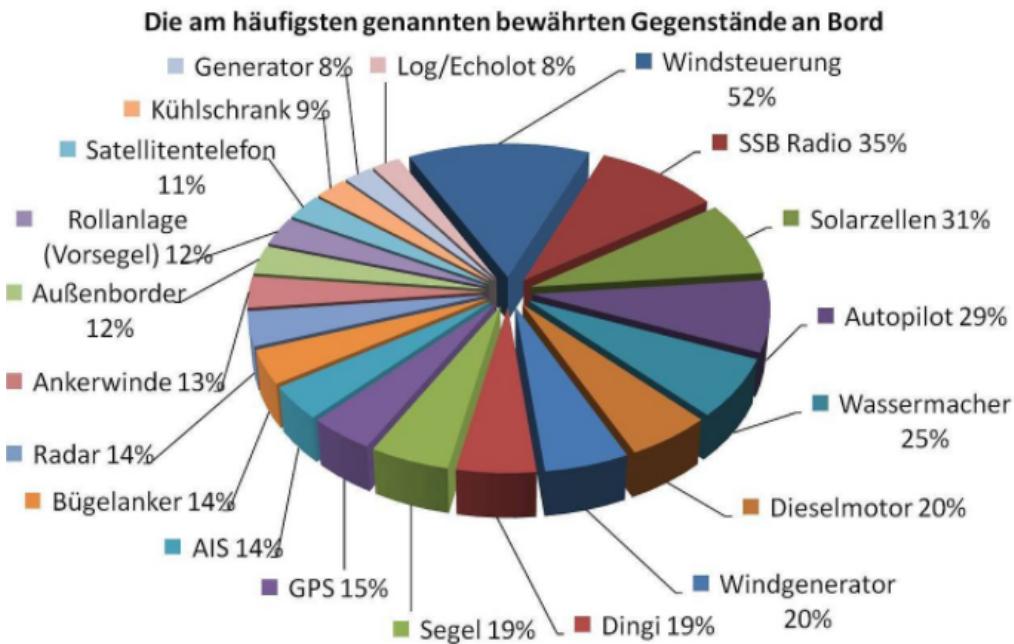
Types of debt

The total owed by the average U.S. household, by debt type.



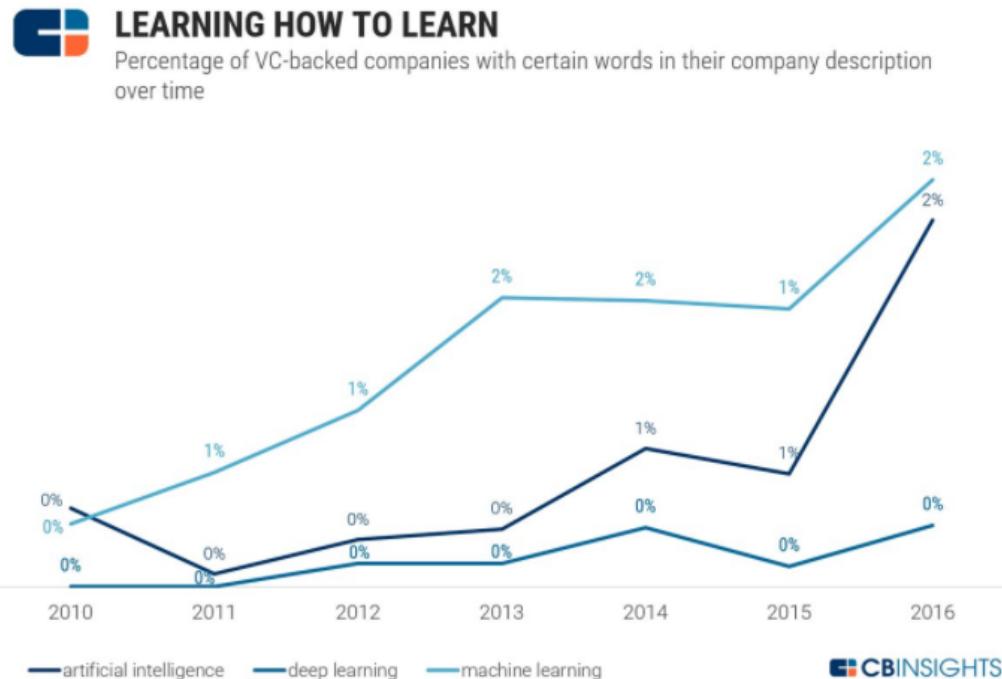
What is wrong with this visual?

- ▶ 3D
- ▶ Too many slices
- ▶ >100%



What is wrong with this visual?

- Decimals

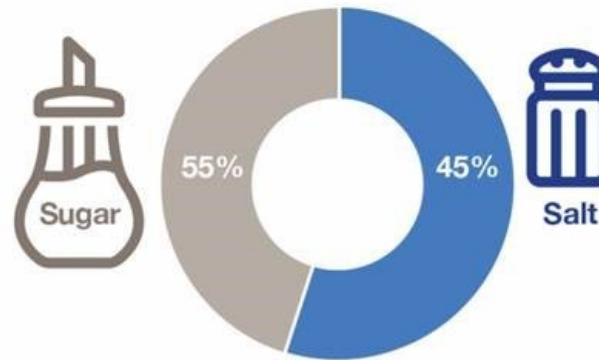


What is wrong with this visual?

- ▶ Slice size

Last Week's Results

Which of these would you have
a harder time giving up, salt or sugar?



Participant Comment

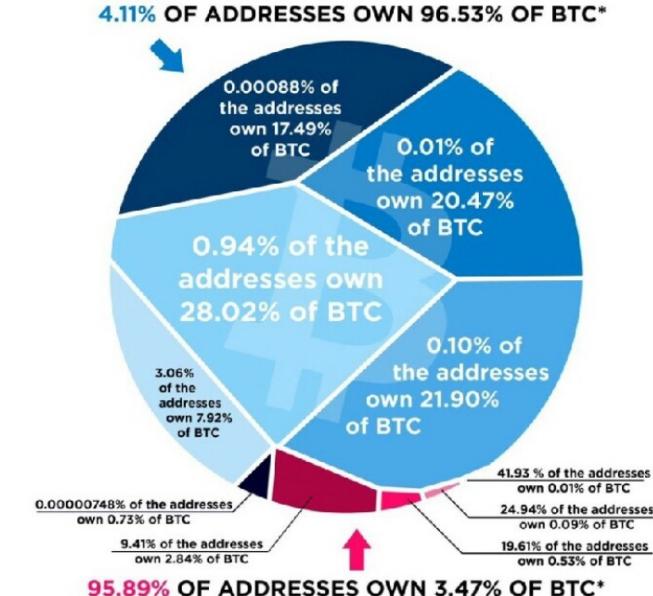
"Salt. Sugar is difficult to give up when you are eating it but after a few days, your body adapts and loses any cravings. Salt is a different story."



What is wrong with this visual?

- ▶ How do you read?

The  **Bitcoin Wealth Distribution**



* Data as of September 12th, 2017

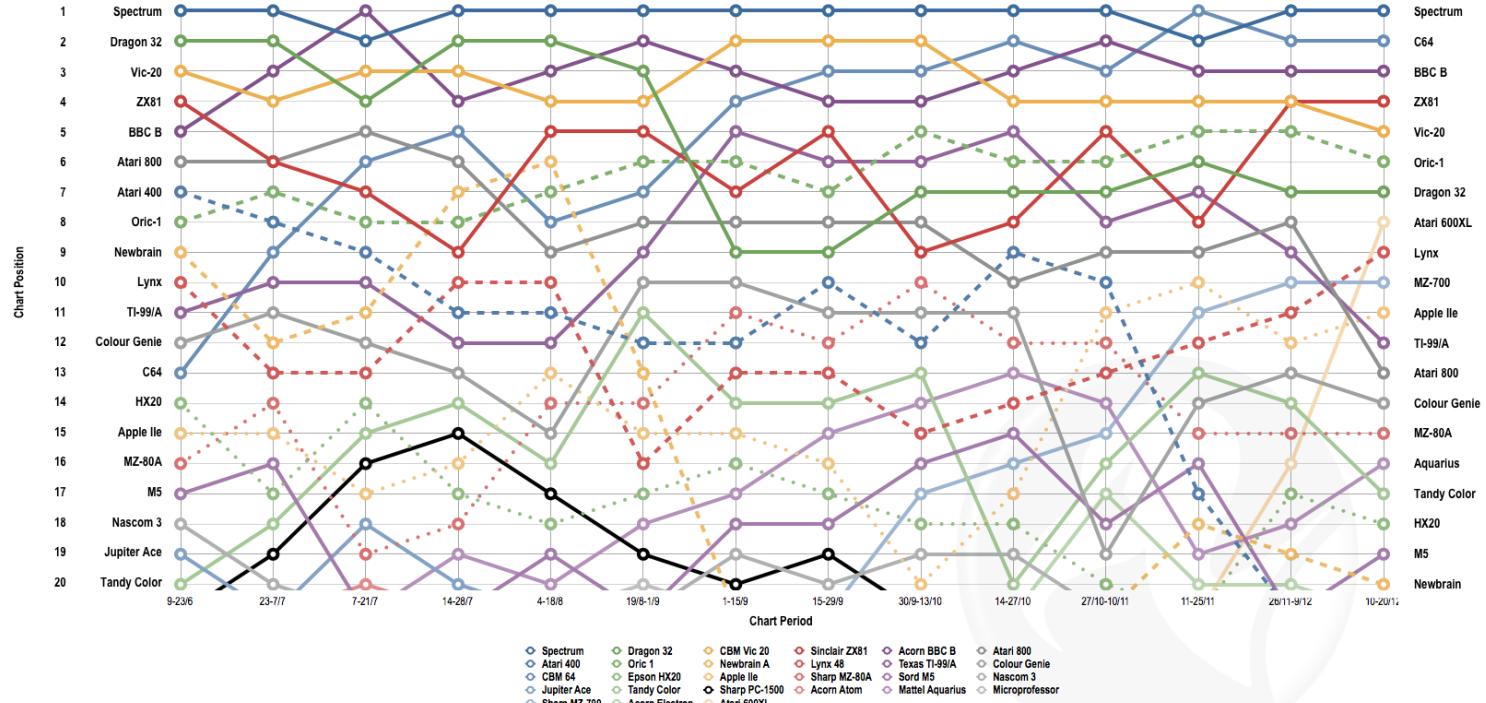
Article and Sources:

<https://howmuch.net/articles/bitcoin-wealth-distribution>

<https://bitcoinvpvcy.net/>



What is the conclusion?





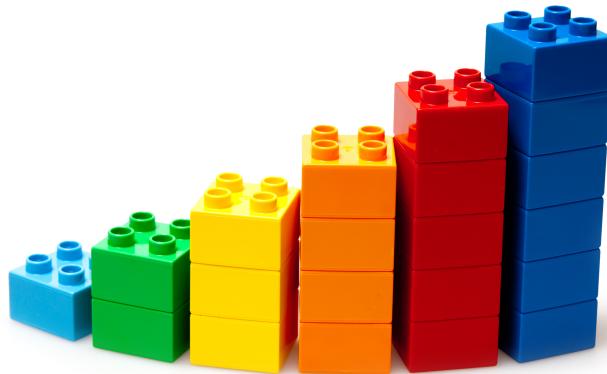
SIMPLIFY YOUR VISUALIZATIONS

The goal of a good visualization

CHAOS

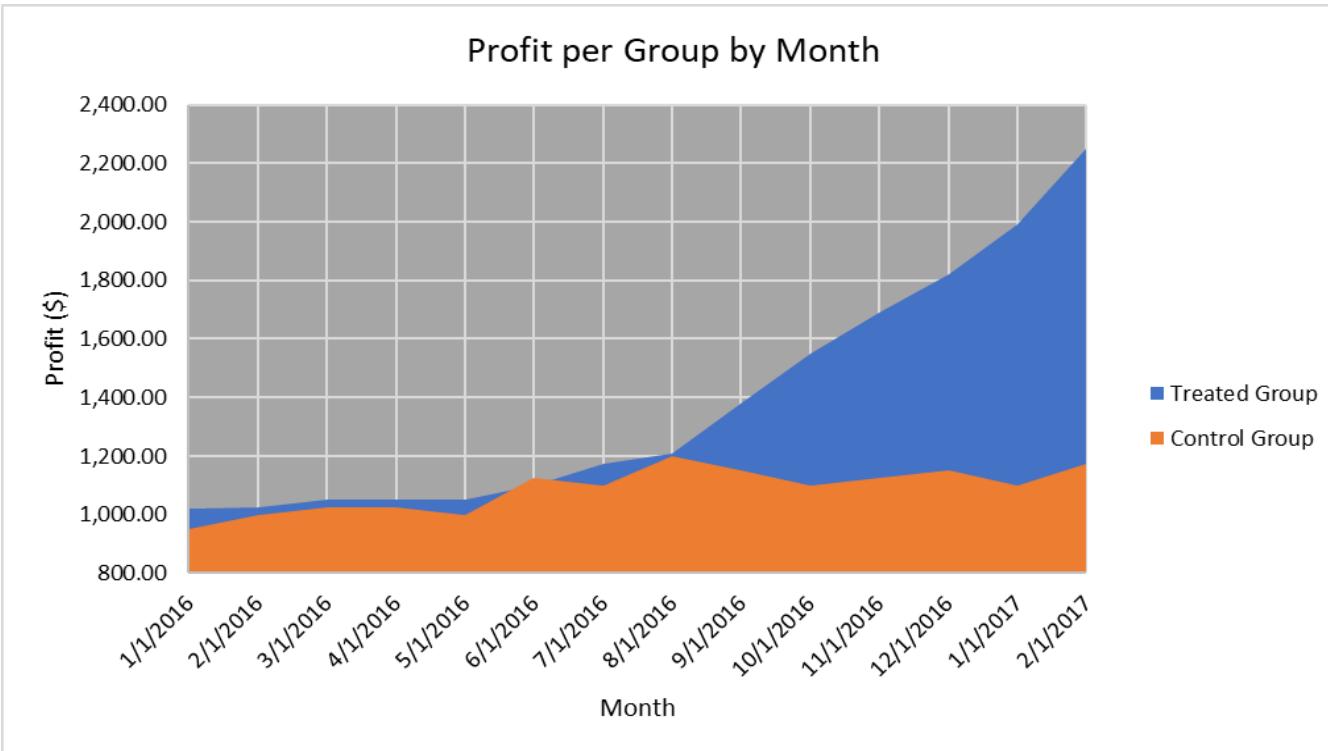


ORDER



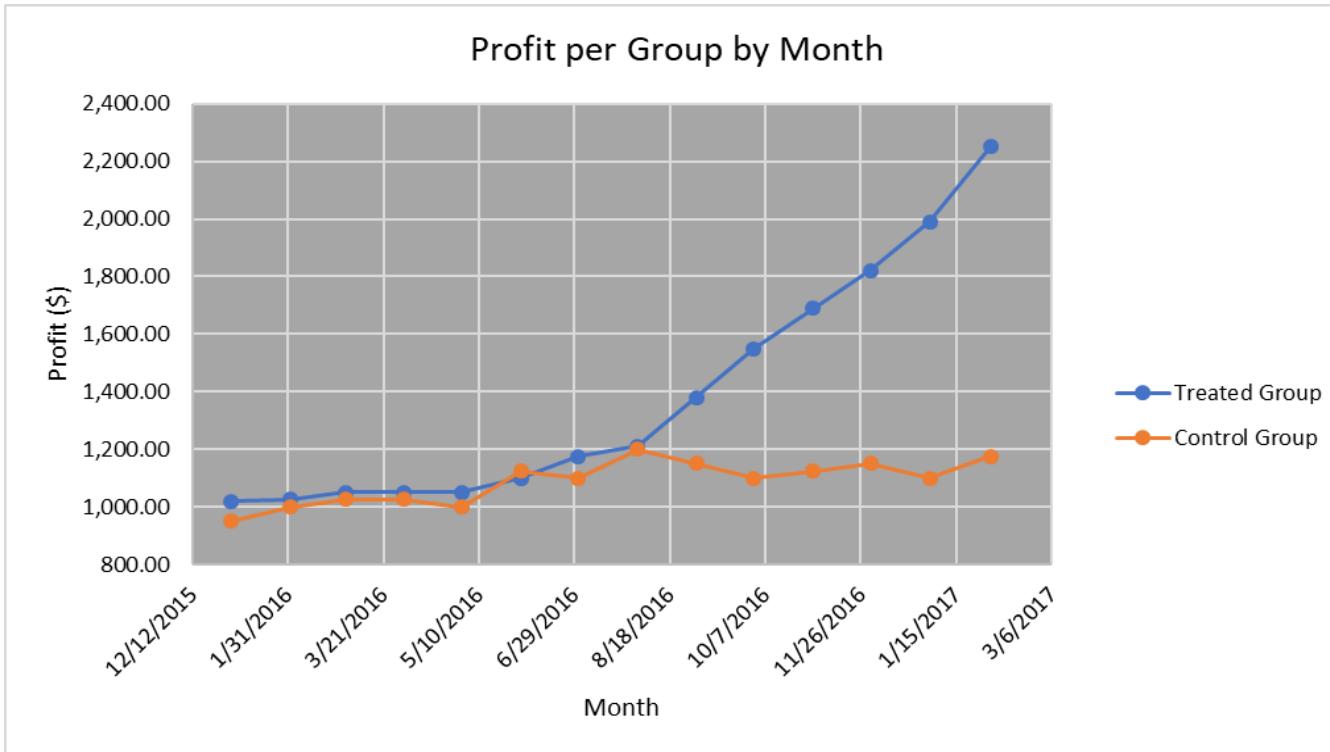
Example

Use line chart!



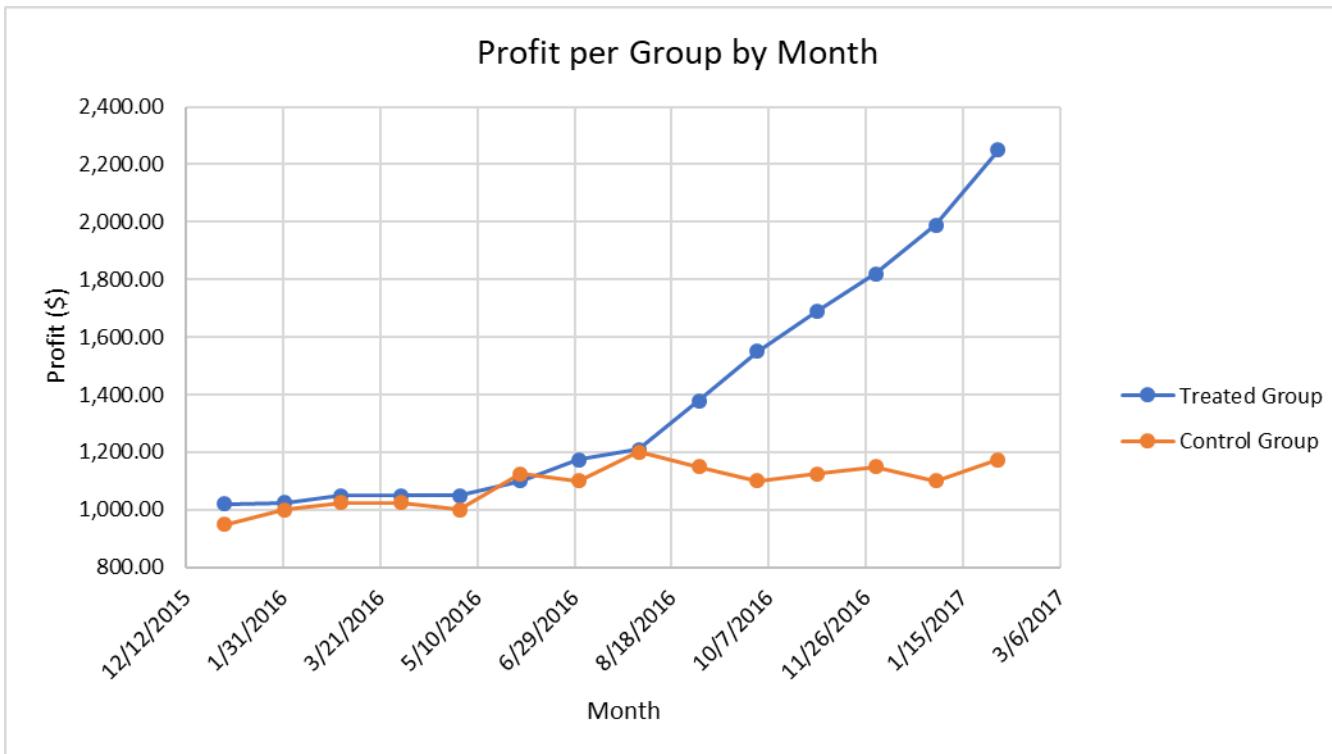
Example

Remove background color!



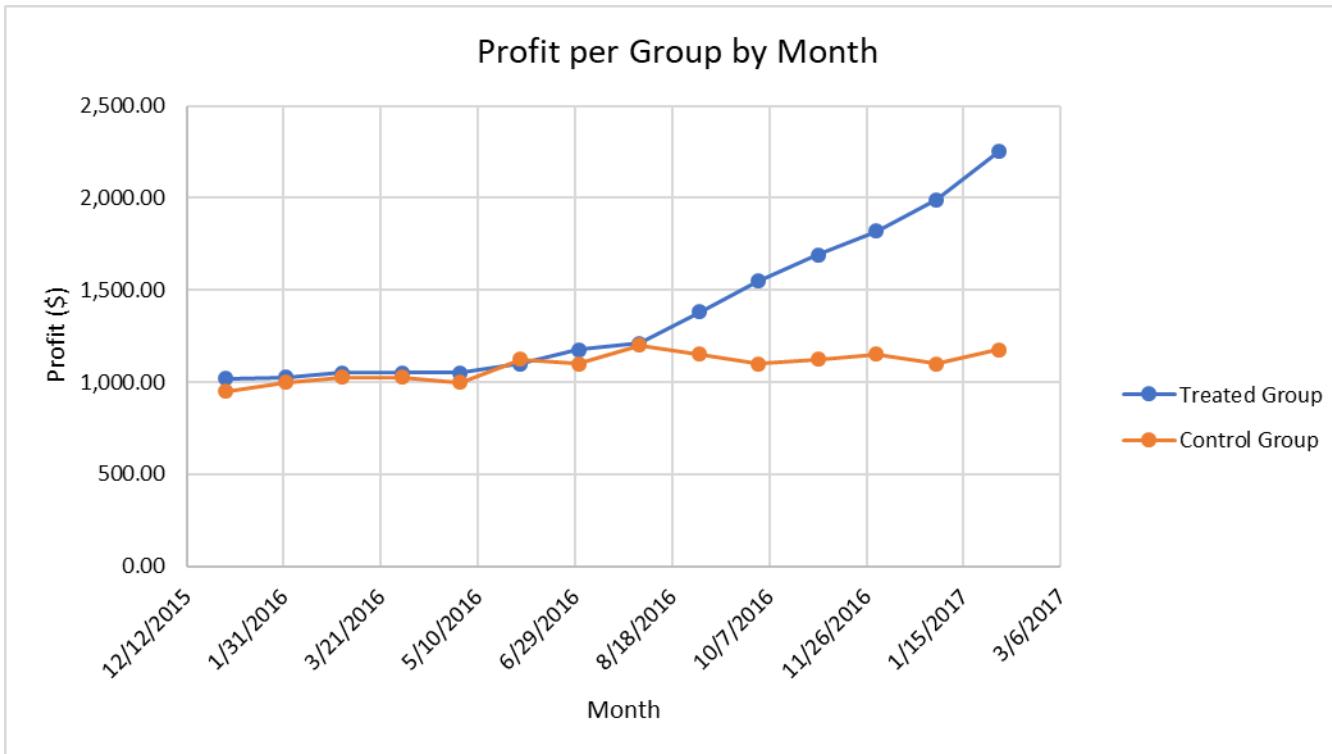
Example

Start axis at 0!



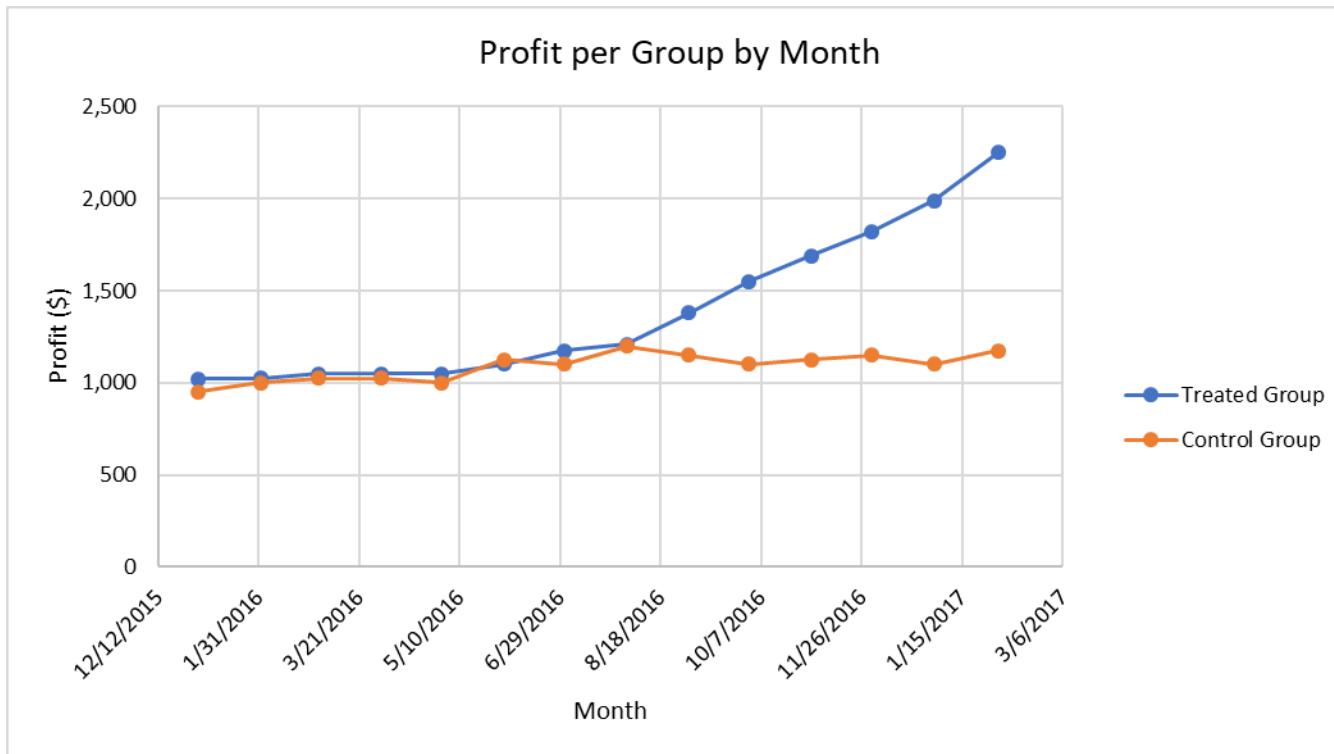
Example

Decimal points!



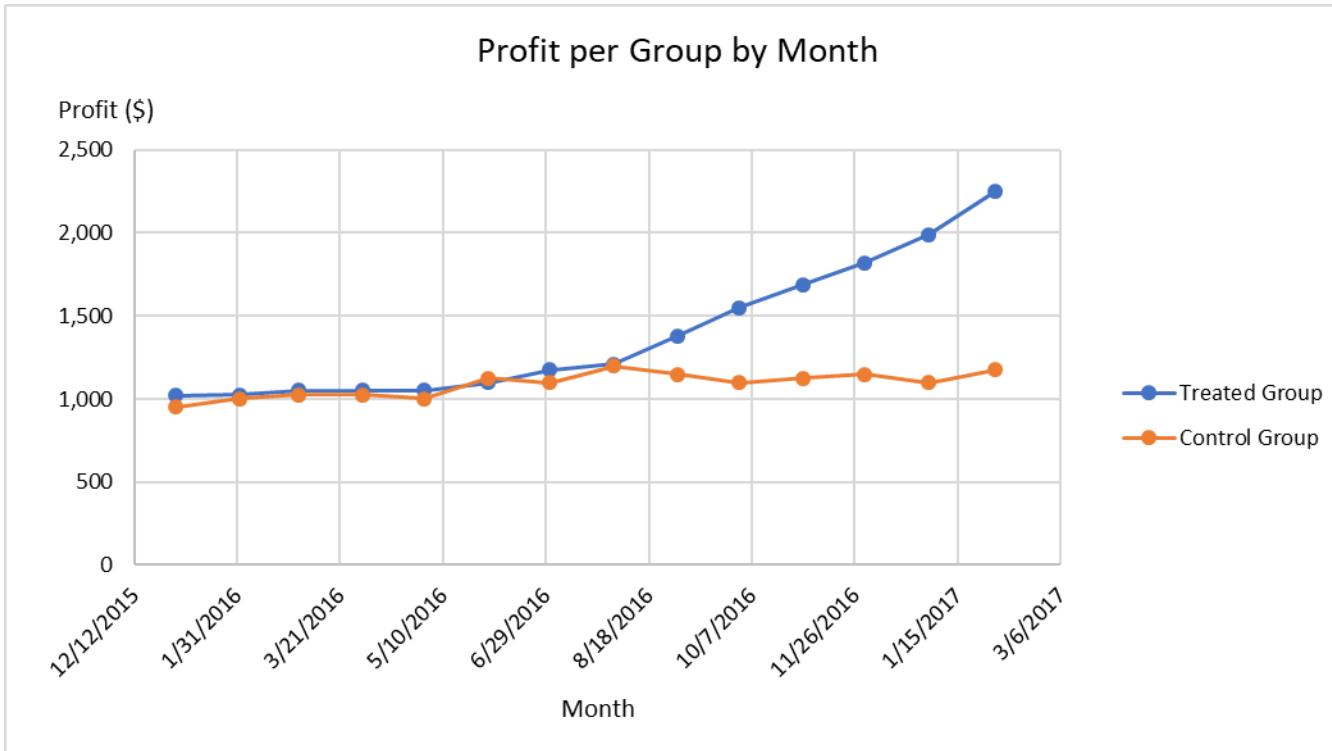
Example

Vertical “Profit” label!



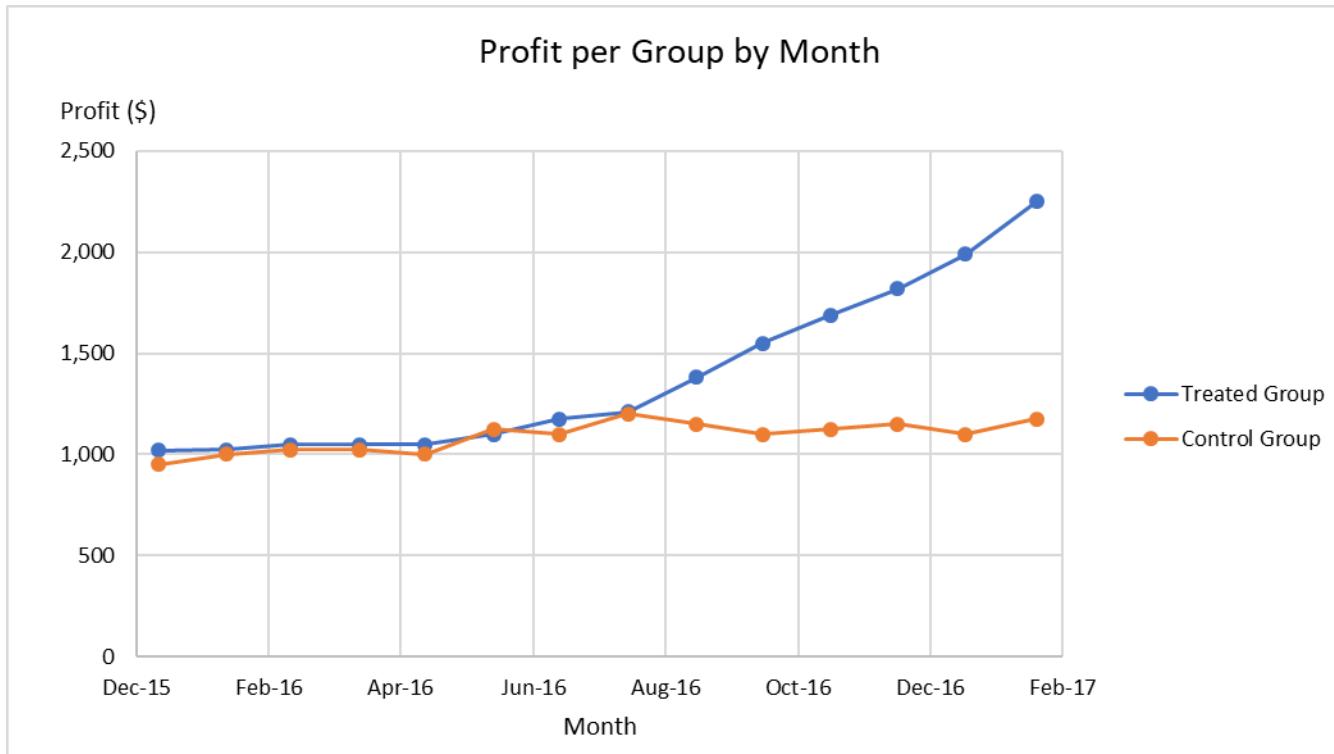
Example

Diagonal x-axis!



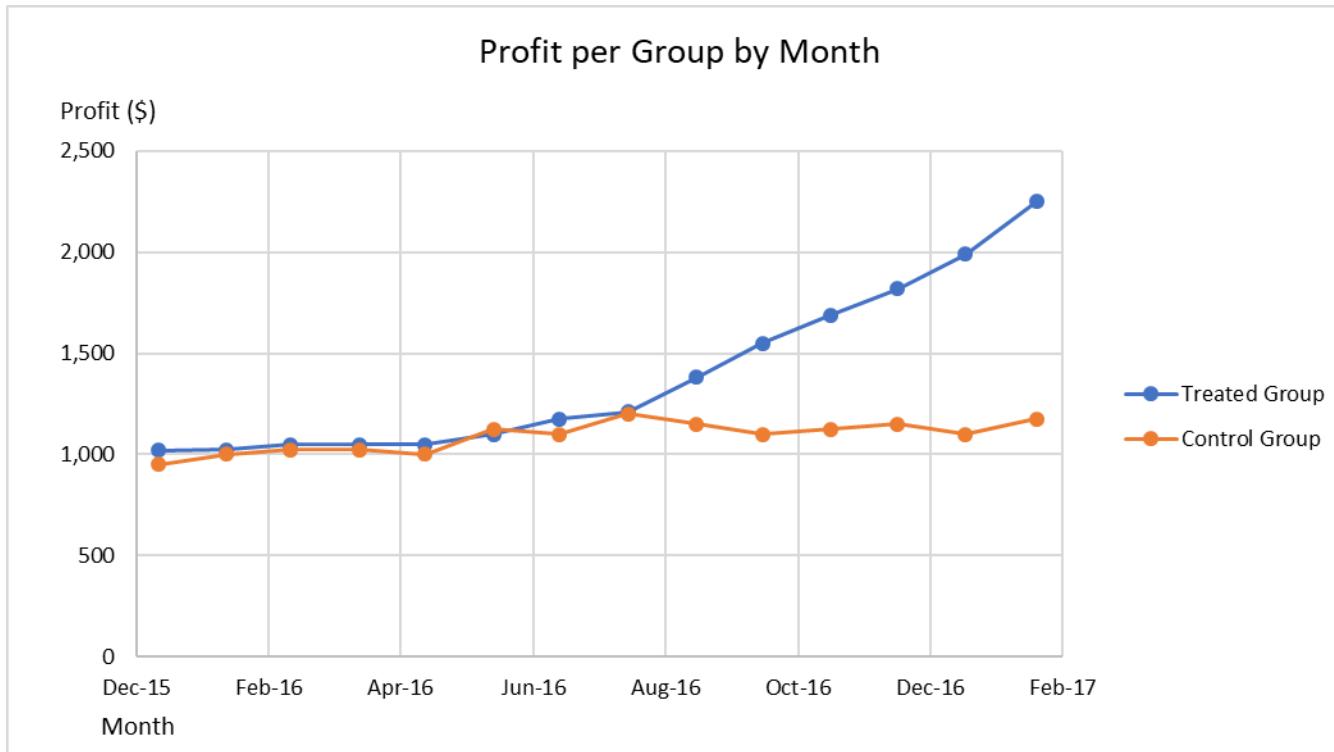
Example

X-Label left justified!



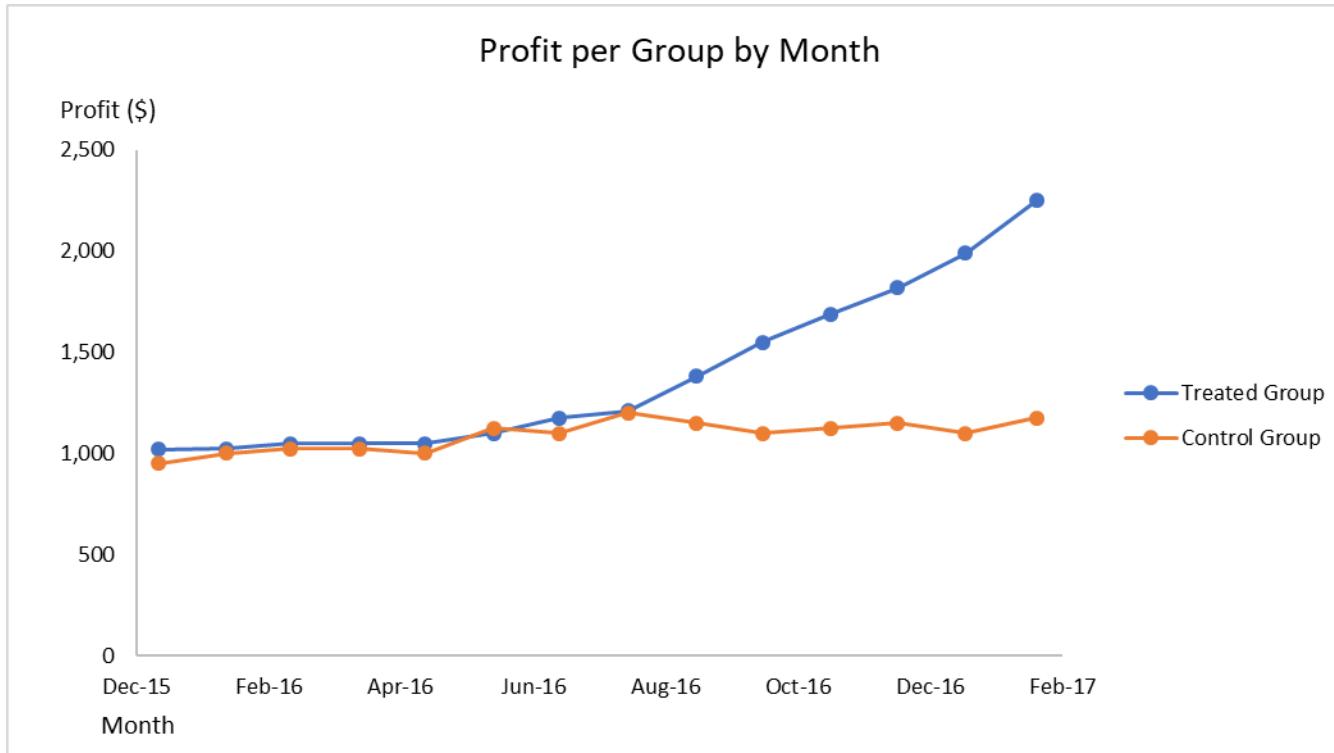
Example

Gridlines!



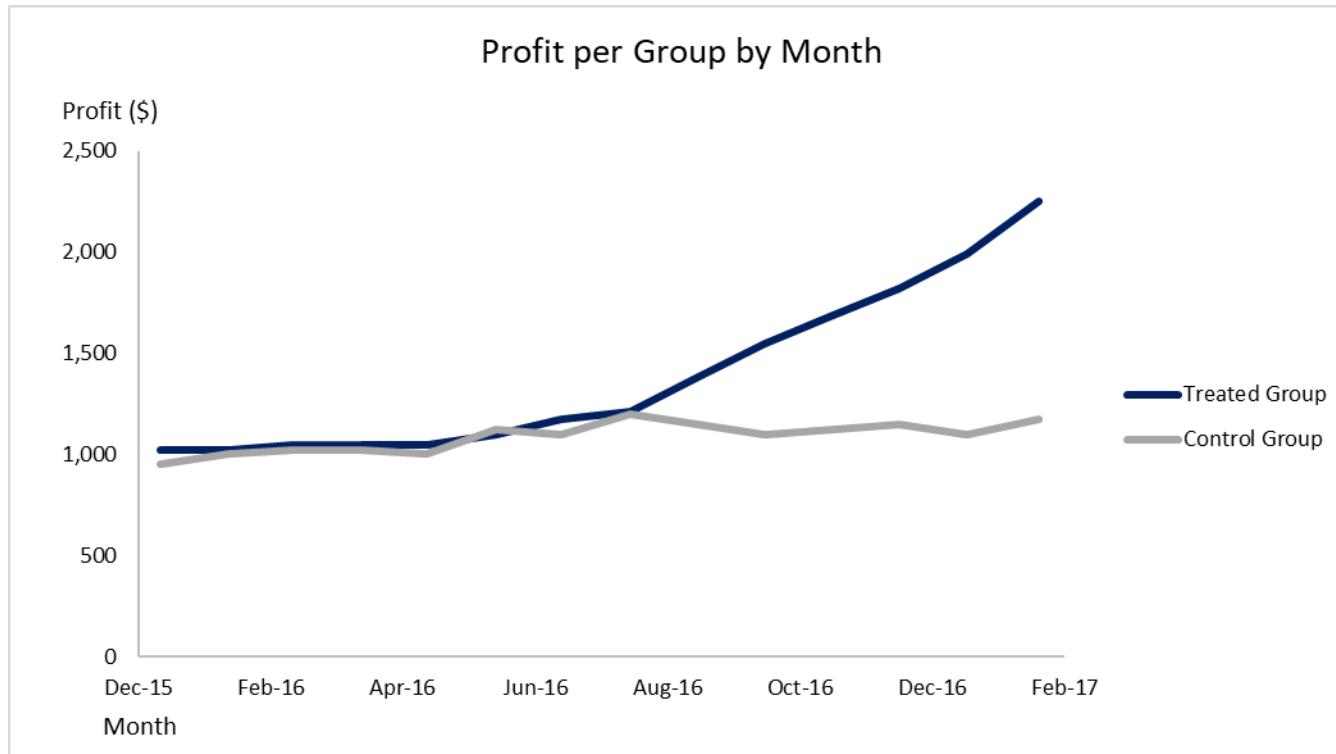
Example

Color and dots!



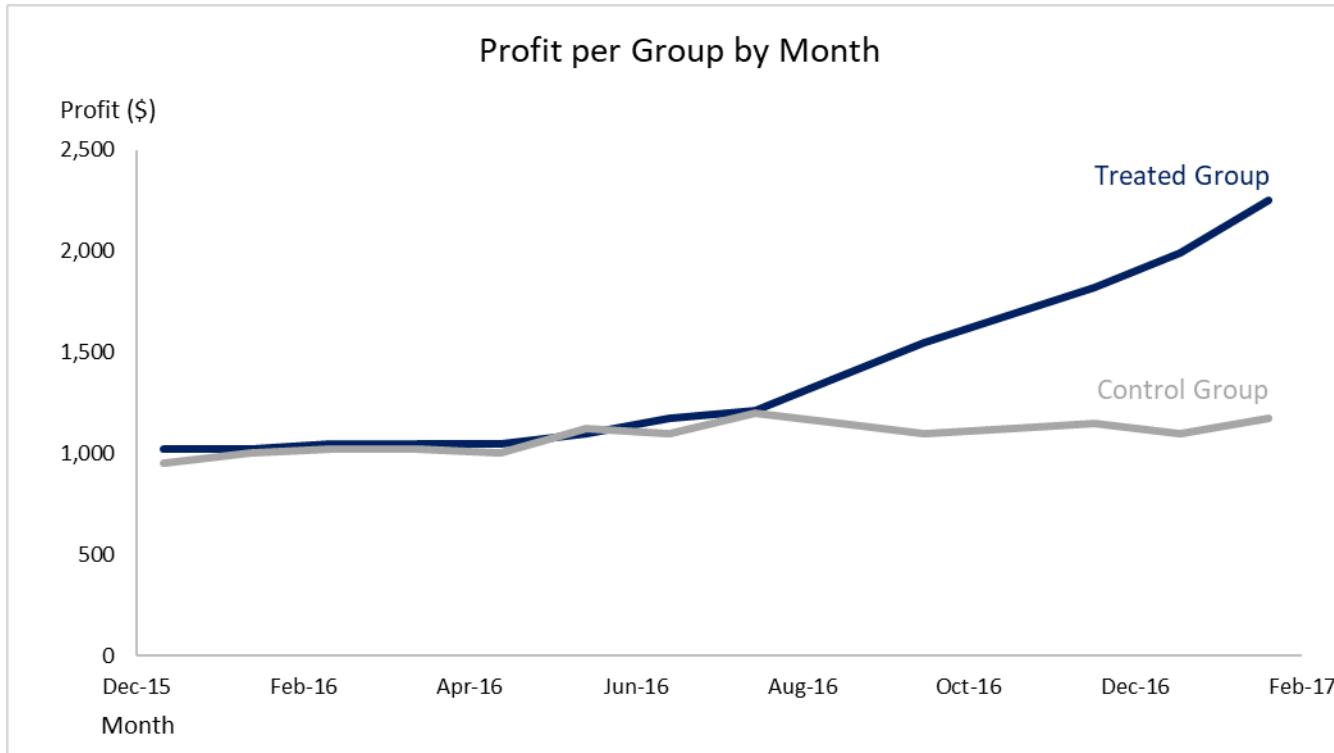
Example

Legend, proximity and space



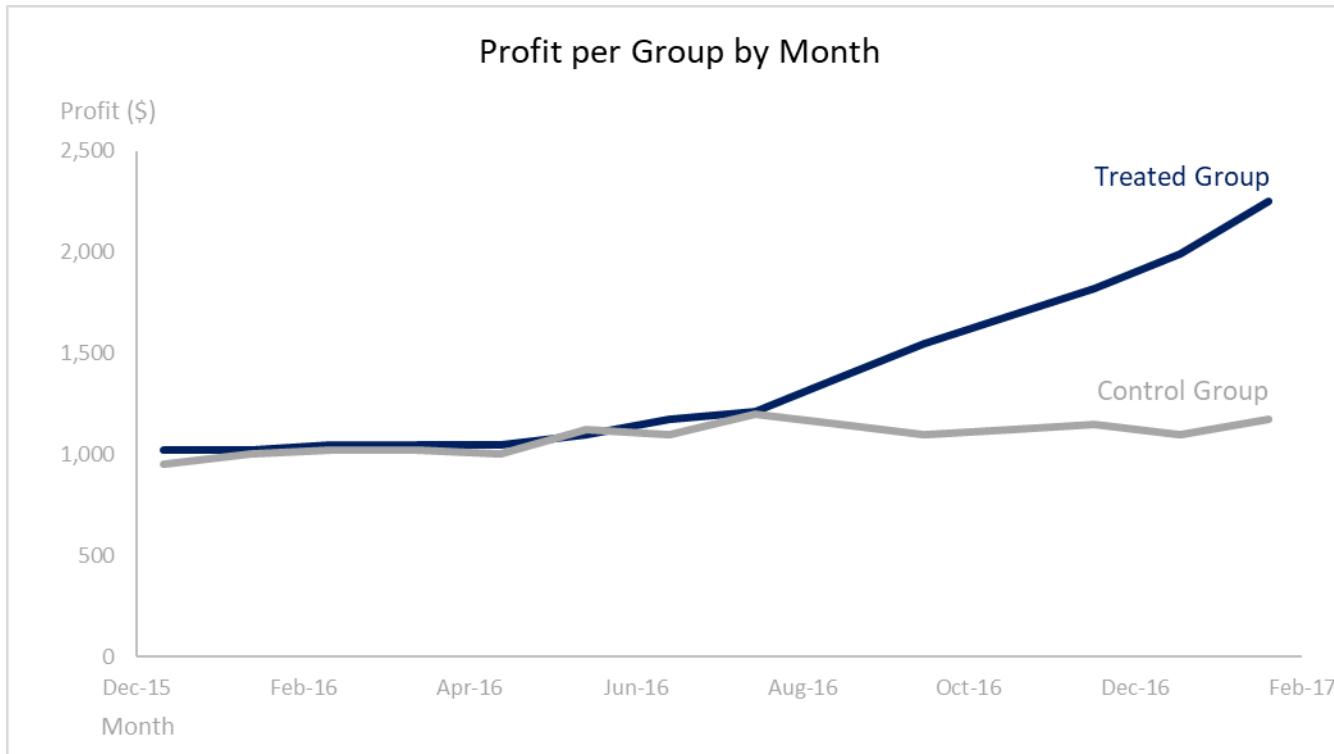
Example

Axis background



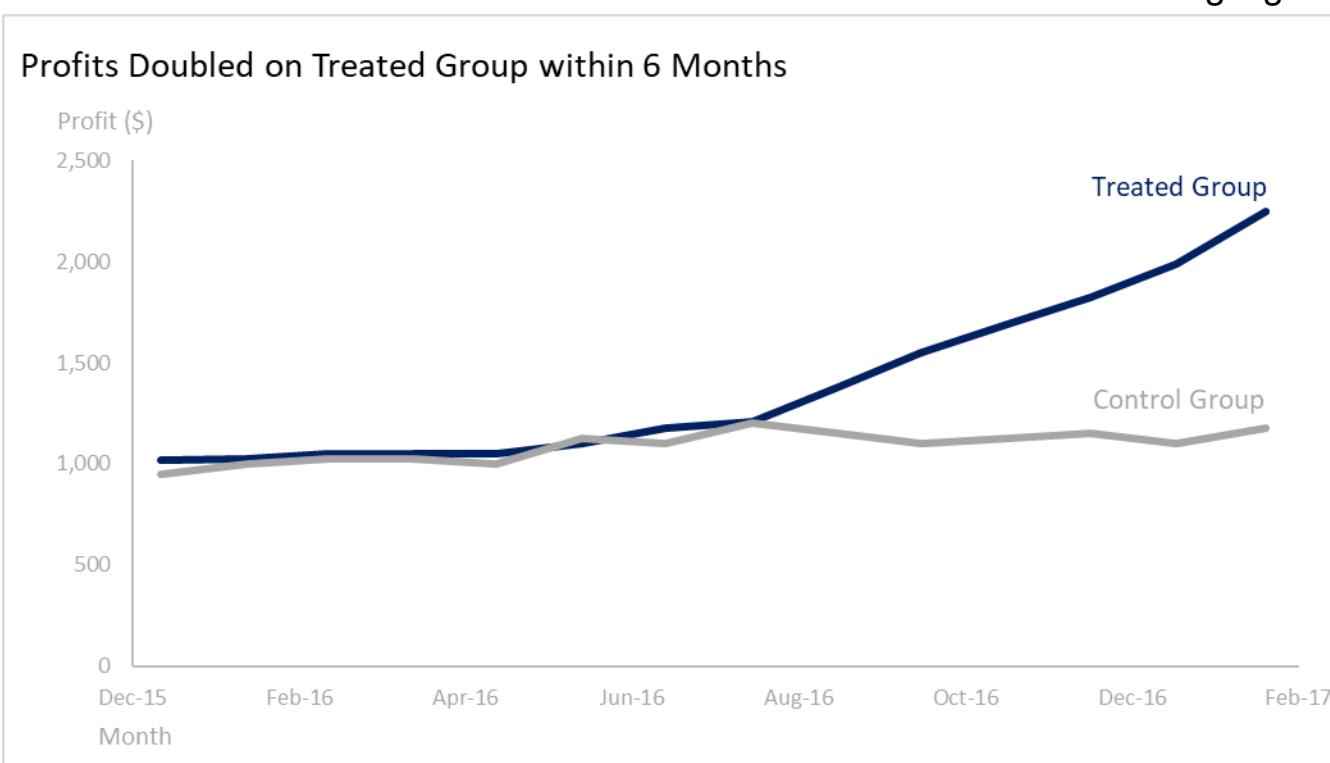
Example

Descriptive title and left!



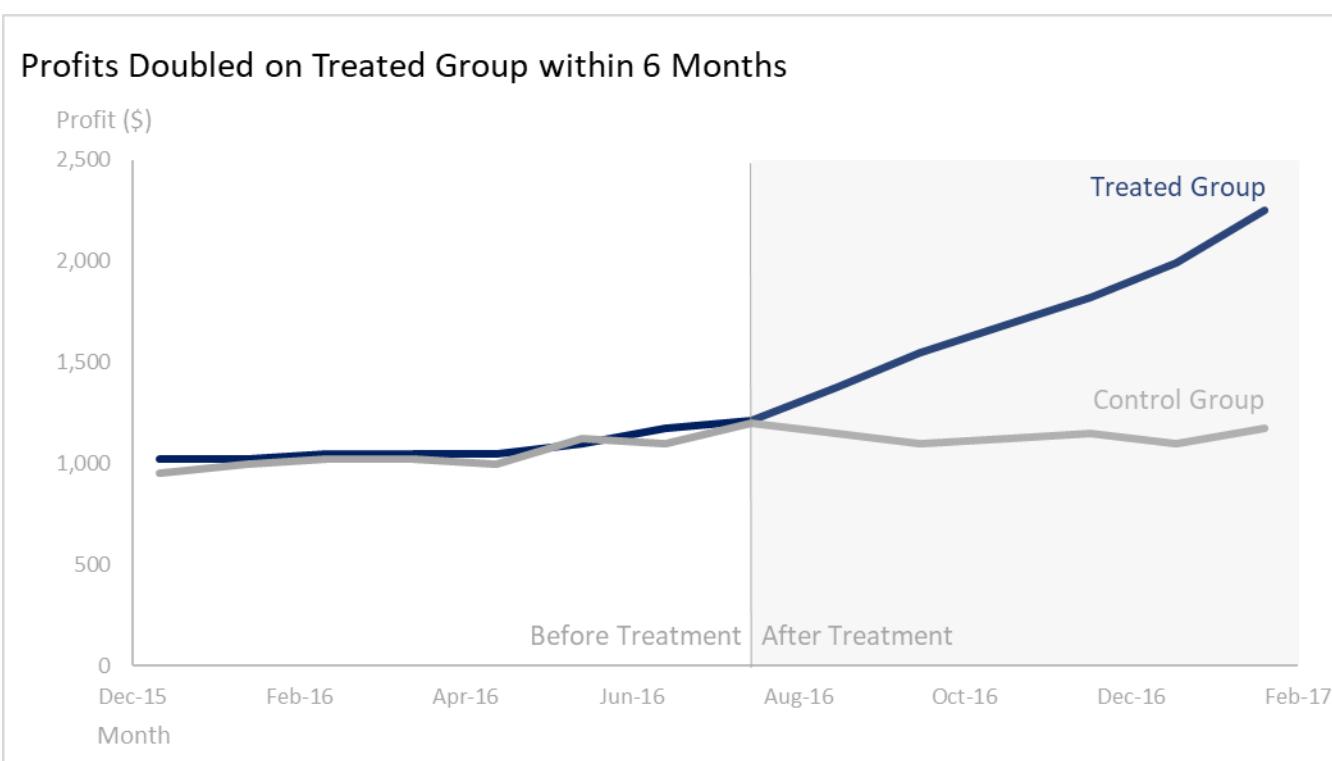
Example

Highlight!



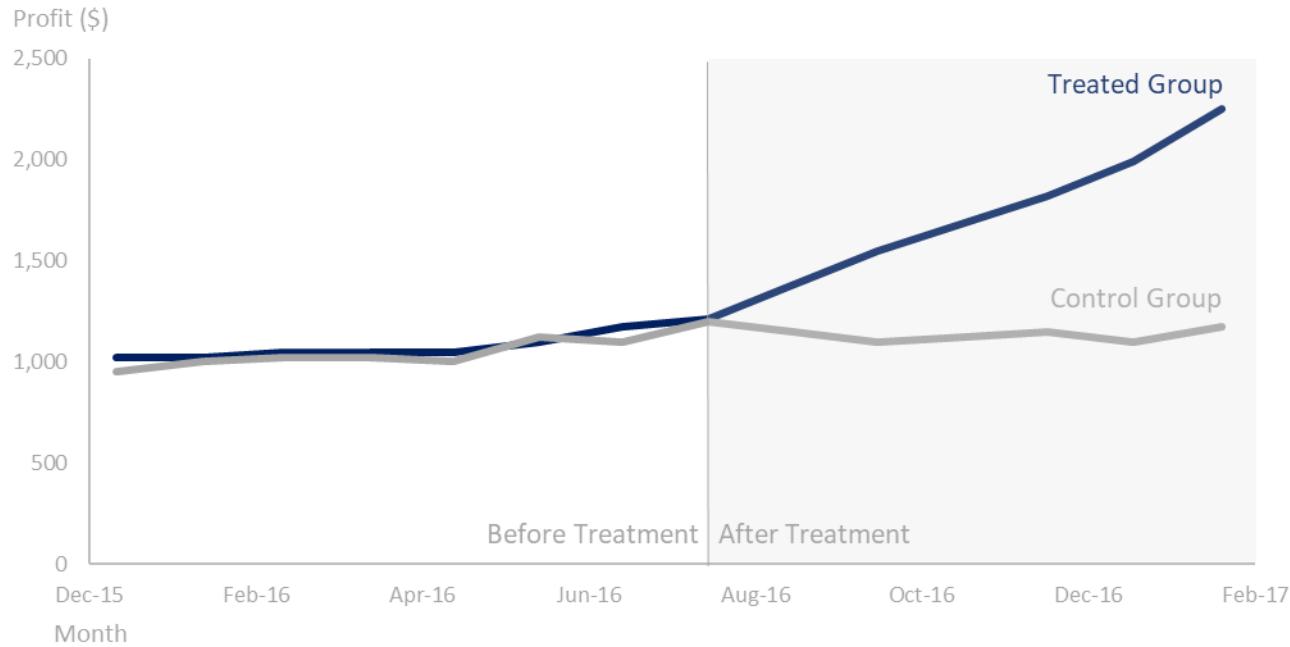
Example

Box!

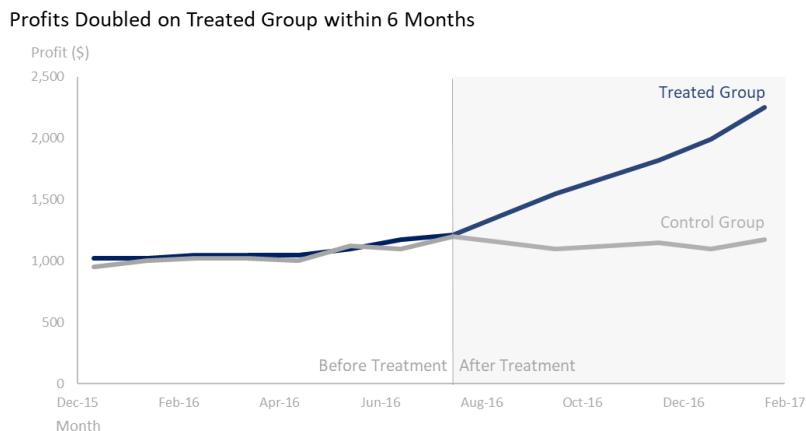
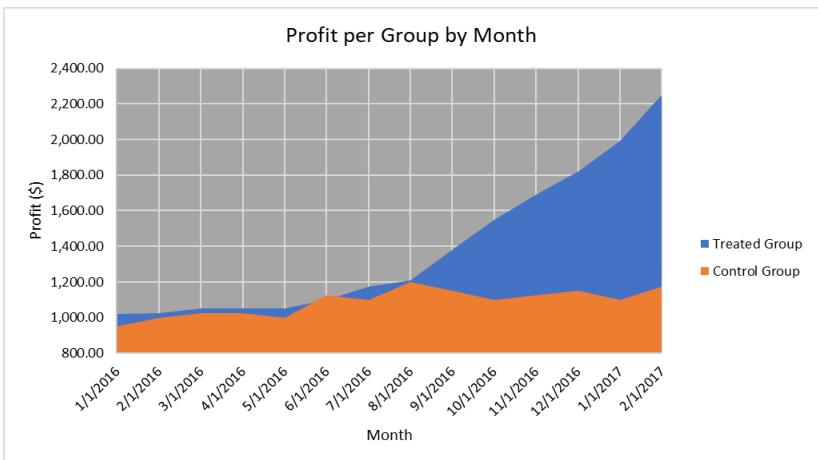


Example

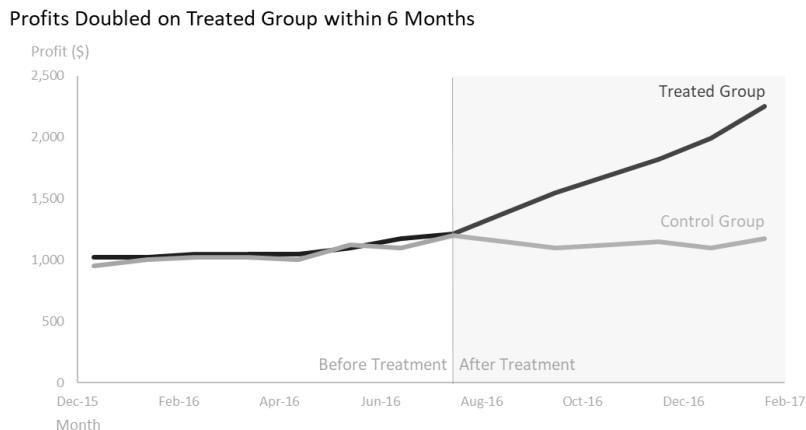
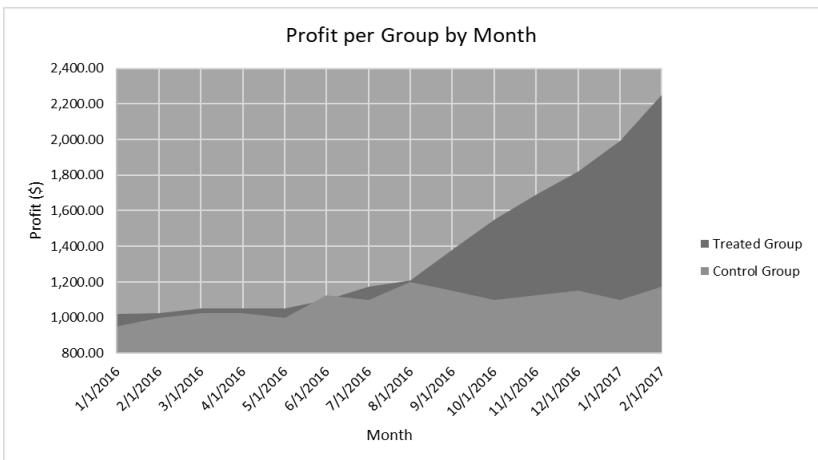
Profits Doubled on Treated Group within 6 Months



Example



Grayscale



Summary

- ▶ Charts should be effective, easy to interpret and understand
- ▶ Use highlight to draw attention
- ▶ Gridlines only if necessary
- ▶ Charts
 - ▶ Avoid 3D
 - ▶ Avoid Pie Charts with too many slices
- ▶ Text
 - ▶ Horizontal
 - ▶ Size easy to read
- ▶ Color
 - ▶ Max 5 colors
 - ▶ Consistency across charts
 - ▶ Be mindful of:
 - ▶ Grayscale
 - ▶ Color Blindness





THANK YOU!
