

# 365 DataScience

BY 365 DATA SCIENCE

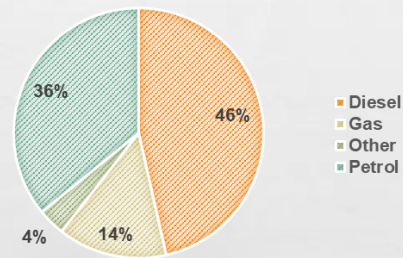
THE COMPLETE DATA VISUALIZATION  
COURSE 

# CHARTS AND DATA TYPES

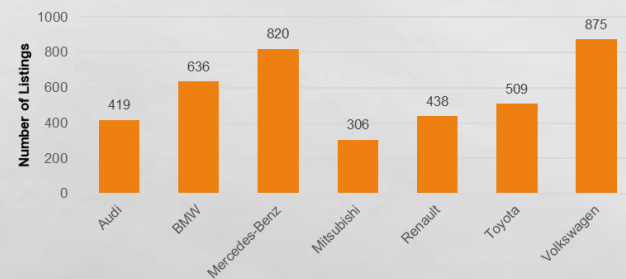
THERE IS NEVER ONLY ONE RIGHT VISUALIZATION

## NUMERICAL & CATEGORICAL

CARS BY ENGINE FUEL TYPE

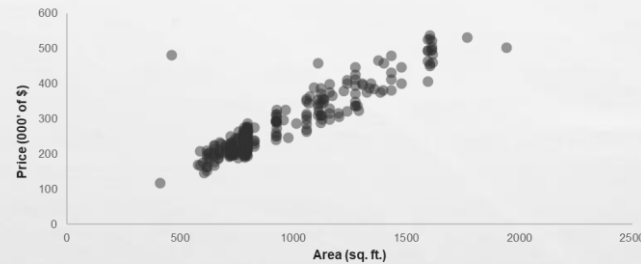


Car Listings by Brand

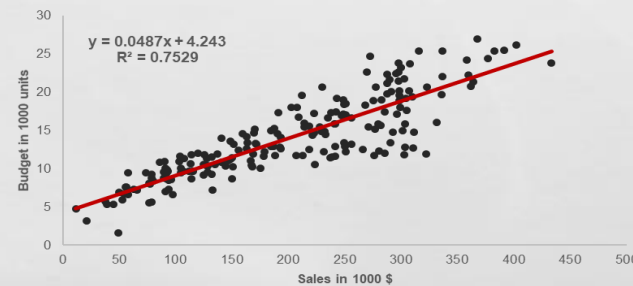


## NUMERICAL & NUMERICAL

Relationship between Area and Price of California Real Estate

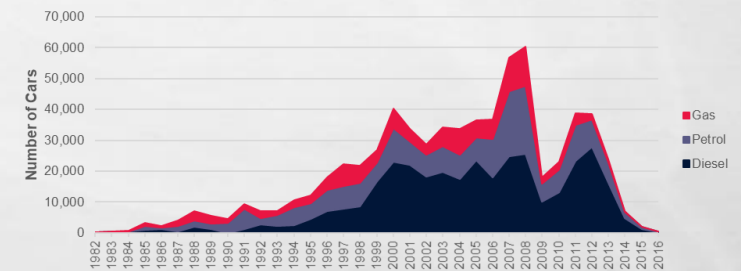


Advertisement vs Sales

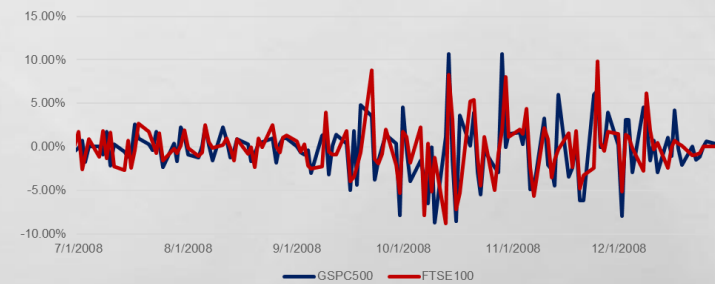


## TIME SERIES

Popularity of engine fuel types (1982-2016)



S&P vs FTSE Returns (H2 2008)



# COLORS

## CHOOSE 2-3 COLORS FOR YOUR CHART



### PREDETERMINED

**COMPANY COLORS**

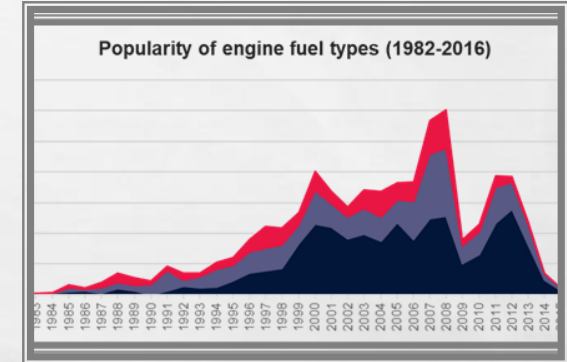
**CLIENTS REQUEST SPECIFIC COLORS**



### ONLINE TOOLS

**CREATE YOUR OWN CUSTOM PALETTE**

**WITH THE AID OF ONLINE TOOLS**



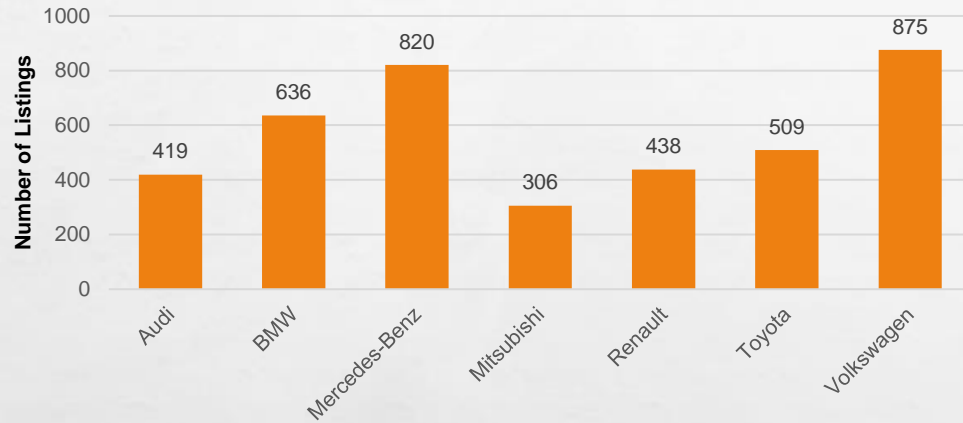
### OUR COLOR PALETTES

**YOU CAN USE ANY OF OUR TEMPLATES**

**TO BUILD YOUR OWN GRAPHS AND CHARTS**

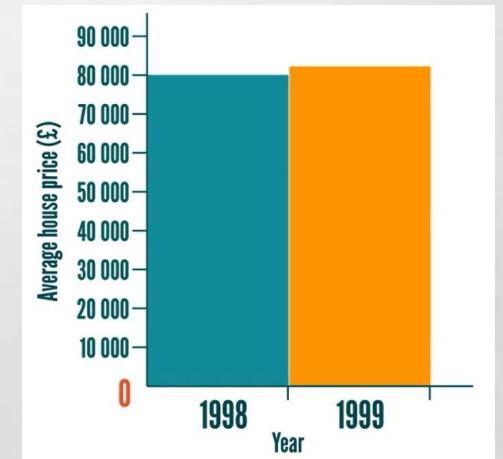
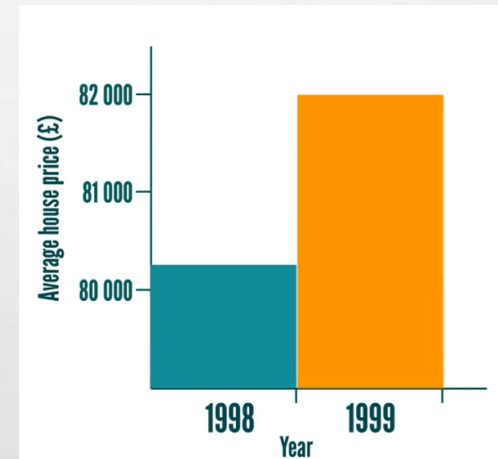
# Bar Chart

Car Listings by Brand



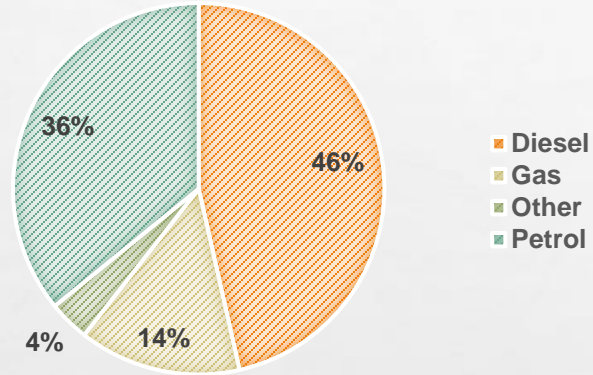
- **INTUITIVE**
- **APPROPRIATE FOR NON-TECHNICAL AUDIENCES**
- **ONE OF THE MOST COMMONLY USED CHARTS**

- **COMMUNICATE YOUR INTENTIONS CLEARLY**
- **MAKE SURE YOUR CHART ISN'T MISLEADING**



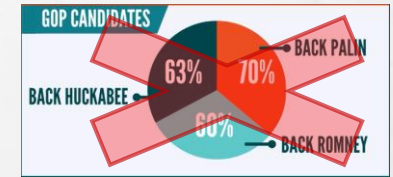
# Pie Chart

CARS BY ENGINE FUEL TYPE

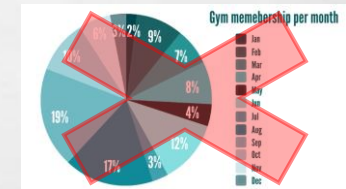


- **APPROPRIATE FOR NON-TECHNICAL AUDIENCES**
- **WIDELY USED, DESPITE CRITICISM**
- **A FEW CATEGORIES**
- **DATA SUMS UP TO 100%**

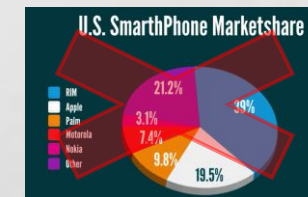
- **DON'T USE WHEN DATA  $\neq$  100%**



- **DON'T USE WHEN THERE ARE TOO MANY CATEGORIES**

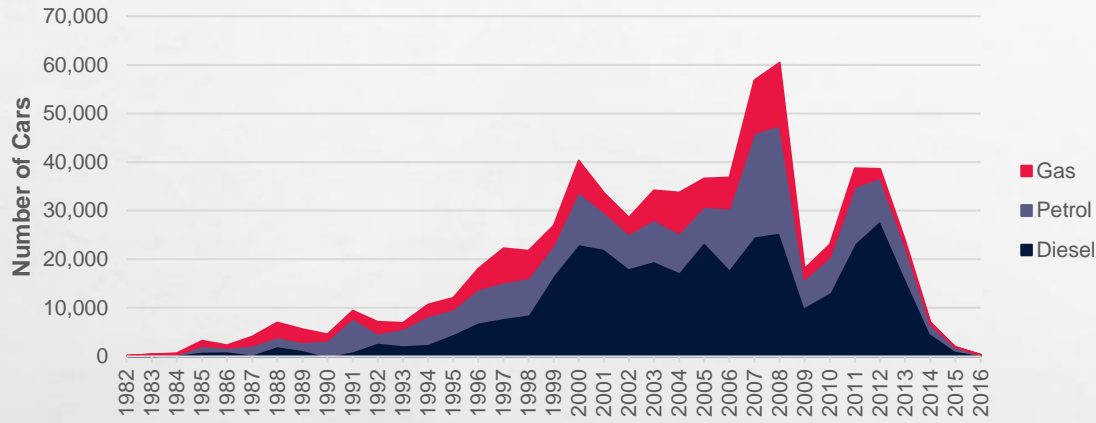


- **NO 3D OR DOUGHNUT**



# Stacked Area Chart

Popularity of engine fuel types (1982-2016)

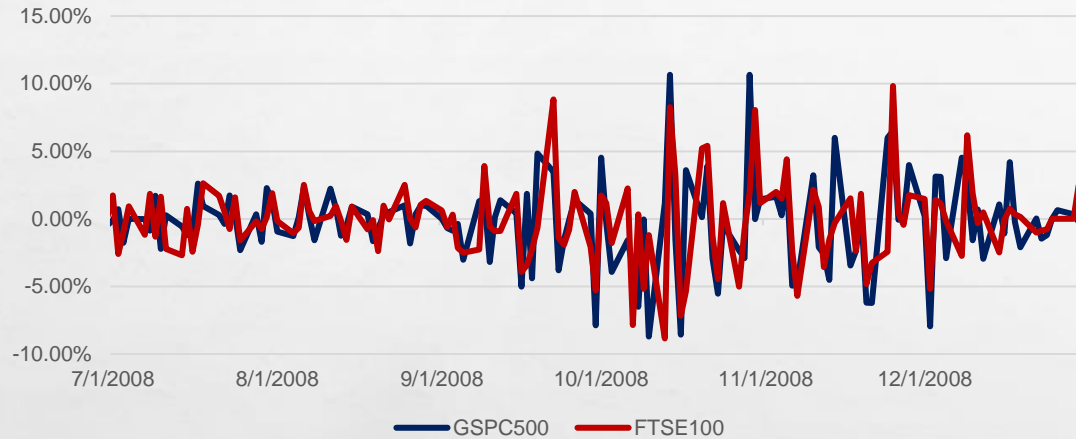


- **COMPARE VOLUME AMONG FEATURES**
- **AT LEAST THREE FEATURES**
  - **ORDERING FOR AT LEAST TWO OF THEM**
- **TIME SERIES DATA**

- **AVOID WHEN YOU HAVE TOO MANY CATEGORIES – A LINE CHART WORKS BETTER**
- **AVOID WITH CATEGORIES OF SIMILAR SIZE – DIFFICULT TO DETERMINE SIZE OF NON-RECTANGULAR SHAPES**
- **ORDER CATEGORIES BY SIZE – TO IMPROVE READABILITY**
- **Y-AXIS MUST START AT 0 – WE'RE MEASURING VOLUME**

# Line Chart

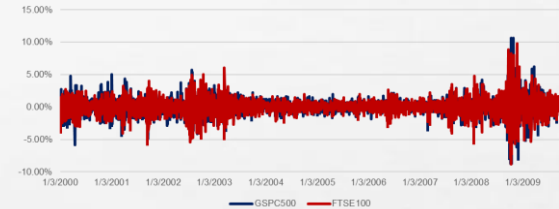
S&P vs FTSE Returns (H2 2008)



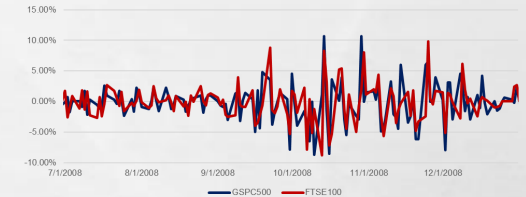
- **UP TO SEVERAL CATEGORIES**
- **TIME SERIES DATA**
- **Y-AXIS DOESN'T HAVE TO START AT 0**

- **WHEN YOU HAVE A LARGE PERIOD OF TIME, NARROW IT DOWN TO GAIN MORE INSIGHT**

S&P vs FTSE Returns (2000 - 2010)



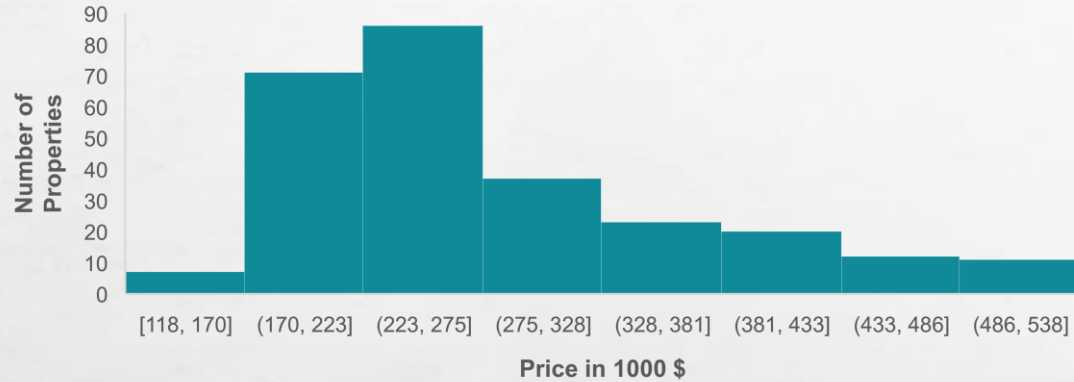
S&P vs FTSE Returns (H2 2008)



- **BE CAREFUL NOT TO INCLUDE TOO MANY CATEGORIES, TO AVOID A SPAGHETTI CHART**

# Histogram

Histogram of Real Estate Prices

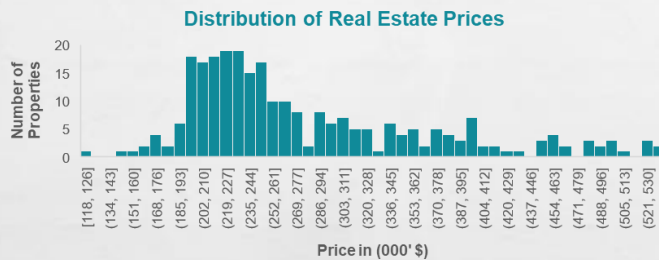


- **DISTRIBUTION OF A NUMERIC VARIABLE**
- **THE VARIABLE'S RANGE OF VALUES IS SPLIT INTO INTERVALS OR BINS**
- **Y-AXIS –NUMBER OF OBSERVATIONS WITHIN EACH INTERVAL(OR DENSITY)**

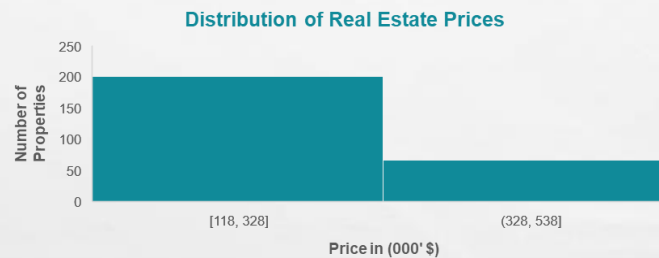
- **SIMILAR TO A BAR CHART, NO GAP BETWEEN BINS**
- **TO CREATE A HISTOGRAM**
  - **DETERMINE THE INTERVAL SIZE**
  - **CHOOSE THE NUMBER OF BINS**



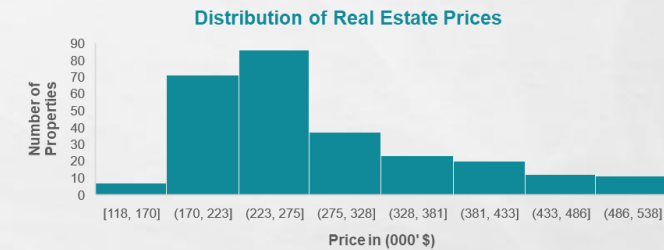
# CHOOSING THE NUMBER OF BINS



**START WITH A VERY LARGE NUMBER TO  
OBSERVE THE DATA PATTERN**



**REDUCE THE NUMBER**



**CHOOSE SEVERAL BINS, SUCH THAT THE  
PATTERN IN THE DATA IS VISIBLE**

**There are scientific approaches, however, they are not often used in practice.  
The reason is that real data has noise, is discrete, etc.**

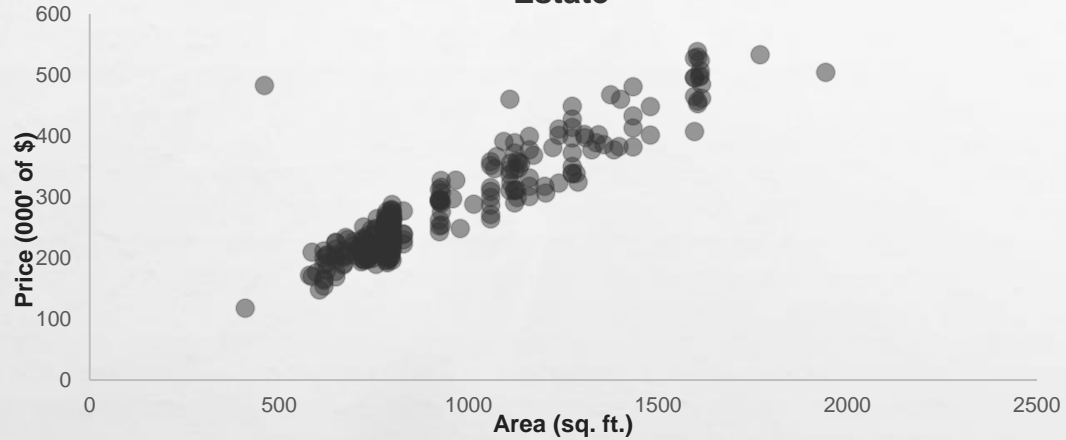
**Scott's rule -  $3.49\sigma n^{-1/3}$**

**Sturge's Rule -  $K = 1 + 3.322 \log_N$**

**Doane's Rule -  $\log_2(n) + 1 + \log_2(1 + \frac{\sqrt{b}}{\sigma\sqrt{b}})$**

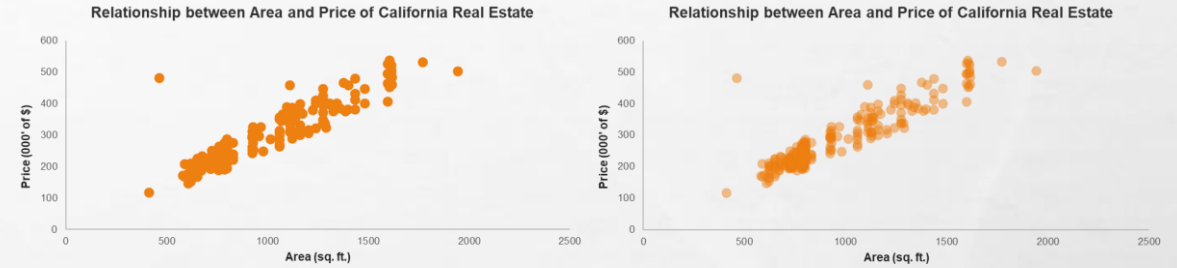
# Scatter Plot

Relationship between Area and Price of California Real Estate



- **DISPLAYS EACH POINT FROM THE DATA, INSTEAD OF SHOWING AGGREGATED FORM**
- **SHOWS RELATIONSHIP BETWEEN VARIABLES**

- **USE TRANSPARENCY TO AVOID OVERPLOTING**

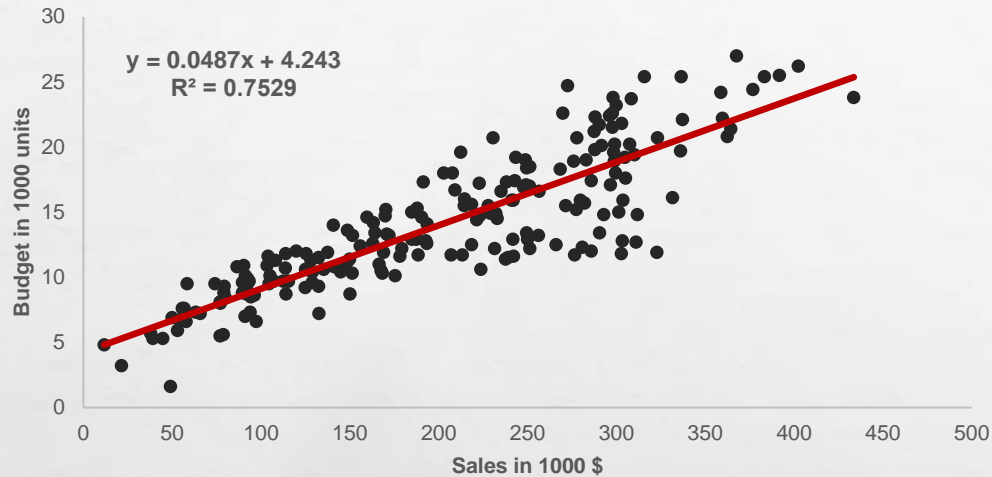


- **A THIRD VARIABLE COULD BE USED WITH A COLOR PARAMETER**



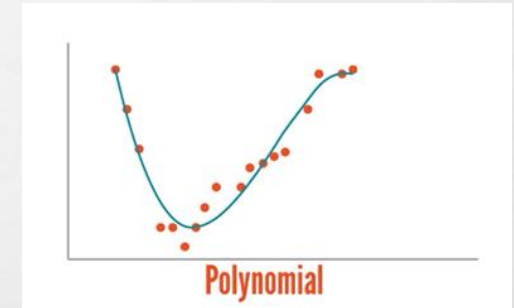
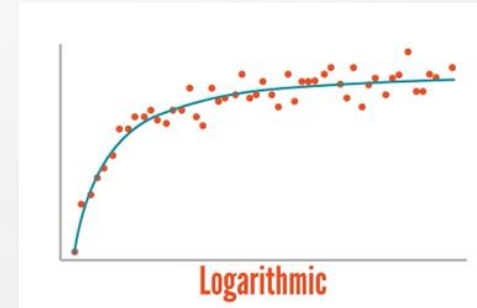
# Regression Plot

Advertisement vs Sales

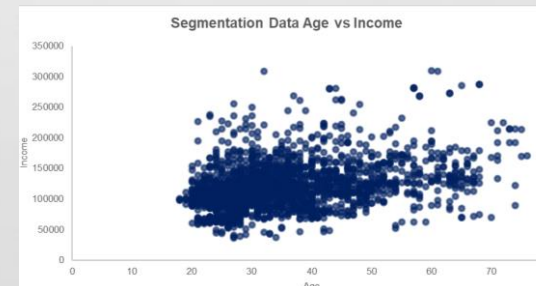


- **USED TO DETERMINE RELATIONSHIPS BETWEEN PREDICTOR(S) AND OUTCOME**
- **REGRESSION LINE & EQUATION HELP US QUANTIFY THE RELATIONSHIP**

- **THERE EXIST MANY TYPES OF RELATIONSHIPS BETWEEN VARIABLES**

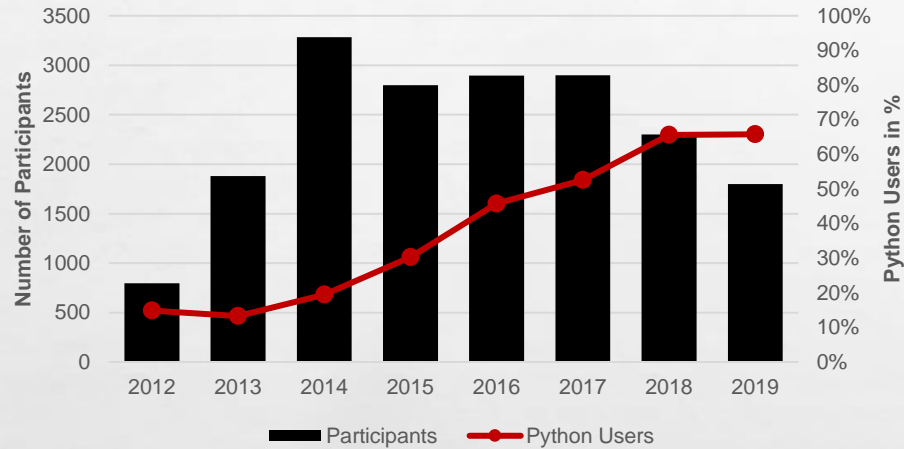


- **SOMETIMES THERE IS NO APPARENT RELATIONSHIP BETWEEN FEATURES**



# Bar and Line Chart

KD Nuggets Survey  
Python Users (2012 - 2019)



- **COMBINATION CHART OF BAR CHART & LINE CHART**
- **DUAL Y-AXIS**
- **IT'S ESSENTIAL FOR THE TWO CHARTS TO BE WELL LABELED**

- **PARETO CHART IS A SPECIFIC CASE OF BAR & LINE CHART**



- **THE SECONDARY Y-AXIS ON A PARETO SHOWS THE CUMULATIVE FREQUENCY AND SUMS UP TO 100%**

# ADDITIONAL RESOURCES

- [HTTPS://SEABORN.PYDATA.ORG/TUTORIAL/REGRESSION.HTML](https://seaborn.pydata.org/tutorial/regression.html)
- [HTTP://WWW.COOKBOOK-R.COM/](http://www.cookbook-r.com/)
- [HTTP://WWW.STAT.COLUMBIA.EDU/~TZHENG/FILES/RCOLOR.PDF](http://www.stat.columbia.edu/~tzheng/files/rcolor.pdf)
- [HTTPS://PYTHON-GRAPH-GALLERY.COM/100-CALLING-A-COLOR-WITH-SEABORN/](https://python-graph-gallery.com/100-calling-a-color-with-seaborn/)
- [HTTPS://WWW.DATA-TO-VIZ.COM/](https://www.data-to-viz.com/)
- [HTTPS://COLORS.CO/](https://colors.co/)