**9-14-23 notes**

Charting a New Course with Excel

Module 1.3 goals:

* Create, modify, and stylize basic charts from start to finish in Microsoft Excel.
* Create scatter plots and trend lines.
* Create charts that contain filtered data.
* Create regressions and calculate moving averages in Excel.

Adding files to GitHub

Add .gitignore- tells GitHub to not track changes for a certain file type.

**Excel Charts-**

Once a chart is created, the + sign in the upper right can be used to add elements

* Axis labels
* Trend line (linear, exponential, etc…)

Filter button can filter lines out

Painbrush can edit style and color

Can cut and paste chart to a new tab in excel

**Data filtering in Excel-**

Select a column and press/hold shift to move a column

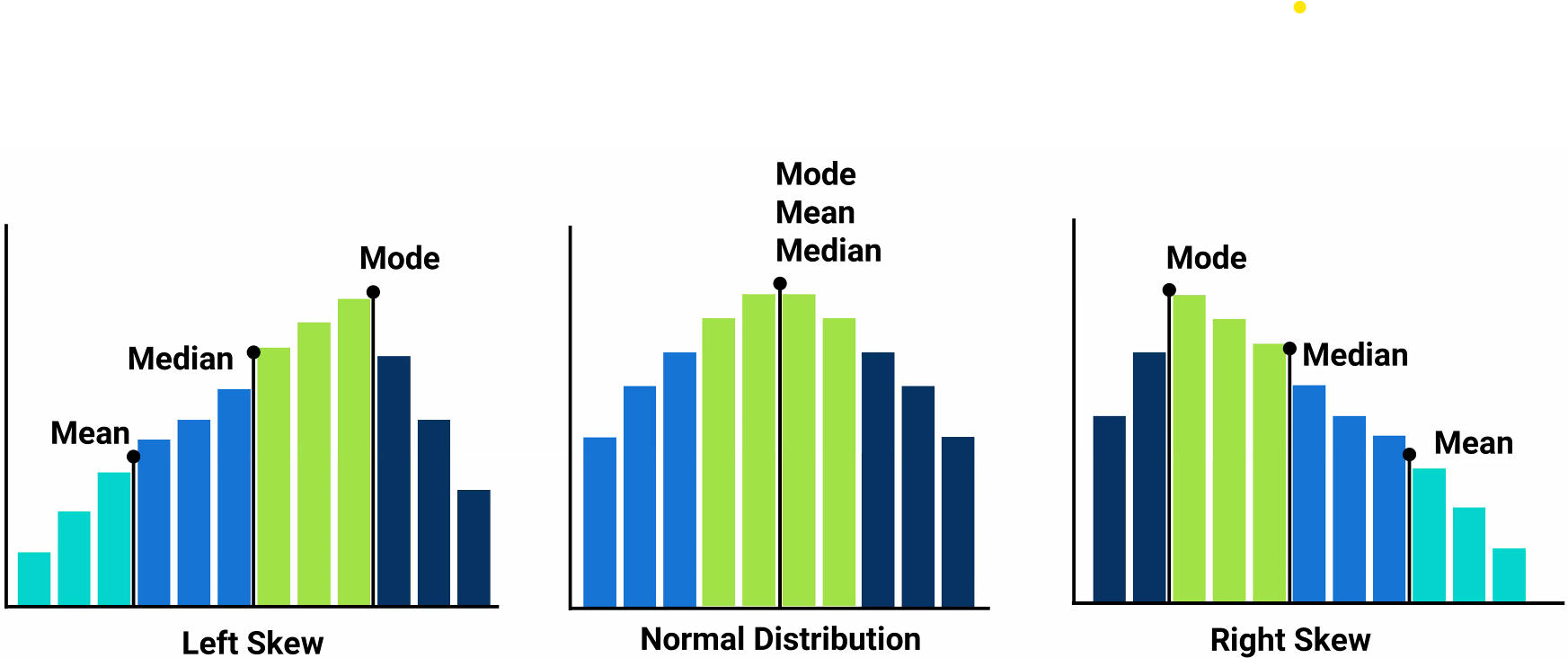
Ribbon/Data/sort to open sort criteria box.

* Can select checkbox to tell excel the data has headers (ignore row 1)

You can sort/filter by cell color as well as contents

Pivot Charts are THE SHIT!!!

**Statistics Stuff-**



For the sake of today’s lesson, we’re working in a normal distribution (Bell curve)

One standard deviation in the diagram above is two columns.

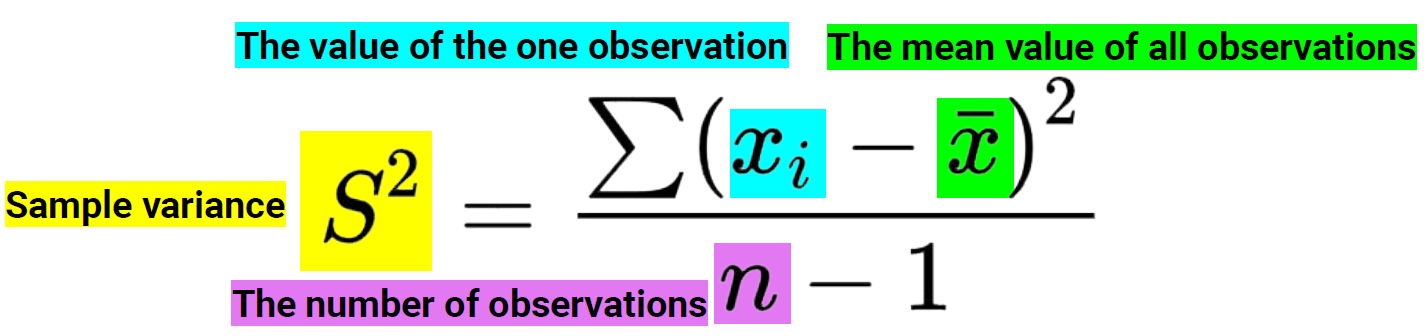
* To the left of the center line is negative SD
* To the right is positive SD

The center of a data set is described by Central Tendency (Mean, Median, Mode)

Variability of a data set is described by Variance, Standard Deviation, and Z-Score

* IE: Low SD is a low risk stock to invest in. High SD could yield higher growth, but also higher potential for the value of the stock to go down

Variance is a solid number. No unit to it.



Variance is S^2, Standard Deviation is root of S^2 aka S.

Lower standard deviation is a better set of data, (more confidence in the data set).

Z-Score describes how far a single point is from the Mean of a data set.

* No unit and can be positive or negative (Right of Mean [+], Left of Mean [-])

In Excel, you can use the formula below to calculate standard deviation.

=stdev.p() vs =stdev.s() {.s and .p stand for Sample/Population}

Small data sets vs Large data sets

=var.p() and =var.s() are used for calculating Variance of a range of cells

No formula for z-score in excel. Use below instead

=(Cell you’re calculating z-score for- Cell for Mean)/Cell for SD

Example- Story of volatility of temperatures between two regions (ie- Austin, TX and San Francisco, CA)

Manish is Six sigma blackbelt certified (Fairly common certification for statistical analysis)

Use Fn+f4 to have Excel automatically put the dollar signs in for absolute reference of the most recent cell in a formula in excel

**Quantiles, Outliers, Box Plots-**

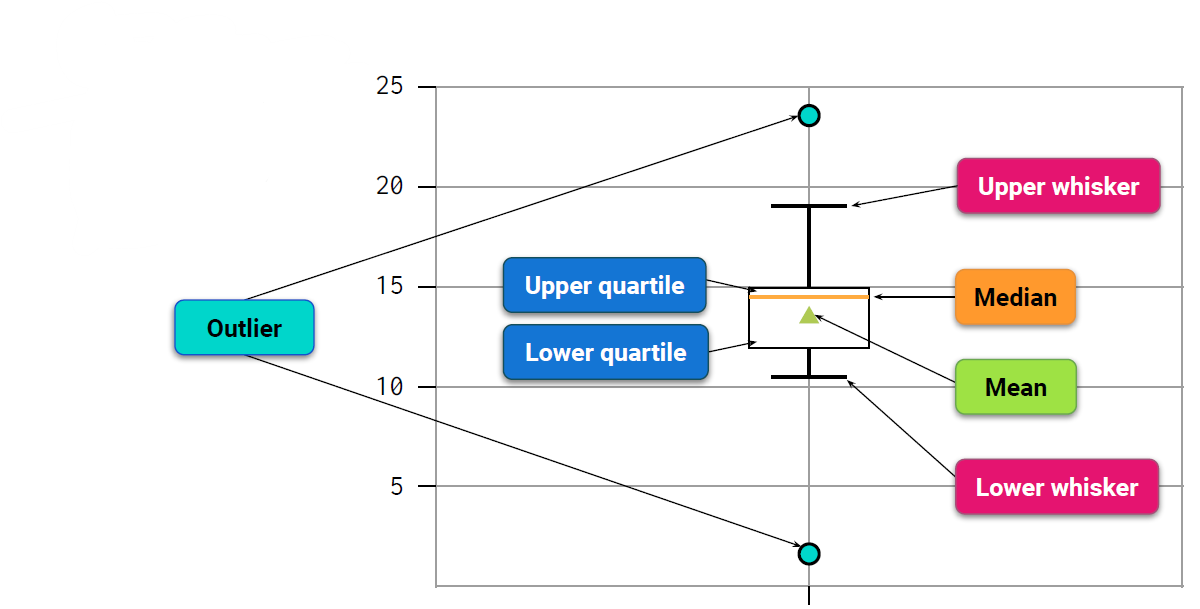
Percentiles are 100 even parts (ie- 95th percentile)

Quartiles are mostly used in Data Science, not in actual business fields.

Quartiles divide the data set into 4 parts

Upper Quartile and lower quartile are the middle quartiles next to the Mean

Upper and lower whiskers are 1.5 IQR outside the upper/lower quartiles

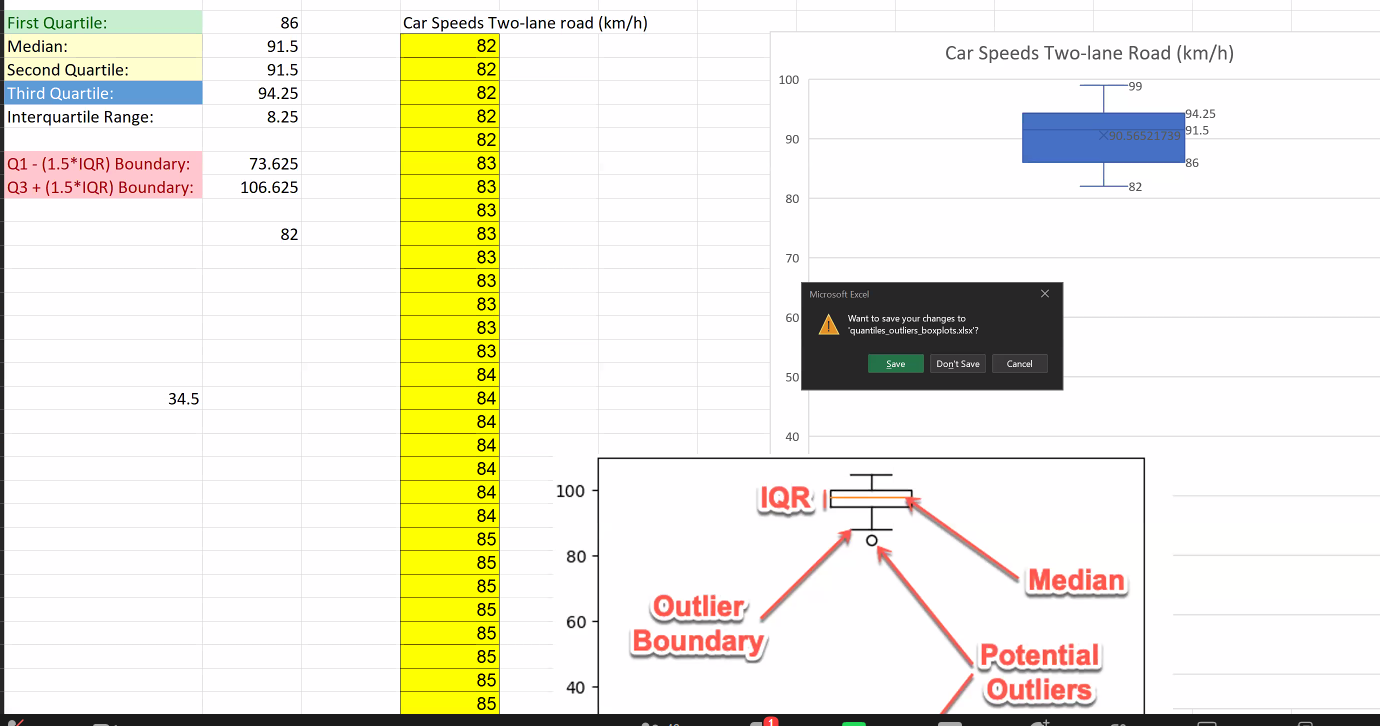


IQR Inter-Quartile Range= Q3-Q1

Whiskers= Q3+IQR or Q1-IQR

Anything outside of the whiskers are outliers

https://www.scribbr.com/statistics/interquartile-range/



Azure DevOps is like a super secure GitHub.

If something is public in GitHub, It’s fair game

Python is opensource, so the community is very supportive, and will bully anyone who will try to make any python code proprietary.

