

Welcome

Friday, September 13, 2019 6:54 PM

Class Objectives

- Students will be able to make, modify, and style bar charts, line graphs, pie graphs, and scatter plots.
- Students will understand how to filter data using Excel.
- Students will know how to calculate moving averages and regressions.

Adding files to Github

Friday, September 13, 2019 6:58 PM

About Github

- Offers a centralized locations where developers can push and pull (upload and download) their code
- Always holds the most up-to-date code and files, handling everyone's updates appropriately
- For now, students will only need to know how to use the GUI for GitHub in order to submit homework. However, In a couple of weeks, students will learn to work with Github through the terminal using git

Submitting Homework through GitHub

- Go to [www.Github.com](https://www.github.com)
- Login into your personal account (or create one)
- Create a new repository with an initialized README.md file
- **Every repository should have a "README" file that explains what the repository contains**

Create a new repository

A repository contains all the files for your project, including the revision history.

Owner / Repository name 

Great repository names are short and memorable. Need inspiration? How about [literate-garbanzo](#).

Description (optional)

 **Public**
Anyone can see this repository. You choose who can commit.

 **Private**
You choose who can see and commit to this repository.

Initialize this repository with a README
This will let you immediately clone the repository to your computer. Skip this step if you're importing an existing repository.

Add .gitignore: **None** | Add a license: **None** 

Create repository

- Create a blank excel file and save
- Go back to GitHub repo and click "Upload files"

1 commit

1 branch

0 releases

1 contributor

Branch: master ▾ New pull request

Create new file Upload files Find file Clone or download ▾

- Select "choose your files" or Drag & Drop

Drag files here to add them to your repository
Or [choose your files](#)

Commit changes

Add files via upload

Add an optional extended description...

- Add commit comment and select "Commit changes", then refresh page

Commit changes

Add files via upload

Add an optional extended description...

➔ Commit directly to the `master` branch.

➔ Create a new branch for this commit and start a pull request. [Learn more about pull requests.](#)

[Commit changes](#) [Cancel](#)

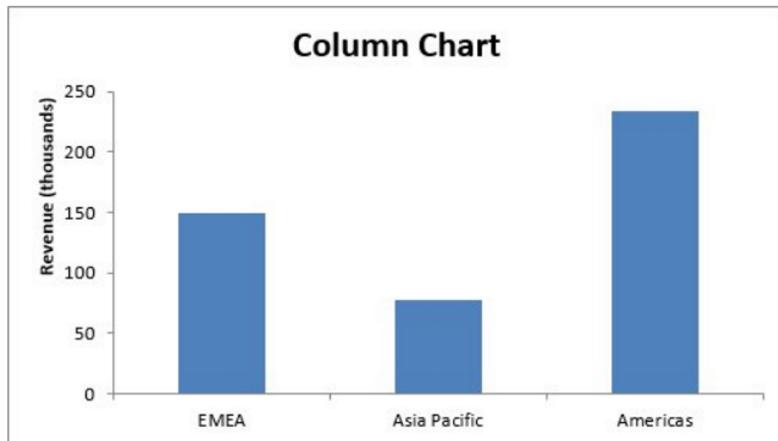
- This will be how you will be submitting the homework. Students will add all of the necessary files to their Github repo then submit the repository link to BCS.

Basic Charting

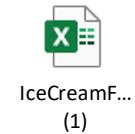
Friday, September 13, 2019 8:09 PM

Column Charts

Column Chart: Column charts are typically used to compare several items in a specific range of values. Column charts are ideal if you need to compare a single category of data between individual sub-items, such as, for example, when comparing revenue between regions.



Flavor	Number of Faves	Percentage of Total
Butter Pecan	96	16%
Mint	94	16%
Rainbow Sherbert	79	13%
Vanilla	67	11%
Strawberry	64	11%
Rocky Road	57	10%
Chocolate	53	9%
Cookie Dough	41	7%
Coffee	34	6%
Neapolitan	15	3%

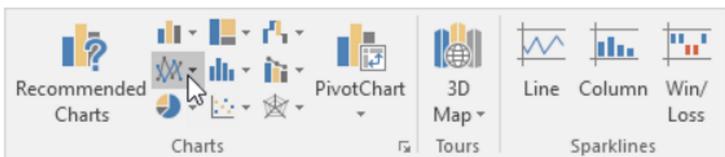


IceCreamF...
(1)

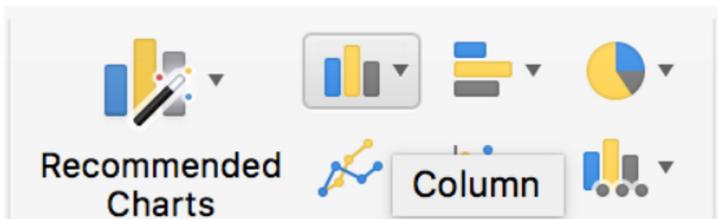
- Select columns A and B
- Go to the insert tab

*Different for PC and Mac

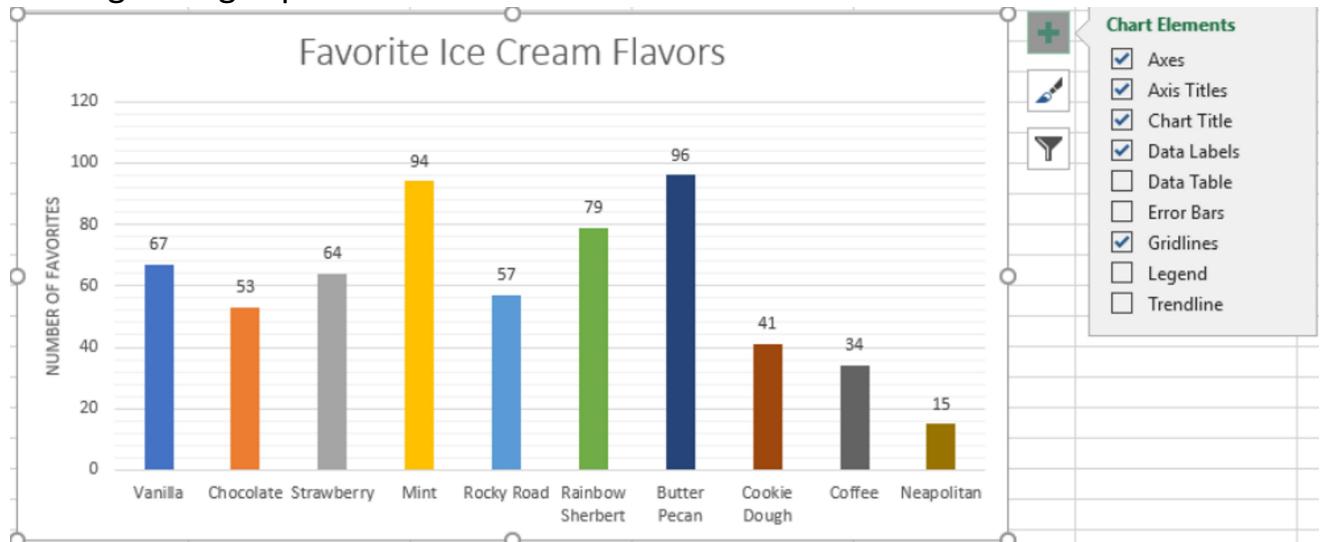
PC



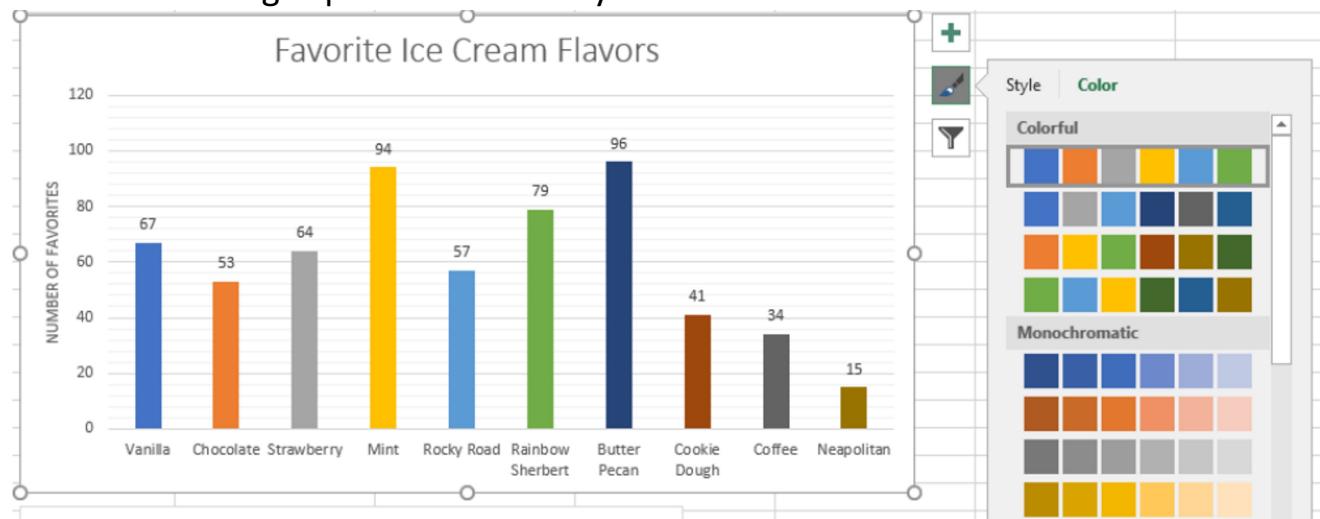
Mac



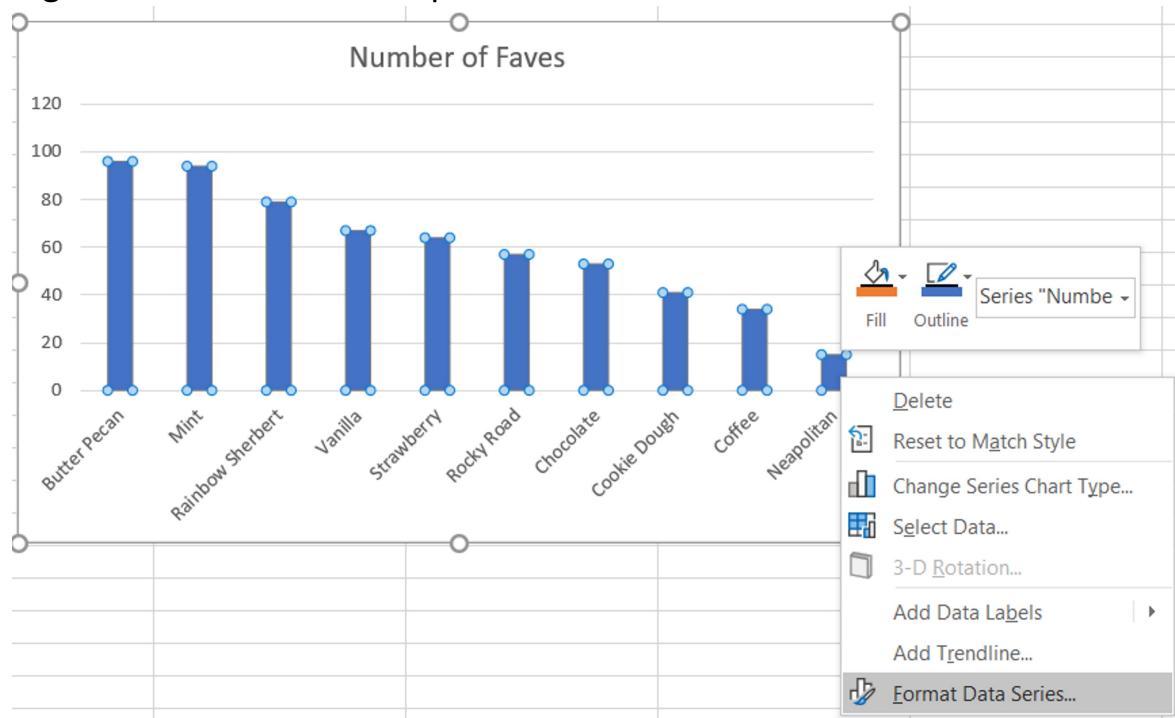
- Select a Column Chart
- For PC
 - Plus sign brings up a list of elements that can be added or remove



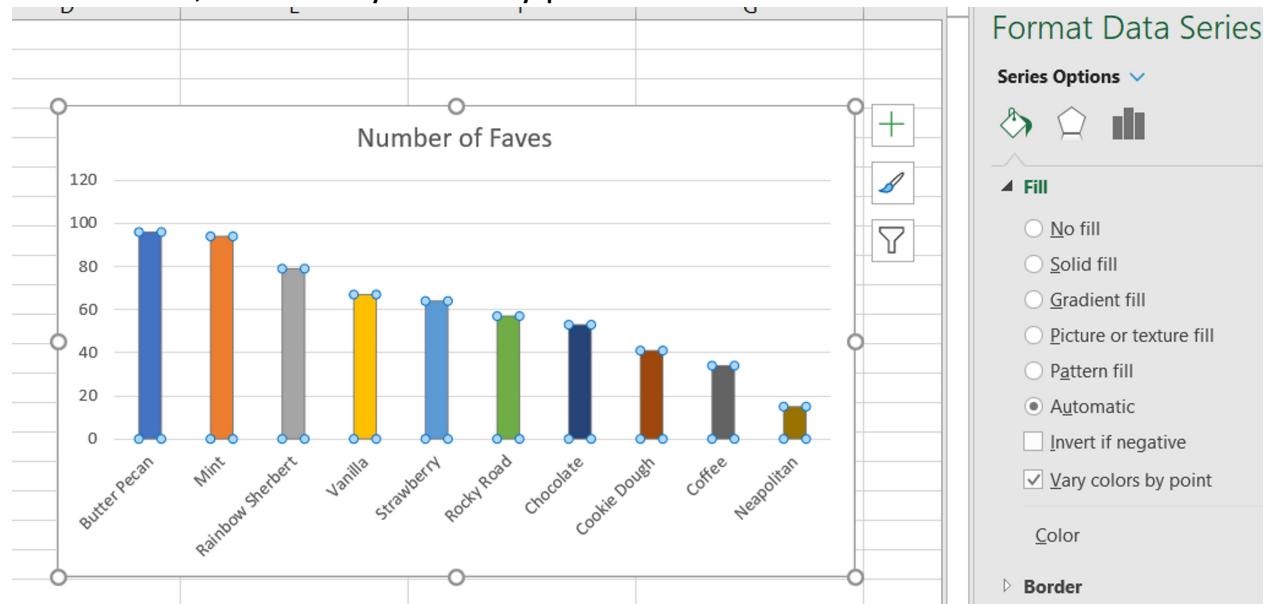
- Paint brush brings up list of color for your chart



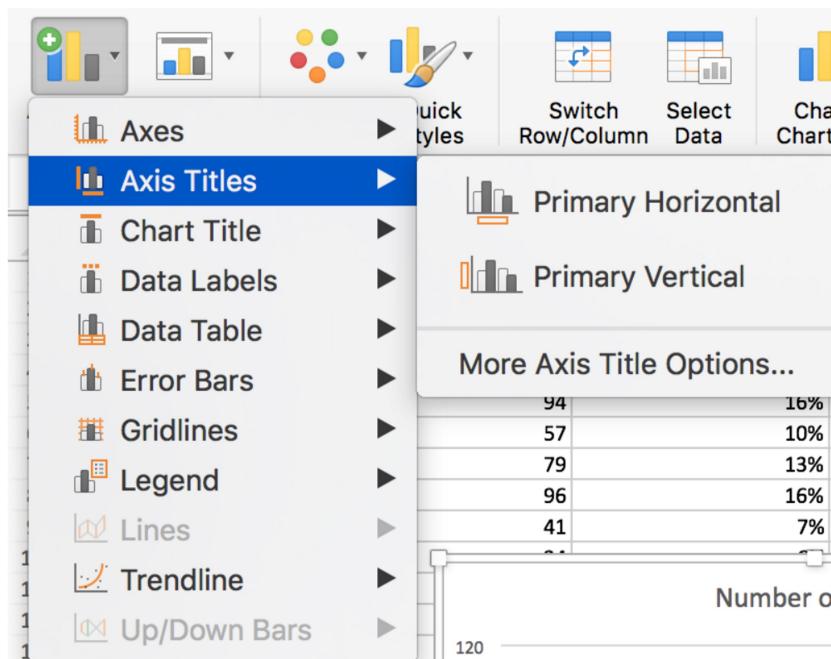
- What if all my data looks the same?
- Right click one of the data points and select "Format Data Series"



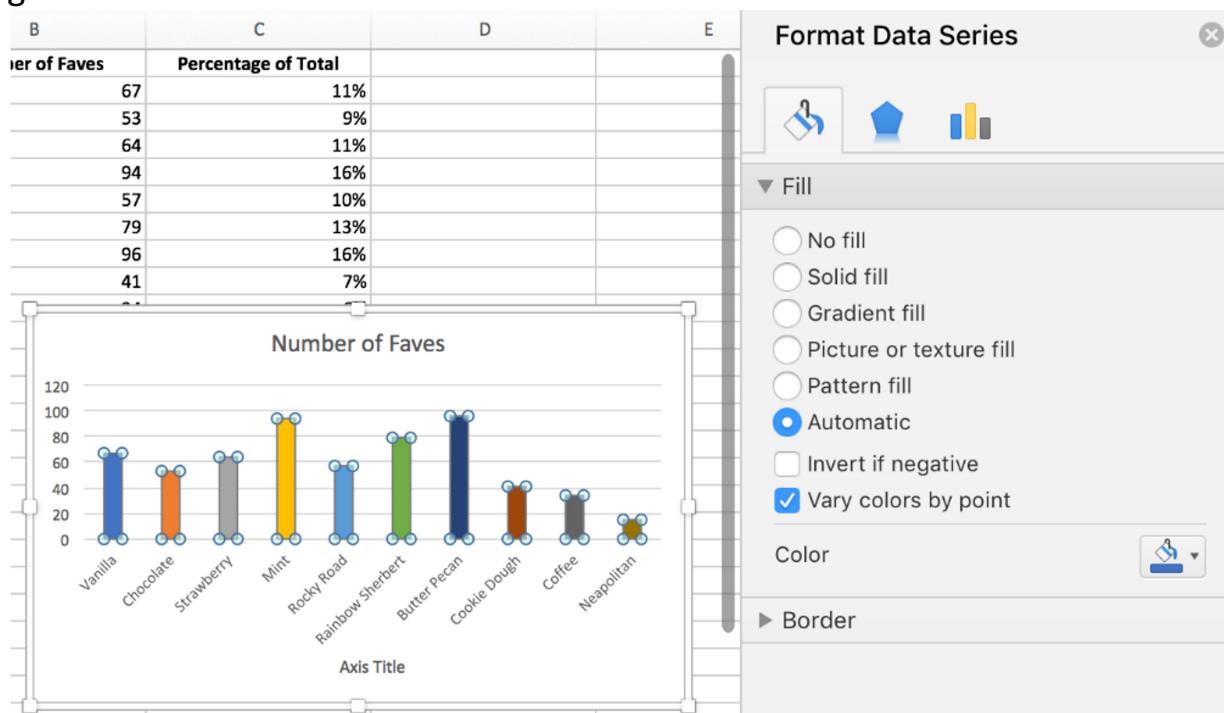
- Under "Fill", have "Very colors by point" checked



- For Mac
 - To add Axis Titles (Horizontal or Vertical), Click on "Add Chart Element" on the left side of the ribbon and moved to "Axis Titles"

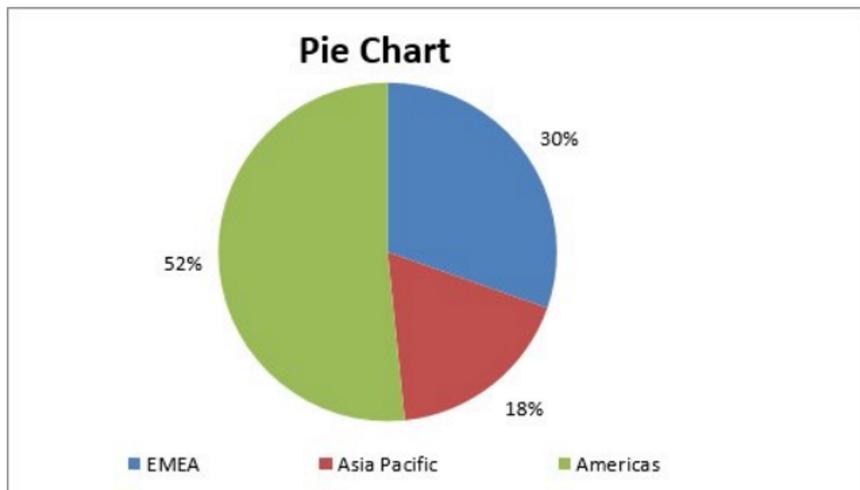


- Click on "Change Colors" on the ribbon to changes the colors of the bar graph (to the right of the "Add Chart Element").
- Double click on any of the bars to bring up the "Format Data Series" menu. Then select "Vary colors by point" option that will give each bar a different color.



Pie Charts

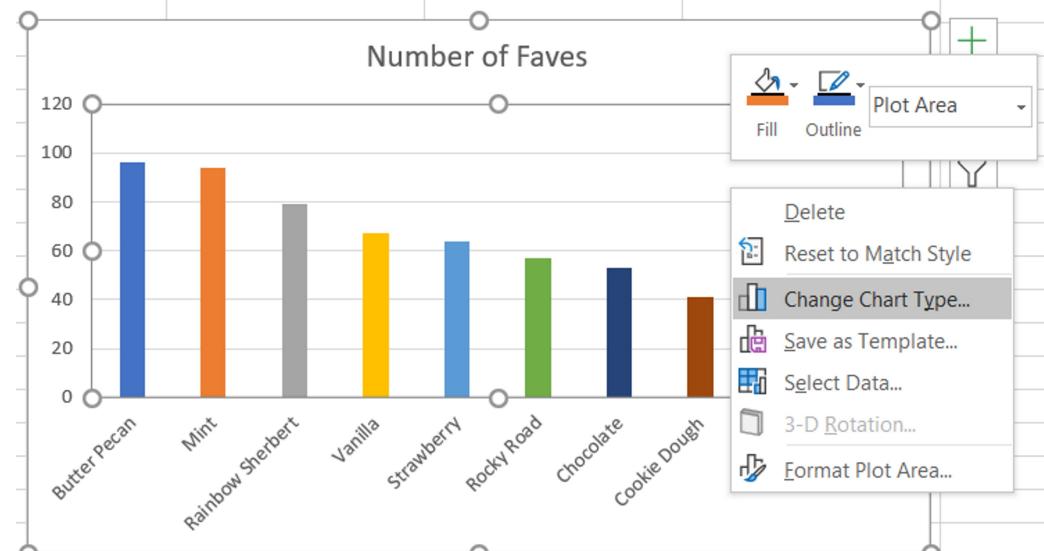
Pie Chart: Another frequently used chart is the old pie chart. A pie chart represents the distribution or proportion of each data item over a total value (represented by the overall pie). A pie chart is most effective when plotting no more than three categories of data.



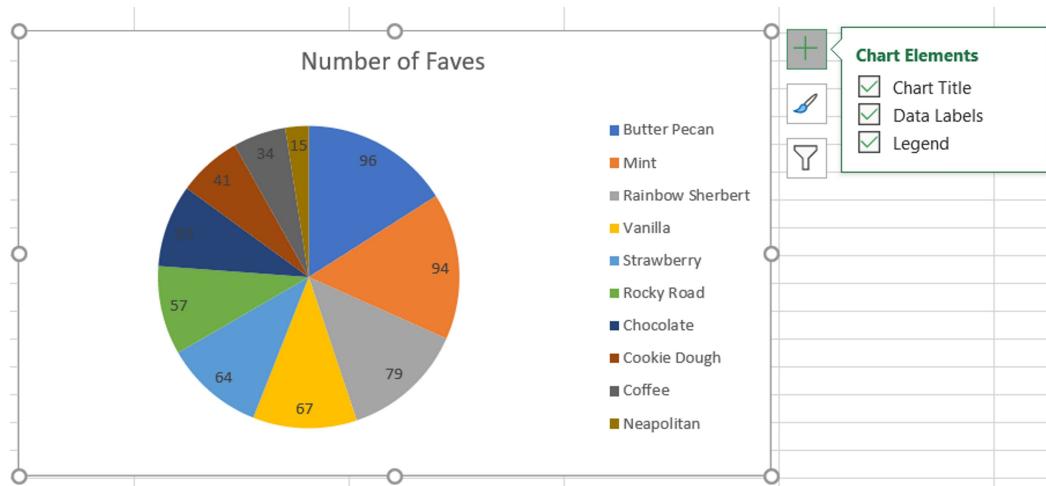
- Converting a Bar Chart to Pie Chart

- PC

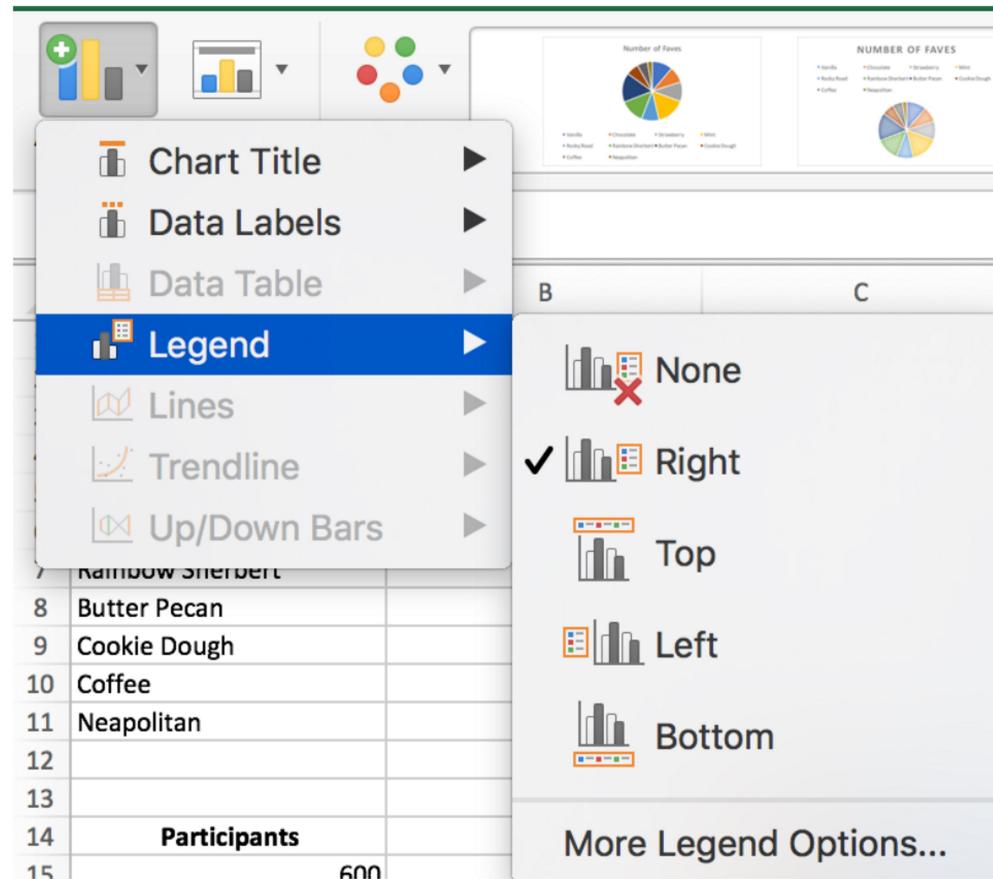
- Right-click chart and selecting "Change chart type"



- Select Pie Chart
 - Add Legend

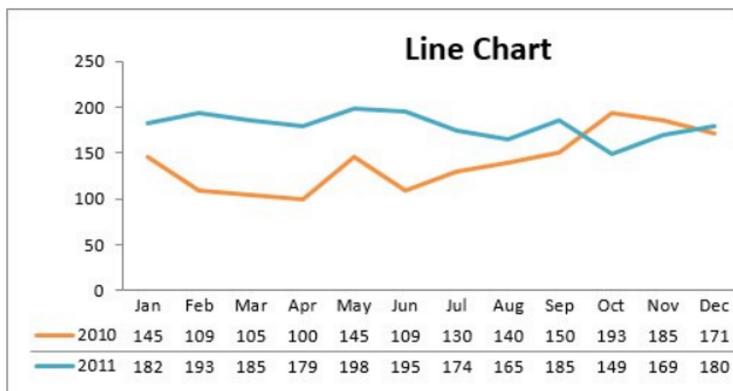


- Alternate way.....selecting the chart, go to the Design tab's Type group, and clicking "Change Chart Type"
- Mac
 - Similar to PC, adding legend is different
 - Clicking on "Add Chart Element" on the ribbon, select add "Legend" and the location.



Line Charts

Line Chart: The line chart is one of the most frequently used chart types, typically used to show trends over a period of time. If you need to chart changes over time, consider using a line chart.

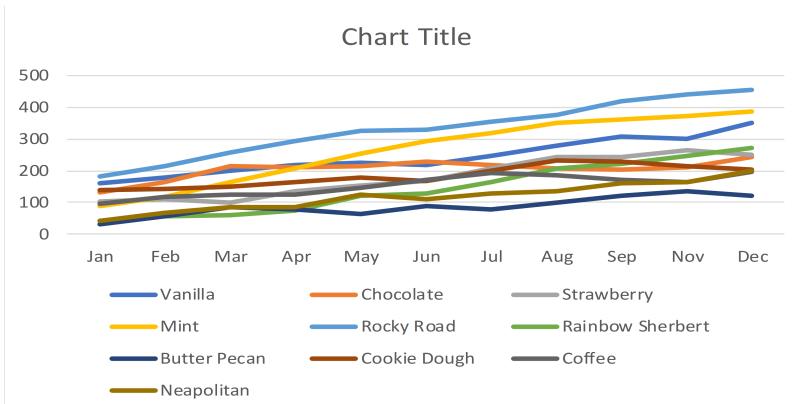


BasicCharts
(1)

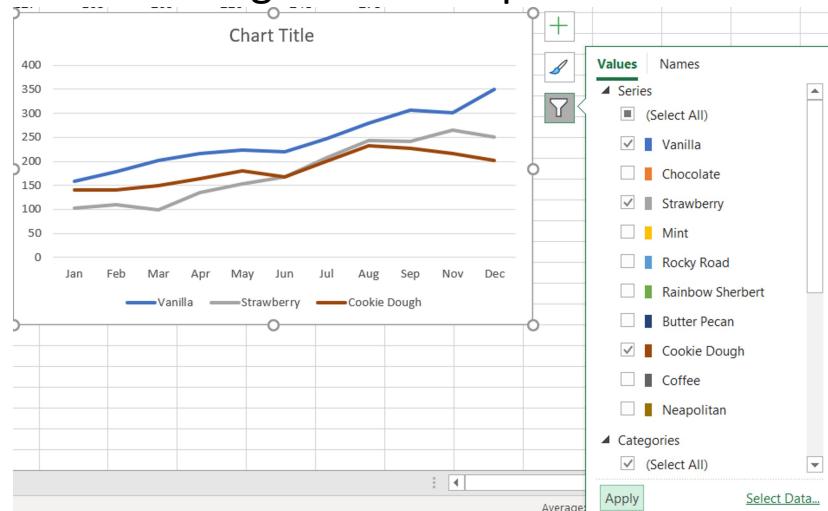
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Nov	Dec
Vanilla	159	178	201	217	224	219	247	280	306	301	350
Chocolate	133	165	215	212	216	227	217	208	203	210	243
Strawberry	103	109	99	135	153	168	207	243	242	265	251
Mint	88	116	163	208	254	292	318	350	360	373	386
Rocky Road	182	214	258	295	327	328	353	374	418	441	453
Rainbow Sherbert	42	55	62	73	121	127	163	209	220	248	273
Butter Pecan	32	55	87	80	65	90	77	99	120	137	123
Cookie Dough	140	141	149	163	180	167	200	232	228	216	202
Coffee	95	118	124	126	148	171	193	185	172	165	198
Neapolitan	44	69	85	85	124	110	130	137	159	165	201

- Select all of the data and choose a 2D line chart from the Charts group on the Insert tab

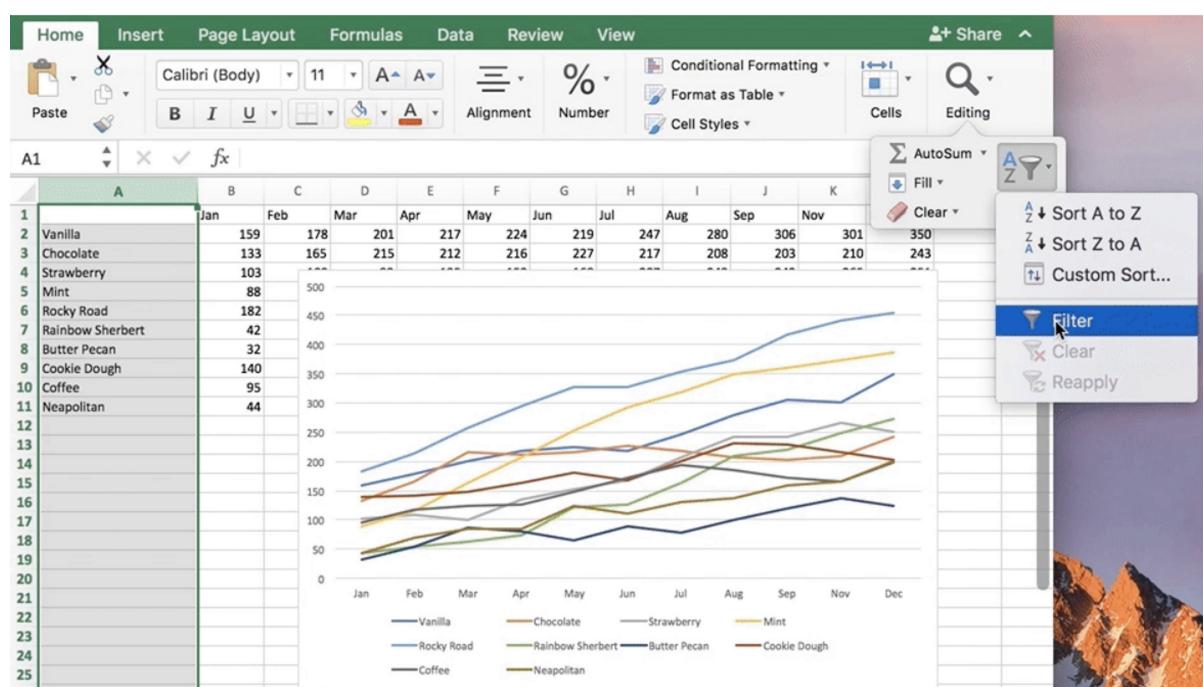
	A	B	C	D	E	F	G	H	I	J	
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug		
2	Vanilla		159	178	201	217	224	219	247	280	
3	Chocolate		133	165	215	212	216	227	217	208	
4	Strawberry		103	109	99	135	153	168	207	243	
5	Mint		88	116	163	208	254	292	318	350	
6	Rocky Road		182	214	258	295	327	328	353	374	
7	Rainbow Sherbert		42	55	62	73	121	127	163	209	



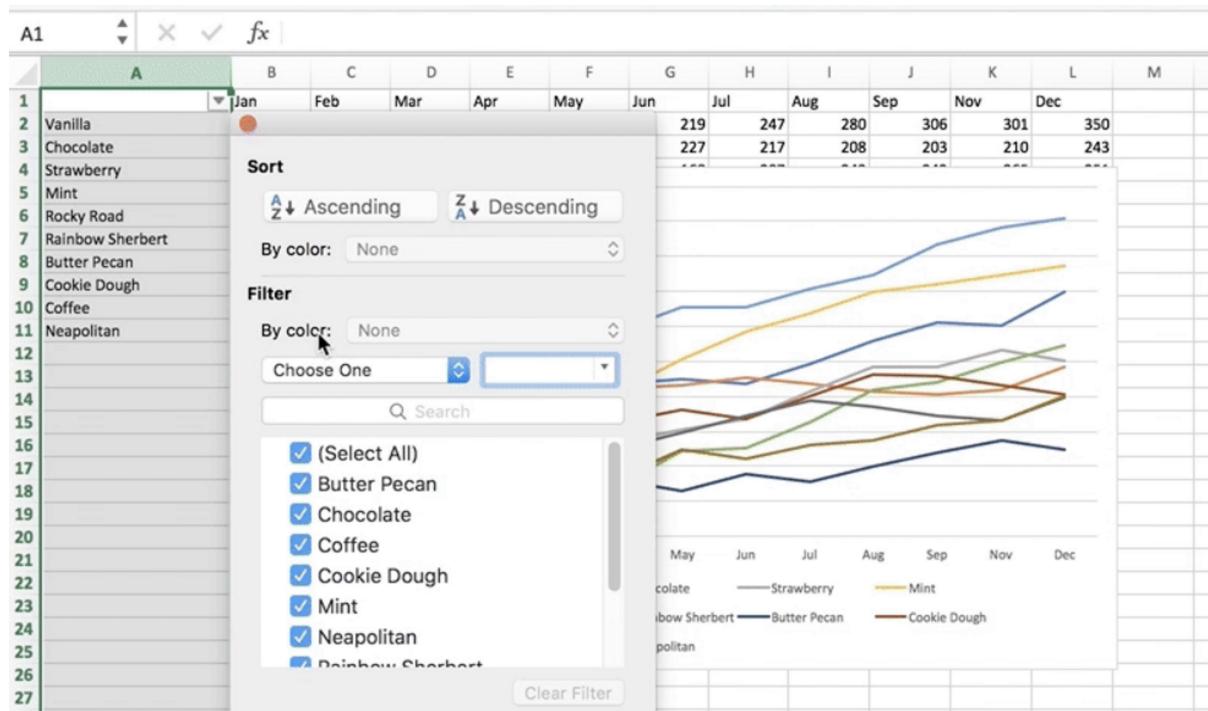
- Use filter to get a clearer picture of the data



- Filter for Mac,
 - Choose Home tab
 - Select Column A
 - Click Sort & Filter in the toolbar



- Once column is on filter mode, click arrow dropdown in its header cell
- Will see options for sorting and filtering



Exercise: The Line and Bar Grades

Friday, September 13, 2019 10:15 PM

- Files Needed: 03-Stu_LineAndBar/README.md and 03-Stu_LineAndBar/StudentGrades_Unsolved.xlsx
- Time: 15 minutes

Line and Bar Grades



StudentGr...

You are going to take the role of a teacher upon yourself for this activity as you create a series of bar and line graphs that visualize the grades of your class over the course of a semester.

Instructions

- Create a series of bar graphs that visualize the grades of all students in the class, one graph for every month.
- Create a line graph using all of the data that can be used to compare students' grades across the semester.
 - Use filtering in the line graph to allow you to drill down to a specific student's progress throughout the semester.

Hint

- When duplicating bar graphs, it pays to get the formatting and look of the chart where you want it for the first graph (e.g. for January), and to then copy that chart and re-select the data for the subsequent copies (keeping the style and format, but just changing the data).
- "Switch Row/Column" from the Chart Design menu on the line graph.

Scatter Plots and Trend Lines

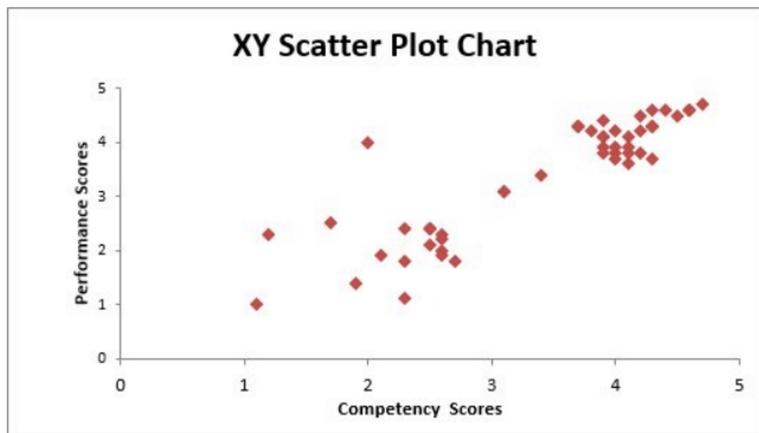
Friday, September 13, 2019 10:24 PM

Scatter Plots

XY Scatter Plot Chart: Scatter charts in Excel (also known as XY scatter plot charts) are excellent for showing correlations between two sets of values. For example an XY scatter plot can be used to illustrate the correlation between employee performance and competency, demonstrating that employee performance rises as competency improves. The x and y axes work together to represent data plots on the chart based on the intersection of x values and y values.



ScatterPlot



- Scatter plots become even more helpful when we then apply a trend line to them. This allows us to further visualize the positive or negative relationship between our variables by plotting a line which best fits our data

Exercise: Video Game Sales

Friday, September 13, 2019 11:45 PM

Partner Up!!!

- Files Needed: 05-Par_GameSales/README.md and VideoGameSales_Unsolved.xlsx
- Time: 15 minutes



VideoGam...

Game Sales

Looking into a product's sales within a region and comparing it to that product's sales worldwide is a great way to determine how important a region is to a company. In this activity, you will pair up with one of your classmates in order to create a series of scatter plots which will compare video game sales across regions.

Instructions

1. Create a scatter plot that compares the NA (North American) sales of games versus the global sales of games. Make sure to add in axis titles, a chart title, and a trend line.
2. Create a scatter plot that compares the EU (European) sales of games versus the global sales of games. Make sure to add in axis titles, a chart title, and a trend line.
3. Create a scatter plot that compares the JP (Japanese) sales of games versus the global sales of games. Make sure to add in axis titles, a chart title, and a trend line.
4. Create a scatter plot that compares other sales of games versus the global sales of games. Make sure to add in axis titles, a chart title, and a trend line.
5. Go back into each of your charts and modify the axes so that they are consistent for each chart.
 - Without consistency of margins between your charts they could be considered misleading.

Hints

- Use ctrl button for non-adjacent columns

NA_Sales													
	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Platform	Year_of_Release	Genre	Publisher	Critic_Score	Critic_Count	User_Rating	Sales	Global_Sales	NA_Sales	EU_Sales	JP_Sales	America_Sales
2	Wii	2006	Sports	Nintendo	76	51	7.6	82.53	41.36	28.96	3.58	2.26	9.14
3	NES	1985	Platform	Nintendo	82	73	7.3	40.24	29.08	12.76	3.58	10.93	11.27
4	Wii	2008	Racing	Nintendo	80	73	7.3	35.52	15.68	12.76	2.26	8.89	10.95
5	Wii	2009	Sports	Nintendo	89	65	7.7	32.77	15.61	10.93	2.26	9.71	7.47
6	GB	1996	Role-Playing	Nintendo	58	41	6.6	31.37	11.27	8.89	9.14	23.2	6.94
7	GB	1989	Puzzle	Nintendo	87	80	8.6	30.26	28.92	13.96	9.05	24.67	26.93
8	DS	2006	Platform	Nintendo	91	64	8.6	594	28.32	14.44	9.71	23.21	26.03
9	Wii	2006	Misc	Nintendo	62	77	7.7	464	23.1	9	6.18	22.71	22.71
10	Wii	2009	Platform	Nintendo	77	116	8.6	22.71	9.05	10.95	9.71	23.21	26.03
11	NES	1984	Shooter	Nintendo	77	116	7.7	22.71	9.71	7.47	23.21	26.03	26.93
12	DS	2005	Simulation	Nintendo	77	116	7.7	22.71	9.05	10.95	23.21	26.03	26.93
13	DS	2005	Racing	Nintendo	77	116	7.7	22.71	9.71	7.47	23.21	26.03	26.93
14	GB	1999	Role-Playing	Nintendo	77	116	7.7	22.71	9.05	10.95	23.21	26.03	26.93
15	Wii	2007	Sports	Nintendo	77	116	7.7	22.71	9.71	7.47	23.21	26.03	26.93

Break (40 Minutes)

Saturday, September 14, 2019 8:10 AM



KEEP
CALM
IT'S
BREAK
TIME

The Need to Filter & Pivot Charts

Saturday, September 14, 2019 1:20 AM

- The previous activity included far more information that we could have utilized.
- Included categories, like "Publisher", which we could use in order to look into the sales for specific companies.
- We can filter data in a spreadsheet and then make a chart out of that data



PigeonRaci...

Pivot Charts

◦ PC

- Pivot charts operate in much the same way as pivot tables do, allowing users to aggregate data of similar types and then create visualizations for them.
- To create a pivot chart, simply navigate into the Charts group of the Insert tab and select "Pivot Chart" from the options available. Once that is done, simply set up the pivot table you desire and a chart will be created based off of it.



Mac

- First create a pivot table using Sex as our row values. "POS" and "Speed" as value.
- Click the ! next to them and switch to **Max** and this will create our pivot table.
- The result should look as follows

PivotTable Builder

FIELD NAME

Search fields

- Arrival
- Speed
- To Win

Filters

Columns

Values

Rows

Values

: Sex (i)

: Max of Pos (i)

: Max of Speed (i)

Drag fields between areas

The screenshot shows the Microsoft Excel PivotTable Builder dialog box. At the top, there's a 'FIELD NAME' section with a search bar. Below it is a list of fields: 'Arrival', 'Speed' (which is checked), and 'To Win'. The 'Filters' section is empty. The 'Columns' section also has an empty 'Values' area. The 'Rows' section contains the field 'Sex'. The 'Values' section contains two items: ': Max of Pos' and ': Max of Speed', both of which are highlighted with a blue border. A general instruction at the bottom says 'Drag fields between areas'.

- Locate "Insert" on the ribbon add any recommended chart to create a Pivot Chart.
- Now when you play around with the filters in our pivot table, the chart will adjust.

Exercise: Filter Game Sales

Saturday, September 14, 2019 1:20 AM

Partner Up!!!

- Files Needed: 07-Par_FilterGamesSales/README.md and VideoGameSales2_Unsolved.xlsx
- Time: 15 minutes
- Only do problems 1 & 2



VideoGam...

Filtering Game Sales

Now that we know how to apply filters to a spreadsheet and create charts based on filtered data sets, we'll take some time to create charts which compare the sales of publishers against one another.

Instructions

1. Create a scatter plot which graphs the critical response (Critic Score) of games published by Nintendo as compared to their global sales.
 - Only chart those games that have been reviewed. Games without any reviews should be ignored.
 - Add a chart title, axis titles, and a trend line to the graph that is created.
 - Copy your chart and paste it into an external program - paint, Microsoft Word, etc - before moving onto making the next chart.
2. Create a scatter plot which graphs the critical response of games published by Electronic Arts as compared to their global sales.
 - Only chart those games that have been reviewed. Games without any reviews should be ignored.
 - Add a chart title, axis titles, and a trend line to the graph that is created.
 - Copy your chart and paste it into an external program - paint, Microsoft Word, etc - before moving onto making the next chart.
3. Select all of the data on the worksheet and create a line chart which can be filtered by publisher, whose rows are set by a game's year of release, and whose values are the sum of global sales for that year.
 - Create a 2D line graph that charts this data.

Moving Averages

Saturday, September 14, 2019 1:58 AM

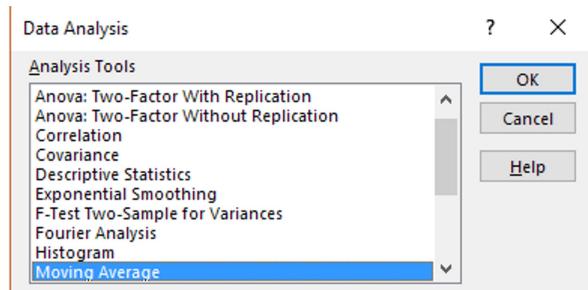
- When creating line graphs in Excel, it is not uncommon for our charts to look rather jagged. There are times in which it would also be beneficial to show smoother progression.
- This is where the concept of a "moving average" comes into play as it can be used to smooth out irregularities in data (large positive/negative spikes) to allow for easily recognizable trends.
 - Imagine a moving average as being similar to a trend line with the only major difference being that, unlike a trend line, the line of a moving average can change direction over time.
 - Moving averages, like trend lines, should not be considered a replacement for more detailed graphs but should be added in on top of graphs to add more clarity.



MovingAv...

- In order to create moving averages, we first need to install a new add-in for Excel called the "Analysis ToolPak": go to the Tools menu and navigate to "Excel Add-Ins...". From the resulting window, select the "Analysis ToolPak".

- Navigate into the Data tab, locate the Analyze group, and select the "Data Analysis" option. Macs just have a "Data Analysis" button.
- From the menu that appears, select "Moving Average".



- Click on the arrow beside "Input Range" and select the cells that you would like to average. In this case, select B2 to M2.
- Set the interval that you would like to take the average of. We will be setting this particular interval to 2 for the time being.
 - The interval of a moving average is the number of data points that Excel will take the average of. This number includes the current data point as well. So, if we set the interval to 6, the moving average will be the average of the previous 5 data points and the current data point.
 - The higher your interval, the smoother your line will be.
- Select an output range for the averages you are calculating. In this case, select B3 to M3.
- Hit "Ok" and Excel will calculate/print the moving average according to your specifications.
 - Notice that the first cell of our range has been filled in with the value "#N/A" - meaning "Not Available". This is because there are not enough data points prior to this one to calculate an average for.

- After creating our moving average, go ahead and make a line chart out of our data to show what the final product looks like.
- Repeat for interval of four to show off how this line is even smoother on our chart.

Exercise: Golfing Targets

Saturday, September 14, 2019 2:26 AM



Golfing Tar...

- Files Needed: 09-Stu_GolfingTargets/README.md and 09-Stu_GolfingTargets/GolfingTargets_Unsolved.xlsx
- Time: 15 minutes

Golfing Targets

What better way to express moving averages than with a data set that is all about calculating the average accuracy of a golfer? We had 5 golfers take 20 shots onto a green and calculated the distance to the hole. It is up to you to find the moving averages of their shots using different intervals.

Instructions

1. Using a moving average with an interval of 2, find the average shot accuracy for each subject. Then create a 2D line graph that charts these values.
 - Copy your chart and paste it into an external program - Paint, Microsoft Word, etc. - before moving onto making the next chart.
2. Using a moving average with an interval of 4, find the average shot accuracy for each subject. Then create a 2D line graph that charts these values.
 - Copy your chart and paste it into an external program - Paint, Microsoft Word, etc. - before moving onto making the next chart.

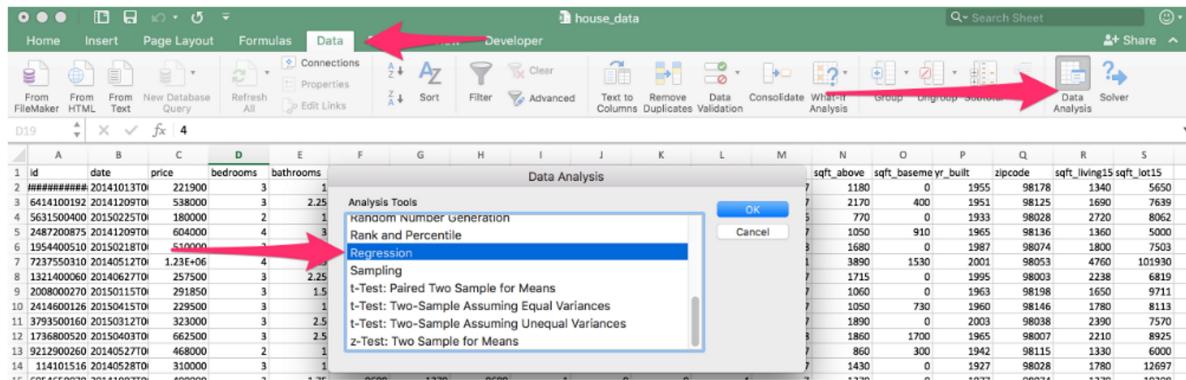
Linear Regression

Saturday, September 14, 2019 2:46 AM

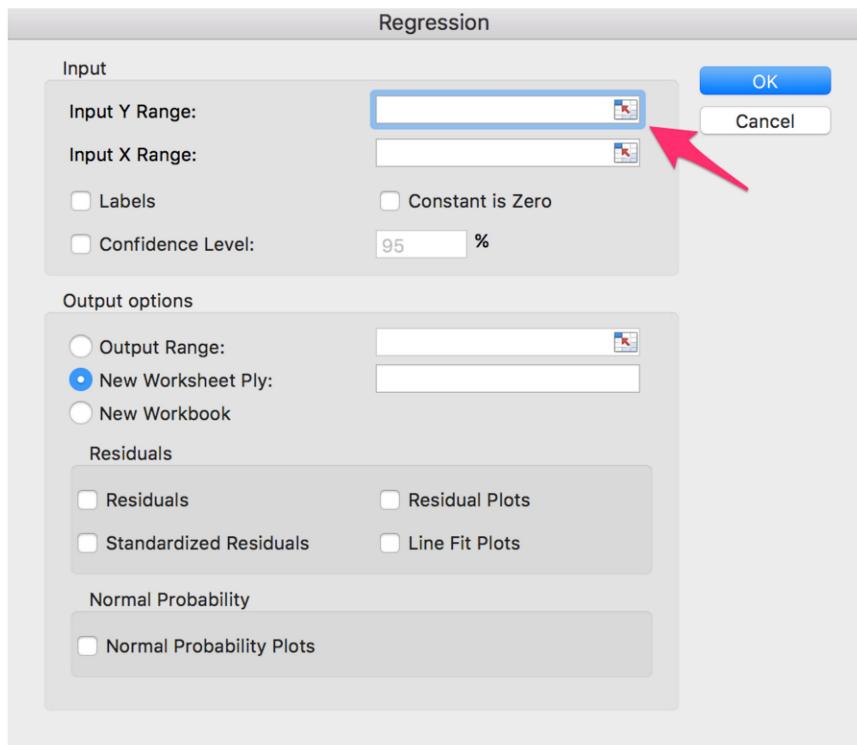


house_data

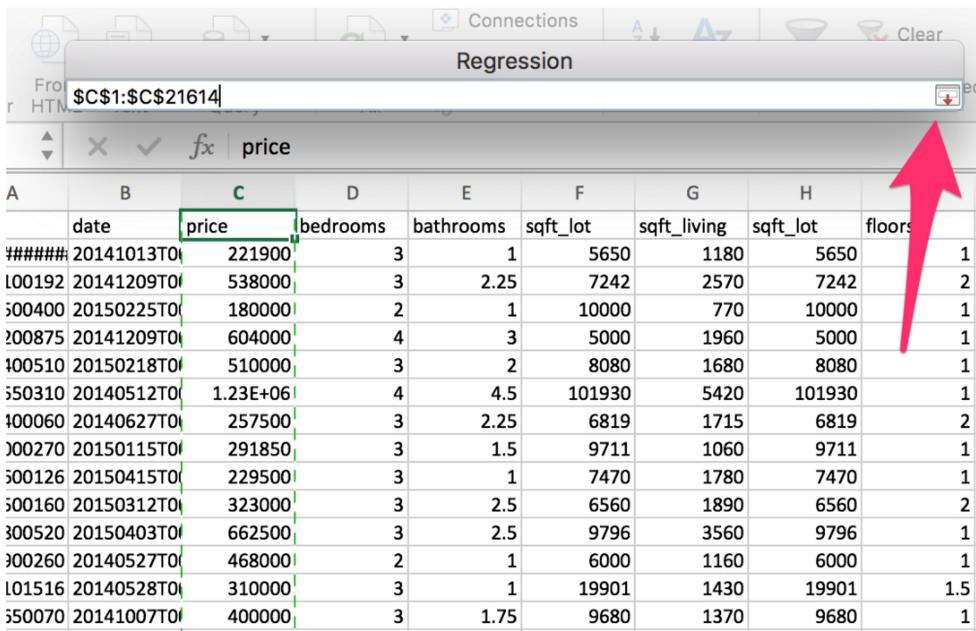
- Linear regression is relationship between a dependent variable and one or more explanatory or independent variables.
 - For example, more ice cream gets sold when the temperature goes up. Ice cream sales are our dependent variable and the temperature is an independent variable.
 - Click on the data tab in the ribbon, then the "Data Analysis" icon and finally then scroll down to Regression.



- Explain that our **Y Range** is our dependent variable or predictor variable.
- Click the box on the right side of the input box.



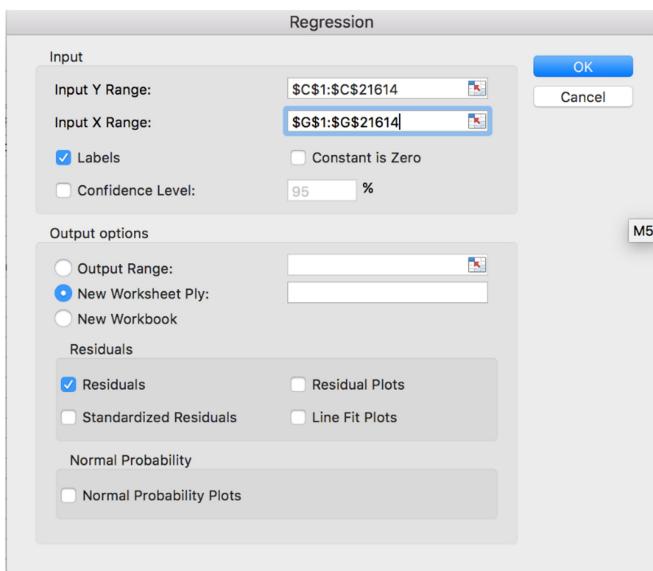
- Click "price" then hit CMD+SHIFT+DOWN (CTRL+SHIFT+DOWN on Windows) to select the entire column and then hit the down arrow in the selection box.



A	B	C	D	E	F	G	H
	date	price	bedrooms	bathrooms	sqft_lot	sqft_living	sqft_lot
#####	20141013T01	221900	3	1	5650	1180	5650
L00192	20141209T01	538000	3	2.25	7242	2570	7242
500400	20150225T01	180000	2	1	10000	770	10000
200875	20141209T01	604000	4	3	5000	1960	5000
100510	20150218T01	510000	3	2	8080	1680	8080
550310	20140512T01	1.23E+06	4	4.5	101930	5420	101930
100060	20140627T01	257500	3	2.25	6819	1715	6819
100270	20150115T01	291850	3	1.5	9711	1060	9711
500126	20150415T01	229500	3	1	7470	1780	7470
500160	20150312T01	323000	3	2.5	6560	1890	6560
300520	20150403T01	662500	3	2.5	9796	3560	9796
300260	20140527T01	468000	2	1	6000	1160	6000
L01516	20140528T01	310000	3	1	19901	1430	19901
550070	20141007T01	400000	3	1.75	9680	1370	9680

- Explain that our `X range` is our independent variable or explanatory variables.
- Click the right of the input box.
- Select "sqft_living" and highlight the whole column and return back to the screen.

Select check the boxes next to `Labels` and `Residuals`.



If you get an error about "Regression - Input range contains non-numeric data.", it means you forgot to check the "Labels" box! Alternately, you can alter your selected range to omit the header row.

- Click `OK` to display the results on a new sheet.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.702043721							
R Square	0.492865387							
Adjusted R Square	0.49284192							
Standard Error	261617.1313							
Observations	21613							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	1.43751E+15	1.43751E+15	21002.93213	0			
Residual	21611	1.47913E+15	68443523397					
Total	21612	2.91665E+15						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-43867.60153	4405.455443	-9.957563321	2.62379E-23	-52502.61916	-35232.58391	-52502.61916	-35232.58391
sqft_living	280.8066899	1.93761499	144.9238839	0	277.0088216	284.6045582	277.0088216	284.6045582

- **R Square** is how close data is to a fitted line, the closer to **1.0** the better, the value in this example is **0.49**, which is not as great as we would like it to be.
- **Significance F** and **P-values** let us tell how reliable the data is. Values under **0.05** are good, and anything higher means new independent variables should be selected. The value here is **0** so the data is likely be very reliable.
- **Coefficients** is saying that for each unit increase in price the square feet of living space increases by **280.623**.

Exercise: Regression Activity

Saturday, September 14, 2019 3:16 AM

- Files Needed: 11-Stu_NBAStats/README.md and 2017_nba_season_stats.csv
- Time: 15 minutes



NBA_regr...

Regression in the NBA

Your turn to run a regression analysis on the 2017 NBA Season.

Instructions

- Run a regression analysis on the 2017 NBA season stats.
- You will use `PTS` (points) as your Y variable, or dependent variable.
- You will determine which X variables, or independent variables, to use.
- After you have run regression, explain why your choice was or was not a good choice.
 - Do the results have a good `R Square` value?
 - Is your choice statistically reliable?
 - Explain what your coefficients mean.
- Finally produce a scatter plot and explain your overall analysis.

Video Guide/Python/End Class

Saturday, September 14, 2019 12:19 PM

- Video Guide
- Python
- End Class

