Barplot Bootcamp

Introduction

Welcome to the first data analysis and figure generation module! We are going to be tackling how to make a basic **barplot** in **excel**. **Barplots** are figures that are often used to compare **numerical variables** with **categorical variables**. Today we are going to take some data collected from 10 subjects about age, BMI and their respective gender. Take a look at our data below!

Data

	A	B	C	D	E	F	G	H
1		age	gender	BMI	smoker	fasted	Caffeine Free	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12								

Data Analysis

Exclusion Criteria

As you can see we used some exclusion criteria in our experiment to minimize the number of confounding variables that may affect BMI measurements. Elise is the subject highlighted in red and it turns out she did not show up to the lab fasted and so she must be excluded from the experiment because the extra bodyweight from her food will give an inaccurate reading for her BMI.

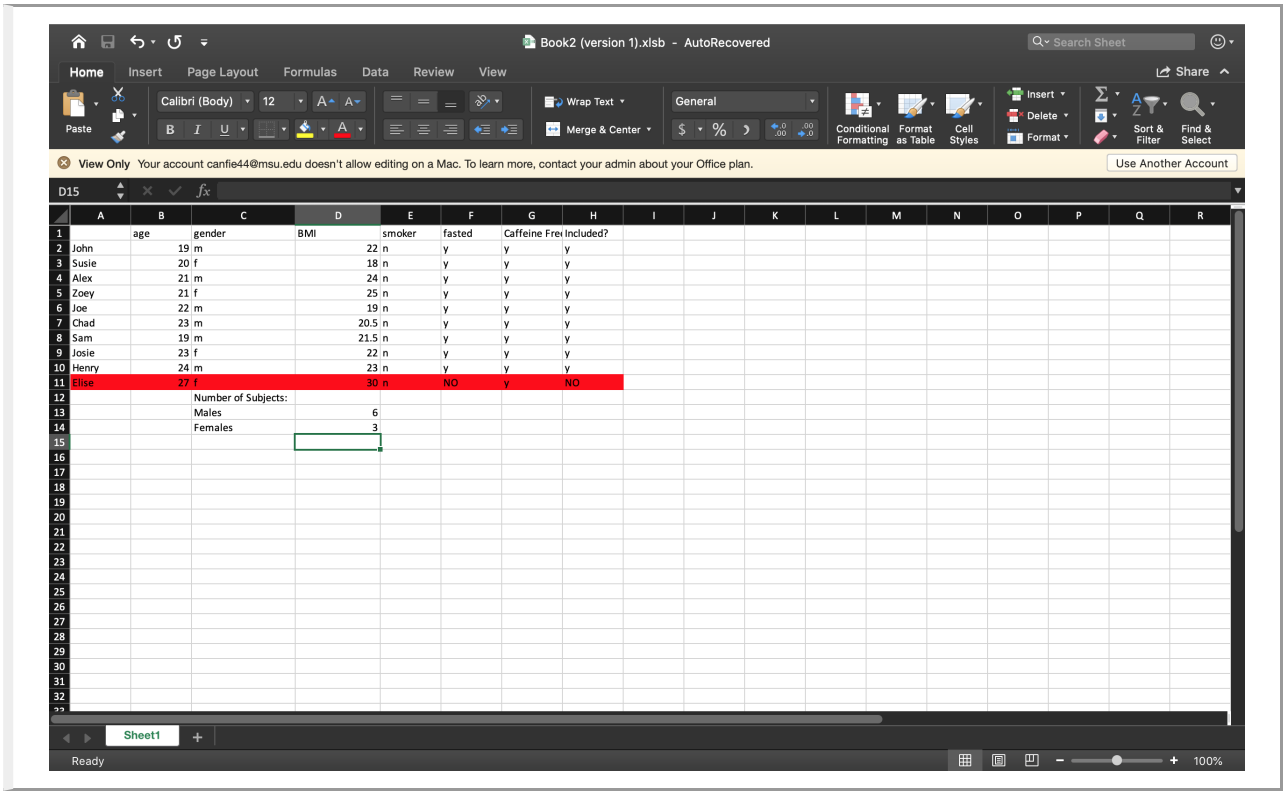
Distribution Statistics

Another aspect of our data that we are going to analyze is the distribution the data creates. Two common features of data distributions that are of particular interest are measures of **center** and **spread**. Measures of **center**, such as the **average (mean)** and **median**, tell us the most prevalent values from the distribution. Measures of **spread**, such as **standard deviation** or **standard error mean (SEM)** on the other hand tell us how much the data varies about the **center**. Both of these types of features of a distribution are heavily influened by the **sample size**, denoted *N*. As the **sample size** from a population increases in size so does our confidence that the measure of **center** represents the true value from the larger population. For example, if we have a sample size of 3 people for estimating average height and each person is 5', 6' and 7' then we would say the average is 6'. Our sample size is very small so there is a high probability that we picked very biased samples from the greater population. As we increase the number of subjects the average will converge on the true value for average height in the population.

Let's Try it!

Number of Samples or Subjects

We are going to calculate the **mean**, **standard deviation** and **SEM** for males and females independetly. We need to count the number of males and females that we have so let's look at our data.



	A	B	C	D	E	F	G	H
1		age	gender	BMI	smoker	fasted	Caffeine Fre	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				

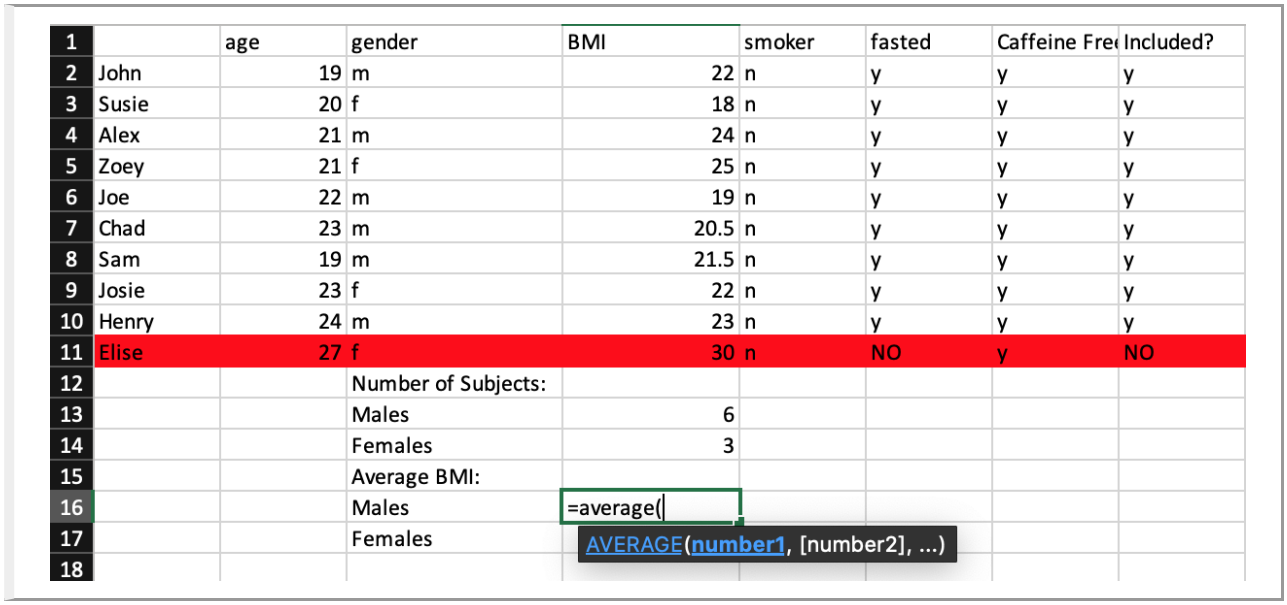
Average or Mean

As you can see in the figure there are 6 Males and 3 Females. Next we want to calculate the **mean** BMI for males and females respectively. **Excel** is handy in that it already has many of these mathematical functions built into it so that we don't have to do it by hand. In order to use a function we must select a cell and add a `=` to tell excel that we want to use a function.

	A	B	C	D	E	F	G	H
1		age	gender	BMI	smoker	fasted	Caffeine Fre	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	=				
17			Females					

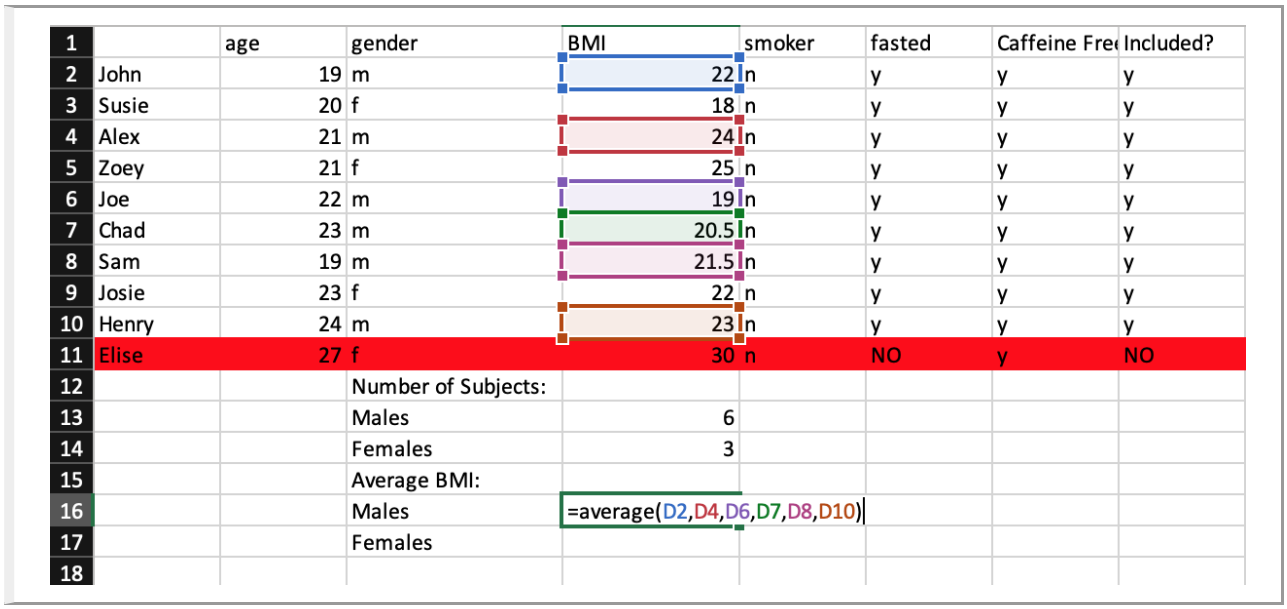
F7

To call one of the functions that already exist in **excel** we just have to start to type the its name. Let's try writing `average(` :



		age	gender	BMI	smoker	fasted	Caffeine Fre	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	=average(				
17			Females	AVERAGE(number1, [number2], ...)				
18								

As you can see **excel** will try to make suggestions for functions it already has built in as you complete the word. Now, we want to calculate the **average** for males exclusively so we can't just select the whole column of BMI data. Using your cursor, while holding down Ctrl (*Windows*) or Command (*MacOS*), and click each of the BMI data points associated with males. When you have selected all of the data make sure you that you close the function parentheses so that it looks like this: `=average(data1, data2, data3...)` . Click enter to calculate the average! After you have done that we will repeat the same steps for females to get their average BMI as well.



		age	gender	BMI	smoker	fasted	Caffeine Fre	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	=average(D2,D4,D6,D7,D8,D10)				
17			Females					
18								

		age	gender	BMI	smoker	fasted	Caffeine Fre	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	21.66666667				
17			Females	21.66666667				
18								
19								

Standard Deviation (STDEV)

Now we need a measure of **spread**! Let's calcualte the **standard deviation** using the same principles. Start by selecting a cell to add the function `=STDEV(` and using *Ctrl* or *Command* to select the male and female BMI data respectively.

		age	gender	BMI	smoker	fasted	Caffeine Fre	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	21.66666667				
17			Females	21.66666667				
18			Standard Deviation BMI:					
19			Males	1.779513042				
20			Females	=stdev(D3,D5,D9)				
21								

Standard Error Mean (SEM)

Lastly, let's calculate the **SEM** as another option for an error bar that we can use in the **Barplot**. The formua for caluclating **SEM** is the **standard deviation** divided by the **sqare root** of **N**, where **N** is the **number of subjects or samples**. In **excel** we can write this by selecting the male and female cell for **standard deviation** and **number of samples** from previous steps by clicking on them. Based on the figure above the function should look like this: `=D19/SQRT(D13)` . Give this a try and do the same for females again.

		age	gender	BMI	smoker	fasted	Caffeine Free	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	21.66666667				
17			Females	21.66666667				
18			Standard Deviation BMI:					
19			Males	1.779513042				
20			Females	3.511884584				
21			SEM BMI:					
22			Males	=D19/sqrt(D13)				
23			Females					
24								

		age	gender	BMI	smoker	fasted	Caffeine Free	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	21.66666667				
17			Females	21.66666667				
18			Standard Deviation BMI:					
19			Males	1.779513042				
20			Females	3.511884584				
21			SEM BMI:					
22			Males	0.726483157				
23			Females	2.02758751				
24								
25								

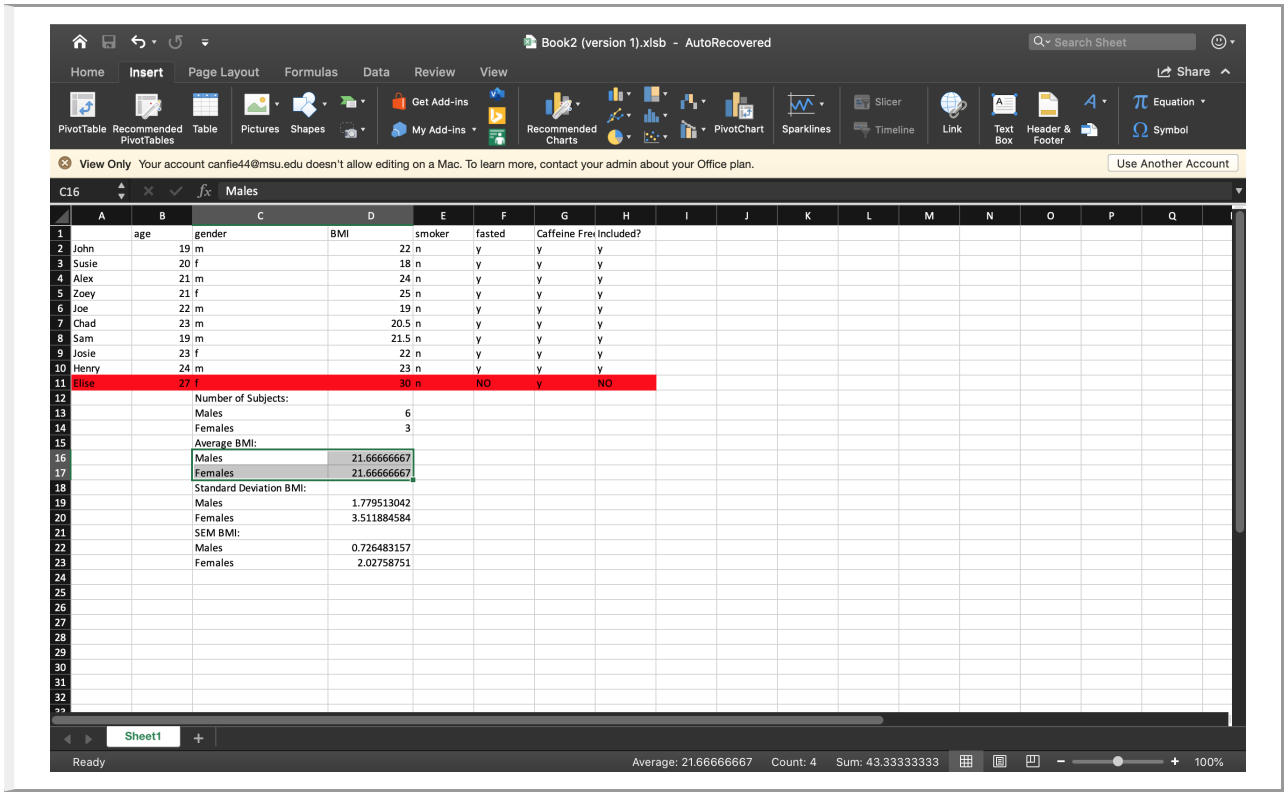
Awesome job! Now we have all of the data that we need to make a **Barplot** that shows the relationship between BMI and gender. Time to make a great figure!

Making the Barplot

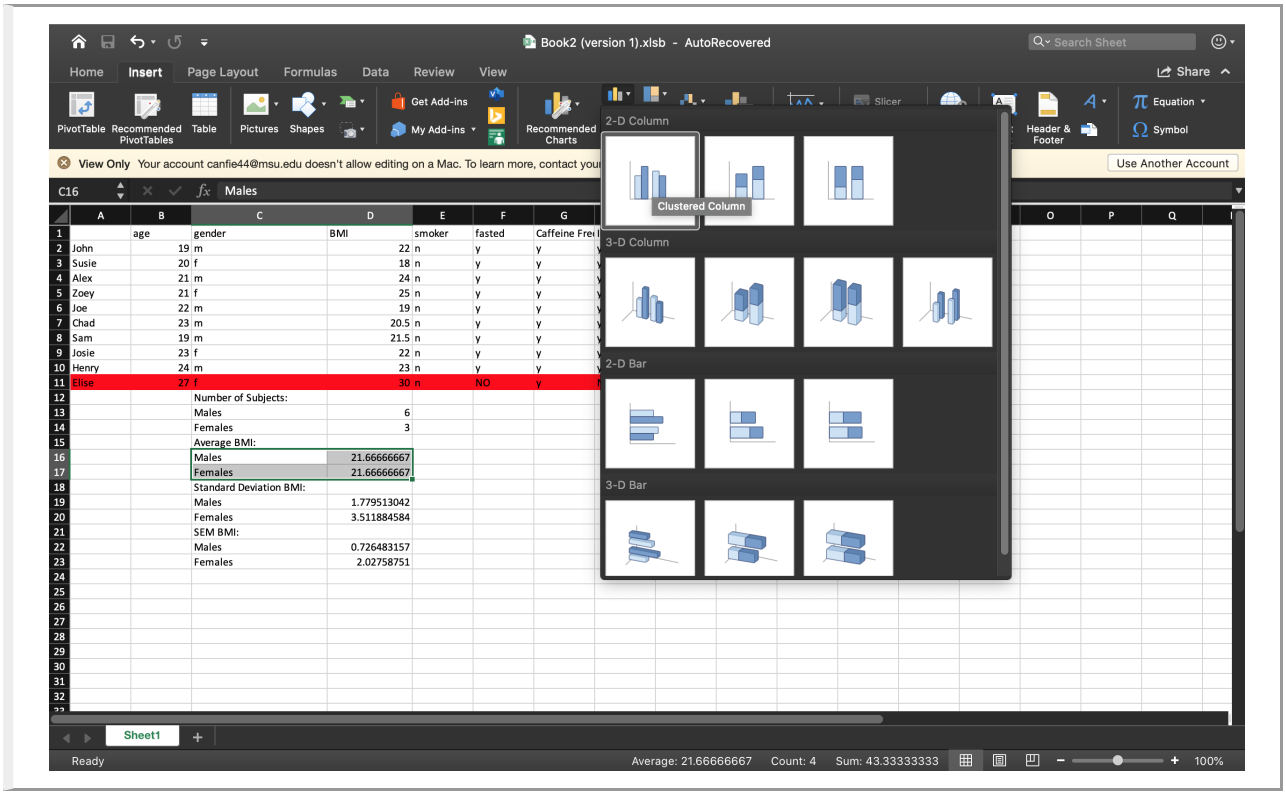
Alright, now we have generated some useful statistics about our data but we want to visualize this in a simple and clean fashion. We are going to use a **barplot** to do this! Let's start by selecting our data about **average** BMI for both males and females like such:

		age	gender	BMI	smoker	fasted	Caffeine Fre	Included?
2	John	19	m	22	n	y	y	y
3	Susie	20	f	18	n	y	y	y
4	Alex	21	m	24	n	y	y	y
5	Zoey	21	f	25	n	y	y	y
6	Joe	22	m	19	n	y	y	y
7	Chad	23	m	20.5	n	y	y	y
8	Sam	19	m	21.5	n	y	y	y
9	Josie	23	f	22	n	y	y	y
10	Henry	24	m	23	n	y	y	y
11	Elise	27	f	30	n	NO	y	NO
12			Number of Subjects:					
13			Males	6				
14			Females	3				
15			Average BMI:					
16			Males	21.66666667				
17			Females	21.66666667				
18			Standard Deviation BMI:					
19			Males	1.779513042				
20			Females	3.511884584				
21			SEM BMI:					
22			Males	0.726483157				
23			Females	2.02758751				
24								
25								

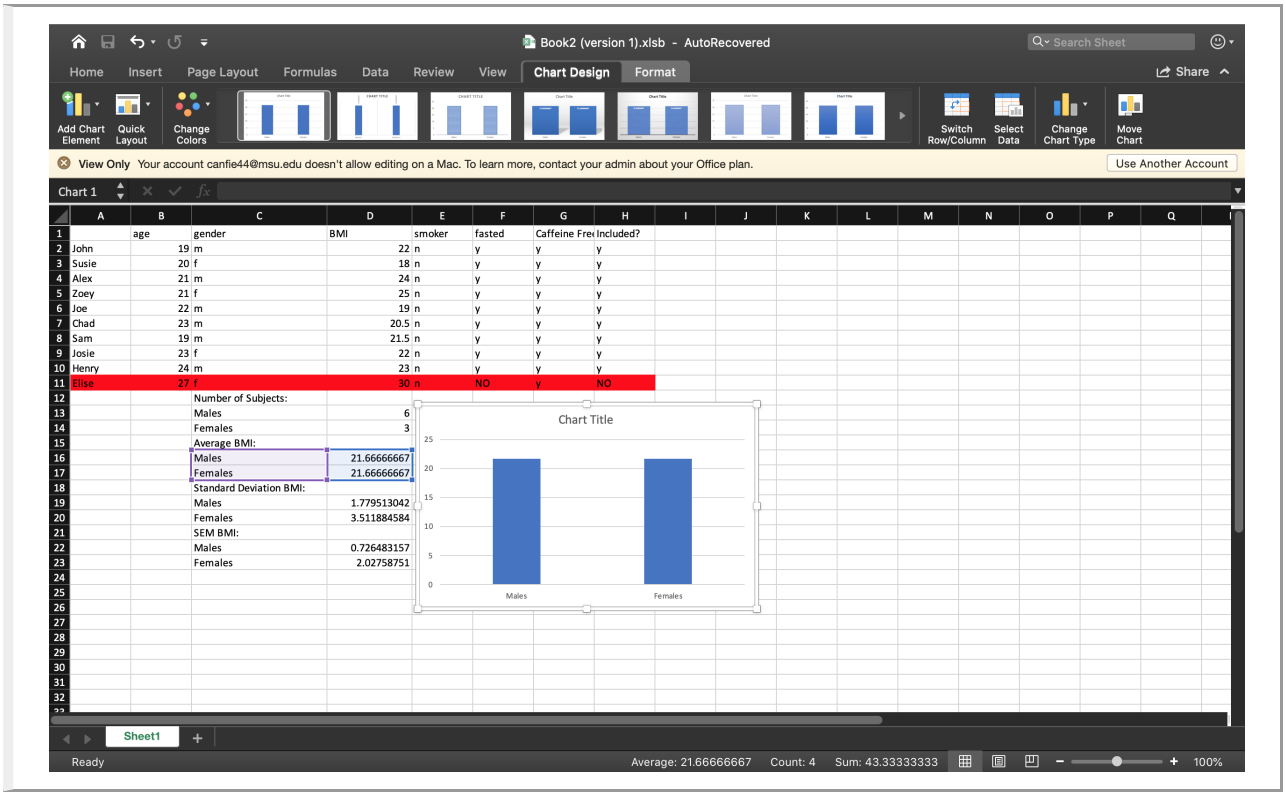
We are now going to go up to the tab in the top of screen and select **Insert**:



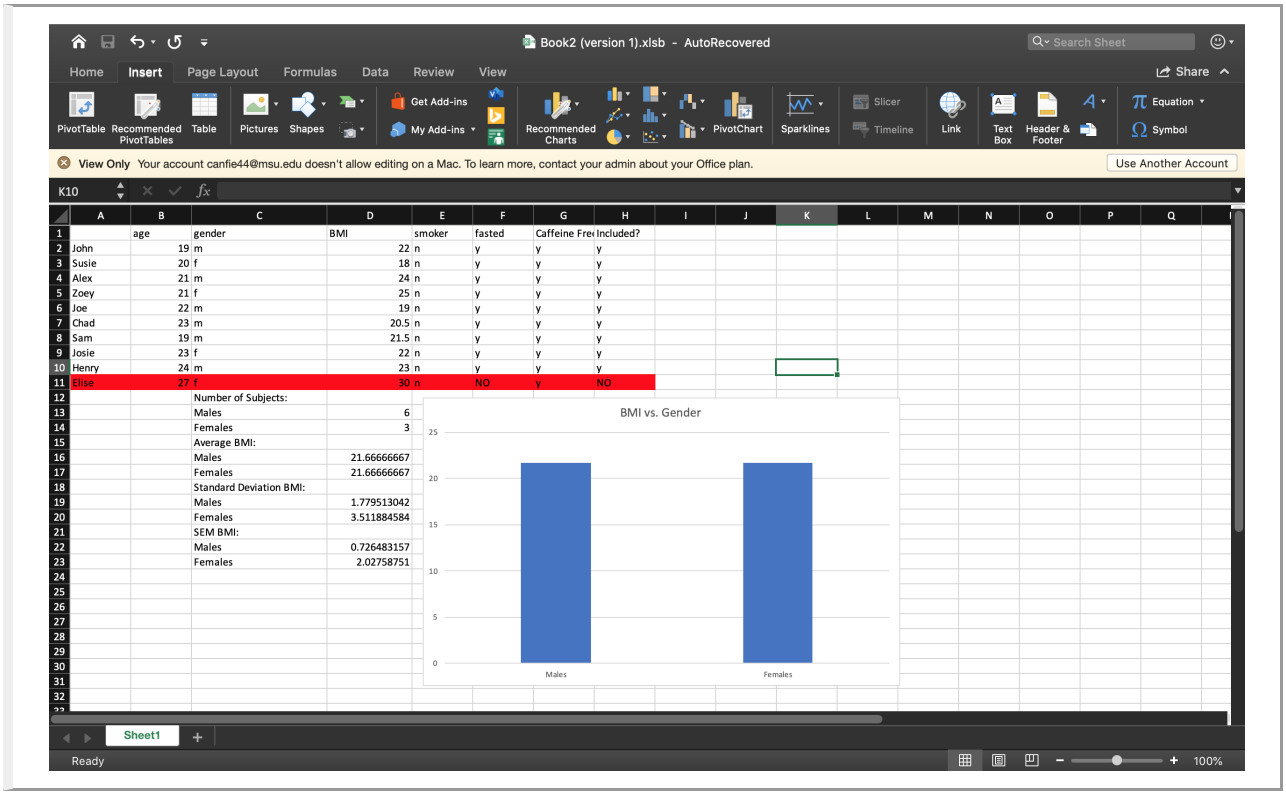
Then select the **column** figure button:



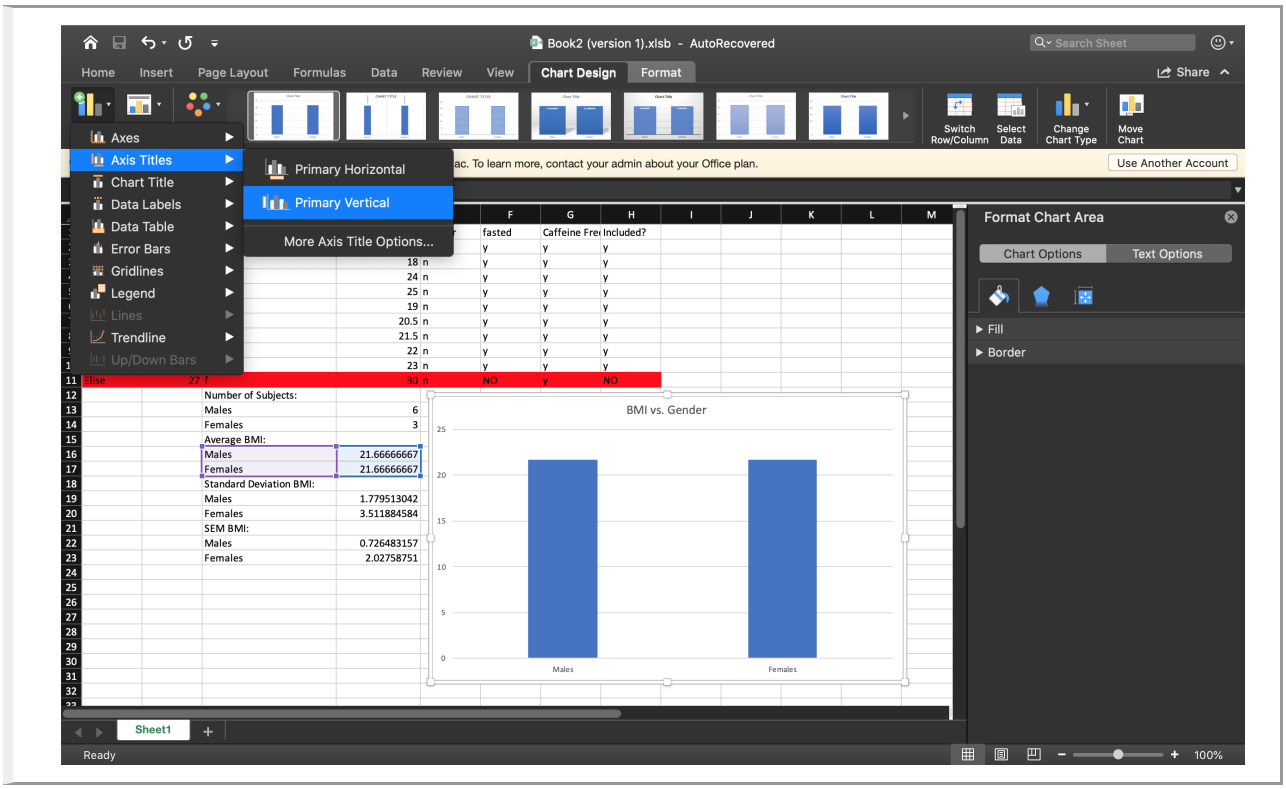
Select the **2D Column (Clustered Column)** figure:

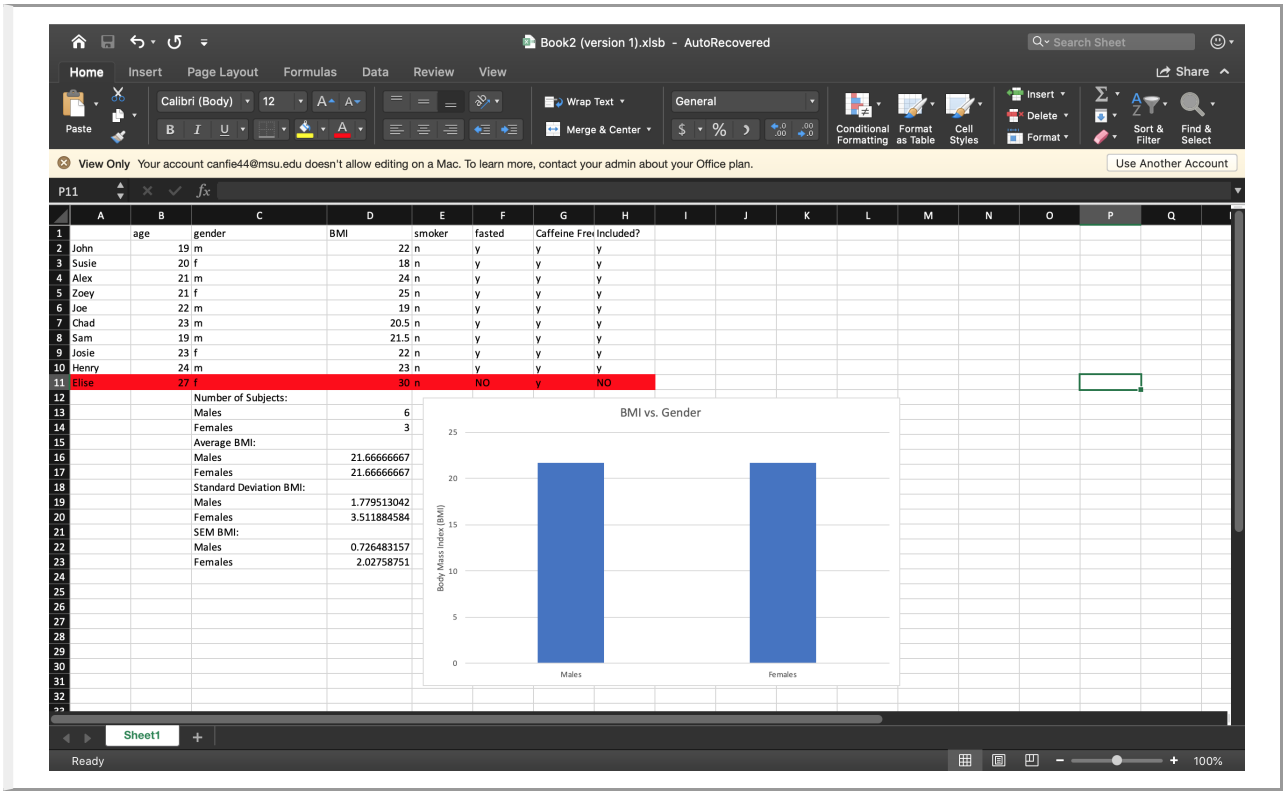


Awesome! Now we have a figure to work with. Now it's time to pretty it up so that it is visually appealing and shows the appropriate information. We already have a *Chart Title* so let's put something meaningful. If you don't want to be super creative the most efficient naming mechanism is: `Independent varaible vs. Dependent variable`. This is generic but will get the information across. Other information pertinent can be included as well. I will use `BMI vs. Gender`.

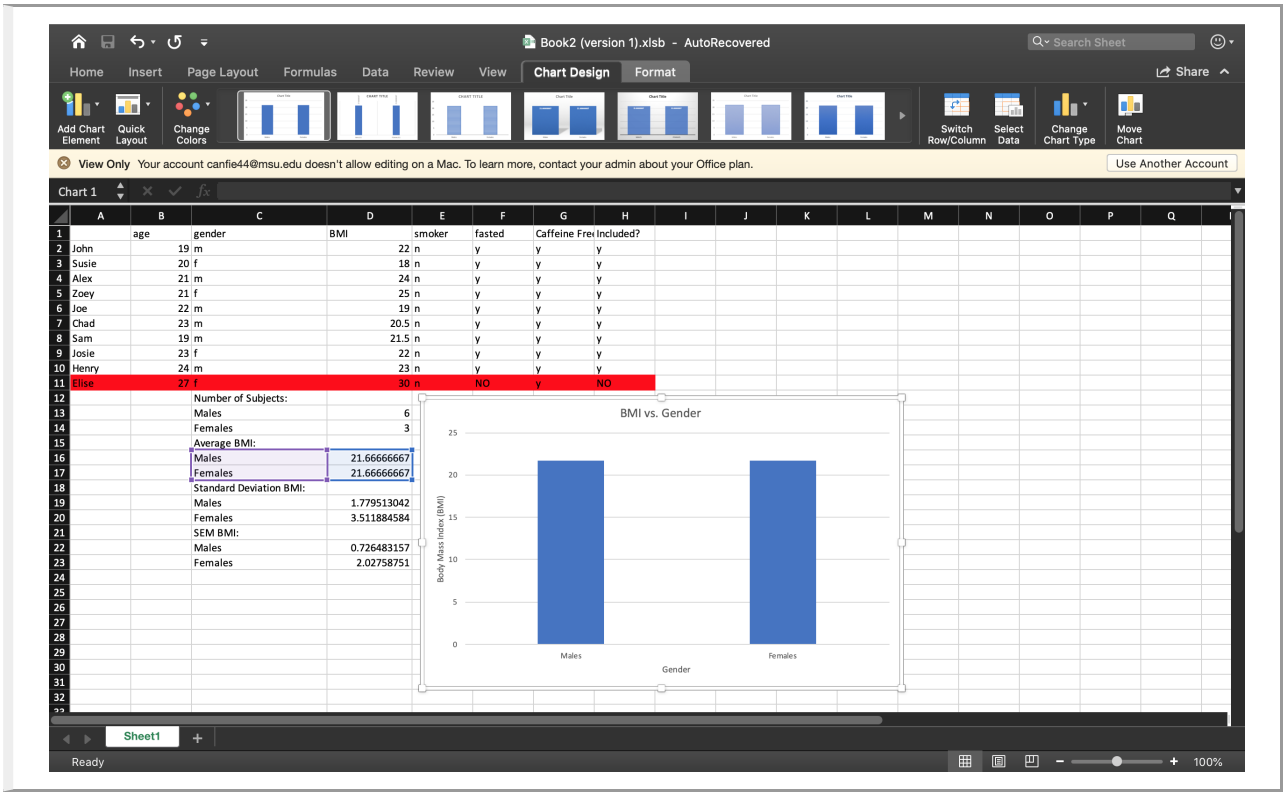


Next we should add some titles to our axes. Typically, you want to include any relevant units or abbreviations so that someone can read your figure without having to read anything else. Let's add an axis title for the y-axis. I'm going to call it `Body Mass Index (BMI)`. To do this we need to double click on some blank space on our figure. This will pull up the **Chart Design** tab. From there we will click **Add Chart Element** (top left). Select **Axis Titles > Primary Vertical**. This will add a textbox on the y-axis of your figure. Change the title of the y-axis like so:

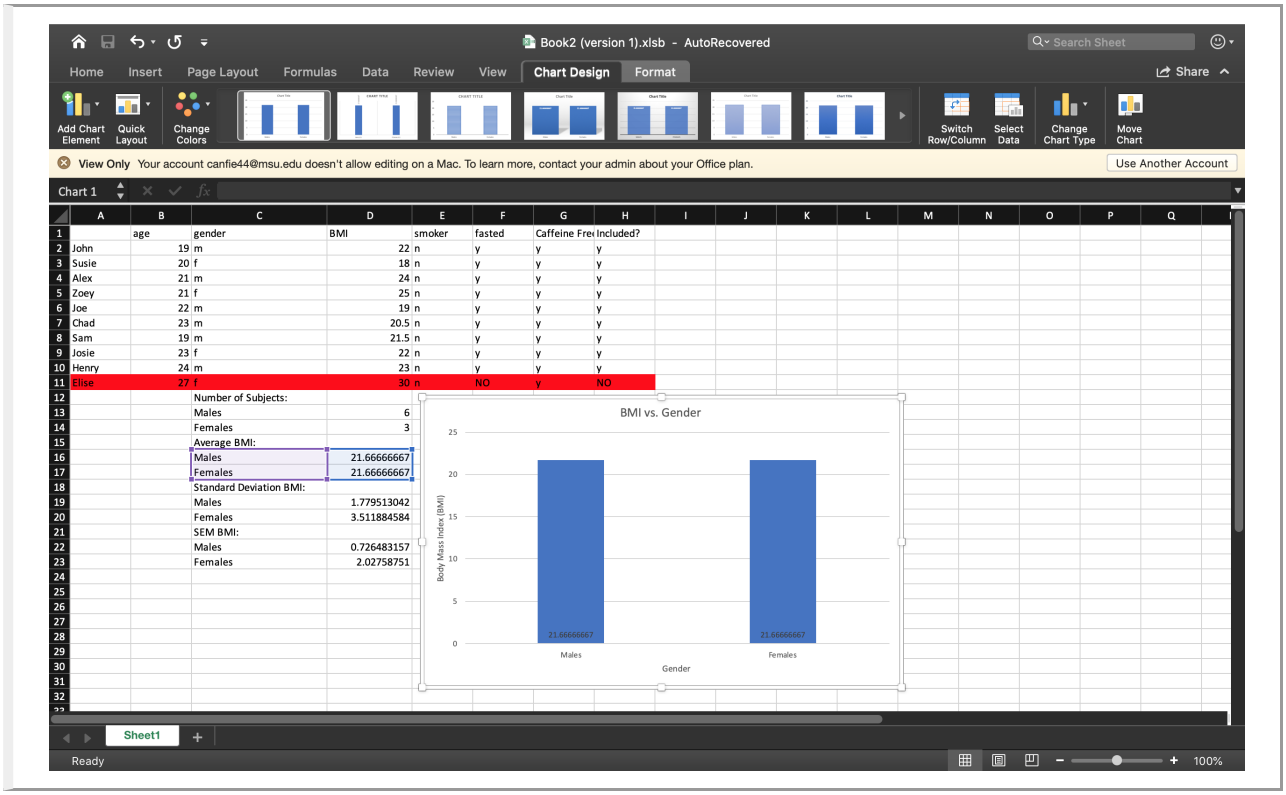




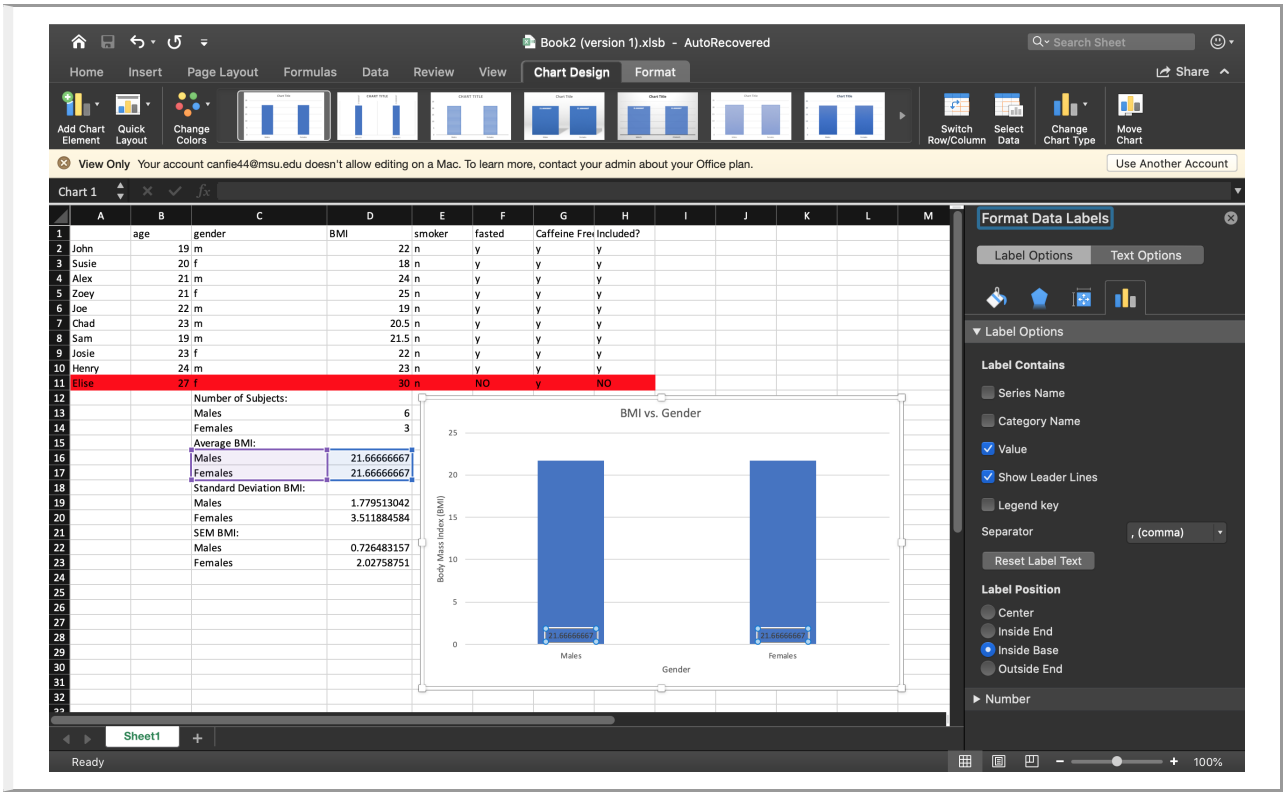
Try doing the same thing with the x-axis! This time select **Add Chart Element > Axis Titles > Primary Horizontal**. I'm going to call my `Gender`



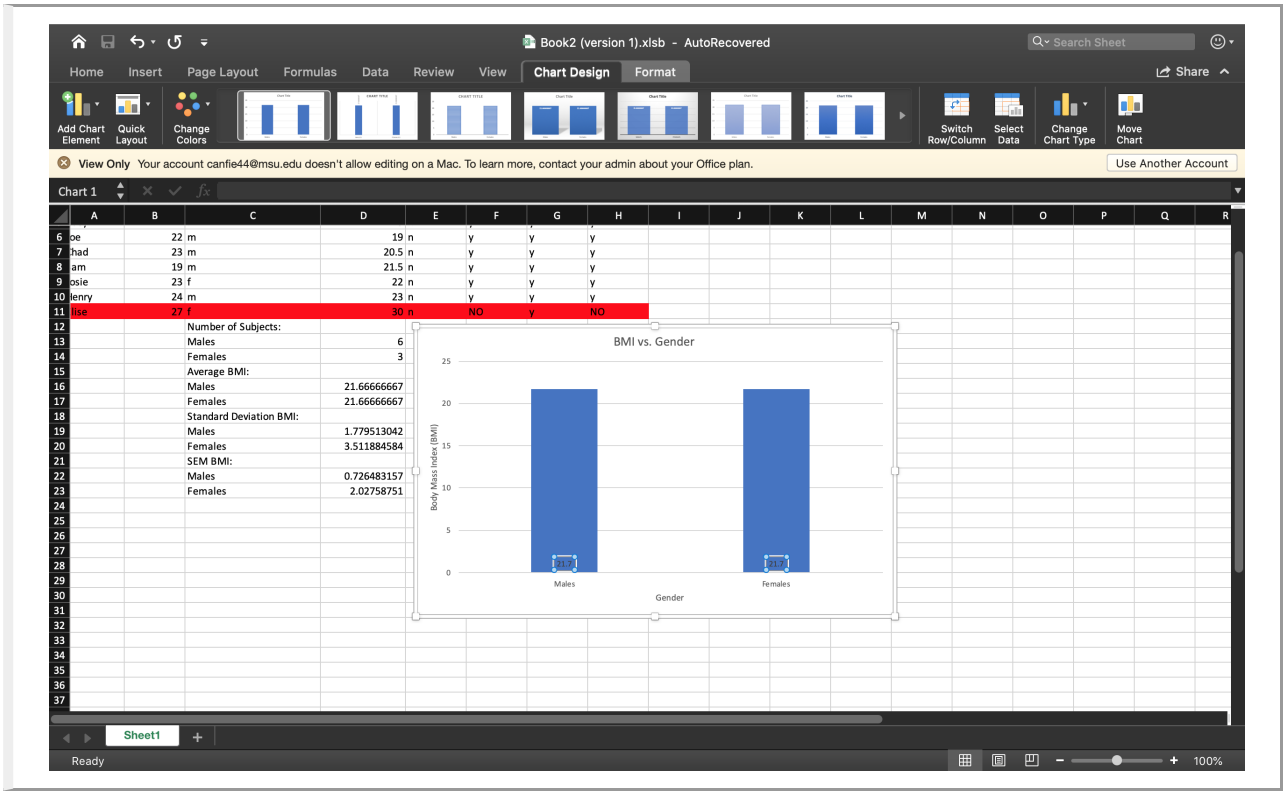
Another thing we can do is add data labels to each bar so that we know what the height of the bar is with high accuracy. To do this select **Add Chart Element > Data Labels > Inside Base**. This will add our value to the bottom of the bar. There are other stylistic ways to add this but the **Inside Base** option is least likely to interfere with some other chart elements that we will be adding later.



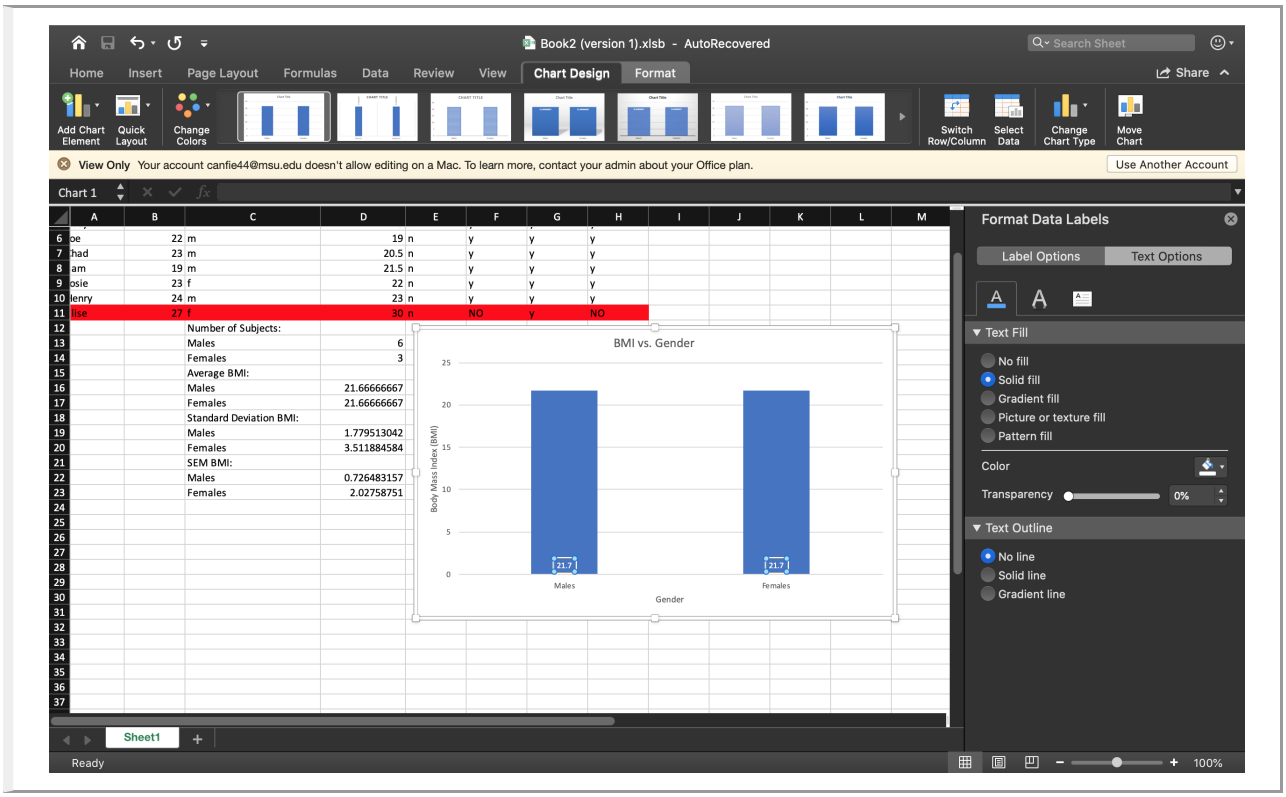
Notice that the color of the text is not very visible with the blue background and there are way too many decimal places? We can format many chart elements by double clicking on them to pull up a menu that allows us to manipulate their properties. Let's do that:



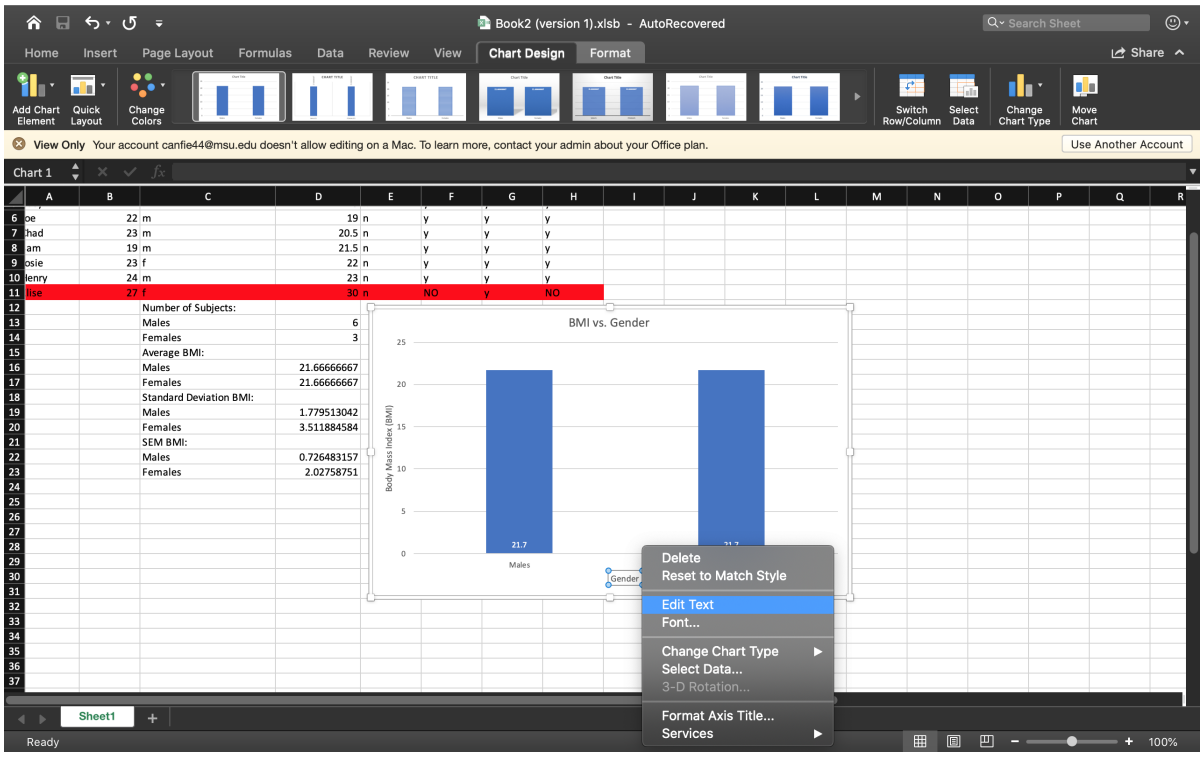
If we go down to where there is an arrow pointing to **Number** and click on it, a drop down menu will appear. There is a field called **Category** which dictates the type of value in a cell or element. Change this to **Number**. This will then allow us to manipulate the **Decimal Places**. The **significant figures** of our data allow 1 decimal place so let's make that change.

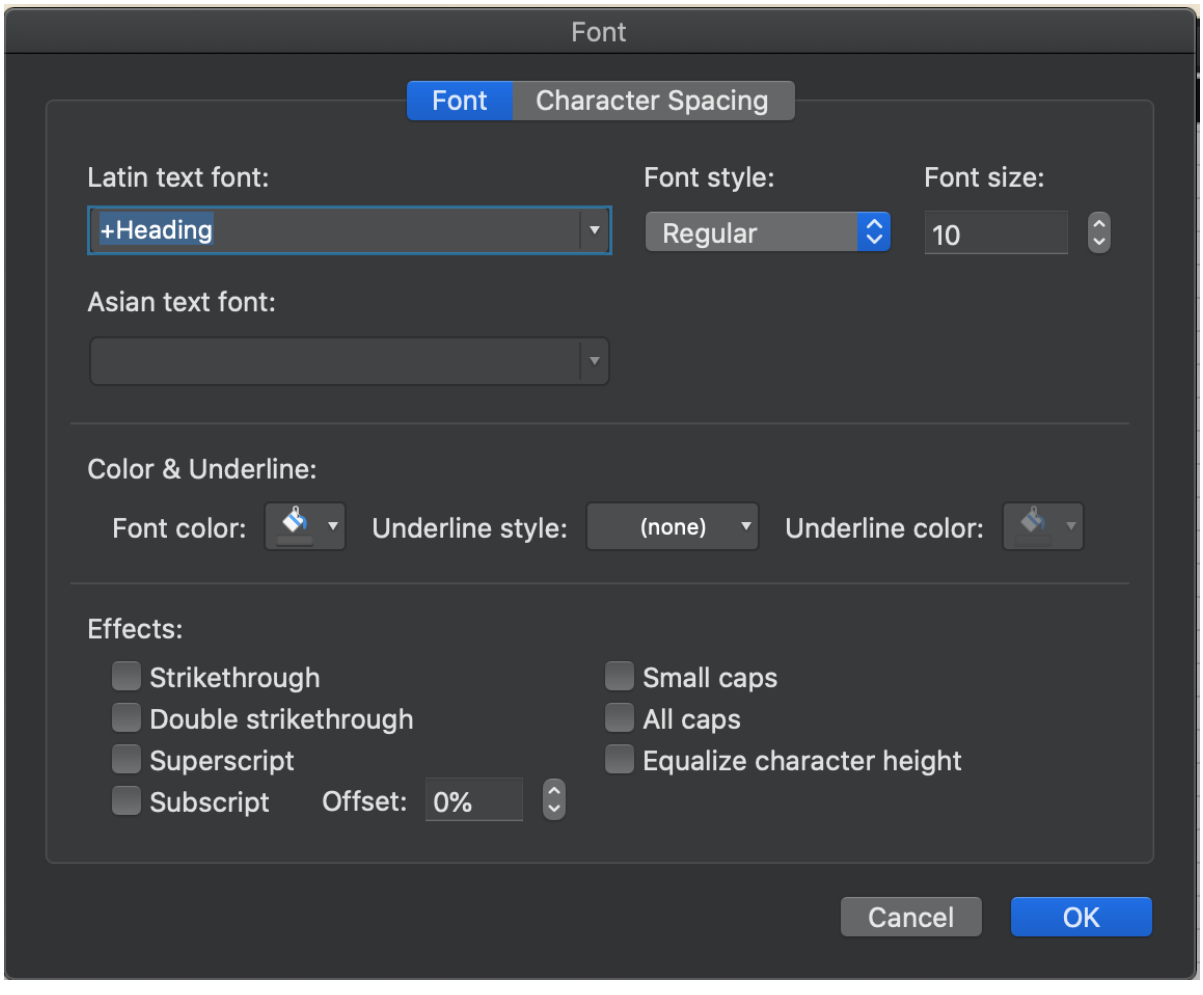


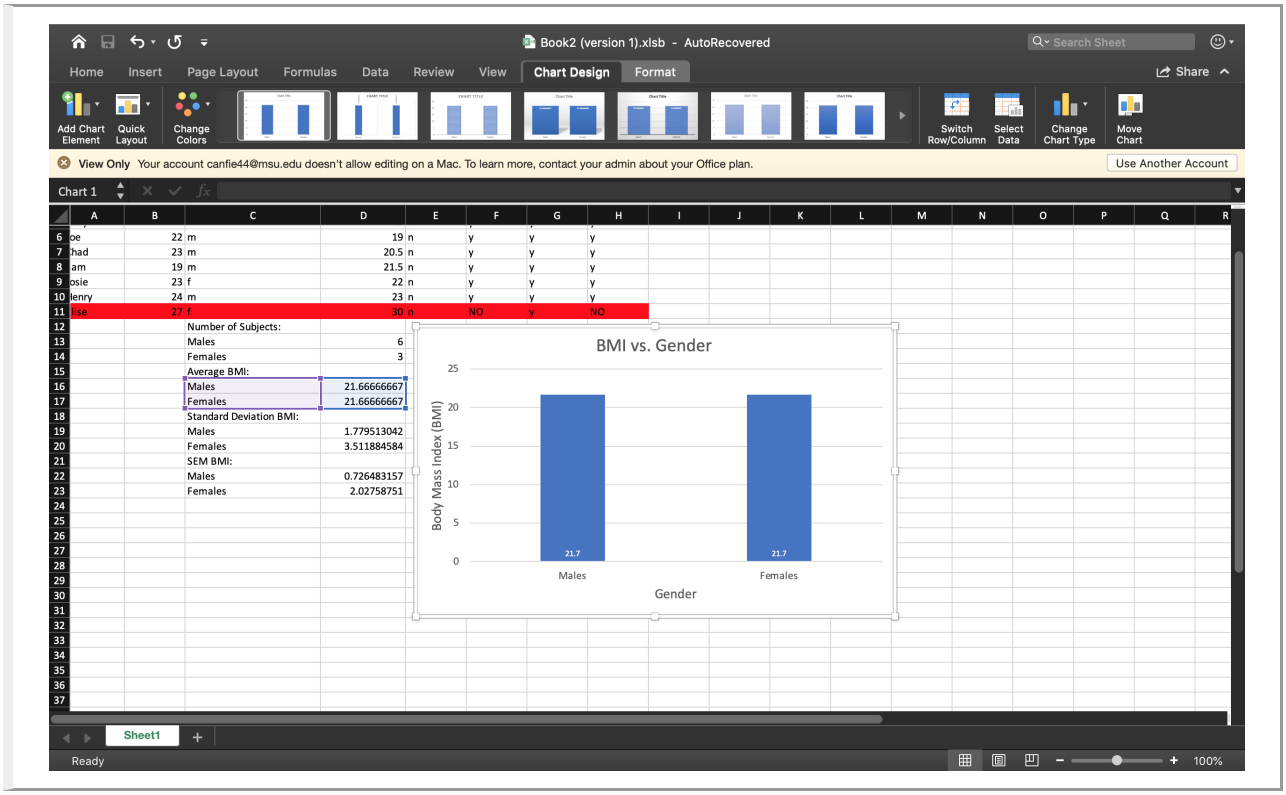
As you can see this solves our **significant figures** problem, however, this doesn't make it any easier to see against the dark background of the bar. Going back to the menu for editing elements by double clicking on the numbers at the bottom of the bar we now see several icons at the top. Currently we are in one that looks like a bar graph. Select the **Paint Can** icon to edit features like text color. Then two tabs will be shown at the top. Select **Text Options**. There will be a smaller **Paint Can** icon on the right of the menu under the drop down menu **Text Fill**. Click on it and select the color of your text (I will use white for contrast).



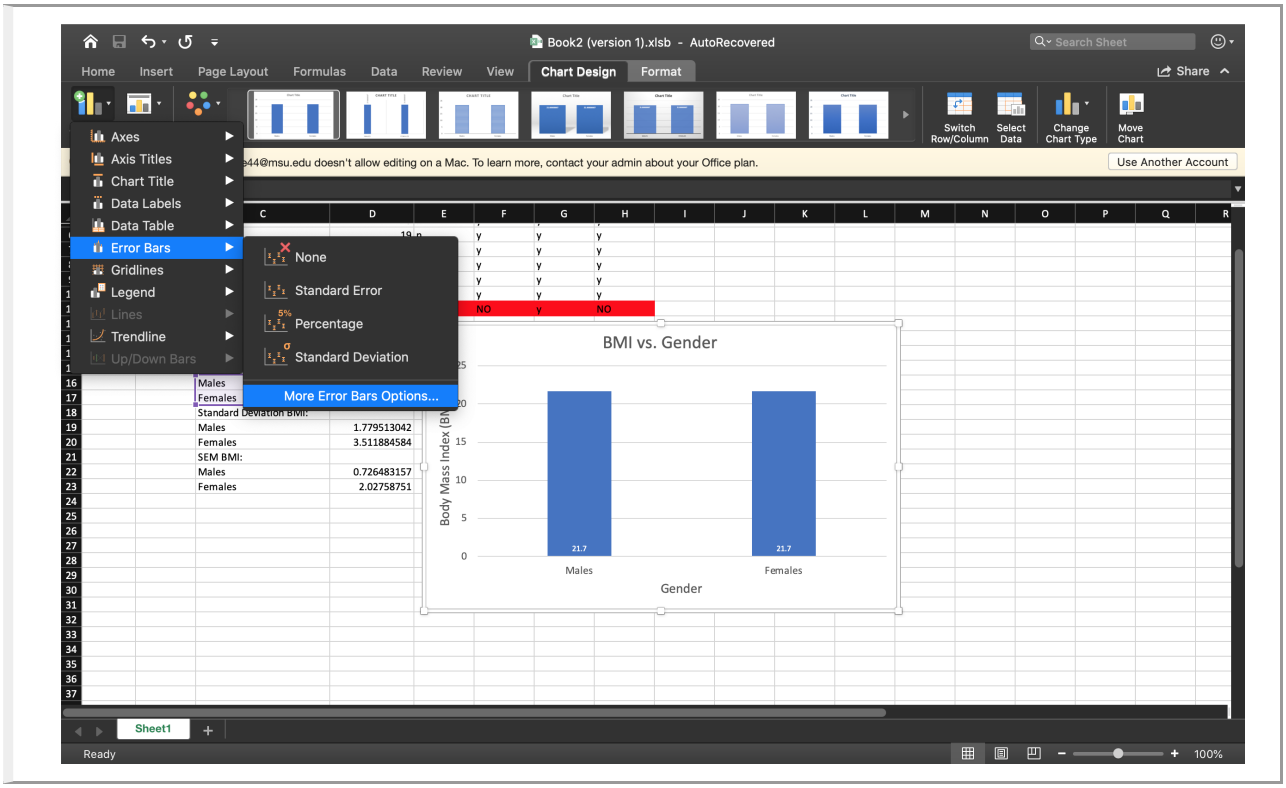
Nice job! Let's keep going. Single click on the **Axis Titles**, **Title** and the **Axis Ticks** and then right click on them. Select **Font**. In the menu displayed increase the size of the text or any other properties so that you can easily see them on your figure.


Book2 (version 1).xlsb - AutoRecovered
Search Sheet
Home
Insert
Page Layout
Formulas
Data
Review
View
Chart Design
Format
Share
Add Chart Element
Quick Layout
Change Colors
Switch Row/Column
Select Data
Change Chart Type
Move Chart
View Only Your account canfie44@msu.edu doesn't allow editing on a Mac. To learn more, contact your admin about your Office plan.
Use Another Account
Chart 1
BMI vs. Gender
Body Mass Index (BMI)
Males
21.7
Gender
Delete
Reset to Match Style
Edit Text
Font...
Change Chart Type
Select Data...
3-D Rotation...
Format Axis Title...
Services
Sheet1
Ready
100%

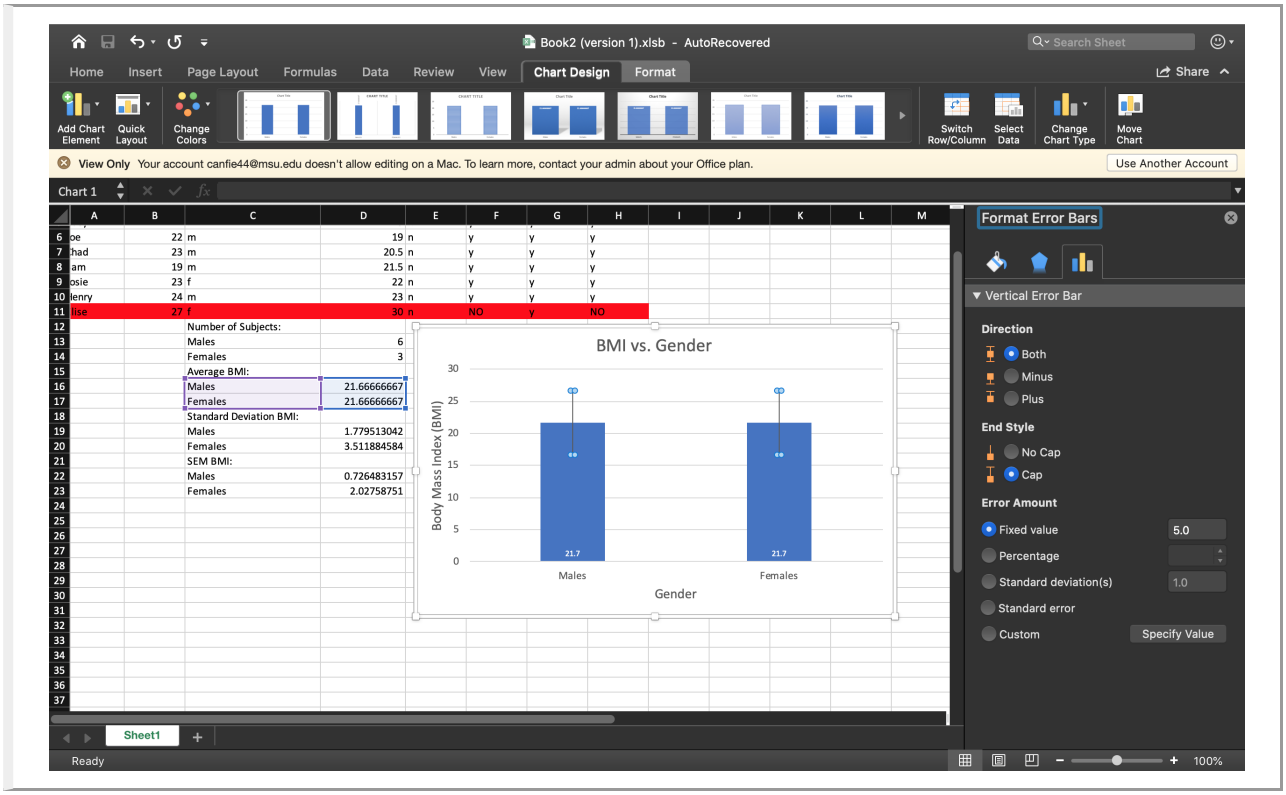

Font
Font
Character Spacing
Latin text font:
+Heading
Font style:
Regular
Font size:
10
Asian text font:
Color & Underline:
Font color:
Underline style:
(none)
Underline color:
Effects:
Strikethrough
Double strikethrough
Superscript
Subscript
Offset:
0%
Small caps
All caps
Equalize character height
Cancel
OK



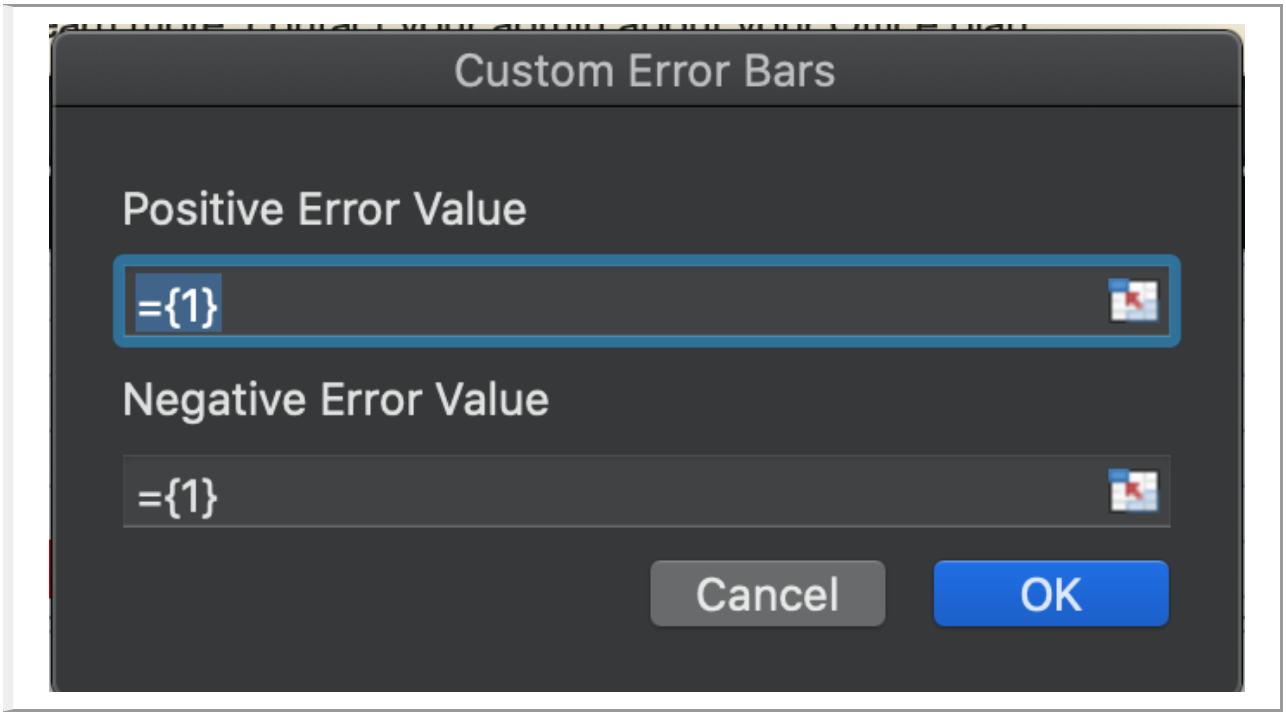
This plot is really coming together now! We're missing something though. Remember, we calculated the **standard deviation** and the **SEM**. Let's incorporate those into our figure in the form of error bars. To do this double click on your figure, select **Add Chart Element > Error Bars > More Error Bar Options...**.



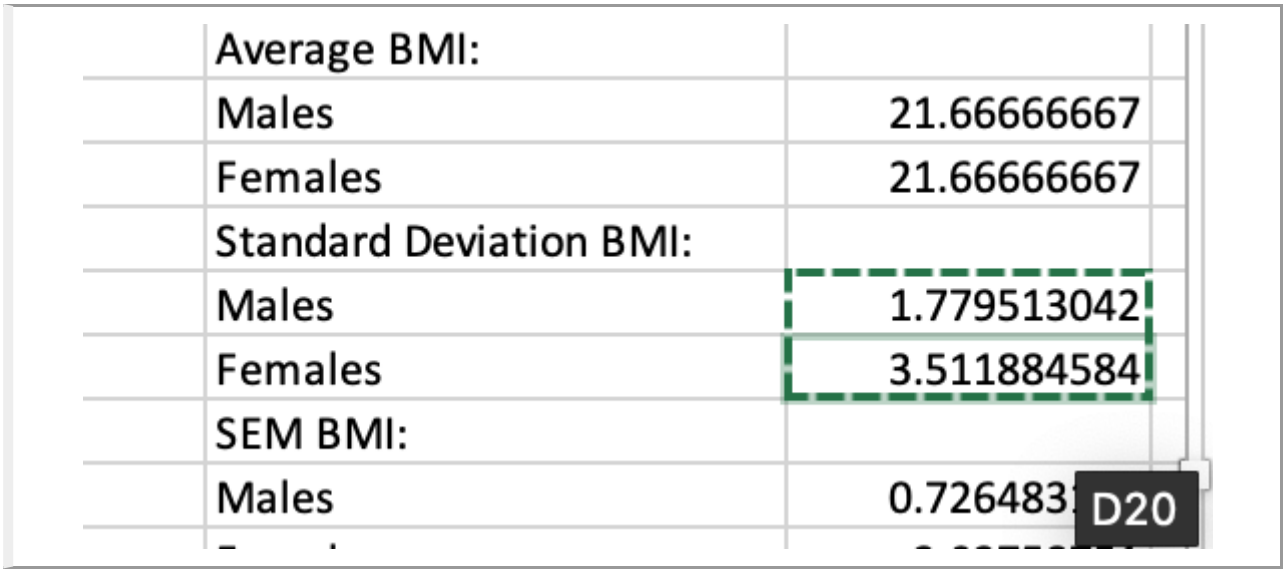
This will pull up a new menu on the right side of your screen:



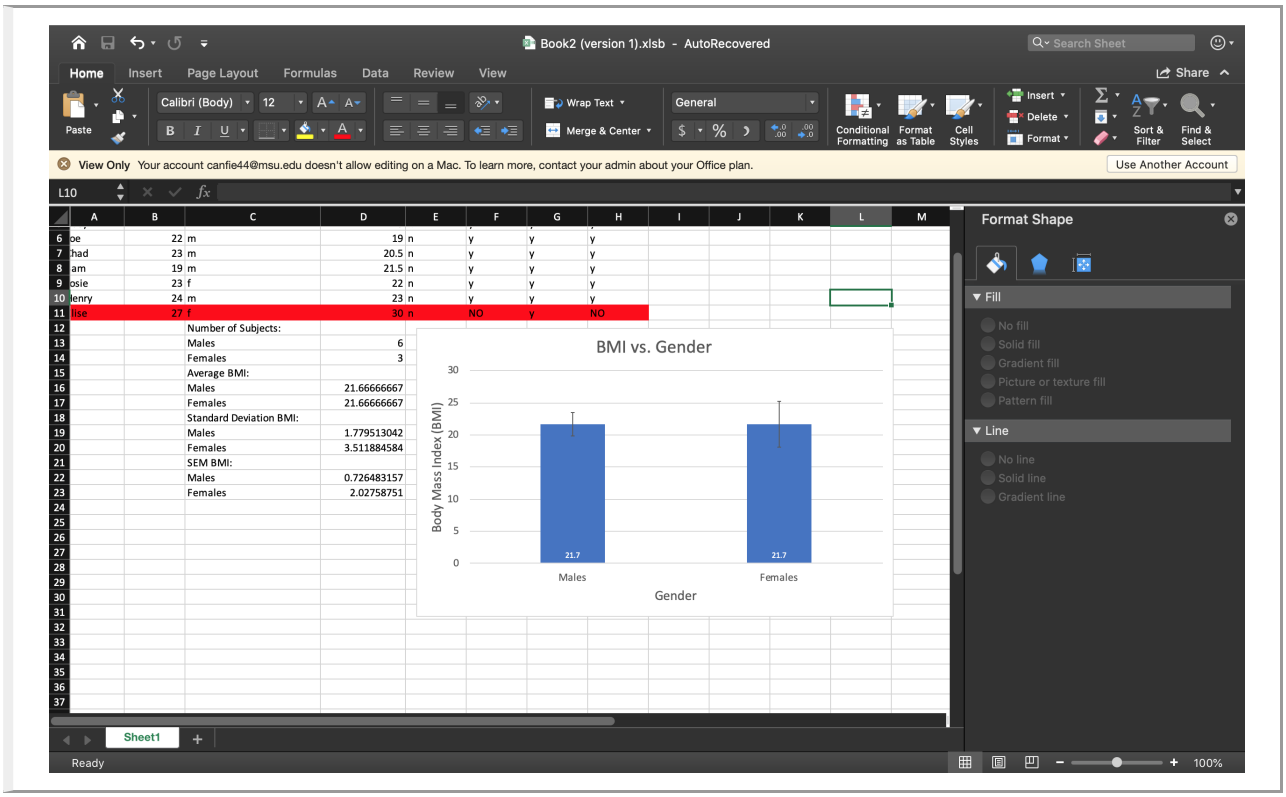
Excel can calculate some of these values for you but because of the nature in which we are displaying our data we are going to do **Custom** error bars. Select **Custom** and then click Specify Value. You will be prompted with this box:



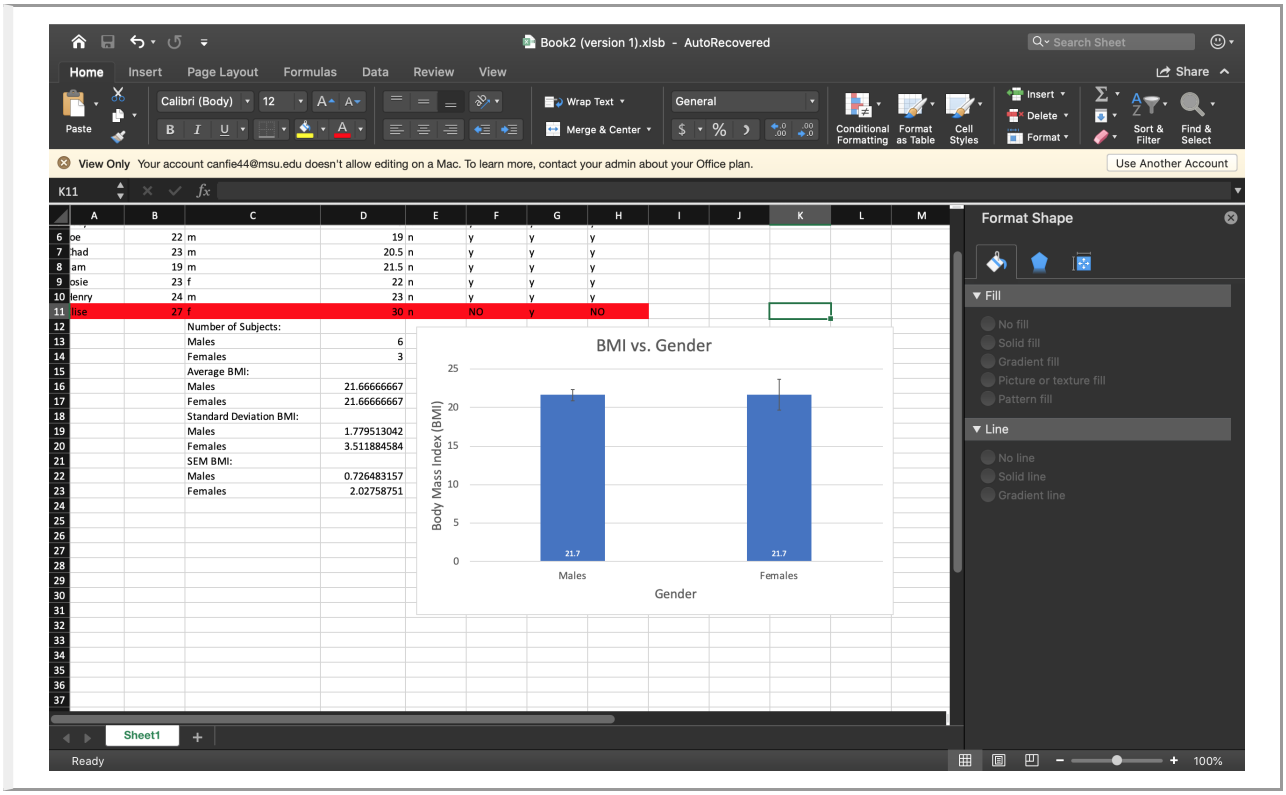
We are now going to click the icon with the red arrow on the right to select the values we want to use for **Positive Error Value** and **Negative Error Value**. Let's select the **standard deviation** that we calculated for both males and females:



Click the icon with the red arrow to add these values. We are going to use this for both our **Positive Error Value** AND **Negative Error Value**. So, repeat this same step, selecting the same values, for the **Negatie Error Values** just as you did for the **Positive Error Value**. Click **OK** when you have selected both sets of values. Now look at your plot with error bars!



We can follow the same steps that we did with **standard deviation** to add error bars using **SEM**. **SEM** is a more common value used for an error bar because it tells us the `average +/- SEM` or basically the range of values surrounding the **estimated average** that the **true average** may fall in. Increasing **sample size, N,** will make **SEM** smaller and thus we can be more confident in our estimation of our **average**.



Conclusion

Congratulations! You have successfully made a **Barplot**! You don't have to be done yet. There are many ways you can stylistically alter your figures in **Excel**. You can change the color, size and font of all the text. You can change the colors and shading methods of the bars. **Excel** also includes many pre-made templates that can look quite interesting. Spend some time exploring and playing around with these so that you can make unique and personalized figures with your own little touch! Hope you enjoyed this tutorial and I look forward to "seeing" you next time! If you want to learn how to make figures like this using the coding language **Python** check out my **Coding Bootcamp** tailored for PSL 475L!