

# Excel training session

API-201, Fall 2021  
TF: Sophie Hill

# Plan for today

- Sophie's 2 Golden Rules of **Excellence**
- Getting oriented with Excel
- Excel for API-201
  - Cleaning data
  - Analyzing data
  - Visualizing data

# Rule #1: Work efficiently

— — —

# Rule #1: Work efficiently

This is a stats class, not an Excel class.

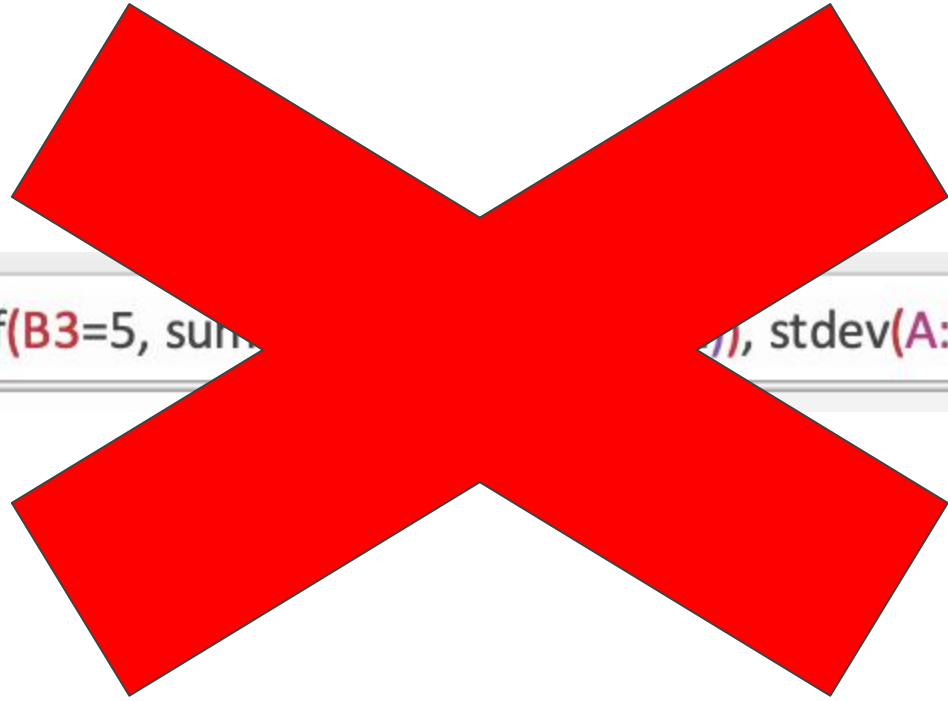
If you are stuck on something Excel-related for more than 5 minutes... it's time to **#askforhelp!**

— — —

# Rule #2: Work transparently

— — —

$f_x$  | =if(A2=1, if(B3=5, sum(A3:C8), avg(C2:D11)), stdev(A:A), "hello world")|



$f_x$  | =if(A2=1, if(B3=5, sum(A:A), stdev(A:A), "hello world"))

# Rule #2: Work transparently

Make your Excel sheet intelligible to your future self / your collaborators / your hypothetical boss!

- Break calculations down into smaller steps
- Use sensible column/sheet names

— — —

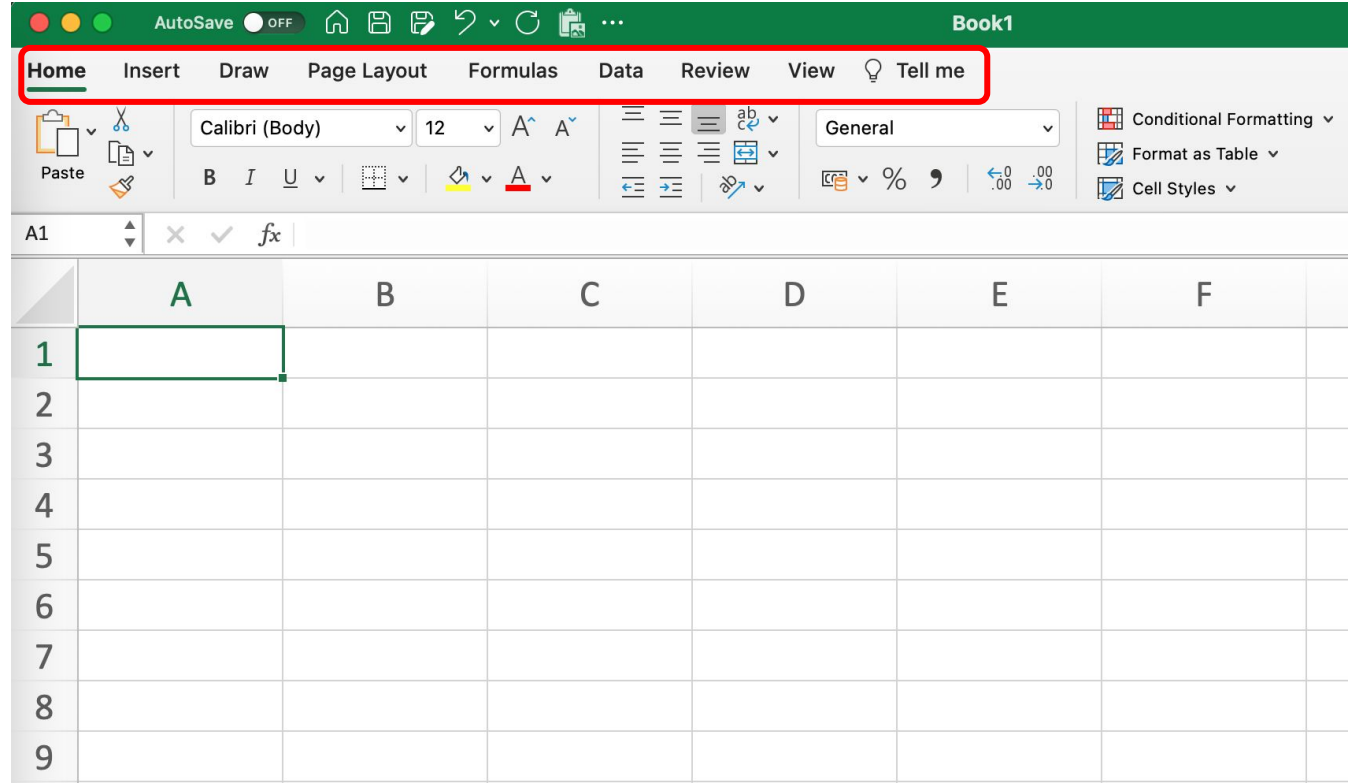


# But first... The Basics

Disclaimer: I am using Microsoft Excel for Mac version 16 (Office 365). Things may look slightly different depending on your version and operating system.

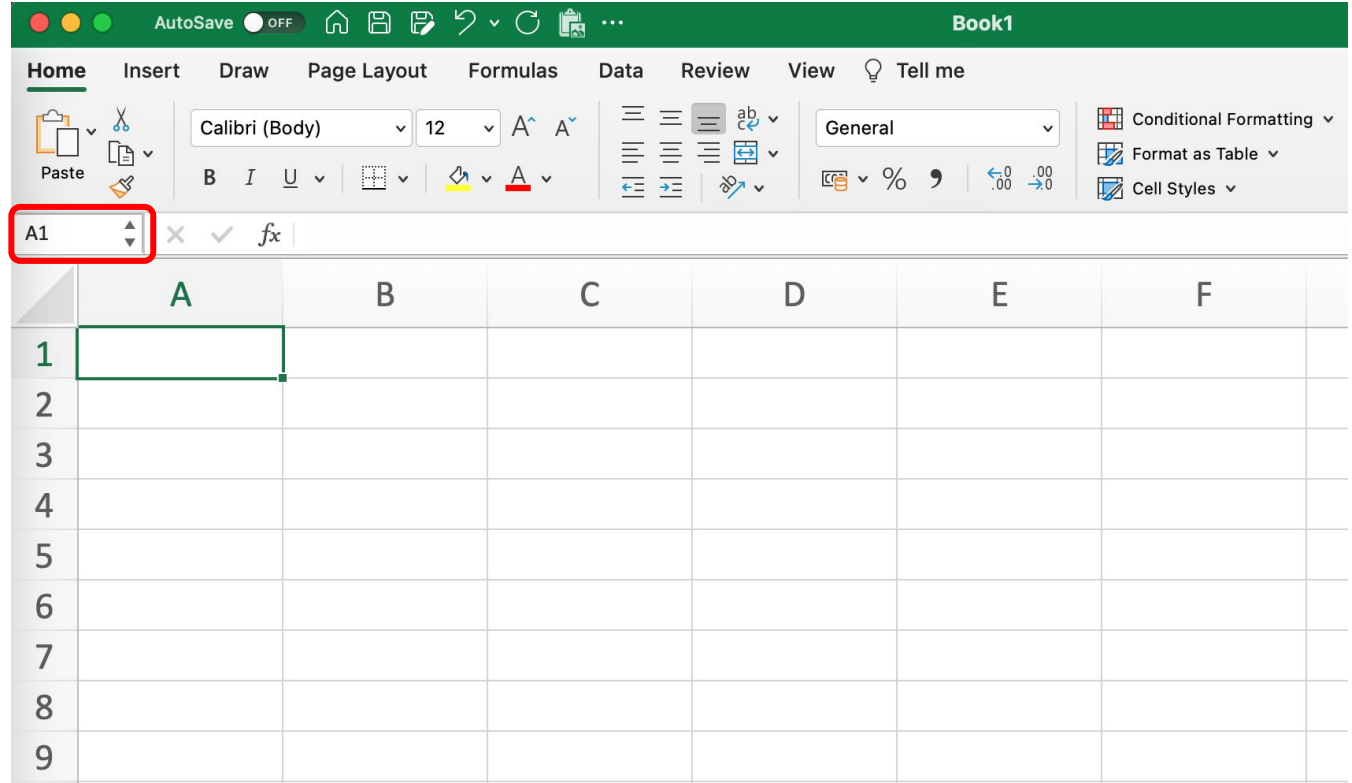
# Getting oriented

Menu bar



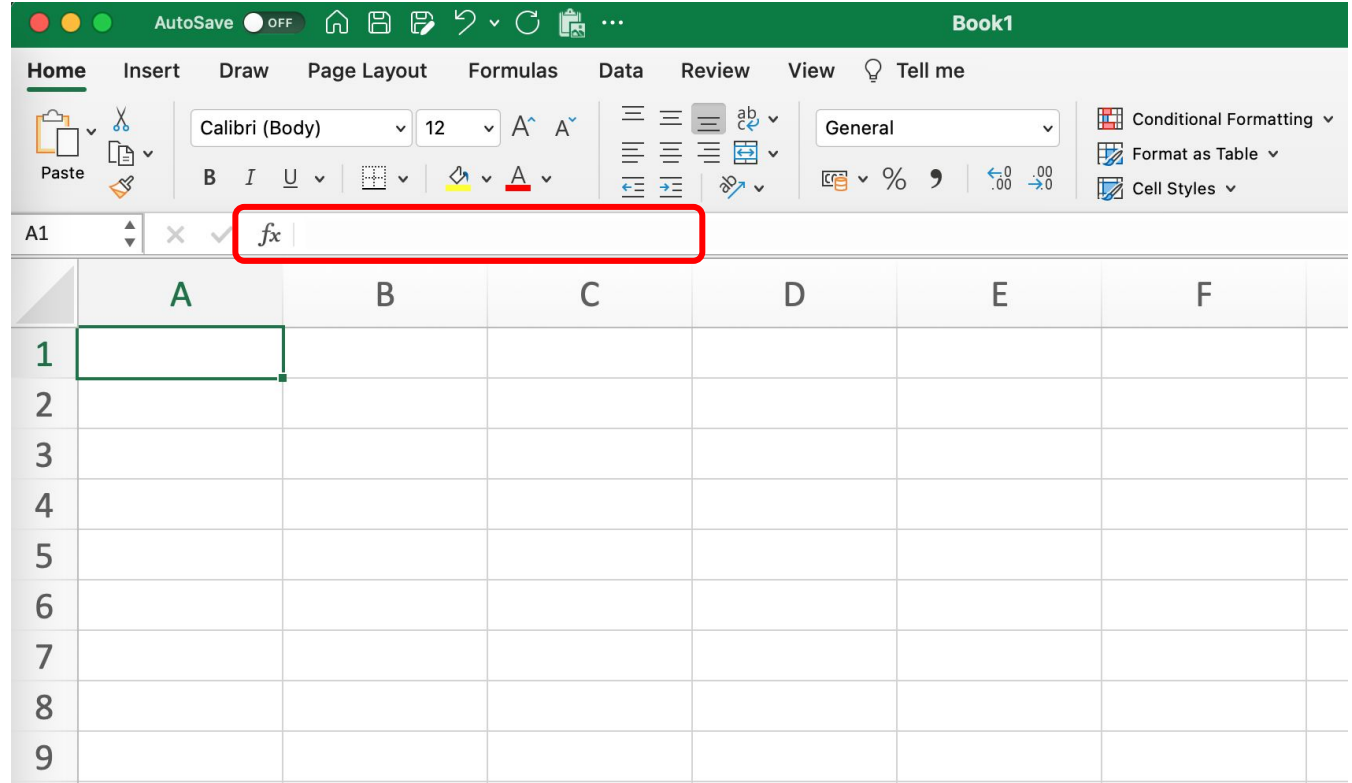
# Getting oriented

Highlighted  
cell



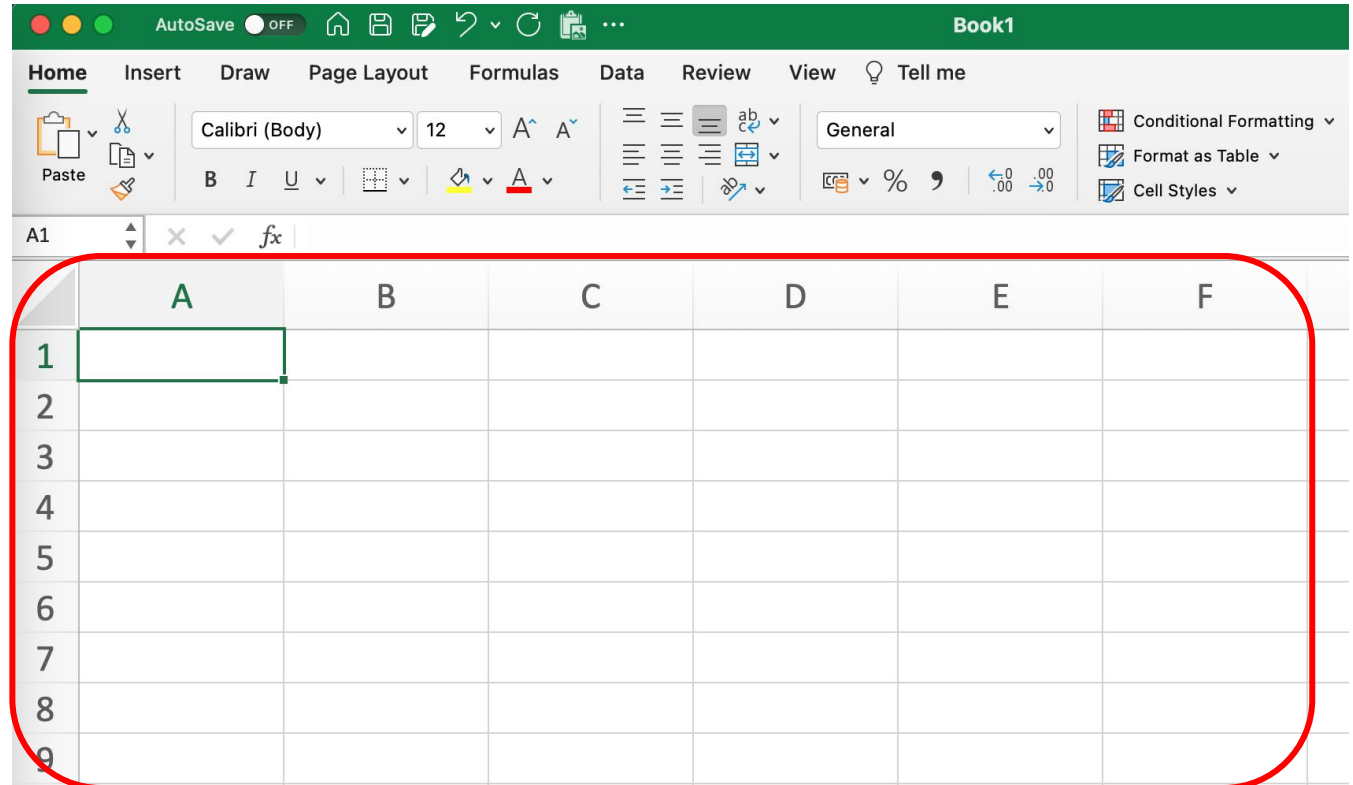
# Getting oriented

Formula bar



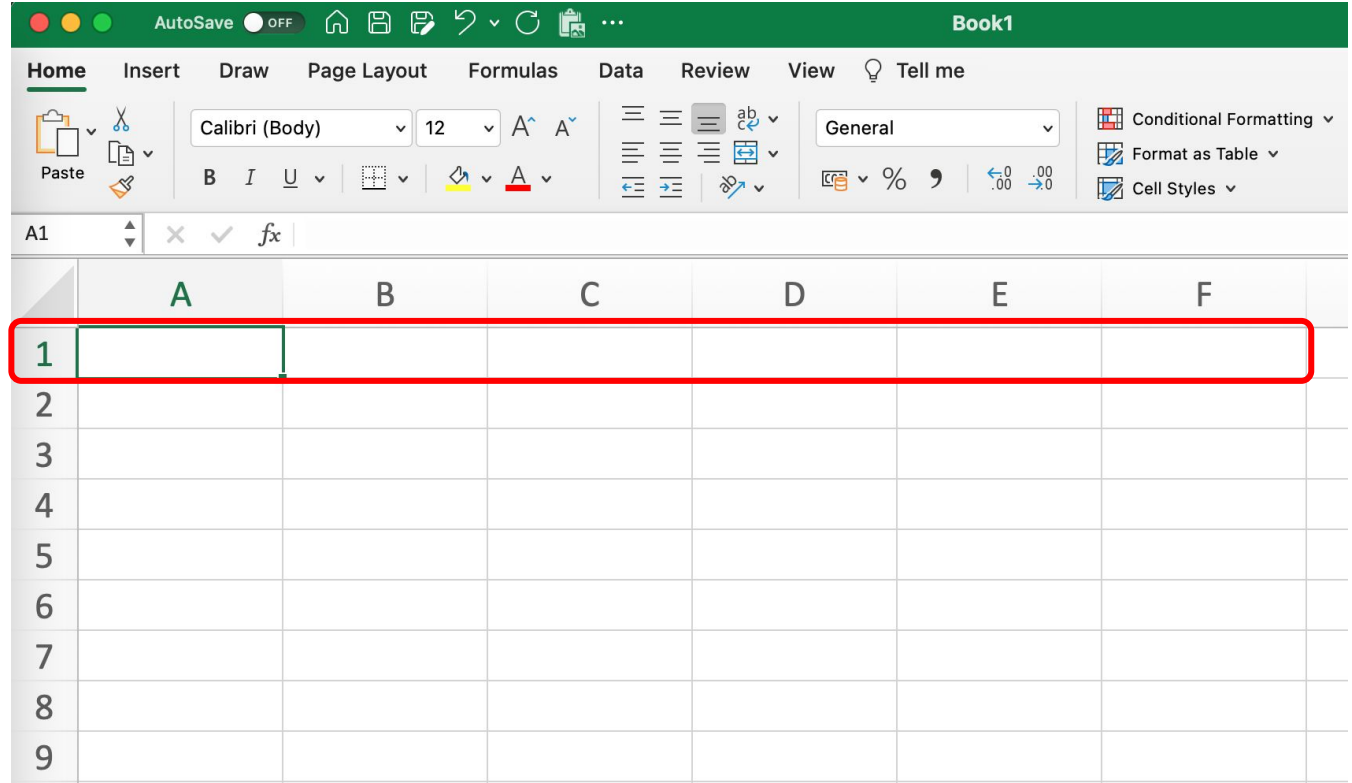
# Getting oriented

Sheet



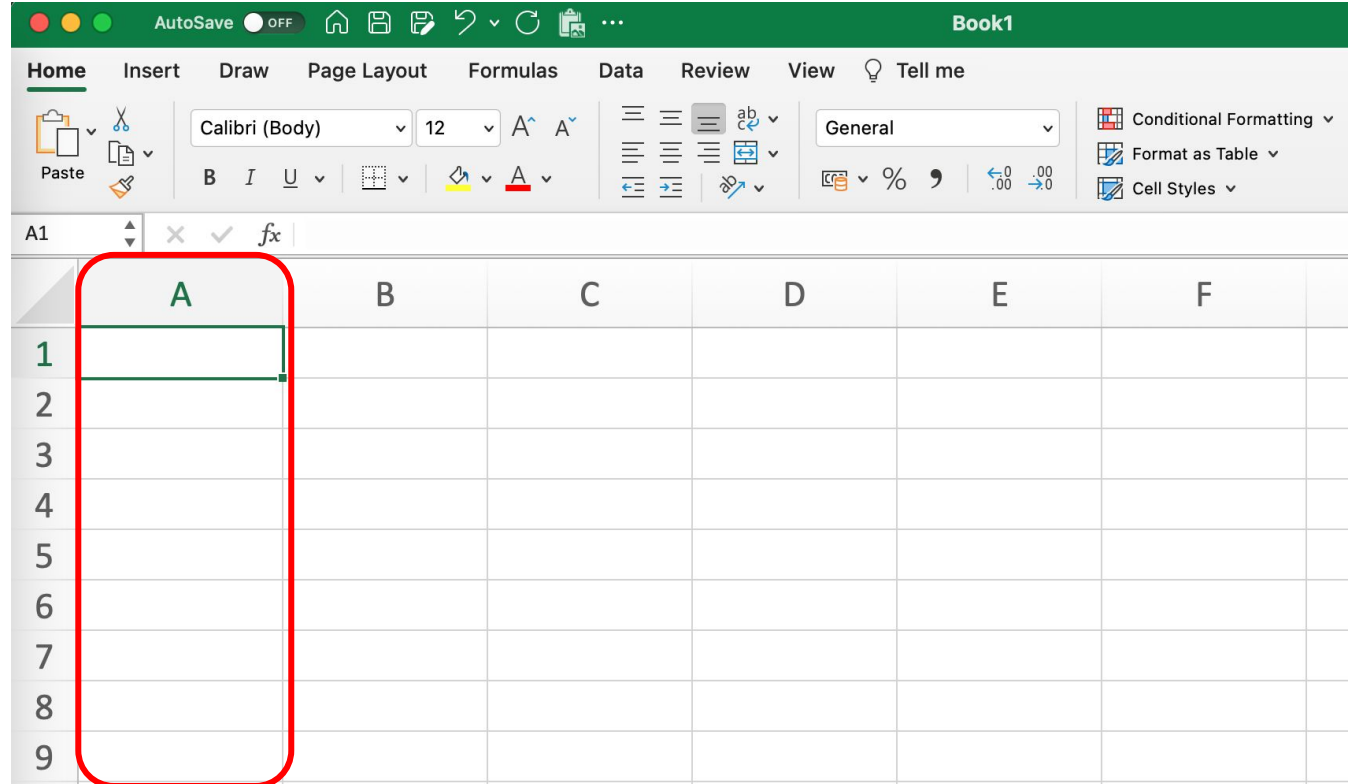
# Getting oriented

Rows  
(numbers)

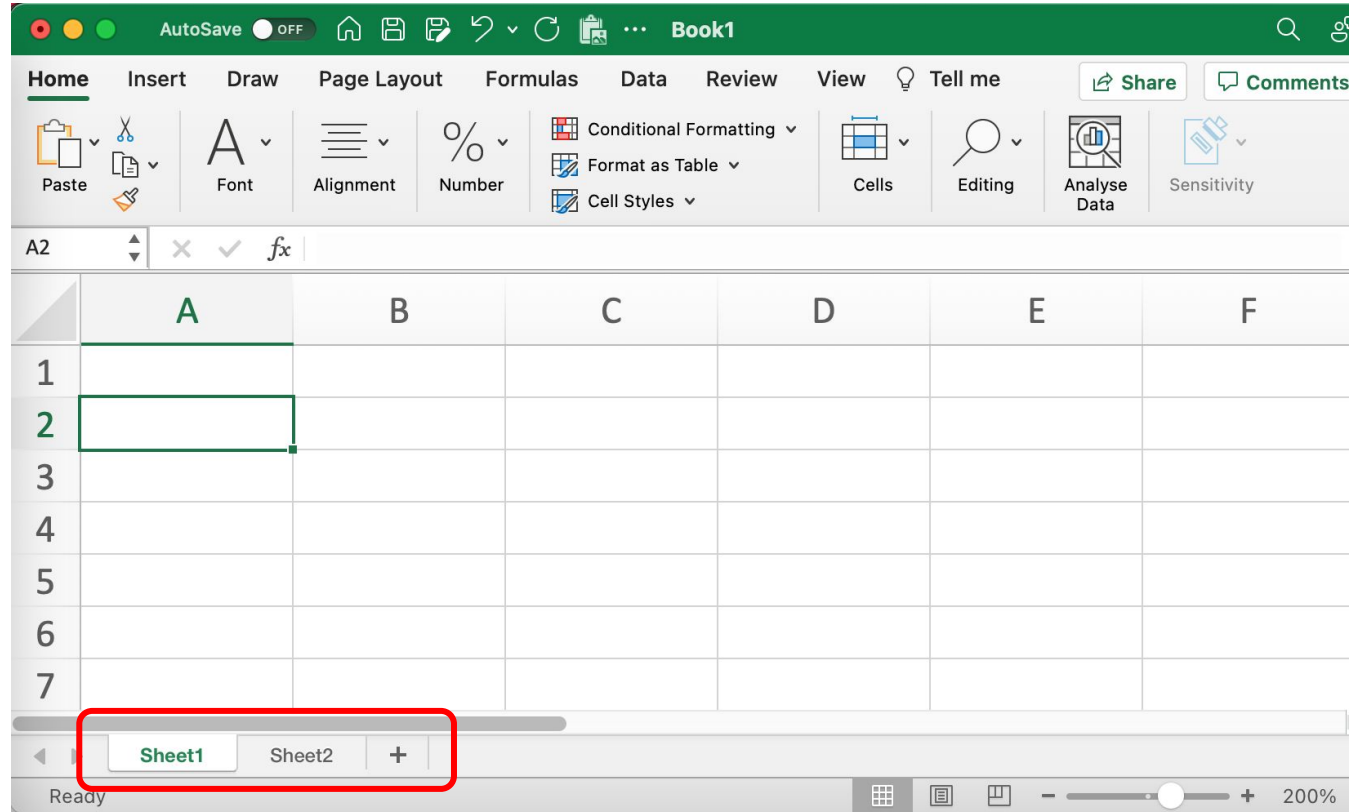


# Getting oriented

Columns  
(letters)



# Getting oriented



Tabs



# Syntax: referencing cells

<b>A1</b>	The cell A1
<b>A1:A3</b>	The cells A1, A2, A3
<b>A:A</b>	
<b>A:B</b>	
<b>A1:B7</b>	
<b>\$A\$1</b>	
<b>Sheet1!A1</b>	

# Syntax: referencing cells

<b>A1</b>	The cell A1
<b>A1:A3</b>	The cells A1, A2, A3
<b>A:A</b>	The whole A column
<b>A:B</b>	
<b>A1:B7</b>	
<b>\$A\$1</b>	
<b>Sheet1!A1</b>	

# Syntax: referencing cells

<b>A1</b>	The cell A1
<b>A1:A3</b>	The cells A1, A2, A3
<b>A:A</b>	The whole A column
<b>A:B</b>	The whole of columns A and B
<b>A1:B7</b>	
<b>\$A\$1</b>	
<b>Sheet1!A1</b>	

# Syntax: referencing cells

<b>A1</b>	The cell A1
<b>A1:A3</b>	The cells A1, A2, A3
<b>A:A</b>	The whole A column
<b>A:B</b>	The whole of columns A and B
<b>A1:B7</b>	The array covering A1 down to A7 and B1 down to B7
<b>\$A\$1</b>	
<b>Sheet1!A1</b>	

# Syntax: referencing cells

<b>A1</b>	The cell A1
<b>A1:A3</b>	The cells A1, A2, A3
<b>A:A</b>	The whole A column
<b>A:B</b>	The whole of columns A and B
<b>A1:B7</b>	The array covering A1 down to A7 and B1 down to B7
<b>\$A\$1</b>	A fixed reference to cell A1 that won't change when you copy a formula to another cell
<b>Sheet1!A1</b>	




# Syntax: referencing cells

<b>A1</b>	The cell A1
<b>A1:A3</b>	The cells A1, A2, A3
<b>A:A</b>	The whole A column
<b>A:B</b>	The whole of columns A and B
<b>A1:B7</b>	The array covering A1 down to A7 and B1 down to B7
<b>\$A\$1</b>	A fixed reference to cell A1 that won't change when you copy a formula to another cell
<b>Sheet1!A1</b>	Cell A1 on Sheet 1

# Syntax: understanding check!

Cell A1 contains a text string (“hello world”).

Cell B1 contains a formula (“=\$A1”).




AVERAGEI    <i>fx</i>   =\$A1		
	A	B
1	hello world	=\$A1
2		

# Syntax: understanding check!

Cell A1 contains a text string (“hello world”).

Cell B1 contains a formula (“=\$A1”).

1. What will B1 evaluate to?

AVERAGEI    <i>fx</i>   =\$A1		
	A	B
1	hello world	=\$A1
2		



# Syntax: understanding check!

Cell A1 contains a text string (“hello world”).

Cell B1 contains a formula (“=\$A1”).

1. What will B1 evaluate to?

B1    ✕    ✓    fx    =\$A1		
	A	B
1	hello world	hello world
2		

# Syntax: understanding check!

Cell A1 contains a text string (“hello world”).

Cell B1 contains a formula (“=\$A1”).

1. What will B1 evaluate to?
2. What will happen if I copy the contents of cell B1 into cell B2?

B1			fx   =\$A1		
	A	B			
1	hello world	hello world			
2					

# Syntax: understanding check!

Cell A1 contains a text string (“hello world”).

Cell B1 contains a formula (“=\$A1”).

1. What will B1 evaluate to?
2. What will happen if I copy the contents of cell B1 into cell B2?

B2		
	A	B
1	hello world	hello world
2		0

# Syntax: understanding check!

Cell A1 contains a text string (“hello world”).

Cell B1 contains a formula (“=\$A1”).

1. What will B1 evaluate to?
2. What will happen if I copy the contents of cell B1 into cell B2?

Why?

Because “\$A1” only fixes the column, not the row!

So cell B2 now contains the formula “=\$A2” and cell A2 is empty.

B2		
	A	B
1	hello world	hello world
2		

# Syntax: fixed references

<b>A1</b>	Neither row nor column is fixed
<b>\$A1</b>	Column is fixed
<b>A\$1</b>	Row is fixed
<b>\$A\$1</b>	Row and column are fixed

# Syntax: formulas

In Excel, formulas must **start with the equals sign (=)**

Otherwise, Excel doesn't know if you want to calculate the SUM or just write the word "sum"!

# Syntax: formulas

In Excel, formulas must **start with the equals sign (=)**

Otherwise, Excel doesn't know if you want to calculate the SUM or just write the word "sum"!

Functions take **arguments** separated by **commas**:

```
=MYFUNCTION(argument1, argument2, argument3)
```

# Syntax: formulas

An **argument** could be:

- a number
- a text string in quotes
- a cell or range of cells
- a logical condition
- the output of another function



# Syntax: formulas

An **argument** could be:

- a number
- a text string in quotes
- a cell or range of cells
- a logical condition
- the output of another function

```
=IF(A1>0, "this is a positive number", 0)
```

# Syntax: formulas

An **argument** could be:

- a number
- a text string in quotes
- a cell or range of cells
- a logical condition
- the output of another function

=IF(A1>0, “this is a positive number”, 0)

# Syntax: formulas

An **argument** could be:

- a number
- a text string in quotes
- a cell or range of cells
- a logical condition
- the output of another function

=IF(A1>0, "this is a positive number", 0)

# Syntax: formulas

An **argument** could be:

- a number
- a text string in quotes
- a cell or range of cells
- a logical condition
- the output of another function

=IF(A1>0, “this is a positive number”, 0)

# Syntax: formulas

An **argument** could be:

- a number
- a text string in quotes
- a cell or range of cells
- a logical condition
- the output of another function

=IF(A1>0, “this is a positive number”, **AVERAGE(A:A)**)

# Syntax: formulas

An **argument** could be:

- a number
- a text string in quotes
- a cell or range of cells
- a logical condition
- the output of another function

=IF(A1>0, “this is a positive number”, AVERAGE(A:A))

“Give an MPP a COUNTIF, and they will complete the problem set today.

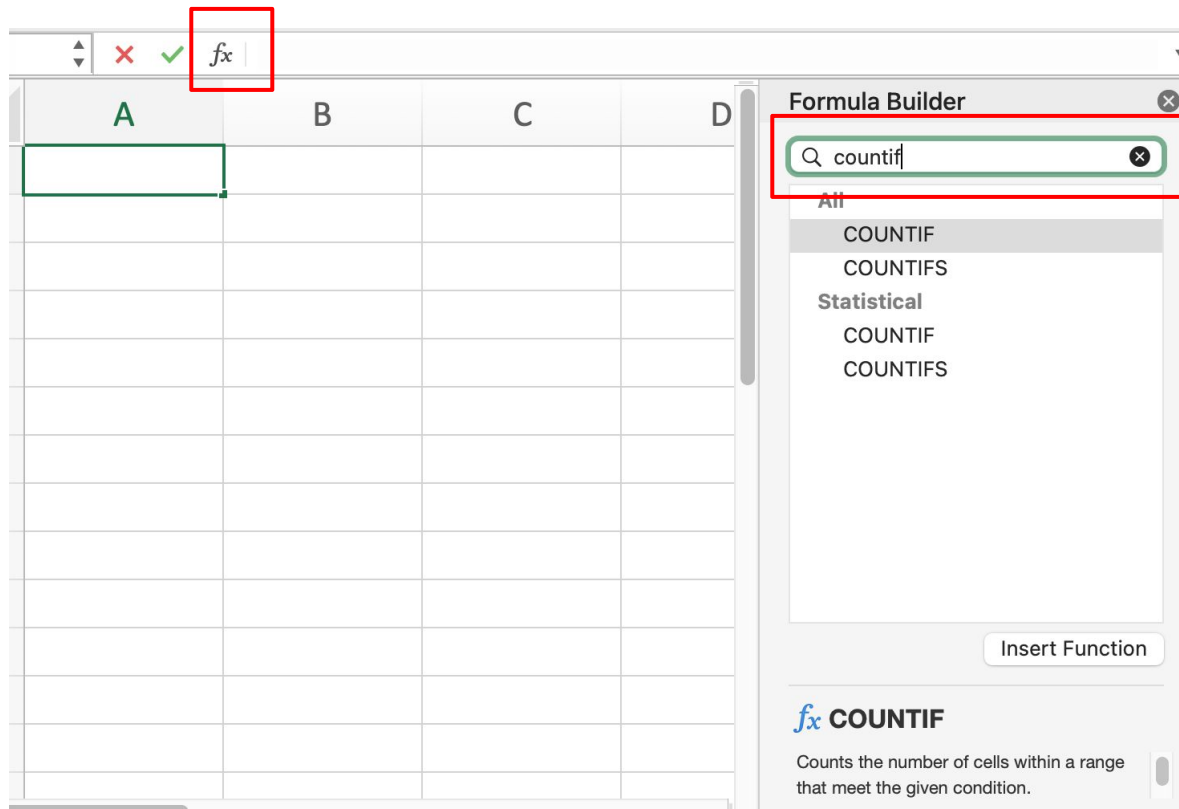
Teach an MPP how to use the Excel help function, and they will complete the problem sets forever.”

— API-201 proverb

# Formula builder

To bring up the Formula Builder, click on the **function button** (fx).

You can **search** for a function by name.





# Formula builder

Click on the function you want, and it will show you the **arguments** you need to enter.

Click on “**More help on this function**” to see examples.

The screenshot shows the Excel Formula Builder interface. The formula bar at the top displays `=COUNTIF()`. The spreadsheet grid shows the formula `=COUNTIF()` entered in cell A1. The Formula Builder pane on the right is titled "Formula Builder" and contains the following elements:

- A button labeled "Show All Functions".
- The function name "COUNTIF".
- A "Range" argument field with the placeholder text "reference". This field is highlighted with a red box.
- A "Criteria" argument field with the placeholder text "any".
- A "Result: {...}" section with a "Done" button.
- A section titled "fx COUNTIF" with a description: "Counts the number of cells within a range that meet the given condition." Below this description is a link labeled "More help on this function", which is also highlighted with a red box.

# Formula builder

Click on the function you want, and it will show you the **arguments** you need to enter.

Click on “**More help on this function**” to see examples.

## Excel for Mac Help

### COUNTIF function

Use COUNTIF, one of the [statistical functions](#), to count the number of cells that meet a criterion; for example, to count the number of times a particular city appears in a customer list.

In its simplest form, COUNTIF says:

- =COUNTIF(Where do you want to look?, What do you want to look for?)

For example:

- =COUNTIF(A2:A5,"London")
- =COUNTIF(A2:A5,A4)

# Cleaning data

i.e., getting the data into a format that is most useful to us

# Recoding variables

There are many ways to encode information.

Sometimes we might want to switch, e.g., from **categorical** variable to a **numeric** variable.

turnout_2020	abstained
voted	0
abstained	1
voted	0
voted	0
voted	0
abstained	1
voted	0

# Recoding variables

We can use the IF function to recode variables.

Note that Excel helpfully gives us the syntax of the function as a pop-up.

fx =IF(A2="abstained", 1,0)	
IF(logical_test, [value_if_true], [value_if_false])	
A	B
<b>turnout_2020</b>	<b>abstained</b>
voted	", 1, 0)
abstained	1
voted	0
voted	0
voted	0
abstained	1
voted	0

# Recoding variables

There are many ways to encode information.

Sometimes we might want to **collapse** a variable down into fewer categories.

vote_2020	vote_2020_simplified
Republican	Republican
Democrat	Democrat
Democrat	Democrat
Independent	Other
Green	Other
Republican	Republican
Libertarian	Other
Democrat	Democrat
Republican	Republican
Republican	Republican

# Recoding variables

But what if we have more than two categories?

We could nest multiple IF functions, but that gets messy!

Instead, we can use the VLOOKUP function with a lookup table.

vote_2020	vote_2020_simplified		
Republican	\$E\$8, 2, FALSE)		
Democrat	Democrat		
Democrat	Democrat		
Independent	Other		
Green	Other		
Republican	Republican		
Libertarian	Other		
Democrat	Democrat		
Republican	Republican		
Republican	Republican		

Lookup table	
Republican	Republican
Democrat	Democrat
Independent	Other
Green	Other
Libertarian	Other

```
=VLOOKUP(A2,$D$4:$E$8, 2, FALSE)
```

```
VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])
```

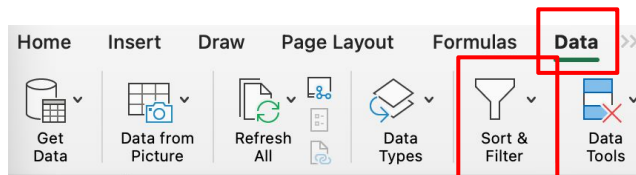
# Analyzing data

i.e., calculating descriptive statistics,  
performing statistical tests



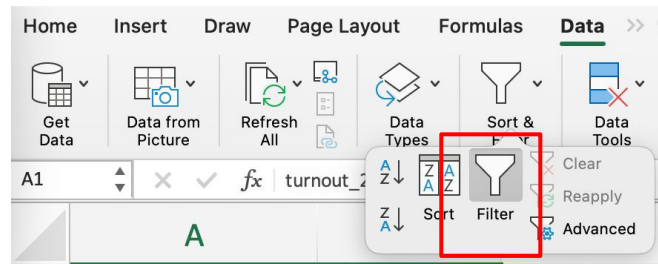
# Tables

Turning our data into a table allows us to sort and filter by different variables



The screenshot shows the Excel ribbon with the 'Data' tab selected. The 'Sort & Filter' icon, which is a funnel, is highlighted with a red rectangular box. Below the ribbon, the formula bar shows 'turnout\_2020'.

	A	B	C
1	turnout_2020	age	
2	voted	24	
3	abstained	55	
4	voted	72	
5	voted	19	
6	voted	44	
7	abstained	37	
8	voted	65	



The screenshot shows the Excel ribbon with the 'Data' tab selected. The 'Filter' icon, which is a funnel, is highlighted with a red rectangular box. A dropdown menu is open, showing options for 'Sort' and 'Filter'. The 'Filter' option is selected, and the table below is filtered to show only rows where 'turnout\_2020' is 'voted'.

	A	B	C
1	turnout_2020	age	
2	voted	24	
3	abstained	55	
4	voted	72	
5	voted	19	
6	voted	44	
7	abstained	37	
8	voted	65	

# Tables

Tables are a great way to **explore** a dataset.

However, if you have a BIG dataset with LOTS of variables, using a table can be slow and confusing.

In these cases, we probably want to use **functions** instead.

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. A table is displayed with the following data:

	A	B
1	turnout_2020	age
2	voted	24
4	voted	72
5	voted	19
6	voted	44
8	voted	65
9		
10		
11		
12		
13		

The 'Sort' dialog box is open, showing the following settings:

- Sort: A-Z Ascending, Z-A Descending
- By colour: None
- Filter: By colour: None
- Equals: voted
- And / Or: And
- Choose One: (empty)
- Search: (empty)
- Filter list: (Select All), ☒ abstained, ☒ voted
- Auto Apply: ☒
- Buttons: Apply Filter, Clear Filter

# Descriptive statistics

Function	What does it do?
COUNT	Counts the number of cells that contain a number
SUM	Adds up the cells
AVERAGE	Takes the mean of the cells
STDEV.P / STDEV.S	Standard deviation for a population / a sample
MEDIAN	Takes the median of the cells

# Descriptive statistics

Function	What does it do?
COUNT / COUNTIF / COUNTIFS	Counts the number of cells that contain a number (or meet a certain condition)
SUM / SUMIF / SUMIFS	Adds up the cells (or the cells that meet a certain condition)
AVERAGE / AVERAGEIF / AVERAGEIFS	Takes the mean of the cells (or the cells that meet a certain condition)
STDEV.P / STDEV.S	Standard deviation for a population / a sample
MEDIAN	Takes the median of the cells

# Statistical tests (don't need to worry about these yet!)



Function	What does it do?
NORM.S.INV	Inverse of the standard normal distribution, used for finding critical values
NORM.S.DIST	Standard normal distribution, used for finding p-values
CHISQ.TEST	Chi-squared test

# Understanding check: let's figure out how to use AVERAGEIF!

Suppose we want to calculate the **average age** of the individuals in our data **who voted** in 2020.

We can use AVERAGEIF! But how does the syntax work?

Let's use the Formula Builder...

turnout_2020 	age 
voted	24
abstained	55
voted	72
voted	19
voted	44
abstained	37
voted	65

# Understanding check: let's figure out how to use AVERAGEIF!

Formula Bar: `=AVERAGEIF(A2:A8,"voted",B2:B8)`

A	B	C	D	E
turnout_2020	age			
voted	24			
abstained	55			
voted	72			
voted	19			
voted	44			
abstained	37			
voted	65			

**Average age of...**

**voters**      **abstainers**

`d",B2:B8)`

### Formula Builder

Show All Functions

**AVERAGEIF**

**Range** = {"voted";"abstained";"voted";"vot...  
`A2:A8`

**Criteria** = "voted"  
"voted"

**Average\_range** = {24;55;72;19;44;37;65}  
`B2:B8`

Result: 44.8      Done

**fx AVERAGEIF**

Finds average (arithmetic mean) for the cells specified by a given condition or

[More help on this function](#)

Sheet3   Sheet2   Sheet4   **Sheet5**   +

## Understanding check: let's figure out how to use AVERAGEIF!

**turnout\_2020**

A	B	C	D	E
voted	24			
abstained	55			
voted	72			
voted	19			
voted	44			
abstained	37			
voted	65			

**Average age of... voters abstainers**

**44.8**

**Formula Builder**

Show All Functions

**AVERAGEIF**

Range = {"voted","abstained","voted","vot..."  
A2:A8

Criteria = "voted"  
"voted"

Average\_range = {24;55;72;19;44;37;65}  
B2:B8

Result: 44.8 Done

**f<sub>x</sub> AVERAGEIF**

Finds average (arithmetic mean) for the cells specified by a given condition or

[More help on this function](#)



# Understanding check: let's figure out how to use AVERAGEIF!

To do the same calculation for the abstainers, I made 2 changes:

1. I added \$ to the cell references to keep it fixed when I copied it!
2. I changed “voted” to “abstained”

Easy!

=AVERAGEIF(\$A\$2:\$A\$8,"abstained",\$B\$2:\$B\$8)				
	B	C	D	E
age				
	24			
	55	Average age of...		
	72	voters		abstainers
	19		44.8	46
	44			
	37			
	65			

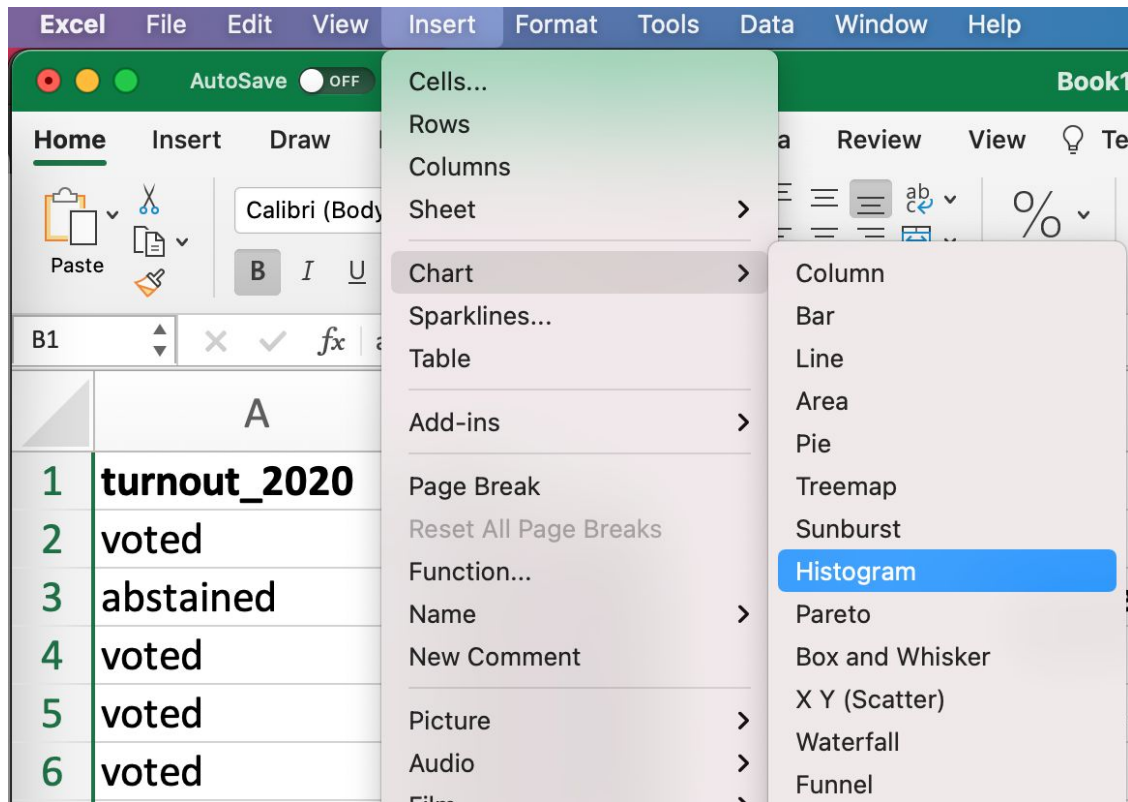
# Visualizing data

i.e., making graphs!

# Charts: one variable

Highlight the column  
we want to display

Go Insert → Chart →  
Histogram

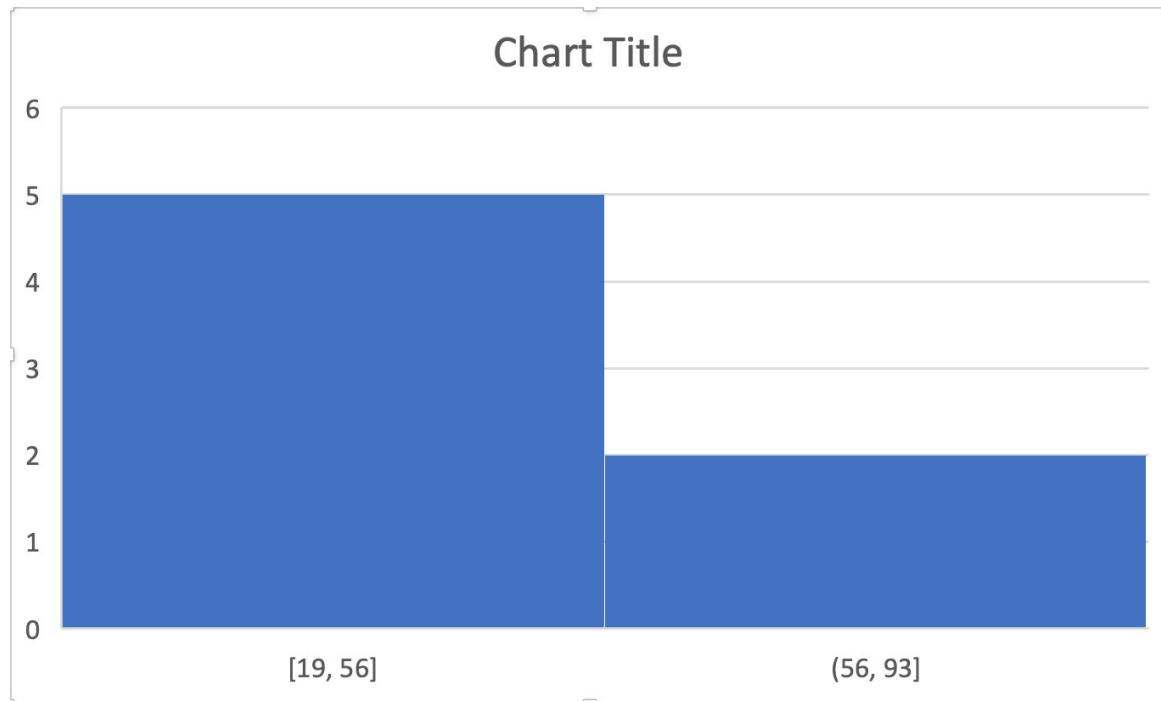


# Charts: one variable

Excel will spit out  
some default chart.

It will not be pretty!

But that's ok.



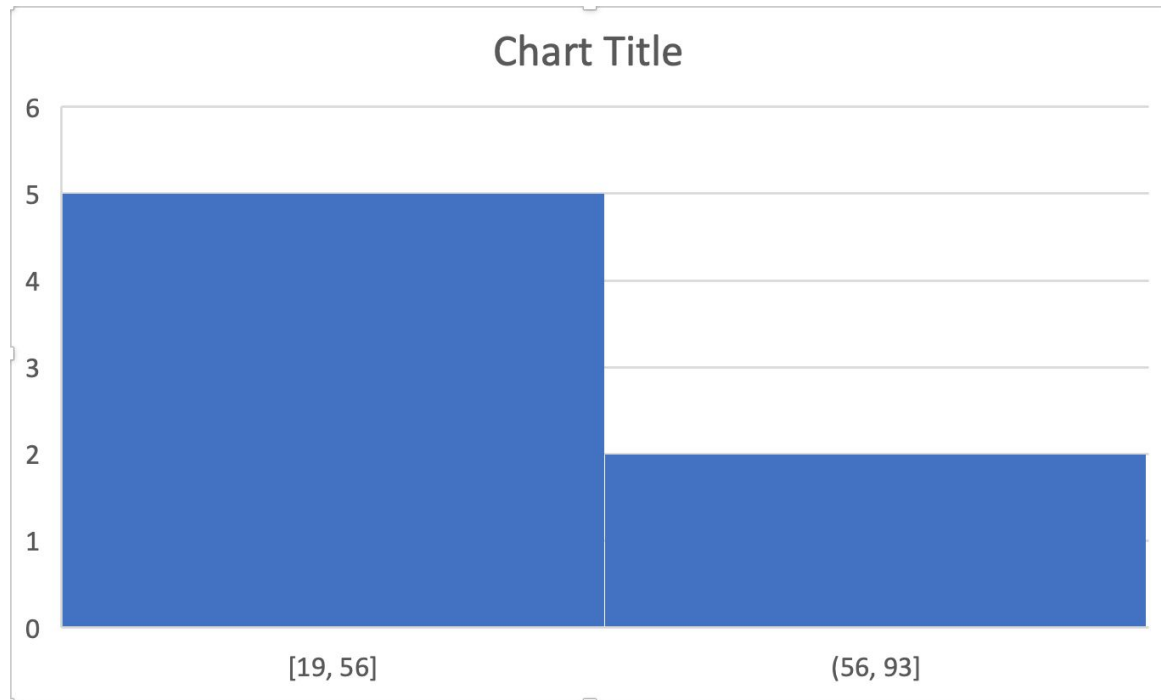
# Charts: one variable

Protip: to change something on a chart, **double-click** on that element!

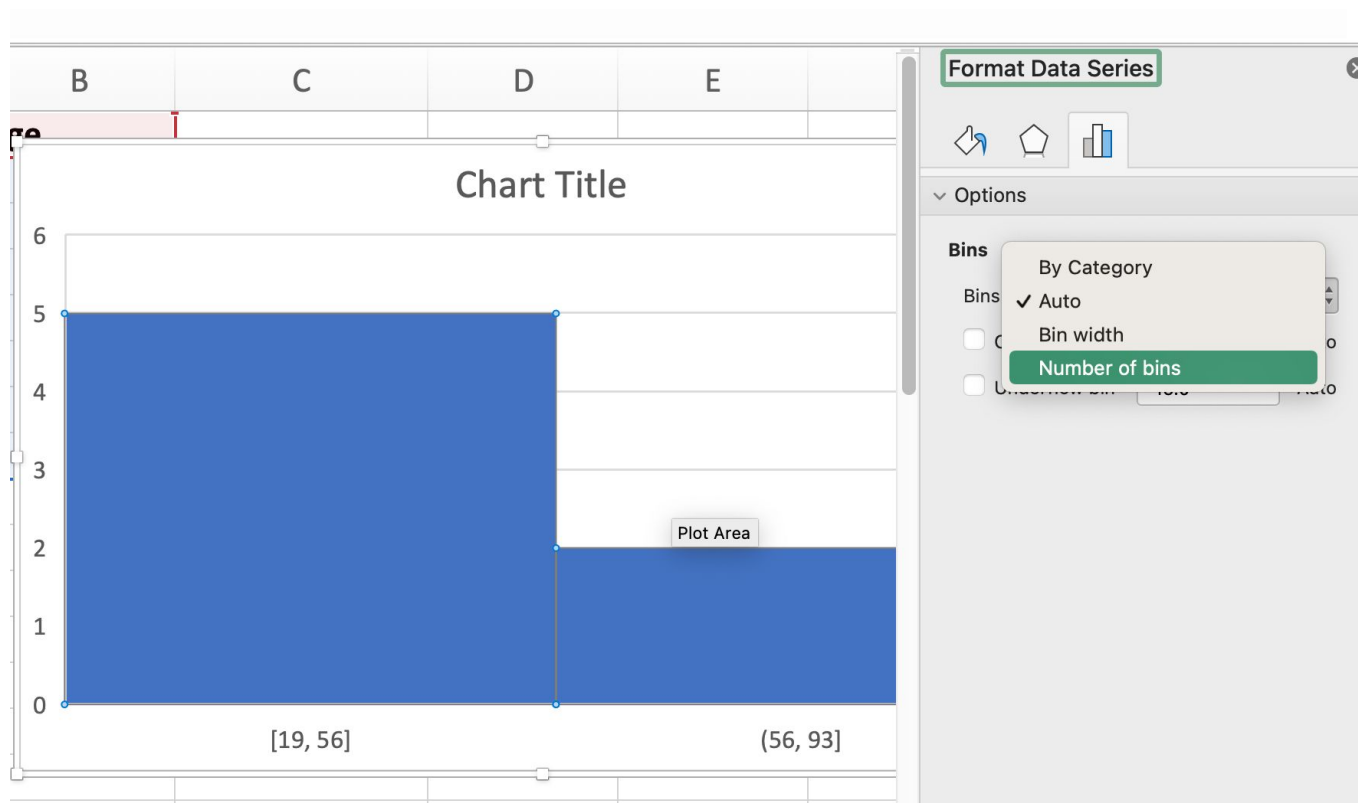
Excel will spit out  
some default chart.

It will not be pretty!

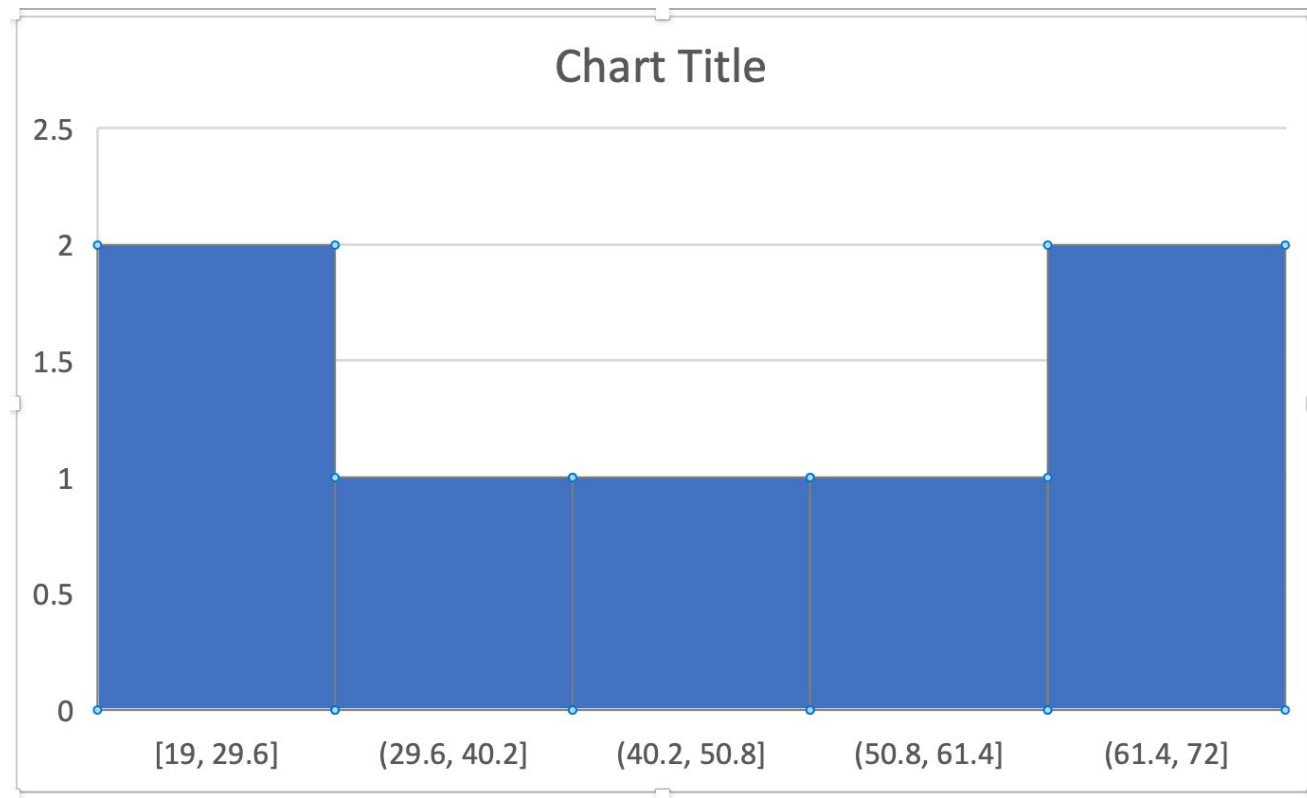
But that's ok.



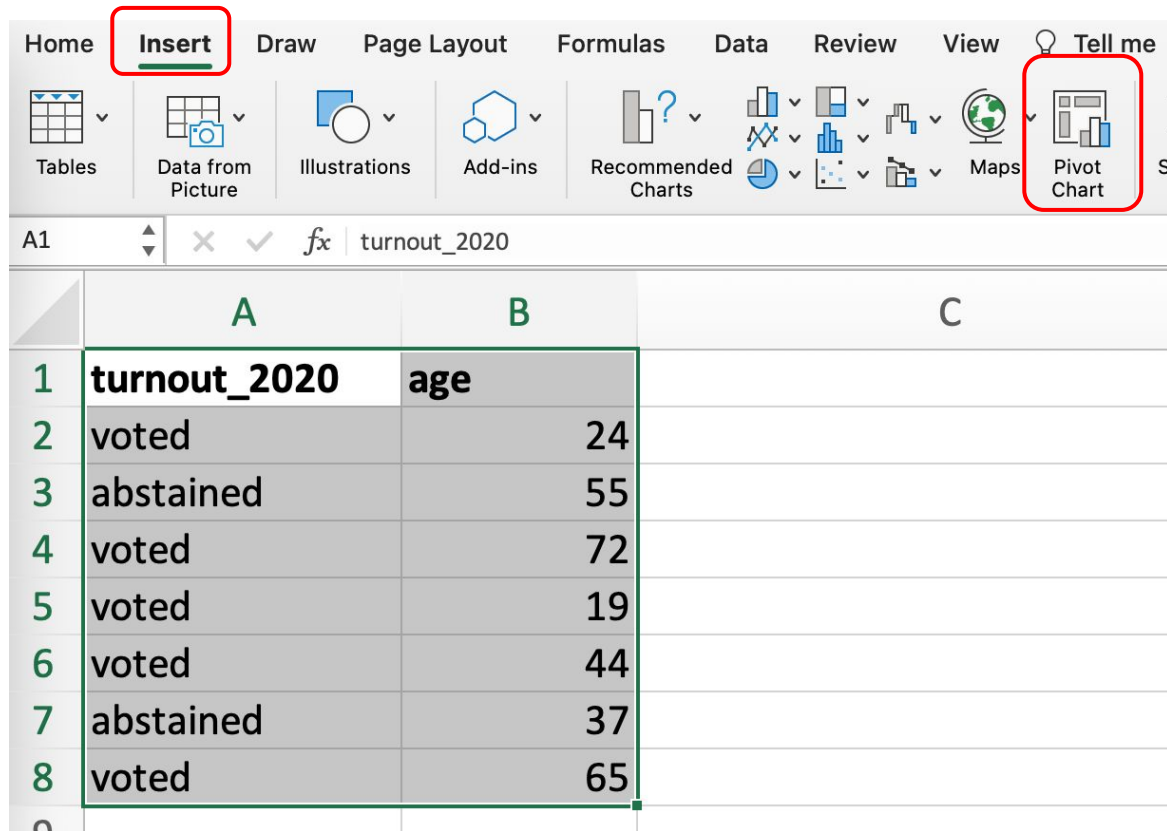
# Charts: one variable



# Charts: one variable



# Charts: multiple variables



The screenshot displays the Microsoft Excel interface. The **Insert** tab is selected in the ribbon, and the **Pivot Chart** icon is highlighted with a red box. Below the ribbon, the active cell is A1, and the formula bar shows the text `turnout_2020`. The data table is as follows:

	A	B	C
1	turnout_2020	age	
2	voted	24	
3	abstained	55	
4	voted	72	
5	voted	19	
6	voted	44	
7	abstained	37	
8	voted	65	



Chart Name:

### Chart 1

Options

Active Field:

Field

 Expand Field

 Collapse Field

Insert

Insert

Filter

Refresh

Change

Clear

Move

Fields,

Field

Chart 1     

PivotTable2

To build a report, choose fields from the PivotTable Field List

To build a PivotChart, choose fields from the PivotChart Field List.

Format Chart Area 

## PivotChart Fields

## Chart Options

Text Options

> Fill

- > Border

# Charts: multiple variables

