Excel training session

API-201, Fall 2021
TF: Sophie Hill

Plan for today

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

Rule #1:
Work efficiently

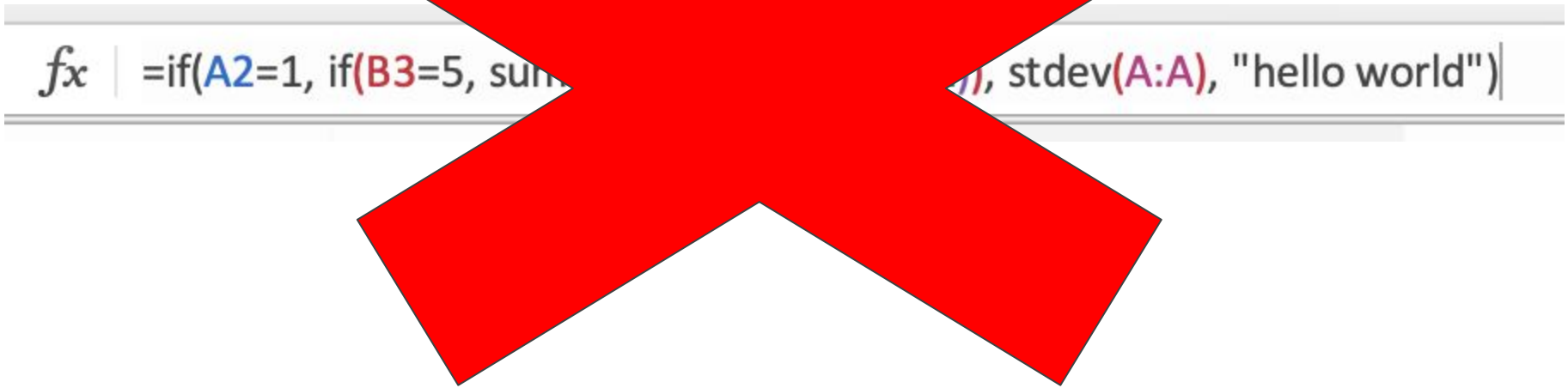
Rule #1: Work efficiently

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


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


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

Rule #2: Work transparently

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

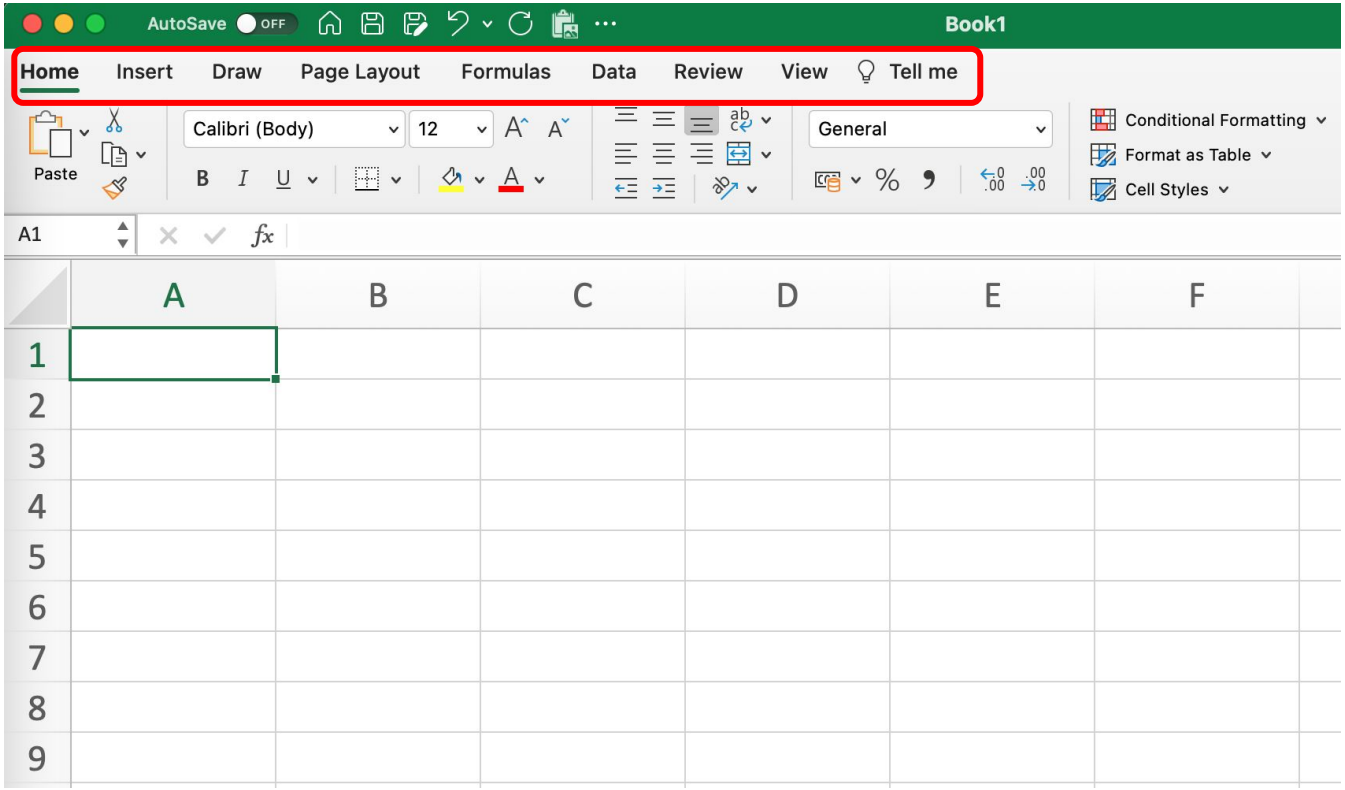
- Break calculations down into smaller steps
- Use sensible column/sheet names

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.

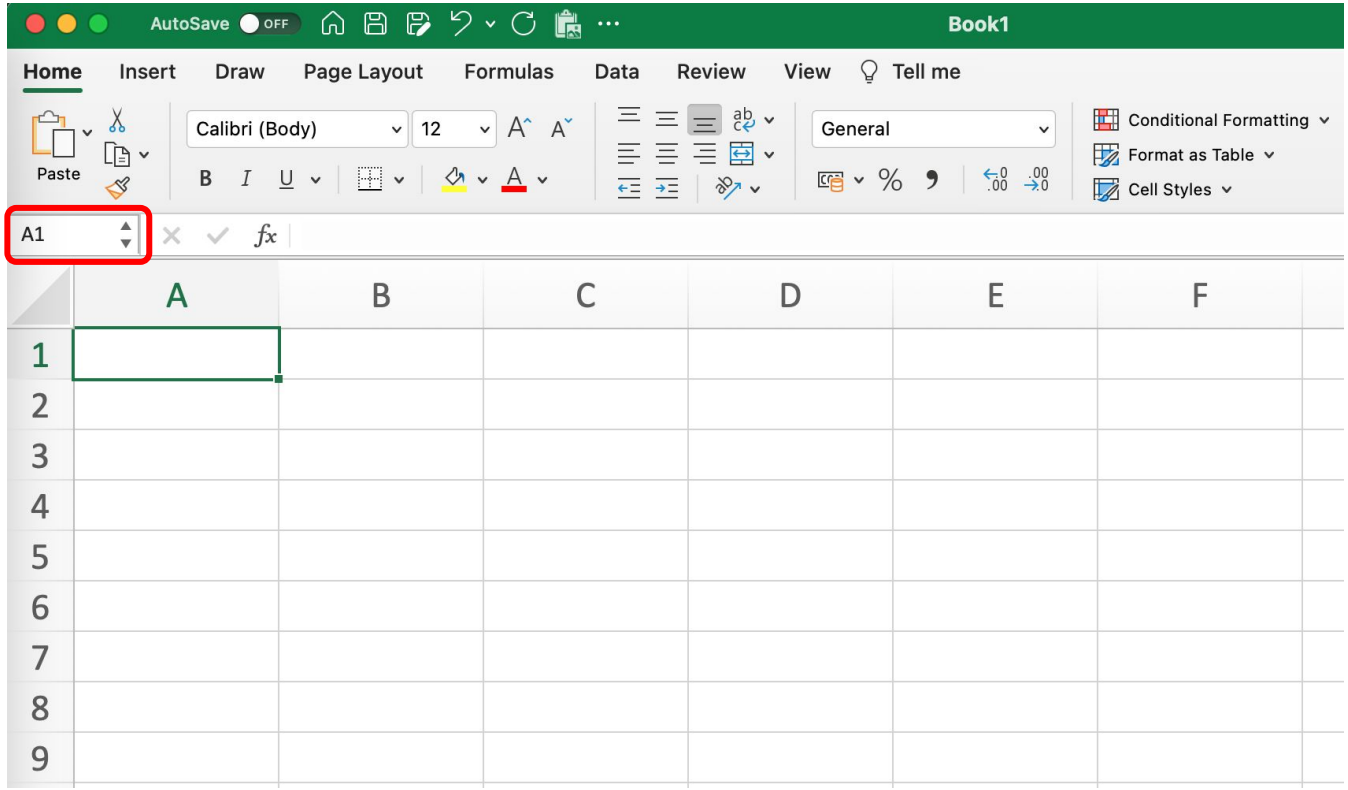
Getting oriented

Menu bar



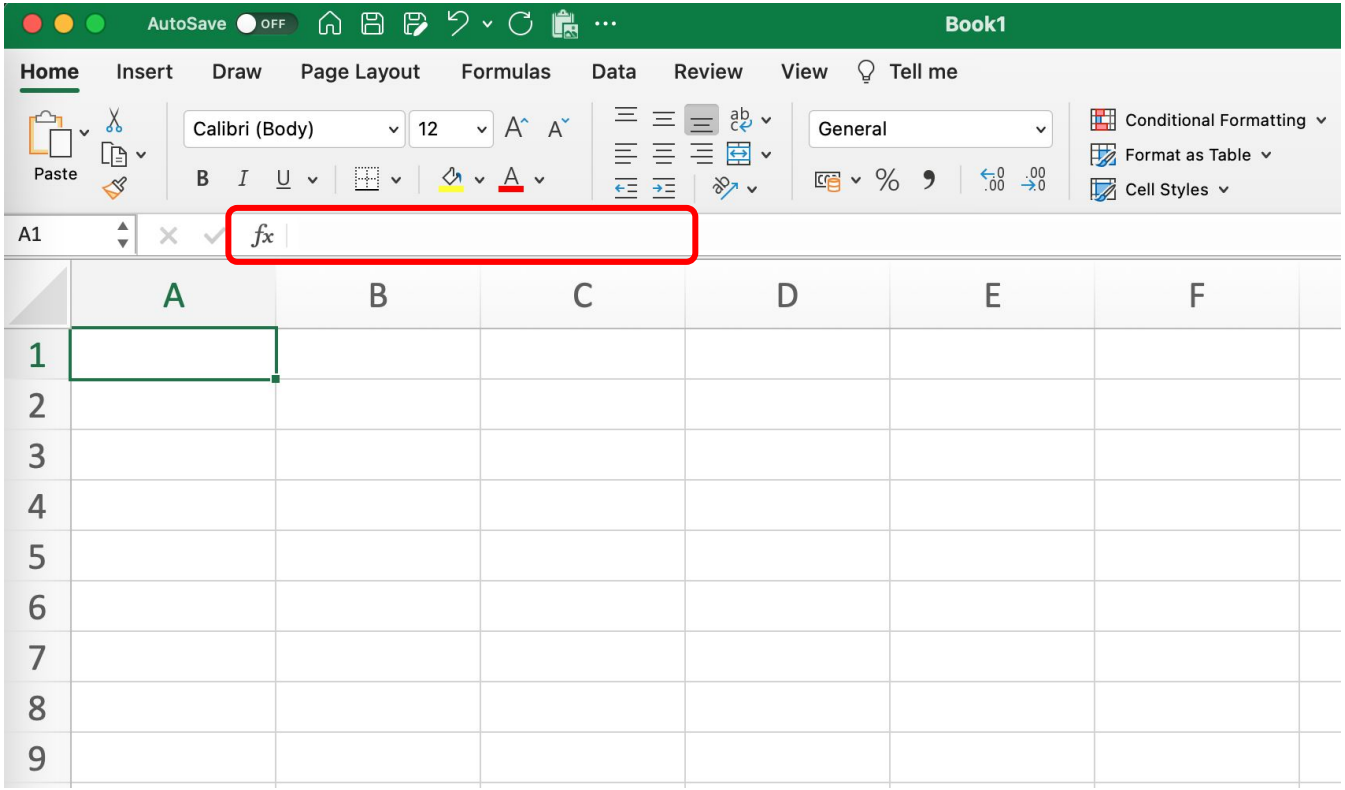
Getting oriented

Highlighted cell



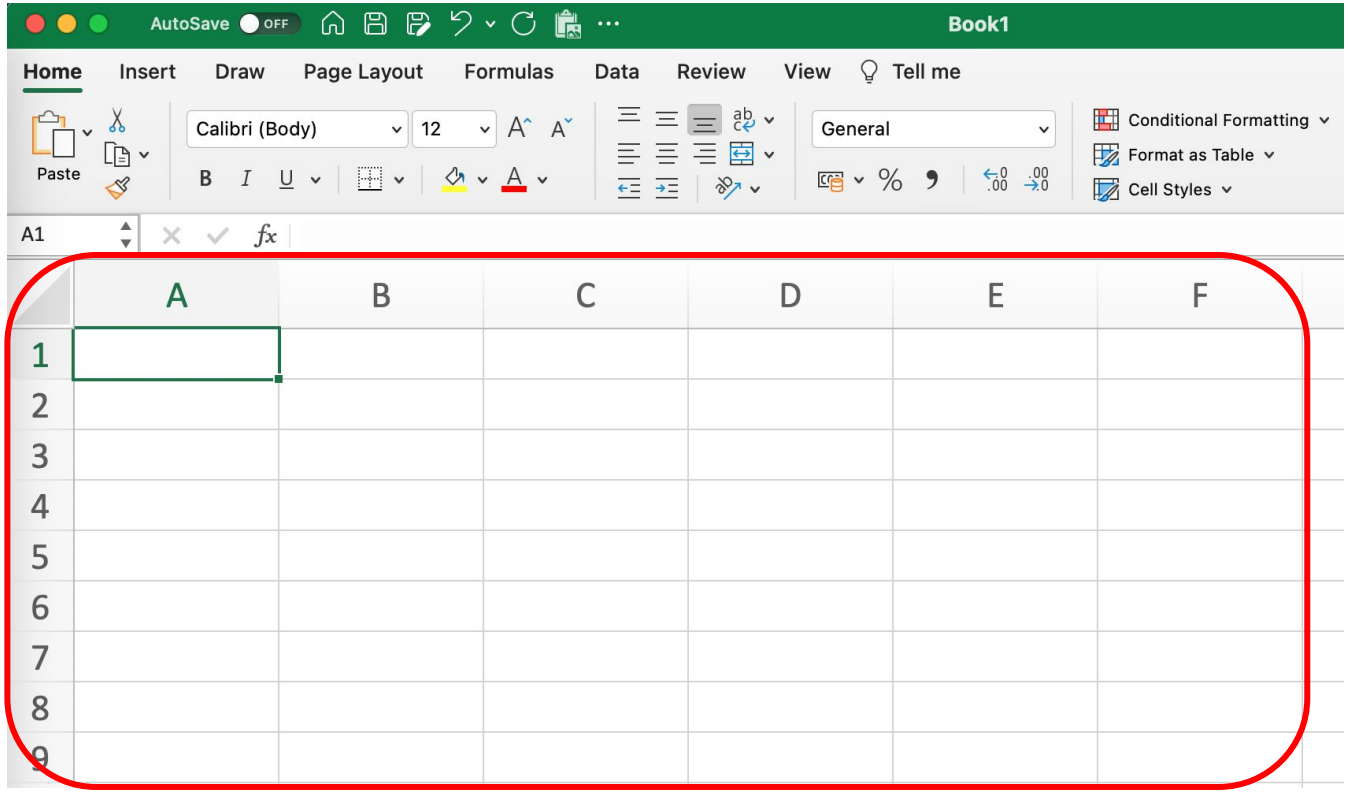
Getting oriented

Formula bar



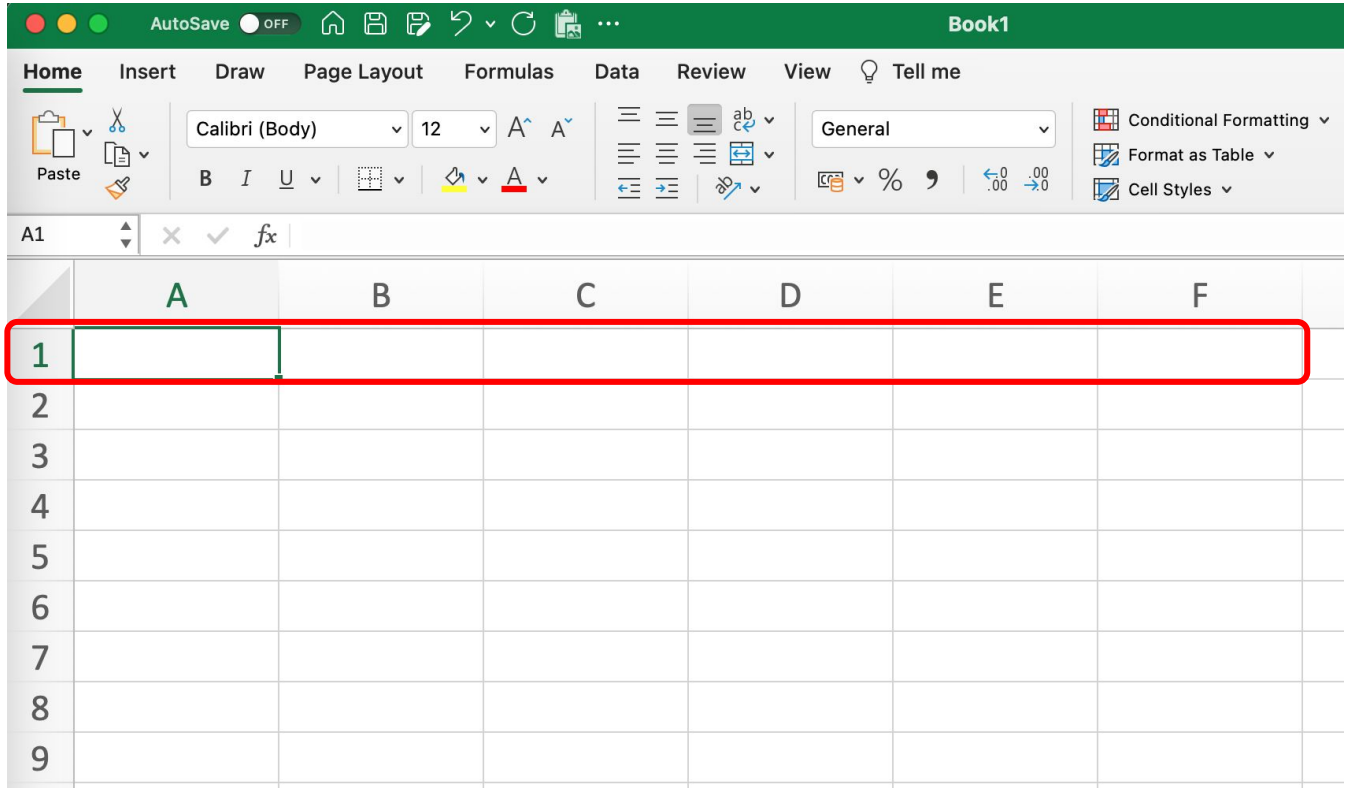
Getting oriented

Sheet



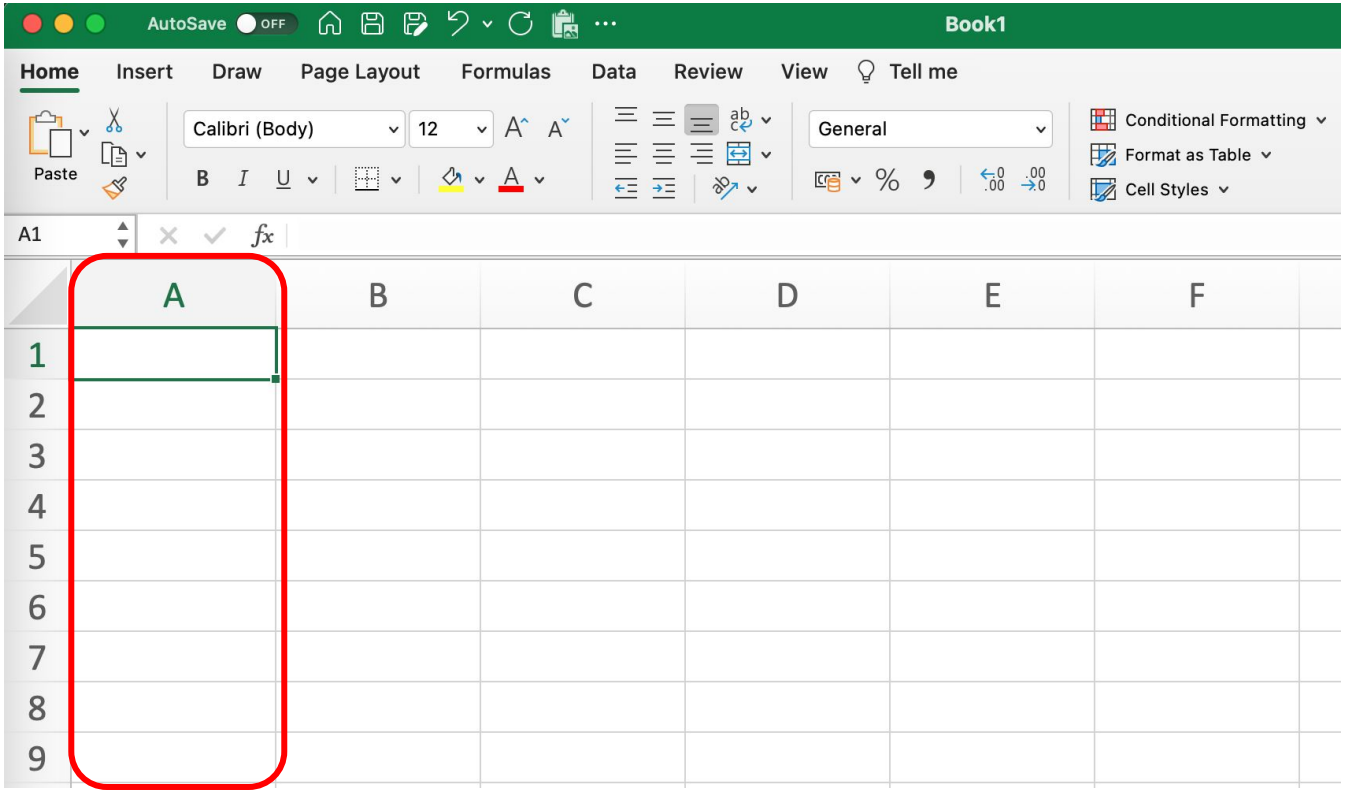
Getting oriented

Rows (numbers)

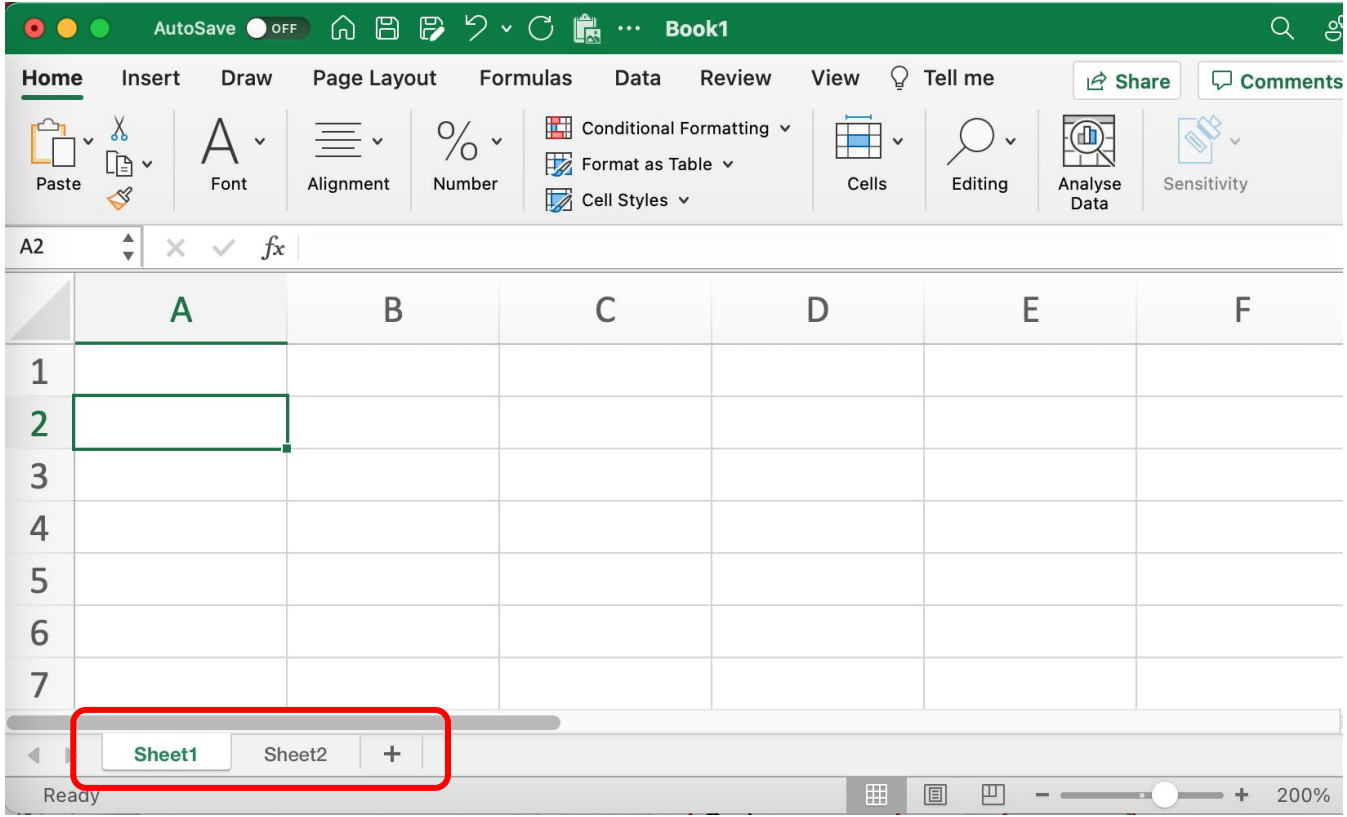


Getting oriented

Columns (letters)



Getting oriented



Tabs

Syntax: referencing cells

A1	The cell A1
A1:A3	The cells A1, A2, A3
A:A	
A:B	
A1:B7	
A1	
Sheet1!A1	

Syntax: referencing cells

A1	The cell A1
A1:A3	The cells A1, A2, A3
A:A	The whole A column
A:B	
A1:B7	
A1	
Sheet1!A1	

Syntax: referencing cells

A1	The cell A1
A1:A3	The cells A1, A2, A3
A:A	The whole A column
A:B	The whole of columns A and B
A1:B7	
A1	
Sheet1!A1	

Syntax: referencing cells

A1	The cell A1
A1:A3	The cells A1, A2, A3
A:A	The whole A column
A:B	The whole of columns A and B
A1:B7	The array covering A1 down to A7 and B1 down to B7
A1	
Sheet1!A1	

Syntax: referencing cells

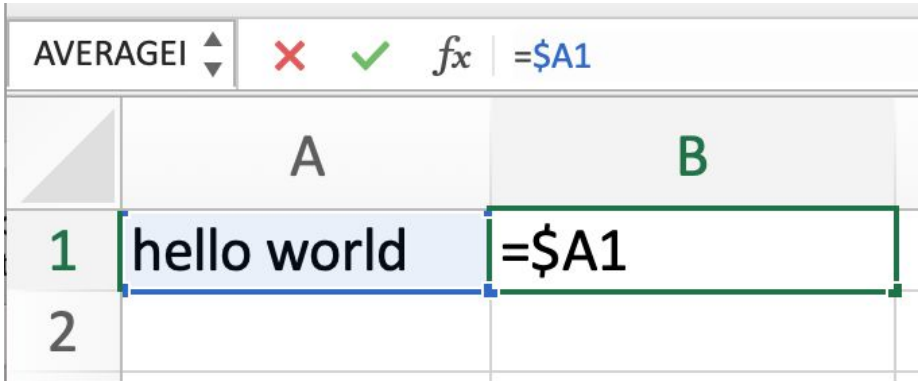
A1	The cell A1
A1:A3	The cells A1, A2, A3
A:A	The whole A column
A:B	The whole of columns A and B
A1:B7	The array covering A1 down to A7 and B1 down to B7
A1	A fixed reference to cell A1 that won't change when you copy a formula to another cell
Sheet1!A1	

Syntax: referencing cells

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

Syntax: understanding check!

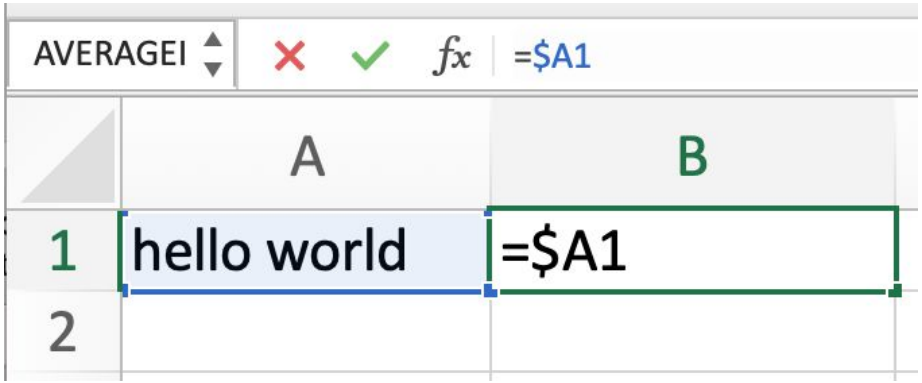
Cell A1 contains a text string (“hello world”).
Cell B1 contains a formula (“=$A1”).



Syntax: understanding check!

Cell A1 contains a text string (“hello world”).
Cell B1 contains a formula (“=$A1”).

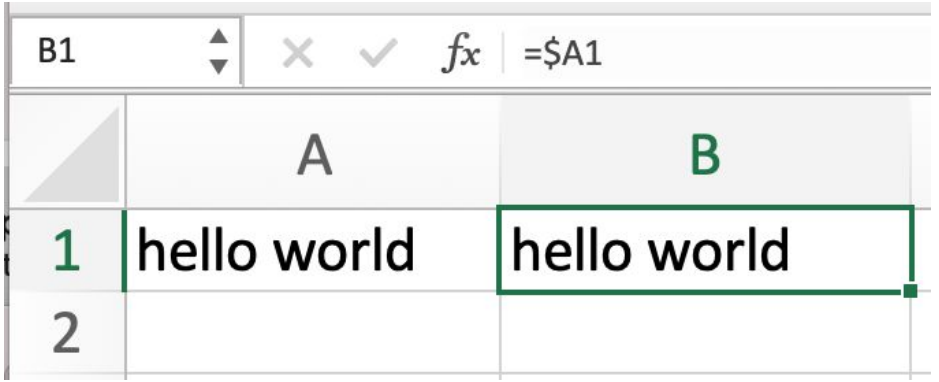
1. What will B1 evaluate to?



Syntax: understanding check!

Cell A1 contains a text string (“hello world”).
Cell B1 contains a formula (“=$A1”).

1. What will B1 evaluate to?

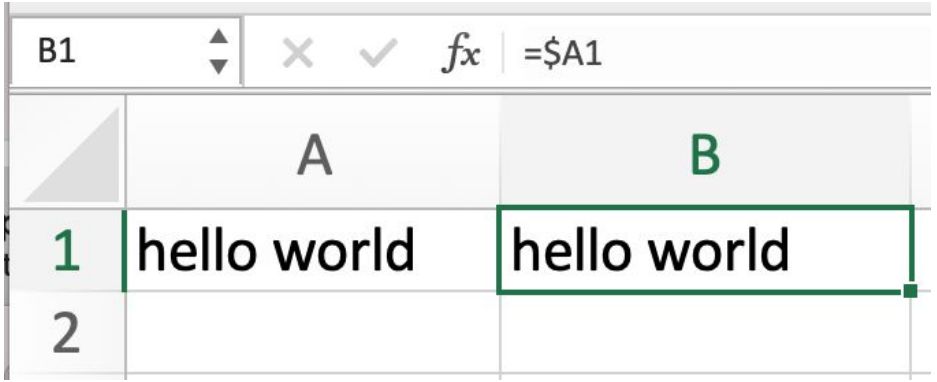


Syntax: understanding check!

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?

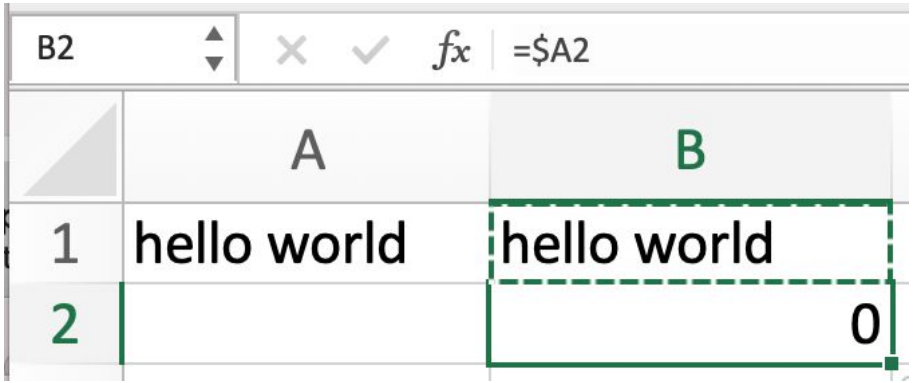


Syntax: understanding check!

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?



Syntax: understanding check!

Cell A1 contains a text string (“hello world”).
Cell B1 contains a formula (“=$A1”).

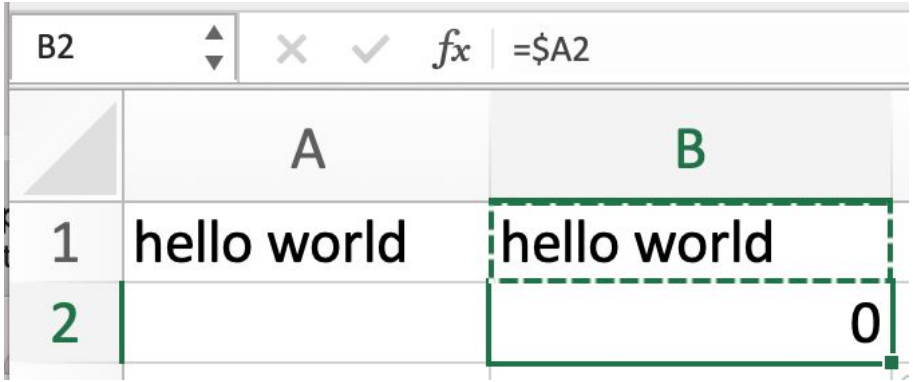
1. What will B1 evaluate to?

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.



Syntax: fixed references

A1	Neither row nor column is fixed
$A1	Column is fixed
A$1	Row is fixed
A1	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.

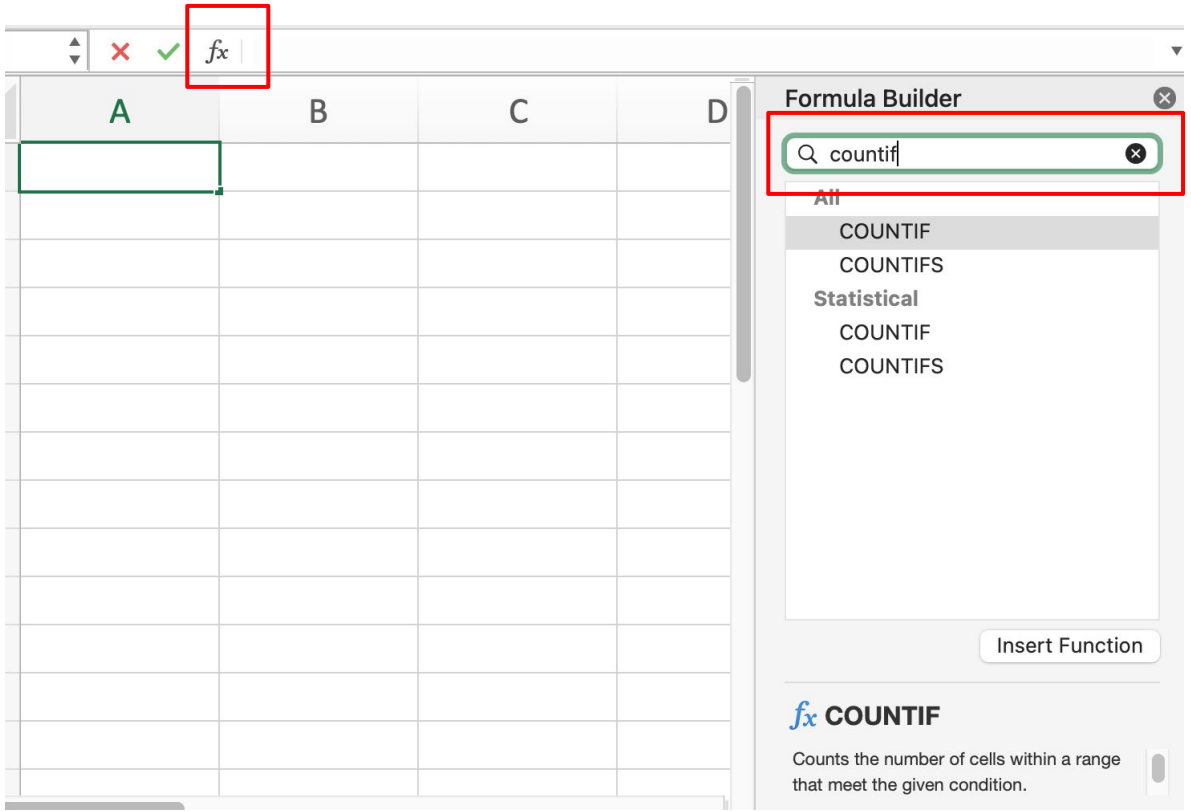
Teach an MPP how to use the Excel help function, and they will complete the problem sets forever.”

— API-201 proverb

Formula builder

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

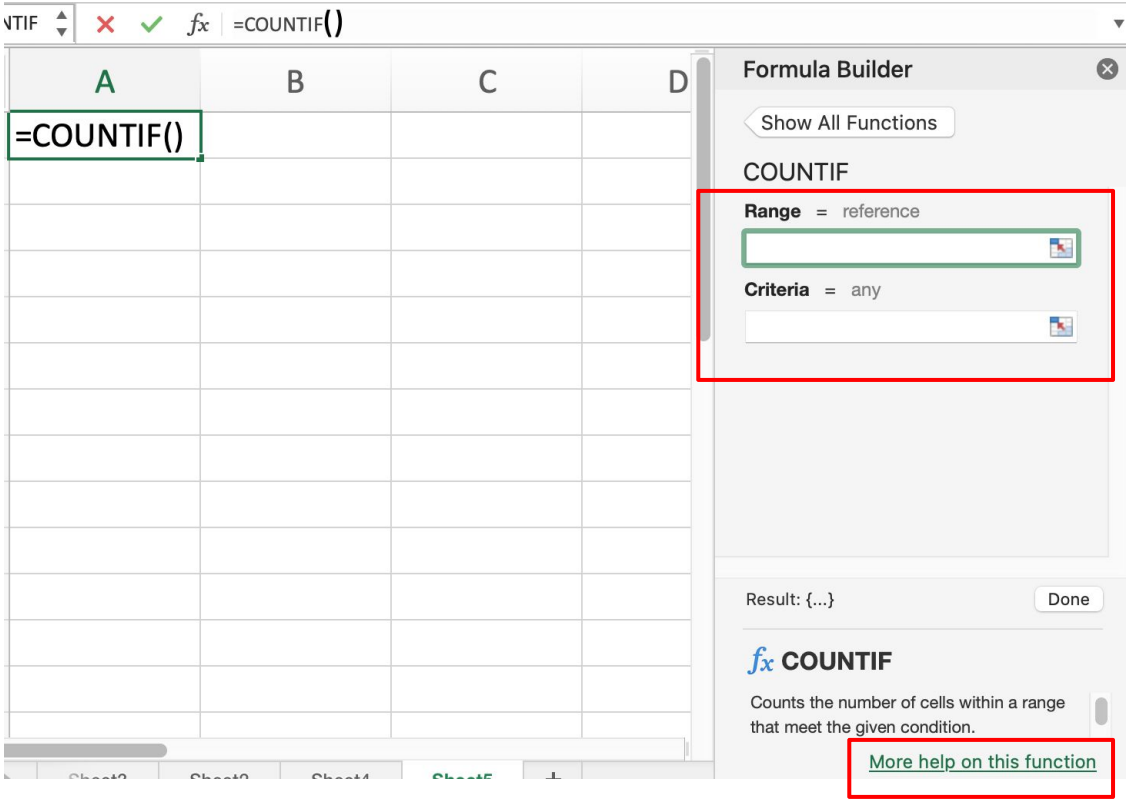
You can search for a function by name.



Formula builder

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

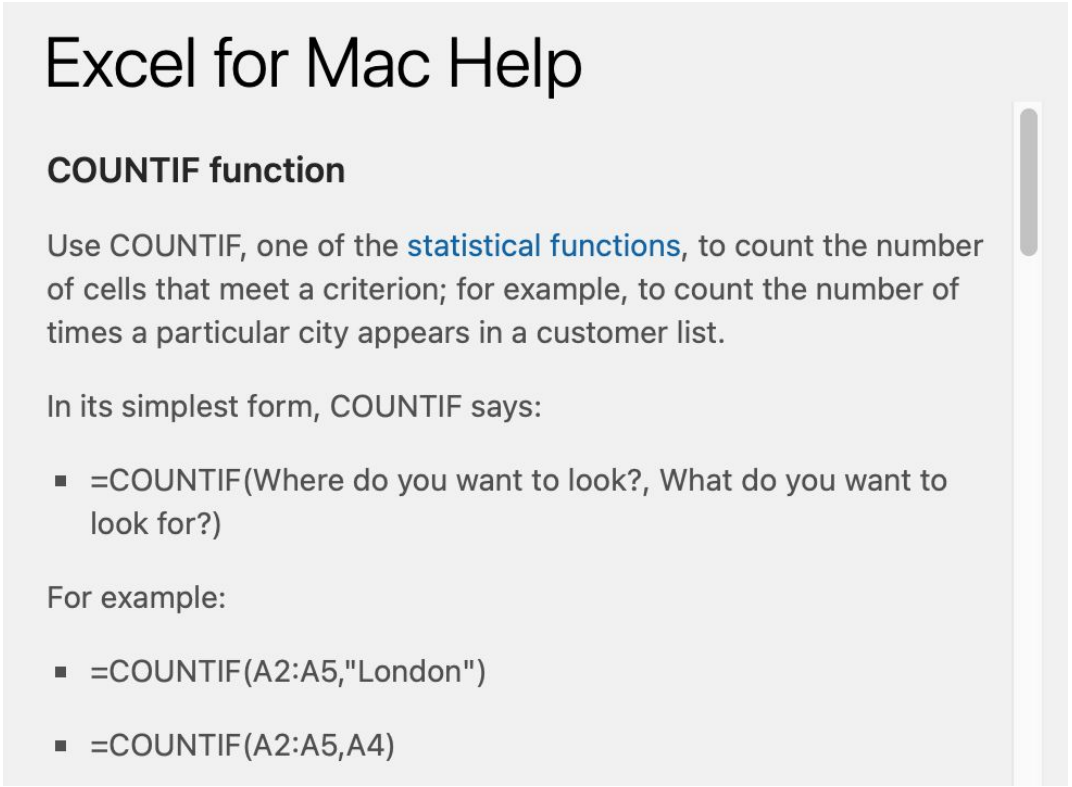
Click on “More help on this function” to see examples.



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.



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.

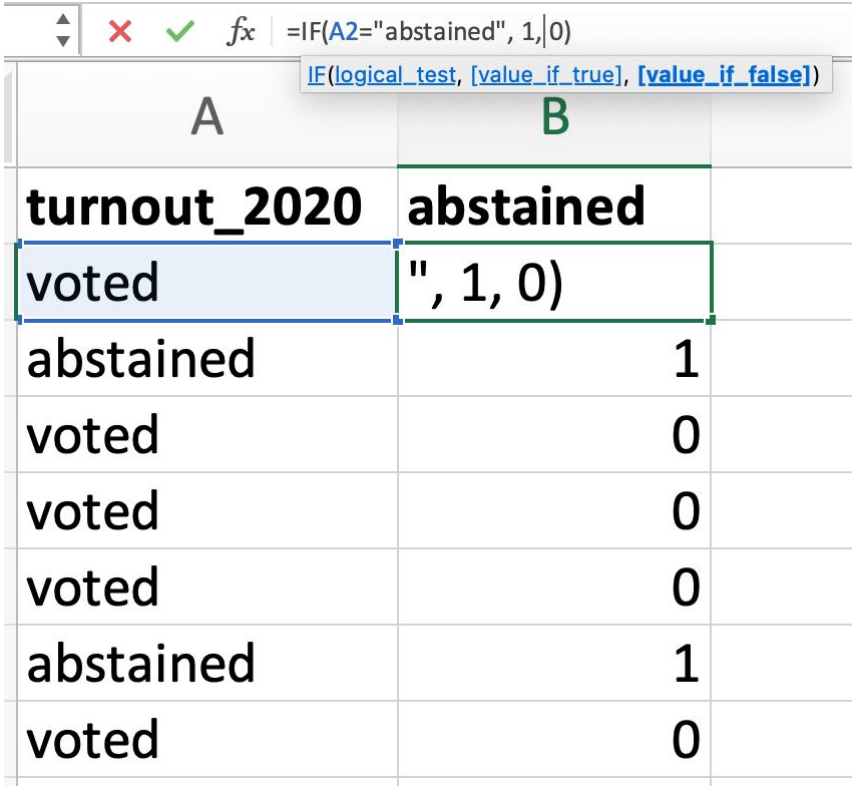
Sometimes we might want to switch, e.g., from **categorical** variable to a **numeric** variable.

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.



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	E8, 2, FALSE)
Democrat	Democrat
Democrat	Democrat
Independent	Other
Green	Other
Republican	Republican
Libertarian	Other
Democrat	Democrat
Republican	Republican
Republican	Republican

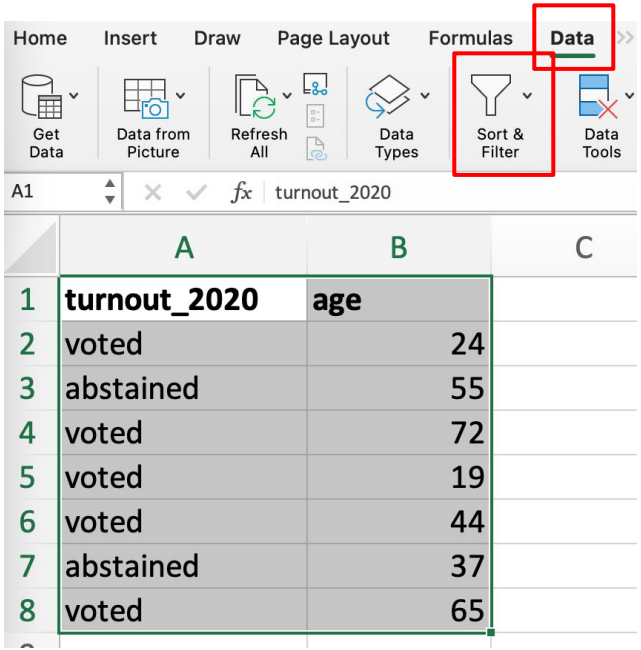
Lookup table	
Republican	Republican
Democrat	Democrat
Independent	Other
Green	Other
Libertarian	Other

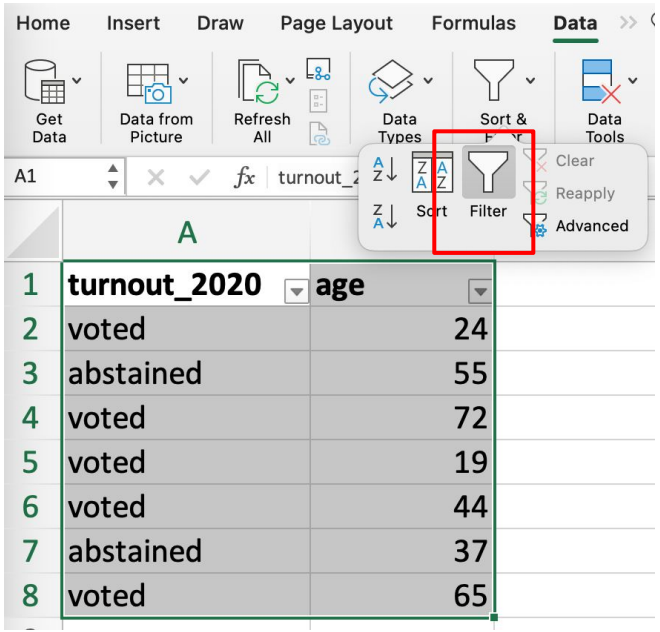
Analyzing data

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

Tables

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



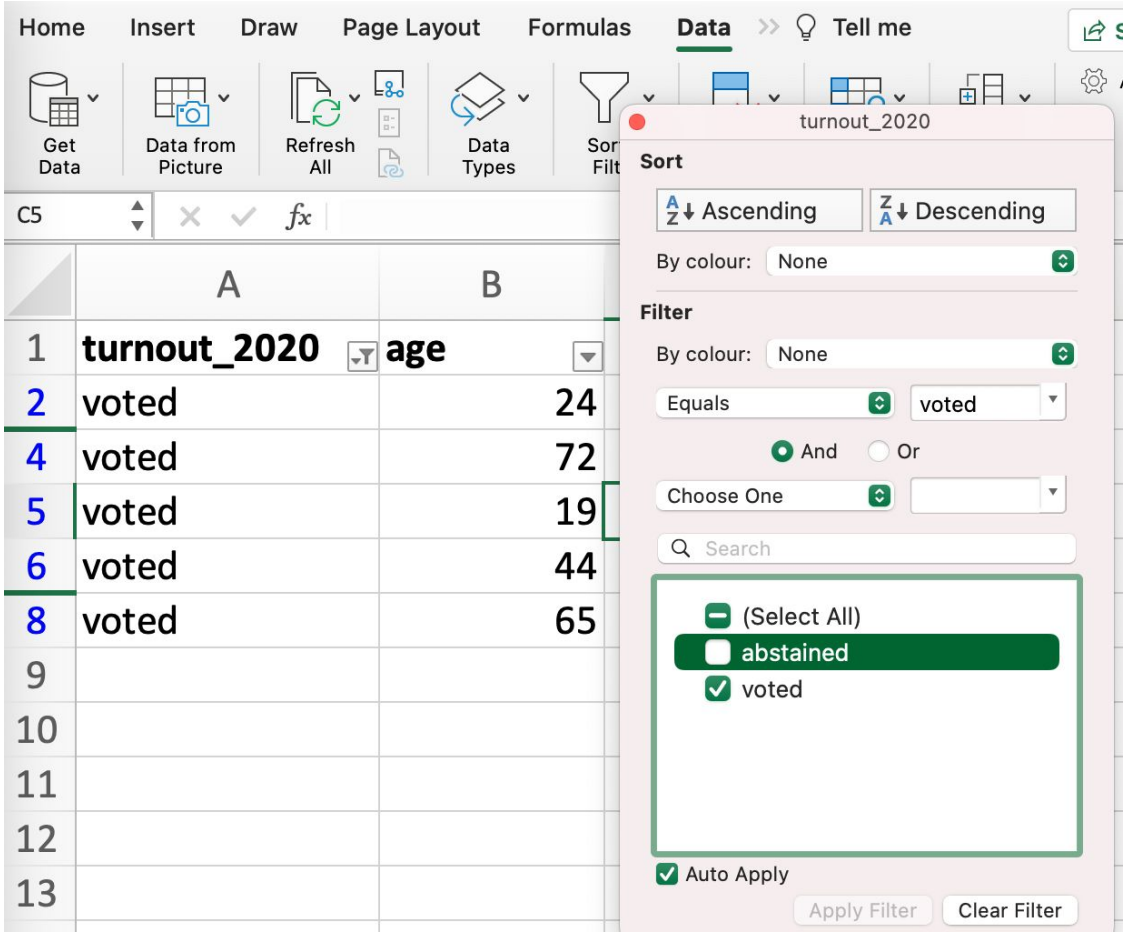


Tables

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.



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!

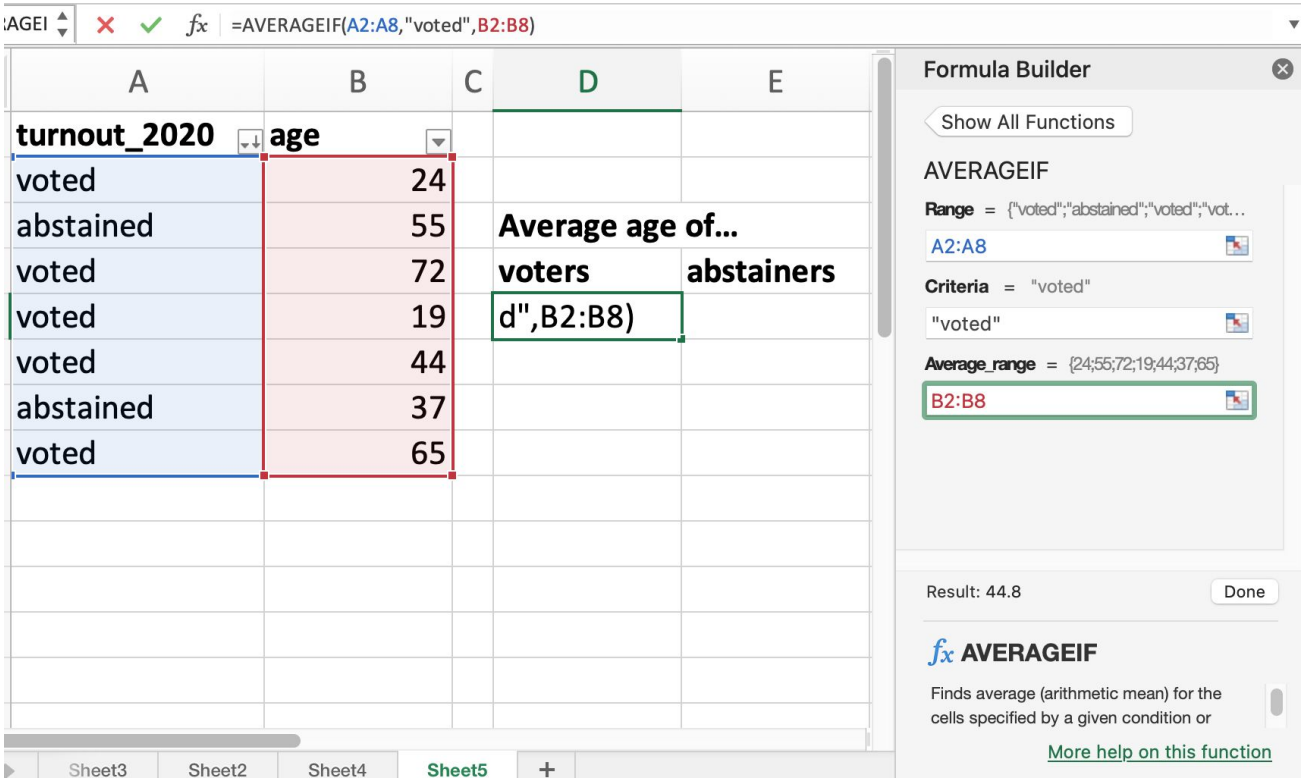
Suppose we want to calculate the **average age** of the individuals in our data **who voted** in 2020.

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

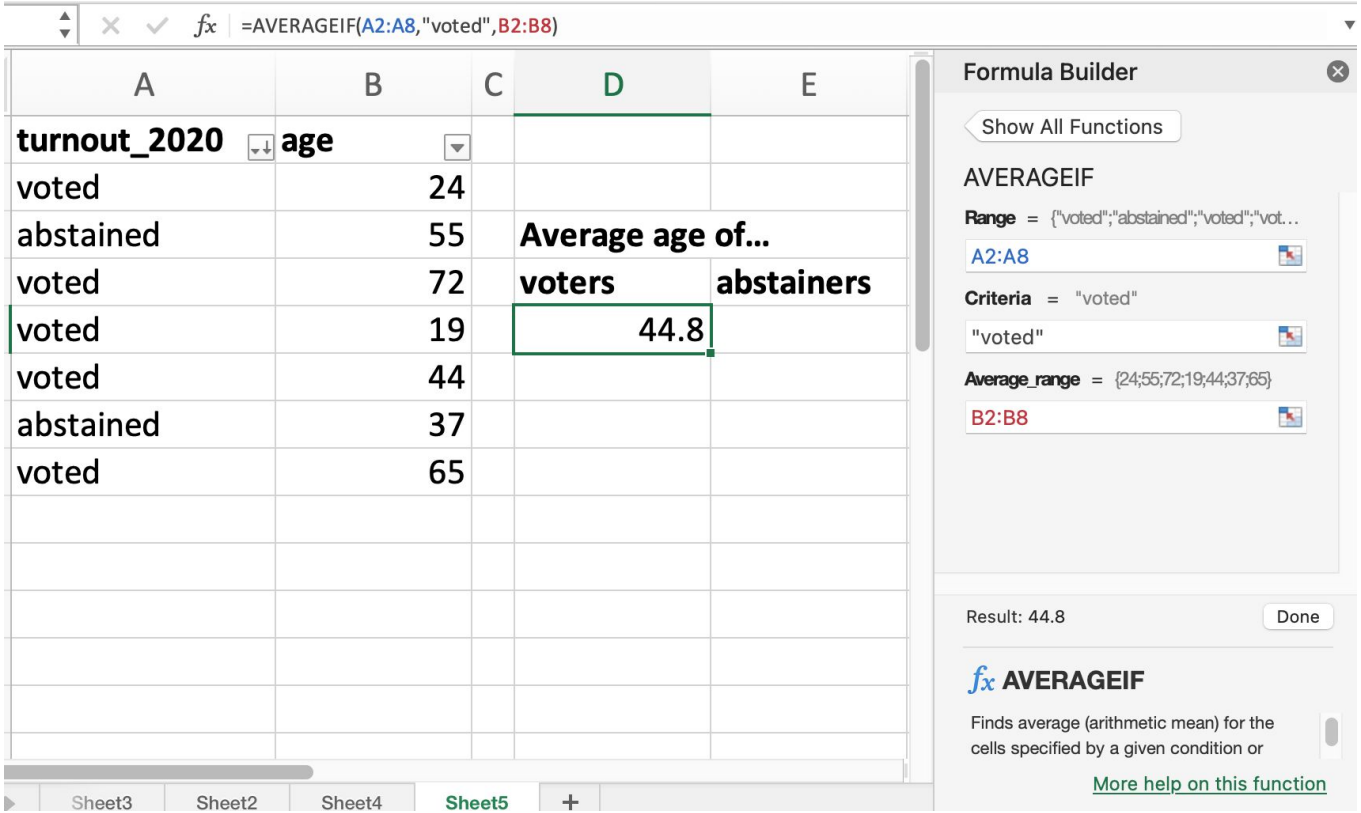
Let's use the Formula Builder...

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!



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

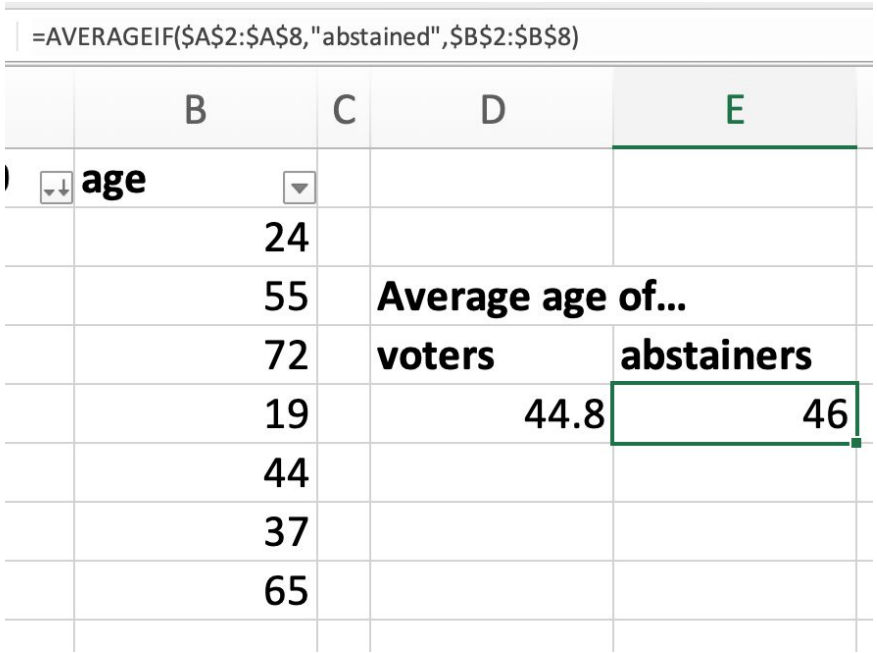


Understanding check: let's figure out how to use AVERAGEIF!

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!



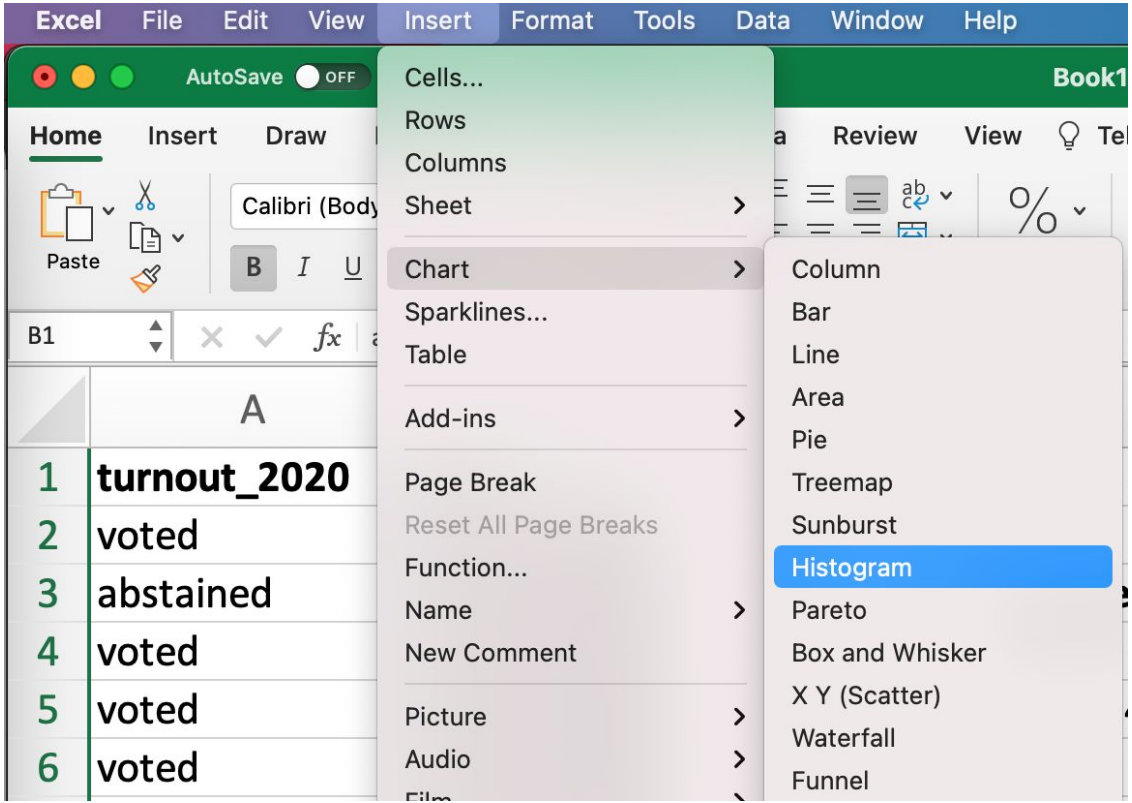
Visualizing data

i.e., making graphs!

Charts: one variable

Highlight the column we want to display

Go Insert → Chart → Histogram

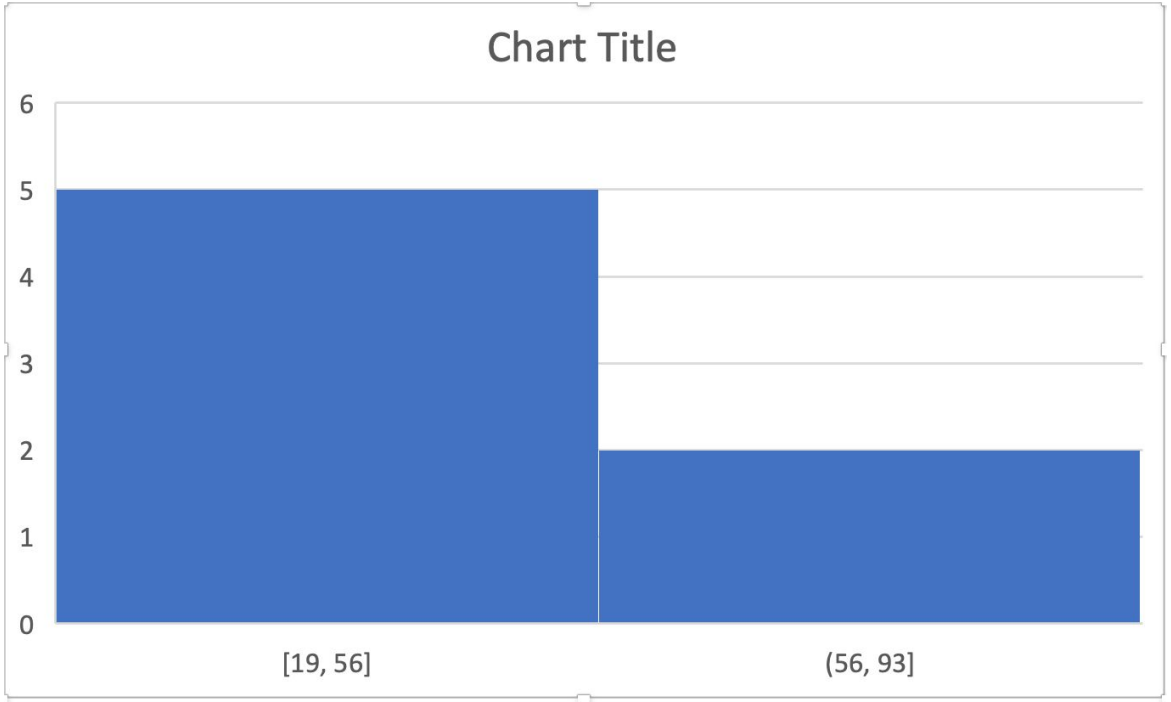


Charts: one variable

Excel will spit out some default chart.

It will not be pretty!

But that's ok.



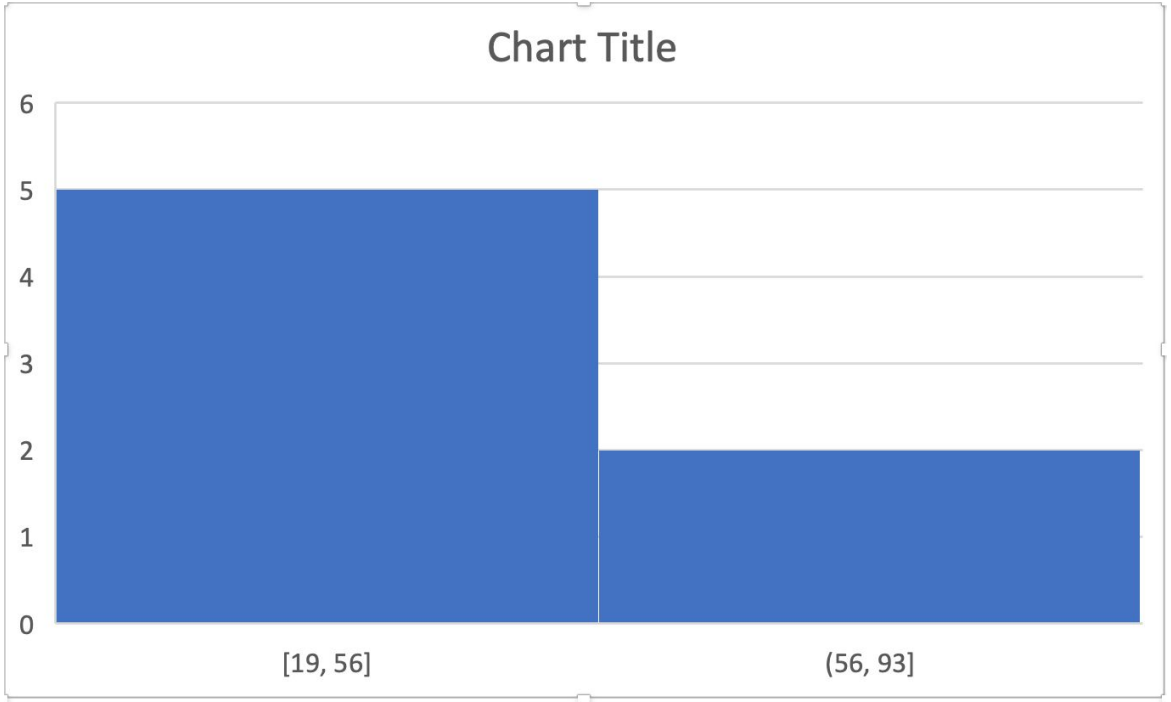
Charts: one variable

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

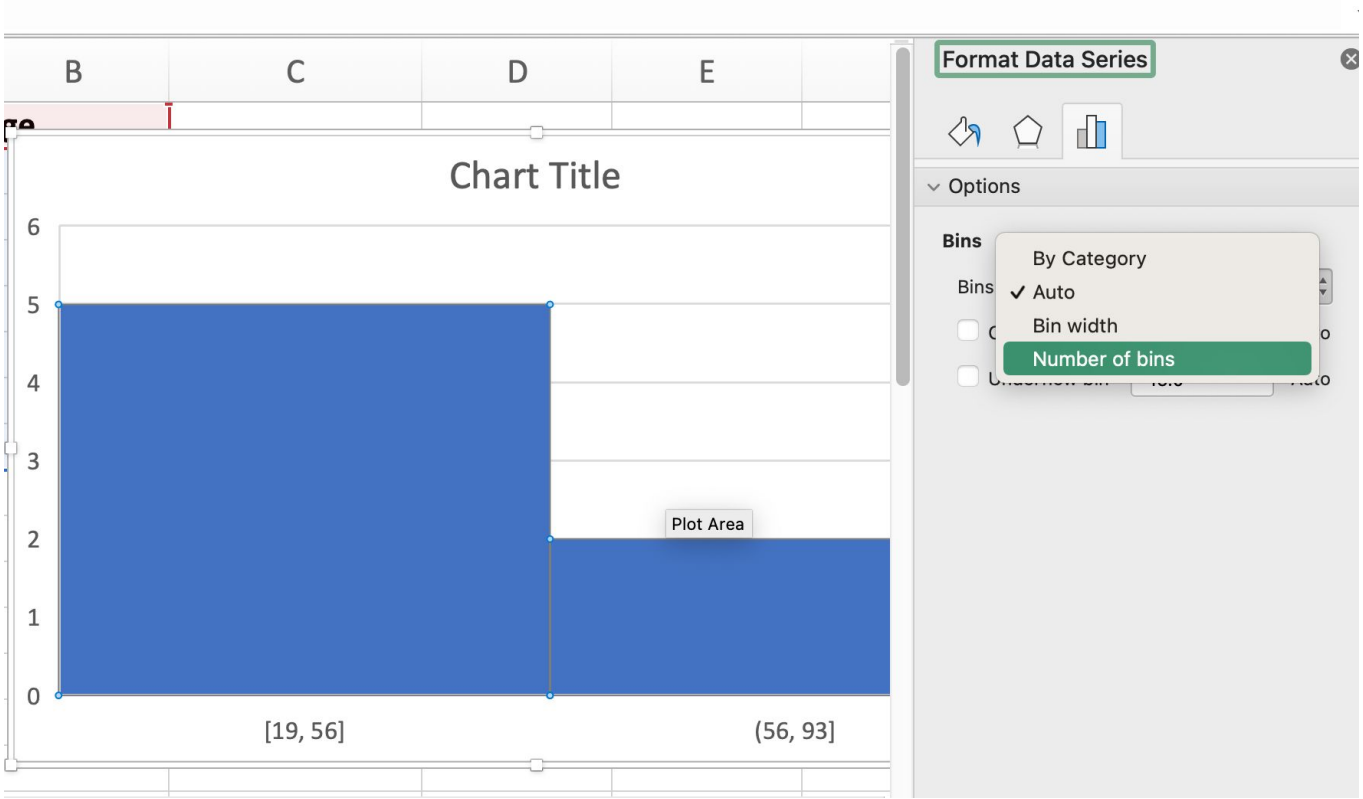
Excel will spit out some default chart.

It will not be pretty!

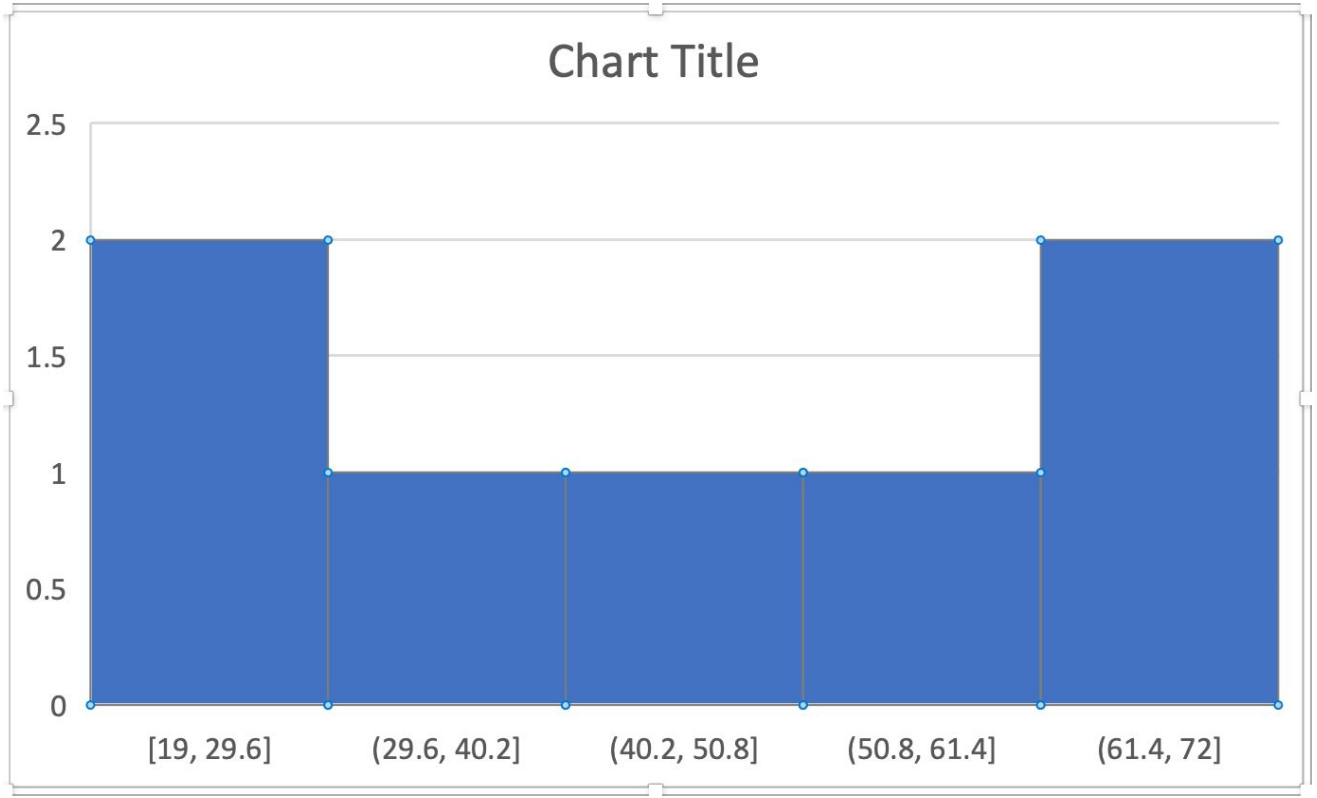
But that's ok.



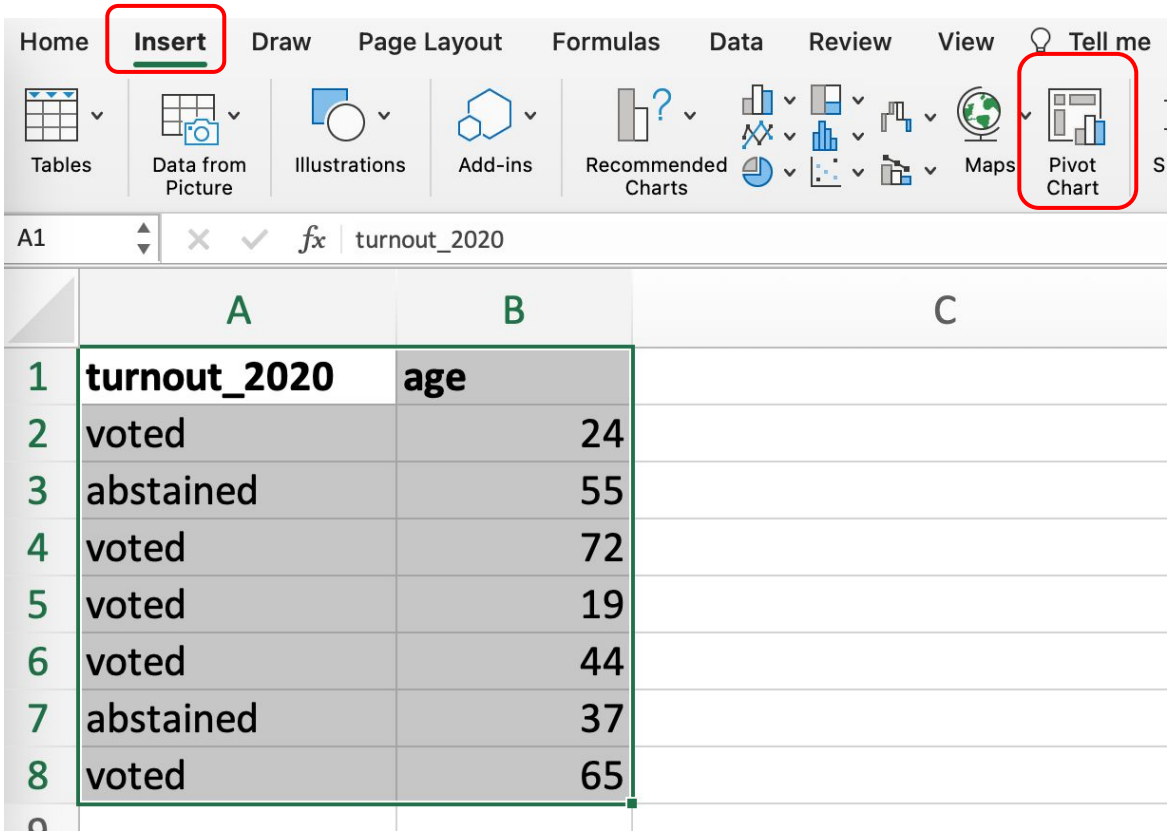
Charts: one variable



Charts: one variable



Charts: multiple variables



	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	

Home Insert Draw Page Layout Formulas Data Review View Pivot Chart Analyse Design Format Tell me Share Comments

Chart Name: Chart 1 | Options | Active Field: | Field Settings | Expand Field | Collapse Field | Insert Slicer | Insert Timeline | Filter Connections | Refresh | Change Data Source | Clear | Move Chart | Fields, Items & Sets | Field List

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

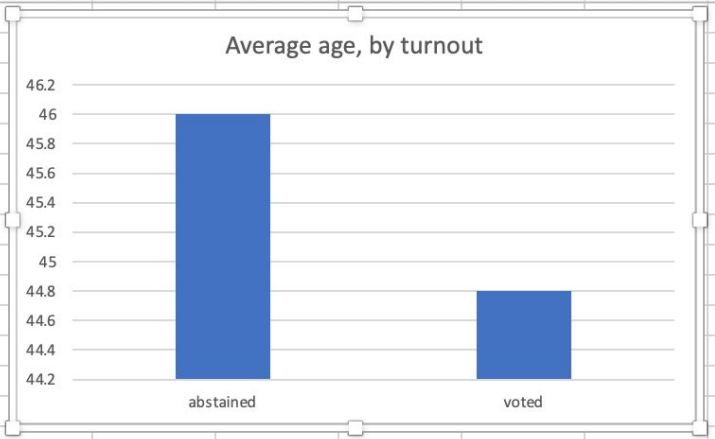
Chart Options | Text Options

Fill

Border

Charts: multiple variables

Row Labels	Average of age
abstained	46
voted	44.8
Grand Total	45.14285714



PivotChart Fields

FIELD NAME

Search fields

turnout_2020

age

Filters

Legend (Series)

Axis (Categories)

turnout_2020

Values

Average of age