BIG DATA Solution of the second sec



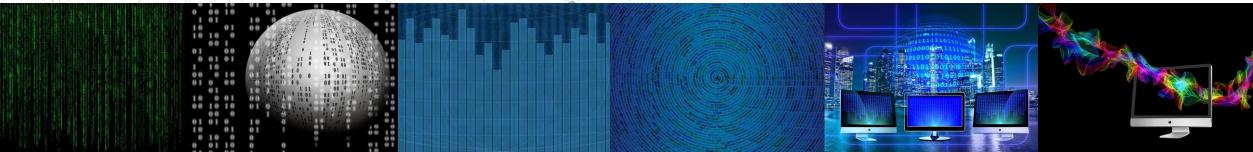
TEXAS A&M UNIVERSITY
Superfund Research Center

Welcome !

superfund.tamu.edu

This Session will Begin at 2:00 pm Eastern US Time "Manipulating Big(ish) Data in Excel, and Reading into R"

Fred Wright – North Carolina State University Candice Brinkmeyer-Langford – Texas A&M University Dillon Lloyd – North Carolina State University

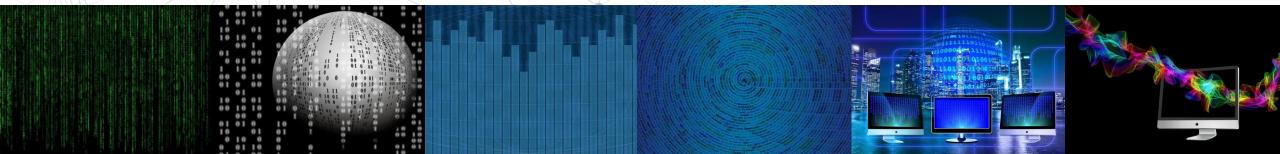


BIG DATA Superfund.tamu.edu IN ENVIRONMENTAL SCIENCE AND TOXICOLOGY



TEXAS A&M UNIVERSITY
Superfund Research Center

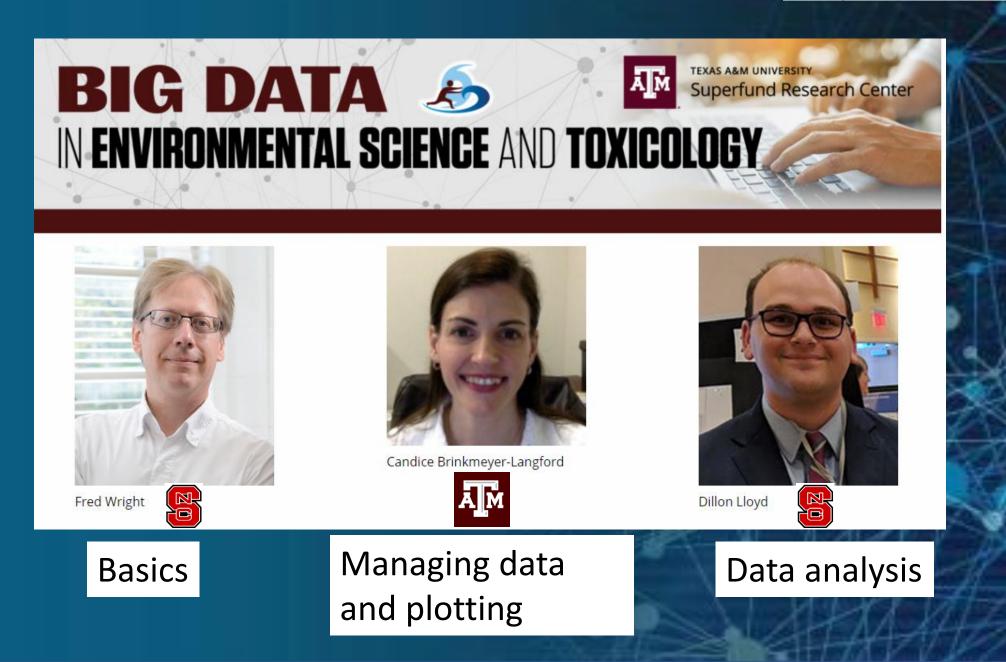
- All participants are muted to enable the speaker to present without interruption.
- Please rename yourself and designate Full Name and Affiliation.
- Last-minute installation issues? Use private chat to Dillon Lloyd.
- Use the reaction icon at the bottom of your screen to raise your hand. However, due to time constraints, most questions will be answered in chat window, with answers posted later as an FAQ.
 - This meeting will be recorded, and posted on the @tamusuperfund website https://superfund.tamu.edu/big-data-series-2021/ in the coming weeks.



BIG DATA SIPERIAL SCIENCE AND TOXICOLOGY



TEXAS A&M UNIVERSITY
Superfund Research Center







TEXAS A&M UNIVERSITY
Superfund Research Center

What is this session about?

- This session is NOT a comprehensive introduction to Excel
- We present a number of topics that have been useful to the speakers for Superfund-relevant data.
- In other words, what do we and collaborators do most often?
- We include links to a number of topics
- We assume only the most basic familiarity with Excel







TEXAS A&M UNIVERSITY
Superfund Research Center

What is this session about?

- Excel is not the main tool for data science
- It is a widely used tool for storing and working with data
- Best practices need to acknowledge reality that Excel is widely used





Basics

Fred Wright, NC State University





TEXAS A&M UNIVERSITY
Superfund Research Center

What is Big Data?

- Big Data invokes and evokes a lot of concepts
- Data we can hold in one spreadsheet is not *that* big
- However, data that we cannot perceive in one screenful is big enough that we need to let the software tell us things





Soundcloud



BIG DATA Solution In Environmental Science and Toxicology



Excel data limits

- Much better than the old days!
- Roughly speaking, big enough for gene expression datasets
- Not big enough for genome scans or sequence data

Worksheet and workbook specifications and limits

Feature	Maximum limit
Open workbooks	Limited by available memory and system resources
Total number of rows and columns on a worksheet	1,048,576 rows by 16,384 columns
Column width	255 characters
Row height	409 points
Page breaks	1,026 horizontal and vertical
Total number of characters that a cell can contain	32,767 characters
Characters in a header or footer	255



Data Window Help

Cells

G

Q.

Editing

Jul

0.9

0.0

0.0

H

0.0

1.0

0.5

+ 100%

View

Jun

0.6

0.3

0.6

rmatting

2+

PC vs. Mac

Tab				
		Ribbon	(may need to tu	urn on the ribbon ເ
Au 🖉 🐨 🗄 🏷 ୯ ୧ 👻 Book1 - Excel 📃 🔎 Search (A	lt+Q) Fred Wright	● ● - □	× É Excel File E	dit View Insert Format Tools D
File Home Insert Page Layout Formulas Data Revie	w View Help JMP	요 S 🗸 🖓 Comme	ints	က ံ Workbook1
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	✓ III Conditional Formatting ✓ III Insert ✓ III Format as Table ✓ III Cell Styles ✓ III Format ✓ III Format ✓ Cells ✓ Cells	Analyze	Paste Street Font	ge Layout Formulas Data Review
E7 👻 : 🗙 🗸 f_x			× A1 + × ✓	f _x Item
	H I J K	L M N O	A B	C D E F
A B C D E F G		L M N (0 1 item Jan 2 item A 1.	Feb Mar Apr May 0 0.5 0.2 0.8
2			3 Item B 0.	
2 3			4 Item C 0.	
4			5	
5			6	
6			7	
7			9 Row Labels	Sum of May
8			10 Item A	0.5
9			11 Item B	0.2
10			12 Item C	0.5
			13 Grand Total	1.2
12 13			14	
Sheet1 (+)	: 4		Sheet1 -	F
Ready		▣ ▥ - ── + 1	00% Ready	010

using View)

CTRL* vs. CMD* * means whatever you are typing!





Got an Excel question?

- The search bar is pretty good!
- If you're still not sure, try googling the question, which often leads to a keyword
- Examples
- format values in Euro currency
- how do I turn my rows into columns.... what is that called?

AutoSav	e 🞯 🗒		~ - E	Book1 - Excel	P	Search (Alt+	Q)			Fred Wright	• •	囨			×
File	Home Ins	ert Pag	e Layout	Formulas	Data	Review	View	Help	JMP			🖻 Sh	are 🖓	Comment	ts
Paste	3 표~	<u>◇</u> ~ <u>A</u> ~		E I I I I I I I I I I I I I I I I I I I	€.0 .00		📆 Form	itional Form at as Table ~ ityles ~ Styles	-	∰ Insert ↔ ₩ Delete ↔ ∰ Format ↔ Cells	 ↓ ~ ↓ ~ ↓ ~ 	2₽~ ₽~	Analyze Data Analysis		~
E7	• :	X 🗸	f _x												~
A	В	с	D	E	F	G	н	I.	J	к	L	м	N	0	ŀ
1															٦
2															
3															
4															-1
5															-
6															-
7				 											-
8 9															-
10															
11															
12															
13															1
	Sheet	1 +							4						Þ
Ready											E P		-	- + 100	196





TEXAS A&M UNIVERSITY
Superfund Research Center

Importing data from a text file

https://support.microsoft.com/en-us/office/import-or-export-text-txt-or-csv-files-5250ac4c-663c-47ce-937b-339e391393ba

- Use file>open>browse>(correct format)>open
- But... this is not an Excel file

 The text import wizard saves the day. Pay attention to delimiters (character that separates variables)

Need to skip the first few lines? There's a box for that!





BOOFED <u>https://support.microsoft.com/en-us/office/undo-redo-or-repeat-an-action-84bdb9bc-4e23-4f06-ba78-</u> <u>f7b893eb2d28#:~:text=To%20undo%20an%20action%20press,want%20to%20undo%20multiple%20steps</u>.

- Never fear, CTRL-z is here
- or undo curvy arrow
- Wait, I didn't goof!
 Other curvy arrow (or CTRL-y)







Jump around

https://support.microsoft.com/en-us/office/move-or-scroll-through-a-worksheet-06fc34b8-64bb-4d78-9b62-34656d700f82

- You can scroll around
- Or you can jump using CTRL* (e.g. CTRL +)
- Try CTRL END or CTRL HOME. What do they do?
- These are called "shortcuts"
- Select stuff using SHIFT
- Combine stuff, e.g. CTRL SHIFT *





Freeze panes!

- Useful for scrolling around while keeping some rows/columns fixed
- View>freeze panes >freeze panes
- Top row, first column, and what else?





DMITRY LOVETSKY/AP

Pixar





Appearance and formatting of cells (basic) <u>https://support.microsoft.com/en-us/office/available-number-formats-in-excel-0afe8f52-97db-41f1-b972-</u> <u>4b46e9f1e8d2</u>

- Cell appearance similar to other MS products
- Boldface, italics, text, size, etc.
- Formatting may appear on the Home action bar
- Or Design>formatting
- Check Alignment and Number
- The formatting doesn't really change the underlying values (look at formula bar)





Sorting

https://support.microsoft.com/en-us/office/sort-data-in-a-range-or-table-62d0b95d-2a90-4610-a6ae-2e545c4a4654

• select value in a column, hope for the best!

OR

- do it deliberately, Home>Editing>Sort and Filter>Custom Sort
- But I have two header rows!
- Sorting is dangerous! Always do a sanity check. Did I sort the region intended, and not just one column?



Adding, deleting, hiding

https://support.microsoft.com/en-us/office/insert-or-delete-rows-and-columns-6f40e6e4-85af-45e0-b39d-65dd504a3246#:~:text=To%20insert%20a%20single%20column,and%20then%20select%20Insert%20Column <u>s</u>. https://support.microsoft.com/en-us/office/hide-or-show-rows-or-columns-659c2cad-802e-44ee-a614dde8443579f8

- Logical for the most part
- The right mouse button is your friend





Highlighting info by conditional formatting https://support.microsoft.com/en-us/office/use-conditional-formatting-to-highlight-information-fed60dfa-1d3f-4e13-9ecb-f1951ff89d7f

- Highlight interesting cells or ranges of cells
- Select cells
- Home>Conditional formatting
- Take a look and explore





Simple functions and formulas

- Make new columns that are log10 of the previous columns
- Note how the cells are relative in the formula
- You can change them to absolute (sometimes called "anchor") values by using \$ or F4 (on PC)





Transforming data (Power Query)

https://www.youtube.com/watch?v=yE8sLnywPAg

- Data tab
- From table/range (make sure correct range is selected)





Drop-down lists

https://support.microsoft.com/en-us/office/create-a-drop-down-list-7693307a-59ef-400a-b769-c5402dce407b

- 1. Select the cells that you want to contain the lists.
- 2. On the ribbon, DATA -> Data Tools->Data Validation.
- 3. Allow to List.
- 4. Click in Source, type the text or numbers (separated by commas, for a comma-delimited list) that you want in your drop-down list, and click OK.



TEXAS A&M UNIVERSITY
Superfund Research Center

Good naming and formatting practices for sharing data via spreadsheet



BIG DATA

IMENTAL SCIENCE AND TOXICOLOGY

```
concept
```



• Data science is 1% inspiration, 99% exasperation (mainly caused by extra spaces, poor heading labels, and unnecessary detail in missing data)





TEXAS A&M UNIVERSITY
Superfund Research Center

Good naming and formatting practices for sharing data via spreadsheet

- DO keep all the data as one row/column rectangle (if possible)
- DO keep things as simple as possible
- DO keep column headings descriptive and CONSISTENT, including utterly consistent capitalization
- DO use a single code (e.g. NA) for missing data if possible
- DO NOT merge cells
- DO NOT add actual commas to numerical values
- DO NOT add hidden spaces in data



Ainsley Seago. doi:10.1371/journal.pbio.1001779.g001

• DO NOT use spaces in column headings. Underscore is preferred.





Transposing data

- Changes rows to columns
- Not exactly "turning the rectangle of data on its side"
- Copy and Paste Special, select Transpose

See <u>https://support.microsoft.com/en-us/office/transpose-function-ed039415-ed8a-4a81-93e9-4b6dfac76027</u> to learn how to do this using the TRANSPOSE function, while preserving formulas





Create tables

https://support.microsoft.com/en-us/office/create-and-format-tables-e81aa349-b006-4f8a-9806-5af9df0ac664

- Don't we already have a table?'
- A table is a specific object for a rectangular data range that Excel can use for analysis
- Hallmark of a table is the coloring/banding, and the drop down button for column headings
- Home > Format as Table





Pivot tables

https://support.microsoft.com/en-us/office/create-a-pivottable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9f99134456576

- Useful for summarizing totals, etc. across variables
- Similar to cross-tabulation
- Somewhat restricted to certain functions, etc.
- Insert>Pivot Table
- The next speaker will provide different ways of summarizing what you have



TEXAS A&M UNIVERSITY
Superfund Research Center

Next up: Manipulating data and counting stuff!

BIG DATA 🍝

IN ENVIRONMENTAL SCIENCE AND TOXICOLOGY

superfund.tamu.e



http://publicdomainaudiovideo.blogspot.com/

Manipulating Big(ish) Data in Excel

Candice Brinkmeyer-Langford

Texas A&M University





Excel version and support

- Excel version being used: Microsoft Excel for Mac, version 16.51
- <u>https://support.microsoft.com/e</u> <u>n-us/office/excel-video-training-</u> <u>9bc05390-e94c-46af-a5b3-</u> <u>d7c22f6990bb</u>

G	ŵ			pport.microsof	t.com/en-us	s/office/ex	xcel-video-	trainin	g-9bc053	390-e 94 ••	• 🛛 🕁	Q Search
	Microsoft	I	Support	Microsoft 365	Office W	/indows	Surface	Xbox	Deals	Buy Micro	osoft 365	
0	ffice support	Ρ	roducts \vee	Devices \lor	What's ne	w Ir	nstall Office	A	account & b	billing \vee	Templates	More support $ \smallsetminus $

Excel video training

Excel for Microsoft 365, Excel 2019, Excel 2016, Excel 2013, Excel 2010, More...

$\overline{\bigcirc}$	K.	Ш	
Quick start	Intro to Excel	Rows & columns	Cells
Ą			<u></u>
Formatting	Formulas & functions	Tables	Charts
	•<		
PivotTables	Share & co-author	Linked data types	Get to know Power



Answering questions with Excel

eous tips and tricks





8	•	P	C	D	F	F	C	Ц		1	K	1	5.4	N	0	D	0	P	c	т	
	A	В	L	D	C	Г	G	П	1	J	ĸ	L	Μ	N	0	P	Q	ĸ	5	1	0
1	Sample_ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Sex	F	F	М	F	М	М	М	F	F	М	F	Μ	М	F	М	F	М	F	М	F
3	Status	N	Y	Y	N	N	N	Y	Y	Ν	Y	Y	Ν	Y	N	N	Y	Y	Y	N	N
4	Cat1	3	7	9	3	1	4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.36314088
7	Cat4	76	75	65	53	89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	71
8	Cat5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	8	7	4	2	9	9
9	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.93651799
10	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847
11	Cat8	59	79	81	65	54	91	50	92	96	51	80	63	78	70	91	60	80	59	81	92
12	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.47184386	0.97195415	0.10610593	0.55886697	0.34184183	0.38806705	0.24976963
13	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	1061	1850	1160	1166	1049	1419	1275	1207
14	Cat11	535	586	543	470	479	502	582	537	412	514	462	487	543	405	504	478	466	419	467	591
15	Cat12	13	19	15	18	11	19	18	10	16	11	15	19	20	18	13	17	16	14	14	14
16																					





IF

- I want to know if individual 8 is female.
 - I could just look at this little table, obviously, but pretend it's a much bigger dataset.
- The IF function can have 2 results (True or False)
- =IF(test,value_if_true,value_if_false)

sι	X 🍃 M	$\checkmark f_x$ =if(
2	A	В	С	D	E
1	Is #8 female?	=if(
2		IF(logical	test, [value	_if_true], [val	ue_if_false])
3					



 TEXAS A&M UNIVERSITY

 Superfund Research Center

IF

- We are looking on a different sheet for our answer, hence the sample_values! before the specific cell I2
- Because we want to know if the contents of cell I2 are a specific value, we use quotation marks: "F"
- Don't forget to add the comma!

sι	JM 🇯	× 🗸	fx = if(sa)	mple_value	s!12="F",				
2	A	В	c IF(ogical_test, [value_if_tru	e], [value_if_	false])	н	I
1	Sample_ID	1	2	3	4	5	6	7	8
2	Sex	F	F	м	F	М	М	М	F
3	Status	N	Y	Y	N	N	N	Y	Y
4	Cat1	3	7	9	3	1	4	7	10
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944
7	Cat4	76	75	65	53	89	66	88	78
8	Cat5	3	10	3	1	7	5	5	3
9	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318
0	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672
1	Cat8	59	79	81	65	54	91	50	92
2	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019
3	Cat10	1646	1436	1486	1218	1151	1278	1329	1942
.4	Cat11	535	586	543	470	479	502	582	537
5	Cat12	13	19	15	18	11	19	18	10
6									
.7									
8									
9									
0									
1									
2									
3									
4									
5									
6									
7									
8									
9									
0									
1									
2									
3									
4									
5									
6									
7									
8									





IF

Sl	JM 🛔	× 🗸	$f_{\mathbf{x}}$ =if(sa	mple_value	es!12="F","Y	ES","NO"				
/	A	В	C IF(Id	<u>ogical_test, [</u>	value_if_true], [<u>value_if_</u> f	f <u>alse]</u>)	Н	I	
1	Sample_ID	1	2	3	4	5	6	7	8	
2	Sex	F	F	М	F	М	м	М	F	F
3	Status	N	Y	Y	N	N	N	Y	Y	Ν
4	Cat1	3	7	9	3	1	4	7	10	
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.3

- If my test is true that is, if column I2 is "F" (in other words, individual #8 is female...
 - "YES",
 - If it's not true, I want Excel to say "NO"
- Don't forget to close the parentheses so Excel knows you're done.





IF

• When I add the close parenthesis (")") and hit Enter, I can go back to my original sheet (Sheet1, in this case) and see my answer.

B1	\$ × ·	fx =	IF(sample_va	lues!12="F"	,"YES","NO")
/	A	В	С	D	E
1	Is #8 female?	YES	7		
7					





How many are in certain categories...?

COUNTIF function counts how many cells in a given range have a certain value

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

		Insert	Draw	Page L	ayour	Formu	iiuo	Dat	u	Revie	vv
1	<mark>ð</mark> - 👗	Calib	ori (Body) 🔹 12	•	A ^ A ▼	=	=	=	200	•
F	Paste 🚕	В	ΙU	•	-	- A -	E	Ξ	Ξ	• =	¢≡
B1	L 🌲	$\times \checkmark$	$f_x =$	COUNTIF(sample	_values!B	2:U2,"F	-")			
4	А	В	С	1	C	E	F		0	i	н
1	Females	1									
2	Males	1	D								
3	Exposed										
4	Unexposed		_								
5											
6											
7							· · · · · · · · · · · · · · · · · · ·				
8			_								
9			_								
10											
11			-								
12											
13			-								
14 15											
15											
10	-		-								
18			-								
19			-								
20											
21			-								
22			-								
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35	-										
36			_								
37											
38					14						
4	> s	ample_valu		Sheet2	+						35



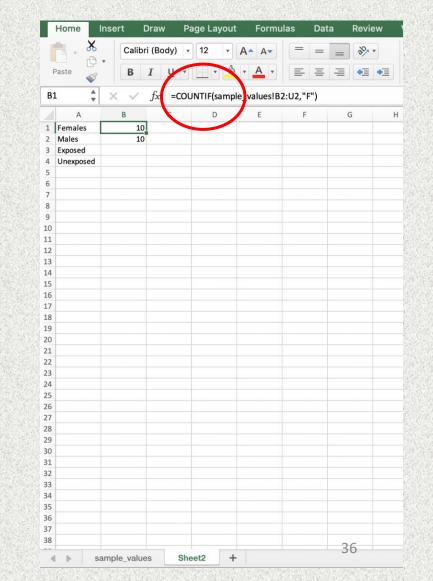


How many are in certain categories...?

COUNTIF function counts how many cells in a given range have a certain value

=COUNTIF (the equals sign tells Excel you want it to carry out some function)

Then, open parenthesis to tell Excel exactly what you want it to search ("cells in a given range") and what you want it to search for ("a certain value")





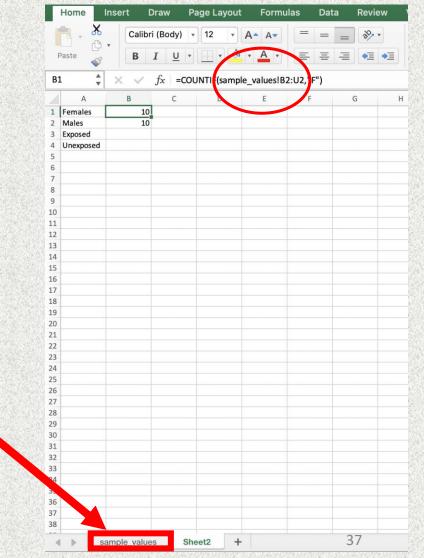


COUNTIF function counts how many *cells in a given range* have a certain value

Where is the "range" of data that contains what we want to count?

Here, the "range" is on a different sheet (sample_values). This is OK.

The exclamation point (!) tells Excel to look on a different sheet: sample_values!





We want to count the number of females.

Range here is the second row of the sheet sample_values:

B2:U2 (cells B2 through U2) or 2:2 (select entire row) The colon (:) means "through"

	Home Kome Compared and the second sec	•	Dra alibri (B I		age Lay	vout A		ulas	Data = =	Revi = ॐ ≣ •≡	•
B1	L 🌲	×	√ J	x =co	UNTIE(sa	ample_v	/alues!B	2:U2,	F")		
/	А	В		С	D		Е	F		G	F
1	Females		10								
2	Males		10								
3	Exposed										
4	Unexposed										
5											
6											
7								a			
J	К	L	М	N	0	Р	Q	R	S	Т	U
	9 10	11	1	2 13	14	15	16	1	7 18	3 19	2

	A	В	C	D	E	F	G	н	1	J	K	L	IVI	N	0	Р	Q	ĸ	5	1	U
1	Sample_ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Sex	F	F	М	F	М	М	М	F	F	М	F	М	М	F	М	F	М	F	М	F
3	Status	N	Y	Y	N	N	N	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	N
4	Cat1	3	7	9	3	1	4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.36314088
7	Cat4	76	75	65	53	89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	71
8	Cat5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	8	7	4	2	9	9
9	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.93651799
10	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847
11	Cat8	59	79	81	65	54	91	50	92	96	51	80	63	78	70	91	60	80	59	81	92
12	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.47184386	0.97195415	0.10610593	0.55886697	0.34184183	0.38806705	0.24976963
13	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	1061	1850	1160	1166	1049	1419	1275	1207
14	Cat11	535	586	543	470	479	502	582	537	412	514	462	487	543	405	504	478	466	419	467	38 591
15	Cat12	13	19	15	18	11	19	18	10	16	11	15	19	20	18	13	17	16	14	14	14
16																					THE REPORT OF THE PARTY OF

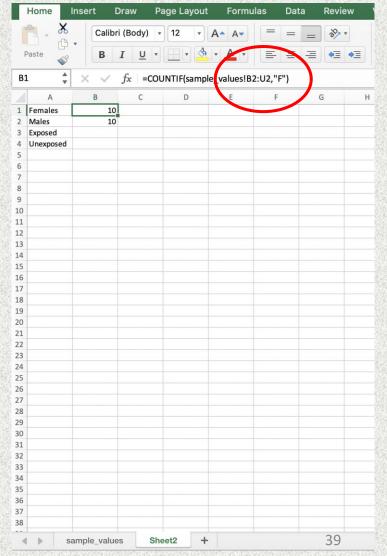


COUNTIF function counts how many cells in a given range *have a certain value*

The value we want to count is number of females.

After we specify the range, add a comma (,) and then the value we want to count.

"F"





COUNTIF function counts how many cells in a given range *have a certain value*

Another option: after we specify the range, add a comma (,) and then a cell containing the value we want to count.

sample_values!C2

("C2" will tell Excel to count cells containing <u>C2</u> rather than the contents of cell C2)

	Home I	nsert D	Draw P	age Layout	Formulas	Data Review
F	Paste	• Calibr	i (Body) I <u>U</u> -	- 12 - A		
B 1	L 🌲	\times \checkmark	f_x =CO	UNTIF(sample_	values!B2:U2,s	ample_values!C2)
1	А	В	С	D	E F	G
1	Females	10				
2	Males	10				
3	Exposed	10				
4	Unexposed	10				
5						
6						
			19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38	Image: Amage:		40
			▲ ▶ S	ample_values Sheet2	+	TO TO





QUESTION 1...

- Which formula should I use if I want to know how many individuals were exposed (status is Y)?
- A: =COUNTIF(B3:U3,"Y")
- B: =COUNTIF(sample_valuesB3:U3,Y)
- C: COUNTIF(sample_values!B3:U3,"Y")
- D: =COUNTIF(sample_values!B3:U3,C3)
- E: =COUNTIF(sample_values!B3:U3,"Y")

	А	В	С	D	E	F	G	Н	I	J	К	L	М	Ν	0	Р	Q	R	S	Т	U
1	Sample_ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Sex	F	F	М	F	М	Μ	М	F	F	М	F	М	М	F	М	F	М	F	М	F
3	Status	N	Y	Y	Ν	N	N	Y	Y	N	Y	Y	N	Y	Ν	N	Y	Y	Y	Ν	N
4	Cat1	3	7	9	3	1	4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.36314088
7	Cat4	76	75	65	53	89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	71
8	Cat5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	8	7	4	2	9	9
9	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.93651799
10	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847
11	Cat8	59	79	81	65	54	91	50	92	96	51	80	63	78	70	91	60	80	59	81	92
12	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.47184386	0.97195415	0.10610593	0.55886697	0.34184183	0.38806705	0.24976963
13	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	1061	1850	1160	1166	1049	1419	1275	1207
14	Cat11	535	586	543	470	479	502	582	537	412	514	462	487	543	405	504	478	466	419	467	591
15	Cat12	13	19	15	18	11	19	18	10	16	11	15	19	20	18	13	17	16	14	41 14	14
16	CARL PROPERTY OF THE OWN																				



A: =COUNTIF(B3:U3,"Y") \leftarrow did not specify the correct sheet where the data was located

Calibri (Body) \cdot 12 \cdot A \cdot $=$ $=$ COUNTIF(sample_values!B3:U3,"Y") B I U \cdot \cdot A \cdot $=$ $=$ $=$ COUNTIF(sample_values!B3:U3,"Y") B3 \cdot f_x =COUNTIF(sample_values!B3:U3,"Y") A B C D E F 1 Females 10 B3 \cdot f_x =COUNTIF(sample_values!B3:U3,"Y") 2 Males 10 B3 \cdot f_x =COUNTIF(sample_values!B3:U3,sample_val		Home I	nsert D	raw	Page Layout	F	ormulas	Data		NTIF(sam	The second s		
Paste B I U I <th>-</th> <th></th> <th></th> <th>(Body)</th> <th>• 12 •</th> <th>A^</th> <th>A▼ =</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>and the second second</th>	-			(Body)	• 12 •	A^	A ▼ =						and the second second
ABCDEF1Females10B3 f_x =COUNTIF(sample_values!B3:U3,	F	Dacto		I <u>U</u>	•	• A		= =	E: =COU	NTIF(sam	ple_value	es!B3:U3	,"Y")
1 Females 10 2 Males 10 3 Exposed 10 4 Unexposed 10 1 Females 10 2 Males 10 3 Exposed 10 4 Unexposed 10 5 Males 10 6 1 7 10 1 Females 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10	B3	3	XV	$f_x \mid = \mathbf{C}$	OUNTIF(sample	_valu	ues!B3:U3,"	Y")					
2 Males 10 B3 Jx =COUNTIF(sample_values!B3:03,sample_values!B3:0	/	A	В	С	D	E	F						
2 Males 10 3 Exposed 10 4 Unexposed 10 1 Females 10 2 Males 10	1	Females	10			B	z	× ./	fr -coi	INTIE(compl		Rill2 comp	
4 Unexposed 10 1 Females 10 10 10 2 Males 10 10 10 10 10	2	Males	10				•		<i>Jx</i> =cot	Sivingsamp	e_values:b.	5.05,samp	e_values:c.
2 Males 10 10 10 10	3	Exposed	10				A	В	С	D	Е	F	G
2 Males 10	4	Unexposed	10			1	Females	10					
3 Exposed 10						2	Males						
						3	Exposed						

Unexposed

10



A: =COUNTIF(B3:U3,"Y")

	Home	Insert	Draw	Page Layout	F	ormulas	Data	exclama	tion
Ŧ	Paste	Cal	ibri (Body) I <u>U</u>					C: COUN D: =COU E: =COU	INTIF
B		, × ×		COUNTIF(sample	_valu			LCOU	
	A	В	С	D	E	F			
1	Females		10		B	,	XV	f	
2	Males		10		D.	•		fx = CO	UNTIF
3	Exposed		10			A	В	С	
5	10.00		10		1	Females	10	1	
4	Unexposed		10						
	Unexposed		10		2	Males	10		

B: =COUNTIF(sample_valuesB3:U3,Y) ← there is no exclamation mark after the sheet name

C: COUNTIF(sample_values!B3:U3,"Y") D: =COUNTIF(sample_values!B3:U3,C3) E: =COUNTIF(sample_values!B3:U3,"Y")

		12.25						
B3		X V	fx =COL	JNTIF(sampl	e_values!B	3:U3,sample	e_values!C3)
	А	В	С	D	E	F	G	
1	Females	10						
2	Males	10						
3	Exposed	10						
4	Unexposed	10						
5								
6								
7								



6 7 A: =COUNTIF(B3:U3,"Y") B: =COUNTIF(sample_valuesB3:U3,Y)

	Home In	isert D	Draw P	Page Layout	F	ormulas	Data	C: COUNTIF(sample_values!B3:U3,"Y") ← there is
			(Body)		A^			equals sign before COUNTIF
F	Paste 💞	В	I <u>U</u>	•	•		= =	D: =COUNTIF(sample_values!B3:U3,C3) E: =COUNTIF(sample_values!B3:U3,"Y")
B3	3 ▲	X 🗸	f_x =CO	UNTIF(sample	_val	ues!B3:U3,"\	Y")	
	А	В	С	D	E	E F		
1	Females	10			В	3 🔺	X V	f_x =COUNTIF(sample_values!B3:U3,sample_values!C3)
2	Males	10				•		
3	Exposed	10				А	В	C D E F G
4	Unexposed	10			1	Females	10	10
					2	Males	10	10
					3	Exposed	10	10
					4	Unexposed	10	10
					1 5			



How to count how many were exposed

	Home	nsert	Draw	Page Lay	out f	orm	ulas	Dat	a
-		Calib	ri (Body)	• 12	• A-	A▼	=	=	
P	Paste	В	<i>Ι</i> <u>υ</u>	•	<u></u>	A •	Ξ	Ξ	
B3	A V	× ✓	$f_x =$	COUNTIF(sa	ample_val	ues!E	33:U3,"	Y")	
\mathbb{Z}	А	В	С	D		E	F		
1	Females	10)		В	3		X	
2	Males	10)				•		
3	Exposed	10					A		В
	Unexposed	10				-			

A: =COUNTIF(B3:U3,"Y") B: =COUNTIF(sample_valuesB3:U3,Y) OUNTIF(sample_values!B3:U3,"Y")

COUNTIF(sample values!B3:U3,C3) \rightarrow The value for C3 will be searched for based on whatever sheet re on, rather than the sheet with the data

E: =COUNTIF	comnla	Values	182.11	2 "V")
LCOUNTIL	Jampie	values	:03.0	J, I J

	· · ·	100 M						
B	3 *	× ✓	f_x =COU	INTIF(sampl	e_values!B	3:U3,sample	e_values!C3)
	А	В	С	D	E	F	G	
1	Females	10						
2	Males	10						
3	Exposed	10						
4	Unexposed	10						
5								
6								
7				104143457781027778600777704477778777				-14345-2216-5-55



	Home	nsert	Draw	Page Layout	F	ormulas	Data
F	aste	• Calib	ori (Body) I <u>U</u>	▼ 12 ▼ ▼ ▼ ▼	A^		= =
B3	A A	× ~	fx =	COUNTIF(sample	e_valu E		
1	Females	- 10			B	3	
2	Males	10	0			•	
3	Exposed	10	D			А	В
4	Unexposed	10	D		1	Females	1
			den ante	tada ne stad	2	Males	1

A: =COUNTIF(B3:U3,"Y")

- B: =COUNTIF(sample_valuesB3:U3,Y)
- C: COUNTIF(sample_values!B3:U3,"Y")
- D: =COUNTIF(sample_values!B3:U3,C3)
- E: =COUNTIF(sample_values!B3:U3,"Y") ← correct! This will count how many individuals have "Y" in the Status row.

Could also search for the value in cell C3 of the sample_values sheet, which is also Y.

L.		1.59						
BB	3 *	× ✓	fx =COL	JNTIF(sampl	e_values!B	3:U3,sample	e_values!C3)	
	А	В	С	D	E	F	G	
1	Females	10						
2	Males	10						
3	Exposed	10						
4	Unexposed	10						
5								
6								
7								



What if I want to know how many females were exposed?

• You can "nest" multiple COUNTIF functions by using COUNTIFS

SL	JM 🛔	×	f_x =COU	JNTIFS(
	А	В	С	D	E
1	Females	10			
2	Males	10			
3	Exposed	10			
1	Unexposed	10			
5	Female+Exp	=COUNTIFS(
6		COUNTIFS	<u>(criteria_ra</u>	nge1, criteri	a1,)
7					

You still need to follow the same basic procedures as before.



Paste

18 19



How many females were exposed?

<u>U</u> ▼ <u>▲</u> ▼ <u>▲</u> ▼ <u>■</u> ≡ ≡ **→** Merge & Center ▼

You can "nest" multiple COUNTIF functions by using COUNTIFS

SUM \Rightarrow X V f_x =COUNTIFS(sample_values!B2:U2,"F",

	А	В	С	<u>cou</u>	NTIFS(crite	ria_range1, c	riteria1, [crit	teria_range2	, criteria2], .)	J	К	L
1	Sample_ID	1		2	3	4	5	6	7	8	9	10	
2	Sex	F	F	ſ	Μ	F	М	М	М	F	F	М	F
3	Status	Ν	Y	١	1	N	N	N	Y	Y	Ν	Y	Y
4	Cat1	3		7	9	3	1	4	7	10	6	9	
5	Cat2	=COUNTIFS(sample_v	alues	:!B2:U2,"F",		0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.1297
6	Cat3	0.01765164	0.740752	219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.6661
7	Cat4	76		75	65	53	89	66	88	78	98	67	
8	Cat5	3		10	3	1	7	5	5	3	3	3	
9	Cat6	0.27817695	0.07110	137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.5291
10	Cat7	0.47111146	0.899778	883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.1735
11	Cat8	59		79	81	65	54	91	50	92	96	51	
12	Cat9	0.15067873	0.696740	031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.848
13	Cat10	1646	14	436	1486	1218	1151	1278	1329	1942	1085	1662	
14	Cat11	535	5	586	543	470	479	502	582	537	412	514	
15	Cat12	13		19	15	18	11	19	18	10	16	11	
16													
17													

Criteria_range1 = where to look (first criteria)

Criteria1

\$ • %) 🔩

= what to look for (first criteria





How many females were exposed?

2.93																					
รเ	UM 🛔	×	f_x =COU	JNTIFS <u>(san</u>	nple values!	B2:U2 <mark>,"F",</mark>	ample value	es!B3:U3 "Y	("")												
	А	В	С	D	E	F	G	Н		J	К	L	М	N	0	Р	Q	R	S	т	U
1	Sample_ID	1	2		3 4	1	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Sex	F	F	М	F	м	м	м	F	F	М	F	М	Μ	F	М	F	М	F	М	F
3	Status	N	Y	Y	N	N	Ň	Υ	Y	N	Y	Υ	N	Y	N	N	Y	Υ	Y	N	N
4	Cat1	3	7		} 3	; 1	4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
5	Cat2	=COUNTIFS(sample_valu	es!B2:U2/F	',sample_value	es!B3:U3,"Y")		0.31730995	0.78559803	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628
6	Cat3	0.01765164	0.74075219	0.42/329	5 0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0 9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.36314088
7	Cat4	76	75	6	5 53	8 89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	71
8	Cat5	3	10		3 1	. 7	5	5	3	3	3	8	6	7	5	8	7	4	2	9	9
9	Cat6	0.27817695	0.07110137	0 3774793	9 0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	2 93419475	0.52918677	0.2503355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.93651799
10	Cat7	0.47111146	0.89977883	0.4250448	2 0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847
11	Cat8	59		8	1 65			50	92	96		80	63		70		60	80	59	81	92
12	Cat9	0.15067873		0.260765		0.54015010		0.31429126		0.87520766	0.0101267		0.81448124				0.10610593	0.55886697	0.34184183		0.24976963
13	Cat10	1646		148					1942	1085		1748	1922		1850	1160	1166				
14	Cat11	535	586	54						412	514	462	487	543	405	504	478	466	419		591
15	Cat12	13	19	1	5 18	; 11	19	18	10	16	11	15	19	20	18	13	17	16	14	14	14
16	NORMAN CANE	MILLION CONTRACTOR	EAST DROLLS	9544070620070	about the second	CARE BERE BERE BERE BERE	ALCONTRACT THE ALCONT	NUTRITICAL DE LA COMPANIA DE LA COMP	10/10/00/00/07/244	CARGON TENNING	100534-00702-0040	Marken and Market Al	ere servera	STAR CHERRY	NAMES OF A COMP	100 100 100 100		0.11753402164	8 5.9 2 M (C) (C) (C)	AT ALL AND THE ADDRESS	NUTSHEEREAN
						Bar Charles															
													1.00								
500	10	citoria .			oro to	look	Critoria	-1	hat to	look f		torio m			ro to l		ritorio	2	aat te	look fo	
		nena	range.	$\mathbf{L} = \mathbf{W}\mathbf{I}$	nere to	юок,	Lineria	$\mathbf{M} = \mathbf{T}$	nat to	IOOK T	or, cri	teria la	ingez	= whe	reioi	00K, C	nteria	z = wr	Iditio	IOOK TO	Л ј
0.55	and the second second			State Land						and the second second	and the second second		CLASSIC STR			1952	and the second		Sector Sector		



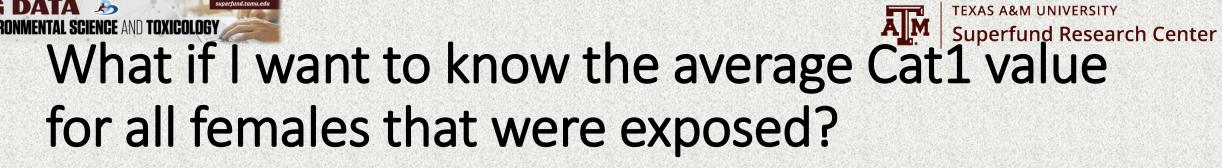


How many females were exposed?

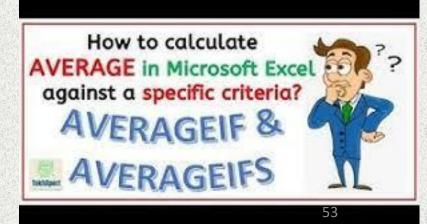
• Our answer, found using COUNTIFS!

SUM	•	×	\checkmark	fx =COUNTIFS(sample_values!B2:U2,"F",sample_values!B3:U3,"Y")	
-----	---	---	--------------	---	--

	А	В	С	D	E	F	G	Н	I	J	К	L	Μ	N	0	Р	Q	R	S	Т	U
1	Sample_ID	1	2	3		4	5 6	5 7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Sex	F	F	М	F	м	м	м	F	F	М	F	М	М	F	М	F	М	F	М	F
3	Status	Ν	Y	Υ	N	N	N	Y	Y	N	Υ	Y	N	Y	N	N	Y	Y	Y	N	N
4	Cat1	3	7	9		3	1 4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
				es!B2:U2,"F",s											0.57328551						
	Cat3						8 0.36549674					0.66614321				0.49873951		0.57727393			
	Cat4	76	75			53 8	9 66	6 88	78	98	67	75	96	90	65	70	89	73	84	85	71
	Cat5	3	10	-		1	7 5	5	3	3	3	8	6	7	5	8	7	4	2	9	9
	Cat6			0.37747939				0.63984189							0.77463581						
	Cat7						9 0.85181114								0.10161224			0.63898328			
	Cat8	59	79	-		65 5												80			
	Cat9	0.15067873					6 62207075	1 21/ 041 25	11/22010			1 0/001 30	1 01/1/01/1/	1/1/20/52	TH A MONDOL		1 10510502				
-	Cat10	1646	1436			_			0									1049			
	Cat11	535 13	586 19			5	- 🗘 🛛 🗙	\sim	fx = C	OUNTIF	S(sample	e_values	!B2:U2,	"F",sam	ple_valu	ies!B3:U	J3,"Y")	466			
15 16	Cat12	13	19	15			*								. –			10	14	14	14
-10						A		В	С		D	Е	F		G	Н		State State			
					1	Female		10													
					-													-			
					2	Males		10													
					3	Exposed		10													
					4	Unexpo	sed	10													
					5	Female	+Exp	5													
		Las and			6																
1081				a series of	7																
252							NAMES OF STREET, SAME			10 11 10 10 10 10 10 10 10 10 10 10 10 1	10000 5 20 20 20 20 10 10 10 10		10 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	194 Sept. 11 49 40% (5.4		10 % TO 29 10 10 10 10 10	State and states			52	And the second
100000																					PARKER FOR THE CAPPED BALL



- AVERAGEIF: averages cells that meet one criteria
- AVERAGEIFS: averages cells that meet multiple criteria
- =AVERAGEIFS(range you want to average, first range of interest, what to look for in first range, second range of interest, what to look for in the second range....)
 - Range we want to average: Cat1
 - First range of interest: sex
 - Look for "F"
 - Second range of interest: exposure status
 - Look for "Y"







What if I want to know the average Cat1 value for all females that were exposed?

	Paste	Calibri (B B I		12 - A			On Sheet1, add
รเ	JM 🌲 :	× 🗸 fs	=averag	geifs(place for
/	A	В	С	D	E	F	AVERAGEIFS
1	Females	10					calculation
2	Males	10					culculation
3	Exposed	10					
4	Unexposed	10					
5	Female+Exp	5					
6	Cat1 avg F+Exp	=averageifs(
7		AVERAGE	FS(average	range, crite	ria_range1.	criteria1,)	
8					- • •		54



TEXAS A&M UNIVERSITY Superfund Research Center What if I want to know the average Cat1 value for all females that were exposed?

•		nsert [ら・ Draw Pa	age Layout	Formu	las Dat	ta Revie	ew Viev	v Acrol		sample
ſ	🛰 ب 🎮	Calibr	i (Body) 🔻	12 🔻	A▲ A▼	= =	- 8		Wrap Text	•	General
P	Paste	в	I U -					•=	Merge & C	enter v	\$ • 9
su	JM 🗍	× 🗸	f_x =ave	rageifs(sam	ple_values!	34:U4,					
	A	В	C AVE	RAGEIFS (av	erage_range,	criteria_ra	<mark>nge1</mark> , criteria	1, [criteria_ra	ange2,)	J	К
1	Sample_ID	1	2	3	4	5	6	7	8	9	
	Sex	F	F	М	F	М	М	м	F	F	м
	Status	N	Y	Y	N	N	N	Y	Y	N	Y
	Cat1	3	7	9	3	1	4	7	10	6	
,	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.228000
	Cat3	0.(=averagei	fs(sample_va	lues!B4:U4,	84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285
'	Cat4	76	75	65	53	89	66	88	78	98	
;	Cat5	3	10	3	1	7	5	5	3	3	
)	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419
0	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.069293
.1	Cat8	59	79	81	65	54	91	50	92	96	
2	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101:
3	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	10
.4	Cat11	535	586	543	470	479	502	582	537	412	
5	Cat12	13	19	15	18	11	19	18	10	16	

Cat1 values are in the sample_values sheet, row 4.

We are telling Excel, "The numbers I want averaged are located here."





What if I want to know the average Cat1 value for all females that were exposed?

ľ	• • • Home	nsert	റ ∙ത് Draw Pa	↓ age Layout	Formu	ilas Dat	a Revie	ew Viev	w Acrob	_	sample_d	ata		Values for Sex
2000			ri (Body) 🔻	12 💌	A▲ A▼	= =	= 8		💫 Wrap Text	*	General			are in the
energi (jili)	Paste	В	I <u>U</u>		- A -	$\equiv \equiv$	=	•=	Merge & C	enter =	\$ 7%)	.00 Coi .0 Foi	sample_values
s	им 🖞	× ×	f_x =ave	rageifs(sam	ple_values!	34:U4,samp	le_values!B	2:U2,"F",				90s.		
	A	В	C AVE	RAGEIFS (av	erage_range,	, <u>criteria_ran</u>	g <u>e1, criteria1</u>	[<u>criteria_ra</u>	nge2, criter	ia2], [criteria	a_range3,)	L	М	sheet, row 2.
1	Sample_ID	1	. 2	3	4	5	6	7	8	9	10	11	1	
2	Sex	F	F	м	F	М	М	М	F	F	М	F	м	
3	Status	N	Y	Y	N	N	N	Y	Y	N	Y	Y	N	VALS and talling
4	Cat1	3	7	3	3	1	4	7	10	6	-	1	0.0000540	We are telling
5	COLL	0.39980403	0.131167 eifs(sample_va		0.99121998		0.4932649 36549674		0.78559809 0.73402944	0.29231847		0.12970948 0.66614321		
7	Cat3	0.1=average 76			ample_values 53	89 sib2:02, F	50549074	0.08057165	0.73402944	98		0.00014321	0.1588000 q	Excel, "When you
8		3	10		1	7	5	5	3	3				
9		0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.250935	calculate the
10	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.0036507	calculate the
11	Cat8	59	79	81	65	54	91	50	92	96	51	80	6	
12	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.8144812	average, only
13		1646			1218	1151	1278	1329	1942	1085				
14		535			470	479	502	582	537	412		462		include values
15		13	19	15	18	11	19	18	10	16	11	15	1	
16														from females."





What if I want to know the average Cat1 value for all females that were exposed?

•	Home	nsert	ら・ Draw P	age Layout	Formu	las Dat	a Revie	ew View	v Acrob		sample_d	ata				Exposure status values
Ρ	Paste	в	ri (Body) • I <u>U</u> •					•=	Wrap Text	enter -	General \$ 💌 %) .00		≠ ditional Form natting as Ta	• Cell ble Styles	are in the sample_value
U	JM 💂	×	fx =ave	rageifs(sam	ple_values!	34:U4,samp	le_values!B	2:U2,"F",sa	mple_value	s!B3:U3,"Y)					
	A	В	С	D	E	F	G	н	1	J	К	L	M	N	0	sheet, row 3.
	Sample_ID	1	. 2	-	4	5	6	7	8	9	10	11	12		14	
	Sex	F	F	M	F	M	M	M	F	F	M	F	M	M	F	
	Status	N	Y	Y	N	N	N	Y	Y 10	N	Y	Y	N	Y	N	
÷	Cat1 Cat2	0.39980403	0.131167	9	0.99121998	0.37915433	0.4932649	0 31730005	0.78559809		0.22800014	0.12970948	0.20895121	8 0.09570753	0 57328551	We are telling
÷	Cat2 Cat3	0.01765164		0.03737433	0.84309817	0.41803328	0.4552045	0.02700000	000000000	0.20202017	0.6285995	0.66614321	0.15886006		0.522676	
ł	Cat4	76			53	89	66	88	78		67	75	96		65	Excel, "When
Ī	Cat5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	
	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	you calculate
ļ	Cat7		0.89977883		0.09046157	0.43938449	0.85181114	0.7906725			0.06929245	0.17350784	0.00365075		0.10161224	
÷	Cat8	59			65	54	91	50	92		51	80	63	78	70	the average,
ł	Cat9	0.15067873		0.2607653	0.54603649		0.63397075		0.1763019		0.0101267	0.8489138				
÷	Cat10	1646				1151	1278	1329	1942		1662	1748			1850 405	only include
+	Cat11 Cat12	535		0.0	470	479	502 19	582	537		514 11	462	487	543 20	405	only merade
t	Catiz	13	19	15	10	11	19	10	10	10	11	15	19	20	10	values for

exposed individuals."

BIG DATA N ENVIRONMENTAL SCIENCE AND TOXICOLOGY We have an answer!



TEXAS A&M UNIVERSITY Superfund Research Center

	A	В	С	D	E	F	G	н	L	J	K
1	Females	10									
2	Males	10									
3	Exposed	10									
4	Unexposed	10									
5	Female+Exp	5									
6	Cat1 avg F+Exp	4.4									
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35											
36											
37											
88											

58





TRUE or FALSE:

Understanding the effect of exposure on average Cat1 values for females could involve an AVERAGEIFS calculation for *unexposed (control) females*.







Let's see...

- Start out the same as before:
 - We still want average Cat1 values
 - We still want data for females

=AVERAGEIFS(sample_values!B4:U4, sample_values!B2:U2, "F",

	JM 🗘	×	f_{x} =ave	agens(sam	pie_valuesit	54.04,5amp	valuesib	2.02, F ,Sd	mple_value	SIB5:05, N				
2	A	В	C AVE	RAGEIFS (av	erage_range,	, <u>criteria_ran</u>	g <u>e1, criteria1</u>	, <u>[criteria_rar</u>	nge2, <u>criteri</u>	a2], [criteria	_range3, crit	eria3],)	М	N
1	Sample_ID	1	2	3	4	5	6	7	8	9	10	11	12	
2	Sex	F	F	М	F	М	М	М	F	F	М	F	М	м
3	Status	N	Y	Y	N	N	N	Y	Y	N	Y	Y	N	Y
4	Cat1	3	7	9	3	1	4	7	10	6	9	1	8	
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.095
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.296
7	Cat4	=avera	geifs(sample	_values!B4:U	4,sample_val	ues!B2:U2,"F	sample_val,	ues!B3:U3,"N	" 78	98	67	75	96	
8	Cat5	3	10	3	1	7	5	5	3	3	3	8	6	
9	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.3910
10	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.622
11	Cat8	59	79	81	65	54	91	50	92	96	51	80	63	
12	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.147
13	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	
14	Cat11	535	586	543	470	479	502	582	537	412	514	462	487	
15	Cat12	13	19	15	18	11	19	18	10	16	11	15	19	





Let's see...

 Now we have calculated the average value of Cat1 for females that were NOT exposed. sample_values!B3:U3, "N")

B7	‡ ×	$\checkmark f_x$	=AVERAGE	FS(sample_	values!B4:U	4,sample_v	alues!B2:U2	2,"F",sample	_values!B3	:U3,"N")	
	A	В	С	D	E	F	G	н	I	J	К
1	Females	10									
2	Males	10									
3	Exposed	10									
4	Unexposed	10									
5	Female+Exp	5									
6	Cat1 avg F+Exp	4.4									
7	Cat1 avg F+Unexp	2.8									
8											
9											
0											

Difference between exposed and unexposed (control)

- Calculate the difference by subtraction.
- This isn't a formula, so no parentheses

	А	В	-ch Cento
1	Is #8 female?	yes	
2	# of Females	10	
3	# of Males	10)
4	# Exposed	10)
5	# Unexposed	10)
6	# of Females Exposed	5	.
7	Cat1 avg F+Exp	4.4	
8	Cat1 avg F+Unexp	2.8	8
9	Difference: Exp-Unexp	= B7-B8	
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23	sample_values Sheet1	names_letters S	her her





TRUE!

• If we want to see how exposure affects the average Cat1 values for females, we could start by doing an AVERAGEIFS calculation for unexposed females.

B9	$ + \times$	$\checkmark f_X$
Ζ	A	В
2	Males	10
3	Exposed	10
4	Unexposed	10
5	Female+Exp	5
6	Cat1 avg F+Exp	4.4
7	Cat1 avg F+Unexp	2.8
8		1.6
٥		

Cat1 average value is <u>higher</u> in exposed females versus unexposed females.





How to find a specific value

LOOKUP function

• When you want to search a specific row or column for some value that corresponds to another row or column

• =LOOKUP(

- the specific thing you already know ["lookup_value"],
- the row/column that contains the information you already know ["lookup_vector"],
- the row/column that contains the value you're interested in ["results_vector"])





What is the Cat7 value for individual 8?

	A	В	С	D	E	F	G	Н	I	J	К	L	М	N	0	Р	Q	R	S	Т	U
1	Sample_ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Sex	F	F	м	F	М	м	м	F	F	м	F	М	М	F	м	F	м	F	м	F
3	Status	N	Y	Y	N	N	N	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	N
4	Cat1	3	7	9	3	1	4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.36314088
7	Cat4	76	75	65	53	89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	71
8	Cat5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	8	7	4	2	9	9
9	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.93651799
1	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847
1	Cat8	59	79	81	65	54	91	50	92	96	51	80	63	78	70	91	60	80	59	81	92
1	2 Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.47184386	0.97195415	0.10610593	0.55886697	0.34184183	0.38806705	0.24976963
1	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	1061	1850	1160	1166	1049	1419	1275	1207
1	Cat11	535	586	543	470	479	502	582	537	412	514	462	487	543	405	504	478	466	419	467	591
1	6 Cat12	13	19	15	18	11	19	18	10	16	11	15	19	20	18	13	17	16	14	14	14
1	5																				

=LOOKUP(the specific thing you already know ["lookup_value"], the row/column that contains the information you already know ["lookup_vector"], the row/column that contains some value that information corresponds to ["results_vector"])

- The lookup_value here is ID number 8
- The lookup_vector here is the row with all the ID numbers
 - IMPORTANT: Lookup_vector must be in ascending order. (Ours is already ordered numerically, left to right.)
- The **results_vector** here is the <u>row with all the Cat7 measurement values</u>.





What is the Cat7 value for individual 8?

su	л 🗘 🗙 🗸	∕ <i>fx</i> =lo	okup(
1	А	В	С	D	E	F
1	Females	10				
2	Males	10				
3	Exposed	10				
4	Unexposed	10				
5	Female+Exp	5				
6	Cat1 avg F+Exp	4.4				
7	Cat1 avg F+Unexp	2.8				
8	Difference: Exp-Unexp	1.6				
9	Cat7: individual 8	=lookup(
10		LOOKUP(I	ookup_valu	, lookup_vec	tor, [result_ve	ector])
11			ookup_valu			
12						



What is the Cat7 value for individual 8? Lookup value: individual 8

Lookup value = 8

I could also select cell containing the sample ID for individual 8.

J		1	н	esult_vector])			KUP(lookur KUP(lookur	L	В	A	
9	3	8	7	0	5	4	3	2	1	Sample_ID	1
	F	F	Λ	M	м	F	м	F	F	Sex	2
	N	Y	,	N Y	N	N	Y	Y	N	Status	3
()	10	7	4	1	3	9	7	3	Cat1	4
923184	ЭО	0.78559809	0.31730995	0.4932649	0.37915433	0.99121998	0.83797435	0.131167	0.39980403	Cat2	5
9952279	1	0.73402944	0.08057165	0.36549674	0.41803328	0.84309817	0.4243295	0.74075219	0.01765164	Cat3	6
98	3	78	88	66	89	53	65	75	76	Cat4	7
1	3	3	5	5	7	1	3	10	3	Cat5	8
602795	3 0	0.64802318	0.63984189	0.1925463	0.31935773	0.13183529	0.37747939	ll1 37	0.278176!es	Cat6	9
819160	2 0	0.03804672	0.7906725	0.85181114	0.43938449	0.09046157	0.42504482	0.89977883	0.47111146	Cat7	10
96	2	92	50	91	54	65	81	79	59	Cat8	11
752076) O	0.1763019	0.31429126	0.63397075	0.94019016	0.54603649	0.2607653	0.69674031	0.15067873	Cat9	12
1085	2	1942	1329	1278	1151	1218	1486	1436	1646	Cat10	13
412	1	537	582	502	479	470	543	586	535	Cat11	14
16)	10	18	19	11	18	15	19	13	Cat12	15
											16
											17
	1020403	1250CTRHOMOUTO	nationalitation	100000000000000000000000000000000000000	anternetternet		10110-1211-1201-1403	and the market state		NORTH AND THE NUMBER OF	17

10	$f_{\rm x}$ \sim $f_{\rm x}$ =LOOKUP(8, sam	ple_values!1:1,sample	values!10:10)
	А	В	С
1	Is #8 female?	yes	
2	# of Females	10	
3	# of Males	10	
4	# Exposed	10	
5	# Unexposed	10	
6	# of Females Exposed	5	
7	Cat1 avg F+Exp	4.4	
8	Cat1 avg F+Unexp	2.8	
9	Difference: Exp-Unexp	1.6	
10	Cat7: individual 8	0.03804672	
			12.00





What is the Cat7 value for individual 8? Lookup vector: all the sample IDs

I select the entire Sample ID row for sheet sample_values by clicking on the row number 1. (If my data were in columns, I could click the column letter to select the entire column.) This also tells Excel what sheet to look at.

SU	JM 🍦	🗙 🗸 j	x =LOO	DKUP(sample	e_values!11,	sample_val	ues!1:1														
	А	В		OKUP(lookup OKUP(lookup			esult_vector]) н	I	J	К	L	М	N	0	Р	Q	R	S	т	U
1	Sample_ID	1		3	4	, >	0	- 7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	Sex	F F		м	F	м	м	M	F	F	м	F	м	м	F	м	F	м	F	м	F
3	Status	N Y		Y	N	N	N	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	N
1	Cat1	3	7	9	3	1	4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628
5	Cat3	0.01765164 0	.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.3631408
7	Cat4	76	75	65	53	89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	7
3	Cat5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	8	7	4	2	9	
Э	Cat6	0.278176!es!1	:1 37	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.9365179
0	Cat7	0.47111146 0	.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847
1	Cat8	59	79	81	65	54	91	50	92	96	51	80	63	78	70	91	60	80	59	81	92
2	Cat9	0.15067873 0	.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.47184386	0.97195415	0.10610593	0.55886697	0.34184183	0.38806705	0.24976963
.3	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	1061	1850	1160	1166	1049	1419	1275	1207
4	Cat11	535	586	543	470	479	502	582	537	412	514	462	487	543	405	504	478	466	419	467	591
5	Cat12	13	19	15	18	11	19	18	10	16	11	15	19	20	18	13	17	16	14	14	14
6																					
17																					





What is the Cat7 value for individual 8? Result vector: all the Cat7 values

I select the entire Cat7 row by clicking on the row number 10 in the sheet sample_values. (Again, if my data were in columns, I could click the column letter to select the entire column.)

	A	В		<u>)KUP(lookup</u>)KUP(lookup			esult_vector	р н	I	J	К	
1	Sample_ID	1	<u></u>	3	<u>-varae, array</u> 4	, ,	0	7	8	9	10	
2	Sex	F	F	М	F	м	м	M	F	F	М	F
3	Status	N	Y	Y	N	N	N	Υ	Y	N	Y	Y
4	Cat1	3	7	9	3	1	4	7	10	6	9	
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	(
7	Cat4	76	75	65	53	89	66	88	78	98	67	
8	Cat5	3	10	3	1	7	5	5	3	3	3	
9	Cat6	0.278176!es	10:10 37	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	. (
10	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	. [0
11	Cat8	59	79	81	65	54	91	50	92	96	51	
12	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	
13	Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	
14	Cat11	535	586	543	470	479	502	582	537	412	514	
15	Cat12	13	19	15	18	11	19	18	10	16	11	
16												

BIG DATA 🍝 WENVIRONMENTAL SCIENCE AND TOXICOLOGY We got the correct answer!

She

names_letters

R

23

sample values

Sheet1

1	Is #8 female?	yes													
2	# of Females	10													
3	# of Males	10	L.	10 🛔	× v	fx 0.038	3046718143	6256							
4	# Exposed	10		A	В	С	D	E	F	G	н	I	J	К	
5	# Unexposed	10	1	Sample_ID	1	2	3	4	5	6	5 7	8	9	10	1
6	# of Females Exposed	5	2		F	F	м	F	м	м	м	F	F	м	F
7	Cat1 avg F+Exp	4.4	3		N 3	Y 7	Y 9	N 3	N	N	Y 7	Y 10	N	Y 9	Y
2			4		0.39980403	4			0.37915433	-	0.31730995				
8	Cat1 avg F+Unexp	2.8	6			0.74075219					0.08057165				
9	Difference: Exp-Unexp	1.6	7		76					66					
10	Cat7: individual 8	0.03804672	8	Cat5	3			-	7	5	-	-	-	3	4
		0.0000.072	0.0545.900	Cat6		0.07110137					0.63984189				-
11				Cat7	1	0.89977883							0.98191601		
12			535 A 54 5 6 5 1	Cat8 Cat9	59	79 0.69674031				91	50 0.31429126		96 0.87520766		
13			10 Carl 1 C 10 C 10 C	Cat9 Cat10	1646					1278		1942			_
14			2000 C C C C C C C C C C C C C C C C C C	Cat10	535										
			(1996)E-41511	Cat12	13					19					_
15			16	5											
16			17												
.7		8	18												
			19 20												
18			20												
19			21 1		ALEMONICATION	Constant Constant		COLORING MARK		al an	t seale states		West of Contraction		6123
20															
21															
22															

TEXAS A&M UNIVERSITY
Superfund Research Center

ĀM





When to use LOOKUP



- LOOKUP can be very useful when you have a large dataset to search.
- LOOKUP is used when the value you want to match is in the first row/column.
- If your lookup_vector isn't in **ascending order**, it is possible you will get a result... But it may not be the correct result!



Ordering data: columns

 Make sure your lookup_vector data is in ascending order by SORTING the lookup_vector row/column before you try to use LOOKUP.

MENTAL SCIENCE ANI

 Data → Sort → choose which column you want to sort, and how you want it sorted (e.g. ascending order [A to Z], descending order [Z to A], etc.)

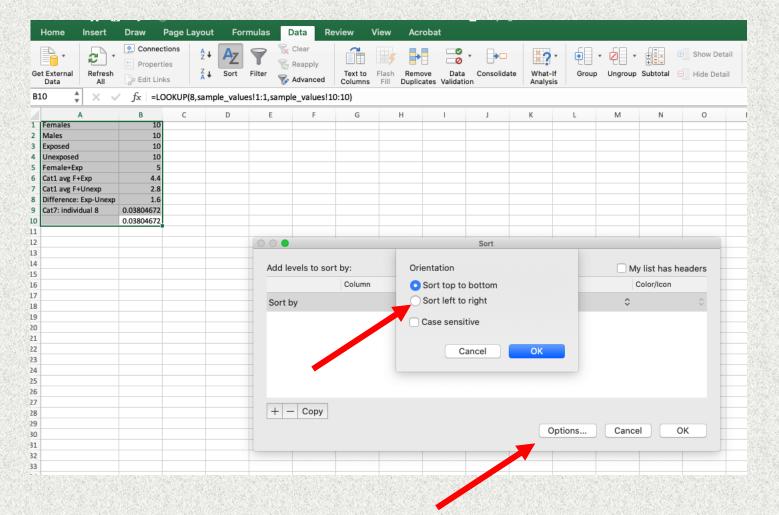
Hom	e Insert		Page Layou	t For	mulas		Review	View	Acrobat						
et Exte Data	rnal Refresh	Connect	es Z	~	T	Clear Reapply	Text to Columns			Ø ▼ →□ Ø ▼ →□		Group	• Ungroup	+ + + Subtotal	Show Det Hide Deta
10	‡ × √		OKUP(8, san	nple_valu					photes va		, maryolo				
(A	В	С	D	E	F	G	н	1	J	К	L	M	N	0
Fema	les	10													
Males		10													
Expos		10													
Unexp		10 5													
	le+Exp														
	avg F+Exp	4.4													
	avg F+Unexp	2.8													
	ence: Exp-Unexp	1.6													
Cat/:	individual 8	0.03804672 0.03804672													
		0.03804672													
-					00					0 t					
-					00					Sort					
-					+										
					Add	l levels to s	ort by:						M;	y list has	headers
							Column			Sort On	Order		(Color/Icon	
					Sor	rt by				🗘 Values	A to Z		\$		\sim
_					-										
-						Corre									
					+	— Сору									
_					-						Opt	tions	Cance		ОК
											op				





Ordering data: rows

 If you need to sort a row instead of a column, click "Options..." and select "Sort left to right"





VLOOKUP

- =VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])
- =VLOOKUP(
 - Lookup_value,
 - Table_array,
 - Col_index_num,
 - Range_lookup)

What you want to look up Where you want to look for it The column number in the range cont

The column number in the range containing the value to return Return an Approximate (TRUE) or Exact (FALSE) match





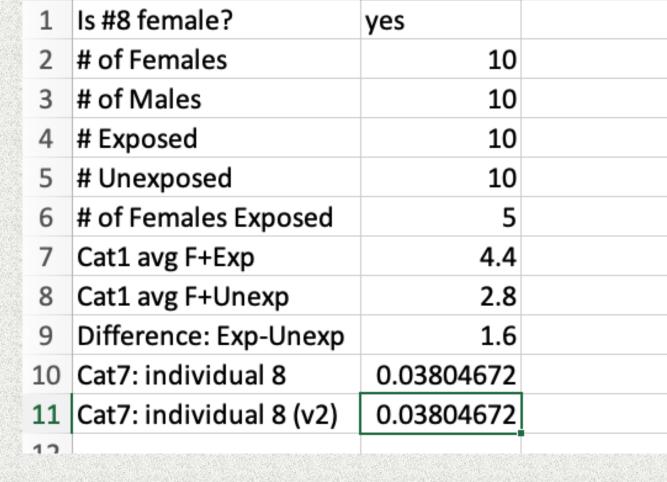
- VLOOKUP doesn't require anything be sorted by ascending order!
- BUT, the value you look up needs to be to the left of the return value you want to find.
- **XLOOKUP** can look in any direction, but it's not available on all versions of Excel.
- INDEX and MATCH, used independently or nested, may be superior to VLOOKUP for some circumstances.
 - https://exceljet.net/index-and-match



TEXAS A&M UNIVERSITY

SCIENCE AND TOXICOLOGY VLOOKUP: What is the Cat7 value for B11 =VLOOKUP(sample_values!A10,sample_values!A1:U15,9,FALSE) individual 8? А

 Repeating the search we did before, but this time pretending nothing is in order







С

В



- We want to know the Cat7 value... so that's my lookup_value
 - Don't highlight the whole row/column, just select the category name

• =VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])

• =VLOOKUP(

- Lookup_value,
- Table_array,
- Col_index_num,
- Range_lookup)

What you want to look up Where you want to look for it The column number in the range containing the value to return Return an Approximate (TRUE) or Exact (FALSE) match

sι	JM 🍦	×	f_x =vloo	kup(sample	_values!A1	D	
/	A	В	С	D	E	F	000.42
1	Sample_ID	1	2	3	4	5	1
2	Sex	F	F	М	F	М	м
3	Status	N	Y	Y	N	N	Ν
4	Cat1	3	7	9	3	1	
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.3
7	Cat4	76	75	65	53	89	
8	Cat5	3	10	3	1	7	
9	Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.
0	Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.8
1	Cat8	A10 59	VLOOKUP(lo	okup_value	, table_array,	col_index_n	um,
2	Cat9	0.10057873		0.2607653	0.54603649	0.94019016	0.6
3	Cat10	1646	1436	1486	1218	1151	
4	Cat11	535	586	543	470	479	
5	Cat12	13	19	15	18	11	
6							
7							
8							
9							
0							
1							
2							
3							
4							
5							
6							
7							
8							
9							
0							
1							
2							
3							
4							
5							-
6							1000
7							
8						80	



32 33

Sheet2

sample_values

+

• =VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])



• =VLOOKUP(**VLOOKUP**

What you want to look up Where you want to look for it The column number in the range containing the value to return Return an Approximate (TRUE) or Exact (FALSE) match

118			0.000		C. ToleView		0.5.000		1. 1				10.5 10.0 10.00	1.3.4.000	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1			1.053.076		1.093.000	ALC: THE REAL PROPERTY OF	1
		2	Connectio	2 +	AZ	Cle		1	7	0 1	•	2.	•		+ - +	Show Detail	¢ [#]					
Ge	et External Data	Refresh All	> Edit Links	Ă∱	Sort Fi	lter 🦷 Ad		Text to Flas Columns Fil		Data s Validation	Consolidate	What-If Analysis	Group	Ungroup Su	ubtotal 😑	Hide Detail	Analysis Tools	Data Analysis				
SU	JM 🛔	×	$f_x = v loo$	kup(sample	_values!A1	0,sample_va	alues A1:U1	5														
	A	В	C VLC	OKUP(looku	<u>ip_value, tab</u>	ole_array, co	l_index_num,	[range_look	up])	J	К	L	М	N	0	Р	Q	R	S	т	U	
1	ample_ID	I	2	3	4	5	6	7	8	9	10	11	1z	13	14	15	10	17	18	19	20	1
2	ex	F	F	м	F	м	м	м	F	F	м	F	м	м	F	м	F	м	F	м	F 5	
3	status	N	Y	Y	N	N	N	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	N	
4	Cat1	3	7	9	3	1	4	7	10	6	9	1	8	8	1	1	2	5	2	8		
5	cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628	Ĺ
6	at3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.36314088	
7	at4	76	75	65	53	89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	71	
8	at5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	8	7	4	2	9		
9	at6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.93651799	
10	Cat7	0.471111/=v	lookup(samp	le_values!A10),sample_val	ues!A1:U15	14	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847	i -
11	Cat8	59	79	81	65	54	91	50	92	96	51	80	63	78	70	91	60	80	59	81	92	
12	Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.47184386	0.97195415	0.10610593	0.55886697	0.34184183	0.38806705	0.2497696	
13	cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	1061	1850	1160	1166	1049	1419	1275	1207	
14	at11	535	586	543	470	479	502	582	537	412	514	462	487	543	405	504	478	466	419	467	591	1
15	at12	13	19	15	18	11	19	18	10	16	11	15	19	20	18	13	17	16	14	14	14	i.
16																		_				
17																						
18																						1
19																						
20																						
21																_						
22																L						

Lookup_value,

Range_lookup)

 Table_array, • Col_index_num,

> The table_array can be the entire table, or whatever section you want to search.





- Individual #8, the one I'm interested in, is 9 columns to the right of my lookup_value (Cat7).
- So here, 9 is my col_index_num

• =VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])

• =VLOOKUP(

- Lookup_value,
- Table_array,
- Col_index_num,
- Range_lookup)

What you want to look up Where you want to look for it

The column number in the range containing the value to return Return an Approximate (TRUE) or Exact (FALSE) match

	A	В	c <u>VLC</u>	OKUP(looku	up_value, tab	le_array, col_	<u>index_num,</u>				J	К
1	Sample_ID	1	2	3	4	5	6		Option	ıs		1
2	Sex	F	F	м	F	М	М	м	TRU	E – Approxi	mate matcl	h
3	Status	N	Y	Y	N	N	N	Y	FALS	E - Exact r	natch	
4	Cat1	3	7	9	3	1	4		7	10	6	
5	Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31	730995	0.78559809	0.29231847	0.2280001
6	Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08	057165	0.73402944	0.9952279	0.628599
7	Cat4	76	75	65	53	89	66		88	78	98	e
8	Cat5	3		3	1	7	5		5	3	3	
9	Cat6		0.07110137								0.26027957	
10	Cat7		lookup(samp	le_values!A10),sample_valu	ues!A1:U15,9		0.7		0.03804672	0.98191601	0.0692924
11	Cat8	59		81	65	54	91		50	92	96	5
12	Cat9		0.69674031				0.63397075	0.314			0.87520766	0.010126
13	Cat10	1646	1436	1486	1218	1151	1278		1329	1942	1085	166
14	Cat11	535	586	543	470	479	502		582	537	412	51
15	Cat12	13	19	15	18	11	19		18	10	16	1
16												
17												
18												
19	1	2	3	4	5	6	7		8	9		
20		<u> </u>	5	-+	J	U	/		0	9		
21												
22												
23												
24 25												
25 26												
20												
27												
20 29												
30												
31												
32												
33												
34												
35												
36												
37												
38												



- Now I put FALSE (or 0) because I want an *exact* match.
- If I'd put TRUE, or 1, I might get an *approximate* match.

B1	10 🜲 🗙 🗸	f_x =VLC	OKUP(san	nple_values	A10, sample	_values!A	1:U15,9,FAL	SE)
/	A	В	С	D	E	F	G	Н
1	Females	10						
2	Males	10						
3	Exposed	10						
4	Unexposed	10						
5	Female+Exp	5						
6	Cat1 avg F+Exp	4.4						
7	Cat1 avg F+Unexp	2.8						
8	Difference: Exp-Unexp	1.6						
9	Cat7: individual 8	0.03804672						
10	Cat7: individual 8 (v2)	0.03804672						
11								
12								
13								
1.4								

• =VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])

• =VLOOKUP(

- Lookup_value, What you want to look up
- Table_array,
- Col_index_num,
- Range_lookup)
- Where you want to look for it The column number in the range containing the value to return
- Return an Approximate (TRUE) or Exact (FALSE) match





 Pey, what are you upto?

 Image: what are you upto?
 <

sι	лм 🌲 🗙 🗸	fx =V	LOOKUP(sa	mple_value	s!A10,samp	le_values!A	1:U15,9,TR	UE)		
/	A	в	/LOOKUP(lo	okup_value,	table_array, o	ol_index_nur	n, [<u>range_lo</u>	okup])	I	
1	Females	10					C	ptions		
2	Males	10						TRUE - App	oroximate n	natch
3	Exposed	10						FALSE - Ex	act match	
4	Unexposed	10								
5	Female+Exp	5								
6	Cat1 avg F+Exp	4.4								
7	Cat1 avg F+Unexp	2.8								
8	Difference: Exp-Unexp	1.6								
9	Cat7: individual 8	0.03804672								
10	Cat7: individual 8 (v2)	0.03804672								
11		9,TRUE)								
12										

(For fun, I tried putting "TRUE" and I got the same result in this situation.)



Nested VLOOKUP

- What if I want to know whether a certain result could be found for a certain individual?
 - Here: is the value "10" found for "Brinkmeyer-Langford"?
- We can nest the VLOOKUP function inside another function, in this example an IF.
 - We just want a true/false answer here, so we start with the IF function.

	BS)	fx =IF(VLOOKUP	(A6,A2:B7,2	2,FALSE)="1	.0","YES","N	10")	
	1	A	В	С	D	E	F	G	
	1	Name	Num_letters						
	2	Fred	4						
	3	Candice	7						NOTE: Excel can also count o
	4	Dillon	6						
	5	Wright	6						look for values that are
	6	Brinkmeyer-Langford	18						
	7	Lloyd	5						greater or less than
34 i	8								
	9		NO						something, using "<" and ">"
1	0								





Nested VLOOKUP

BS) \$ ×∽	√ <i>f</i> x =I	F(VLOOKUP	(A6,A2:B	7,2,FALSE)="1	0","YES","I	NO")
7	А	В	С	D	E	F	G
1	Name	Num_letters			VIOOKUP	to detern	nine the value in the
2	Fred	4					
3	Candice	7			_		n corresponding to
4	Dillon	6			"Brinkmey	ver-Langfo	ord"
5	Wright	6					
6	Brinkmeyer-Langford	18				alue "Bri	nkmeyer-Langford" or A6
7	Lloyd	5			· · -		, .
8					Lookup_ve		oie table
9		NO			Col_index	_num: 2	
10							





Nested VLOOKUP

BS) =IF	VLOOKU	P <mark>(A6,A</mark> 2	::B7,2,F/	ALSE <mark>)</mark> ="2	10","YE	s","NO")
/	A	В	С	D	E	F	G
1	Name	Num_letters					
2	Fred	4					
3	Candice	7					
4	Dillon	6	=IF(VLO	OKUP(A6,	A2:B7,2,F	ALSE)="1	0","YES","NO")
5	Wright	6					
6	Brinkmeyer-Langford	18					
7	Lloyd	5					
8							
9		NO					
0							

TIP: KEEP AN EYE ON THE PARENTHESES!

The VLOOKUP part is all together, within the IF part. Excel color-codes them for you automatically. Excel sees you want to do an IF (true/false) test, but will first complete the VLOOKUP part, then use the result value it gets to finish working out the IF part. If parentheses aren't in the proper place, you'll get an error message.





QUESTION 3:

• Which of these formulas will calculate the total number of letters in "Candice Brinkmeyer-Langford"?

A =(VLOOKUP("Candice",A2:B7,2,FALSE))+(VLOOKUP("Brinkme yer-Langford",A2:B7,2,FALSE))

B =LOOKUP(A3,B2:B7)+LOOKUP(A6,B2:B7)

C =(VLOOKUP(A3,A2:B7,2,FALSE))+(VLOOKUP(A6,A2:B7,2,FALS E))

D =VLOOKUP((A3,A2:B7,2,FALSE)+(A6,A2:B7,2,FALSE))

/	А	В
1	Name	Num_letters
2	Fred	4
3	Candice	7
4	Dillon	6
5	Wright	6
6	Brinkmeyer-Langford	18
7	Lloyd	5
SAVENI MENERALI		



QUESTION 3:

• Which of these formulas will calculate the total number of letters in "Candice Brinkmeyer-Langford"?

A =(VLOOKUP("Candice",A2:B7,2,FALSE))+(VLOOKUP("Brinkmeyer -Langford",A2:B7,2,FALSE))

B -=LOOKUP(A3,B2:B7)+LOOKUP(A6,B2:B7)

C =(VLOOKUP(A3,A2:B7,2,FALSE))+(VLOOKUP(A6,A2:B7,2,FALSE))

/	A	В
1	Name	Num_letters
2	Fred	4
3	Candice	7
4	Dillon	6
5	Wright	6
6	Brinkmeyer-Langford	18
7	Lloyd	5

Ð

-=VLOOKUP((A3,A2:B7,2,FALSE)+(A6,A2:B7,2,FALSE))

=VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])⁸⁹



Another example of a nested VLOOKUP

- How can I determine whether Brinkmeyer-Langford is the longest name?
- And if it is longest, how much longer is "Brinkmeyer-Langford" than the average name?

/	A	В
1	Name	Num_letters
2	Fred	4
3	Candice	7
4	Dillon	6
5	Wright	6
6	Brinkmeyer-Langford	18
7	Lloyd	5



93

Another example of a nested VLOOKUP

- How can I determine whether Brinkmeyer-Langford is the longest name?
- And if it is longest, how much longer is "Brinkmeyer-Langford" than the average name?

1 Name Nun 2 Fred 3 Candice 4 4 Dillon	В
3 Candice	n_letters
	4
4 Dillon	7
	6
5 Wright	6
6 Brinkmeyer-Langford	18
7 Lloyd	5

=IF((VLOOKUP(A6,A2:B7,2,FALSE))= MAX(B2:B7)), VLOOKUP(A6,A2:B7,2,FALSE)-(AVERAGE(B2:B7)),0)

If the value for # of letters in Brinkmeyer-Langford is the same as the maximum value in the Num_letters column, then how many more letters does Brinkmeyer-Langford have than the average value of the Num_letters column? Or if the # of letters in Brinkmeyer-Langford is NOT equal to the greatest value for Num_letters, return "0".





Charting data

- A good hypothesis will take you a long way in knowing:
 - What to chart
 - What type of chart to use
 - What data to represent on X and Y axes
 - If a secondary axis could be useful for telling the story

If you don't know exactly where you're going, how will you know when you get there?

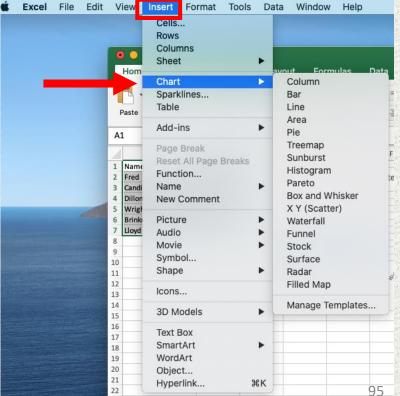
> ~Steve Maraboli From: RaiseYourMind.com





- Start by selecting the data you wish to graph.
- Then, Insert (either Insert menu gets you to Chart)

• • •	6	ن ک	ি 🖛								6	🖻 sample_data							
Home	Insert	Draw	Page La	yout l	Formulas	Data	Review	View	Acrobat										
PivotTable R	Recommende PivotTables	d Table	Pictures	~	Sn	Models * nartArt * reenshot *		t Add-ins Add-ins ▼	🔹 Visio Data 🗾 Bing Maps		Recommende	ו•	եւ հեր թ. հեր	Maps	PivotChar				
A1	\$ × ~	fx	Name																
	A	В	С	D	E	F	G	н	I	J	К	L	М	N	0				
Name		Num_lette	rs																
Fred			4																
Candice			7																
Dillon			6																
Wright			6																
Brinkmey	er-Langford	1	8																
Lloyd			5																
Constant in									NATION S										



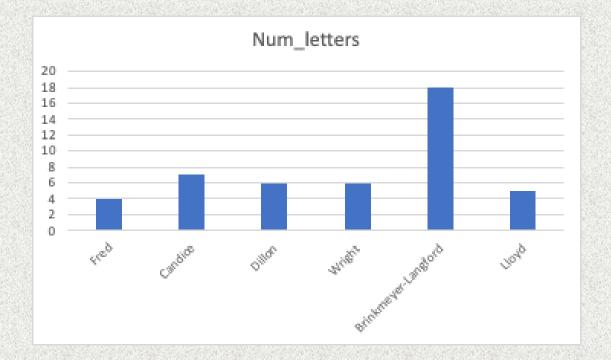


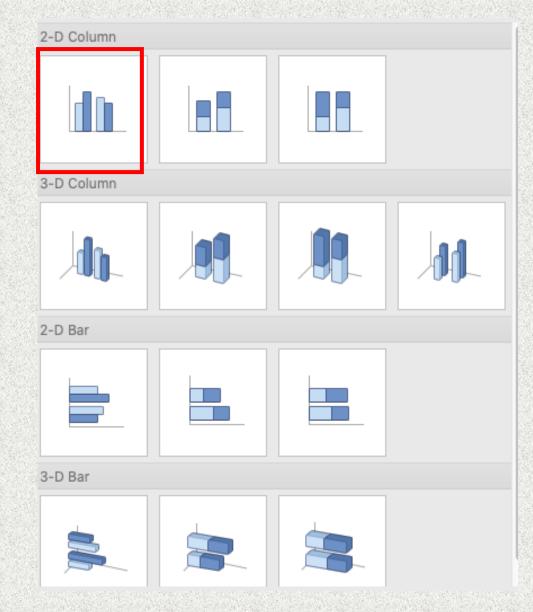
 TEXAS A&M UNIVERSITY

 Superfund Research Center

Bar charts

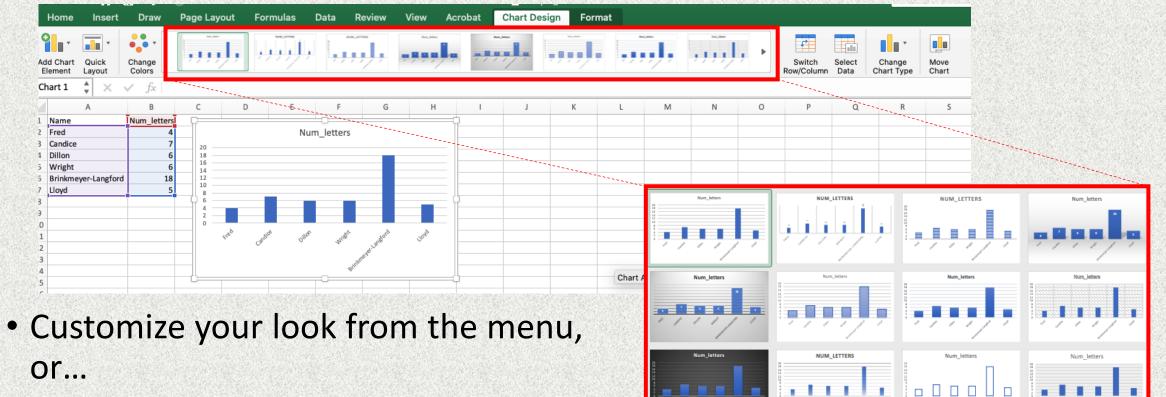
- Excel has multiple options for bar charts.
- Simplest: 2-D Column











Num_letters

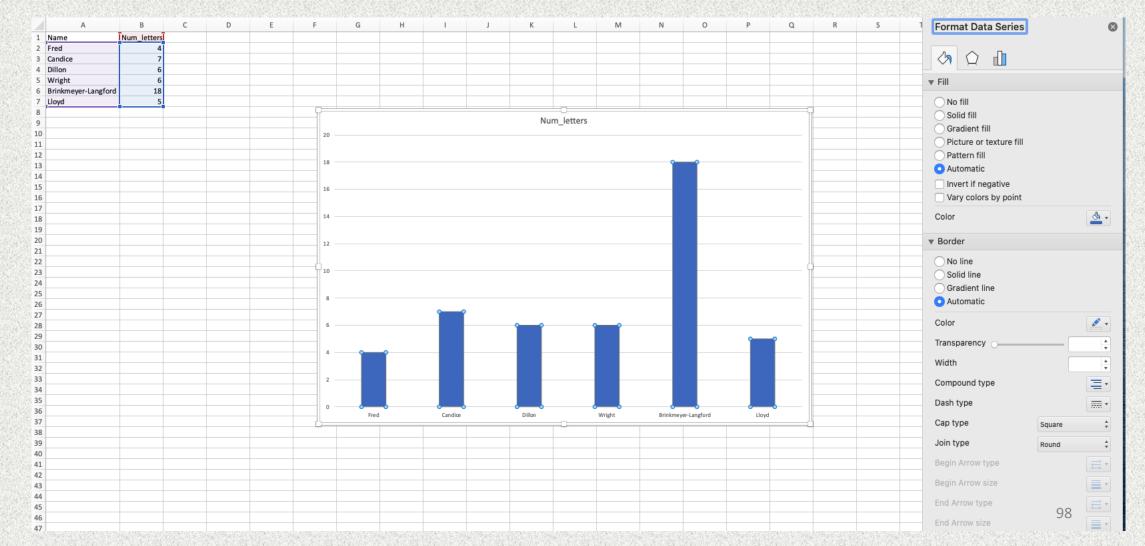
NUM LETTE

Num letter





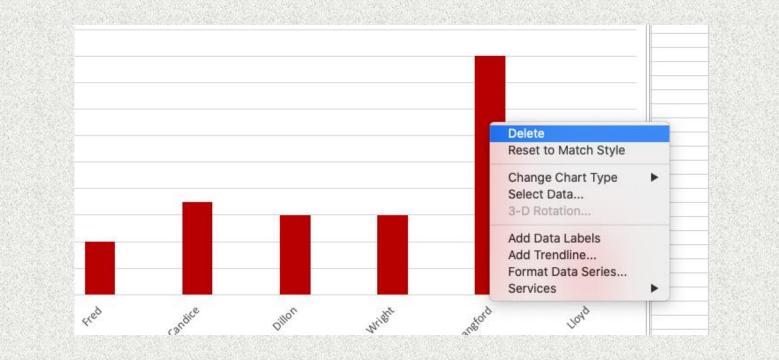
• Change colors, font, etc. by selecting the item.







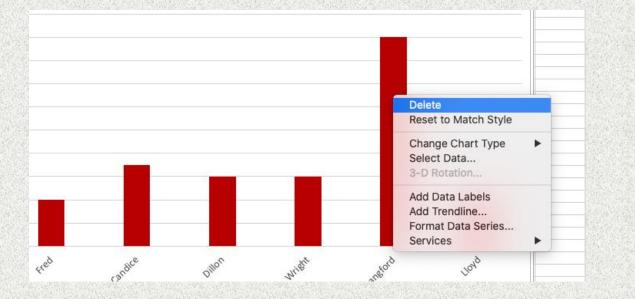
- Change colors, font, etc. by selecting the item.
- Right click to get a menu.
- Choose "Format [whatever]"

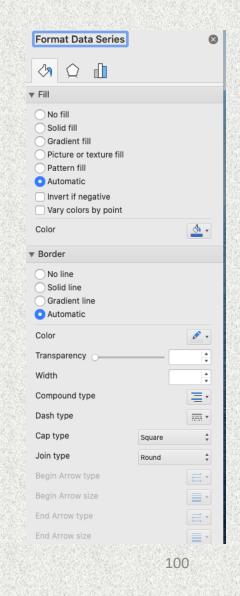






- Change colors, font, etc. by selecting the item.
- Right click to get a menu.
- Here, I right clicked on one of the bars and chose "Format Data Series" to change the bar colors.
- A menu appears to the right.





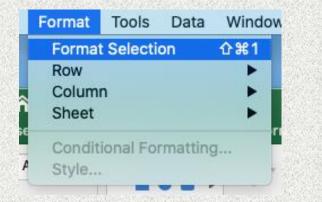




• Another option: select the Format menu at the top of the spreadsheet window.

$\blacksquare \bullet \blacksquare \bullet$	•••	6	ب ک	৫	Ŧ										🔊 san	nple_data		
	Home	Insert	Draw	Pa	ge La	ayout	Form	ulas	Data	Review	View	Acrob	at	Chart Design	Format			
	Vertical (Va	alue) Axis	•	A	\sum	2	12-				<hr/>	<u></u>	. 🖉	•	•		Α.	A .
sinape in the set to Match Style Shape Sha	Reset to	o Match Style			11	•	Change Shape					Shape Fill	4	. A	A	A	Text Fill	۰ ۵

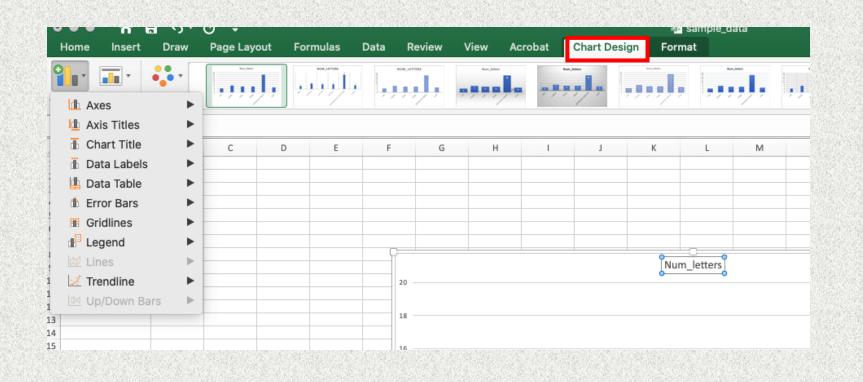
• You can also use the Format menu at the top of the Excel window.







• Add additional items like axes titles, error bars, etc. using Chart Design menu at the top.







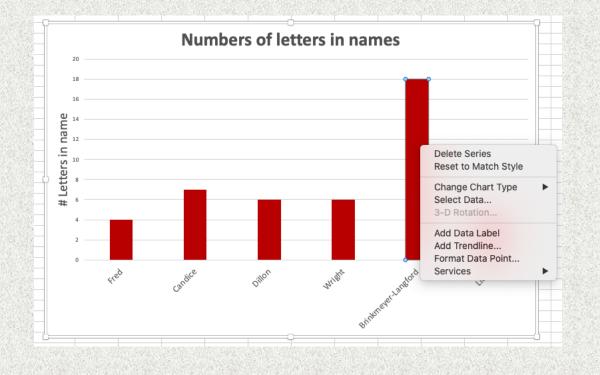
• To change font options for any part of the graph, select what you want to changes, then use use the **Home** menu.

🔍 🦳 🕐 🖪	ד שימים						×∎ sar	mple_data		11219125668.003	0.0000000000000000000000000000000000000	11219125668.003
Home Insert	Draw Page Layout	Formulas [Data Re	eview View	Acrobat	Chart Design	n Format					
Cut	Calibri (Body) 🔻 9	• A• A•	= = =	= * *	📑 🖉 Wrap Text 🔻	Gene	ral	T			•	× - 1
Paste Sormat	BII	<u></u> • <u>A</u> •		≣♦≣♥≣	↔ Merge & Cer	ter • \$ •	%)	00. 00.	Conditional Format Formatting as Table	Cell Styles		Delete For
Chart 2 🌲 🗙 🗸	/ fx											
A	B C	D E	F	G	н і	J	К	L	M N	0	Р	Q
	Num_letters											
2 Fred	4											
3 Candice	7											
4 Dillon	6											
5 Wright 6 Brinkmeyer-Langford	6											
6 Brinkmeyer-Langford 7 Lloyd	18											
8	5		0		1							
9							- Alst	+ 1.				
10						umper	s ot let	ters II	n names			-
11				20								
12												
13				18								
14												
15				16								
16				u								
17			ě	14								
18				12								
19												
20 21				10								
22			#									
23			—0 t	8								ü
24				۲ ۲								
25			#	\$								
26				4								
27												
28				2								
29												
30												
31				4red	cand	.e	Dillon	Mile	nt Lanet	6	11040	
32				440	and		Oille	Nill	- 19	J.	10,	
33					0			TATI NUT IN A	2			





- Here, I have selected one data point (Brinkmeyer-Langford).
 - Format Data Point lets me change color, pattern, etc. just for that one bar.
 - Add Data Label shows the values for each selected data point (or for everything, if you choose).

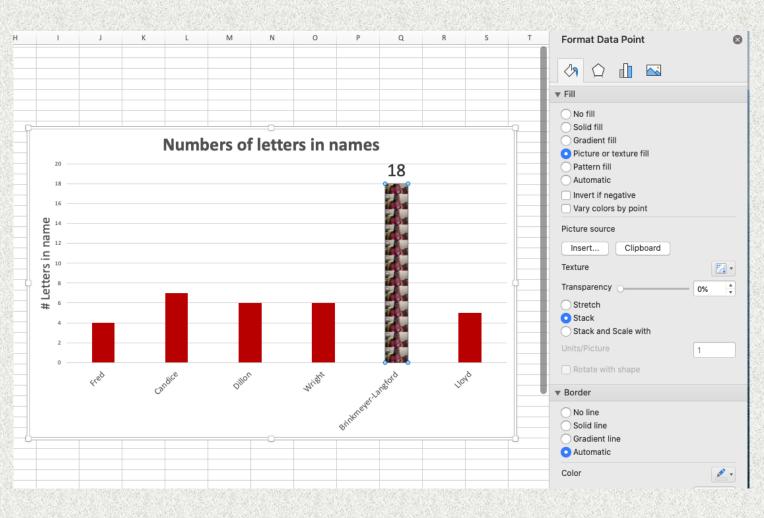




BIG DATA

IN ENVIRONMENTAL SCIENCE AND TOXICOLOGY

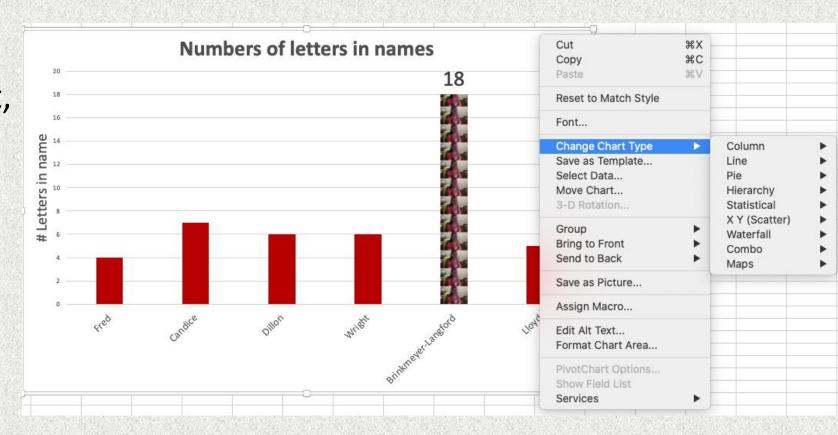
- Here, I have selected one data point (Brinkmeyer-Langford).
 - Format Data Point lets me change color, pattern, etc. just for that one bar.
 - Add Data Label shows the values for each selected data point (or for everything, if you choose).







- Changing your chart type entirely:
 - Select the whole chart, right click, Change Chart Type







- Changing your chart type entirely:
 - Select the whole chart, right click, Change Chart Type
- OR, menu at top of spreadsheet:
 - Chart Design

Chart Design

Format

Change Chart Type

56

Select

Data

Change

Chart Typ:

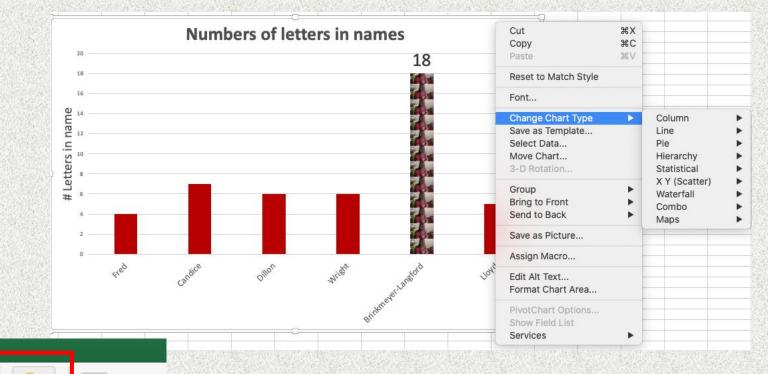
Switch

Row/Column

Move

C

ange Chart Type



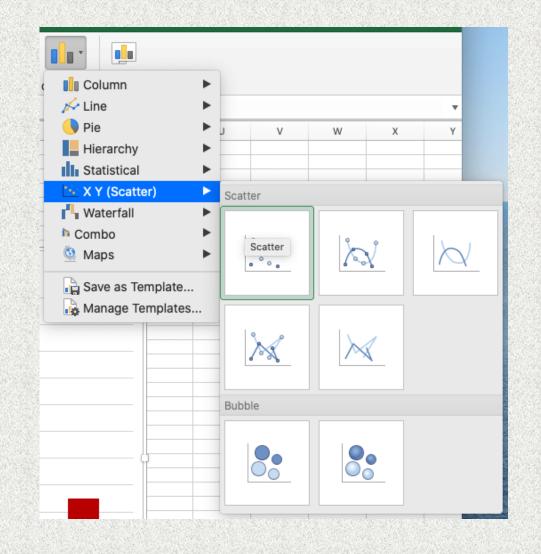


 TEXAS A&M UNIVERSITY

 Superfund Research Center

Changing your chart

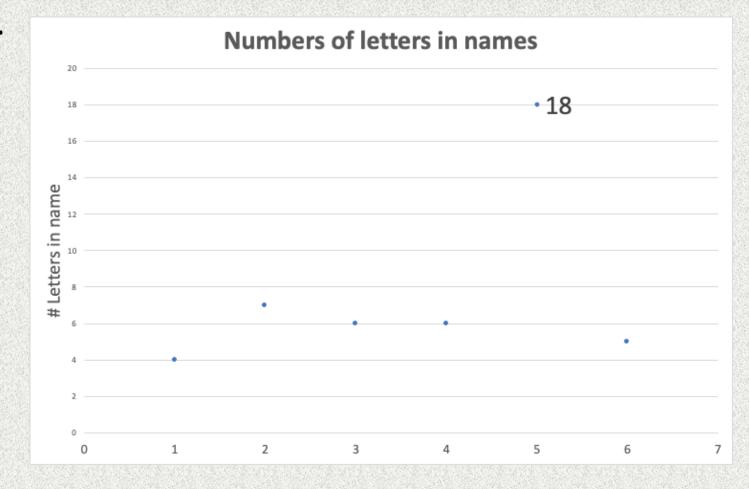
• Let's try scatter plotting...







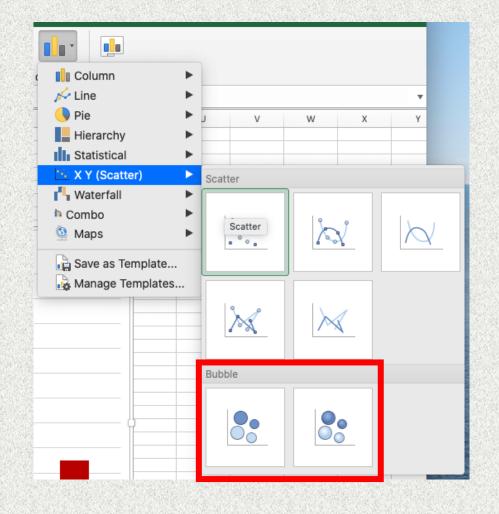
- Let's try scatter plotting...
- Not ideal here.







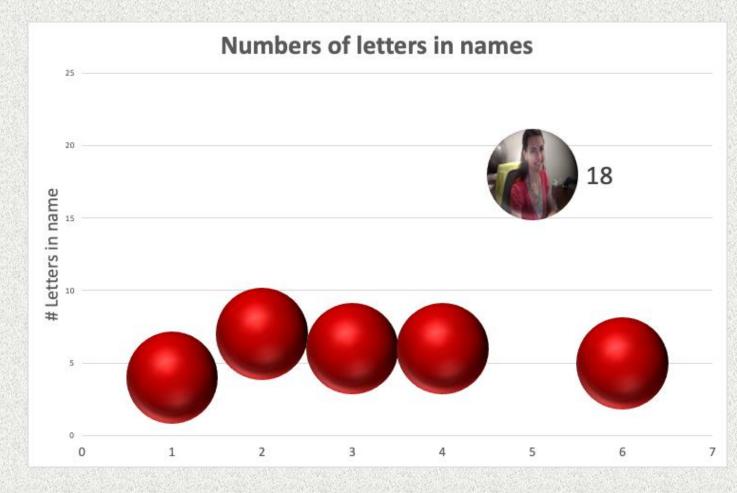
• Bubble scatter plot?!







- Bubble scatter plot
 - Can vary bubble sizes, colors, data labels, etc. as before







Scatterplots

 Simple scatterplot to compare Cat10 values

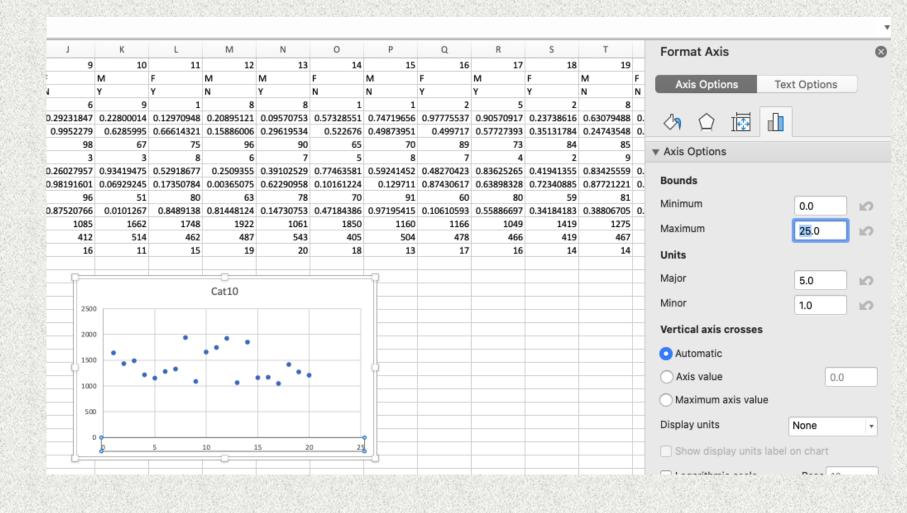
0		6	∙ ∿	- -								🔊 🔊 🔊 🔊	mple_data								
	Home	Insert	Draw P	age Layout	Formu	ulas Dat	ta Revie	ew Viev	v Acrob	oat Cha	art Design	Format									
	ement L	Quick Cha ayout Co	ange olors						Cet0		-					► S	witch Sel Column Da	ect Cha		ve	
h	art 1	\times \checkmark	fx																		
1	A	В	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р	Q	R	S	Т	U
	Sample_ID			2 3		-	-		-						14						
-	Sex	F	F		F			м						м	F				•		F
	Status	N	Y		N			•	-		•	-		Y	N				Y		N
	Cat1	3		⁷ 9		-									1		-		2		
	Cat2	0.39980403		0.83797435																	
	Cat3		0.74075219			0.41803328						0.66614321				0.49873951		0.57727393			
	Cat4	76							78	98	67		96	90	65		89	73	84	85	
	Cat5	3			-		-	-	-	-	3	-	6		-		7		-	9	
	Cat6			0.37747939																	
	Cat7			8 0.42504482														0.63898328			
	Cat8	59									51		63	78	70		60	80	59	81	
-	Cat9	1646	0.69674031			0.94019016			1942	1085	1662		0.81448124	0.14/30/53	1850		1166	1049	1419	1275	
_	Cat10 Cat11	535								412			487	543	405		478		419	467	
	Cat11 Cat12	13											487				478			407	-
6	Catiz	15		, 15	10	11	15	10	10	10		15	15	20	10	15	1/	10	14	14	
7																					
8													Cat10			T.					
9													CallO								
0										2500											
1																					
2										2000		•									
3											•		• •								
4											• •			• •		Ċ.					
5										1000	•	ŧ٠ •	•	••. *							
6										1000											
7										500											
8																					
9										0											
											0	5	10 1	15 2	0 2	5					
0										Ŭ=			-0			-0 -					
0 1 2																					





Scatterplots

- Format chart as before...
 - Here, updating x-axis bounds.

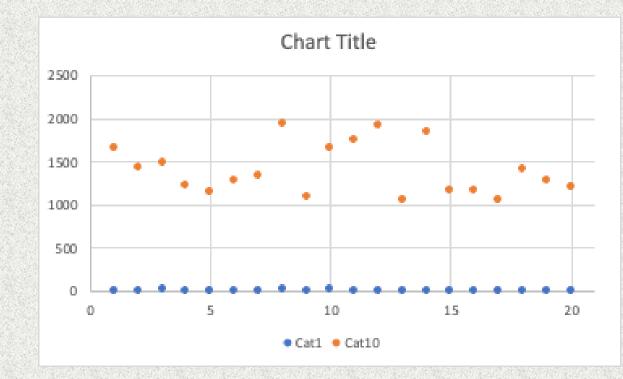






Scatterplots

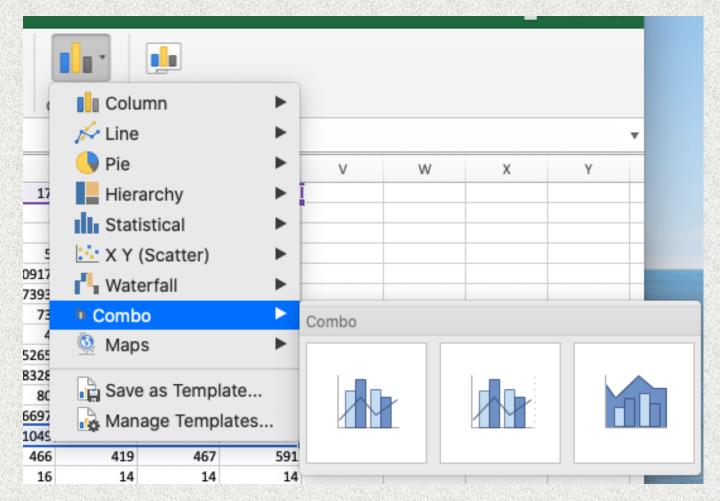
- We can add >1 categories to a scatterplot.
- Excel makes them different colors by default, but this can be changed using the Format menus as before.
- But here, Cat1 and Cat10 are on very different scales...





Adding a secondary axis to show data with different scales

• To represent data on separate axes (e.g. y- and z-z-axes), we start by changing the chart type to Combo.



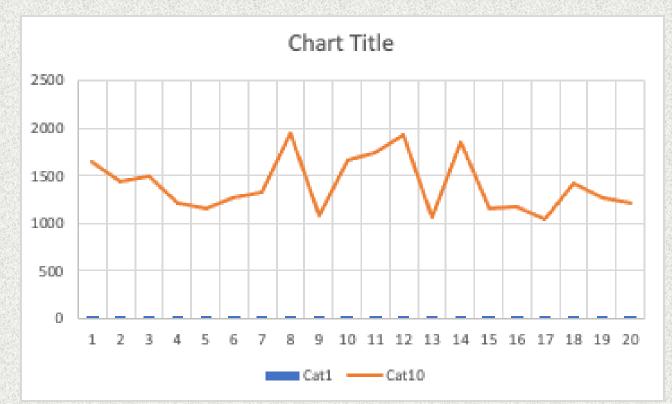
TEXAS A&M UNIVERSITY





Adding a secondary axis

• Here is what it looks like at first:

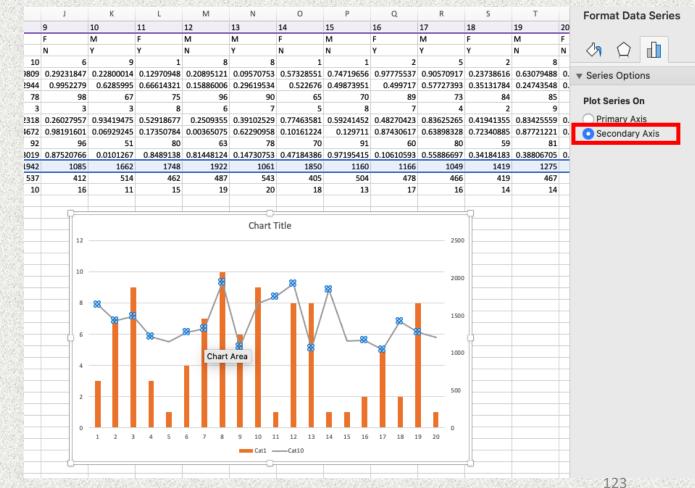






Adding a secondary axis

- Select one of the two categories.
- Right click Format Data Series
- Plot series on Secondary Axis

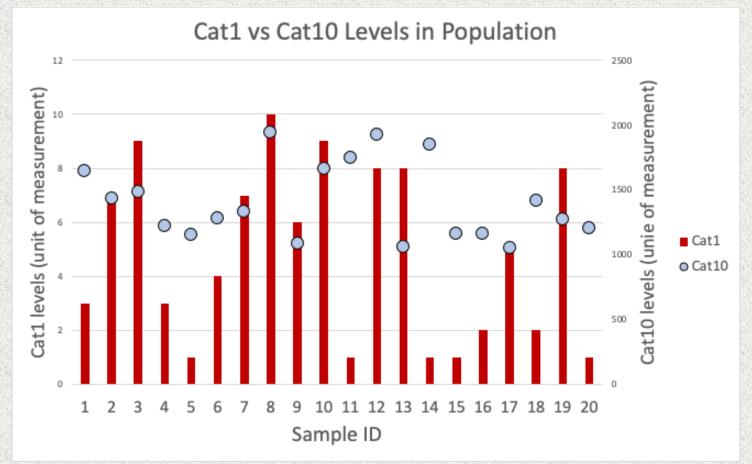






Adding a secondary axis

• Format further to tell the story effectively.







• Here, I have calculated the average values for each of the 12 categories in exposed females and unexposed females.

	A	В	с	D		E	F	G	н і	J
1	Is #8 temale?	YES								
2	Females	10								
3	Males	10								
4	Exposed	10								
5	Unexposed	10				А	В	С	D	E
6	Female+Exp	5		-				ç	0	-
7	Cat1 avg F+Exp	4.4				Sample_ID	1	2	3	4
8	Cat1 avg F+Unexp	2.8			2	Sex	F	F	м	F
9	Difference: Exp-Unexp	1.6			3	Status	N	Y	Y	N
10	Cat7: individual 8	0.03804672			4	Cat1	3	7	9	3
11	Cat7: individual 8 (v2)	0.03804672			5 (Cat2	0.39980403	0.131167	0.83797435	0.99121998
12					6 0	Cat3	0.01765164	0.74075219	0.4243295	0.84309817
13					7 0	Cat4	76	75	65	53
14				_	-	Cat5	3			
15			Cat avg F+Une	exp	-	Cat6	0.27817695		-	-
16		4.4	2.8		-					
17	Cat2		0.45858686			Cat7	0.47111146			
18	Cat3		0.54835892		_	Cat8	59			
19	Cat4	80.2	72.6	1	-	Cat9	0.15067873			
20	Cat5	6	4.2	1	13 (Cat10	1646	1436	1486	5 1218
21	Cat6		0.47628912	1	14	Cat11	535	586	543	3 470
22	Cat7		0.47810595	1	15	Cat12	13	19	19	5 18
23	Cat8	74	76.4	- 1	16					
24	Cat9	0.43398075								
25	Cat10	1542.2	1401.2		_					
26	Cat11	496.4	482.6		_					
27	Cat12	15	15.8							

Notice where the \$ symbols are in the formula...

\$ indicates what you do NOT
want to change when you "drag
and fill"

For example: column identity.





Stacked bar charts: Drag and fill

• "Drag and fill"

- Move your mouse cursor over lower right corner until it becomes "+",
- Hold down the right mouse button
- drag and fill in the same formula for multiple cells located immediately below (or to the right).

	A	В	с	D	E	F	G	н	1	J
1	Is #8 female?	YES								
2	Females	10								
3	Males	10								
4	Exposed	10								
5	Unexposed	10								
6	Female+Exp	5								
7	Cat1 avg F+Exp	4.4								
8	Cat1 avg F+Unexp	2.8								
9	Difference: Exp-Unexp	1.6								
10	Cat7: individual 8	0.03804672			_					
11	Cat7: individual 8 (v2)	0.03804672								
12										
13										
14										
15		Cat avg F+Ex	Cat avg F+Une	хр						
16	Cat1	4.4	2.8						Т	
17	Cat2	0.45232322	0.45858686							
18	Cat3	0.59839193	0.54835892							
19	Cat4	80.2	72.6							
20	Cat5	6	4.2							
	Cat6	0.43008582	0.47628912							
21	Cat7	0.54180968	0.47810595							
	0.10	74	76.4							
22	Cat8									
22 23		0.43398075	0.45870727							
21 22 23 24 25	Cat9	0.43398075	0.45870727 1401.2							
22 23 24	Cat9 Cat10									



QUESTION 4

• I used AVERAGEIFS to calculate the Cat1 values, and then dragged the "fill handle" (lower right corner) to fill in the formula for all the other categories. Which of these formulas should I see for the selected cell?

A.=AVERAGEIFS(B4:U4,sample_values!\$B2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

B.=AVERAGEIFS(sample_values!B4:U4, sample_values!\$B2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

C.=AVERAGEIFS(sample_values!\$B11:\$U11,sample_values!\$ B\$2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

D.=AVERAGEIFS(sample_values!B\$11:U\$11,sample_values!\$ B\$2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

	A	В	С	D		E	te
1	Sample_ID	1	2	3		4	!
2	Sex	F	F	м		F	1
3	Status	N	Y	Y		N	1
4	Cat1	3	7		9	3	
5	Cat2	0.39980403	0.131167	0.8379		0.99121998	
6	Cat3	0.01765164	0.74075219	0.424	3295	0.84309817	
7	Cat4	76	75		65	53	
8	Cat5	3	10		3	1	-98
9	Cat6	0.27817695	0.07110137	0.3774			
10	Cat7	0.47111146	0.89977883	0.4250		0.09046157	
11	Cat8	59	79		81	65	
12	Cat9	0.15067873	0.69674031	0.260		0.54603649	
13	Cat10	1646	1436		1486	1218	
14	Cat11	535	586		543	470	
15	Cat12	13	19		15	18	
16					2020		
	eample values	Shoot1					
<u>.</u>	sample_values	Sheet1					
	sample_values	A Sheet1	В			С	
 15	sample_values		B Cat avg		Cat	C avg F+Une:	хр
15 16	cat1				Cat	avg F+Une	xp 2.8
			Cat avg	F+Exp	Cat	avg F+Une	2.8
16	Cat1		Cat avg 0.4523	F+Exp 4.4	Cat	avg F+Une	2.8 355
16 17	Cat1 Cat2		Cat avg 0.4523	F+Exp 4.4 323217	Cat	avg F+Une 0.4585868 0.5483589	2.8 355
16 17 18	Cat1 Cat2 Cat3		Cat avg 0.4523	F+Exp 4.4 323217 391934	Cat	avg F+Une: 0.4585868 0.5483589 7	2.8 355 917
16 17 18 19	Cat1 Cat2 Cat3 Cat4		Cat avg 0.4523 0.5983	F+Exp 4.4 323217 391934 80.2	Cat	avg F+Une: 0.4585868 0.5483589 7	2.8 355 917 2.6 4.2
16 17 18 19 20	Cat1 Cat2 Cat3 Cat4 Cat5		Cat avg 0.4523 0.5983	F+Exp 4.4 323217 391934 80.2 6	Cat	avg F+Une 0.4585868 0.5483589 7	2.8 355 917 2.6 4.2 122
16 17 18 19 20 21	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6		Cat avg 0.4523 0.5983	F+Exp 4.4 323217 391934 80.2 6 085816	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762891 0.4781059	2.8 355 917 2.6 4.2 122
16 17 18 19 20 21 22	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6 Cat7		Cat avg 0.4523 0.5983 0.4300 0.5418	F+Exp 4.4 323217 391934 80.2 6 085816 309681	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762891 0.4781059	2.8 355 917 2.6 4.2 122 953 6.4
16 17 18 19 20 21 22 23	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6 Cat7 Cat8		Cat avg 0.4523 0.5983 0.4300 0.5418 0.4339	F+Exp 4.4 323217 391934 80.2 6 085816 309681 74	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762893 0.4781059 7	2.8 355 917 2.6 4.2 122 953 6.4 272
16 17 18 19 20 21 22 23 24	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6 Cat7 Cat8 Cat9		Cat avg 0.4523 0.5983 0.4300 0.5418 0.4339	F+Exp 4.4 323217 391934 80.2 6 085816 309681 74 980755	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762893 0.4781059 7 0.4587072 140	2.8 355 917 2.6 4.2 122 953 6.4 272

128



QUESTION 4

• I used AVERAGEIFS to calculate the Cat1 values, and then dragged the "fill handle" (lower right corner) to fill in the formula for all the other categories. Which of these formulas should I see for the selected cell?

A.=AVERAGEIFS(**B4:U4**,sample_values!\$B2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

B.=AVERAGEIFS(sample_values!B4:U4, sample_values!\$B2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

C.=AVERAGEIFS(sample_values!\$B11:\$U11,sample_values!\$ B\$2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

D.=AVERAGEIFS(sample_values!B\$11:U\$11,sample_values!\$ B\$2:\$U\$2,"F",sample_values!\$B\$3:\$U\$3,"Y")

	A	В	С	D		E	t
1	Sample_ID	1	2	3		4	:
2	Sex	F	F	м		F	1
3	Status	N	Y	Y		N	1
4	Cat1	3	7		9	3	
5	Cat2	0.39980403	0.131167	0.8379	7435	0.99121998	
6	Cat3	0.01765164	0.74075219	0.4243	3295	0.84309817	
7	Cat4	76	75		65	53	
8	Cat5	3	10		3	1	
9	Cat6	0.27817695	0.07110137	0.3774			
10	Cat7	0.47111146	0.89977883	0.4250	4482	0.09046157	
11	Cat8	59	79		81	65	_83
12	Cat9	0.15067873	0.69674031	0.260		0.54603649	
13	Cat10	1646	1436	:	1486	1218	
14	Cat11	535	586		543	470	_83
15	Cat12	13	19		15	18	
16							
					201810.0		
88 <u>.</u>	sample_values	Sheet1					
	sample_values	Sheet1	В			с	
	sample_values						
15			B Cat avg	F+Exp	Cat	avg F+Unex	•
1. 15 16					Cat	avg F+Unex	•
			Cat avg	F+Exp	Cat	avg F+Unex	2.8
16	Cat1		Cat avg	F+Exp 4.4	Cat	avg F+Une	2.8 355
16 17	Cat1 Cat2 Cat3		Cat avg	F+Exp 4.4 323217	Cat	avg F+Unex 0.4585868 0.5483589	2.8 355 917
16 17 18	Cat1 Cat2 Cat3 Cat4		Cat avg	F+Exp 4.4 323217 391934	Cat	avg F+Unex 0.4585868 0.5483589 7	2.8 355 917 2.6
16 17 18 19	Cat1 Cat2 Cat3 Cat4		Cat avg 0.4523 0.5983	F+Exp 4.4 323217 391934 80.2	Cat	avg F+Unex 0.4585868 0.5483589 7	2.8 355 917 2.6 4.2
16 17 18 19 20	Cat1 Cat2 Cat3 Cat4 Cat5		Cat avg 0.4523 0.5983	F+Exp 4.4 323217 391934 80.2 6	Cat	avg F+Unex 0.4585868 0.5483589 7	2.8 355 917 2.6 4.2
16 17 18 19 20 21	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6		Cat avg 0.4523 0.5983	F+Exp 4.4 323217 391934 80.2 6 085816	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762891 0.4781059	2.8 355 917 2.6 4.2 122
16 17 18 19 20 21 22	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6 Cat7 Cat8		Cat avg 0.4523 0.5983 0.4300 0.5418	F+Exp 4.4 323217 391934 80.2 6 085816 309681	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762891 0.4781059	2.8 355 917 2.6 4.2 122 953 6.4
16 17 18 19 20 21 22 23	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6 Cat7 Cat8		Cat avg 0.4523 0.5983 0.4300 0.5418 0.4339	F+Exp 4.4 323217 391934 80.2 6 305816 309681 74	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762891 0.4781059 7	2.8 355 917 2.6 4.2 122 953 6.4 272
16 17 18 19 20 21 22 23 24	Cat1 Cat2 Cat3 Cat4 Cat5 Cat6 Cat7 Cat8 Cat9		Cat avg 0.4523 0.5983 0.4300 0.5418 0.4339	F+Exp 4.4 323217 391934 80.2 6 085816 309681 74 980755	Cat	avg F+Unex 0.4585868 0.5483589 7 0.4762891 0.4781059 7 0.4587072 140	2.8 355 917 2.6 4.2 122 953 6.4 272

129

sample values





• Select data \rightarrow Insert \rightarrow 2-D Column \rightarrow Stacked Column

vo	ome Insert	d Table	Page Layo Pictures Sh	apes Icons	The Small Sm	Data Models T hartArt T reenshot T	ᇬ Му	View t Add-ins Add-ins *	Acrobat	h	Recommended Charts	2-D Column	• 🕅 • 📲		
316	• × <	$f_x = A$	VERAGEIFS(sample_val	ues!\$B4:\$	U4,sampl	e_values!\$	B\$2:\$U\$2	,"F",sample_valu	es!\$B\$3:	\$U\$3,"Y")				
	A	В	С	D	E	F	G	1	4 I	J	к				
. Is	#8 female?	YES											Stacked	Column	
F	emales	10													
	1ales -	10										3-D Column		-	
	xposed	10				_						5-D Column			
	nexposed	10				_									
	emale+Exp	5													
_	at1 avg F+Exp	4.4													11
	at1 avg F+Unexp	2.8													
	ifference: Exp-Unexp	1.6													
-	at7: individual 8	0.03804672													
	at7: individual 8 (v2)	0.03804672										2-D Bar			
2												2 0 001			
3												-			
.4															
5		Cat1 avg F+E		nexp											
6 C		4.4													
	at2	0.45232322													
	at3	0.59839193										-			
	at4	80.2	72.6									3-D Bar			
	at5 at6	6										o o o ou			
1 C		0.43008582 0.54180968										-			
3 0		0.54180968											-		
4 C		0.43398075													
	at9 at10	1542.2	1401.2												
	at10	496.4	482.6												
	at11 at12	496.4													
28	d112	15	15.8												





- The data does not fit well using the same y-axis.
- Can transform the data in some way, or add a secondary axis for 1 category, or...

				h													_
	Cat1 avg F+E	Cat1 avg F+L	Jnexp	Т					ch	art T	itla						
Cat1	4.4	2.8	Ī						CII	Iditi	nue						
Cat2	0.45232322	0.45858686		3500													_
Cat3	0.59839193	0.54835892		3000													_
Cat4	80.2	72.6		2500													
Cat5	6	4.2		2000													
Cat6	0.43008582	0.47628912		44													_
Cat7	0.54180968	0.47810595		1500													-
Cat8	74	76.4		1000													
Cat9	0.43398075	0.45870727		500										_	_		_
Cat10	1542.2	1401.2		0													_
Cat11	496.4	482.6			1	2	3	4	5	6	7	8	9	10	11	12	
Cat12	15	15.8							Ser	ies1 📕	Series	2					
				₿ —						=0=						131	_



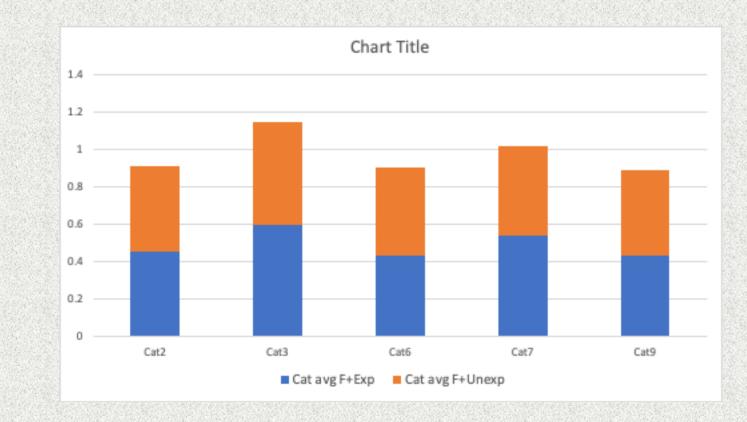
- Much better!
- Next, update the Series labels by clicking in the Name area...

								TEXA: Sup	erfu		esea	arc	h Cei
		Cat1 avg F+	E Cat1 ave	ξ F+Un	exp								1
Cat1		4.4		2.8		_		Chart Ti	tle				
Cat2		0.45232322	-			1.4 -							
						_							-232
Cat3		0.59839193				1.2							
Cat4		80.2	2	72.6		1 -							
Cat5		6	6	4.2		0.8							1.5
Cat6		0.43008582	2 0.47628	3912									
Cat7		0.54180968	8 0.47810)595		0.6							
Cat8		74		76.4		0.4							22
Cat9		0.43398075	-										
						0.2							
Cat10		1542.2		01.2		0							
Cat11		496.4	4 4	82.6		_	Cat2 Cat3	Cate	5	Cat7	Ca	t9	22
Cat12		19	5	15.8				Series1	Series2				
nent Layout	Colors								Data Source				Ro
ent Layout	Colors				5		Range Details	heet2!\$A\$1		100t21\$A\$	21.\$C\$22	2 Sheet	Ro
A A A A A A A A A A A A A A A A A A A	Colors fx B	c	D	E	F	G	-	heet2!\$A\$1		neet2!\$A\$	21:\$C\$22	2,Sheet	Ro
A #8 female?	Colors				F	G	-	heet2!\$A\$1		neet2!\$A\$	21:\$C\$22	2,Sheet	Ro
A #8 female? males ales	fx B YES 10 10				F	G	-	heet2!\$A\$1		neet2!\$A\$	21:\$C\$22	2,Sheet	Ro
A A A A A A A A A A A A A A A A A A A	Colors fx B B YES 10 10 10				F	G	-	heet2!\$A\$1		neet2!\$A\$	21:\$C\$22	2,Sheet	Ro
A #8 female? mmales ales posed hexposed	fx B YES 10 10 10 10 10				F	G	-	heet2!\$A\$1		neet2!\$A\$	21:\$C\$22	2,Sheet	Ro
A #8 female? males ales posed nexposed male+Exp	Colors fx B B YES 10 10 10				F	G	Chart data range: =St Legend entries (Series):		7:\$C\$18,SH	_	21:\$C\$22	2,Sheet	
A #8 female? males ales posed nexposed male+Exp t1 avg F+Exp t1 avg F+Unexp	fx B YES 10 10 10 10 10 10 20 10 20				F	G	Chart data range: =SH Legend entries (Series): Series1			_	21:\$C\$22	2,Sheet	
A #8 female? males ales posed nexposed male+Exp tt1 avg F+Exp tt1 avg F+Unexp fference: Exp-Unexp	fx B YES 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10				F	6	Chart data range: =St Legend entries (Series):		7:\$C\$18,SH	_	21:\$C\$22	2,Sheet	
A #8 female? males ales posed exposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp ffrerence: Exp-Unexp t7: individual 8	fx B YES 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 0.03804672 10				F	<u>G</u>	Chart data range: =SH Legend entries (Series): Series1		7:\$C\$18,SF				
A #8 female? males ales posed exposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp ffrerence: Exp-Unexp t7: individual 8	fx B YES 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10				F	<u>G</u>	Chart data range: =SH Legend entries (Series): Series1		7:\$C\$18,SF	_			
A #8 female? males ales posed exposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp ffrerence: Exp-Unexp t7: individual 8	fx B YES 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 0.03804672 10				F	6	Chart data range: =SH Legend entries (Series): Series1		7:\$C\$18,SF				
A #8 female? males ales posed exposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp ffrerence: Exp-Unexp t7: individual 8	fx B YES 10 10 10 10 10 10 5 4.4 2.8 1.6 0.03804672 0.03804672 0.03804672		D		F	6	Chart data range: =SH Legend entries (Series): Series1		7:\$C\$18,SF				
A #8 female? males ales posed male+Exp t1 avg F+Exp t1 avg F+Unexp fference: Exp-Unexp fference: Exp-Unexp t7: individual 8 t7: individual 8 (v2)	fx B YES 10 100 100 100 100 100 100 100 100 0.03804672 0.03804672 Cat1 avg F+E Cat1 avg F+E	Catl avg F+Unex	D		F	G G Chart 1	Chart data range: =SH Legend entries (Series): Series1	*	7:\$C\$18,SF				
A ##8 female? males ales posed nexposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp fference: Exp-Unexp t7: individual 8 (v2)	fx B YES 10 10 10 10 10 10 5 4.4 2.8 1.6 0.03804672 0.03804672 0.03804672 Catl avg F+E 4.4 0.45232322 4.4	Catl avg F+Unex 2.8 0.45858686	D		F	0	Chart data range: =St Legend entries (Series): Series1 Series2 + - Switch Row/C	Column	7:\$C\$18,SF Name Y values	:: =(Shee	t2!\$B\$17	:\$B\$18	
A #8 female? males ales posed texposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp fference: Exp-Unexp t7: individual 8 (v2) t1 t1 t2 t3	fx B YES 10 10 10 10 5 4.4 2.8 1.6 0.03804672 0.03804672 0.03804672 Catl avg F+E 4.4 0.452322 0.59839193	Cat1 avg F+Unex 2.8 0.45858686 0.54835892	D		F	0	Chart data range: =St Legend entries (Series): Series1 Series2	Column	7:\$C\$18,SF Name Y values	:: =(Shee	t2!\$B\$17	:\$B\$18	
A #8 female? males ales posed exposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp fference: Exp-Unexp fference: Exp-Unexp t7: individual 8 t7: individual 8 t1 t2 t3 t4	fx B fx 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 5 4.4 0.3804672 0.03804672 Cat1 avg F+E 4.4 0.4522322 0.59839193 80.2 10	Cat1 avg F+Unex 2.8 0.45858686 0.54835892 72.6	D		F	0	Chart data range: =St Legend entries (Series): Series1 Series2 + - Switch Row/C	Column	7:\$C\$18,SF Name Y values	:: =(Shee	t2!\$B\$17	:\$B\$18	
A ##8 female? males ales posed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp fference: Exp-Unexp fference: Exp-Unexp t1 avg F+Unexp t1 avg F+Unexp t1 avg F+Unexp t1 avg F+Unexp t1 avg F+Unexp t1 avg F+Unexp t2 individual 8 (v2)	fx B fx 10 10 10 10 10 10 10 10 10 0.03804672 0.03804672 Catl avg F+E 4.4 0.45232322 0.59839193 80.2 6	Catl avg F+Unex 2.8 0.45858686 0.54835892 72.6	P 14 12 1 08		F	0	Chart data range: =St Legend entries (Series): Series1 Series2 + - Switch Row/C	Column	7:\$C\$18,SF Name Y values	:: =(Shee	t2!\$B\$17	:\$B\$18	
A #8 female? males ales posed nexposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Unexp fference: Exp-Unexp t7: individual 8 (v2) t1 t1 t2 t3 t4 t5 t6	fx B fx 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 0.03804672 0.03804672 Catl avg F+E 4.4 0.45232322 0.59839193 80.2 6 0.43008582 6	Cat1 avg F+Unex 2.8 0.45858686 0.54835892 72.6 4.2	D		F	0	Chart data range: =SH Legend entries (Series): Series1 Series2 + - Switch Row/C Horizontal (Category) axis Hidden and Empty Cells	Column labels: =(7:\$C\$18,SF Name Y values Sheet2!\$A	:: =(Shee	t2!\$B\$17	:\$B\$18	
A #8 female? males ales ales posed exposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Exp t2 individual 8 t7: individual 8 t7: individual 8 t7: t1 t1 t2 t3 t4 t5 t6 t7 t8	fx B fx B YES 10 100 100 100	Catl avg F+Unex 2.8 0.45858686 0.54835892 72.6 4.2 0.47628912 0.47810595 76.4	D		F	0	Chart data range: =St Legend entries (Series): Series1 Series2 + - Switch Row/O Horizontal (Category) axis	Column labels: =(7:\$C\$18,SF Name Y values	:: =(Shee	t2!\$B\$17	:\$B\$18	
A #8 female? males ales sposed emale+Exp tt1 avg F+Exp tt1 avg F+Exp tt2 individual 8 tt7: individual 8 (v2) ett1 tt2 tt2 tt3 tt4 tt5 tt6 tt7 tt8 tt9	fx B fx B YES 10 100 100 100	Cat1 avg F+Unex 2.8 0.45858686 0.5485892 72.6 4.2 0.47628912 0.47628912 0.47810595 76.4 0.45870727	D D 1.4 1.2 1 0.8 0.6 0.4 0.2		F	0	Chart data range: =SH Legend entries (Series): Series1 Series2 + - Switch Row/C Horizontal (Category) axis Hidden and Empty Cells	Column labels: =(7:\$C\$18,SF Name Y values Sheet2!\$A	:: =(Shee	t2!\$B\$17	:\$B\$18	
A #8 female? males ales posed exposed male+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Exp t1 avg F+Comexp ffreence: Exp-Unexp ffreence: Exp-Unexp t7: individual 8 t7: individual 8 t8	fx B fx B YES 10 100 100 100	Cat1 avg F+Unex 2.8 0.45858686 0.54835892 72.6 4.2 0.47628912 0.47810595 76.4 0.45870727 1401.2	D		F	0	Chart data range: =St Legend entries (Series): Series1 Series2 + - Switch Row/C Horizontal (Category) axis Hidden and Empty Cells Show empty cells as: G	Column labels: =(7:\$C\$18,SF Name Y values Sheet2!\$A	:: =(Shee \$17:\$A\$18	t2!\$B\$17	:\$B\$18	





 Now we can show how each category was affected by exposure in females.



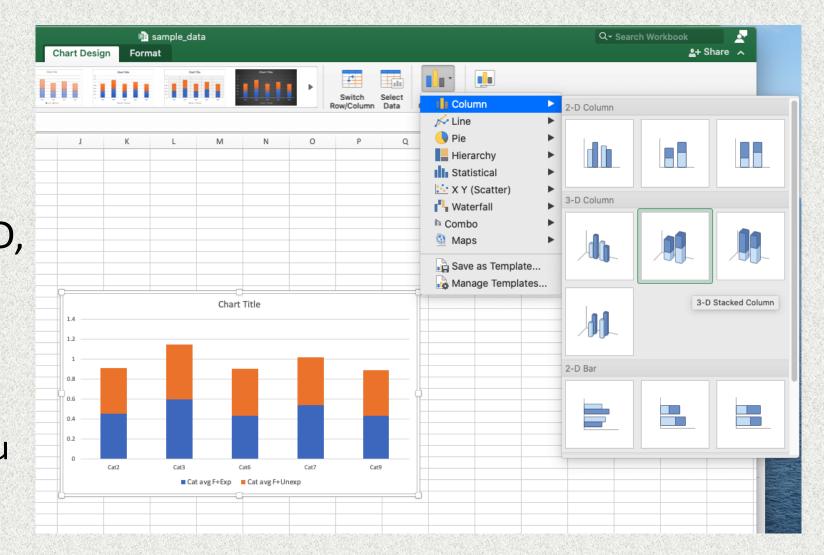


3-D graphs

BIG DATA

IN ENVIRONMENTAL SCIENCE AND TOXICOLOGY

- Excel offers 3-D options for different types of charts.
- You can start off with 3-D, or you can try it on an existing chart
 - Select the chart
 - Change Chart Type
 - Select the 3-D option you want

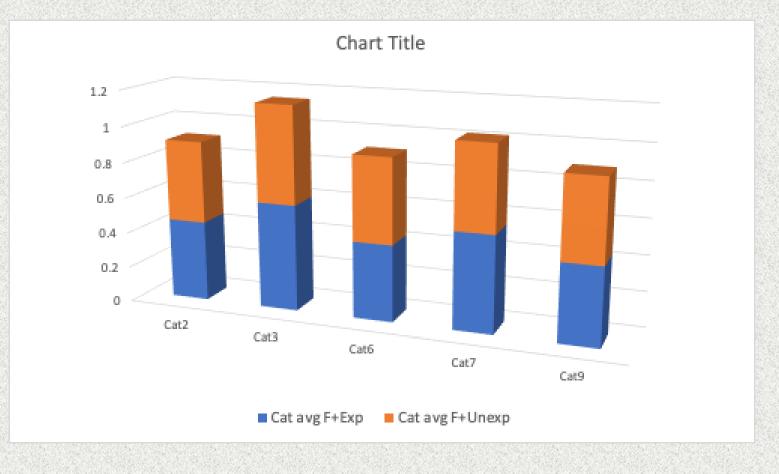






3-D graphs

- You can change colors, fonts, labels, etc. as before.
- Can't add a secondary axis.







Box and whisker plots

- Select the data you want to plot.
 - Here, I've chosen data that will produce histograms in the same scale range.

Sample_ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Sex	F	F	м	F	м	м	м	F	F	м	F	м	м	F	м	F	м	F	м	F
Status	N	Y	Y	N	N	N	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	N
Cat1	3	7	9	3	1	4	7	10	6	9	1	8	8	1	1	2	5	2	8	1
Cat2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.22800014	0.12970948	0.20895121	0.09570753	0.57328551	0.74719656	0.97775537	0.90570917	0.23738616	0.63079488	0.03630628
Cat3	0.01765164	0.74075219	0.4243295	0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.6285995	0.66614321	0.15886006	0.29619534	0.522676	0.49873951	0.499717	0.57727393	0.35131784	0.24743548	0.36314088
Cat4	76	75	65	53	89	66	88	78	98	67	75	96	90	65	70	89	73	84	85	71
Cat5	3	10	3	1	7	5	5	3	3	3	8	6	7	5	8	7	4	2	9	9
Cat6	0.27817695	0.07110137	0.37747939	0.13183529	0.31935773	0.1925463	0.63984189	0.64802318	0.26027957	0.93419475	0.52918677	0.2509355	0.39102529	0.77463581	0.59241452	0.48270423	0.83625265	0.41941355	0.83425559	0.93651799
Cat7	0.47111146	0.89977883	0.42504482	0.09046157	0.43938449	0.85181114	0.7906725	0.03804672	0.98191601	0.06929245	0.17350784	0.00365075	0.62290958	0.10161224	0.129711	0.87430617	0.63898328	0.72340885	0.87721221	0.74542847
Cat8	59	79	81	65	54	91	50	92	96	51	80	63	78	70	91	60	80	59	81	92
Cat9	0.15067873	0.69674031	0.2607653	0.54603649	0.94019016	0.63397075	0.31429126	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.47184386	0.97195415	0.10610593	0.55886697	0.34184183	0.38806705	0.24976963
Cat10	1646	1436	1486	1218	1151	1278	1329	1942	1085	1662	1748	1922	1061	1850	1160	1166	1049	1419	1275	1207
Cat11	535	586	543	470	479	502	582	537	412	514	462	487	543	405	504	478	466	419	467	591
Cat12	13	19	15	18	11	19	18	10	16	11	15	19	20	18	13	17	16	14	14	14
																				WWW.WARDOWN.COM COM

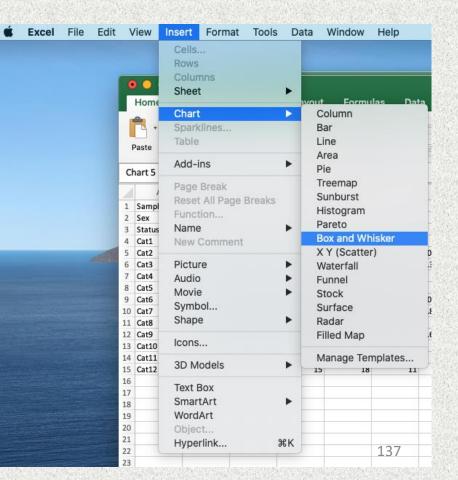




Histograms: Box and Whisker plot

• As before: Insert \rightarrow your chart of choice (Box and Whisker plot here).

Но	ome	nsert	Draw P	age Layout	Formu	ılas Dat	ta Revie	ew Viev	v Acrol	pat							
ivotT	🗲 Table Rec	ommended votTables	Table Pi	ctures Shap	• G	C 3D Mod	rt 🔻 🛔	Get Add-ir My Add-in	ns 🗾 Bir	sio Data Visua ng Maps ople Graph		ecommended Charts	∎∎* ,‰* ● *	Histogram	∎• <u>()</u> 		
12	*	\times \checkmark	fx Cat9														
	A	В	с	D	E	F	G	н	1	J	к	L	M			L	
Sar	mple_ID	1	2	3	4	5	6	7	8	9	10	11	12			2h	10
Sex	x	F	F	м	F	м	м	м	F	F	м	F	м				F
Sta	atus	N	Y	Y	N	N	N	Y	Y	N	Y	Y	N				Y
Cat	t1	3	7	9	3	1	4	7	10	6		9 1			_		1
Cat	t2	0.39980403	0.131167	0.83797435	0.99121998	0.37915433	0.4932649	0.31730995	0.78559809	0.29231847	0.2280001	4 0.12970948	0.208	Box and \	Vhisker		6 0
Cat	t3	0.01765164	0.74075219		0.84309817	0.41803328	0.36549674	0.08057165	0.73402944	0.9952279	0.628599	5 0.66614321	0.158				1
Cat		76			53	89	66	88	78	98	6						0
Cat		3	10	-	1	7	5	5	3	3		3 8		I I I	İ.		8
Cat		0.27817695		0.37747939								5 0.52918677		ļļļ	Ţ		2 0
Cat						0.43938449	0.85181114	0.7906725	0.03804672			5 0.17350784			- II.		.1 (
Cat Cat		59	79 0.69674031		65 0.54603649	54 0.94019016	91	50	92 0.1763019		0.010126						.5 0
	t10	1646	1436		1218	1151	1278	1329	1942	1085	166			1922	1061	1850	1160
Cat		535	586		470	479	502	582	537	412	51			487	543	405	504
	413	13	10		470	4/5	10	10		412	51			10	345	405	10

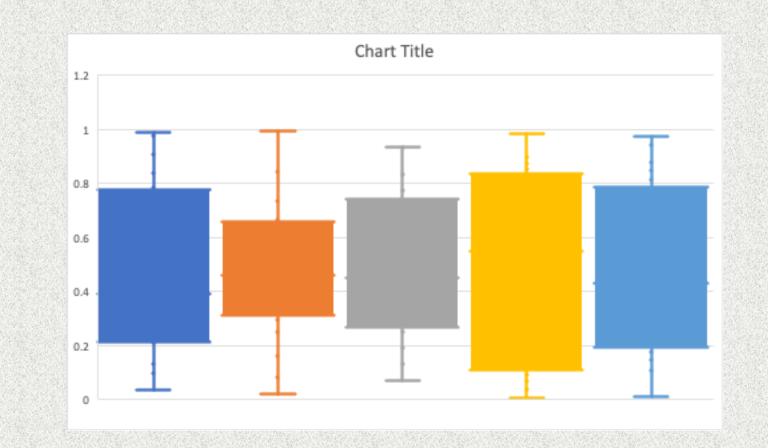






Histograms: Box and Whisker plot

• Caution: Nothing is labeled by default.







Histograms: Box and Whisker plot

Format Data Ser

Box and Whisker

🗸 Show inner poin

🗸 Show outlier poi

🗸 Show mean mai

Show mean line

Quartile Calculation

Options

- Format → Shape Outline to show individual data points and median line.
- We can change what points are shown by right clicking a plot and choosing Format Data Series.
 - We can only edit one plot at a time.
- Add Legends, Labels, Titles, etc. the same way as other charts:
 - Chart Design --> Add Chart Element



nlot	Acrob	Chat Cha	art Design	Format			
plot	Shape Fill	• 🖍 •	pe Outline	Α	A	Text F	
		J	к	1	М	N	0
				L		N	
	8 <u>-</u>	9 F	10	11 F	12	13	14 F
		F N	M Y	F Y	N	M Y	F N
	10	N 6		1		1 8	
	1 <u> </u>	0.29231847	-	0.12970948			
	0.73402944	0.9952279				0.29619534	0.5226
	78	98	67	75	96	90	
	3	3	3	8	6	7	
		0.26027957	-	0.52918677	0.2509355		0.774635
	0.03804672					0.62290958	
	92	96	51	80	63	78	
8	0.1763019	0.87520766	0.0101267	0.8489138	0.81448124	0.14730753	0.471843
-	1942	1085	1662	1748	1922	1061	18
	537	412	514	462	487	543	4
	10	16	11	15	19	20	
						Chart Title	
						Chart Title	
	L	1.2					
		1		•	••		
		0.8			Ŷ		
			1		•		
				1			
		0.6			<u>ه</u>		
					<u>Å</u>		
		0.4			•		
clusive median 🌲					2		
					ě.	1	
•					1		
•		0.2			9		
•		0.2			Į		





Miscellaneous tips and tricks

If I copy these two cells...

Na	me Box A	В
1	Is #8 female?	YES
2	Females	10
3	Males	10
4	Exposed	10
5	Unexposed	10
6	Female+Exp	5
7	Cat1 avg F+Exp	4.4
8	Cat1 avg F+Unexp	2.8
9	Difference: Exp-Unexp	1.6
10	Cat7: individual 8	0.03804672
11	Cat7: individual 8 (v2)	0.03804672
1.0		

And try to paste them directly below:

ERROR!

1	А	В
1	Is #8 female?	YES
2	Females	10
3	Males	10
4	Exposed	10
5	Unexposed	10
6	Female+Exp	5
7	Cat1 avg F+Exp	4.4
8	Cat1 avg F+Unexp	2.8
9	Difference: Exp-Unexp	1.6
10	Cat7: individual 8	0.03804672
11	Cat7: individual 8 (v2)	0.03804672
12		
13		
14		
15		
16	Cat1 avg F+Exp	#DIV/0!
17	Cat1 avg F+Unexp	#DIV/0!
18	1.00 - 1.000 m 20 20 1.000 M	

The "addresses" in the formulas of the cells update automatically.

Excel will either give an error message (if it can't calculate the formula because some value is missing), or you'll get a bogus result.





Miscellaneous tips and tricks

Solution: add \$ before any part you don't want to

change.

SU	JM 🌲	×	~	f_X	=AVE	RAGEIF	S(sample	_valu	es!B\$4:U\$	4,sample_v	alues!	B2:U2	,"F",san	nple	_values!B3	:U3,"Y")		
	A			В	AVE	RAGEIF	S(<u>averag</u>	e_ranç	i <u>e</u> , <u>criteria</u>	<u>range1, crite</u>	eria1, [c	riteria_	range2,	crite	eria2], [crite	ria_range3,	criteria3], .)
1	Is #8 female?	11755 (1741, 175	YI	ES	ara periorenta	auxonnoner ben	1210113412102007573	CENTRAL 20	RANDATINA DIGEN MARKAN SANJAR	ANOTH DOT WILL BALL BLAND	eter sen was a	al element the	NAME OF ALCOMPANY	CT DOWNER	THE REPORT OF THE PARTY			u economi fi

Another solution: if you don't need Excel to recalculate anything (for example, you're just copying and pasting some data in another place):

Copy the data, then right click where you want to paste it, and choose Paste Special \rightarrow Values

Cut Copy Paste	жх жс жv						
Paste Special	>	Paste	жv				
Smart Lookup Thesaurus	^て第L ^て第R	R Formulas & Number Formattin					
Insert Copied Cells. Delete Clear Contents		Keep Source Format No Borders Keep Source Column Transpose	-				
Filter	2	Values					
Insert Comment	Copy 第C Paste 第V Paste Special ・ Smart Lookup ^て第L Thesaurus ^て第L Clear Contents Filter ト Sort ・	Values & Number Formatting Values & Source Formatting					
Pick From Drop-dov Define Name	wn List	Paste Link Paste Picture					
	•	Paste Special 14	1 ^#V				





Miscellaneous tips and tricks

To change the direction of some list: **copy** the list (NOT cut)...

Then Paste Special --> Transpose

6 Cat3 0.01 7 Cat4	2	А	
3 Status N 4 Cat1 1 5 Cat2 0.39 6 Cat3 0.01 7 Cat4 1 8 Cat5 1 9 Cat6 0.27 10 Cat7 0.47 11 Cat8 1 12 Cat9 0.15 13 Cat10 1 14 Cat11 1	1	Sample_ID	1
4 Cat1 5 Cat2 0.39 6 Cat3 0.01 7 Cat4 0.39 8 Cat5 0.27 9 Cat6 0.27 10 Cat7 0.47 11 Cat8 0.15 12 Cat9 0.15 13 Cat10 1	2	Sex	F
5 Cat2 0.39 6 Cat3 0.01 7 Cat4	3	Status	N
6 Cat3 0.01 7 Cat4	4	Cat1	
7 Cat4 8 Cat5 9 Cat6 0.27 10 Cat7 0.47 11 Cat8 12 12 Cat9 0.15 13 Cat10 1	5	Cat2	0.39
8 Cat5 9 Cat6 0.27 10 Cat7 0.47 11 Cat8 12 12 Cat9 0.15 13 Cat10 14	6	Cat3	0.01
9 Cat6 0.27 10 Cat7 0.47 11 Cat8 11 12 Cat9 0.15 13 Cat10 11 14 Cat11 11	7	Cat4	
10 Cat7 0.47 11 Cat8 0.15 12 Cat9 0.15 13 Cat10 0.15 14 Cat11 0.15	8	Cat5	
11 Cat8 12 Cat9 0.15 13 Cat10 14 14 Cat11 14	9	Cat6	0.27
12 Cat9 0.15 13 Cat10 14 Cat11	10	Cat7	0.47
13 Cat10 14 Cat11	11	Cat8	
14 Cat11	12	Cat9	0.15
	13	Cat10	
	14	Cat11	-
15 Cat12	15	Cat12	
16	16		

Cut Copy Paste	米X 米C 米V							
Paste Special	►	Paste	۳۱)					
Smart Lookup. Thesaurus	^\\$L ^\\%R	S						
Insert Copied Delete Clear Contents		Keep Source Formatting No Borders Keep Source Column Widths Transpose Values Values & Number Formatting						
Clear Contents	>							
Filter Sort	*							
Insert Comme	nt	Values & Source Fo	ormatting					
Format Cells Pick From Drop Define Name Hyperlink	o-down List	Formatting Paste Link Paste Picture Paste Picture Link						
Services	•	Paste Special	^#V					



Cat1



Miscellaneous tips and tricks

To change the direction of some list: **copy** the list (NOT cut)...

Then Paste Special --> Transpose

and the second

	1	A Sample_ID	1			Cut Copy Paste Paste Special	¥X ¥C ¥V	-		π
	2	Sex Status	F Smart Lookup N Thesaurus				^ጊ ೫L ^ጊ ೫R			
	4 5 6	Cat2 Cat3	0.39 0.01			Insert Copied Ce Delete Clear Contents		Keep Se No Bord	ource Formatti ders ource Column '	ng
	7 8 9	Cat4 Cat5 Cat6 Cat7	0.27			Filter Sort Insert Comment	;	Values Values	& Number Forr & Source Form	
	11 Cat8 12 Cat9 0.15				Format Cells Pick From Drop- Define Name	Formatting Paste Link Paste Picture				
t2	Cat3	Cat4	Cat5	Cat6	Cat7	Cat8	Cat9	Cat10	Cat11	Cat12
	15		å					A Statistics	Ner ales S	143

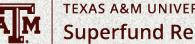




Miscellaneous tips and tricks...

 Excel can also calculate the minimum, maximum, and average values, along with standard deviation, for a data range: =MIN(data_range); =MAX(data_range); =AVERAGE(data_range); =STDEV(data_range)









Next up: Analyzing Big(ish) Data in Excel



Analyzing Big(ish) Data in Excel

Dillon Lloyd

North Carolina State University





Analysis ToolPak Installation

- The Analysis ToolPak is an extension of Excel that can be used to perform common statistical tests quickly and easily
- Instructions on installation for both Windows and Mac can be found at the link below
 - <u>https://support.microsoft.com/en-us/office/load-the-analysis-toolpak-in-excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4</u>

BIG DATA STIENCE AND TOXICOLOGY

Excel Options

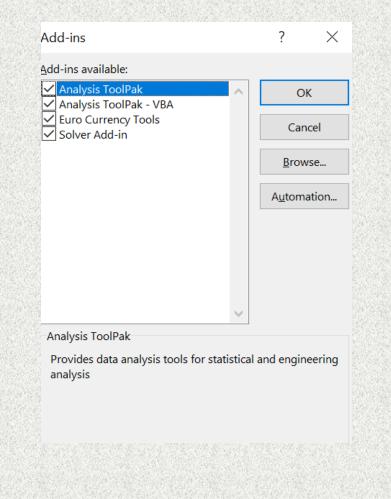


? ×

TEXAS A&M UNIVERSITY Superfund Research Center

a	Add-ins								
oofing	Name 🔺	Location	Туре						
ve	Active Application Add-ins								
	Analysis ToolPak	C:\Office16\Library\Analysis\ANALYS32.XLL	Excel Add-in						
iguage	Analysis ToolPak - VBA	C:\ice16\Library\Analysis\ATPVBAEN.XLAM	Excel Add-in						
se of Access	Euro Currency Tools	C:\root\Office16\Library\EUROTOOL.XLAM	Excel Add-in						
vanced	Solver Add-in	C:\Office16\Library\SOLVER\SOLVER.XLAM	Excel Add-in						
	Inactive Application Add-ins								
ustomize Ribbon	Date (XML)	C:\Microsoft Shared\Smart Tag\MOFL.DLL	Action						
uick Access Toolbar	Inquire	C:\soft Office\Office16\DCF\NativeShim.dll	COM Add-in						
	Microsoft Actions Pane 3		XML Expansion Pack						
ld-ins	Microsoft Data Streamer for Excel	C:\softDataStreamerforExcel.vsto vstolocal	COM Add-in						
ust Center	Microsoft Power Map for Excel	C:\ap Excel Add-in\EXCELPLUGINSHELL.DLL	COM Add-in						
	Microsoft Power Pivot for Excel	C:\I Add-in\PowerPivotExcelClientAddIn.dll	COM Add-in						
	Microsoft Power View for Excel	C:\el Add-in\AdHocReportingExcelClient.dll	COM Add-in						
	Document Related Add-ins								
	No Document Related Add-ins								
	Add-in: Analysis ToolPak								
	Publisher: Microsoft Office								
	Compatibility: No compatibility inform	ation available							
	Location: C:\Program Files\Micros	oft Office\root\Office16\Library\Analysis\ANALYS32	2.XLL						
	Description: Provides data analysis to	ools for statistical and engineering analysis							
	Description: Provides data analysis tools for statistical and engineering analysis								

superfund.tamu.edu





Home Insert Page Layout Formulas Data Review View Help



Data Analysis	? ×
<u>A</u> nalysis Tools	ОК
Anova: Single Factor Anova: Two-Factor With Replication Anova: Two-Factor Without Replication Correlation	Cancel
Covariance Descriptive Statistics Exponential Smoothing F-Test Two-Sample for Variances Fourier Analysis Histogram	Help





Histogram

- In the Data Ribbon click data analysis
- Scroll down and click on "Histogram"
- Select the data you want to see the distribution of
- Check "Chart Output" to get a plot in the output
- Click Ok and the histogram will show up in a new sheet





TEXAS A&M UNIVERSITY AM Superfund Research Center

alysis Tools lova: Single Factor lova: Two-Factor With Replication lova: Two-Factor Without Replication lova: Two-Factor Without Replication lovariance escriptive Statistics ponential Smoothing Test Two-Sample for Variances urier Analysis		Histogram		?	
nistogram		Input Input Range:	\$B\$8:\$AT\$8	ОК	
Data Analysis	? ×	<u>B</u> in Range:	1	Cancel	
Analysis Tools Anova: Single Factor Anova: Two-Factor With Replication Anova: Two-Factor Without Replication Correlation Covariance Descriptive Statistics Exponential Smoothing	OK Cancel <u>H</u> elp	 Labels Output options Output Range: New Worksheet <u>Ply:</u> New <u>W</u>orkbook P<u>a</u>reto (sorted histogram) Cumulative Percentage)	Help	
F-Test Two-Sample for Variances Fourier Analysis Histogram	A E 1 Bin Frequ		F G H		
	2 17373 3 34347.17 4 51321.33 5 68295.5 6 85269.67	1 18 18 5 0 2 2 15 10 5 0 0 15 10 5 0 0 0 0 0 0 0 0 0 0 0 0 0	Histogram		
	7 102243.8 8 More 9		³ 3 ³ 2 ^{1,1,} , ^{3,1,2,1,3,} , ^{3,1,2,1,3,} , ^{NOFE} Bin	Frequency	
	2 3			152	





T-Test

- Filter Data to the dose row and the row of your gene of interest
- Click Data Analysis on the data tab
- Click T-Test: Paired Two Sample for Means
- Under Variable 1 Range, select the columns in the row of your gene that correspond to the control dose
- Under Variable 2 Range, selects the columns in the row of your gene that corresponds to the max dose
- Click ok to view the results





T-Test

t-Test: Paired Two Sample for Me	ans	? ×	
Input			
Variable <u>1</u> Range:	\$B\$5:\$F\$5	ОК	
Variable <u>2</u> Range:	\$AP\$5:\$AT\$5	Cancel	
Hypothesized Mean Difference:		<u>H</u> elp	
<u>A</u> lpha: 0.05			
Output options			
Output Range:	1		
• New Worksheet <u>P</u> ly:			
O New <u>W</u> orkbook			

Control vs. Max Dose T-Test Results

4		_	
	A	В	C
1	t-Test: Paired Two Sample for Means		
2			
3		Variable 1	Variable 2
4	Mean	18	30
5	Variance	32.5	45.5
6	Observations	5	5
7	Pearson Correlation	0.461584247	
8	Hypothesized Mean Difference	0	
9	df	4	
10	t Stat	-4.115966043	
11	P(T<=t) one-tail	0.007330294	
12	t Critical one-tail	2.131846786	
13	P(T<=t) two-tail	0.014660589	
14	t Critical two-tail	2.776445105	
15		ייני באיני איז איז איז איז איז איז איז איז איז אי	





ANOVA

- Filter Data to the dose row and the row of your gene of interest
- Copy and Transpose Data and create columns for each different dose
 - ANOVA requires a bit more data manipulation due to the number of groups
 - Transpose data by highlighting the Dose and Gene of Interest Row, pressing ctrl + shift + right
 arrow key, and copying that highlighted data
 - Then go to Home \rightarrow Paste \rightarrow Paste Transpose in a new worksheet
- Click Data Analysis on the data tab
- Click on ANOVA
- Under Input Range, select the dose columns created in the previous step
- Group by columns and click ok to get the ANOVA output





TEXAS A&M UNIVERSITY Superfund Research Center

ANOVA

	В											N
	1USG0000000028				Dose0.03							
0	21		21									
0	10		10									
0	22		22	21			15			13		
0	23		23	13	25	23	10	15	21	34		
0	14		14	115	15	68	24	18	44	49	27	
0.01	39											
0.01	10											
0.01	21											
0.01	13											
0.01	115											
0.03	10											
0.03	38											
0.03	10											
0.03	25											
0.03	15											
0.1	13											
0.1	11											
0.1	9											
0.1	23											
0.1	68											
0.3	17											
0.3	17											
0.3	15											
0.3	10											
0.3	24											
	13											
1	38											
1												
1	13 15											
1	15											
1												
3	17											
3	66											
3	22											
3	21											
3	44											
10	41											
10	24											
10	13											
10	34											
10	49											
30	28											
30	26											
30	42											
30	27											
30	27											
		NAMES OF TAXABLE AND								STATISTICS OF THE OWNER		LO STOTEMENTS





TEXAS A&M UNIVERSITY Superfund Research Center

ANOVA

				D	C	D	L	Г	G	1
			1 Anova: Single Factor							
			2							
			3 SUMMARY							
			4 Groups	Count	Sum	Average	Variance			
nova: Single Factor		1 Anova: Single Factor Image: Single Fact								
Input			6 Column 2	5	198	39.6	1903.8			
	\$E\$2:\$M\$6	OK	7 Column 3	5	98	19.6	143.3			
ova: Single Factor out out Range: \$E\$2:\$M\$6 ouped By:		Cancel	8 Column 4	5	124	24.8	612.2			
nput Range: \$E\$2:\$M\$6		9 Column 5	5	83	16.6	25.3				
	O <u>R</u> ows	<u>H</u> elp	10 Column 6	5	97	19.4	112.3			
<u>L</u> abels in first row			11 Column 7	5	170	34	431.5			
Alphai 0.05			12 Column 8	5	161	32.2	199.7			
			13 Column 9	5	150	30	45.5			
	Factor ? X 4 Groups Count Sum Average M Factor ? X 5 Column 1 5 90 18 0 SES2:\$M\$56 OK Cancel 9 Column 3 5 98 19.6 © Columns Each Cancel 9 Column 4 5 124 24.8 24 9 Column 5 5 83 16.6 3 97 19.4 10 Column 7 5 170 34 12 Column 7 5 161 32.2 13 Column 9 5 150 30 14 12 13 14									
Output options			15	Count Sum $Average$ $Variance$ $Iaction (Count) Sum Average Variance Sount Sum Average Variance Iaction (Count) Iaction (Count) Sount Sum Average Variance Iaction (Count) Iaction (Count) Sount Sum Average Variance Iaction (Count) Iaction (Count) Sount Sount Sount Sount Sount Iaction (Count) Iaction (Count) Sount Sount Sount Sount Iaction (Count) Iaction (Count) Iaction (Count) Sount Sount Sount Sount Iaction (Count) Iaction (Count) Iaction (Count) Sount Sount Sount Iaction (Count) Iaction (Coun$						
Output Range:	Image: second									
			17 Source of Variation	SS	df	MS	F	P-value	F crit	
			18 Between Groups	2708.578	8	338.5722	0.8691	0.550983	2.208518	
New Workbook			19 Within Groups	14024.4	36	389.5667				
		2 SUMMARY Image: Sum of the second of t								
			21 Total	16732.98	44					
			22							
			23							
			24							





Regression

- Filter the data to dose and gene of interest
- Copy the dose and gene row by highlighting the first two cells and clicking ctrl+shift+right arrow
- Create a new worksheet and transpose the data In the data analysis tab select regression
- Choose the Y variable as the gene expression value, and the X variable as the dose
- Do this by highlighting the numeric value under the header row and clicking ctrl+shift+down arrow to select just the numeric cells
- Click ok to review results





Regression

		A	В	С	D	E	F	G	Н	
gression	? ×	1 SUMMARY	OUTPUT							
put	ОК	3 Regression	Statistics							
nput <u>Y</u> Range: \$B\$2:\$B\$47		4 Multiple R								
pput <u>X</u> Range: \$A\$2:\$A\$47	Cancel	5 R Square								
	Help	6 Adjusted F	-0.01009							
Labels Constant is Zero	Terb	7 Standard I	19.48447							
Confidence Level: 95 %		8 Observation	46							
put options		9								
) Output Range:		10 ANOVA								
New Worksheet <u>Ply</u> :		11	df	SS	MS		gnificance	F		
) New <u>W</u> orkbook		12 Regression	1	208.9637	208.9637	0.550419	0.462089			
esiduals		13 Residual	44	16704.36	379.6446					
Residuals Residual Plots		14 Total	45	16913.33						
ormal Probability		15 16	oefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
<u>N</u> ormal Probability Plots		17 Intercept								
		18 X Variable								
		19								
		20								



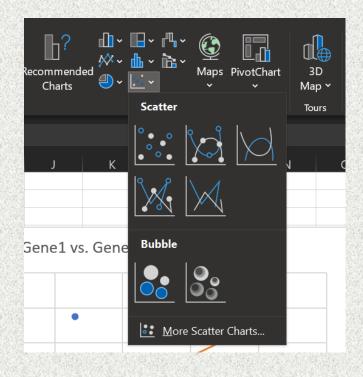


Scatter Plot

- Select Two Genes of interest and the Dose column through the filtering step and transpose them into a new worksheet
- Select the two gene columns
- Click Insert Scatter
- The Scatter Plot will then appear
- The add a trend line, right click on a point and click add trend line
- The trend line can then be edited to make it more clear
- These steps can be repeated to make a scatter plot of Dose vs. Gene Expression to visualize the regression analysis



Scatter Plot



A	В	С	D	E	F	G	Н	<u> </u>		K	_	L M	N	V 0	P
	ENSMUSG0000000028	ENSMUSG0000000088													
0	21	1911													
0	10	1770													
0	22	1547								_	_				
0	23	1931							Gene1	vs. Ger	ie2				
0	14	2413			8000										
0.01	39	2998													
0.01	10	2190			7000					_					
0.01	21	1912								•					
0.01	13	1614			6000										
0.01	115									•					
0.03	10				5000								•		
0.03	38										T				
0.03	10				4000										
0.03	25														
0.03	15				3000										
0.1	13						· ·		•						
0.1	11	1651			2000			-							
0.1	9					-									
0.1	23				1000	•	•								
0.1	68														
0.3	17	1549			0										
0.3	17	2341				0	20	40	60		80	100	120	140	
0.3	15														
0.3	10														
0.3	24														
1	13							~	ono Eve	roccia	« Dee-				
1	38							G	ene Exp	ression	Dose				
1	13				140										
1	15														
1	18				120										
3	17 66	1106 6746				•									
3					100										
3	22	3177 1629			100										
3	44	2907													
10	44	2907			80		_								
10	24	1385				• •									
10	13	1385			60	-					_				
10	34														
10	49				40	•									
30	28				40	• •		•							
30	28												•		
30	42				20			Ī							
30	27					5 •		•							
30	27	1539			0										
50	27	1555				0	5	10	15		20	25	30	35	
CALIFICATION OF COMPANY		CONTRACTOR A PROPAGATION OF	CTELESCORT SPACE			CT. CONTRACTOR		SURPLUP AND		CONTRACTOR OF	00170300-20		CT. CONCINCTION		20167 (Delate

TEXAS A&M UNIVERSITY

Superfund Research Center





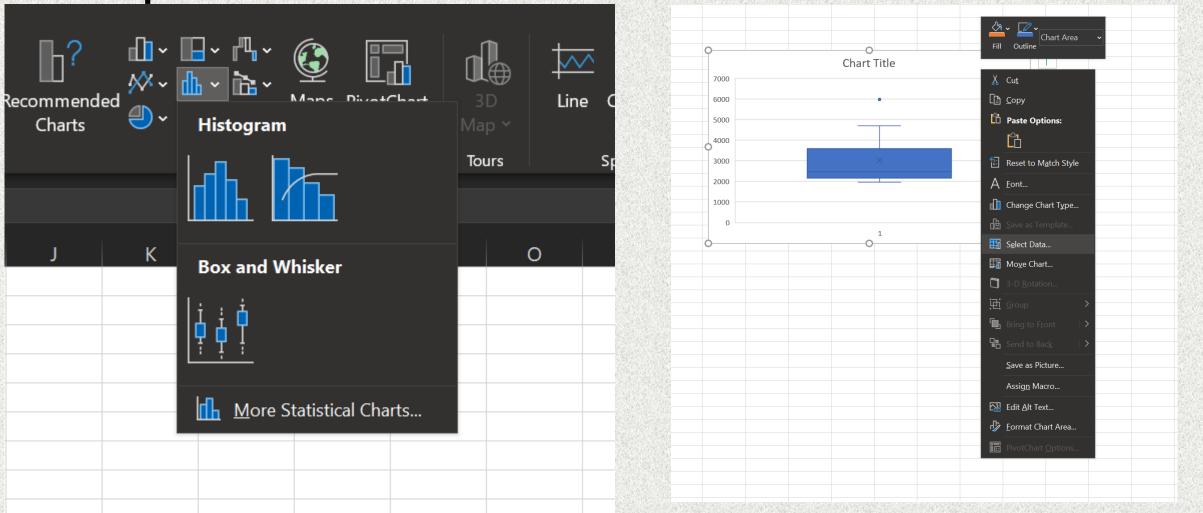
Boxplot

- Filter to Dose column or row and gene of interest
- Highlight gene column
- Click Insert \rightarrow Statistics Chart \rightarrow Box and Whisker Plot
- Right click and select Select Data
- Under series, click edit and select the data at Dose 1, then do the same for Dose 2, Dose 3 etc.
- Under Chart Design click Add Chart Element and add a legend to the plot to distinguish between doses
- Can then right click on the plot to edit other aesthetic properties





Boxplot





AM

TEXAS A&M UNIVERSITY Superfund Research Center

Boxplot ? \times Select Data Source 1 Chart data range: =Sheet17!\$B\$2:\$B\$12 Chart Title Switch Row/Column 000 Legend Entries (Series) Horizontal (Category) Axis Labels 000 <u>ia A</u>dd Edit 📅 Edi<u>t</u> X <u>R</u>emove 000 Series1 000 000 000 000 Hidden and Empty Cells OK Cancel 0 1 Δ В D G

1	CLASS:DOSE	ENSMUSG0000000001	ENSMUSG0000000049						
2	0	2289	30854						
3	0	2171	Edit S	eries			? ×		
4	0	2124	Curre	enes					
5	0	2538	Series	<u>n</u> ame:					
6	0	2397	Dose1			1 Sel	ect Range		and a second
7	0.01	4707	Series	values:					
8	0.01	2797	=Shee	t17!\$B\$2:\$B\$6	<u>↑</u> = 2	289, 2171, 21.			
9	0.01	3235						-	
10	0.01	1955				l	OK		
11	0.01	5983	92976)					100
12									
13									
14								164	
15								-0.	
2 Car	Contact for the Contact of the				sey contact y a real second	et estador		KORGE / EN LOUIS CONTRACTOR	0724





TEXAS A&M UNIVERSITY Superfund Research Center

Boxplot • Quick Add Chart Change Colors ~ Element ~ Layout ~ h A<u>x</u>es Boxplot *f*x Image: Image 140 ☐ <u>C</u>hart Title > В MUSG0000000001 ENSMUSG 120 Data Labels > 2289 Dose0 2171 100 <u>G</u>ridlines Dose0.01 124 Dose0.03 ∎ <u>L</u>egend 80 38 لىلىك اللە <u>N</u>one Dose0.1 0 97 6 Dose 0.3 60 0.01 07 Dose1 սե 🗉 <u>R</u>ight 0.01 97 8 Dose3 40 0.01 35 9 Dose10 55 0.01 Dose30 <u> T</u>op 83 20 0.01 Enh Left 0 13 1 14 վր 15 <u>B</u>ottom 16 17 18 19 165





R Introduction

- R is a statistical computing language that is very flexible and can do everything we have talked about up to this point and more!
- Can create publication ready plots fairly easily
- RStudio is the primer Interactive Development Environment and has a host of helpful features that makes coding in R as simple as possible
- Can download R and RStudio from the links below
 - R: <u>https://cran.r-project.org/</u>
 - RStudio: https://www.rstudio.com/products/rstudio/





Importing Excel Data into R

- Make sure column names are consistent across files
 - Column names should not have special characters or spaces
 - Capitalization needs to be consistent, case matters
- Know what kind of file you are reading in
 - Are you reading in a csv file? Or a .xlsx file that contains multiple pages?
- Know where your data is located
 - What folder? Is it on a google drive or some other network drive?
- Have a plan for your data
 - Statistical Tests? Plots? Have a plan before you start





Importing .csv and .xlsx files into R

- First, locate the file that you want to import on your computer
- Next, open RStudio and click file \rightarrow New File \rightarrow R Script
- Then, type the code in the next slide
- Make sure to replace the file path in the green quotations with the location of the example file on your computer
- Highlight the code, and press run to have R read in the files
- You can view the files in the viewing pane or by printing the dataset
- Now, they will be saved into your R workspace and can be used for analysis





169

Importing .csv and .xlsx files into R

RStudio						- 0 ×
	t Code View Plots Session Build Debug Profile Tools Help					
*1 -	🐮 📹 🔹 📑 📑 🧀 🦽 Go to file/function 🔡 👻 🖌 Addins 🔹					🔋 Project: (None) 🝷
💽 🖻 e	ccel_import.R × ■ csv_file ×			Environment History	Connections	
	🕨 🛲 🔳 Source on Save 🛛 🔍 🎢 📲	📑 Run 💁	📑 Source 👻 🗎	📹 🔚 🖬 Import D	ataset 👻 🞻	🗏 List - 🛛 🖸 -
	#install.packages("readx1")			dlobal Environment		
				Data		
	library(readx1)			💿 csv_file	39185 obs. of 46 variable	s
				🔘 xlsx_file	39185 obs. of 46 variable	s 🗌
	<pre>csv_file <- read.csv(file = 'G:\\My Drive\\Dillon_LLoyd_BRC\\TAMU\\Big Data Workshop\\mouse.csv')</pre>					
8	<pre>xlsx_file <- read_excel('G:\\My Drive\\Dillon_LLOyd_BRC\\TAMU\\Big Data Workshop\\mouse.xlsx', sheet = 'mouse')</pre>					
10						
11	view(csv_file)					
12	view(xlsx_file)					
2						
28						
13						
8						
<u> 1</u>						
S.						
				Files Plots Package		
				🗢 📫 🔑 Zoom	🗏 Export 👻 🔌	C
12						
18 I.						
<u>R</u>						
5:1	(Top Level) ÷		R Script 🔅			
Cons	vie Terminal × Jobs ×					
~/			4			
2 >						
<u> 1</u>						
S.						
20						
19						
8						
<u>R</u>						
9						





Importing .csv and .xlsx files into R

	Linnant D. 🔪 🛄 ulau fi											
exce	l_import.R × xlsx_fi	le × csv_file	^								Q	
		GSM1535917 🗘	GSM1535918 🗘	GSM1535919 🗘	GSM1535920 🗘	GSM1535921 🗘	GSM1535922 🗘	GSM1535923 🗘	GSM1535924 🗘	GSM1535925 🗘	GSM1535926 [‡]	GSM1535927
1	CLASS:DOSE	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01	0.03
2	ENSMUSG00000000001	2289	2171	2124	2538	2397	4707.00	2797.00	3235.00	1955.00	5983.00	1918.00
3	ENSMUSG0000000003	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
4	ENSMUSG0000000028	21	10	22	23	14	39.00	10.00	21.00	13.00	115.00	10.00
5	ENSMUSG0000000031	7	8	5	8	7	18.00	3.00	102.00	8.00	51.00	8.00
6	ENSMUSG0000000037	2	5	4	0	1	7.00	0.00	7.00	3.00	4.00	0.00
7	ENSMUSG0000000049	30854	34567	38467	34859	46227	59527.00	37223.00	42136.00	31393.00	92976.00	31710.00
8	ENSMUSG0000000056	1005	1183	1043	1417	1004	2712.00	1110.00	1318.00	1204.00	2113.00	1036.00
9	ENSMUSG0000000058	88	77	87	75	91	188.00	95.00	66.00	47.00	264.00	83.00
10	ENSMUSG0000000078	244	218	171	171	232	480.00	162.00	603.00	210.00	803.00	210.00
11	ENSMUSG0000000085	92	113	140	126	149	251.00	142.00	154.00	113.00	386.00	108.00
12	ENSMUSG0000000088	1911	1770	1547	1931	2413	2998.00	2190.00	1912.00	1614.00	5048.00	1512.00
13	ENSMUSG0000000093	35	26	44	38	55	68.00	37.00	32.00	36.00	145.00	33.00
14	ENSMUSG0000000094	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
15	ENSMUSG0000000103	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
16	ENSMUSG0000000120	22	17	14	24	38	63.00	54.00	54.00	13.00	54.00	15.00
17	ENSMUSG0000000125	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00
18	ENSMUSG0000000126	13	11	6	3	9	27.00	9.00	7.00	7.00	31.00	8.00
19	ENSMUSG0000000127	94	118	143	157	107	308.00	153.00	158.00	111.00	398.00	95.00





Questions??





Thank you for joining us !

The next session is on October 6, 2021 2:00 – 4:00 pm Eastern US Time "Manipulating and Displaying Big(ish) Data in R"

> Fred Wright – North Carolina State University Burcu Beykal – University of Connecticut Allison Dickey – North Carolina Sate University