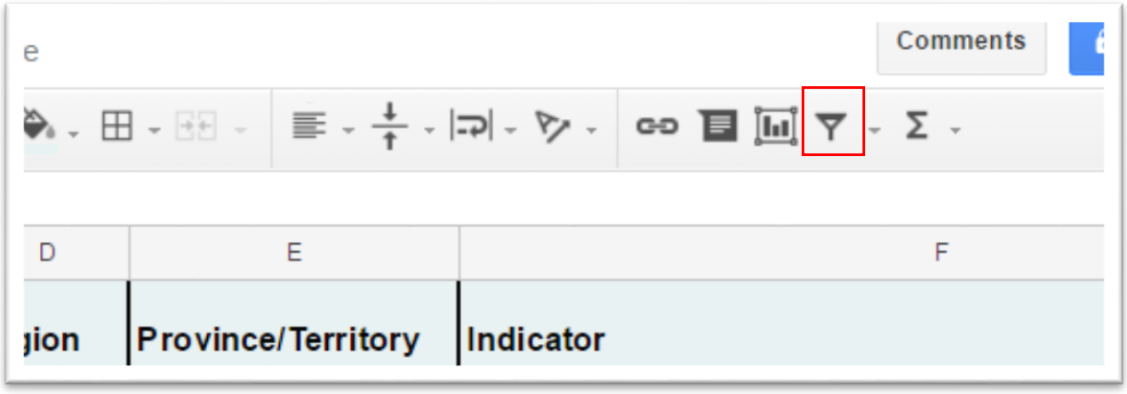
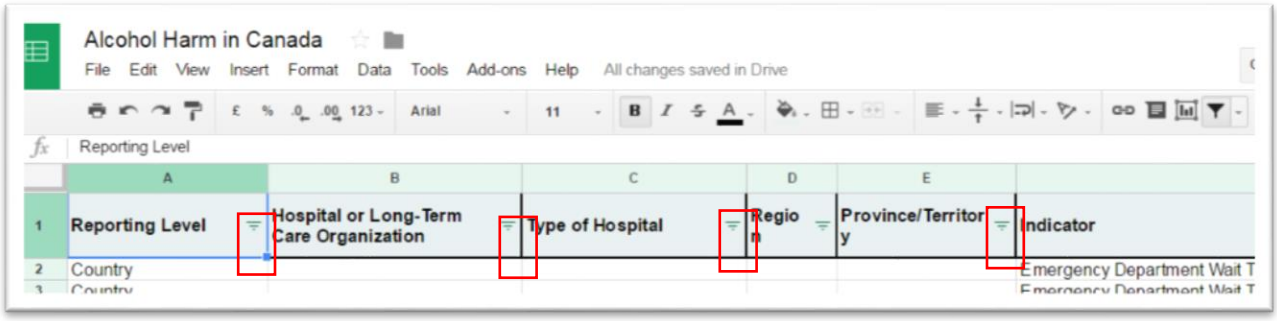
Filtering Canadian Institute for Health Information hospitalization data, using Google Sheets

1. You'll the Ottawa Citizen's story about the dataset by clicking here. For this tutorial, you will have the option of using Google's spreadsheet, or Excel.
2. The 77,000 number the reporter uses in the lead, is taken directly from the institute's news release.
3. Before figuring out if there's a better number we can use for a stronger story, let's determine the origin of the 77,000 hospitalizations caused by heart attacks.
4. The institute initially refused to release the dataset on which its findings are based, but eventually relented, agreeing to provide the data in Excel format, which can be downloaded by clicking here.
5. After reducing the table's size, it was uploaded to Google Sheets. If you can open the file in Excel, then use it to complete the rest of the tutorial. If not, then use the smaller dataset uploaded to Google Sheets. Pages 55-57 of The Data Journalist explains the differences in the interfaces in Google Sheets and Excel, and the circumstances under which each type of spreadsheet is used. The screen grabs in this tutorial are for Google Sheets, though you can also use Excel. The choice is yours.
6. After opening the table in Excel or Google Sheets, take a minute to answer these basic questions: How many rows does it have? Do the columns contain text or numbers. What do the column titles mean? Is there a dictionary or file that contains information about the table? After spending a few minutes, delving into these questions, you're ready to find the number used in the news release and newspaper story.

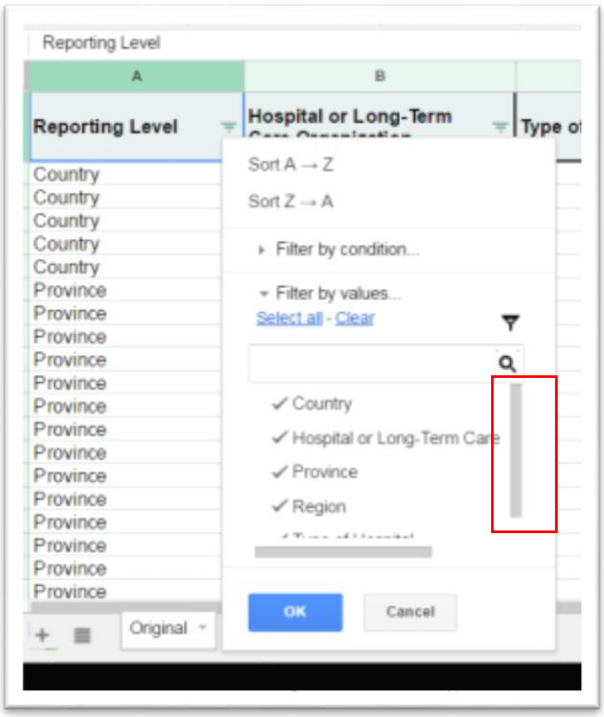
7. We must first apply the filter. Click your mouse anywhere within the table, and select the Filter, the to the right of the menu across the top that looks like a filter.



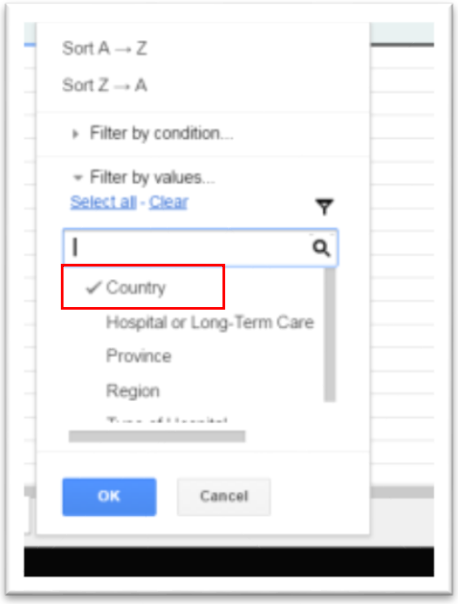


8. You'll see the filters to the right of the column tables.
9. Filters sort textual information in alphabetical order; numbers from smallest to largest. You can use the vertical scroll bar to the right to navigate up and down.

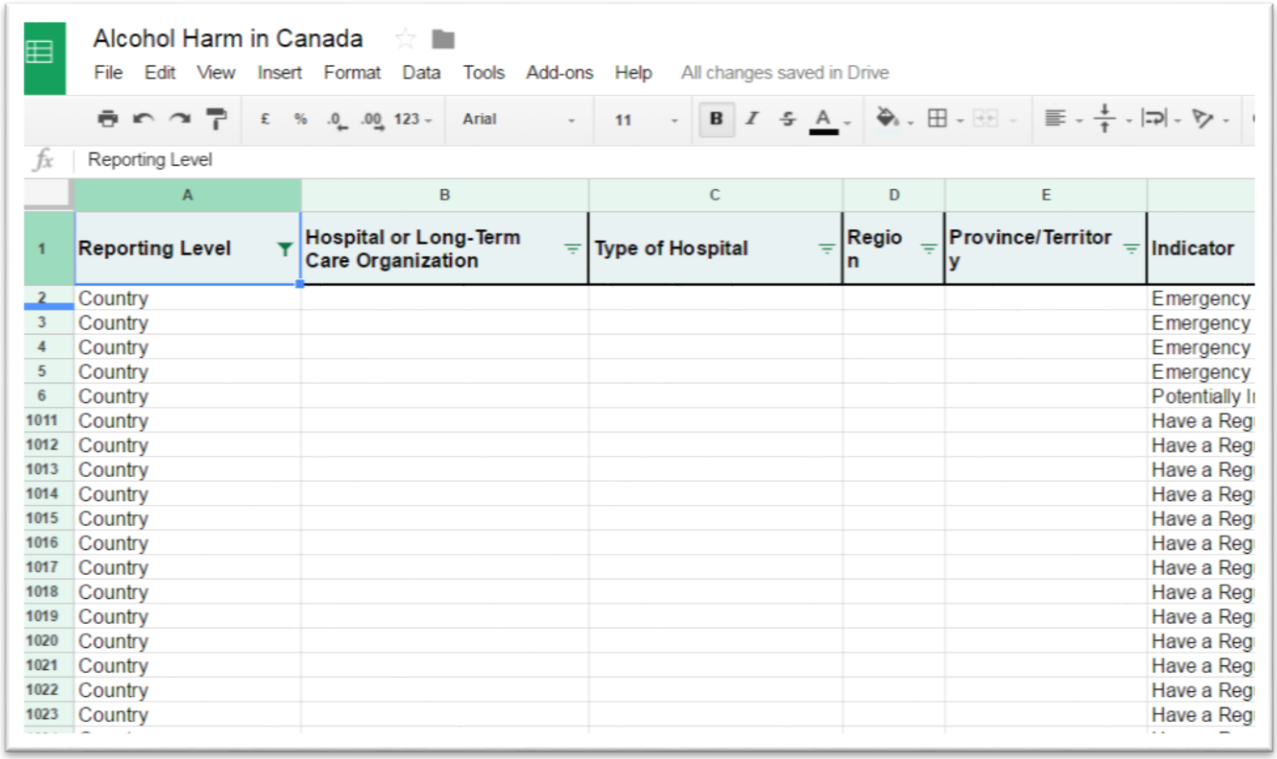
10. Click column A’s filter.



11. “Clear” the selection, and click the space to the left of “Country”.



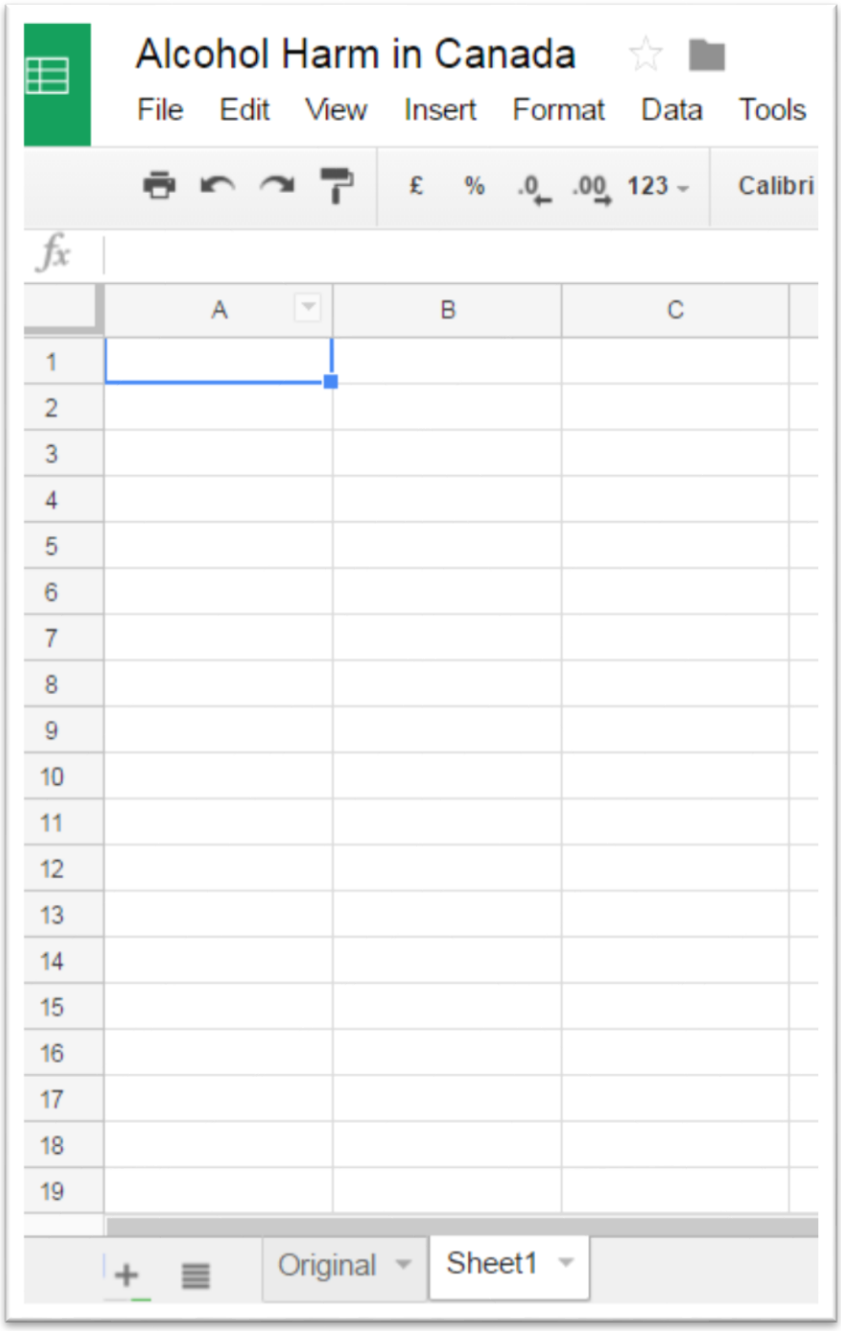
12. Select the “OK” tab.



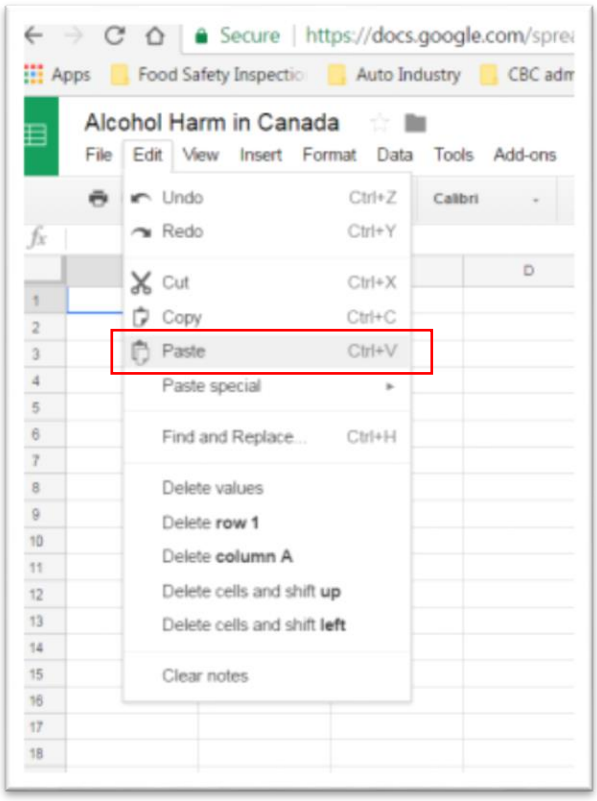
	A	B	C	D	E	
1	Reporting Level	Hospital or Long-Term Care Organization	Type of Hospital	Region	Province/Territory	Indicator
2	Country					Emergency
3	Country					Emergency
4	Country					Emergency
5	Country					Emergency
6	Country					Potentially I
1011	Country					Have a Reg
1012	Country					Have a Reg
1013	Country					Have a Reg
1014	Country					Have a Reg
1015	Country					Have a Reg
1016	Country					Have a Reg
1017	Country					Have a Reg
1018	Country					Have a Reg
1019	Country					Have a Reg
1020	Country					Have a Reg
1021	Country					Have a Reg
1022	Country					Have a Reg
1023	Country					Have a Reg

13. Use the horizontal scroll bar at the bottom of the table, to navigate to column G (Data Year).
14. Since the story was based on the most recent year for which the data is available, filter for 2015-2016.
15. For column H ( ), select “All”
16. For Column I (Sex), select “Both”
17. Sort column N (Numerator) in descending order and find the 77,064 number that was used in the story.
18. Now we want to put this table into a new worksheet.
19. Click the plus sign (+) at the bottom left of the table to add a new worksheet.

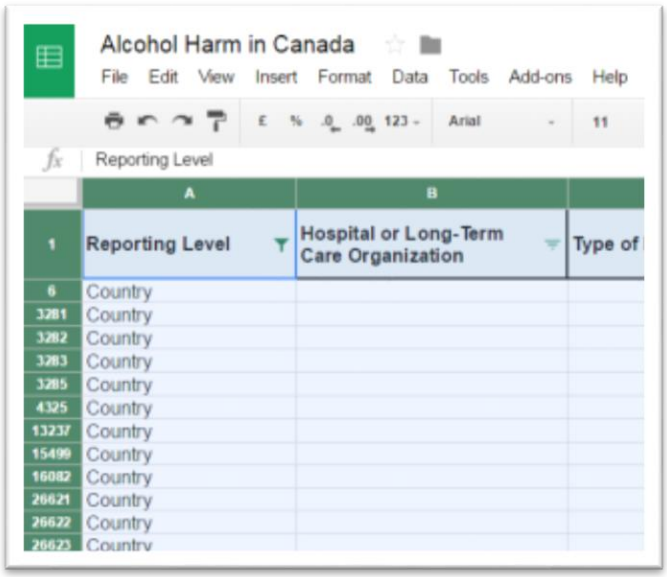
A
B
1
Reporting Level
Hospital or Long-Care Organizatio
6
Country
281
Country
282
Country
283
Country
285
Country
325
Country
3237
Country
5499
Country
5682
Country
5621
Country
5622
Country
5623
Country
3425
Country
9604
Country
9375
Country
3528
Country
3529
Country
7736
Country
Add Sheet
Original


Alcohol Harm in Canada
File Edit View Insert Format Data Tools
£ % .0 .00 123 Calibri
A B C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Original Sheet1

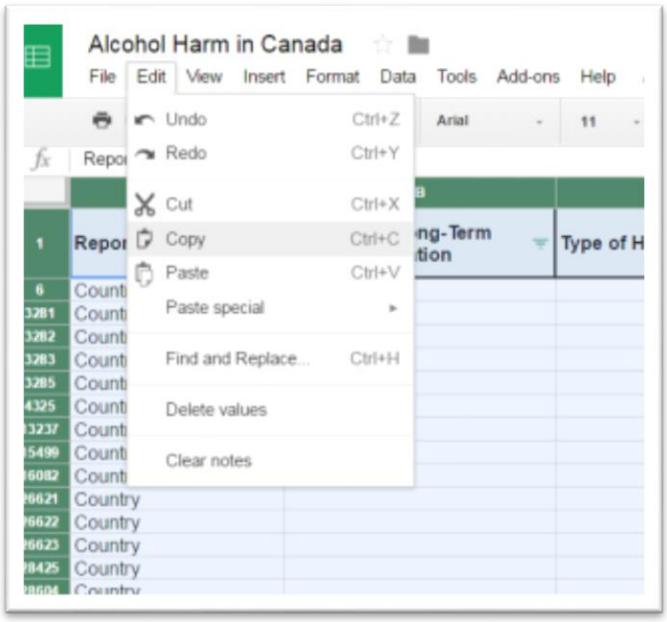
20. Paste the new table.



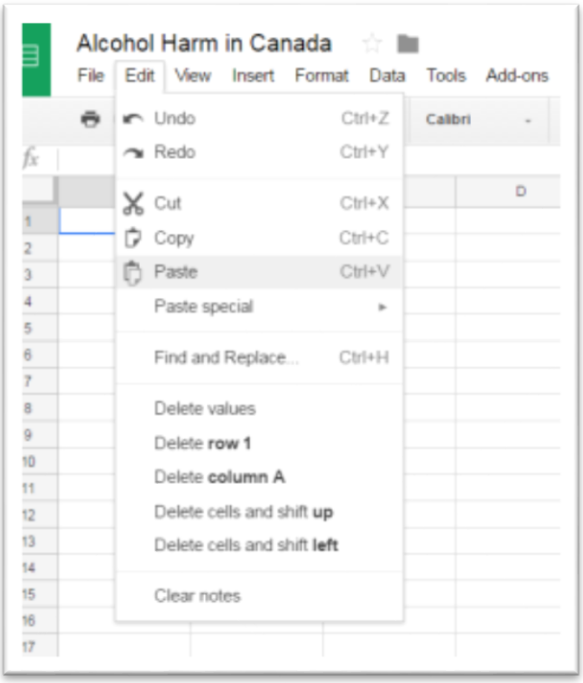
21. Return to the original worksheet and select the table by clicking rectangle between label A in the first column and the number 1, indicating the first row.

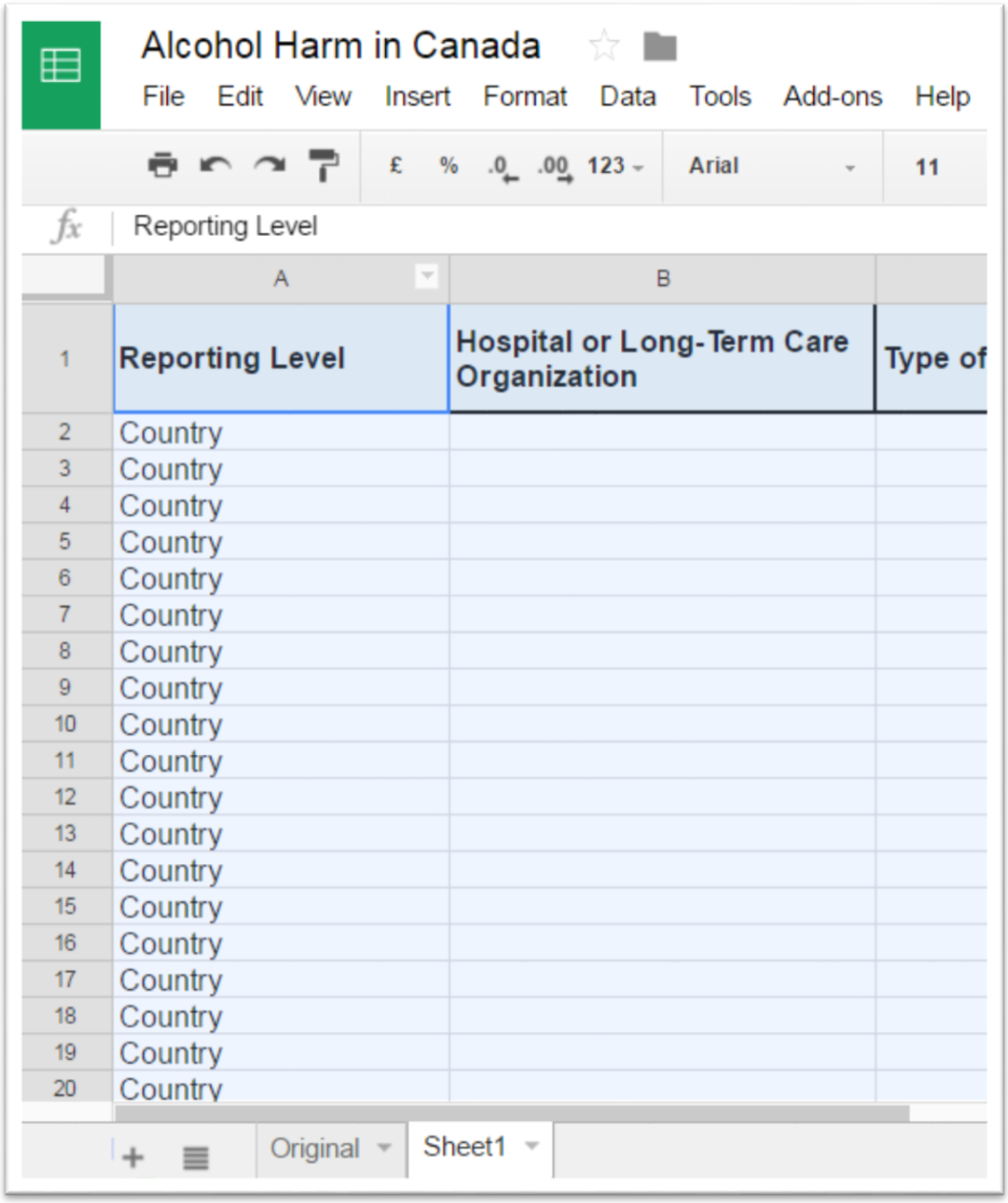


22. Copy the filtered table.

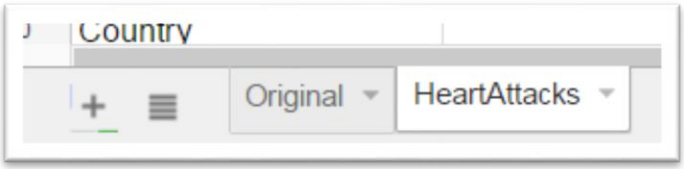


23. Return to the new worksheet, and paste the table that we've just copied.

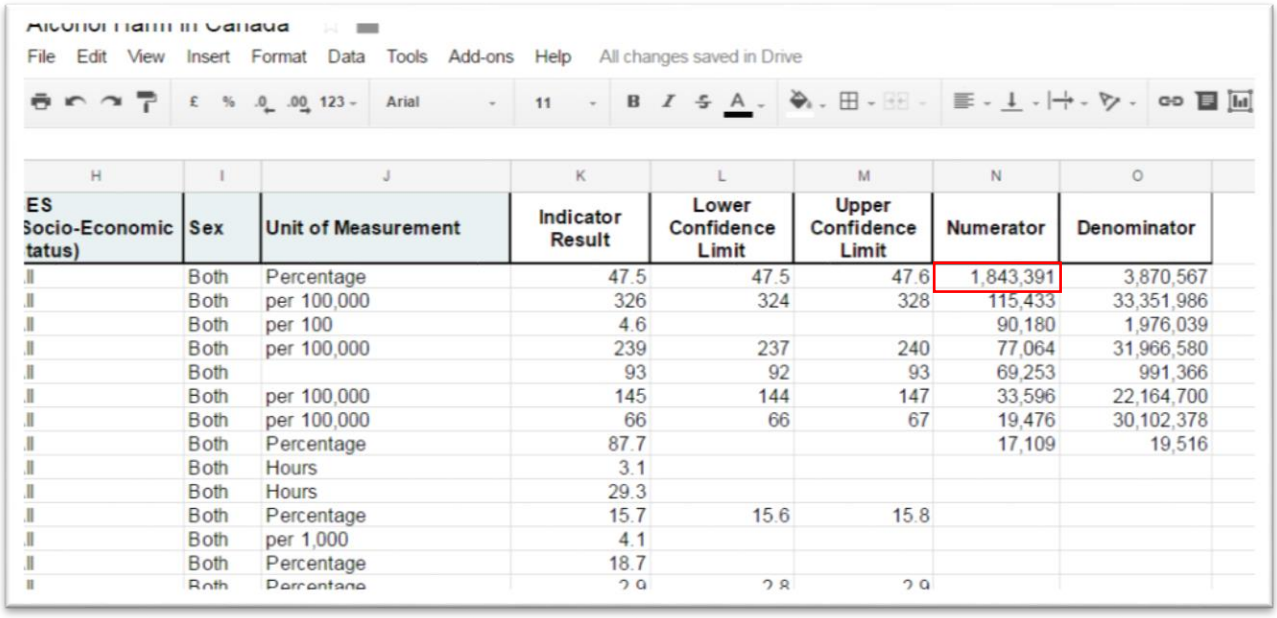

Alcohol Harm in Canada
File
Edit
View
Insert
Format
Data
Tools
Add-ons
Undo Ctrl+Z
Redo Ctrl+Y
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Paste special
Find and Replace... Ctrl+H
Delete values
Delete row 1
Delete column A
Delete cells and shift up
Delete cells and shift left
Clear notes
Calibri
D



24. Double click on “Sheet1” and type in a new name.

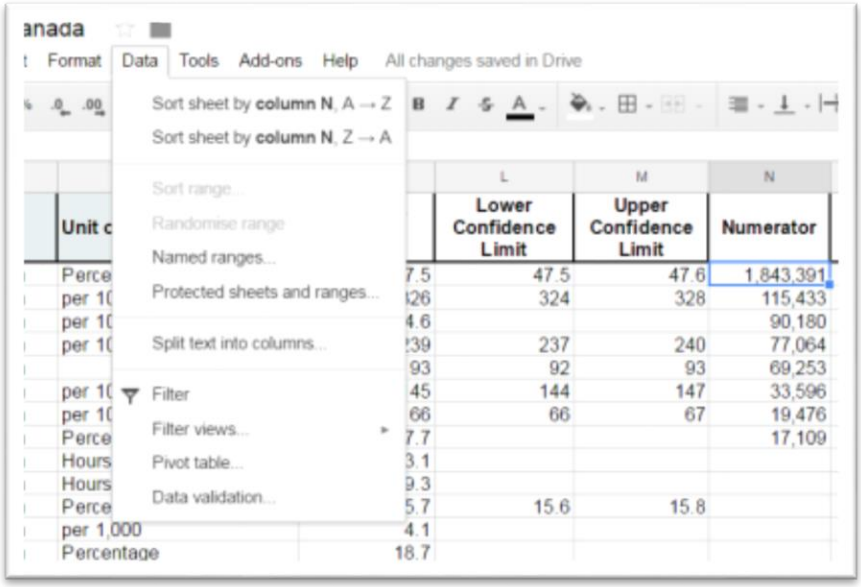


25. Use your horizontal scroll bar to navigate to column N (Numerator).



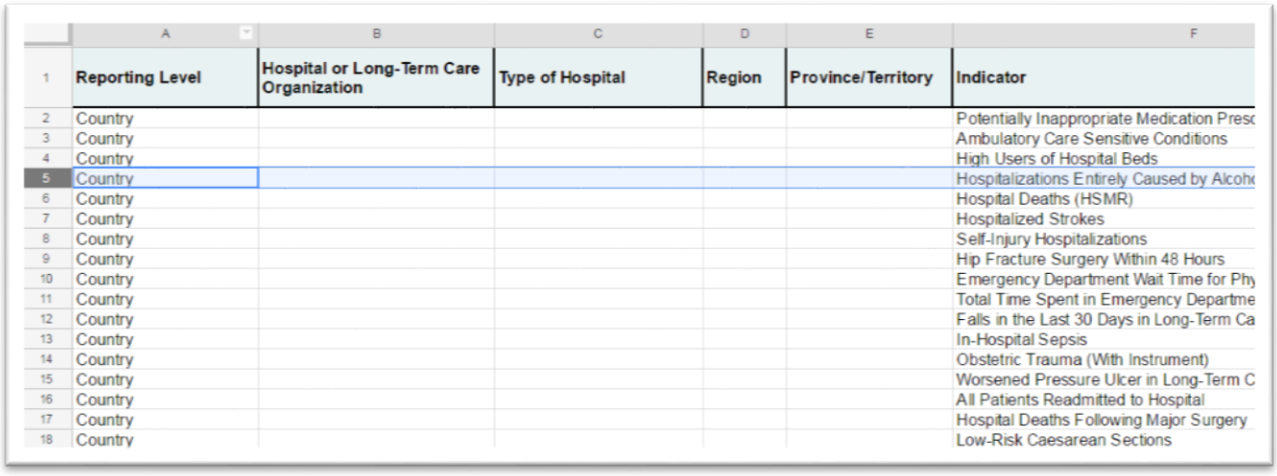
H	I	J	K	L	M	N	O
ES Socio-Economic tatus)	Sex	Unit of Measurement	Indicator Result	Lower Confidence Limit	Upper Confidence Limit	Numerator	Denominator
	Both	Percentage	47.5	47.5	47.6	1,843,391	3,870,567
	Both	per 100,000	326	324	328	115,433	33,351,986
	Both	per 100	4.6			90,180	1,976,039
	Both	per 100,000	239	237	240	77,064	31,966,580
	Both		93	92	93	69,253	991,366
	Both	per 100,000	145	144	147	33,596	22,164,700
	Both	per 100,000	66	66	67	19,476	30,102,378
	Both	Percentage	87.7			17,109	19,516
	Both	Hours	3.1				
	Both	Hours	29.3				
	Both	Percentage	15.7	15.6	15.8		
	Both	per 1,000	4.1				
	Both	Percentage	18.7				
	Both	Percentage	2.9	2.8	2.9		

26. Sort the values in descending order, by selecting the first cell under the column's label.
27. Click the "Data" category in the menu above the table.



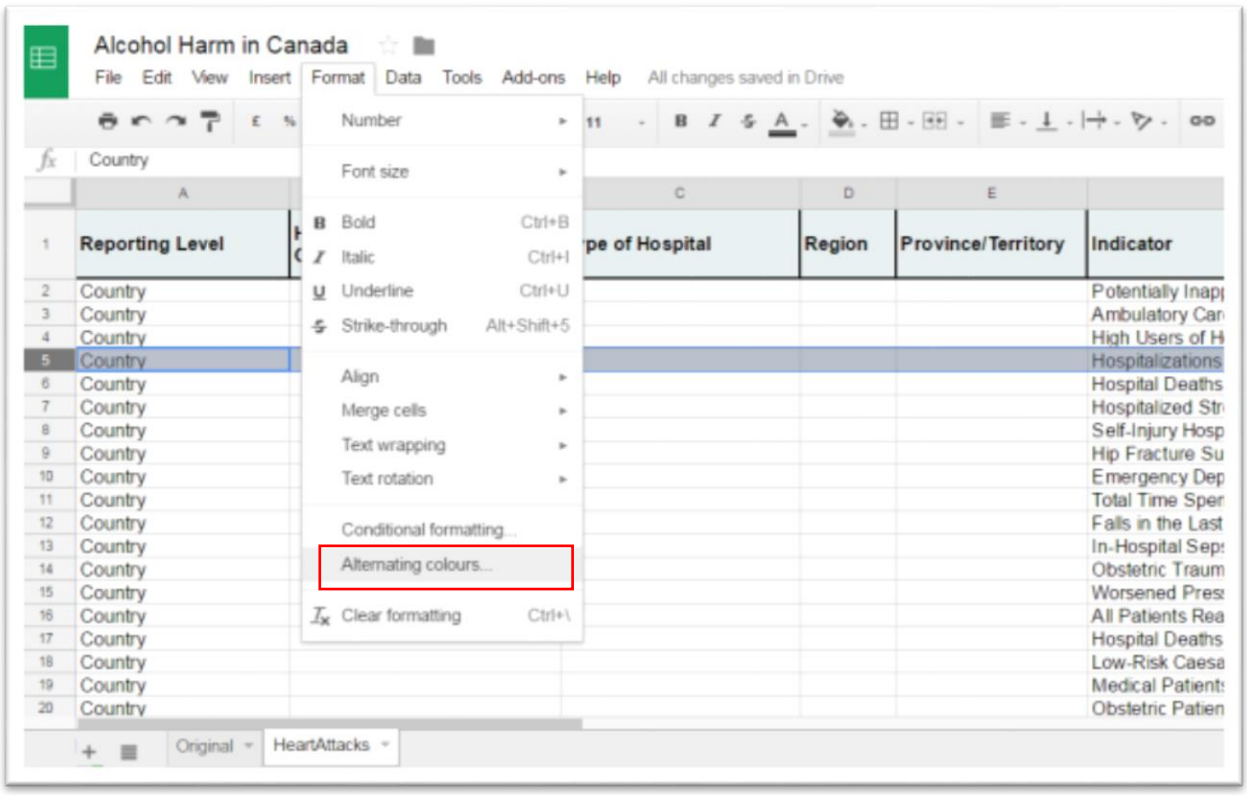
28. Choose the second option "Sort sheet by column N, Z-A, which means descending order. (Note: A-Z is ascending order) (NOTE: In this case, the values are already in descending order.)

29. Scroll back to the left, and click the number reference for row 5 so that the entire row is highlighted.

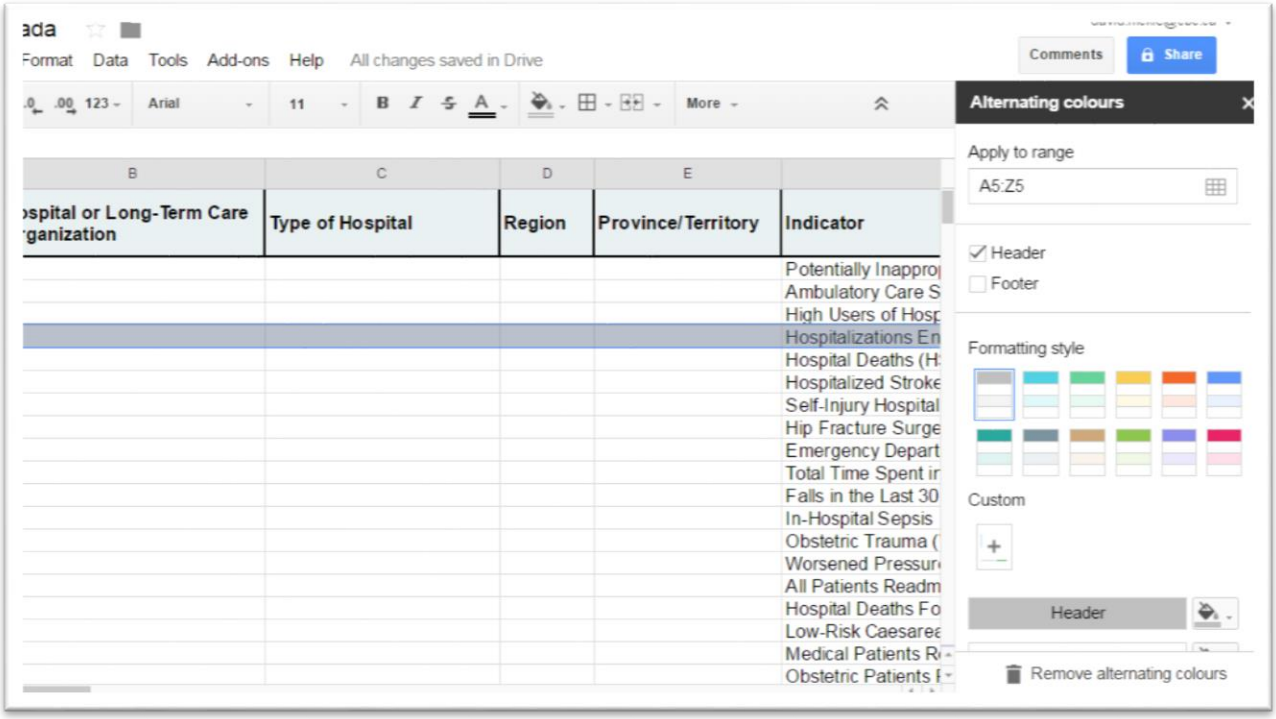


	A	B	C	D	E	F
1	Reporting Level	Hospital or Long-Term Care Organization	Type of Hospital	Region	Province/Territory	Indicator
2	Country					Potentially Inappropriate Medication Presc
3	Country					Ambulatory Care Sensitive Conditions
4	Country					High Users of Hospital Beds
5	Country					Hospitalizations Entirely Caused by Alcoh
6	Country					Hospital Deaths (HSMR)
7	Country					Hospitalized Strokes
8	Country					Self-Injury Hospitalizations
9	Country					Hip Fracture Surgery Within 48 Hours
10	Country					Emergency Department Wait Time for Phy
11	Country					Total Time Spent in Emergency Departme
12	Country					Falls in the Last 30 Days in Long-Term Ca
13	Country					In-Hospital Sepsis
14	Country					Obstetric Trauma (With Instrument)
15	Country					Worsened Pressure Ulcer in Long-Term C
16	Country					All Patients Readmitted to Hospital
17	Country					Hospital Deaths Following Major Surgery
18	Country					Low-Risk Caesarean Sections

30. If the information in this row formed the basis of your story, you might want to highlight it by clicking on “Format” on the menu above the table.

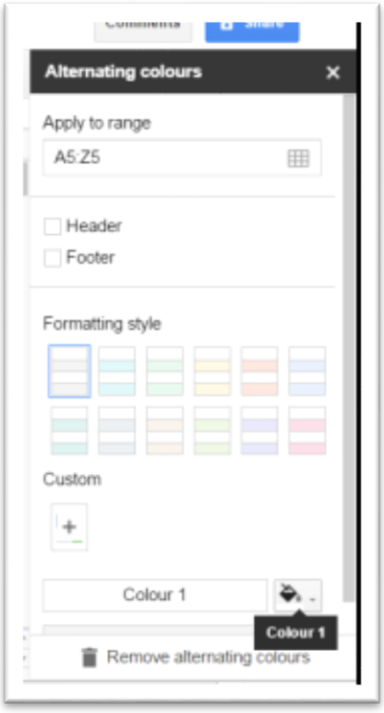


31. Select “Alternating colours”.

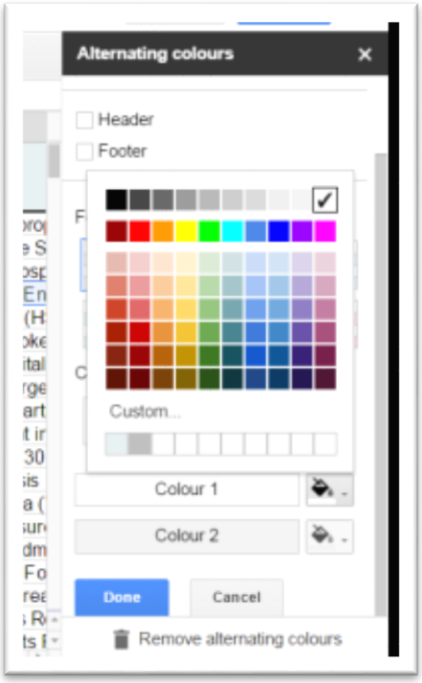


32. In the “Alternating colours” dialogue box, you’ll notice that all the cells in the row, or the “range”, is identified with the cell references A5 to Z5. For those using Excel, you can simply right-click to obtain the short-cut menu and select the colour that you want.

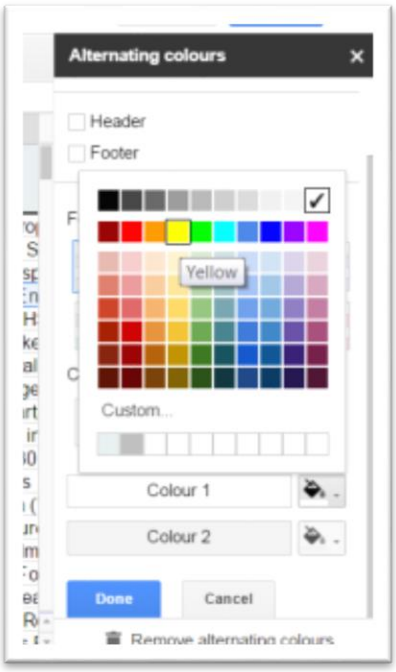
33. Uncheck the box to the left of “Header”.

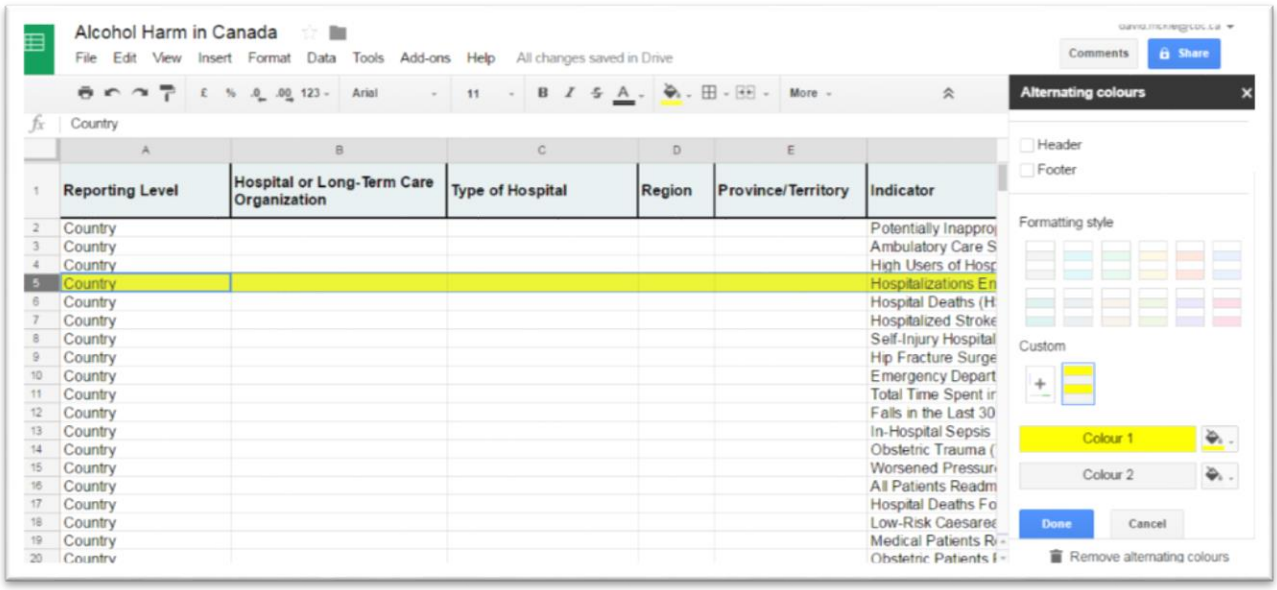


34. Click the arrow next to the icon to the right of the “Colour 1” section.



35. Choose a colour from the palette, let's say yellow.





36. If you're satisfied with the selection, click "Done" to return to the table.
37. Close the "Alternating colours" dialogue box.
38. Take a minute to study the new table, especially the categories rows 2-4. These also look serious, and might be worth taking a closer look by returning to the original worksheet and filtering the

data for categories such those in column H, the Socio-Economic Status, or “Sex” to see if the results are more serious for males of females. Or you might be interested in other categories in column F, “Indicator.” All of a sudden, you are in charge. So, instead of being guided by a news release, you’re in a position to make different – and possibly better -- choices using the institution’s own data.

39. At the beginning of this tutorial, we mentioned the importance of reading documentation such as data dictionaries, or “readme” files that help explain the numbers like these notations that I’ve made in the report uploaded to DocumentCloud.