



## Coding & Data Skills for Communicators

Dr. Cindy Royal

Texas State University - San Marcos

School of Journalism and Mass Communication

---

### Spreadsheet Basics

Excel is a powerful productivity tool. It's a spreadsheet that allows you to organize and analyze numbers and text. It is the foundational skill set of many jobs that require basic to advanced analysis. Knowing how to use Excel can be helping in securing a job, but also in performing in a variety of positions. It is part of the Microsoft Office productivity suite, along with Word and Powerpoint. Google Drive also offers a powerful spreadsheet option in the cloud, so it is good to understand the functions of each.

### Inputting and Basic Math

1. Open Excel and use File, New Workbook. Like other programs, the file menu is where you would Open, Close and Save documents.
2. Notice the layout of the spreadsheet, columns with letters, rows with numbers.
3. You identify a cell by the intersection of the column and row, i.e. A1, B25, C3, etc.
4. Start inputting 10 numbers going down in a column, starting in A1.
5. Sum a column. Excel has numerous functions (we will cover several). You invoke a function by typing in an "=" sign, then the function, then you put any parameters of the function in parentheses. For sum, you include the range of cells, separated by a colon.  
`=sum(A1:A10)`
6. If you begin the function and open the parentheses, you can select the range.
7. Now type in a second column of numbers in the B column and sum it. Let's add each row in C. Simply type "=" then select the first number, type a "+" and select the 2<sup>nd</sup> number. Hit enter.
8. Copy cells. To copy the function to all the rest of the column, find the dot in the bottom right corner of the cell. Click and drag to fill with all the sums. You can look in each cell to see the formulas. You can also use basic copy/paste feature under Edit menu.
9. Use columns D, E, F to subtract, multiply and divide the A & B columns.
10. Labeling Cells – it's easy to work with numbers in a spreadsheet, but unless you label your cells, you have no idea what you are looking at. Let's insert a row in 1 to add columns for years, and a column at A to add labels for the categories (use anything you want). Select the cell where you want to insert the row or column, and choose Insert, Row or Insert, Column. Add labels. You can use any of the formatting icons on the formatting toolbar (center, etc).
11. FYI, sometimes you need to select an entire row or column. You can do this by clicking on the number or letter. You can select everything in the spreadsheet by clicking the small cell that is in the top left corner.

## Percentages

Percentages are often the best way to compare data in a relative manner.

1. Let's remove the items in the cells in C, D, E, F. Select the cells (click and drag over them to highlight) and choose Edit, Clear, All
2. To get the percentage of the total, we need to divide each cell by the sum. Do this in C1 for the first number in the A column. However, if you repeat the formula by dragging, you get an error. You have to set the divisor so it will be the same, an absolute rather than relative cell reference. You do this with the "\$" sign and can hold either the row or the column. In this case, we want to hold the row. In the first formula, put a "\$" in front of the number. Now, you can drag and copy the do the rest of the percentages. The bottom one should be 100% (or 1, if not formatted yet). If you drag over to the next column, you get the same for the next column. You can see where this saves time in dealing with a large amount of data.
3. You can also use Command-T to toggle the absolute reference of the cell until the "\$" is in the right position.

## Formatting Cells

We have these nice cells with our percentages, but they aren't formatted as percentages. You format a cell under Format, Cells. Select the cells and apply an appropriate formatting.

## Sorting

Now, let's put our categories in alphabetical order. You have to think about what you want to order and what needs to stay in place.

1. Select everything except the sum row. Use the box in the top left of the sheet to select everything.
2. Choose Data, Sort. Notice the checkbox My List Has Headers. Check this if your list has headers. You have to choose your sort Column. Choose A, and then run the sort. Your items should now be sorted alphabetically.

## Functions

There are many functions you can use in Excel. We'll go over a few valuable ones, but feel free to experiment with them under Insert, Function.

Experiment with  
=count(range)  
=average(range)  
=median(range)  
=max(range)  
=min(range)

A helpful set of functions have to do with the financial functions. You can figure out payments on a loan or annuity, or get the present or future value.

Payment =pmt(rate, nper,pv,fv,type)

Future Value =fv(rate, nper, pmt, pv, type)

Present Value =pv(rate, nper, pmt)

You can also use the function interface to adjust variables in the function.

Excel can also do some of the functions that SPSS can perform. You are typically comparing two data sets to determine correlation, significant differences, etc.

=correl(array1, array2); gain use with actual numbers (Pearson) or rank-order (Spearman); degree of relationship between two datasets.

=chisq.test(actual, expected) – categorical between two or more independent groups; observed vs. expected values (calculate – row total\*column total/grand total). Probability that differences happened by chance.

=t.test(array1, array2, tails, type) – significant difference of means (related to variability of data) in two populations.

=rand or =randbetween generates random numbers – helpful for research

Note: An ANOVA test is used when testing significance of difference amongst more than two groups. This is no longer standard in Excel, best to use SPSS for ANOVA.

## Working with Data

We are going to use data from <http://datasets.flowingdata.com/hot-dog-places.csv>

1. This is a comma-separated file that shows years and then the results of the Nathan's Hot Dog Eating Contest. It provides the number of hot dogs eaten by the top three finishers. Obviously, there are many other datasets online, and some are very complicated, but this will work for illustration.
2. Use your browser to save the file. Make sure you save as a csv. File, Save Page As.
3. Use Excel to open the CSV. You may have to choose All Files to select the CSV. Sometimes you have to give Excel more information about the separators in the file before it opens.
4. Once the data is in, you can label, sum, etc. as we did above. Add labels to the rows for 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> place.

## Charting

1. Now that we have the data, we can use Excel's charting features to chart it. Let's consider how we'd like to show it. Select the data you want to chart, then choose the Chart tab. Pick the type of chart you want to use. I selected the 2-D stacked column chart. We want to show each finisher and the totals.
2. A chart is inserted on the page. You now can format it. What's wrong with it? We

need to adjust the series and the x-axis values. Choose Chart, Source Data to make any adjustments to the data. You can change the Chart Type by choosing Chart Type. You can decide if you want the Chart to be in a new sheet or its own object in the current sheet under Move Chart.

3. Chart Layout lets you adjust how the chart looks. Title, axis, legend, gridlines, etc. And if you double-click on part of the graph, you can change other features such as scale, number format, color, etc.
4. You can type in titles or use references to cells. Select the title, type an "=" in the formula bar, click the cell with the title and select enter.
5. What other ways could we envision this data? Is a stacked column the most effective method? Change the Chart Type to see what happens.
6. Once you setup a chart, the references are relative. Adding a row or column adjusts so that the chart stays intact.

### **Other Excel Features**

- You can save multiple spreadsheets in a Workbook. You can see each sheet associated with a workbook as a tab at the bottom. You can reference across spreadsheets. This is very helpful if you have multiple spreadsheets associated with a project.
- If you want to number a column or row, start with putting 1 and 2 in the cells, then select them and drag. If you just had one cell, it would copy the one number to each cell.
- You can copy formulas or just their values. Use Paste Special.
- You can put an if statement in a cell to test for something =if(c1=2012,"yes","no")
- You can use VLookup or HLookup to access data in cells that meets certain criteria.
- =VLOOKUP("apples", A1:C12,2) looks up an item and returns the value of the cell in the array defined by the index in the last item. Notice that if you are indexing a text string, it must be in quotes.
- =HLOOKUP(2012,A1:C11,3)
- Pivot Tables allow you to summarize data. If you have a spreadsheet that lists data by salesperson and region, you can summarize the region by applying a pivot table. The pivot table becomes a new sheet in the Workbook. I will demonstrate.

### **More Complicated Datasets**

We will be using data Texas population data.

Find this data at: <http://demographics.texas.gov/Data/TPEPP/Estimates/>

Save the last one in the list for Age, Sex, Race/Ethnicity . Save the csv version for 2015

You can open the csv by choosing it when you do a File, Open. You may have to change the file type.

## Filtering

1. First, let's filter the data to show only the rows that include data from Hays and Travis counties.
2. Choose Data, Filter to turn on the filter.
3. You will see an arrow at the top of each column that allows you to set the filter. Clear all filters, then search for Hays and Travis Counties. and turn it on. Choose OK to apply the filter. You should now see only the rows for Hays and Travis.

## Hiding Columns or Rows

1. We don't need to see all the columns. Let's hide the FIPS code.
2. Select the columns by clicking on the letter at the top of the column.
3. Right-click (or ctrl-click) and choose Hide. You will no longer see this range of columns.
4. You can select the range of columns that include it to Unhide when you need to see them again.
5. There are a lot of columns in this sheet and we could spend a lot of time focusing on a variety of them. But for the purposes of this exercise, let's just use the Total\_Pop, Total\_Male, Total\_Female. Hide everything else. You can look at the File Layout document to see exactly which columns represent which groups.

## Copy and Paste into a New Sheet

1. Now that we have the columns and rows we want to use, it is probably a good idea to copy this information into a new sheet, so we don't disrupt the layout of the original sheet with any new columns or rows we wish to insert.
2. Select all the data that is currently displaying and choose Edit, Copy.
3. Use the + sign at the bottom of the sheet to add a new Sheet.
4. In the new Sheet, paste (Edit, Paste) the data. Now we can focus on exactly what we want to study.
5. Let's also remove the line that has All Ages. Including this would double our results. We'll check back on this in the original file later.

## Percentages

Get % of total to see the percentage of male/female in each age group.

## Sorting

Use Data, Sort to see which age groups have the smallest male percentages.. Select the entire sheet (top left) and choose, Data, Sort. Sort on the % male column you just made. Pick Smallest to Largest. Make sure to select that your list has headers. What observations can you make about this data.

## **Pivot Table**

Pivot Tables are an advanced spreadsheet feature, but are very powerful in quickly summarizing a large dataset. Let's summarize the data for Travis and Hays that we have collected.

1. Select all the data by clicking on the box in the top left.
2. Under Data on the top menu, choose Summarize with Pivot Table.
3. A new sheet is created and you can use the wizard to select the rows, columns and values to be shown in the table.
4. For this data, choose County Name, Age Group and Total Population for fields.
5. Drag County Name to Column Labels; Age Group to the Row Labels; Sum of Total Population in the Values.
6. You now have a nice chart that shows you the Total Population by Age Group for the two counties. You can check the totals against the All Ages total from the original data.

## **Additional Functions – Google Spreadsheets**

### Import Functions

Google Spreadsheet has some very powerful import functions that allow you to essentially scrape data from the Web.

=ImportHTML(url, query, index)

- url – the address of the data
- query – “list” or “table”
- index – the number of the list or table on the page, if there are multiple.

For example:

=IMPORTHTML("http://governor.state.tx.us/music/musicians/talent/talent/", "table", 1)

Type the above code in a Google Spreadsheet and see what happens.

=ImportFeed(url of RSS feed)

=ImportData(url of csv)