**Excel Training Session for Reno**, Aug. 23, 2017

**How data journalism works**

* STEP ONE: Save the raw copy of the file. Do not use.
* STEP TWO: Start a document to keep track of your data analysis.
* STEP THREE: Check the data for basic errors. Answer any lingering questions.
  + Always ask for a “data dictionary”
* STEP FOUR: “Clean” the data
  + The data should be square (no blank entries)
  + Every column represents a single concept
  + No unnecessary spaces, commas or other formatting issues
* STEP FIVE: Analyze the data. Find the story.
* STEP SIX: Report on what you see. Get to the bottom of it.
* STEP SEVEN: Visualize the data with maps, charts, etc.

**Two important Excel buttons**

SORT: Find it on the “data” tab. Organize your data alphabetically, biggest to smallest or smallest to biggest. When Excel asks, always choose the option to “expand the selection.”

FILTER: Find it on the “data” tab. When you press the filter button, you will see little triangles appear. Select or un-select the check marks to add or remove data. Press the filter button again to return to your original dataset.

**An Excel function:** A formula you put in Excel that will give you back a value

* Always starts with an equals sign and is contained in parentheses

=SUM(8+12)

* Can refer to actual values or other cells

=SUM(A8+B14) or =MAX(A1:A20)

* More advanced functions are “logical tests”

=IF(criteria, true, false)

* Excel helps you by hinting to what you should include and giving you error messages

**Useful Excel functions to know**

MONTH-DAY-YEAR: It allows you to grab just one piece of a date e.g. you have a series of dates and you want a new field that just gives the year or you want a new field to just list the month.

=Year(Datefield)

=Month(Datefield)

=Day(Datefield)

DATE: Compile a date based on a year, month and day (in that order.)

=DATE(year,month,day)

WEEKDAY: This works much the same way as the above formula, but instead it returns the actual day of the week (Monday, Tuesday, etc). However the results come out as 1 (for Sunday), 2 (for Monday). Displaying words instead of numbers: Go to Format > Cells and choose Custom and type “ddd” in the Type box provided.

=Weekday(Datefield)

DATEDIF: Useful for calculating ages from DOB. Gives you the difference between two dates.

=Datedif(Date 1, Date 2, Unit of Measure)

Units of Measure: “y” – years “m” – months “ym” –months since the last year

You can use the TODAY() function to refer to today’s date. Or you could put a specific date in there (with quotes around it)

=Datedif(b2, today(), “y”)

=Datedif(b2, “1/1/2004”, “y”)

SUBSTITUTE: Allows you to mass replace or eliminate of a specific word or phrase in a column.

=(cell, oldtext, newtext)

LEFT: This tells the computer to start at the first byte on the left side of the field. Then we have to tell it how many bytes (or characters) to take.

=LEFT(celladdress, number of bytes to take)

=LEFT(B5, 5) – this will extract the first 5 characters of the contents of cell B5

\*there is a RIGHT function that operates the same way\*

MID: To use this function, you have to tell the computer which cell to work on, where to start and where to stop. If you want to take everything that remains in the field, just put a really big number in that will likely encompass all possibilities.

=MID(celladdress, byte number to start at, number of bytes to take)

=MID(B5,10,4) – this will start at the 10th character and take 4 characters.

SEARCH: This works as a tool to tell the computer to either start or stop taking a “string” at a certain character (or space). This is how we can tell the program to split a name field at the comma, for example. For this type of work, it is used in conjunction with the MID function. The character you what to find should be enclosed in quotes.

=SEARCH(“character we want to find”, celladdress)

=SEARCH(“,”,B5)

You can combine this with Mid to explain that you either want to start or stop at a certain character (even if the character isn’t located at the same byte in every record).

=MID(b5, search(“,”, b5), 100)

CONCATENATE: Combine strings of data from other cells into a single cell

=CONCATENATE(first value, second value, third value)

TRIM: Eliminate spaces before or after the characters in a cell.

=TRIM(text)

**More Advanced Excel**

IF STATEMENTS:

These are one of several LOGICAL functions in Excel. It’s an extremely powerful tool for a variety of tasks, most notably for assigning categories to your data based on certain criteria and for some data cleanup functions that require looking for patterns. They allow you to do one thing if your criteria is true, and another thing if your criteria is false.

A basic IF statement consists of:

1) What we’re going to measure as being either true or false

2) What to do if it’s true

3) What to do if it’s false

=IF(criteria, true, false)

=IF(e2>f2, “visit”, “home)

=IF(e2>f2, “visit”, IF(e2=f2, “tie”, “home”))

PIVOT TABLES:

Pivot tables are one of Excel's most powerful features. A pivot table allows you to extract the significance from a large, detailed data set by grouping your data together in different ways so you can more easily make comparisons.

The "pivot" part of a pivot table stems from the fact that you can rotate (or pivot) the data in the table in order to view it from a different perspective. You're not adding to, subtracting from, or otherwise changing your data – just reorganizing it.

Step 1: On a PC, click anywhere inside your data array, click the Insert tab, select "PivotTable" and click "OK" on the pop-up menu. Always start it in a new worksheet.

Step 2: Excel will create a blank pivot table for you. Next, drag and drop a field – labeled based on the names of the columns in your spreadsheet – into the "Row Labels" area. This will determine what unique identifier the pivot table will organize your data by.

Step 3: Add in some values you want to see, grouped by your unique identifier, by dragging a field into "Values." You can also group the data into columns by dragging a field into “Columns.”

Step 4: Adjust the results. The “sum” of the values will be calculated by default, but you can easily change this to something like average, maximum, or minimum. Click on the small upside-down triangle next to your value and select "Value Field Settings" in order to access the menu. You can also right click in your results and press “group” to group values together. You can also right click in the data and press “show value as” and “column total” to see percent values.