A few of my favorite data things

Lam Vo @lamthuyvo

Why are you panicking? | **This may be your solution**

CLEAN

There are different spellings for the same thing!

Open Refine | Facets and Clusters
Allows for consolidation, cleaning and mass editing of data.

Open Refine | Edit cells > Transform
With this formula you can replace specific values with another one.
value.replace('value that needs replacing' ,'value you want instead')

FORMAT

There's garbled stuff in my data or I don't like the words that are used repeatedly!

Excel, Spreadsheets & Open Refine | Dates, names, addresses
For custom date formatting in Spreadsheets, select the column you want to modify, then go to *format > number > more formats > more date and time formats*. Then add and drop formats.
For name formats in Excel or Spreadsheets, *=PROPER("john smith")* results in "John Smith"
For splitting columns in Open Refine, select *split into several columns > by separator and into two columns* in the column's dropdown menu

MERGE

Data comes from different places and needs to be merged!

Excel or Spreadsheets | column or row-based joining of data sets
Use vertical lookup to join data based on common column values. This formula searches down the first column of a range for a key and returns the value of a specific cell in the row found.
=VLOOKUP (value, table, col_index, [range_lookup])

Or combine index with match. The formula *=index(range, row_or_column)* gets you a value based on the index/position of a value and *=match(lookup_value, lookup_range, match_type)* gets you the position of a value. Use them together:
=index(range, MATCH(lookup_value, lookup_range, match_type))match_type))

ANALYZE

I don't know what to make of the data!

Excel or Spreadsheets | sorting
Sort it to find outliers or to rank a data set. Do so by highlighting and then going to *data > sort range > selecting from "A to Z" or "Z to A"* (optionally, you can first sort by one column and then by another).

Excel or Spreadsheets | summarize raw data
If you have raw data (each row represents a data entry, which includes individuals, survey respondents or households), then you can summarize and count entries by using pivot tables (*data> pivot table*). With a pivot table you can count or add values by common categories. Fun things include adding *row* and *value* categories.

Excel or Spreadsheets | isolating and filtering
If you're curious about a specific category subset in your data, you can highlight your data and then select *data > filter*. Then you can check and uncheck values in your filter to show or hide data (there's also a sort function!).

Excel or Spreadsheets | visualizing for analysis
Seeing a visual aids can help you make sense of your data. There's *insert > chart* to look at your data through different charts. There's *=SPARKLINE(range)* to show tiny graphics inside lines. And there's conditional formatting (*format > conditional formatting > color scale with min and max values*) which turns your spreadsheet into a heat map!