**Preparing a Dataset**

Prepping a Dataset

Once you have a dataset of results from patients, there are several things you need to do to a dataset to make it ready to be analyzed. This involves the following:

* Checking for missing data
* Checking for data that is incorrect or has been inputted incorrectly
* Making variable names (the heads of columns) one word so that you can analyze them in a program.
* Making sure there are no special characters (commas, parenthesis, slashes) in your variable names

Let’s take a practice data set. Let’s say our goal is to see the affect of symptoms of anxiety on pain. You can download it on the website.



There are 5 problems with this dataset: there is data that is inputted incorrectly in 2 cells, and the variable names (or the heads of the columns) for 3 of the columns have multiple words in them. Because we do a lot of surveys and questionnaires in redcap, data often comes out looking like this: there are certain cells which do not have the correct variables in them because the answer to a question may be a empty cell that a patient can answer with a 0 or 1 even though the question is asking yes or no. In addition, the variable names are often more than 1 word which does not work when importing to SAS.

So if I was going to make changes here, I would:

* Change the cell in A1 to “PtNo”, Change the cell in E1 to “SxAnxiety” and change the cell in F1 to “Pain”
* Find the true value of the cells in E16 and F13. If you do not know what the true value is, you can delete that cell and count them as missing values. Let’s say we have that information and we are able to fill it in with E16 being “No” and F13 being “2”.

The last thing you need to do when you are done is save your file as a CSV file instead of an excel file. A CSV is essentially a text file where each column of a specific row of your excel file is separated by a comma. This is a file that is much more easily read by the SAS system. In order to do that, click “Save As” instead of “Save” and where it says “save as type” make sure “CSV (Comma delimited)” is selected before you save. Pictures below

