* So I am guessing if you are taking this course, then you spend a good amount your time cleaning data, but to really put this into perspective I like to cite this article from the New York times, that data scientists spend 50 to 80% of their time cleaning data. Yikes, that is a lot of time doing kind of boring work!
* A lot of has probably had that report where we are deleting X columns, hiding Y rows, the whole process is really tedious, and we do it every week. Power Query can definitely help to automate that, but we want to make sure that what we are automating is actually going to help us analyze data, and that each time we size up a new dataset that we aren’t reinventing the wheel.
* And the paradigm we will borrow from to do that comes from Hadley Wickham, a statistician and prolific author of R packages. Wickham wrote a paper called Tidy Data that gave some pointers for how to clean datasets so that they are optimized for analysis.
* It’s really pretty brilliant and like a lot of brilliant ideas, it’s rather simple: We want to break our data into two categories, observations and variables. There are also sometimes called dimensions and measures. A good way to think about this is that observations are going to be fixed by the design of our data collection, and variables will be measured during the course of .
* So, let’s take an example here.
* Now, here are the rules we want to follow, and I will admit that Wickham’s paper mentions a third but we won’t get into that one:
* First, each variable should form a column, and next
* Each observation forms a row.
* So, let’s take an example here and the issue is…

[Work on this later]

* One way to think of this is if you are measuring the same unit across different columns, that’s probably a bad sign, because we don’t want .
* Another sign, if we are measuring different categories in the same row, that’s probably bad too, because now we can’t [] (Think about this)
* So, let’s take a couple of exercises here and consider what we need to do to make them tidy. I am taking these from Wickham’s paper so please don’t poke and look!
* Now before we get into Power Query I wanted to share this Abraham Lincoln quote, because I think it’s so true with data analysis, that we want to spend a good amount of time up-front really understanding what exactly we want to do with our data, and how it should change, before we jump into cleaning it up in Excel.