How Not To Be Wrong

With liberal borrowing from Rob Gebeloff’s slides [here](https://github.com/gebelo/training2017/blob/master/wrong.pptx).

* Go in assuming the data exists, and if it doesn’t, you can make your own
* It can be true, but you can be wrong
  + You can’t hedge with data stories. If it’s not true, it’s wrong
* Math is correct, but...
  + Wrong context
  + Wrong description of your findings
  + Wrong on uncertainties or caveats
  + Wrong focus; missing the forest through the trees

**Among Things to Consider...**

* Download data ASAP, may not be available in the future
* Share findings with your targets, including hostile experts
* Talk to the people who collected the data
* Know how many records you should have
* Check your results against published totals
* Record your steps(!!)
  + Include clearly where you got the data. Hidden sections of websites can be difficult to find
* Keep a clean copy of the data
* Consistency check your fields (company names) and know what you have in each column you’re using
  + Sort, filter and pivot on columns first
* Check for duplicates and understand how they got there
* Understand the codebook and any caveats that come with the data
* Go through the story and highlight each number, fact check it, ideally from a clean copy of the data
* Think of all the ways you could be wrong
* What confounding factors could you skew this data?
* Account for outliers. Could it be a mistake?
* Make sure you have the right denominator in percentages or division
* Standardize dates, including breaking them up if necessary
* Longer date range, the better; don’t cherry pick a base year
* Get the underlying record or understand where the data came from
* Do not type things. Use formulas, scraping, PDF manipulation, etc. Hand-entering data introduces mistakes
* Spot check your data after doing large changes