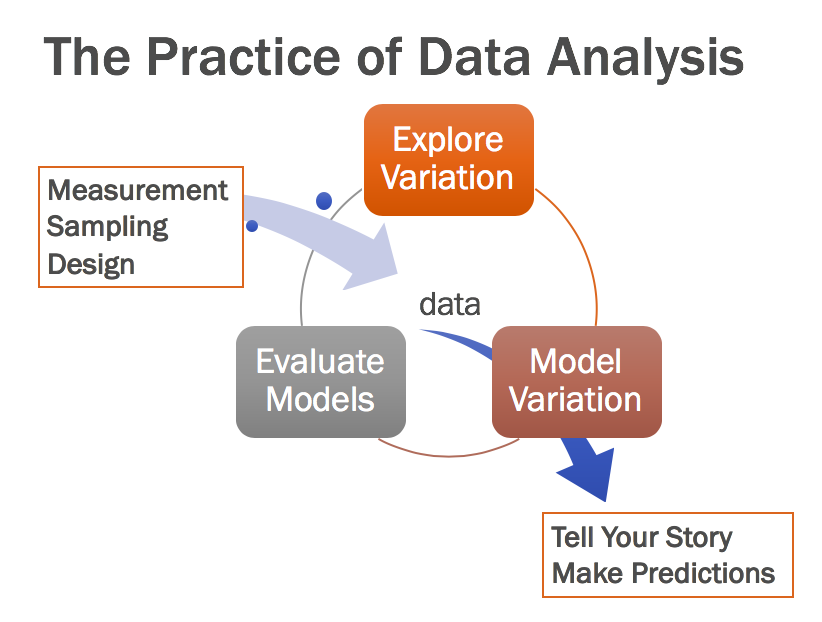
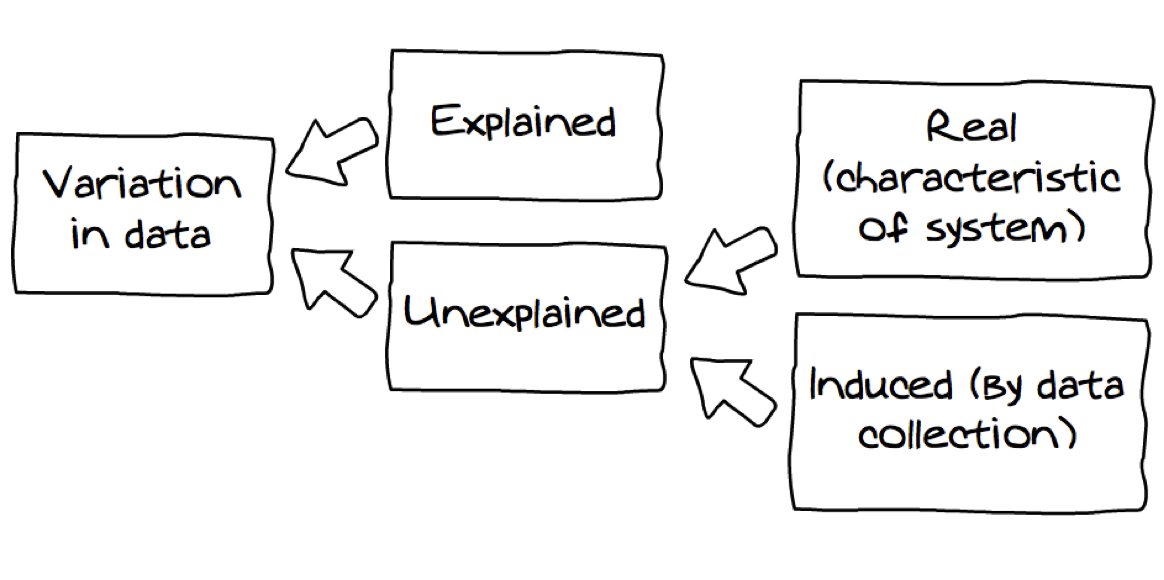
Week 3.1

* R practice
  + Age in days (rather than years) - put in histogram
  + How would you figure out what percentage of this sample is male?
  + Income as histogram or bar?
  + Divide data up into rich, middle, and poor
  + Filter for unmarried people
  + What are some interesting research questions you can ask with this dataframe?
  + *[didn’t ask this] How would we look at just a few people’s Age and Height and Sexual preference*
* Return to “practice of data analysis” - why do we engage in this cycle? “To EXPLAIN variation”
  + 
* Return to variation in **height**
  + Histogram, 5 number summary (on a number line under histogram)
  + What is “equal” across the two quartiles? (PE)
  + Why aren’t they equally spaced? (PE)
  + What explains variation in **height**?
  + 
  + Does **sex** explain variation in **height**?
  + What does it mean to “explain variation”? (think-pair-PE)
    - Knowing someone’s value on sex helps you make a better guess about height
    - Height = sex + other stuff
    - Explained (with sex), what about unexplained (induced by data collection - Mistakes? People being off by a little bit? Counting their hair, shoes off/on, etc? ; real characteristics - genes? Nutrition? etc)
    - What about sampling variation?
* Return to variation in age
  + Does sex explain variation in age as well as variation in height?
* Does marital status explain variation in age?
* Different sources of variation:
  + Variation that we can explain
  + Variation we haven’t explained yet
  + Measurement error
  + Mistakes
  + Sampling variation