review

- big picture of applied stats: see 36200 image idk
- we have statistics ($\overline{x}, \hat{p}, \ldots$) and standard error ($\text{SE}_{\overline{x}}, \text{SE}_{\hat{p}}, \ldots$)
- population: literally everyone, hard to measure
- sample: subset of population
- parameter: perfect summary (e.g. mean height)
- statistic: measurable summary (e.g. mean height of sample)
- stderr of stat: typical variation due to random sampling.
  - diff error formulae for each stat.
  - simply calc with software
- inference: give estimate and measure of how far off it might be
  - if statistic is random and sampling distribution known, we have probabilistic inference; can get p-value or margin or err

1 variable EDA

- categorical
  - bar graph
  - percent summaries
- quantitative
  - histogram
  - center: $\overline{x}$, median
  - spread: stddev, IQR, range
  - five number summary/box plot

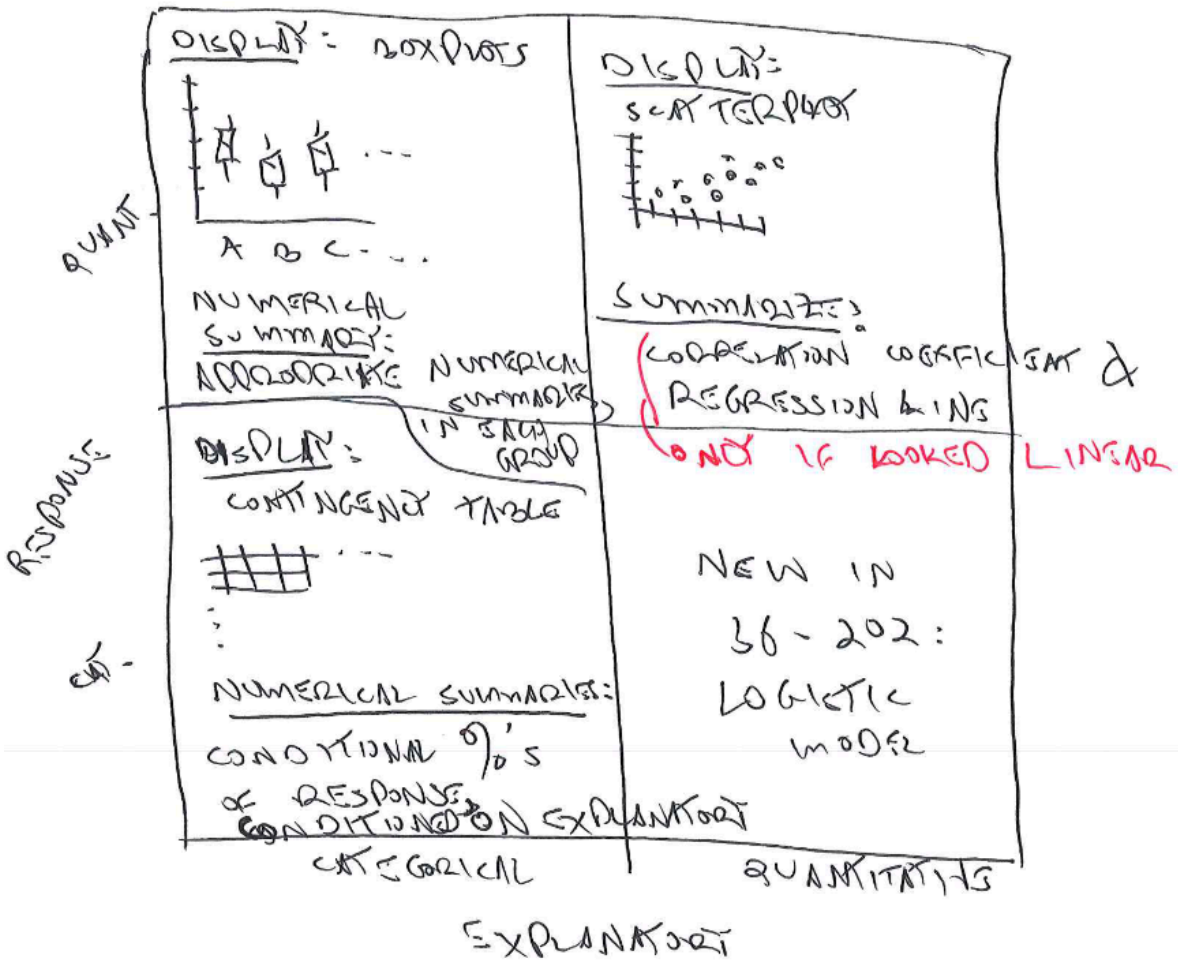
1 variable transformations

- need normal distributions?
- $x^{\frac{1}{n}}$, $\log(x + c)$ so everything is > 1.
- the above's inverses
- quantile plots (qqplot) can help us diagnose if normal enough (look for straight line)

2 variable EDA

- explanatory x axis $\rightarrow$ response y axis

Review of 2 Variable EDA (graphs and summaries to explore bivariate relationships)
[Reference: prerequisite course]



1 variable inference

- statistics ($\overline{x}$)