Notes

# Main Idea

* I want to test that when we are visually comparing the distributions of a certain variable (say Income in a year) across two groups (Males and Females say), using a histogram specifically, can the choice of bins offer two contrasting views?
  + Say one bin size makes it seem that more Males are concentrated in the higher Income bins, than women.  
      
    But another bin size makes it seem that there isn’t any / much difference in Males vs Females in this respect.  
      
    Yet another bin size might give the opposite view? More Females are concentrated in higher income bins that men (This seems unlikely, if say the underlying distribution implies that men have higher income on general for example)
* I think Simpson’s Paradox is something different. That requires another variable along which you can combine / segregate data to create contrasting interpretations (https://en.wikipedia.org/wiki/Simpson%27s\_paradox)

# Technical stuff

1. read.csv messes up column names for some reason – the names become “i..A” and “B” – can be verified by calling attributes(<name of variable in which you stored it>)
2. read\_excel from the readxl packages does fine (but technically this is for Excel files)
3. read\_csv from readr packages also works fine