I ILLINOIS

STAT107 Data Science Discovery

Lab: Simpson's Paradox

Man Fung (Heman) Leung

Fall, 2021

University of Illinois at Urbana-Champaign

- Please work in a group of 2–4 students
  - collaboration is important in data science!
  - meet new friends and discuss :)
  - let us know if you have any questions

Random fact of the day

Do you know why grocery stores produce and sell their own brands, e.g., Good & Gather (Target) or Great Value (Walmart)? This is related to the anchoring effect.

- Overview of grading procedure before we start
- Main page
- Retrieve the lab using git
- Complete the notebook
    - hint for puzzle 3.1, e.g., discovery_mpercent
    - the number of males can be counted by
      `sum(df_discoverypizza["Gender"].isin(male))`
    - the number of males who recommend Discovery Pizza can be
      counted by `sum((df_discoverypizza["Recommend"] ==`
      `"Yes") & df_discoverypizza["Gender"].isin(male))`
- Submit your work. Feel free to:
    - ask us questions
    - leave whenever you finish the lab

- Default total number of cells: 46
    - 1.1 in cell 7
    - 1.2 in cell 13
    - 1.3 in cell 18
    - 2.1 in cell 20
    - 2.2 in cell 23 (reflection)
    - 3.1 in cell 26
    - 3.2 in cell 31 (reflection)
    - 4.1 in cell 36
    - 4.2 in cell 39
    - 4.3 in cell 42
    - 4.4 in cell 45 (reflection)