

Data

- **census_usa.csv**
 - US state population data, from Homework Week 8.
- **CS171_TournamentData.csv**
 - This is our tournament data file containing tournament results from all the tournaments we are looking at. It contains 10 rows for every individual debater, 1 for each tournament in the year. This data was manually scraped and created. Code for scraping and cleaning this dataset can be found here:
<https://colab.research.google.com/drive/1sZCpos93lYksi9liCAQt3ZJMDdYQmRpX#scrollTo=FIEmxH-pwgb>
 - The columns we use are:
 - Wins: The number of wins that debaters won at a given tournament. If they did not attend this tournament, this variable is NA.
 - Name: First and Last name of the specific debater.
 - Code: School name and initials, ie. Harvard GK
 - School: School Name
 - Tournament: The name of the tournament that row is representing
 - Elo: The debater's Elo at a given tournament, this changes as the year goes on
 - Gender: girl/boyunknown - classified using <https://pypi.org/project/gender-guesser/>
 - Private/Public: Whether they are a private or public school student. Magnet schools are public under this classification.
 - State: The state from which the debater is coming from.
- **per-pupil-spending-by-state-2024.csv**
 - Data, per state, showing how much is spent on K-12 students. This dataset was sourced from
<https://worldpopulationreview.com/state-rankings/per-pupil-spending-by-state>
 - We only use the "Spending" column.
- **probs.csv**
 - A 4x4 matrix that is flattened with weights for Male/Female and Public/Private performance. Retrieved from empirical win odds. Used in the Debate Simulation tool to calculate win odds.
 - The percent listed is the probability that the first group beats the second group. ie. PrivateBoysvsPrivateGirls, then the probability is the probability that private boys wins.
- **win_percentages_df.csv**
 - For each tournament, the percentages at which Affirmative or Negation teams win rounds. Code for scraping and cleaning this dataset can be found here:
<https://colab.research.google.com/drive/1sZCpos93lYksi9liCAQt3ZJMDdYQmRpX#scrollTo=FIEmxH-pwgb>
 - The columns we use are:

- Tournament: The specific tournament name.
- Aff Win %: The number of times that Aff won in this tournament.
- Neg Win %: The number of times that Neg won in this tournament.