Mike Slovenski
INF 385T.9 Data Wrangling
3 February 2019

Project Ideas and Sources

Food Inspection Score and Type of Establishment

This project would make use of the dataset on the city of Austin's website that contains all of the food establishment inspection scores for all inspections in the city in the last three years. My idea is to compare inspection scores with establishment categorizations such as cafe/diner, bar & grill, convenience store, etc.

Data for establishment type has not been located in a specific dataset but could possibly be retrieved through using the Yelp or the Google API or some scraping method. Inspection score data is easily accessible but contains location redundancies in scores over time (the last three years) which could be dealt with by averaging or using the most recent score for each establishment.

Sources:
- Inspection Scores: https://data.austintexas.gov/Health-and-Community-Services/Food-Establishment-Inspection-Scores/ecmv-9xxi
- Establishment Types: Yelp, Google, or other source

Baseball Manager Ejections and Gameday Temperature

This project would compare baseball games in which manager ejections occurred to the gameday temperature as retrieved from NOAA's National Climatic Data Center. Games that took place in indoor stadiums could be used as a control or excluded from the comparison.

The data on manager ejections is not currently downloadable as a ready-made file but is available from the following index by way of "The Directory of Managers" with each manager's page listing their instances of ejection. The historical weather data is available by request from the NCDC – data could be requested for each city for the relevant time period in which ejections occurred there.

Sources:
- Manager Ejections: https://www.retrosheet.org/boxesetc/index.html
- Historical Weather Data: https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND

The Inc. 5000 and Population growth by State

This project would use Inc.'s 2018 list of the fastest growing companies in America as well as official census data to compare the locations of the fastest growing companies in America to the general population growth of those locations. Potentially, more specific population growth data could be used (i.e. city/town population comparisons) but have not been found.

Sources:
- The Inc. 5000: https://data.world/aurielle/inc-5000-2018
- Population Growth by State: https://www.census.gov/content/census/en/data/tables/time-series/demo/popest/2010s-state-total.html