

Capstone 2 Project Ideas and Datasets - Kirsten Regier

1. Decline in the number of US Colleges and Universities

Data sources:

- US College Scorecard Data: <https://www.kaggle.com/noriuk/us-college-scorecard-data-19962015>
- Number of higher education institutions in the U.S. 1980-2017:
<https://www.statista.com/statistics/240833/higher-education-institutions-in-the-us-by-type/>

According to the “Number of higher education institutions in the U.S. 1980-2017” data, between 2012 and 2017, the number of higher ed institutions decreased by roughly 400. I am aware of at least 2 colleges in my region that have closed in the last several years. I would like to see if there are identifiable factors that can predict school closings. This is particularly relevant as colleges and universities face financial difficulties in the face of the coronavirus, which compound the previously existing demographic trends that challenge college enrollment.

2. Text analysis of some sort

Data sources:

- Google N-gram viewer: <https://aws.amazon.com/datasets/google-books-ngrams/>
- Corpus of Contemporary American English: <https://www.english-corpora.org/coca/>
I’m not sure what question to pose - but I would like to work on something related to text analysis.

3. Traffic accidents and bike lanes.

Data sources:

- Chapel Hill, North Carolina: <https://catalog.data.gov/dataset/pedestrian-crashes>,
<https://opendata-townofchapelhill.hub.arcgis.com/datasets/bike-map-lines>,
<https://opendata-townofchapelhill.hub.arcgis.com/datasets/chapel-hill-greenways-and-bicycle-map>
- Allegheny County, PA: <https://catalog.data.gov/dataset/allegheny-county-crash-data> &
<https://catalog.data.gov/dataset/bikepghs-pittsburgh-bike-map-geographic-data>

I would like to look at the number and severity of bicycle-involved traffic accidents with respect to the presence or absence of bike lanes (or other bike-related infrastructure).