Available Datasets for Final Project

Below are links to datasets that can be used, along with some sample questions that can be asked of the dataset. Each of the datasets can be made available in .Rdata format using the tools we developed in the class. You are not limited to these datasets, but we can provide support for the use of these datasets.

Educational Longitudinal Study of 2002

https://nces.ed.gov/onlinecodebook
(Can use version on course webpage https://rafael-a-garcia.github.io/LLO8200-61-Spring2022/els.Rdata]
Unit of analysis: high school students
Time frame: sophomore year of high school through 10 years later
Example questions:
What predicts college enrollment?
What predicts first-year college retention?
What factors predict early labor market participation?

Delta Cost Project Dataset

https://nces.ed.gov/ipeds/deltacostproject/
Unit of analysis: colleges and universities
Time frame: 1987–2015
Example questions:
What predicts tuition revenue for private colleges?
What predicts enrollment levels at community colleges?
What predicts bachelor's degree completion for all colleges in Utah?

St. Louis Federal Reserve Economic Data (FRED)

https://github.com/business-science/tidyquant
Example, consumer sentiment for United States, 1978–present:

```
install.packages("tidyquant")
library(tidyquant)
sentiment<- tq_get("UMCSENT",get="economic.data" ,from="1978-01-01")
```

Unit of analysis: county, state, country
Time frame: variable, from 1950 to present, depending on series
Example questions:
What predicts changes in consumer sentiment?
What are the predictors of unemployment in counties in Alabama?
What predicts changes in consumer expenditures in all 50 states?

United States Bureau of the Census, Survey of Business Owners

https://cran.r-project.org/web/packages/censusapi/vignettes/getting-started.html
Unit of analysis: business owners
Time frame: every 5 years

Example Questions:
What predicts firm receipts?
What predicts firm size?
Do owner characteristics predict firm age?

Common Core of Data

https://github.com/UrbanInstitute/education-data-package-r
Example:

```
install.packages("devtoosls")
devtools::install_github('UrbanInstitute/education-data-package-r')
df <- get_education_data(level = 'schools',
source = 'ccd',
topic = 'enrollment',
by = list('race', 'sex'),
filters = list(year = 2008,
grade = 9:12,
ncessch = '340606000122'),
add_labels = TRUE)
```

Unit of analysis: K–12 schools
Time frame: yearly
Example questions:
What predicts school size?
Which states are likely to have the largest increases/decreases in teacher labor force?
Where will the most school principals be hired next year?

Program for the International Assessment of Adult Competencies (PIAAC)

https://nces.ed.gov/surveys/piaac/datafiles.asp

Unit of analysis: Individuals/Prisoners

Time frame: 4 cross-sections

Example questions:
What predicts Readiness to Learn?
How do Reading, Writing, Numeracy skills predict earned income?
Do higher levels of education predict greater use of skills inside/outside of work?