Data Sources

R Club

November 20, 2023

Abstract

These are some data sources that might be interesting as you begin developing your own projects. Most are free to access, either as a member of the public or as a student/faculty member at SJU. Some sources have paid versions or require a subscription. A general rule of thumb when it comes to collecting data is not to expect that you will be able to simply click one link and receive everything in a single .csv file in the exact format you desire. Often, it requires a process of 'data cleaning', wherein you determine the data you want, read through the documentation, understand its format, and finally import it into R. This is one of the more tedious aspects of data analysis, but it is also a necessary step! Don't get discouraged if you can't find exactly the data you are looking for. Be patient, be creative, and be open to modifying your research goals based on data availability.

Economics & Finance

- Federal Reserve Economic Data (FRED)
  Provides a vast array of economic data, including interest rates, inflation, and employment figures.
  Account Required: No

- Bureau of Labor Statistics
  Offers extensive data on employment, wages, inflation, productivity, and many other economic indicators.
  Account Required: No

- The World Bank
  Access to comprehensive data about global development, including economic and political topics.
  Account Required: No

- Yahoo Finance
  A resource for historical stock prices and financial data, which can be easily imported into R.
  Account Required: No (Some data might be limited)

- Quandl
  Offers a wide range of financial and economic datasets, including stock prices and GDP data.
  Account Required: Yes (Free and premium datasets available)

Politics

- Pew Research Center
  Offers datasets on political issues, public opinion, and demographic trends.
  Account Required: Yess

- Harvard Dataverse
  Hosts a variety of political science datasets, including electoral data and political surveys.
  Account Required: No (Some datasets may have specific conditions)

Sports

- MLB Savant
  Extensive source of baseball sabermetrics data.
  Account Required: No

- NFL Savant
  Extensive source of football data, including team performance, player stats, and play-by-play data.
  Account Required: No

- Sports-Statistics.com
  Wide-ranging sports data, including UFC, NBA, NFL, MLB, NHL, and more.
  Account Required: No

Media and Culture

- The Internet Movie Database (IMDb)
  Contains data on movies, TV shows, actors, and film ratings.
  Account Required: No (IMDbPro is paid)

- Spotify's Web API
  Provides data on music tracks, artists, and user listening habits.
  Account Required: Yes

- Google Trends
  Offers data on the latest search trends, useful for analyzing media and pop culture trends.
  Account Required: No

General Topics

- Statista
  Comprehensive platform for statistics, market data, and industry reports across numerous sectors and regions.
  Account Required: Yes (Free through St. John's Library)

- Kaggle Datasets
  Source for general datasets on wide variety of topics, including education, health, food, travel, etc.
  Account Required: Yes

- UCI Machine Learning Repository
  Hosts datasets specifically for machine learning, suitable for advanced students.
  Account Required: No

- Data.gov
  The U.S. government's open data portal, providing data on various topics.
  Account Required: No

- Gapminder
  Offers global development data, useful for cross-disciplinary projects.
  Account Required: No