Problem Set 4

Emilian Zavala

February 24, 2026

1 Data Sources of Interest

One data source I would be interested in scraping is baseball statistics data. Baseball is one of the most statistically rich sports in the world, and there is a wealth of freely available data on the internet. For example, websites such as Baseball Reference and FanGraphs provide detailed statistics on player performance, team performance, and historical data going back over a century. This data could be used to analyze player performance trends, compare players across different eras, or build predictive models for future performance.

2 Question 5 Results

2.1 Part (d): Object Types

The `class(mydf)` function returned `"matrix"`, meaning the dataframe is stored as a matrix object in R. The `class(mydf$date)` function returned `"NULL"`, which is because in a matrix object, columns cannot be accessed with the `$` operator like in a data frame.

2.2 Part (e): First Rows of Data

The `head(mydf)` command returned the first several rows of the dataset. The data contains columns for `date`, `description`, `lang`, `category1`, `category2`, and `granularity`. The first few events are from the Roman Empire era, categorized by place with a granularity of year. For example, one entry describes Tiberius quelling revolts in Germania, and another records Gaius Caesar and Lucius Aemilius Paullus being appointed consuls.