CaseStudy2

Ruhaab Markas – 281 253 1820

Email: [Rmarkas@gmail.com](mailto:Rmarkas@gmail.com)

Google Doc:

<https://docs.google.com/spreadsheets/d/1GsVHn3Bb-jAqGEnVRUZ8DUQt_r0frUwO9FtH1RJfK8M/edit#gid=0>

Link to Scrape site:

<https://en.wikipedia.org/wiki/List_of_countries_by_Human_Development_Index#Complete_list_of_countries>

<https://docs.google.com/spreadsheets/d/1fIeUdw0CEzXG4VzV1POcHfgmGbQWUZNpes4n6kfjjY8/edit#gid=0>

Notes:

Keep md file (know how to create md file)

Add 5 new columns for mean of each survey by employee

Variable names under 12 chars

Round years

Codebook

**Codebook - Doing Data Science Live Code Datasets**

**Introduction**

The following datasets are referenced in the Live Demos. Some are simulated for instructional purposes. Others are obtained from public, accessible avenues, or scraped for educational purposes. Please be sure to give credit if warranted.

Disclaimer: These data sets are not meant to be used as a substitute for real data or construed as advice for shaping legal, business, or political decisions. They are intended to be used for educational purposes only.

This data folder contains the following:

* Beers.csv - a data frame concerning alcoholic beverages around the United States.
* Breweries.csv - a data frame concerning breweries around the United States.
* OWLexamresults.csv - a data frame constructed of performance on Wizarding exams.
* read\_in.txt - a tab-delimited data frame of demographic information.

**Beers.csv**

A subset of alcoholic beverages around the United States.

* Type: Comma-separated value file
* Dimensions: 2410 observations X 7 variables
* Unit Used: Case Study 1; Midterm Review
* Public: Yes

Variable Information: Correct Class

* Name: The name of the beer: character
* Beer\_ID: A unique identifier of beer: factor
* ABV: The alcohol by volume in decimal format: numeric
* IBU: The international bitterness units: numeric
* Brewery\_ID: The unique identifier of brewery associated with this beer: factor
* Style: The beer's sub-typing, such as whether it is a cider, porter, etc.: factor
* Ounces: The ounces of beer in a single serving: numeric

Miscellaneous Information

* This is not all the beer in the United States.
* Note that some beers have the same name but different years. There may be a revision in recipe from year-to-year.
* Brewery\_ID can be matched to a column in Breweries.csv

**Breweries.csv**

A subset of breweries around the United States.

* Type: Comma-separated value file
* Dimensions: 558 observations X 4 variables
* Unit Used: Case Study 1; Midterm Review
* Public: Yes

Variable Information: Correct Class

* Brew\_ID: Unique identifier for the brewery: factor
* Name: The name of the brewery: character
* City: City level information of the brewery: character
* State: Abbreviated state indicator (i.e., "NC" for North Carolina): character

Miscellaneous Information

* These are not all the breweries in the United States
* This dataset can be joined with Beers.csv's Brewery\_ID

**OWLexamresults.csv**

Ordinary Wizarding Level exam results

* Type: Comma-separated value file
* Dimensions: 206 observations X 4 variables
* Unit Used: Reproducible Research
* Public: Simulated for demonstration

Variable Information: Correct Class

* Student: Unique identifier for a Hogwarts student: factor
* House: Hogwarts House belonged to (e.g., Slytherin): factor
* Pet: Did they have a pet or not: boolean
* Pass: How many O.W.L. exams the student passed: numeric

**read\_in.txt**

Sample demographic information to be cleaned

* Type: Tab-separated text file
* Dimensions: 154 observations X 5 variables
* Unit Used: Tidy Data
* Public: Simulated for demonstration

Variable Information

* Name: The name of the person, containing first and last name separated by an underscore: character
* ID: A unique identifier of the individual, in case there were repeat names: factor
* IQ: The intelligence quotient score of the individual; IQ scores as a population are always centered on 100 with a standard deviation of 15: numeric
* Grades: University averaged score on advanced engineering tests. Scores are bounded between 0 and 100. It can be thought of as a percentage: numeric
* Legs: Number of legs that an individual has: numeric.