Importing_Data_Crena

Andrew Crena

2023-05-19

R Markdown

Importing Data in the form of CSV Files

```
example_brain_data <- read.csv("Brain_Data.csv", header=T, sep=",",
                               na.strings=c("NA", "888", "999"))
example_behavior_data <-read.csv("Bx_Data.csv", header=T, sep=",",
                                 na.strings=c("NA", "888", "999"))
```

This chunk of code is specifically for importing csv files into your working directory. This is operating under the assumption that your working directory contains the csv files being pulled in this chunk. The arguments are good opportunity to tell your system about any incorrect/missing values, so make sure to memorize them (atleast the common and relevant ones)!

It's important to remember that th data you are importing could have missing values, incorrect values, and other pieces of data that need to be attended to by the software. Using the arguments in this chunk is a good example of weeding out the missing values and other data that isn't wanted.

What if we are downloading csv files from the internet to do some exploratory data analysis? For example, what if i found an interesting data set on Kaggle?

This next code chunk is an example of downloading a data set from the internet so that it is expressed as a data frame in your working directory. It is a depression data set randomly chosen on Kaggle, and I want to write the code that will grab it from the specific URL that refers to the specific data set, or some code that will achieve the same goal but in different ways. Using Kaggle data usually requires installing the Kaggle package, and having permission or an API token to use the data for exploratory analysis. the data set can be found with this link: https://www.kaggle.com/datasets/arashnic/the-depression-dataset ALWAYS REMEMBER DIRECTORIES!!!! Where's your data coming from/going to?

```
# Assuming the csv file has been extracted and is in your working directory....
# We won't focus on additional arguments, so that we can focus on one thing at a time
```

```
library(readr)
scores <- read_csv("scores.csv")
```

In this example, we will show how to get the data set into a df after downloading it and extracting the zip file (I find this to be easiest at my current skill level)

```
## Rows: 55 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): number, age, edu
## dbl (9): days, gender, afftype, melanch, inpatient, marriage, work, madrs1, ...
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
View(scores)

# Next, view your assigned variable to make sure that your data frame was made properly
View(scores)
```

```


# Assuming the ZIPPED csv file has been extracted and is in your working directory....

# Unzip the downloaded file
unzip("DEP_DATA_KAGGLE.zip")
```

Remember that the bottom right quadrant is excellent for doing these same tasks, just not programatically. However, let's practice unzipping a file obtained from the internet!

```
## Warning in unzip("DEP_DATA_KAGGLE.zip"): error 1 in extracting from zip file
```

```
# Read the unzipped CSV file into a dataframe
data <- read.csv("scores.csv")

# Make sure to check if your data frame was created and looks good!

View(data)
```