**Assessment Part 2: Data Import**

In this part of the assessment, you will import real datasets and learn more about useful arguments to **readr** functions. You will encounter common issues that arise when importing raw data. This part of the assessment will require you to program in R.

Use the **readr** package in the **tidyverse**library:

library(tidyverse)

**Question 14**

1.0/1.0 point (graded)

Inspect the file at the following URL:

<http://mlr.cs.umass.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data>

Which **readr** function should be used to import this file?

read\_table

read\_csv

read\_csv2

read\_tsv

None of the above

correct

Answer

Correct:

Correct - this is a comma-separated value file.

You have used 2 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 15**

0.0/1.0 point (graded)

Check the documentation for the **readr** function you chose in the previous question to learn about its arguments. Determine which arguments you need to the file from the previous question:

url <- "http://mlr.cs.umass.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data"

Does this file have a header row? Does the **readr** function you chose need any additional arguments to import the data correctly?

Yes, there is a header. No arguments are needed.

Yes, there is a header. The header=TRUE argument is necessary.

Yes, there is a header. The col\_names=TRUE argument is necessary.

No, there is no header. No arguments are needed.

No, there is no header. The header=FALSE argument is necessary.

No, there is no header. The col\_names=FALSE argument is necessary.

incorrect

Answer

Incorrect:

Try again. Do you see variable names in the first row of the file? Read the documentation for the **readr** function.

You have used 2 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 16**

2.0/2.0 points (graded)

Inspect the imported data from the previous question.

How many rows are in the dataset?

correct

569 Loading

How many columns are in the dataset?

correct

32 Loading

You have used 5 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 17**

2.0/3.0 points (graded)

Read in the table from the following URL using a function from the **readr** library and save it as co2\_mauna\_loa:

url <- "ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2\_annmean\_mlo.txt"

Use the skip argument to skip all of the documentation rows so that the column names are c("year", "mean", "unc") or c("#", "year", "mean", "unc"). (The "#" column could be removed after import with, for example, select.

How many rows must be skipped?

This may require some experimentation - this is normal for data wrangling.

incorrect

60 Loading

Which **readr** function correctly imports the table as a data frame after skipping the documentation rows?

read\_csv

read\_csv2

read\_tsv

read\_table

read\_lines

correct

How many rows are in co2\_mauna\_loa?

correct

60 Loading

You have used 1 of 10 attempts