**Assessment Part 1: Data Import**

In this part of the assessment, you will answer several multiple choice questions that review the concepts of data import. You can answer these questions without using R, although you may find it helpful to experiment with commands in your console.

In the second part of the assessment on the next page, you will import real datasets and learn more about useful arguments to **readr** functions. The second part of the assessment will require you to program in R.

**Question 1**

1/1 point (graded)

Which of the following is NOT part of the data wrangling process?

Importing data into R

Formatting dates/times

Checking correlations between your variables

Tidying data

correct

Answer

Correct:

Data analyses, such as checking correlations or creating visualizations, are done AFTER the data has been processed into a tidy format.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 2**

1/1 point (graded)

Which files could be opened in a basic text editor?

Select ALL that apply.

data.txt

data.csv

data.xlsx

data.tsv

correct

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 3**

1/1 point (graded)

You want to analyze a file containing race finish times for a recent marathon. You open the file in a basic text editor and see lines that look like the following:

initials,state,age,time  
vib,MA,61,6:01  
adc,TX,45,5:45  
kme,CT,50,4:19

What type of file is this?

A comma-delimited file without a header

A tab-delimited file with a header

A white space-delimited file without a header

A comma-delimited file with a header

correct

Answer

Correct:

This file has a header that describes the contents of each column. Values are separated by commas.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 4**

1/1 point (graded)

Assume the following is the full path to the directory that a student wants to use as their working directory in R: "/Users/student/Documents/projects/"

Which of the following lines of code CANNOT set the working directory to the desired "projects" directory?

setwd("~/Documents/projects/")

setwd("/Users/student/Documents/projects/")

setwd(/Users/student/Documents/projects/)

dir <- "/Users/student/Documents/projects"   
setwd(dir)

correct

Answer

Correct:

Correct: this code will not work. You need to use quotation marks when providing a full path to the working directory.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 5**

1/1 point (graded)

We want to copy the "murders.csv" file from the dslabs package into an existing folder "data", which is located in our HarvardX-Wrangling projects folder. We first enter the code below into our RStudio console.

> getwd()

[1] "C:/Users/UNIVERSITY/Documents/Analyses/HarvardX-Wrangling"

> filename <- "murders.csv"

> path <- system.file("extdata", package = "dslabs")

Which of the following commands would NOT successfully copy “murders.csv” into the folder “data”?



file.copy(file.path(path, "murders.csv"), getwd())



setwd("data")

file.copy(file.path(path, filename), getwd())



file.copy(file.path(path, "murders.csv"), file.path(getwd(), "data"))



file.location <- file.path(system.file("extdata", package = "dslabs"), "murders.csv")  
file.destination <- file.path(getwd(), "data")  
file.copy(file.location, file.destination)

correct

Answer

Correct:

This command does NOT copy the “murders.csv” file into your “data” folder; instead it copies it into the parent directory, “HarvardX-Wrangling”. You need to specify the “data” folder, either by changing the working directory or by adding it to the file path.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 6**

1/1 point (graded)

You are not sure whether the murders.csv file has a header row. How could you check this?

Select ALL that apply.

Open the file in a basic text editor.

In the RStudio “Files” pane, click on your file, then select “View File”.

Use the command read\_lines (remembering to specify the number of rows with the n\_max argument).

correct

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 7**

1/1 point (graded)

What is one difference between read\_excel and read\_xlsx?

read\_excel also reads meta-data from the excel file, such as sheet names, while read\_xlsx only reads the first sheet in a file.

read\_excel reads both .xls and .xlsx files by detecting the file format from its extension, while read\_xlsx only reads .xlsx files.

read\_excel is part of the readr package, while read\_xlsx is part of the readxl package and has more options.

read\_xlsx has been replaced by read\_excel in a recent readxl package update.

correct

Answer

Correct:

You can use read\_excel for both .xls and .xlsx files, while read\_xlsx only reads .xlsx files.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 8**

1/1 point (graded)

You have a file called “times.txt” that contains race finish times for a marathon. The first four lines of the file look like this:

initials,state,age,time

vib,MA,61,6:01

adc,TX,45,5:45

kme,CT,50,4:19

Which line of code will NOT produce a tibble with column names “initials”, “state”, “age”, and “time”?

race\_times <- read\_csv("times.txt")

race\_times <- read.csv("times.txt")

race\_times <- read\_csv("times.txt", col\_names = TRUE)

race\_times <- read\_delim("times.txt", delim = “,”)

correct

Answer

Correct:

This code will import the comma-separated values table called “times.txt”, but the base R function read.csv does not produce a tibble. It creates a data frame.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 9**

1/1 point (graded)

You also have access to marathon finish times in the form of an Excel document named “times.xlsx”. In the Excel document, different sheets contain race information for different years. The first sheet is named “2015”, the second is named “2016”, and the third is named “2017”.

Which line of code will NOT import the data contained in the “2016” tab of this Excel sheet?

times\_2016 <- read\_excel("times.xlsx", sheet = 2)

times\_2016 <- read\_xlsx("times.xlsx", sheet = “2”)

times\_2016 <- read\_excel("times.xlsx", sheet = "2016")

times\_2016 <- read\_xlsx("times.xlsx", sheet = 2)

correct

Answer

Correct:

When the sheet argument is contained within quotes, the function expects a sheet name. There is no sheet named “2” in this spreadsheet, so the code will not work.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 10**

1/1 point (graded)

You have a comma-separated values file that contains the initials, home states, ages, and race finish times for marathon runners. The runners’ initials contain three characters for the runners’ first, middle, and last names (for example, “KME”).

You read in the file using the following code.

race\_times <- read.csv(“times.csv”)

What is the data type of the initials in the object race\_times?

integers

characters

factors

logical

correct

Answer

Correct:

If you don’t supply the argument stringsAsFactors = F, the read.csv function will automatically convert characters to factors.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 11**

1/1 point (graded)

Which of the following is NOT a real difference between the readr import functions and the base R import functions?

The import functions in the readr package all start as read\_, while the import functions for base R all start with read.

Base R import functions automatically convert character columns to factors.

The base R import functions can read .csv files, but cannot read files with other delimiters, such as .tsv files, or fixed-width files.

Base R functions import data as a data frame, while readr functions import data as a tibble.

correct

Answer

Correct:

This statement is false. The base R import functions can read files with other delimiters like .tsv using read.delim and can read fixed-width files using read.fwf.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 12**

1/1 point (graded)

You read in a file containing runner information and marathon finish times using the following code.

race\_times <- read.csv(“times.csv”, stringsAsFactors = F)

What is the class of the object race\_times?

data frame

tibble

matrix

vector

correct

Answer

Correct:

The read.csv function generates a data frame containing these data.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 13**

1/1 point (graded)

Select the answer choice that summarizes all of the actions that the following lines of code can perform. Please note that the url below is an example and does not lead to data.

url <- "https://raw.githubusercontent.com/MyUserName/MyProject/master/MyData.csv "

dat <- read\_csv(url)

download.file(url, "MyData.csv")

Create a tibble in R called dat that contains the information contained in the csv file stored on Github and save that tibble to the working directory.

Create a matrix in R called dat that contains the information contained in the csv file stored on Github. Download the csv file to the working directory and name the downloaded file “MyData.csv”.

Create a tibble in R called dat that contains the information contained in the csv file stored on Github. Download the csv file to the working directory and randomly assign it a temporary name that is very likely to be unique.

Create a tibble in R called dat that contains the information contained in the csv file stored on Github. Download the csv file to the working directory and name the downloaded file “MyData.csv”.

correct

Answer

Correct:

The read\_csv command creates the tibble in R and the download.file command downloads the csv file from the internet to the working directory with the specified name.

You have used 1 of 2 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.