Please complete this check on learning quiz after reviewing all of the required readings for this module. You may take this quiz as many times as you like.

# 

# Q1. File paths

Let’s say I'm using a Mac and we moved a file called diabetes.csv to a different folder on my desktop called my\_studies. What file path would we need to give R to tell it how to find the data?

| ✅ | /Users/bradcannell/Desktop/my\_studies/diabetes.csv | |
| --- | --- | --- |
|  | /Users/bradcannell/Desktop/diabetes.csv | |
|  | /my\_studies/diabetes.csv | |
|  | /Users/bradcannell/my\_studies/diabetes.csv | |

The correct answer is /Users/bradcannell/Desktop/my\_studies/diabetes.csv. [Click here for a review of file paths](https://www.r4epi.com/file-paths.html).

# Q2. Packages

Which of the following R packages are commonly used for importing data?

|  | convertr | |
| --- | --- | --- |
| ✅ | readr | |
|  | dplyr | |
|  | importr | |

readr is the primary package we use in this module for importing plain text files. We use dplyr for manipulating data frames, but not for importing data frames. As far as I know, converter and importr don't actually exist.

# Q3. Import csv

Which of the following functions is BEST for importing a csv file into R?

|  | readxl() | |
| --- | --- | --- |
|  | import() | |
| ✅ | read\_csv() | |
|  | read.csv() | |

readxl() and import() cannot be used to import .csv files.

read.csv() is base R's primary function for importing csv files; however, readr's read\_csv() function is approximately 10 times faster, doesn't convert character variables to factors by default, and behaves more consistently across operating systems and geographic locations.

# Q4. Data types

Which of the following is NOT a commonly used file type for storing data?

|  | .csv | |
| --- | --- | --- |
|  | .xlsx | |
|  | .dta | |
| ✅ | .doc | |

Comma separated value (.csv) files, Microsoft Excel (.xlsx) files, and Stata data files (.dta) are all commonly used for storing data. Microsoft Word (.doc) files are not typically used for *storing* data.

# Q5. Factors

Which of the following are reasons that you may want to use a factor variable to represent a categorical variable in R?

|  | Factors levels are displayed alphabetically when summarized | |
| --- | --- | --- |
|  | Factors can only take on values that are observed in the data | |
|  | Factors can easily be exported to plain text files | |
| ✅ | Factors can improve the efficiency of data entry | |

Factors levels are NOT displayed alphabetically when summarized. The unique levels of character vectors are.

One of the advantages of using factors is that they can contain values that are NOT observed in the data.

Factors cannot be exported to plain text files. In order to retain factors, you must export to a binary file format like .Rds.

By allowing you to enter a numeric vector and later apply labels, factors can improve the efficiency of data entry.