



<u>Course</u> > <u>Section 1: Data Import</u> > <u>1.1: Data Import</u> > Assessment Part 2: Data Import

Audit Access Expires Mar 19, 2020

You lose all access to this course, including your progress, on Mar 19, 2020.

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

1/1 point (graded)

Inspect the file at the following URL:

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

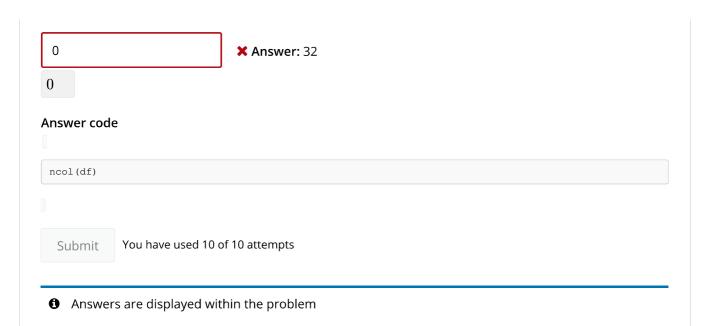
Which **readr** function should be used to import this file?

O read_table
● read_csv
read_csv2
C read_tsv
None of the above
✓
Answer Correct: Correct - this is a comma-separated value file.
Submit You have used 2 of 2 attempts

Correct (1/1 point)

Generating Speech Output

Question 15	
	mentation for the readr function you chose in the previous question to learn about its ermine 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.dat
Does this file ha import the data	ve a header row? Does the readr function you chose need any additional arguments to correctly?
Yes, there	is a header. No arguments are needed.
Yes, there	is a header. The <code>header=TRUE</code> argument is necessary.
Yes, there	is a header. The col_names=TRUE argument is necessary.
No, there is	s no header. No arguments are needed.
No, there i	s no header. The [header=FALSE] argument is necessary.
No, there is	s no header. The col_names=FALSE argument is necessary. 🗸
Submit Yo	ou have used 2 of 2 attempts
• Answers ar	re displayed within the problem
Question 16	
0/2 points (graded Inspect the impo	orted data from the previous question.
How many rows	are in the dataset?
0	★ Answer: 569
Answer code	
_	<pre>/mlr.cs.umass.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data" - read_csv(url, col_names = FALSE) (df)</pre>



Question 17

1/3 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.

3 **X** An

X Answer: 56

\(\)

Answer code

```
co2_mauna_loa <- read_table(url, skip = 56) # skip documentation
co2_mauna_loa</pre>
```

read	_csv
read	_csv2
read	tsv
	_table
read	lines
co2_mauna_	loa <- read_table(url, skip = 56) # skip documentation
co2_mauna_	
co2_mauna_ low many 2	rows are in co2_mauna_loa ? ** Answer: 60

© All Rights Reserved