

<u>Course</u> > <u>Comprehensive Ass...</u> > <u>Comprehensive Ass...</u> > Puerto Rico Hurrica...

Puerto Rico Hurricane Mortality: Part 1

Question 1

1/1 point (graded)

In the extdata directory of the **dslabs** package, you will find a PDF file containing daily mortality data for Puerto Rico from Jan 1, 2015 to May 31, 2018. You can find the file like this:

```
fn <- system.file("extdata", "RD-Mortality-Report_2015-18-180531.pdf", package="dslabs")</pre>
```

Find and open the file or open it directly from RStudio. On a Mac, you can type:

```
system2("open", args = fn)
```

and on Windows, you can type:

```
system("cmd.exe", input = paste("start", fn))
```

Which of the following best describes this file?

- It is a table. Extracting the data will be easy.
 - It is a report written in prose. Extracting the data will be impossible.
 - ullet It is a report combining graphs and tables. Extracting the data seems possible. ullet
 - It shows graphs of the data. Extracting the data will be difficult.

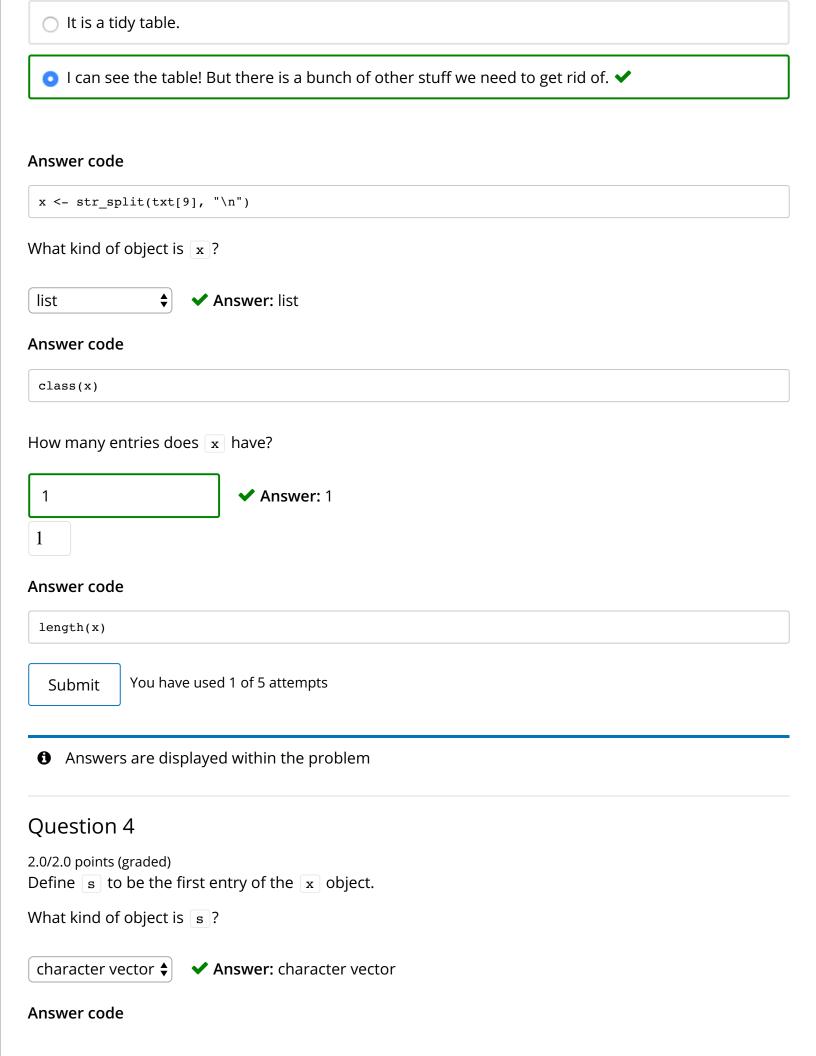
Submit

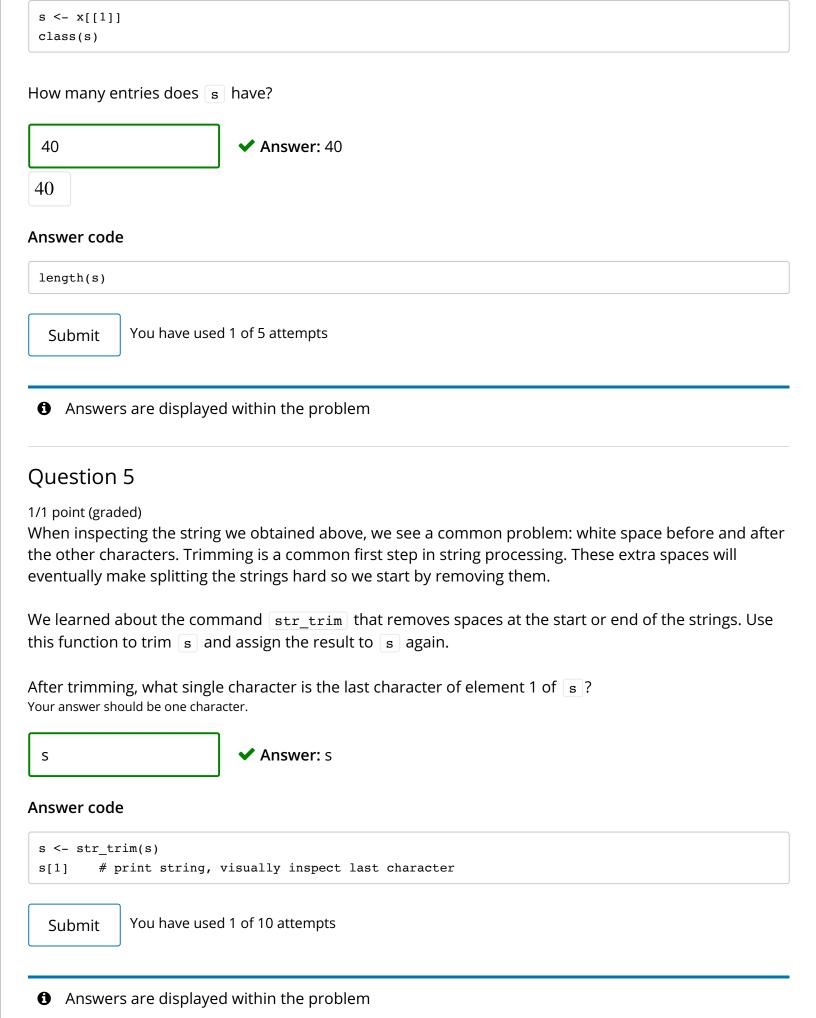
You have used 1 of 2 attempts

1 Answers are displayed within the problem

1/1 point (graded) We are going to create a tidy dataset with each row representing one observation. The variables in this dataset will be year, month, day and deaths.
Use the pdftools package to read in fn using the pdf_{text} function. Store the results in an object called fxt .
Describe what you see in txt.
A table with the mortality data.
 A character string of length 12. Each entry represents the text in each page. The mortality data is in there somewhere. ✓
A character string with one entry containing all the information in the PDF file.
O An html document.
Answer code
<pre>txt <- pdf_text(fn)</pre>
Submit You have used 1 of 2 attempts
Answers are displayed within the problem
Question 3
3.0/3.0 points (graded) Extract the ninth page of the PDF file from the object <code>txt</code> , then use the <code>str_split</code> function from the stringr package so that you have each line in a different entry. The new line character is <code>\n</code> . Call this string vector \mathbf{x} .
Look at x. What best describes what you see?
○ It is an empty string.
○ I can see the figure shown in page 1.

Question 2





Question 6

1/1 point (graded)

We want to extract the numbers from the strings stored in s. However, there are a lot of non-numeric characters that will get in the way. We can remove these, but before doing this we want to preserve the string with the column header, which includes the month abbreviation.

Use the <code>str_which</code> function to find the row with the header. Save this result to <code>header_index</code>. Hint: find the first string that matches the pattern <code>"2015"</code> using the <code>str_which</code> function.

What is the value of header_index ?



Answer code

```
header_index <- str_which(s, "2015")[1]
header_index</pre>
```

Submit

You have used 1 of 10 attempts

1 Answers are displayed within the problem

Question 7

2/2 points (graded)

We want to extract two objects from the header row: month will store the month and header will store the column names.

Save the content of the header row into an object called <code>header</code>, then use <code>str_split</code> to help define the two objects we need.

What is the value of month?

Use header_index to extract the row. The separator here is one or more spaces. Also, consider using the simplify argument.

SEP **✓ Answer:** SEP

Answer code

```
tmp <- str_split(s[header_index], "\\s+", simplify = TRUE)
month <- tmp[1]
header <- tmp[-1]
month</pre>
```

What is the third value in header?
2017 ✓ Answer : 2017
2017
Answer code
header[3]
Submit You have used 1 of 10 attempts
Answers are displayed within the problem
This assessment continues on the next page. Make sure the variable ${\tt s}$ is defined as in the exercises above.

© All Rights Reserved