**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")

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.

It is a report combining graphs and tables. Extracting the data seems possible.

It shows graphs of the data. Extracting the data will be difficult.

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 2**

0/1 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 txt.

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.

An html document.

incorrect

Answer

Incorrect:

Try again. This is not a character string with one entry.

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 3**

3.0/3.0 points (graded)

Extract the ninth page of the PDF file from the object txt, then use the str\_split function from the **stringr** package so that you have each line in a different entry. The new line character is \n. Call this string vector x.

Look at x. What best describes what you see?

It is an empty string.

I can see the figure shown in page 1.

It is a tidy table.

I can see the table! But there is a bunch of other stuff we need to get rid of.

correct

What kind of object is x?

correct

How many entries does x have? correct

1 Loading

You have used 1 of 5 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 4**

2.0/2.0 points (graded)

Define s to be the first entry of the x object.

What kind of object is s?

correct

How many entries does s have? correct

40 Loading

You have used 1 of 5 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)

When inspecting the string we obtained above, we see a common problem: white space before and after the other characters. Trimming is a common first step in string processing. These extra spaces will eventually make splitting the strings hard so we start by removing them.

We learned about the command str\_trim that removes spaces at the start or end of the strings. Use this function to trim s and assign the result to s again.

After trimming, what single character is the last character of element 1 of s?

Your answer should be one character.

correct

You have used 1 of 10 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)

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 str\_which function to find the row with the header. Save this result to header\_index. Hint: find the first string that matches the pattern "2015" using the str\_which function.

What is the value of header\_index? correct

2 Loading

You have used 1 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 7**

0/2 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 header, then use str\_split 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.

incorrect

What is the third value in header? incorrect

2016 Loading

You have used 1 of 10 attempts