Final Assessment due Jun 15, 2021 14:59 +03

This assessment continues from the previous page and assumes that you have defined all of the variables from those questions. In particular, make sure `s` is defined as in the previous exercises.

Question 8

1/1 point (graded)

Notice that towards the end of the page defined by `s` you see a *"Total"* row followed by rows with other summary statistics. Create an object called `tail_index` with the index of the *"Total"* entry.

What is the value of `tail_index`?

35

✔ **Answer:** 35

35

Answer code

```
           tail_index  <- str_which(s, "Total")
tail_index
```

Submit

You have used 1 of 10 attempts

Answers are displayed within the problem

Question 9

1/1 point (graded)

Because our PDF page includes graphs with numbers, some of our rows have just one number (from the y-axis of the plot). Use the `str_count()` function to create an object `n` with the count of numbers in each row.

How many rows have a single number in them?
You can write a regex for a number like this `\\d+`.

2

✔ **Answer:** 2

2

Answer code

```
            n <- str_count(s, "\\d+")
sum(n == 1)
```

Submit

You have used 1 of 10 attempts

Answers are displayed within the problem

Question 10

1/1 point (graded)
We are now ready to remove entries from rows that we know we don't need. The entry `header_index` and everything before it should be removed. Entries for which `n` is 1 should also be removed, and the entry `tail_index` and everything that comes after it should be removed as well.

How many entries remain in `s`?

30

✔ **Answer:** 30

30

Answer code

```
            out <- c(1:header_index, which(n==1), tail_index:length(s))
s <- s[-out]
length(s)
```

Submit | You have used 1 of 10 attempts

Answers are displayed within the problem

Question 11

1/1 point (graded)

Now we are ready to remove all text that is not a digit or space. Do this using regular expressions (regex) and the `str_remove_all()` function.

In regex, using the `^` inside the square brackets `[]` means *not*, like the `!` means not in `!=`. To define the regex pattern to catch all non-numbers, you can type `[^\\d]`. But remember you also want to keep spaces.

Which of these commands produces the correct output?

- ```
  s <- str_remove_all(s, "[^\\d]")
  ```
- ```
  s <- str_remove_all(s, "[\\d\\s]")
  ```
- ```
  s <- str_remove_all(s, "[^\\d\\s]")
  ```
- ```
  s <- str_remove_all(s, "[\\d]")
  ```

Submit | You have used 1 of 2

attempts

Answers are displayed within the problem

Question 12

4/4 points (graded)

Use the `str_split_fixed` function to convert `s` into a data matrix with just the day and death count data:

```
s <- str_split_fixed(s, "\\s+", n = 6)[,1:5]
```

Now you are almost ready to finish. Add column names to the matrix: the first column should be `day` and the next columns should be the `header`. Convert all values to numeric. Also, add a column with the month. Call the resulting object `tab`.

What was the mean number of deaths per day in September 2015?

75.3

✔ **Answer:** 75.3

75.3

Answer code

```
            tab <- s %>%
    as_data_frame() %>%
    setNames(c("day", header)) %>%
    mutate_all(as.numeric)
mean(tab$"2015")
```

What is the mean number of deaths per day in September 2016?

78.9

✔ **Answer:** 78.9

78.9

Answer code

```
mean(tab$"2016")
```

Hurricane María hit Puerto Rico on September 20, 2017. What was the mean number of deaths per day from September 1-19, 2017, before the hurricane hit?

83.7 ✔ **Answer:** 83.7

83.7

Answer code

```
mean(tab$"2017"[1:19])
```

What was the mean number of deaths per day from September 20-30, 2017, after the hurricane hit?

122 ✔ **Answer:** 122

122

Answer code

```
mean(tab$"2017"[20:30])
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Question 13

1/1 point (graded)

Finish it up by changing `tab` to a tidy format, starting from this code outline:

```
      tab <- tab %>% _____(year, deaths, -day) %>%
    mutate(deaths = as.numeric(deaths))
tab
```

What code fills the blank to generate a data frame with columns named "day", "year" and "deaths"?

- ○ `separate`
- ○ `unite`
- ● `gather`
- ○ `spread`

✔

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 14

2/2 points (graded)

Make a plot of deaths versus day with color to denote year. Exclude 2018 since we have no data. Add a vertical line at day 20, the day that Hurricane María hit in 2017.

Which of the following are TRUE?

Check all correct answers.

- ☑ September 2015 and 2016 deaths by day are roughly equal to each other.
- ☐ The day with the most deaths was the day of the hurricane: September 20, 2017.

- [x] After the hurricane in September 2017, there were over 100 deaths per day every day for the rest of the month.
- [x] No days before September 20, 2017 have over 100 deaths per day.

✔

Submit

You have used 1 of 2 attempts

✔ Correct (2/2 points)