

GY7702 CW1

Jacob Embery

05/11/2020

```
library(tidyverse)
```

Question 1:

Create the vector of 25 numbers between 1 and 7 listed below, using the function `c`. Assume those values

represent the answers to a survey question, following the scale presented further below.

```
[1] NA 3 4 4 5 2 4 NA 6 3 5 4 0 5 7 5 NA 5 2 4 NA 3 3 5 NA
```

1 = completely disagree

2 = disagree

3 = somehow disagree

4 = neither agree nor disagree

5 = somehow agree

6 = agree

7 = completely agree

NA = missing value

Question 1.1: write the code necessary to check whether all participants to the survey either completely

disagree or completely agree, once the missing values are excluded.

```
#First make the vector containing all the answers to the question
survey_answer <- c(NA, 3, 4, 4, 5, 2, 4, NA, 6, 3, 5, 4, 0, 5, 7, 5, NA, 5, 2, 4, NA, 3, 3, 5, NA)

#Before doing anything you must remove NA values from the vector
survey_answer = survey_answer[
```

```

!is.na(survey_answer)]

#1 = completely disagree so this asks the vector if all values are equal to 1.
print(paste(
  "Do all participants completely disagree?", all(survey_answer == 1)
))

## [1] "Do all participants completely disagree? FALSE"

#7 = completely agree, same as above.
print(paste(
  "Do all participants completely agree?", all(survey_answer == 7)
))

## [1] "Do all participants completely agree? FALSE"

```

Question 1.2: write the code necessary to extract the indexes related to the participants in the survey who

at least somehow agree or more.

```

#the which function used to extract 5, 6 and 7 (in the vector c) out of the survey_answer vector
which(survey_answer %in% c(5,6,7))

## [1] 4 7 9 12 13 14 15 20

```

Question 2:

Question 2.1: Install the library palmerpenguins.

```

library(palmerpenguins)
library(dplyr)
library(tidyverse)

#install.packages("palmerpenguins")
pengtable <- palmerpenguins::penguins

```

Question 2.2: write the code necessary to create a table showing species, island, bill length and body mass

of the 10 Gentoo penguins in the penguins table with the highest body mass.

```

pengtable %>%
  dplyr::select(species, island, bill_length_mm, body_mass_g) %>%
  dplyr::filter(species == "Gentoo") %>%
  dplyr::slice_max(body_mass_g, n = 10)

## # A tibble: 11 x 4
##   species island bill_length_mm body_mass_g

```

```
##      <fct>   <fct>           <dbl>      <int>
##  1 Gentoo  Biscoe           49.2        6300
##  2 Gentoo  Biscoe           59.6        6050
##  3 Gentoo  Biscoe           51.1        6000
##  4 Gentoo  Biscoe           48.8        6000
##  5 Gentoo  Biscoe           45.2        5950
##  6 Gentoo  Biscoe           49.8        5950
##  7 Gentoo  Biscoe           48.4        5850
##  8 Gentoo  Biscoe           49.3        5850
##  9 Gentoo  Biscoe           55.1        5850
## 10 Gentoo  Biscoe           49.5        5800
## 11 Gentoo  Biscoe           48.6        5800
```

Question 2.3: write the code necessary to create a table showing the average bill length per island, ordered

by average bill length.

```
pengtable %>%
  dplyr::select(bill_length_mm, island) %>%
  dplyr::group_by(island) %>%
  dplyr::mutate(
    avg_bill_length_mm = mean(bill_length_mm)
  ) %>%
  dplyr::arrange(avg_bill_length_mm)
```

```
## # A tibble: 344 x 3
## # Groups:   island [3]
##   bill_length_mm island avg_bill_length_mm
##           <dbl> <fct>           <dbl>
##  1           39.5 Dream             44.2
##  2           37.2 Dream             44.2
##  3           39.5 Dream             44.2
##  4           40.9 Dream             44.2
##  5           36.4 Dream             44.2
##  6           39.2 Dream             44.2
##  7           38.8 Dream             44.2
##  8           42.2 Dream             44.2
##  9           37.6 Dream             44.2
## 10           39.8 Dream             44.2
## # ... with 334 more rows
```

Question 2.4: write the code necessary to create a table showing the minimum, median and maximum

proportion between bill length and bill depth by species.

```
pengtable %>%
  na.omit(pengtable) %>%
  dplyr::select(bill_length_mm, bill_depth_mm, species) %>%
  dplyr::mutate(
```

```

    min_proportion = min(bill_length_mm) / min(bill_depth_mm)
  ) %>%
  dplyr::mutate(
    med_proportion = median(bill_length_mm) / median(bill_depth_mm)
  ) %>%
  dplyr::mutate(
    max_proportion = max(bill_length_mm) / max(bill_depth_mm)
  ) %>%
  dplyr::group_by(species)

```

```

## # A tibble: 333 x 6
## # Groups:   species [3]
##   bill_length_mm bill_depth_mm species min_proportion med_proportion
##           <dbl>         <dbl> <fct>         <dbl>         <dbl>
## 1             39.1           18.7 Adelie           2.45           2.57
## 2             39.5           17.4 Adelie           2.45           2.57
## 3             40.3           18   Adelie           2.45           2.57
## 4             36.7           19.3 Adelie           2.45           2.57
## 5             39.3           20.6 Adelie           2.45           2.57
## 6             38.9           17.8 Adelie           2.45           2.57
## 7             39.2           19.6 Adelie           2.45           2.57
## 8             41.1           17.6 Adelie           2.45           2.57
## 9             38.6           21.2 Adelie           2.45           2.57
## 10            34.6           21.1 Adelie           2.45           2.57
## # ... with 323 more rows, and 1 more variable: max_proportion <dbl>

```

Question 3

Question 3.1: write the code necessary to load the data from covid19_cases_20200301_20201017 to

a variable named covid_data.

the code line below doesn't work in rserver.

I uploaded the file using the upload button in the files tab.

```
read_csv("C:/Users/mirif/OneDrive/Desktop/MASTERS/R/covid19_cases_20200301_20201017.csv")
```

```

#Using this code i can import the uploaded dataset and call it covid_data
library(readr)

covid_data <- read_csv("covid19_cases_20200301_20201017.csv")
#View(covid_data)

```