Practice for Exam, Answers

Misja Mikkers, Gertjan Verhoeven & Has van Vlokhoven

Packages needed

```
library(tidyverse)
library(readxl)
library(writexl)
library(palmerpenguins)
library(gapminder)
```

Introduction

In this notebook we will show you exam questions and show the answers to the questions.

Assignment 0

Pull the repository https://github.com/misjamikkers/Rprogramming21_student to find the data needed for the exam, the exam and the answers.

Assignment 1

Preparation

First we prepare some data and write the data as an excel file

```
set.seed(123)
ex_1a <- tibble(t = seq(0, 2*pi, by = 0.1)) %>%
  mutate(Group = "Group1") %>%
  mutate(Variable_1 = 16 * sin(t)^3) %>%
  mutate(Variable_2 = 13*cos(t) - 5*cos(2*t) - 2*cos(3*t) - cos(4*t))

ex_1b <- tibble(t = seq(0, 2*pi, by = 0.001)) %>%
  mutate(Group = "Group2") %>%
  mutate(Variable_1 = rnorm(n = length(t), mean = 0, sd = 6)) %>%
  mutate(Variable_2 = rnorm(n = length(t), mean = 0, sd = 6))


ex_1 <- rbind(ex_1a, ex_1b)

write_xlsx(ex_1, "../Sourcedata/dataQ1.xlsx")
```

Question 1

In the Sourcedata folder you will find the excelfile with the name dataQ1.xlsx

Your task is to

1. read this data into a R notebook
2. keep the rows of Group1
3. make a scatterplot with Variable_1 on the x-axis and Variable_2 on the y_axis

Which figure do you recognize in the graph?

a. Dinosaur
b. Heart
c. the letter “L”
d. the number “1”

Answer 1

First we will read the data. Please inspect the data. No surprises there

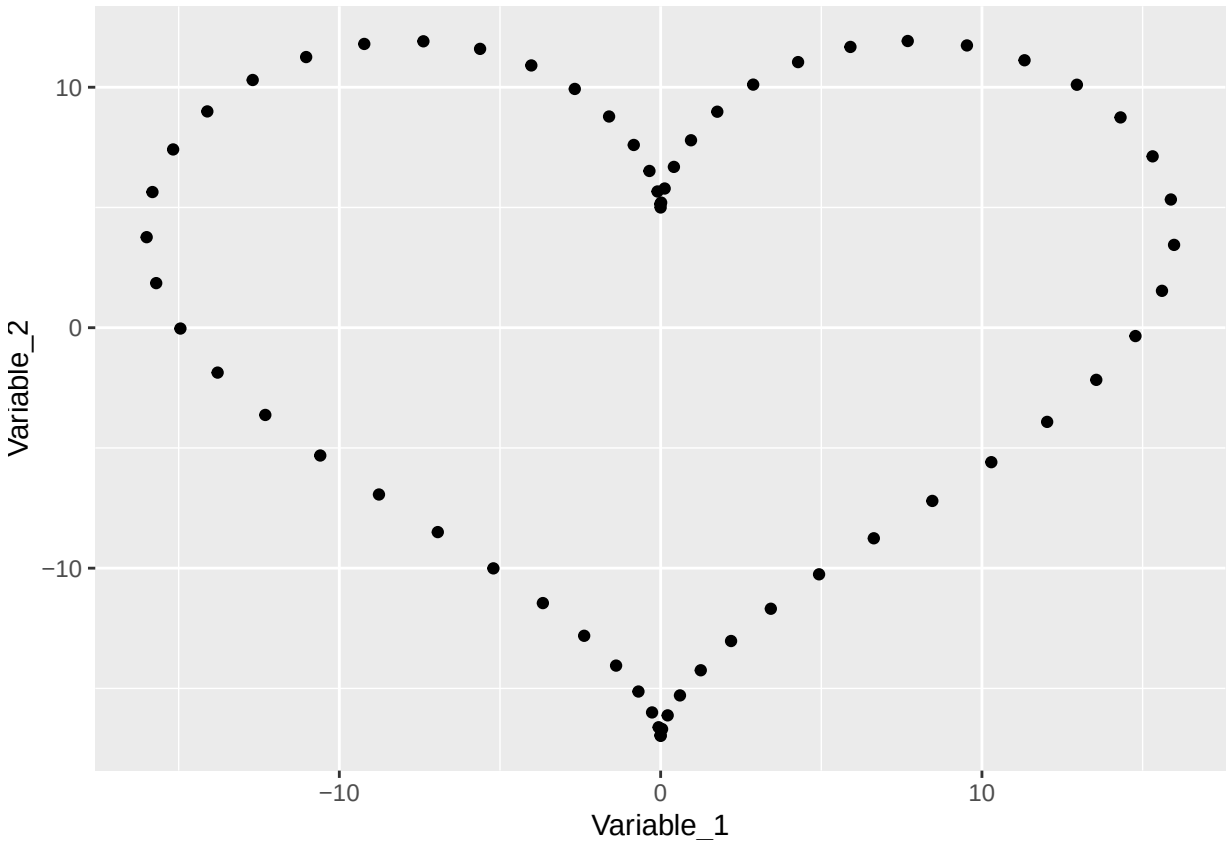
```
q1 <- read_xlsx("../Sourcedata/dataQ1.xlsx")
```

Next we will filter the data

```
q2 <- q1 %>%
  filter(Group == "Group1")
```

And then we can graph

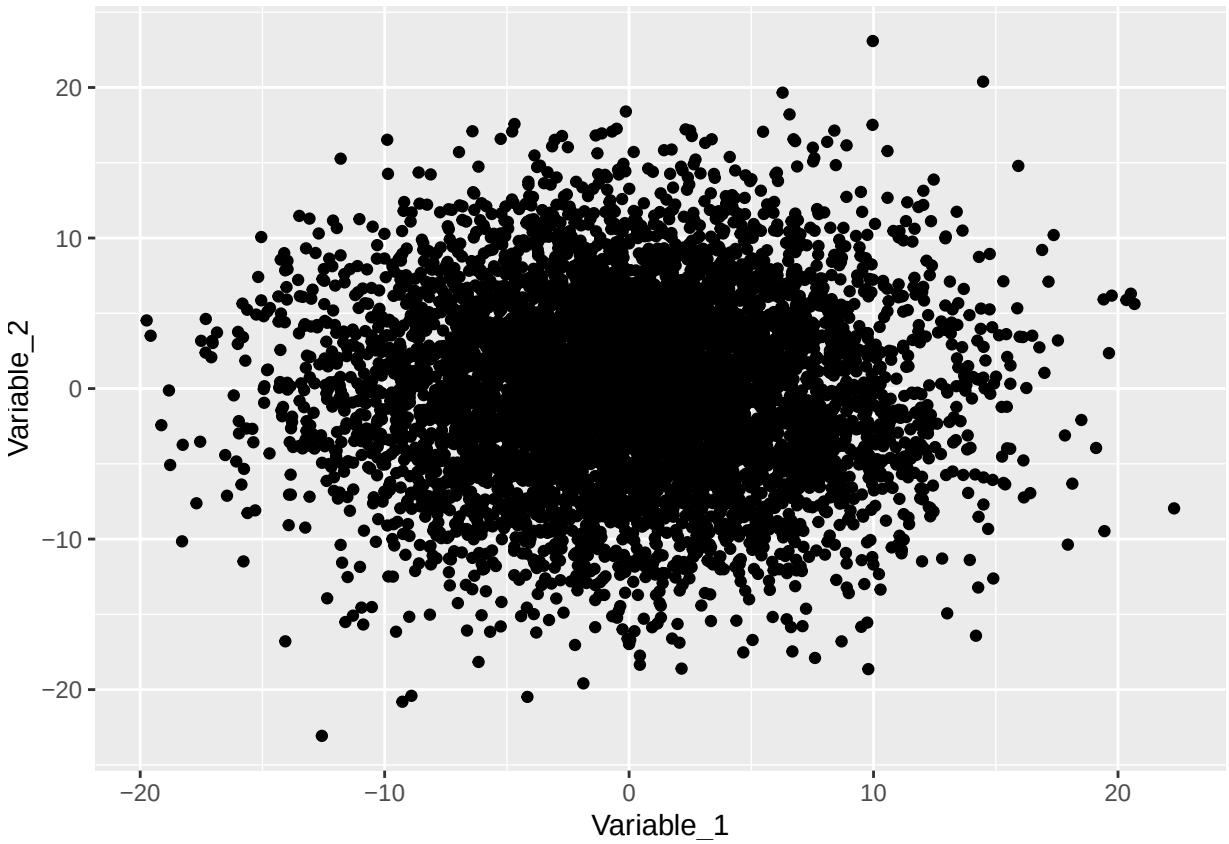
```
ggplot(data = q2, aes(x = Variable_1, y = Variable_2)) +
  geom_point()
```



And see that the answer is b.

However, if you would not filter, the heart would be very hard to detect.

```
ggplot(data = q1, aes(x = Variable_1, y = Variable_2)) +
  geom_point()
```



Assignment 2

Question 2

Your task is to

- use the penguins data from the palmerpenguins package
- keep the rows with year 2009
- create a boxplot with on the x-axis Species per sex and on the y-axis body_mass_g.

Which species per sex has the highest median body mass?

a. females of the Adelie species
b. males of the Adelie species
c. females of the Chinstrap species
d. males of the Chinstrap species
e. females of the Gentoo species
f. males of the Gentoo species
g. the NA group

Answer 2

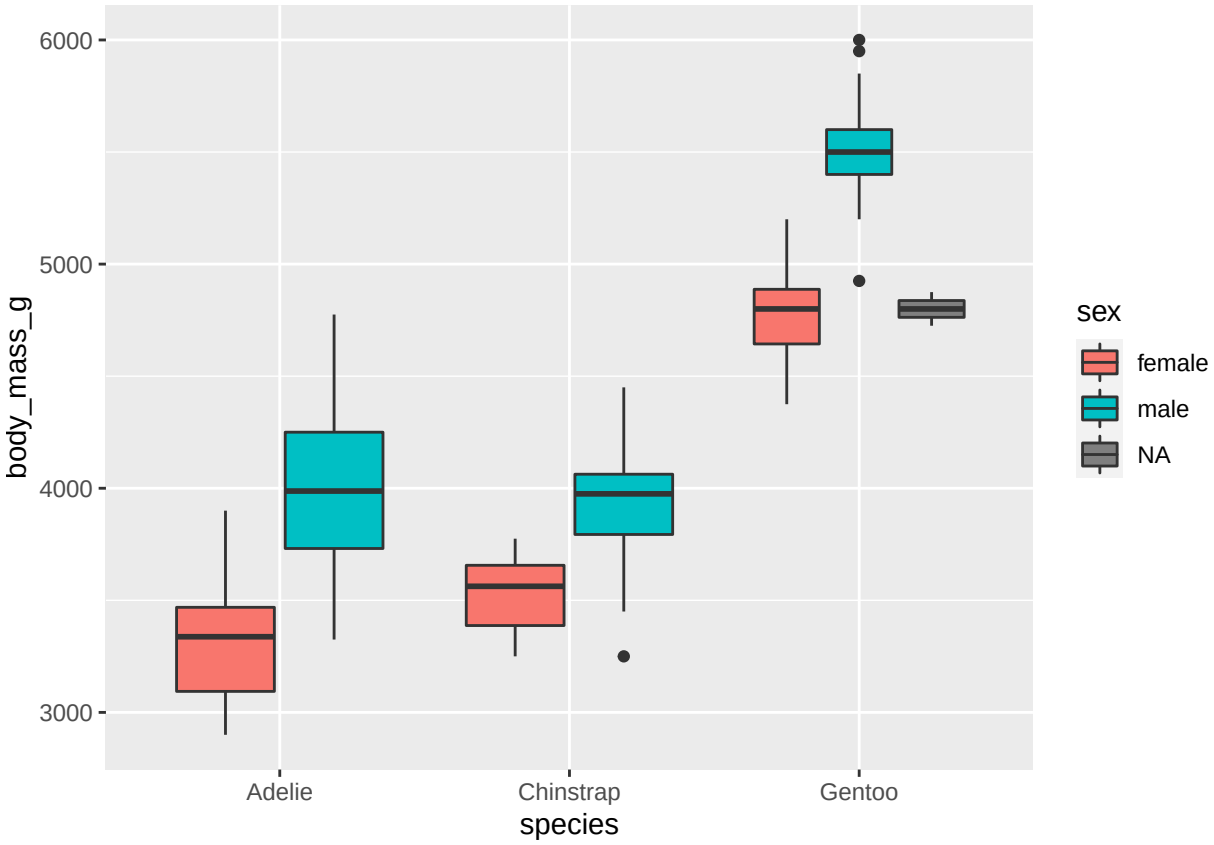
First we load the data and filter for the year 2009.

```r
p1 <- penguins %>%
  filter(year == 2009)
```

Then we can plot the boxplot and look:

```r
ggplot(data = p1, aes(x = species, y = body_mass_g, fill = sex)) +
  geom_boxplot()
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```



The correct answer is f

Assignment 3

Question 3

Your task is

- use the data of the gapminder package
- keep the rows of year 2007
- calculate the medium life expectancy (lifeExp) per continent

What is the median life expectancy of Oceania? Round your answer to the nearest integer.

Answer 3

```
gap <- gapminder%>%
  filter(year == 2007) %>%
  group_by(continent) %>%
  summarise(Median_LE = median(lifeExp))

kableExtra::kable(gap)
```

continent	Median_LE
Africa	52.9265
Americas	72.8990
Asia	72.3960
Europe	78.6085
Oceania	80.7195

The answer is 81.

Assignment 4

Question 4a

Your task is to

- Copy the code below (NOTE: include the set.seed!!!) into a notebook
- Use left_join to merge m1 and m2 (in that order!) in to a dataframe with the name "m3"

How many observations (rows) does m3 have?

Question 4b

Your task is to use m3 from question 4a to

- create a long dataframe m4 with a new colum "Var" that indicates whether the row concerned shows a value of "x" or "y" and the values in column "Val".

So your dataframe should look like:

id	Var	Val
31	x	-0.1639310
31	y	14.3968207
...	...	....
....	...	....

- Then delete all NA's.

How many observations (rows) does m4 have?

```r
set.seed(123)

m1 <- tibble(id = sample(x= 1:100, size = 70, replace = TRUE)) %>%
  mutate(x = rnorm(n = length(id), mean = 0, sd = 1))

m2 <- tibble(id = sample(x= 1:100, size = 80, replace = TRUE)) %>%
  mutate(y = rnorm(n = length(id), mean = 10, sd = 5 ))
```

Answers question 4

```r
m3 <- left_join(m1, m2, by = "id")


m4 <- m3 %>%
  pivot_longer(cols = c(x, y), names_to = "Var", values_to = "Val") %>%
  na.omit()
```

Dataframe m3 has 97 observations.

Dataframe m4 has 166 observations

Assignment 5

Question 5

What does this code do?

```
knitr::opts_chunk$set(echo = FALSE)
```

a. The code makes sure that this code chunk will not be executed
b. The code makes sure that all the code chunks after this command in the Rmd will not be executed
c. The code makes sure that this code chunk will not be shown in the output document (e.g. the pdf)
d. The code makes sure that all the code chunks after this chunk will not be shown in the output document (e.g. the pdf)

Answer 5

The correct answer is d.

Assignment 6

Question 6

Your task is to

- open a notebook

- simulate some data:
  - USE `set.seed(123)` [This is really important!]
  - generate an x-variable in a tibble with a normal distribution with mean 0, and standard deviation 15. We need 1000 observations
  - make a new column y with the formula $sin(x^2) + 33 * x^4$
  - calculate the mean of the variable y.

Answer 6

```
set.seed(123)


fd <- tibble(x= rnorm(n = 1000, mean = 100, sd = 10)) %>%
  mutate(y = log(x) + 33 * x^5) %>%
  summarise(mean_y = mean(y))

kableExtra::kable(fd)
```

mean_y
365927858869

Assignment 7

Create a function that for the numbers x to x + 2 * π calls the function `sin(x)` and adds the results.

The answer should be 0.9878401

```
my_func <- function(x){

  total <- 0

  for (i in x: (x + 2* pi)) {
    total <- total + sin(x)

    print(total)
  }

  total
}

my_func(3)
```

```
## [1] 0.14112
## [1] 0.28224
## [1] 0.42336
## [1] 0.56448
## [1] 0.7056
## [1] 0.84672
## [1] 0.9878401

## [1] 0.9878401
```