Homework 2 by Md Nahian Imtiaz Hasan

Stefan Konigorski

October 26, 2022

Download this R Markdown file, save it on your computer, and perform all the below tasks by inserting your answer in text or by inserting R chunks below. After you are done, upload this file with your solutions on Moodle.

Exercise 1

a) Create an R chunk here to insert R code. Add R code in this R chunk to perform a simple calculation (e.g. calculate the sum of 1 and 2).
b) Create an R chunk with a basic calculation (e.g. 1+1). Try out the different ways how to include this in the knitted report.
c) Knit this Rmd file to html and to pdf.

Perform a simple calculation

```
a=5
b=7
print(a*b)
```

```
## [1] 35
```

Exercise 2: Manipulating variables and data frames

Load the Pima Indian dataset:

```
dat_ex2 <- read.csv(file = url("https://www.dropbox.com/s/tqrauwuxyi03kee/Pima_diabetes.csv?dl=1"))
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

and answer the following questions:

How many women have Glucose levels 0?

Answer = 5

```
table(dat_ex2$Glucose==0)
```

```
##
## FALSE  TRUE
##   763     5
```

How many women have Insulin levels 0?

Answer = 374

```
table(dat_ex2$Insulin==0)
```

```
##
## FALSE  TRUE
##   394   374
```

How many women have both Glucose levels as well as Insulin levels 0?

Answer = 4

```
table(dat_ex2$Glucose==0 & dat_ex2$Insulin==0)
```

```
##
## FALSE  TRUE
##   764     4
```

How many women have either Glucose levels or Insulin levels 0?

Answer = 375

```
table(dat_ex2$Glucose==0 | dat_ex2$Insulin==0)
```

```
##
## FALSE  TRUE
##   393   375
```

How many women have missing BMI values?

Answer = 0

```
table(is.na(dat_ex2$BMI))
```

```
##
## FALSE
##   768
```

How many women have BMI larger than 40?

Answer = 96

```
table(dat_ex2$BMI>40)
```

```
##
## FALSE  TRUE
##   672    96
```

Build a dataset that only includes the women with BMI>40

```
dat_BMI40_vs1 <- dat_ex2[dat_ex2$BMI > 40, ]


# Create a new variable named BMIOutlier, which has the value 0 if a women has BMI smaller or equal 50,
```

```
dat_ex2$BMIOutlier <- 0
dat_ex2$BMIOutlier[dat_ex2$BMI > 50] <- 1
table(dat_ex2$BMIOutlier)
```

```
##
##   0   1
## 760   8
```

Exercise 3 (optional)

Explore merging two datasets.

As a preparation, execute the following code to create different data frames

```
# import data
dat_ex3 <- read.csv(file = url("https://www.dropbox.com/s/tqrauwuxyi03kee/Pima_diabetes.csv?dl=1"))

# extract two smaller data sets
```

```r
dat3_1 <- dat_ex3[1:100, 1:3]
dat3_2 <- dat_ex3[101:300, 1:3]

dat3_3 <- dat_ex3[1:100, 1:3]
dat3_4 <- dat_ex3[1:100, 4:6]
```

Task 3a: Think about how you can use the [.] operator to respectively piece dat3_1 and dat3_2, and dat3_3 and dat3_4 together into one data frame.

Task 3b: Explore the help of the merge() function in R in order to achieve the same goal of combining dat3_3 and dat3_4 together into one data frame. Hint: first create an ID variable in each data frame, then use this in the "by" argument.

Exercise 4 (optional): Times and dates in R

Create an Excel file with 5 observations of 2 variables. Variable 1 is just an ID variable (number 1-5 or character string etc.), and variable 2 is a date/time variable. Use variable 2 to describe the time (and day) you had lunch in the last 5 days. Then try to import the Excel file with both variables into R and/or transform the variables in R to Date or POSIXct variables.