Homework 2

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).

```
a <- 1
b <- 2
c <- a + b
c
```

```
## [1] 3
```

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.

a = 1
b = 2

1. a * b is 2.
2. a + b equals 3.
3. a / b yields 0.5.

c) Knit this Rmd file to html and to pdf.

"both html and pdf files are added with the assignment."

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"))
```

and answer the following questions:

```
# How many women have Glucose levels 0?
table(dat_ex2$Glucose == 0)[2] # [2] means only true values are showing
```

```
## TRUE
##    5
```

```
# How many women have Insulin levels 0?
table(dat_ex2$Insulin == 0)[2]
```

```
## TRUE
##  374
```

```
# How many women have both Glucose levels as well as Insulin levels 0?
table(dat_ex2$Glucose ==0 & dat_ex2$Insulin == 0)[2]
```

```
## TRUE
##    4
```

```
# How many women have either Glucose levels or Insulin levels 0?
sum(dat_ex2$Glucose == 0 | dat_ex2$Insulin ==0)
```

```
## [1] 375
```

```
# How many women have missing BMI values?
sum(is.na(dat_ex2$BMI))
```

```
## [1] 0
```

```
# How many women have BMI larger than 40?
sum(dat_ex2$BMI > 40)
```

```
## [1] 96
```

```
# Build a dataset that only includes the women with BMI>40
dat_n <- dat_ex2[dat_ex2$BMI > 40,]
dat_n <- data.frame(dat_n)
dat_n
```

```
##     Pregnancies Glucose BloodPressure SkinThickness Insulin  BMI
## 5             0     137            40            35     168 43.1
## 17            0     118            84            47     230 45.8
## 19            1     103            30            38      83 43.3
## 42            7     133            84             0       0 40.2
## 44            9     171           110            24     240 45.4
## 46            0     180            66            39       0 42.0
```

```
## 58    0   100   88   60   110 46.8
## 59    0   146   82    0     0 40.5
## 60    0   105   64   41   142 41.5
## 68    2   109   92    0     0 42.7
## 73   13   126   90    0     0 43.4
## 79    0   131    0    0     0 43.2
## 85    5   137  108    0     0 48.8
## 93    7    81   78   40    48 46.7
## 100   1   122   90   51   220 49.7
## 121   0   162   76   56   100 53.2
## 126   1    88   30   42    99 55.0
## 127   3   120   70   30   135 42.9
## 154   1   153   82   42   485 40.6
## 155   8   188   78    0     0 47.9
## 156   7   152   88   44     0 50.0
## 160  17   163   72   41   114 40.9
## 163   0   114   80   34   285 44.2
## 174   1    79   60   42    48 43.5
## 178   0   129  110   46   130 67.1
## 179   5   143   78    0     0 45.0
## 194  11   135    0    0     0 52.3
## 202   1   138   82    0     0 40.1
## 212   0   147   85   54     0 42.8
## 214   0   140   65   26   130 42.6
## 216  12   151   70   40   271 41.8
## 230   0   117   80   31    53 45.2
## 231   4   142   86    0     0 44.0
## 232   6   134   80   37   370 46.2
## 236   4   171   72    0     0 43.6
## 238   0   179   90   27     0 44.1
## 248   0   165   90   33   680 52.3
## 271  10   101   86   37     0 45.6
## 276   2   100   70   52    57 40.5
## 288   1   119   86   39   220 45.6
## 293   2   128   78   37   182 43.3
## 294   1   128   48   45   194 40.5
## 304   5   115   98    0     0 52.9
## 329   2   102   86   36   120 45.5
## 333   1   180    0    0     0 43.3
## 336   0   165   76   43   255 47.9
## 350   5     0   80   32     0 41.0
## 351   4    92   80    0     0 42.2
## 355   3    90   78    0     0 42.7
## 379   4   156   75    0     0 48.3
## 380   0    93  100   39    72 43.4
## 388   8   105  100   36     0 43.3
## 392   5   166   76    0     0 45.7
## 406   2   123   48   32   165 42.1
## 410   1   172   68   49   579 42.4
## 413   1   143   84   23   310 42.4
## 421   1   119   88   41   170 45.3
## 423   0   102   64   46    78 40.6
## 425   8   151   78   32   210 42.9
## 429   0   135   94   46   145 40.6
```

```
## 436           0     141             0             0       0 42.4
## 446           0     180            78            63      14 59.4
## 470           6     154            78            41     140 46.1
## 471           1     144            82            40       0 41.3
## 485           0     145             0             0       0 44.2
## 486           0     135            68            42     250 42.3
## 487           1     139            62            41     480 40.7
## 488           0     173            78            32     265 46.5
## 532           0     107            76             0       0 45.3
## 533           1      86            66            52      65 41.3
## 547           5     187            76            27     207 43.6
## 559          11     103            68            40       0 46.2
## 562           0     198            66            32     274 41.3
## 578           2     118            80             0       0 42.9
## 581           0     151            90            46       0 42.1
## 591          11     111            84            40       0 46.8
## 597           0      67            76             0       0 45.3
## 609           0     152            82            39     272 41.5
## 623           6     183            94             0       0 40.8
## 624           0      94            70            27     115 43.5
## 639           7      97            76            32      91 40.9
## 662           1     199            76            43       0 42.9
## 674           3     123           100            35     240 57.3
## 682           0     162            76            36       0 49.6
## 683           0      95            64            39     105 44.6
## 690           1     144            82            46     180 46.1
## 692          13     158           114             0       0 42.3
## 700           4     118            70             0       0 44.5
## 713          10     129            62            36       0 41.2
## 733           2     174            88            37     120 44.5
## 741          11     120            80            37     150 42.3
## 745          13     153            88            37     140 40.6
## 747           1     147            94            41       0 49.3
## 748           1      81            74            41      57 46.3
## 754           0     181            88            44     510 43.3
## 762           9     170            74            31       0 44.0
##     DiabetesPedigreeFunction Age Outcome
## 5                      2.288  33       1
## 17                     0.551  31       1
## 19                     0.183  33       0
## 42                     0.696  37       0
## 44                     0.721  54       1
## 46                     1.893  25       1
## 58                     0.962  31       0
## 59                     1.781  44       0
## 60                     0.173  22       0
## 68                     0.845  54       0
## 73                     0.583  42       1
## 79                     0.270  26       1
## 85                     0.227  37       1
## 93                     0.261  42       0
## 100                    0.325  31       1
## 121                    0.759  25       1
## 126                    0.496  26       1
```

```
## 127                 0.452  30       0
## 154                 0.687  23       0
## 155                 0.137  43       1
## 156                 0.337  36       1
## 160                 0.817  47       1
## 163                 0.167  27       0
## 174                 0.678  23       0
## 178                 0.319  26       1
## 179                 0.190  47       0
## 194                 0.578  40       1
## 202                 0.236  28       0
## 212                 0.375  24       0
## 214                 0.431  24       1
## 216                 0.742  38       1
## 230                 0.089  24       0
## 231                 0.645  22       1
## 232                 0.238  46       1
## 236                 0.479  26       1
## 238                 0.686  23       1
## 248                 0.427  23       0
## 271                 1.136  38       1
## 276                 0.677  25       0
## 288                 0.808  29       1
## 293                 1.224  31       1
## 294                 0.613  24       1
## 304                 0.209  28       1
## 329                 0.127  23       1
## 333                 0.282  41       1
## 336                 0.259  26       0
## 350                 0.346  37       1
## 351                 0.237  29       0
## 355                 0.559  21       0
## 379                 0.238  32       1
## 380                 1.021  35       0
## 388                 0.239  45       1
## 392                 0.340  27       1
## 406                 0.520  26       0
## 410                 0.702  28       1
## 413                 1.076  22       0
## 421                 0.507  26       0
## 423                 0.496  21       0
## 425                 0.516  36       1
## 429                 0.284  26       0
## 436                 0.205  29       1
## 446                 2.420  25       1
## 470                 0.571  27       0
## 471                 0.607  28       0
## 485                 0.630  31       1
## 486                 0.365  24       1
## 487                 0.536  21       0
## 488                 1.159  58       0
## 532                 0.686  24       0
## 533                 0.917  29       0
## 547                 1.034  53       1
```

```
## 559                    0.126  42       0
## 562                    0.502  28       1
## 578                    0.693  21       1
## 581                    0.371  21       1
## 591                    0.925  45       1
## 597                    0.194  46       0
## 609                    0.270  27       0
## 623                    1.461  45       0
## 624                    0.347  21       0
## 639                    0.871  32       1
## 662                    1.394  22       1
## 674                    0.880  22       0
## 682                    0.364  26       1
## 683                    0.366  22       0
## 690                    0.335  46       1
## 692                    0.257  44       1
## 700                    0.904  26       0
## 713                    0.441  38       1
## 733                    0.646  24       1
## 741                    0.785  48       1
## 745                    1.174  39       0
## 747                    0.358  27       1
## 748                    1.096  32       0
## 754                    0.222  26       1
## 762                    0.403  43       1
```

```r
# Create a new variable named BMIOutlier, which has the value 0 if a women has BMI smaller or equal 50,
dat_ex2["BMIOutlier"] = ifelse(dat_ex2$BMI<=50,0,1)
```

Exercise 3 (optional)

Explore merging two datasets.

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

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

# extract two smaller data sets
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.