

Data Programming Exam

Please reply to the following questions in an R script called “surname_name.R” (e.g., Mario Rossi will return a file named “rossi_mario.R”). Please comment your answers using the symbol: #, before the comments (e.g., # comment).

Exercise 1

- (2pt) Create a vector, named `vec`, containing the following values: 1, 5, 12, 14, 6, 78, 68, 34, 34, 32, 56, 75
- (2pt) Select the 3-rd element of `vec`.
- (2pt) Select all elements of `vec` apart from the 1st.

Exercise 2

Given the following list, named `l1`:

```
l1 <- list(numbers = c(2,3,5,6,7), letters = c("z", "x", "y", "t"))
l1
```

```
## $numbers
## [1] 2 3 5 6 7
##
## $letters
## [1] "z" "x" "y" "t"
```

- (2pt) Extract the element named `letters`.
- (2pt) Extract the object contained in the first element of `l1`.
- (2pt) Add the following element to `l1`:

```
mat <- matrix(1:12, ncol = 3)
```

Exercise 3

- a. (2pt) Generate a data frame, named `df`, corresponding to:

country	population	continent
Italy	59801004	Europe
France	64668129	Europe
China	1382323332	Asia
Japan	126323715	Asia
Chile	18131850	South America
Colombia	48654392	South America
United States	324118787	North America
Canada	36286378	North America
Libya	6330159	Africa
Cameroon	23924407	Africa

Use `data.frame()` function and remember to maintain character vectors as they are, specifying `stringsAsFactors = FALSE`.

- b. (2pt) Select the third row of `df`.
- c. (2pt) Select the 2-nd and 3-rd columns of `df`

Exercise 4

- a. (3 pt) Import the file `2008.txt` using the command `read.table()` in a data frame named `flights`. Open the text file before importing it to control if the first row contains column names and to control the field and the decimal separator characters. Remember to not import the character columns as factors, specifying the argument `stringsAsFactors` as `FALSE`.

This dataset contains information about flight arrival and departure details for all commercial flights within the USA in 2008.

Load `dplyr` library:

```
require(dplyr)
```

- b. (1 pt) Convert `flights` data frame to a `tbl_df` using `tbl_df()` function.
- c. (2 pt) Starting from `flights` data frame, select `ArrDelay` and `Dest` variables and filter the records for which `ArrDelay` variable is greater than 120.
- d. (2 pt) Starting from `flights` data frame, compute the mean delay at departure (`DepDelay` variable) groupig by `Origin` variable. Remember to add `na.rm=TRUE` option to mean computation.

Exercise 5

Load `mtcars` dataset in this way:

```
data("mtcars")
```

`mtcars` data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

```
head(mtcars)
```

```
##           mpg cyl  disp  hp  drat    wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110  3.90  2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21.0   6  160 110  3.90  2.875 17.02  0  1    4    4
## Datsun 710      22.8   4  108  93  3.85  2.320 18.61  1  1    4    1
## Hornet 4 Drive  21.4   6  258 110  3.08  3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175  3.15  3.440 17.02  0  0    3    2
## Valiant         18.1   6  225 105  2.76  3.460 20.22  1  0    3    1
```

To achieve more information about `mtcars` dataset type `?mtcars` on R console.

Load `ggplot2` library:

```
require(ggplot2)
```

- (1 pt) Calculate the number of rows and columns of the `mtcars` dataset.
- (3 pt) Build a scatterplot to analyze the relationship between `mpg` and `wt` variables. Use `ggplot()` and `geom_point()` functions.
- (3 pt) Represent the distribution of `mpg` variable with an histogram. Use `ggplot()` and `geom_histogram()` functions.