

BioSys PhD | Earthsystems PhD - Exploratory Data Analysis

EXERCISES

1. A meteorologist installed a rain gauge (pluviometer) in his backyard to measure precipitation for one year. Monthly totals obtained (in mm) are recorded in the following table:

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
90.2	100.5	70.2	80.6	30.3	10.9	8.2	2.4	3.7	50.8	50.7	99.4

- a) Create vector `prec.tot`, for the 12 monthly totals shown in the table above.
- b) Create vector `month`, with the names of the 12 months.
- c) Associate to each rainfall register, the name of the respective month.
- d) Calculate, using predefined functions of **R**, the following quantities:
 - i. Total annual precipitation;
 - ii. Average monthly precipitation;
 - iii. Median monthly precipitation;
 - iv. Standard deviation of monthly precipitation;
 - v. Minimum and maximum monthly precipitation.
- e) Build, from vector `prec.tot`, the sub-vector of rainfall in the summer months (June to September).
- f) Build, from vector `prec.tot`, the sub-vector of months with precipitation above the average.
- g) Identify, using **R** commands, the month where there was the minimum total precipitation.
- h) Represent graphically the 12 monthly rainfall totals. Include the convenient legend and use colors. Comment the result.
- i) Execute command `boxplot(prec.tot,col='orange')`. Identify the resulting graphical representation and make a comment.

2. Twenty five people were asked about their favourite brand of beer. The available brands are: (1) BeerDear, (2) BlondieBeer, (3) ForeignBeer (4) LoveBeer.

3 4 1 1 3 4 3 3 1 3 2 1 2 1 2 3 2 3 1 1 1 1 4 3 1

- (a) Construct a bar plot:

- with the relative frequencies;
- with the absolute frequencies;
- insert a title and other details you consider important.

- (b) Construct a circular plot:

- including the brands' names (argument `label`);
- with four colors of your preference.

3. A study was carried out to assess whether the students who smoke study less. The data collected are in the following table:

Student	Smokes?	Study hours (week)
1	Y	less than 5 hours
2	N	5 - 10 hours
3	N	5 - 10 hours
4	Y	more than 10 hours
5	Y	more than 10 hours
6	Y	less than 5 hours
7	Y	5 - 10 hours
8	N	less than 5 hours
9	N	more than 10 hours
10	Y	5 - 10 hours

- (a) Create vector `smoke`, constituted by "Y" and "N", and vector `hours`, composed by categories 1, 2 and 3, for *less than 5 hours*, *5-10 hours* and *more than 10 hours*, respectively. View the summary table.
- (b) Use the function `barplot()` to represent graphically the hours of study in groups of smokers and non-smokers. Include legend and title.
- (c) Place side-by-side bars for each group.