

Assignment 2 - Histograms, Medians, Means, Boxplots and Standard Deviation. Due September 21, 11:59pm 2018

EPIB607 - Inferential Statistics^a

^aFall 2018, McGill University

This version was compiled on September 11, 2018

The first step in understanding data is to hear what the data say, to “let the data speak for themselves”. Numbers speak clearly only when we help them speak by organizing, displaying, and summarizing. In this assignment you will explore how to visualize your data and summarize it using summary statistics. All graphs and calculations are to be completed in an R Markdown document using the provided template. Please submit both the compiled HTML report and the source file (.Rmd) to myCourses by September 21, 2018, 11:59pm. Both HTML and .Rmd files should be saved as ‘IDnumber_LastName_FirstName_EPIB607_A2’.

Histograms | Means | Medians | Boxplots | Standard Deviation | Summary statistics | mosaic package

Template

The .Rmd template for Assignment 2 is available [here](#)

1. The mosaic package (optional)

The mosaic package provides a consistent and user-friendly interface for descriptive statistics, plots and inference. You may find it useful to complete an interactive tutorial on its plotting functions. (note: this is optional and will not be counted for any marks). First install the following packages:

```
install.packages(c("learnr", "mosaic"),
                 dependencies = TRUE)
```

Then, from RStudio, run the following command which will open a new page in your web browser:

```
learnr::run_tutorial("introduction",
                    package = "ggformula")
```

A more advanced tutorial on customizing your plots is available also:

```
learnr::run_tutorial("refining",
                    package = "ggformula")
```

2. Age-structures of Populations, then and now

Please try to complete this step before class on Thursday September 6:

1. Sign up for a free DataCamp account at [this link](#). Note: you are required to sign up with a @mail.mcgill.ca or @mcgill.ca email address.
2. Sign in to <https://github.mcgill.ca/> using your McGill email address and corresponding password.

```
dat <- read.table("~/git_repositories/epib607/data/age_sex_frequ
library(dplyr)
dat <- dat %>% mutate(Male = case_when(Male == 1 ~ "Male", Male
colnames(dat) <- c("Gender", "Age", "Freq")
write.csv(dat, file = "data/age_sex_frequencies_ireland.csv", qu

library(rmarkdown)
draft("mypaper.Rmd", template="pdf", package="pinp", edit=TRUE)
```

3. Install Git

You need to first install the [git](#) version control system on your system. Follow Chapter 1: Installing Git at [this link](#) for step-by-step installation instructions with screenshots.

4. Install R and RStudio

This short course will guide you through installing both R and RStudio. RStudio is a software application that facilitates how you interact with R. Click on the link the in the figure caption join the course.



Fig. 1. Working with the RStudio IDE (Part 1)

5. Introduction to R

In this course you will get a hands-on introduction to the basic commands in R. With the knowledge gained in this course, you will be ready to perform a data analysis. Click on the link the in the figure caption join the course.



Fig. 2. Introduction to R

6. R Markdown

You will learn how to create reproducible reports using R and Markdown. All assignments for this course must be submitted in this format. Click on the link the in the figure caption join the course.



Fig. 3. Reporting with R Markdown

7. GitHub using RStudio

You will learn how to use RStudio to version control your code. All assignments for this course must be submitted to a GitHub repository. Click on the link the in the figure caption join the course (Chapter 2 only).



Fig. 4. Version Control with RStudio IDE (Chapter 2 only)