JJiang_HW3

Joe Jiang

November 22, 2018

Question 1: Clone a repo from GITHUB

Screening shot showing how to clone a repo from GITHUB

Question 2

```
df <- read.csv("titanic.csv")
```

```
sum(df$Sex=="male")
```

```
## [1] 577
```

```
sum(df$Sex=="female")
```

```
## [1] 314
```

```
plot(df$Sex, main = "Frequency of male and female passenges of Titanic", xlab = "Passenges", ylab = "Fr
```

```
apply(df[c("Survived", "Age", "Fare")], 2, mean, na.rm = TRUE)
```

```
##   Survived        Age       Fare
##  0.3838384 29.6991176 32.2042080
```

Question 3

```
#Load the sleep data file into R
df <- read.csv("http://talklab.psy.gla.ac.uk/L1_labs/lab_1/homework/sleep_data_01.csv")

func <- function(x){
  medianAge = median(x$Age, na.rm = TRUE)
  minSleep = min(x$Duration, na.rm = TRUE)
  maxSleep = max(x$Duration, na.rm = TRUE)
  meanRSES = mean(x$RSES, na.rm = TRUE)
  sdRSES = sd(x$RSES, na.rm = TRUE)

  report <- data.frame(medianAge, meanRSES/5, sdRSES/5, maxSleep-minSleep)
  colnames(x) <- c("MedianAge", "SelfEsteem", "SE_SD","DurationRange")
  round(report, 2)
}

#Run the function
func(df)
```

```
##   medianAge meanRSES.5 sdRSES.5 maxSleep...minSleep
## 1        14       3.62     1.24                   7
```

Question 4

```
# install fivethirtyeight package and load the library
#install.packages("fivethirtyeight")
```

```
$ pwd
/c/Users/Hannah/Desktop/Courses/Doing Data Science/Unit3/testrepo

Hannah@JoeJiang  ~/Desktop/Courses/Doing Data Science/Unit3/testrepo (master)
$ cd ..

Hannah@JoeJiang  ~/Desktop/Courses/Doing Data Science/Unit3
$ ls
DDSUnit3Winter.pdf  HW3.pdf  __MACOSX  sleep_data_01.csv  testrepo

Hannah@JoeJiang  ~/Desktop/Courses/Doing Data Science/Unit3
$ mkdir homework

Hannah@JoeJiang  ~/Desktop/Courses/Doing Data Science/Unit3
$ cd homework

Hannah@JoeJiang  ~/Desktop/Courses/Doing Data Science/Unit3/homework
$ git clone  https://github.com/caesar0301/awesome-public-datasets
Cloning into 'awesome-public-datasets'...
remote: Enumerating objects: 3, done.
remote: Counting objects: 100% (3/3), done.
remote: Compressing objects: 100% (3/3), done.
remote: Total 1487 (delta 0), reused 1 (delta 0), pack-reused 1484
Receiving objects: 100% (1487/1487), 585.14 KiB | 204.00 KiB/s, done.
Resolving deltas: 100% (866/866), done.

Hannah@JoeJiang  ~/Desktop/Courses/Doing Data Science/Unit3/homework
$ ls
awesome-public-datasets
```

Figure 1: Fig. 1 ScreenShot showing how to create a new dir and navigate into the right dir and clone a repo from GITHUB

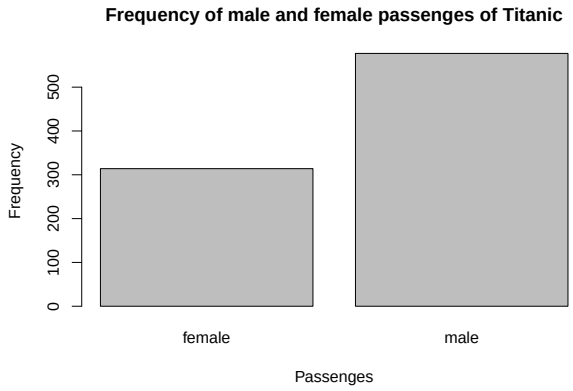

Figure 2: Fig. 2 Plot showing the frequency of male and feamle passengers on Titanic.

```
#library("fivethirtyeight")

# get title of the 22nd datasets in the library and assign the dataset to df
#data(package = 'fivethirtyeight')$results[22,'Item']
#df <- college_recent_grads

# get the URL from a more detailed list of datasets for a link to a related news story
#vignette("fivethirtyeight", package = "fivethirtyeight")

# The URL for the related news story: http://fivethirtyeight.com/features/the-economic-guide-to-picking
```

```
dim(df)
```

```
## [1] 101  15
```

```
colnames(df)
```

```
##  [1] "SubjID"   "Age"      "Gender"   "PSQI"     "PSAScog"  "FOMO"
##  [7] "SMUISa"   "SM1"      "SM2"      "HADSa"    "HADSd"    "RSES"
## [13] "Bedtime"  "Latency"  "Duration"
```

Question 5

```
# Column names and the number of columns
colnames(df)
```

```
##  [1] "SubjID"   "Age"      "Gender"   "PSQI"     "PSAScog"  "FOMO"
##  [7] "SMUISa"   "SM1"      "SM2"      "HADSa"    "HADSd"    "RSES"
## [13] "Bedtime"  "Latency"  "Duration"
```

```
length(colnames(df))
```

```
## [1] 15
```

```
#load the dplyr and use the function count()
#library(dplyr)
#major_count <- count(df, df$major_category)
```