Lab 1

Sharleen Price spp2122

January 18, 2017

Instructions

Before you leave lab today make sure that you upload an RMarkdown file to the Canvas page (this should have a .Rmd extension) as well as the PDF output (or HTMF) after you have knitted the file (this will have a .pdf or .html extension). Note that since you have already knitted this file, you should see both a **Lab1_UNI.pdf** and a **Lab1_UNI.Rmd** file in your UN2102 Lab1 folder. Click on the **Files** tab to the right to see this. The files you upload to the Canvas page should be updated with commands you provide to answer each of the questions below. You can edit this file directly to produce your final solutions. Please do not waste space by printing the dataset or any vector over, say, length 20.

Goals

The goals of this lab are to perform some basic tasks using **R** and **R Markdown**. The primary goal is to guarantee that every student is successfully knitting a markdown file. Secondary goals include; (1) uploading a dataset, (2) looking at the **head** of a dataset, (3) investigating the structure of a dataset, (4) assigning variables names, (5) perform a basic subsetting task, and (6) creating a basic scatter plot.

Background: Strike's Dataset

We consider a dataset on 18 countries over 35 years (compiled by Bruce Western, in the Sociology Department at Harvard University). The measured variables are:

- `country`, `year`: country and year of data collection
- `strike.volume`: days on strike per 1000 workers
- `unemployment`: unemployment rate
- `inflation`: inflation rate
- `left.parliament`: leftwing share of the government
- `centralization`: centralization of unions
- `density`: density of unions

Tasks

1) Create a folder on your desktop (or wherever) labeled Lab1. Inside the folder you should have the **Lab1.Rmd** file and the **strikes.csv** dataset.

2) Uncomment and run the following code. Briefly explain what the two functions are doing.

```
#-- R code goes here ----
strikes <- read.csv("strikes.csv", as.is = TRUE)
dim(strikes)
```

```
## [1] 625   8
```

the read.csv() function reads in the entire csv file and assigns it to the variable strikes

the dim() function returns the dimensions of the csv file

3) Look at the first 4 observations of this dataframe using the **head** function. To investigate the **head** function, run the code **?head**.

```
#-- R code goes here ----
head(strikes)
```

```
##     country year strike.volume unemployment inflation left.parliament
## 1 Australia 1951           296          1.3      19.8            43.0
## 2 Australia 1952           397          2.2      17.2            43.0
## 3 Australia 1953           360          2.5       4.3            43.0
## 4 Australia 1954             3          1.7       0.7            47.0
## 5 Australia 1955           326          1.4       2.0            38.5
## 6 Australia 1956           352          1.8       6.3            38.5
##   centralization density
## 1      0.3748588      NA
## 2      0.3751829      NA
## 3      0.3745076      NA
## 4      0.3710170      NA
## 5      0.3752675      NA
## 6      0.3716072      NA
```

```
?head
```

4) Look at the structure of the **strikes** dataset using the **str** function.

```
#-- R code goes here ----
str(strikes)
```

```
## 'data.frame':    625 obs. of  8 variables:
##  $ country        : chr  "Australia" "Australia" "Australia" "Australia" ...
##  $ year           : int  1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 ...
##  $ strike.volume  : int  296 397 360 3 326 352 195 133 109 208 ...
##  $ unemployment   : num  1.3 2.2 2.5 1.7 1.4 1.8 2.3 2.7 2.6 2.5 ...
##  $ inflation      : num  19.8 17.2 4.3 0.7 2 6.3 2.5 1.3 1.8 3.8 ...
##  $ left.parliament: num  43 43 43 47 38.5 38.5 38.5 36.9 36.9 36.9 ...
##  $ centralization : num  0.375 0.375 0.375 0.371 0.375 ...
##  $ density        : num  NA NA NA NA NA NA NA NA NA 50.2 ...
```

5) Run the following code and briefly describe what the **summary** function is doing.

```
#-- R code goes here ----
summary(strikes)
```

```
##    country               year      strike.volume     unemployment
##  Length:625         Min.   :1951   Min.   :   0.0   Min.   : 0.000
##  Class :character   1st Qu.:1959   1st Qu.:  19.0   1st Qu.: 1.200
##  Mode  :character   Median :1968   Median : 127.0   Median : 2.500
##                     Mean   :1968   Mean   : 288.7   Mean   : 3.555
```

```
##                    3rd Qu.:1977   3rd Qu.: 360.0   3rd Qu.: 5.500
##                    Max.   :1985   Max.   :5918.0   Max.   :17.000
##
##    inflation      left.parliament centralization        density
##  Min.   :-2.900   Min.   : 8.16   Min.   :0.000005   Min.   :13.60
##  1st Qu.: 2.700   1st Qu.:32.20   1st Qu.:0.248274   1st Qu.:32.52
##  Median : 4.800   Median :42.50   Median :0.379830   Median :42.00
##  Mean   : 5.957   Mean   :40.85   Mean   :0.456375   Mean   :44.98
##  3rd Qu.: 8.200   3rd Qu.:49.70   3rd Qu.:0.749203   3rd Qu.:58.10
##  Max.   :27.500   Max.   :78.70   Max.   :0.999788   Max.   :81.30
##                                                      NA's   :179
```

```
?summary
```

The summary function groups the variables and calculates various statistical results such as mean, median, etc for the variables containing numbers. When the variable contains strings it calculates the length class and mode of the objects.

6) Create a logical vector of length 625 (same number of rows in the strikes dataset) which gives a **TRUE** when the country corresponds to Switzerland and **FALSE** otherwise. Assign the logical vector as **Switzerland.logical**. How many cases correspond are measure on Switzerland?

```
#-- R code goes here ----
Switzerland.logical <- (strikes$country == "Switzerland")
table(Switzerland.logical)
```

```
## Switzerland.logical
## FALSE  TRUE
##   590    35
```

There are 35 cases that corrsponds to Switzerland

7) Create a new sub-dataset (of dataframe) that consists only of the cases corresponding to Switzerland. Call this dataframe *Switzerland.strikes**. Also display the head of** Switzerland.strikes** and identify how many rows are in this new dataset?

```
#-- R code goes here ----
Switzerland.strikes <- subset(strikes, country == "Switzerland")
dim(Switzerland.strikes)
```

```
## [1] 35  8
```

There are 35 rows in this new dataset

8) Create a time-series plot (or scatter plot) that shows Switzerland's inflation rate as a function of time. The code is explicitly given below. Change the title of the plot and label the axes appropriately.

```
#-- R code goes here ----
plot(Switzerland.strikes$year,Switzerland.strikes$inflation,
```

```
      main="Switzerland's Inflation Rate vs Time", xlab="Time",ylab="Inflation Rate",
      col="blue",type="l")
```