R Competency

Milana Stetsenko

Spetember 2020

Step 1

I have downloaded the data on my computer and improted it into the R Studio as a csv file. I have also removed any empty values and converted them to N/As. Then, I have checked the dimensions of the data to see what it looks like and to get an understanding of the rows and columns so then I could see whether any changes and trims occured.

```
rm(list=ls())
gender_data <- read.csv("/Users/milanastetsenko/Documents/Coding/R/R Assignments/R Competency/odp_contri
dim(gender_data)
```

```
## [1] 147631      9
```

```
head(gender_data)
```

```
##   Contribution_ID ISOCode3 M49_Code Contributing_Country Mission_Acronym
## 1          427903      DZA       12              Algeria         MONUSCO
## 2          427904      ARG       32            Argentina         MINURSO
## 3          427905      ARG       32            Argentina         UNFICYP
## 4          427906      ARG       32            Argentina         UNFICYP
## 5          427907      ARG       32            Argentina          UNMISS
## 6          427908      ARG       32            Argentina           UNTSO
##       Personnel_Type Female_Personnel Male_Personnel Last_Reporting_Date
## 1 Experts on Mission                0              2          31/07/2020
## 2 Experts on Mission                0              2          31/07/2020
## 3             Troops               15            219          31/07/2020
## 4      Staff Officer                3              6          31/07/2020
## 5  Individual Police                1              5          31/07/2020
## 6 Experts on Mission                0              4          31/07/2020
```

Step 2

I am now omititng all the rows with the N/A values so we could have a full analysis without any missing data in it for easier inference. As we can see from the difference in the dim(gender_data) and dim (gender_data1), we have ommited 5 rows

```
gender_data1<- na.omit(gender_data)
```

Step 3

I am converting the class of the data column to the date class. I had some issues and the date was outputting NA so i changed the sustem locale and it worked.

```r
class(gender_data1$Last_Reporting_Date)
```

```
## [1] "character"
```

```r
gender_data1$Last_Reporting_Date<- as.Date(gender_data1$Last_Reporting_Date, "%d/%m/%Y")

class(gender_data1$Last_Reporting_Date)
```

```
## [1] "Date"
```

```r
head(gender_data1)
```

```
##   Contribution_ID ISOCode3 M49_Code Contributing_Country Mission_Acronym
## 1          427903      DZA       12              Algeria         MONUSCO
## 2          427904      ARG       32            Argentina         MINURSO
## 3          427905      ARG       32            Argentina         UNFICYP
## 4          427906      ARG       32            Argentina         UNFICYP
## 5          427907      ARG       32            Argentina          UNMISS
## 6          427908      ARG       32            Argentina           UNTSO
##       Personnel_Type Female_Personnel Male_Personnel Last_Reporting_Date
## 1 Experts on Mission                0              2          2020-07-31
## 2 Experts on Mission                0              2          2020-07-31
## 3             Troops               15            219          2020-07-31
## 4      Staff Officer                3              6          2020-07-31
## 5  Individual Police                1              5          2020-07-31
## 6 Experts on Mission                0              4          2020-07-31
```

Step 4

Here I will analyze whether the goal of 20+% of women serving on mission has been achieved in July 2019. I will create a subset of the dataset where the last reporting date is in July, then I will sum the number of women serving in Formed Police Units and the number of men (this tep is optional I could just divide it by the total sum of both), and then will calucate the precentage of women out of all the units.

```r
july_2020<-subset(gender_data1, gender_data1$Last_Reporting_Date =="2020-07-31")

#women_police_units <- gender_data1$Personnel_Type == "Formed Police Units"
women_police_units<-sum(july_2020[which(july_2020$Personnel_Type== "Formed Police Units"),7])
women_police_units
```

```
## [1] 741
```

```r
men_police_units<-sum(july_2020[which(july_2020$Personnel_Type== "Formed Police Units"),8])
men_police_units
```

```
## [1] 6039
```

```r
percentage_women<-(women_police_units/(men_police_units+women_police_units))*100
percentage_women
```

```
## [1] 10.9292
```

The percentage turns out to be 10.9% so their goal was not achieved but the number of women serving is increasing and hopefully, they will achieve their goal by 2028.

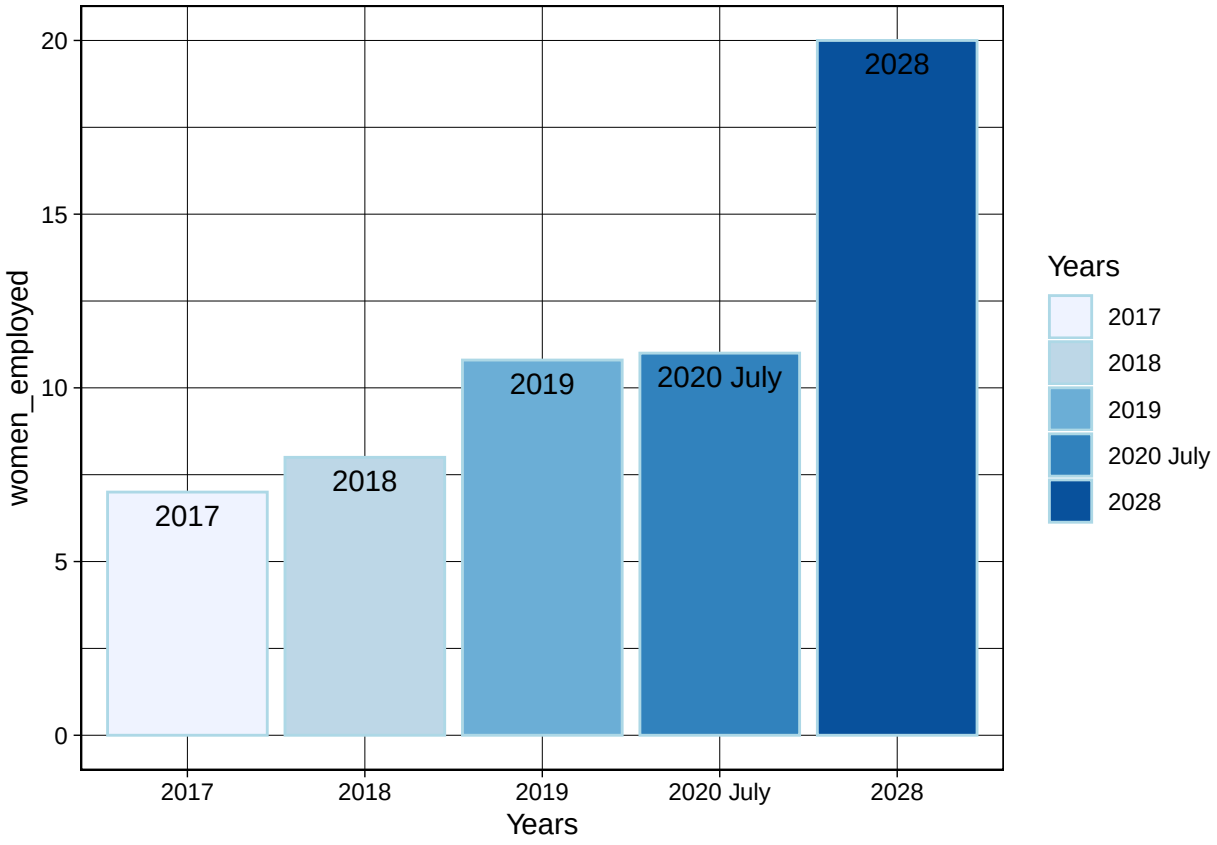
Step 5

Here, I will plot the outcomes of the step 4 together with the other prediction and exisitng data provided by UN.

```
#creating a dataframe with necessary values
Years<-c('2017', '2018', '2019', '2020 July', '2028')
women_employed<-c(7,8,10.8,11,20)
percentage_rate<- data.frame(Years, women_employed)
percentage_rate
```

```
##       Years women_employed
## 1      2017            7.0
## 2      2018            8.0
## 3      2019           10.8
## 4 2020 July           11.0
## 5      2028           20.0
```

```
p<-ggplot(data = percentage_rate, aes(x = Years, y = women_employed, fill = Years)) +
  geom_bar(stat="identity", color = "lightblue") +
  geom_text(aes(label=Years), vjust=1.6, color = 'black', size=4)+
  theme_linedraw()
p+scale_fill_brewer()
```



Step 6

I am trying to calculate country-dependent unique missions in Minerva cities. Here we can see that in Deutschland, we have 24 unique missions and they are also liste din the form of the list. You can cahnge the

input to the function to learn about other Minerva countries.

```
gender_data1$ISOCode3<- trimws(gender_data1$ISOCode3,which='both', whitespace = "[ \t\r\n]")
minerva_missions<-function(ISO){
  unique_missions<- list(unique(factor((gender_data1[which(gender_data1$ISOCode3 == ISO), 5]))))
  output<- append(unique_missions, lengths(unique_missions))
  return (output)
}
minerva_missions("DEU")
```

```
## [[1]]
##  [1] MINURSO
##  [2] MINUSMA
##  [3] UNAMID
##  [4] UNIFIL
##  [5] UNMIK
##  [6] UNMISS
##  [7] UNSOM
##  [8] UNMHA
##  [9] MINUJUSTH
## [10] UNSMIL
## [11] UNMIL
## [12] UNAMA
## [13] MINUSTAH
## [14] UNMIS
## [15] UNMIS
## [16] UNIFIL
## [17] UNMIL
## [18] UNAMID
## [19] UNMIK
## [20] UNOMIG
## [21] UNAMSIL
## [22] MINURSO
## [23] UNIKOM
## [24] UNMIBH
## 24 Levels: MINUJUSTH ... UNSOM
##
## [[2]]
## [1] 24
```

```
minerva_countries <- c("USA", "DEU", "KOR", 'IND', "ARG", "GBR")
for (country in minerva_countries) {
  print(country)
  print(minerva_missions(country))
  }
```

```
## [1] "USA"
## [[1]]
##  [1] BINUH
##  [2] MINUSCA
##  [3] MINUSMA
##  [4] MONUSCO
##  [5] UNMISS
##  [6] UNSMIL
##  [7] UNTSO
##  [8] MINUJUSTH
```

```
##  [9] UNMIL
## [10] MINUSTAH
## [11] UNAMA
## [12] MINURCAT
## [13] UNMIS
## [14] MINUSTAH
## [15] UNMIL
## [16] UNMIK
## [17] UNTSO
## [18] UNIOSIL
## [19] UNAMID
## [20] UNMIT
## [21] UNOTIL
## [22] UNAMSIL
## [23] UNMISET
## [24] UNIKOM
## [25] UNMEE
## [26] UNMIBH
## 26 Levels: BINUH ... UNTSO
##
## [[2]]
## [1] 26
##
## [1] "DEU"
## [[1]]
##  [1] MINURSO
##  [2] MINUSMA
##  [3] UNAMID
##  [4] UNIFIL
##  [5] UNMIK
##  [6] UNMISS
##  [7] UNSOM
##  [8] UNMHA
##  [9] MINUJUSTH
## [10] UNSMIL
## [11] UNMIL
## [12] UNAMA
## [13] MINUSTAH
## [14] UNMIS
## [15] UNMIS
## [16] UNIFIL
## [17] UNMIL
## [18] UNAMID
## [19] UNMIK
## [20] UNOMIG
## [21] UNAMSIL
## [22] MINURSO
## [23] UNIKOM
## [24] UNMIBH
## 24 Levels: MINUJUSTH ... UNSOM
##
## [[2]]
## [1] 24
##
```

```
## [1] "KOR"
## [[1]]
##  [1] MINURSO
##  [2] UNAMID
##  [3] UNIFIL
##  [4] UNMISS
##  [5] UNMOGIP
##  [6] UNMHA
##  [7] MINUJUSTH
##  [8] UNMIL
##  [9] UNOCI
## [10] MINUSTAH
## [11] UNMIT
## [12] UNISFA
## [13] UNMIS
## [14] UNMIN
## [15] UNAMA
## [16] UNMIS
## [17] UNMIL
## [18] UNIFIL
## [19] UNAMID
## [20] UNMIT
## [21] MINURSO
## [22] UNMISET
## [23] UNFICYP
## 23 Levels: MINUJUSTH ... UNOCI
##
## [[2]]
## [1] 23
##
## [1] "IND"
## [[1]]
##  [1] MINURSO
##  [2] MONUSCO
##  [3] UNDOF
##  [4] UNFICYP
##  [5] UNIFIL
##  [6] UNISFA
##  [7] UNMISS
##  [8] UNTSO
##  [9] UNSOM
## [10] MINUJUSTH
## [11] MINUSTAH
## [12] UNMIL
## [13] UNAMA
## [14] UNOCI
## [15] UNAMI
## [16] UNMIT
## [17] UNMIS
## [18] MONUC
## [19] UNDOF
## [20] UNFICYP
## [21] UNIFIL
## [22] MONUC
```

```
## [23] MINUSTAH
## [24] UNMIT
## [25] UNMIL
## [26] UNMIS
## [27] UNMIK
## [28] UNIOSIL
## [29] UNMEE
## [30] UNOCI
## [31] ONUB
## [32] UNOMIG
## [33] UNAMSIL
## [34] MINURSO
## [35] UNIKOM
## [36] UNMIBH
## 36 Levels: MINUJUSTH ... UNTSO
##
## [[2]]
## [1] 36
##
## [1] "ARG"
## [[1]]
##  [1] MINURSO
##  [2] UNFICYP
##  [3] UNMISS
##  [4] UNTSO
##  [5] UNVMC
##  [6] MINUSCA
##  [7] MINUJUSTH
##  [8] UNAMI
##  [9] MINUSTAH
## [10] UNMC
## [11] UNOCI
## [12] UNMIL
## [13] UNMIS
## [14] UNMIL
## [15] UNFICYP
## [16] UNOCI
## [17] MINUSTAH
## [18] MONUC
## [19] UNMIK
## [20] UNMIS
## [21] UNMISET
## [22] UNIKOM
## [23] UNIMOG
## [24] UNMIBH
## 24 Levels: MINUJUSTH MINURSO MINUSCA ... UNVMC
##
## [[2]]
## [1] 24
##
## [1] "GBR"
## [[1]]
##  [1] MINUSMA
##  [2] UNAMA
```

```
##  [3] UNFICYP
##  [4] UNMISS
##  [5] UNSMIL
##  [6] UNSOM
##  [7] UNSOS
##  [8] MONUSCO
##  [9] UNVMC
## [10] UNMC
## [11] UNMIL
## [12] MINUSTAH
## [13] UNAMI
## [14] UNISFA
## [15] UNMIS
## [16] MONUC
## [17] UNFICYP
## [18] UNAMID
## [19] UNMIS
## [20] UNMIL
## [21] UNMIK
## [22] UNMEE
## [23] UNIOSIL
## [24] UNAMSIL
## [25] MONUC
## [26] UNMISET
## [27] UNIKOM
## [28] UNMIBH
## 28 Levels: MINUSMA MINUSTAH ... UNVMC
##
## [[2]]
## [1] 28
```

Step 7

This step is about descriptive stats oof the data, mostly focusing on the personnel and dates. The highest and the lower points can be found looking at the lists, but the lowest points had multiple dates, when there was only one troop on the mission.

```
library("dplyr" )
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
gender_data1$Total_personnel = gender_data1$Female_Personnel+gender_data1$Male_Personnel
#creating a subset of the df
minusma <- subset(gender_data1, Mission_Acronym == "MINUSMA", select=c(ISOCode3, Total_personnel, Last_
head(minusma)
```

```
##    ISOCode3 Total_personnel Last_Reporting_Date
```

```
## 8       ARM               1          2020-07-31
## 17      AUT               2          2020-07-31
## 29      BGD            1265          2020-07-31
## 30      BGD               2          2020-07-31
## 31      BGD             280          2020-07-31
## 32      BGD               1          2020-07-31
```

```
#calculating the descriptive stats of the column Total_personnel
round(mean(minusma$Total_personnel))
```

```
## [1] 114
```

```
median(minusma$Total_personnel)
```

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

```
sapply(minusma, class)
```

```
##            ISOCode3     Total_personnel Last_Reporting_Date
##         "character"           "integer"              "Date"
```

```
quantile(minusma$Total_personnel, probs=c(25/100, 75/100))
```

```
## 25% 75%
##   2  52
```

```
min_value<-min(minusma$Total_personnel)
max_value<-max(minusma$Total_personnel)
lowest_point<- minusma[which(minusma$Total_personnel == 1), 3]
head(lowest_point) #too many values for the output, remove head to see all the dates
```

```
## [1] "2020-07-31" "2020-07-31" "2020-07-31" "2020-07-31" "2020-07-31"
## [6] "2020-07-31"
```

```
highest_point<- minusma[which(minusma$Total_personnel == 1726), 3]
highest_point
```

```
## [1] "2017-01-31"
```