

CS112_assignment1_Rcompetency

Stevedavies Wambui

9/19/2020

Question 1: Loading the dataset

```
#load the dataset into a dataframe
```

```
un_data <- read.table(file = "https://tinyurl.com/UNpckpdata", header=TRUE, sep = ',', fill=TRUE, quote=)
```

```
#Print out the first 5 rows
```

```
head(un_data)
```

```
##   i..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
```

```
#Shows the structure of the dataframe
```

```
str(un_data)
```

```
## 'data.frame':   147631 obs. of  9 variables:
## $ i..Contribution_ID : int  427903 427904 427905 427906 427907 427908 427909 427910 427911 427912
## $ ISOCode3           : chr  "DZA" "ARG" "ARG" "ARG" ...
## $ M49_Code           : int  12 32 32 32 32 32 32 51 51 51 ...
## $ Contributing_Country: chr  "Algeria" "Argentina" "Argentina" "Argentina" ...
## $ Mission_Acronym     : chr  "MONUSCO" "MINURSO" "UNFICYP" "UNFICYP" ...
## $ Personnel_Type      : chr  "Experts on Mission" "Experts on Mission" "Troops" "Staff Officer" ...
## $ Female_Personnel    : int  0 0 15 3 1 0 5 0 2 0 ...
## $ Male_Personnel      : int  2 2 219 6 5 4 7 1 30 1 ...
## $ Last_Reporting_Date : chr  "31/07/2020" "31/07/2020" "31/07/2020" "31/07/2020" ...
```

Question 2: Null values

```
#Check if there are any null/empty values and print the total in each column
colSums(is.na(un_data))
```

```
##      i..Contribution_ID      ISOCode3      M49_Code
##              0              0              5
## Contributing_Country      Mission_Acronym      Personnel_Type
##              0              0              0
##      Female_Personnel      Male_Personnel      Last_Reporting_Date
##              4              4              0
```

```
#remove the empty/null values
un_data <- na.omit(un_data)

#confirm values have been removed
colSums(is.na(un_data))
```

```
##      i..Contribution_ID      ISOCode3      M49_Code
##              0              0              0
## Contributing_Country      Mission_Acronym      Personnel_Type
##              0              0              0
##      Female_Personnel      Male_Personnel      Last_Reporting_Date
##              0              0              0
```

Question 3: Formatting date column

```
#Change the column with dates to class 'date'
un_data$Last_Reporting_Date <- as.Date(un_data$Last_Reporting_Date, format = "%d/%m/%Y")
head(un_data)
```

```
##      i..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
```

```
#confirm that the class of the column has changed to 'Date' and show the structure of the dataframe
class(un_data$Last_Reporting_Date)
```

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

```
str(un_data)
```

```
## 'data.frame': 147626 obs. of 9 variables:
## $ i..Contribution_ID : int 427903 427904 427905 427906 427907 427908 427909 427910 427911 427912
## $ ISOCode3 : chr "DZA" "ARG" "ARG" "ARG" ...
## $ M49_Code : int 12 32 32 32 32 32 32 51 51 51 ...
## $ Contributing_Country: chr "Algeria" "Argentina" "Argentina" "Argentina" ...
## $ Mission_Acronym : chr "MONUSCO" "MINURSO" "UNFICYP" "UNFICYP" ...
## $ Personnel_Type : chr "Experts on Mission" "Experts on Mission" "Troops" "Staff Officer" ...
## $ Female_Personnel : int 0 0 15 3 1 0 5 0 2 0 ...
## $ Male_Personnel : int 2 2 219 6 5 4 7 1 30 1 ...
## $ Last_Reporting_Date : Date, format: "2020-07-31" "2020-07-31" ...
## - attr(*, "na.action")= 'omit' Named int [1:5] 30645 147337 147338 147507 147508
## ..- attr(*, "names")= chr [1:5] "30645" "147337" "147338" "147507" ...
```

```
##Question 4: Women in formed police units
```

```
#load the dplyr package to use the filter function
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
```

```
#confirm the naming of formed police units for accurate referencing
unique(un_data$Personnel_Type)
```

```
## [1] "Experts on Mission" "Troops" "Staff Officer"
## [4] "Individual Police" "Formed Police Units"
```

```
#filter the data by the Personnel type and the last reporting data so as to calculate the percentage
subset <- un_data %>%
  filter(Personnel_Type == "Formed Police Units", Last_Reporting_Date >= "2020-07-1")

#confirm dataset is filtered
head(subset)
```

```
## i..Contribution_ID ISOCode3 M49_Code Contributing_Country Mission_Acronym
## 1 427933 BGD 50 Bangladesh MINUSMA
## 2 427938 BGD 50 Bangladesh MONUSCO
## 3 427942 BGD 50 Bangladesh UNAMID
## 4 427960 BEN 204 Benin MINUSMA
```

```
## 5          428014      BFA      854      Burkina Faso      MINUSMA
## 6          428020      BFA      854      Burkina Faso      UNAMID
##      Personnel_Type Female_Personnel Male_Personnel Last_Reporting_Date
## 1 Formed Police Units          43          237      2020-07-31
## 2 Formed Police Units          77          103      2020-07-31
## 3 Formed Police Units          29          110      2020-07-31
## 4 Formed Police Units           5          133      2020-07-31
## 5 Formed Police Units          10          130      2020-07-31
## 6 Formed Police Units           7          133      2020-07-31
```

```
#calculate the total number of female personnel in formed police units
female_fpolic <- sum(subset$Female_Personnel)

#calculate the total number of total personnel in formed police units
male_fpolic <- sum(subset$Male_Personnel)

#calculate the total number of personnel in formed police units
total_fpolic <- female_fpolic + male_fpolic

#calculate the percentage of female personnel formed police units in 2020
percentage_2020 <- (female_fpolic/total_fpolic * 100)

percentage_2020
```

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

Based on the result (10.93%), we can see that the goal for 2028 has not yet been met and there is only a slight change from the previous year.

Question 5: Barplots for Female formed police units

```
#load the ggplot2 package to create a bar plot
library(ggplot2)
```

```
#filter 2019 data on formed police units
subset1 <- un_data %>%
  filter(Personnel_Type == "Formed Police Units", Last_Reporting_Date <= "2019-12-31" & Last_Reporting_Date >= "2019-01-01")

#calculate the percentage of female personnel in formed police units in 2019
percent_2019 <- (sum(subset1$Female_Personnel)/(sum(subset1$Male_Personnel) + sum(subset1$Female_Personnel)) * 100)
percent_2019
```

```
## [1] 10.38213
```

```
#filter 2018 data on formed police units
subset2 <- un_data %>%
  filter(Personnel_Type == "Formed Police Units", Last_Reporting_Date <= "2018-12-31" & Last_Reporting_Date >= "2018-01-01")

#calculate the percentage of female personnel in formed police units in 2018
percent_2018 <- (sum(subset2$Female_Personnel)/(sum(subset2$Male_Personnel) + sum(subset2$Female_Personnel)) * 100)
percent_2018
```

```
## [1] 7.654012
```

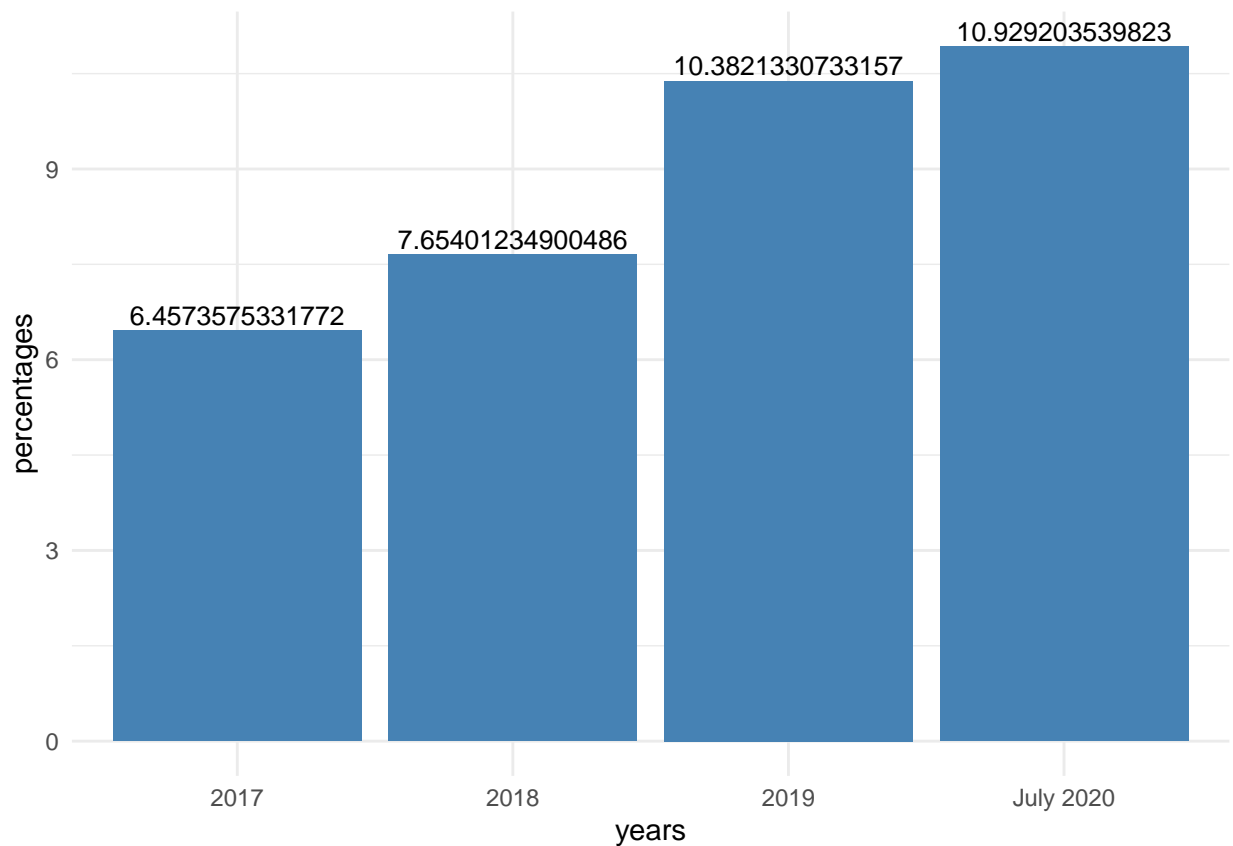
```
#filter 2017 data on formed police units
subset3 <- un_data %>%
  filter(Personnel_Type == "Formed Police Units", Last_Reporting_Date <= "2017-12-31" & Last_Reporting_Date <= "2019-12-31")

#calculate the percentage of female personnel in formed police units in 2018
percent_2017 <- (sum(subset3$Female_Personnel)/(sum(subset3$Male_Personnel) + sum(subset3$Female_Personnel)))
percent_2017
```

```
## [1] 6.457358
```

```
#create dataframe with years and percentages
df_bar <- data.frame(percentages = c(percent_2017, percent_2018, percent_2019, percent_2020),
  years = c('2017', '2018', '2019', 'July 2020'))

#Barplot showing the percentages of women in 2017, 2018, 2019 and 2020
ggplot(data = df_bar, aes(x=years, y=percentages)) +
  geom_bar(stat="identity", fill="steelblue") +
  geom_text(aes(label=percentages), vjust=-0.3, size=3.5) +
  theme_minimal()
```



```
## Question 6: Unique missions by country
```

```

#the function below takes the ISO code of a country and returns with a list of unique missions
#and number of missions personnel have been sent by the country
unique_missions_country <- function(iso){
  iso_df <- un_data %>%
    filter(ISOCode3 == iso)
  #trim whitespace in the missions column
  mission <- trimws(iso_df$Mission_Acronym, 'r')
  #get a list with unique missions
  unique_miss_list <- list(unique(mission))
  total_unique <- length(unique(mission))
  print(iso)
  print(unique_miss_list)
  print(total_unique)
}

#confirm that function works by testing on what we already know
unique_missions_country('ARG')

```

```

## [1] "ARG"
## [[1]]
## [1] "MINURSO" "UNFICYP" "UNMISS" "UNTSO" "UNVMC" "MINUSCA"
## [7] "MINUJUSTH" "UNAMI" "MINUSTAH" "UNMC" "UNOCI" "UNMIL"
## [13] "UNMIS" "MONUC" "UNMIK" "UNMISSET" "UNIKOM" "UNIMOG"
## [19] "UNMIBH"
##
## [1] 19

```

```

#create a list with the countries we are interested in
iso_list <- list('USA', 'KOR', 'IND', 'DEU', 'ARG', 'GBR')
#loop through the list above and call the unique missions functions
for(count in iso_list){
  unique_missions_country(count)
}

```

```

## [1] "USA"
## [[1]]
## [1] "BINUH" "MINUSCA" "MINUSMA" "MONUSCO" "UNMISS" "UNSMIL"
## [7] "UNTSO" "MINUJUSTH" "UNMIL" "MINUSTAH" "UNAMA" "MINURCAT"
## [13] "UNMIS" "UNMIK" "UNIOSIL" "UNAMID" "UNMIT" "UNOTIL"
## [19] "UNAMSIL" "UNMISSET" "UNIKOM" "UNMEE" "UNMIBH"
##
## [1] 23
## [1] "KOR"
## [[1]]
## [1] "MINURSO" "UNAMID" "UNIFIL" "UNMISS" "UNMOGIP" "UNMHA"
## [7] "MINUJUSTH" "UNMIL" "UNOCI" "MINUSTAH" "UNMIT" "UNISFA"
## [13] "UNMIS" "UNMIN" "UNAMA" "UNMISSET" "UNFICYP"
##
## [1] 17
## [1] "IND"
## [[1]]
## [1] "MINURSO" "MONUSCO" "UNDOF" "UNFICYP" "UNIFIL" "UNISFA"
## [7] "UNMISS" "UNTSO" "UNSOM" "MINUJUSTH" "MINUSTAH" "UNMIL"

```

```
## [13] "UNAMA"      "UNOCI"      "UNAMI"      "UNMIT"      "UNMIS"      "MONUC"
## [19] "UNMIK"      "UNIOSIL"    "UNMEE"      "ONUB"       "UNOMIG"     "UNAMSIL"
## [25] "UNIKOM"     "UNMIBH"
##
## [1] 26
## [1] "DEU"
## [[1]]
## [1] "MINURSO"    "MINUSMA"    "UNAMID"     "UNIFIL"     "UNMIK"      "UNMISS"
## [7] "UNSOM"      "UNMHA"      "MINUJUSTH"  "UNSMIL"     "UNMIL"      "UNAMA"
## [13] "MINUSTAH"   "UNMIS"      "UNOMIG"     "UNAMSIL"    "UNIKOM"     "UNMIBH"
##
## [1] 18
## [1] "ARG"
## [[1]]
## [1] "MINURSO"    "UNFICYP"    "UNMISS"     "UNTSO"      "UNVMC"      "MINUSCA"
## [7] "MINUJUSTH"  "UNAMI"      "MINUSTAH"   "UNMC"       "UNOCI"      "UNMIL"
## [13] "UNMIS"      "MONUC"      "UNMIK"      "UNMISSET"   "UNIKOM"     "UNIMOG"
## [19] "UNMIBH"
##
## [1] 19
## [1] "GBR"
## [[1]]
## [1] "MINUSMA"    "UNAMA"      "UNFICYP"    "UNMISS"     "UNSMIL"     "UNSOM"
## [7] "UNSOS"      "MONUSCO"    "UNVMC"      "UNMC"       "UNMIL"      "MINUSTAH"
## [13] "UNAMI"      "UNISFA"     "UNMIS"      "MONUC"      "UNAMID"     "UNMIK"
## [19] "UNMEE"      "UNIOSIL"    "UNAMSIL"    "UNMISSET"   "UNIKOM"     "UNMIBH"
##
## [1] 24
```

Question 7: MINUSMA dataframe summary statistics

```
minusma_df <- un_data%>%
  filter(Mission_Acronym == "MINUSMA")

#sum up the female and male personnel into one column
personell_total <- minusma_df$Female_Personnel + minusma_df$Male_Personnel

#Create dataframe with the personnel total by the Last_Reporting_Date
minusma_person <- aggregate(personell_total, by=list(
  DATE=minusma_df$Last_Reporting_Date), FUN=sum)

#Rename the columns in the new dataframe
minusma_person %>%
  rename(
    Total_Number_Of_Personnel = x
  )%>%
  head
```

```
##          DATE Total_Number_Of_Personnel
## 1 2013-07-31                6294
## 2 2013-08-31                6010
## 3 2013-09-30                6005
## 4 2013-10-31                5872
```

```
## 5 2013-11-30          6347
## 6 2013-12-31          6439
```

a) Average total personnel over time

```
#calculate the average of personnel over time
mean_minusma <- mean(minusma_person$x)
mean_minusma
```

```
## [1] 11768.85
```

b)Median of personnel

```
#calculate the median personnel value
median_minusma <- median(minusma_person$x)
median_minusma
```

```
## [1] 12039.5
```

c)Quantile of personnel over time

```
#calculate the 25th and 75th percentile for personnel
perc <- c(0.25, 0.75)
quantile_minusma <- quantile(minusma_person$x, perc)
quantile_minusma
```

```
##      25%      75%
## 10125.25 13882.00
```

d)Lowest number of personnel and when

```
#calculate the lowest personnel value and print out when it was
lowest_minusma <- min(minusma_person$x)
lowest_minusma_df <- minusma_person[minusma_person$x == lowest_minusma, c('DATE', 'x')]
lowest_minusma_df
```

```
##      DATE      x
## 4 2013-10-31 5872
```

e)Highest number of personnel and when

```
#calculate the highest personnel value and print out when it was
highest_minusma <- max(minusma_person$x)
highest_minusma_df <- minusma_person[minusma_person$x == highest_minusma, c('DATE', 'x')]
highest_minusma_df
```


| ## | DATE | x |
|-------|------------|-------|
| ## 66 | 2019-03-31 | 14871 |