EMATM0061-Coursework

Zhen Liu

2022-12-14

Section A

A.1

Answer

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✔ ggplot2 3.3.6      ✔ purrr   0.3.4
## ✔ tibble  3.1.8      ✔ dplyr   1.0.10
## ✔ tidyr   1.2.1      ✔ stringr 1.4.1
## ✔ readr   2.1.3      ✔ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
# load the csv file into a dataframe called "data_original"

data_original<-read.csv("finance_data_2022.csv")


#check the shape of dataframe

dim(data_original)
```

```
## [1] 1580    8
```

The number of rows is 1580 and the number of columns is 8.

A.2

Answer

```
#select the columns and rename

finance_data<-data_original%>%
  select(IFC.Region,IFC.Cumulative.Commitments..US..Thousands.,Country,Loan...Guarant
ee.participations.Cumulative.Commitments..US..Thousands.,As.of.Date)%>%
  rename(IFC=IFC.Region,IFC_CC=IFC.Cumulative.Commitments..US..Thousands.,Country=Cou
ntry, Loan_Guarantee_CC=Loan...Guarantee.participations.Cumulative.Commitments..US..T
housands. ,Date = As.of.Date )
```

```
#show the new dataframe 'finance_data' column names
head(finance_data)
```

```
##                          IFC     IFC_CC            Country Loan_Guarantee_CC
## 1 East Asia and the Pacific  316463.25          Cambodia             155000
## 2 East Asia and the Pacific 8199672.61             China            1830109
## 3 East Asia and the Pacific   52493.22              Fiji               2500
## 4 East Asia and the Pacific 4068991.86         Indonesia            2512055
## 5 East Asia and the Pacific    1798.00          Kiribati                  0
## 6 East Asia and the Pacific  868449.18 Korea, Republic of             195700
##         Date
## 1 06/30/2015
## 2 06/30/2015
## 3 06/30/2015
## 4 06/30/2015
## 5 06/30/2015
## 6 06/30/2015
```

A.3

Answer

```
# filter the dataframe based on conditions
data_part1<-finance_data%>%
  filter(IFC_CC>=300000 & Loan_Guarantee_CC<=500000)
```

```
# sort the IFC_CC in descending order
data_part1<-data_part1%>%
  arrange(desc(IFC_CC))
```

```
# display the first 4 rows of selected columns
data_part1%>%
  select(IFC,IFC_CC,Loan_Guarantee_CC)%>%
  head(4)
```

```
##                   IFC   IFC_CC Loan_Guarantee_CC
## 1           Worldwide 13280154          330206.2
## 2           Worldwide 11399022          330206.0
## 3 Sub-Saharan Africa 10426234          477155.0
## 4 Sub-Saharan Africa  9863582          456155.0
```

A.4

Answer

```
# add a new column by mutate and map2_dbl

finance_data<-finance_data%>%
  mutate(IFC_ratio= map2_dbl(.x = IFC_CC, .y = Loan_Guarantee_CC, ~ (.x/(.x+.y))))
```

```
#Display a subset consisting of first 5 rows and 4 columns

finance_data%>%
  select(IFC,IFC_CC,Loan_Guarantee_CC,IFC_ratio)%>%
  head(5)
```

```
##                          IFC     IFC_CC Loan_Guarantee_CC IFC_ratio
## 1 East Asia and the Pacific  316463.25            155000 0.6712363
## 2 East Asia and the Pacific 8199672.61           1830109 0.8175325
## 3 East Asia and the Pacific   52493.22              2500 0.9545399
## 4 East Asia and the Pacific 4068991.86           2512055 0.6182895
## 5 East Asia and the Pacific    1798.00                 0 1.0000000
```

A.5

Answer

```
# split the date columns into three individual column day, month, year

# make sure the three columns are numeric type

finance_data<-finance_data%>%
  separate(Date, into =c('month','day','year'), sep = '/',convert = TRUE)
```

```
# display the first rows and 4 columns

finance_data%>%
  select(IFC_CC,day,month,year)%>%
  head(5)
```

```
##       IFC_CC day month year
## 1  316463.25  30     6 2015
## 2 8199672.61  30     6 2015
## 3   52493.22  30     6 2015
## 4 4068991.86  30     6 2015
## 5    1798.00  30     6 2015
```

A.6

Answer

```
# generate a summary data frame and have seven rows and 7 columns
# the missing values should not be taken into account

summary_data<-finance_data %>%
  select(IFC,IFC_CC,Loan_Guarantee_CC) %>%
  group_by(IFC) %>%
  summarise(ifc_mn=mean(IFC_CC,na.rm=TRUE),
            ifc_21q=quantile(IFC_CC,probs=0.21,na.rm=TRUE),
            ifc_var=var(IFC_CC,na.rm = TRUE),
            lg_mn = mean(Loan_Guarantee_CC,na.rm=TRUE),
            lg_21q = quantile(Loan_Guarantee_CC,probs=0.21,na.rm=TRUE),
            lg_var = var(Loan_Guarantee_CC,na.rm = TRUE))
```

```
# display the summary data frame
summary_data
```

```
## # A tibble: 7 × 7
##   IFC                              ifc_mn ifc_21q ifc_var  lg_mn lg_21q  lg_var
##   <chr>                             <dbl>   <dbl>   <dbl>  <dbl>  <dbl>   <dbl>
## 1 East Asia and the Pacific       1481717.  19678. 6.32e12 4.18e5      0 5.71e11
## 2 Europe and Central Asia         1280212. 122080. 6.51e12 3.06e5      0 5.92e11
## 3 Latin America and the Caribbean 1791922.  76417. 1.06e13 7.06e5      0 2.48e12
## 4 Middle East and North Africa    1114380. 206004. 2.34e12 2.06e5      0 7.84e10
## 5 South Asia                      2854665. 168250  2.28e13 2.89e5      0 3.18e11
## 6 Sub-Saharan Africa               640224.  26152. 2.61e12 5.91e4      0 2.49e10
## 7 Worldwide                       1307271.   1233. 1.55e13 3.89e4      0 1.19e10
```

A.7

Answer

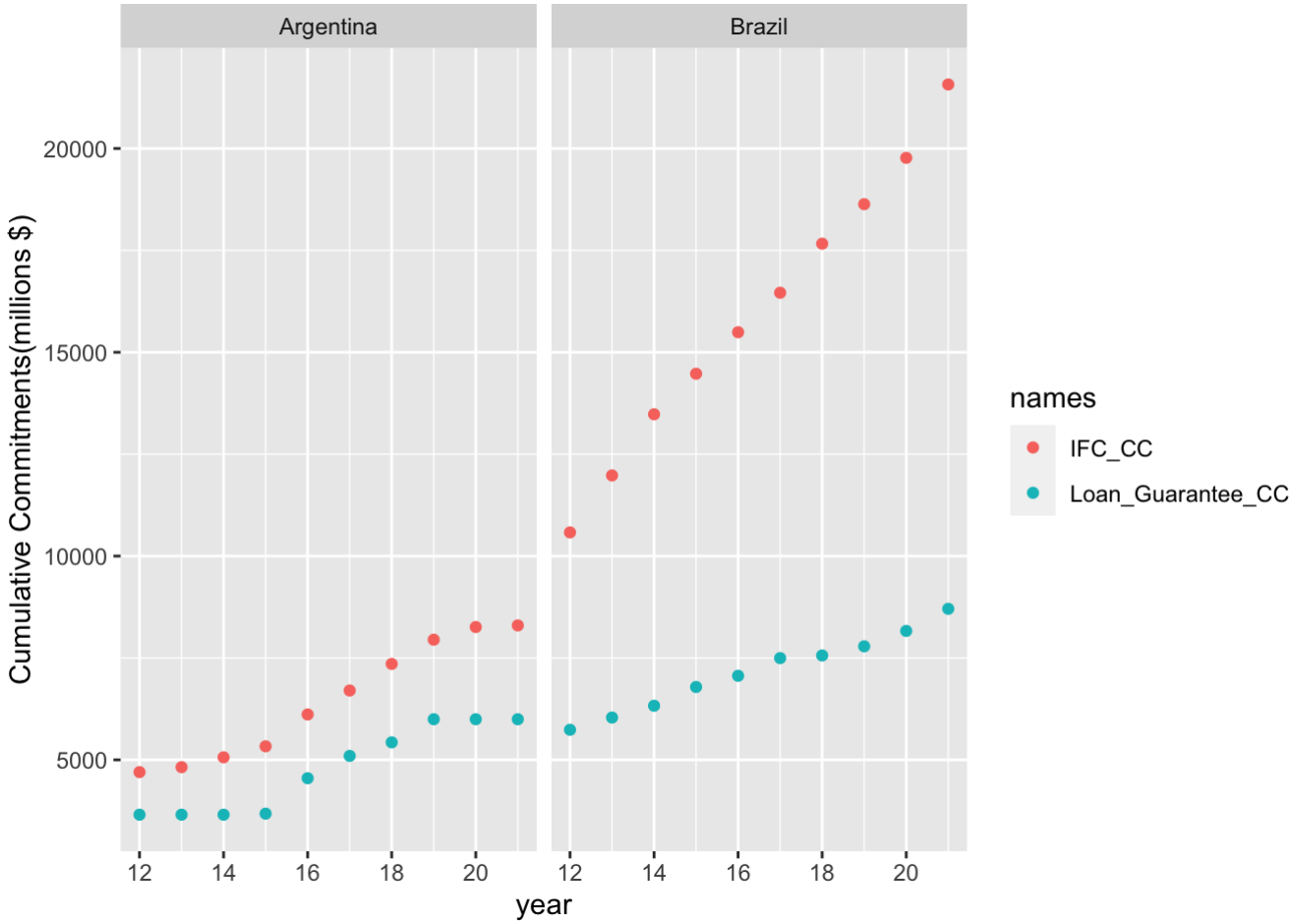
```
# creat a plot to display two panels including two countries and cumulative commitmen
ts
# the years are presented by last two digits
# IFC_CC and Loan_Guarantee_CC should be in the unit of million dollars rather than t
housand


plot_df<-finance_data%>%
  select(Country,year,IFC_CC,Loan_Guarantee_CC)%>%
  filter(Country=='Argentina' | Country=='Brazil')%>%
  mutate(year = year - round(year, -2),
         IFC_CC = IFC_CC/1000,
         Loan_Guarantee_CC = Loan_Guarantee_CC/1000)
```

```
# Collapses the two names 'IFC_CC' and 'Loan_Guarantee_CC' into key-value pairs

plot_df<-plot_df%>%
  select(year,Country,IFC_CC,Loan_Guarantee_CC)%>%
  gather(key = "names",value = "value", -year ,-Country)
```

```
# plot the data frame
ggplot(plot_df,aes(x=year,y=value, color=names)) + geom_point()+
  facet_wrap(~Country) +
xlab('year') + ylab('Cumulative Commitments(millions $)')
```



A.8

Answer

```
#build up a function to replace NA with 0.9 quantile
impute_by_quantile<-function(x){
  qu<-quantile(x,probs=0.9,na.rm=TRUE)

  impute_f<-function(z){
    if (is.na(z)){
      return(qu)
  }else{
    return(z)
  }
}
return(map_dbl(x,impute_f))
}
```

```
# Apply the function to finance_data dataframe
finance_data<-finance_data%>%
  mutate(IFC_CC = impute_by_quantile(IFC_CC),
         Loan_Guarantee_CC = impute_by_quantile(Loan_Guarantee_CC),
         IFC_ratio = impute_by_quantile(IFC_ratio))
```

```
#display a dataframe contain 3 columns and one row, it will show the mean value of th
e 3 columns from finance_data.

summary_finance<-finance_data%>%
  summarise(IFC_CC = mean(IFC_CC,na.rm=TRUE),
            Loan_Guarantee_CC = mean(Loan_Guarantee_CC,na.rm=TRUE),
            IFC_ratio = mean(IFC_ratio,na.rm=TRUE))

summary_finance
```

```
##    IFC_CC Loan_Guarantee_CC IFC_ratio
## 1 1290654            301778 0.8884476
```