````
---
title: "AssignmentRlab01"
author: "Ehsan Eslmai Shafigh"
date: "2023-04-13"
output:
  html_document: default
  word_document: default
  pdf_document: default
---

## Assignment 1

### Exercise 1: American Airlines Employees

#### 1)

We start by importing the data:

```{r}
aa = read.delim(file = "american_airline_empl.txt" , sep = "")
da = read.delim(file = "delta_airline_empl.txt" , sep = "")
fe = read.delim(file = "federal_express_empl.txt" , sep = "")
ua = read.delim(file = "united_airline_empl.txt" , sep = "")
```

There are a couple of columns which contain strings, which we would like to convert to numbers:

```{r}
aa$Full.time = as.numeric(gsub("," , "" , aa$Full.time ))
aa$Part.time = as.numeric(gsub("," , "" , aa$Part.time ))
aa$Grand = as.numeric(gsub("," , "" , aa$Grand ))

da$Full.time = as.numeric(gsub("," , "" , da$Full.time ))
da$Part.time = as.numeric(gsub("," , "" , da$Part.time ))
da$Grand = as.numeric(gsub("," , "" , da$Grand ))

fe$Full.time = as.numeric(gsub("," , "" , fe$Full.time ))
fe$Part.time = as.numeric(gsub("," , "" , fe$Part.time ))
fe$Grand = as.numeric(gsub("," , "" , fe$Grand ))

ua$Full.time = as.numeric(gsub("," , "" , ua$Full.time ))
ua$Part.time = as.numeric(gsub("," , "" , ua$Part.time ))
ua$Grand = as.numeric(gsub("," , "" , ua$Grand ))
```

We convert the imported data into a tibble:

```{r}
aa = tibble(aa)
da = tibble(da)
fe = tibble(fe)
ua = tibble(ua)
```

#### 2)

To successfully merge the tibbles, we need to change the names of some columns:

```{r}
aa = rename(aa , 'American Airlines FT' = 'Full.time' , 'American Airlines PT' =
'Part.time' , 'American Airlines Grand' = 'Grand' )
da = rename(da , 'Delta Airlines FT' = 'Full.time' , 'Delta Airlines PT' = 'Part.time'
, 'Delta Airlines Grand' = 'Grand' )
fe = rename(fe , 'Federal Express FT' = 'Full.time' , 'Federal Express PT' =
'Part.time' , 'Federal Express Grand' = 'Grand' )
ua = rename(ua , 'United Airlines FT' = 'Full.time' , 'United Airlines PT' =
'Part.time' , 'United Airlines Grand' = 'Grand' )
```
````

We need to combine the two columns corresponding to the year and month, to create a new column of the date type. For a practical reason, we need also to add a column of days:

```{r}
aa$day = 1
aa = unite(aa, date , Year , Month , day , sep = "-")
aa$date = ymd(aa$date)

da$day = 1
da = unite(da, date , Year , Month , day , sep = "-")
da$date = ymd(da$date)

fe$day = 1
fe = unite(fe, date , Year , Month , day , sep = "-")
fe$date = ymd(fe$date)

ua$day = 1
ua = unite(ua, date , Year , Month , day , sep = "-")
ua$date = ymd(ua$date)

df = merge(aa , da , by = 'date')
df = merge(df , fe , by = 'date')
df = merge(df , ua , by = 'date')
```

3)

one plot for the full-time employees:

```{r}
df = select(df , -Total.x , -Total.y)

ggplot(df, aes(x = date)) +
  geom_line(aes(y = `American Airlines FT`, color = "American Airlines")) +
  geom_line(aes(y = `Delta Airlines FT`, color = "Delta Airlines")) +
  geom_line(aes(y = `Federal Express FT`, color = "Federal Express")) +
  geom_line(aes(y = `United Airlines FT`, color = "United Airlines")) +
  scale_color_manual(name = "Legends", values = c("American Airlines" = "red", "Delta
Airlines" = "blue" , "Federal Express" = "green" , "United Airlines" = 'orange')) +
  labs(x = "Date", y = "Number of Full-Time Employees by  Airline") +
  theme_classic()+
  scale_x_date(date_breaks = "1 year", date_labels = "%Y")+ theme(
    panel.grid.major = element_line(color = "gray", linetype = "dashed"),
    panel.grid.minor = element_blank(),
    panel.background = element_blank()
  )

ggsave("full time employees.png", width = 15, height = 6)

```

Another plot for the part-time workers:

```{r}
ggplot(df, aes(x = date)) +
  geom_line(aes(y = `American Airlines PT`, color = "American Airlines")) +
  geom_line(aes(y = `Delta Airlines PT`, color = "Delta Airlines")) +
  geom_line(aes(y = `Federal Express PT`, color = "Federal Express")) +
  geom_line(aes(y = `United Airlines PT`, color = "United Airlines")) +
  scale_color_manual(name = "Legends", values = c("American Airlines" = "red", "Delta
Airlines" = "blue" , "Federal Express" = "green" , "United Airlines" = 'orange')) +
  labs(x = "Date", y = "Number of Part-Time Employees by Airline") +
  theme_classic()+
  scale_x_date(date_breaks = "1 year", date_labels = "%Y")+ theme(
    panel.grid.major = element_line(color = "gray", linetype = "dashed"),
    panel.grid.minor = element_blank(),
```

```
    panel.background = element_blank()
  )


ggsave("part time employees.png", width = 15, height = 6)
```

4)

```{r}
print(df[which.max(df$`American Airlines Grand`),1])
print(df[which.max(df$`Delta Airlines Grand`),1])
print(df[which.max(df$`Federal Express Grand`),1])
print(df[which.max(df$`United Airlines Grand`),1])

print(df[which.min(df$`American Airlines Grand`),1])
print(df[which.min(df$`Delta Airlines Grand`),1])
print(df[which.min(df$`Federal Express Grand`),1])
print(df[which.min(df$`United Airlines Grand`),1])
```

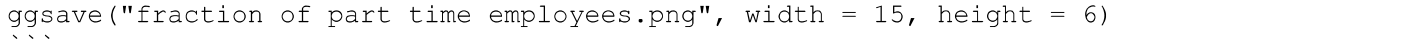
5)

We create a new column containing the fraction of part-time workers to the total workers:

```{r}
df$ptf_aa = df$`American Airlines PT`/(df$`American Airlines FT` + df$`American Airlines PT`)
df$ptf_da = df$`Delta Airlines PT`/(df$`Delta Airlines FT` + df$`Delta Airlines PT`)
df$ptf_fe = df$`Federal Express PT`/(df$`Federal Express FT` + df$`Federal Express PT`)
df$ptf_ua = df$`United Airlines PT` / (df$`United Airlines FT` + df$`United Airlines PT`)
```

```{r}
ggplot(df, aes(x = date)) +
  geom_line(aes(y = `ptf_aa`, color = "American Airlines")) +
  geom_line(aes(y = `ptf_da`, color = "Delta Airlines")) +
  geom_line(aes(y = `ptf_fe`, color = "Federal Express")) +
  geom_line(aes(y = `ptf_ua`, color = "United Airlines")) +
  scale_color_manual(name = "Legends", values = c("American Airlines" = "red", "Delta Airlines" = "blue" , "Federal Express" = "green" , "United Airlines" = 'orange')) +
  labs(x = "Date", y = "Fraction of Part-Time Employees by Airline") +
  theme_classic()+
  scale_x_date(date_breaks = "1 year", date_labels = "%Y")+ theme(
    panel.grid.major = element_line(color = "gray", linetype = "dashed"),
    panel.grid.minor = element_blank(),
    panel.background = element_blank()
  )

ggsave("fraction of part time employees.png", width = 15, height = 6)
```



6)

We have to take a look at the total employees of the four companies:

```{r}
ggplot(df, aes(x = date)) +
  geom_line(aes(y = `American Airlines Grand`, color = "American Airlines")) +
  geom_line(aes(y = `Delta Airlines Grand`, color = "Delta Airlines")) +
  geom_line(aes(y = `Federal Express Grand`, color = "Federal Express")) +
  geom_line(aes(y = `United Airlines Grand`, color = "United Airlines")) +
  scale_color_manual(name = "Legends", values = c("American Airlines" = "red", "Delta
```

```
Airlines" = "blue" , "Federal Express" = "green" , "United Airlines" = 'orange')) +
  labs(x = "Date", y = "Total Number of Employees by Airline") +
  theme_classic()+
  scale_x_date(date_breaks = "1 year", date_labels = "%Y")+ theme(
    panel.grid.major = element_line(color = "gray", linetype = "dashed"),
    panel.grid.minor = element_blank(),
    panel.background = element_blank()
  )


ggsave("total employees.png", width = 15, height = 6)
```

We see that unlike the other companies, the number of employees in the Federal Express has increased during the period, while for the others it has declined initially and then returned almost to the pre-covid values.

Exercise 2: Data Frames & Tibble

we begin with importing the data:

```{r}
nyc_flights = flights
```

In order to plot the data we form a column corresponding to the date of each flight:

```{r}
nyc_flights = unite(nyc_flights , date , year , month , day , sep = "-" )
nyc_flights$date = ymd(nyc_flights$date)
```

1.1)

We group by the dates and the origin, and summarize the results as the counts of flights:

```{r}
flight_count_summary = nyc_flights %>%
  group_by(date , origin) %>%
  summarize(flight_counts = n())
```

we plot the results:

```{r}
ggplot(data = flight_count_summary, aes(x = date, y = flight_counts, color = origin)) +
  labs(x = "Date", y = "Flight Counts per Date") +
  geom_line()+
  scale_x_date(date_breaks = "20 day", date_labels = "%D")


ggsave("flights per day.png", width = 15, height = 6)
```