

library(tidyverse)

```
## --- Attaching core tidyverse packages --- tidyverse 2.0.0 ---
## ✓ dplyr 1.1.0 ✓ readr 2.1.4
## ✓ forcats 1.0.0 ✓ stringr 1.5.0
## ✓ ggplot2 3.4.1 ✓ tibble 3.2.0
## ✓ lubridate 1.9.2 ✓ tidyr 1.3.0
## ✓ purrr 1.0.1
## --- Conflicts --- tidyverse_conflicts() ---
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag() masks stats::lag()
## ! Use the ð;8;http://conflicted.r-lib.org/ð;8; to force all conflicts to become errors
```

```
df_amer_air <- read.delim(file='/Users/alessandrobreccia/Desktop/1_ANNO_MAGISTR/2_semester/AdvStat/Lab/Rlab01/ame
rican_airline_empl.txt', header=TRUE)
df_delta_air <- read.delim(file='/Users/alessandrobreccia/Desktop/1_ANNO_MAGISTR/2_semester/AdvStat/Lab/Rlab01/de
lta_airline_empl.txt', header=TRUE)
df_federal_express_air <- read.delim(file='/Users/alessandrobreccia/Desktop/1_ANNO_MAGISTR/2_semester/AdvStat/La
b/Rlab01/federal_express_empl.txt', header=TRUE)
df_united_airline_empl <- read.delim(file='/Users/alessandrobreccia/Desktop/1_ANNO_MAGISTR/2_semester/AdvStat/La
b/Rlab01/united_airline_empl.txt', header=TRUE)
```

```
df_amer_air <- df_amer_air %>% add_column(Company="American Airline")
df_delta_air <- df_delta_air %>% add_column(Company="Delta Airline")
df_federal_express_air <- df_federal_express_air %>% add_column(Company="Federal Express Airline")
df_united_airline_empl <- df_united_airline_empl %>% add_column(Company="United Airline")
```

```
df_tot <- full_join(df_amer_air, (full_join(df_delta_air, full_join(df_federal_express_air, df_united_airline_emp
l))))
```

```
## Joining with `by` = join_by(Month, Year, Full.time, Part.time, Grand.Total,
## Company)`
## Joining with `by` = join_by(Month, Year, Full.time, Part.time, Grand.Total,
## Company)`
## Joining with `by` = join_by(Month, Year, Full.time, Part.time, Grand.Total,
## Company)`
```

```
df_tot$Full.time <- gsub('','',df_tot$Full.time) %>% strtoi()
df_tot$Part.time <- gsub('','',df_tot$Part.time) %>% strtoi()
df_tot$Grand.Total <- gsub('','',df_tot$Grand.Total) %>% strtoi()
```

```
df_tot$date <- as.Date(with(df_tot, paste(Year, Month,1,sep="-")), "%Y-%m-%d")
```

library(tidyverse)

ggplot(data = df_tot, mapping = aes(x = date, y = Full.time)) +
 geom_point(aes(color = Company)) + geom_line(aes(color = Company))

ggplot(data = df_tot, mapping = aes(x = date, y = Part.time)) +
 geom_point(aes(color = Company)) + geom_line(aes(color = Company))

companylist <- c('American Airline','Delta Airline','Federal Express Airline','United Airline')

for(i in 1:4){
 maxemplame <- which.max(df_tot[df_tot\$Company == companylist[i],]\$Grand.Total)
 minemplame <- which.min(df_tot[df_tot\$Company == companylist[i],]\$Grand.Total)

 cat("The maximum number of employees of the", companylist[i], "is at", as.character(df_tot[df_tot\$Company ==
 companylist[i],]\$date[maxemplame]), "and is", df_tot[df_tot\$Company == companylist[i],]\$Grand.Total[maxempl
 e], "employees\n")

 cat("The minimum number of employees of the", companylist[i], "is at", as.character(df_tot[df_tot\$Company ==
 companylist[i],]\$date[maxemplame]), "and is", df_tot[df_tot\$Company == companylist[i],]\$Grand.Total[maxempl
 e], "employees\n")
}

```
## The maximum number of employees of the American Airline is at 2018-06-01 and is 109171 employees  
## The minimum number of employees of the American Airline is at 2018-06-01 and is 109171 employees  
## The maximum number of employees of the Delta Airline is at 2023-01-01 and is 94675 employees  
## The minimum number of employees of the Delta Airline is at 2023-01-01 and is 94675 employees  
## The maximum number of employees of the Federal Express Airline is at 2021-03-01 and is 278383 employees  
## The minimum number of employees of the Federal Express Airline is at 2021-03-01 and is 278383 employees  
## The maximum number of employees of the United Airline is at 2001-03-01 and is 102046 employees  
## The minimum number of employees of the United Airline is at 2001-03-01 and is 102046 employees
```

ggplot(data = df_tot, mapping = aes(x = date, y = Part.time/Grand.Total)) +
 geom_point(aes(color = Company)) + geom_line(aes(color = Company))

ggplot(data = df_tot, mapping = aes(x = date, y = Full.time/Grand.Total)) +
 geom_point(aes(color = Company)) + geom_line(aes(color = Company))

ggplot(data = subset(df_tot, df_tot\$date > '2019-01-01'), mapping = aes(x = date, y = Full.time)) +
 geom_point(aes(color = Company)) + geom_line(aes(color = Company))

----- Exercise 2 -----

library(nycflights13)
library(tidyverse)

flights\$date <- as.Date(with(flights, paste(year, month, day, sep="-")), "%Y-%m-%d")

```
library(dplyr)  
flights$date <- weekdays(as.Date(flights$date))  
  
totflight <- flights %>% group_by(date,origin,wdate) %>% summarise(count=n(), .groups = 'drop')  
  
ggplot(data = totflight, mapping = aes(x = date, y = count)) +  
  geom_point(aes(color = origin))
```

```
totflight$week <- strftime(totflight$date, format = "%V")  
  
avgweekend <- totflight[ totflight$wdate == 'Saturday' | totflight$wdate == 'Sunday' , ] %>% group_by(week) %>%  
  summarise(mean=mean(count), .groups='drop')  
  
avgweek <- totflight[ totflight$wdate == 'Monday' | totflight$wdate == 'Tuesday' | totflight$wdate == 'Wednesda  
y' | totflight$wdate == 'Thursday' | totflight$wdate == 'Friday' , ] %>% group_by(week) %>% summarise(mean=mean  
(count), .groups='drop')  
  
avgweekend <- avgweekend %>% add_column(id='week-end')  
avgweek <- avgweek %>% add_column(id='week')  
  
avg= rbind(avgweekend,avgweek)  
  
ggplot(data = avg, mapping = aes(x = week, y = mean)) +  
  geom_point(aes(color = id)) + geom_line(aes(color = id))
```

```
## 'geom_line()': Each group consists of only one observation.  
## ! Do you need to adjust the group aesthetic?
```

library(ggplot2)
library(reshape2)

Attaching package: 'reshape2'

```
## The following object is masked from 'package:tidyr':  
##     smiths
```

```
maxdel <- flights %>% group_by(date) %>% summarise(max=max(dep_delay, na.rm=TRUE), min=min(dep_delay, na.rm=TRU  
E), avg=mean(dep_delay, na.rm=TRUE))  
  
maxdel <- melt(maxdel, id.vars = 'date', variable.name = 'type')  
  
ggplot(data = maxdel, mapping = aes(x = date, y = value)) + geom_line(aes(color = type))
```

```
flights$speed <- flights$distance/((flights$sair_time %/% 100) + (flights$sair_time %/% 100)/60)  
  
avgspeed <- flights %>% group_by(date,origin) %>% summarise(avg=mean(speed, na.rm=TRUE), .groups='drop')  
  
ggplot(data = avgspeed, mapping = aes(x = date, y = avg)) + geom_point(aes(color= origin))
```

```
flights$week <- strftime(flights$date, format = "%V")  
  
perday <- flights %>% group_by(date,carrier) %>% summarise(flights_per_day=n(), .groups='drop')  
perday <- perday %>% group_by(carrier) %>% summarise(mean_flights_per_day=mean(flights_per_day))  
  
perweek <- flights %>% group_by(week,carrier) %>% summarise(flights_per_week=n(), .groups='drop')  
perweek <- perweek %>% group_by(carrier) %>% summarise(mean_flights_per_week=mean(flights_per_week))  
  
permonth <- flights %>% group_by(month,carrier) %>% summarise(flights_per_month=n(), .groups='drop')  
permonth <- permonth %>% group_by(carrier) %>% summarise(mean_flights_per_month=mean(flights_per_month))  
  
distpermonth <- flights %>% group_by(month,carrier) %>% summarise(dist=max(distance, na.rm=TRUE), .groups='drop')  
distpermonth <- distpermonth %>% group_by(carrier) %>% summarise(mean_dist_per_month=mean(distpermonth))  
  
cat("The company offering the largest number of flights per day is",perday$carrier[which.max(perday$mean_flights_  
per_day)],"with a value of",perday$mean_flights_per_day[which.max(perday$mean_flights_per_day)],"\n")  
  
## The company offering the largest number of flights per day is UA with a value of 160.726  
  
cat("The company offering the largest number of flights per week is",perweek$carrier[which.max(perweek$mean_fli  
ghts_per_week)],"with a value of",perweek$mean_flights_per_week[which.max(perweek$mean_flights_per_week)],"\n")  
  
## The company offering the largest number of flights per week is UA with a value of 1128.173  
  
cat("The company offering the largest number of flights per month is",permonth$carrier[which.max(permonth$mean_f  
lights_per_month)],"with a value of",permonth$mean_flights_per_month[which.max(permonth$mean_flights_per_mont  
h)],"\n")  
  
## The company offering the largest number of flights per month is UA with a value of 4888.75  
  
cat("The company offering the largest covered distance per month is",distpermonth$carrier[which.max(distpermonth$me  
an_dist_per_month)],"with a value of",distpermonth$mean_dist_per_month[which.max(distpermonth$mean_dist_per_mont  
h)],"\n")  
  
## The company offering the largest covered distance per month is HA with a value of 4983
```