# Module 6 - Assignment 1

## Pierce, Michaela

### Lubridate

# Load necessary packages  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lubridate)

library(readr)  
Appointments <- read\_csv("Appointments.csv")

## Rows: 61214 Columns: 12  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (12): Age, Sex, ReservationMonth, ReservationDay, ReservationHour, Reser...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#View(Appointments)

# Working with Dates

# set the ReservationYear to 2019 for all the rows  
Appointments <- Appointments %>%  
 mutate(ReservationYear = 2019)  
  
  
# Combine ReservationYear, ReservationMonth, and ReservationDay to create ReservationDate  
# Combine CreationYear, CreationMonth, and CreationDay to create CreationDate  
Appointments <- Appointments %>%  
 mutate(ReservationDate = make\_date(year = ReservationYear, month = ReservationMonth, day = ReservationDay),  
 CreationDate = make\_date(year = CreationYear, month = CreationMonth, day = CreationDay))  
  
  
# Calculate the difference in days between ReservationDate and CreationDate  
# Convert the difftime object to numeric  
Appointments <- Appointments %>%  
 mutate(ReservationSpan = as.numeric(difftime(ReservationDate, CreationDate, units = "days")))

# Exploratory Analysis

# View summary of data before converting ReservationSpan to numeric  
summary(Appointments)

## Age Sex ReservationMonth ReservationDay   
## Min. : 0.00 Min. :1.000 Min. :1.000 Min. : 1.00   
## 1st Qu.: 16.00 1st Qu.:1.000 1st Qu.:2.000 1st Qu.: 7.00   
## Median : 34.00 Median :2.000 Median :3.000 Median :15.00   
## Mean : 34.64 Mean :1.597 Mean :2.527 Mean :14.51   
## 3rd Qu.: 51.00 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:21.00   
## Max. :101.00 Max. :2.000 Max. :4.000 Max. :28.00   
## ReservationHour ReservationMinute CreationMonth CreationDay   
## Min. : 0.00 Min. :0 Min. : 1.000 Min. : 1.00   
## 1st Qu.:11.00 1st Qu.:0 1st Qu.: 2.000 1st Qu.: 7.00   
## Median :14.00 Median :0 Median : 2.000 Median :14.00   
## Mean :13.85 Mean :0 Mean : 5.265 Mean :14.51   
## 3rd Qu.:17.00 3rd Qu.:0 3rd Qu.:12.000 3rd Qu.:22.00   
## Max. :21.00 Max. :0 Max. :12.000 Max. :28.00   
## CreationYear CreationHour Creation Minute Show   
## Min. :2018 Min. : 0.00 Min. :0 Min. :0.0000   
## 1st Qu.:2018 1st Qu.:10.00 1st Qu.:0 1st Qu.:1.0000   
## Median :2019 Median :13.00 Median :0 Median :1.0000   
## Mean :2019 Mean :13.42 Mean :0 Mean :0.7898   
## 3rd Qu.:2019 3rd Qu.:16.00 3rd Qu.:0 3rd Qu.:1.0000   
## Max. :2019 Max. :23.00 Max. :0 Max. :1.0000   
## ReservationYear ReservationDate CreationDate ReservationSpan   
## Min. :2019 Min. :2019-01-01 Min. :2018-10-03 Min. : 1.00   
## 1st Qu.:2019 1st Qu.:2019-02-01 1st Qu.:2018-12-21 1st Qu.: 25.00   
## Median :2019 Median :2019-03-04 Median :2019-01-20 Median : 35.00   
## Mean :2019 Mean :2019-03-01 Mean :2019-01-21 Mean : 38.53   
## 3rd Qu.:2019 3rd Qu.:2019-03-28 3rd Qu.:2019-02-19 3rd Qu.: 49.00   
## Max. :2019 Max. :2019-04-28 Max. :2019-03-28 Max. :207.00

# Convert ReservationSpan to numeric  
Appointments <- Appointments %>%  
 mutate(ReservationSpan = as.numeric(ReservationSpan))  
  
# View summary of data after converting ReservationSpan to numeric  
summary(Appointments)

## Age Sex ReservationMonth ReservationDay   
## Min. : 0.00 Min. :1.000 Min. :1.000 Min. : 1.00   
## 1st Qu.: 16.00 1st Qu.:1.000 1st Qu.:2.000 1st Qu.: 7.00   
## Median : 34.00 Median :2.000 Median :3.000 Median :15.00   
## Mean : 34.64 Mean :1.597 Mean :2.527 Mean :14.51   
## 3rd Qu.: 51.00 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:21.00   
## Max. :101.00 Max. :2.000 Max. :4.000 Max. :28.00   
## ReservationHour ReservationMinute CreationMonth CreationDay   
## Min. : 0.00 Min. :0 Min. : 1.000 Min. : 1.00   
## 1st Qu.:11.00 1st Qu.:0 1st Qu.: 2.000 1st Qu.: 7.00   
## Median :14.00 Median :0 Median : 2.000 Median :14.00   
## Mean :13.85 Mean :0 Mean : 5.265 Mean :14.51   
## 3rd Qu.:17.00 3rd Qu.:0 3rd Qu.:12.000 3rd Qu.:22.00   
## Max. :21.00 Max. :0 Max. :12.000 Max. :28.00   
## CreationYear CreationHour Creation Minute Show   
## Min. :2018 Min. : 0.00 Min. :0 Min. :0.0000   
## 1st Qu.:2018 1st Qu.:10.00 1st Qu.:0 1st Qu.:1.0000   
## Median :2019 Median :13.00 Median :0 Median :1.0000   
## Mean :2019 Mean :13.42 Mean :0 Mean :0.7898   
## 3rd Qu.:2019 3rd Qu.:16.00 3rd Qu.:0 3rd Qu.:1.0000   
## Max. :2019 Max. :23.00 Max. :0 Max. :1.0000   
## ReservationYear ReservationDate CreationDate ReservationSpan   
## Min. :2019 Min. :2019-01-01 Min. :2018-10-03 Min. : 1.00   
## 1st Qu.:2019 1st Qu.:2019-02-01 1st Qu.:2018-12-21 1st Qu.: 25.00   
## Median :2019 Median :2019-03-04 Median :2019-01-20 Median : 35.00   
## Mean :2019 Mean :2019-03-01 Mean :2019-01-21 Mean : 38.53   
## 3rd Qu.:2019 3rd Qu.:2019-03-28 3rd Qu.:2019-02-19 3rd Qu.: 49.00   
## Max. :2019 Max. :2019-04-28 Max. :2019-03-28 Max. :207.00

# Calculate average time between appointment creation and reservation date  
average\_reservation\_span <- mean(Appointments$ReservationSpan)  
  
# Find maximum time between appointment creation and reservation date  
max\_reservation\_span <- max(Appointments$ReservationSpan)  
  
# Find minimum time between appointment creation and reservation date  
min\_reservation\_span <- min(Appointments$ReservationSpan)  
  
# Print results  
cat("a. Average time between appointment creation and reservation date:", average\_reservation\_span, "days\n")

## a. Average time between appointment creation and reservation date: 38.52842 days

cat("b. Maximum time between appointment creation and reservation date:", max\_reservation\_span, "days\n")

## b. Maximum time between appointment creation and reservation date: 207 days

cat("c. Minimum time between appointment creation and reservation date:", min\_reservation\_span, "days\n")

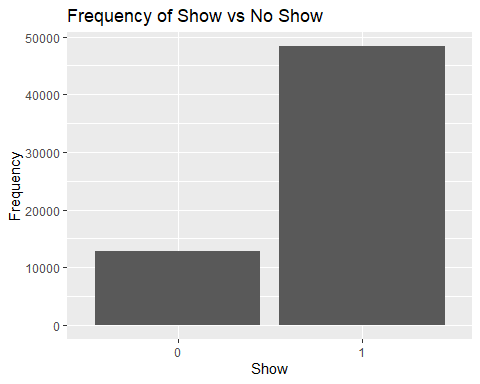
## c. Minimum time between appointment creation and reservation date: 1 days

**a. What is the average time between appointment creation and the actual reservation date?**   
Approximately 38.53 days.   
**b. What was the max time?**   
207 days.   
**c. What was the min time?**   
1 day.

# Calculate correlation between ReservationSpan and Show  
correlation <- cor(Appointments$ReservationSpan, Appointments$Show)  
print(correlation) # Print correlation coefficient

## [1] 0.002693853

# Create bar chart for Show variable  
ggplot(Appointments, aes(x = factor(Show))) +  
 geom\_bar() +  
 labs(x = "Show", y = "Frequency", title = "Frequency of Show vs No Show")



**a. Are ReservationSpan and Show highly correlated?**  
The correlation coefficient between ReservationSpan and Show is 0.002693853. Since this value is very close to zero, it indicates a very weak correlation between these two variables. Therefore, ReservationSpan and Show are not highly correlated.

**b. Based on 0 being “no show” and 1 being “show”, did most people make their appointments?**  
Yes, most people made their appointments.