# Module 6 - Assignment 1

## Angelo, Elizabeth

### Lubridate

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ 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

library(dplyr)  
  
Appointments <- Appointments %>%  
 mutate(ReservationYear = 2019)  
  
Appointments <- Appointments %>%  
 mutate(  
 ReservationDate = make\_date(year = ReservationYear, month = ReservationMonth, day = ReservationDay)  
 )  
  
Appointments <- Appointments %>%  
 mutate(  
 CreationDate = make\_date(year = CreationYear, month = CreationMonth, day = CreationDay)  
 )  
  
Appointments <- Appointments %>%  
 mutate(  
 ReservationSpan = as.numeric(difftime(ReservationDate, CreationDate, units = "days"))  
 )

### Exploratory Analysis

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

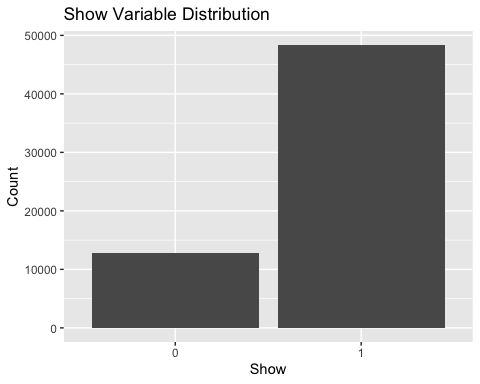
Appointments$ReservationSpan <- as.numeric(Appointments$ReservationSpan)  
summary(Appointments)

## 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

correlation <- cor(Appointments$ReservationSpan, Appointments$Show)  
cat("Correlation between ReservationSpan and Show:", correlation, "\n")

## Correlation between ReservationSpan and Show: 0.002693853

bar\_chart <- Appointments %>%  
 ggplot(aes(x = as.factor(Show))) +  
 geom\_bar() +  
 labs(title = "Show Variable Distribution", x = "Show", y = "Count")  
print(bar\_chart)



The average time between appointment creation and actual reservation date 38.53 days. The max amount of time is 207 days. The min amount of time is 1 day.  
  
With a correlation of 0.002693853 being close to 0, Reservation Span and Show are not highly correlated. This suggest little to no linear relationship between the number of days in advance an appointment is made, and whether the person showed up.

Based on the bar chart, most people did make it to their appointments and very few did not.