Module 6 Assignment 1: Data Wrangling with lubridate

Tabitha Hagen

2022-08-13

# Module 6 - Assignment 1

## Hagen, Tabitha

### Lubridate

# Part 1: Open libraries and Importing the dataset using readr

knitr::opts\_chunk$set(echo = TRUE)  
library(tidyverse)  
library(lubridate)  
library(readr)  
Appointments <- read\_csv("Appointments.csv")

# Part 2: Working with Dates

Appointments<- mutate(Appointments,ReservationDate = make\_date (month=ReservationMonth, day=ReservationDay, year='2019') )  
# Create a new column at the end of the table and inserted 2019 as ReservationYear which didn't exist  
  
Appointments<- mutate(Appointments,CreationDate = make\_date (month=CreationMonth, day=CreationDay, year=CreationYear) ) # Create a new column at the end of the table   
  
Appointments <-Appointments %>%  
 mutate (  
 ReservationSpan = interval(CreationDate,ReservationDate), # subtract dates  
 ReservationSpan = as.period(ReservationSpan), # put it in easy format  
 )

# Part 3 - Exploration Analysis

## Summary

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   
## ReservationDate CreationDate   
## Min. :2019-01-01 Min. :2018-10-03   
## 1st Qu.:2019-02-01 1st Qu.:2018-12-21   
## Median :2019-03-04 Median :2019-01-20   
## Mean :2019-03-01 Mean :2019-01-21   
## 3rd Qu.:2019-03-28 3rd Qu.:2019-02-19   
## Max. :2019-04-28 Max. :2019-03-28   
## ReservationSpan   
## Min. :1d 0H 0M 0S   
## 1st Qu.:25d 0H 0M 0S   
## Median :35d 10H 30M 0S   
## Mean :38d 18H 42M 27.6524977944791S   
## 3rd Qu.:49d 10H 30M 0S   
## Max. :207d 15H 0M 0S

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   
## ReservationDate CreationDate ReservationSpan   
## Min. :2019-01-01 Min. :2018-10-03 Min. : 86400   
## 1st Qu.:2019-02-01 1st Qu.:2018-12-21 1st Qu.: 2160000   
## Median :2019-03-04 Median :2019-01-20 Median : 3061800   
## Mean :2019-03-01 Mean :2019-01-21 Mean : 3350548   
## 3rd Qu.:2019-03-28 3rd Qu.:2019-02-19 3rd Qu.: 4271400   
## Max. :2019-04-28 Max. :2019-03-28 Max. :17938800

What is the average time between appointment creation and the actual reservation date? Mean : 3350548

What was the max time? Max. :17938800

What was the min time? Min. : 86400

## Correlation Test

dat <- data.frame(x=Appointments$Show, y=Appointments$ReservationSpan)  
  
head(dat) # Show first few rows

## x y  
## 1 1 4012200  
## 2 1 2073600  
## 3 1 2419200  
## 4 1 2802600  
## 5 1 3839400  
## 6 1 3234600

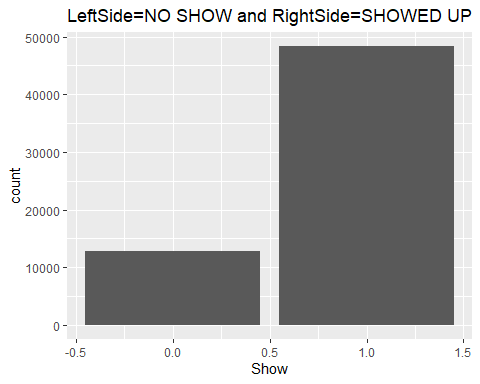
cor(dat$x,dat$y) # tells there is a high correlation between x and y at 0.002761478,

## [1] 0.002761478

Are ReservationSpan and Show highly correlated? A p-value coefficient of 0.002761478 says there is not much of a linear relationship between ReservationSpan and Show

## Box plot of the Variable Show

ggplot(Appointments)+  
 geom\_bar(aes(x=Show, fill=Show))+  
 ggtitle(label = "LeftSide=NO SHOW and RightSide=SHOWED UP")



#coord\_cartesian(ylim = c(-1, 2)) # I set the y axis scale so the plot looks better.

Based on 0 being “no show” and 1 being “show”, did most people make their appointments? YES