Exercise 3 (Web Scraping)

Georg Pretzner (wi19b013)

26 9 2021

Contents

Libraries

Load the tidyverse library

```
library(tidyverse)
```

1 Assignment

1. Create a scatterplot of the number of crossings per day volume against the high temperature that day
2. Separate your plot into facets by weekday
3. Add regression lines to the two facets

2 Dataset

Load the dataset `RailTrail` from the package `mosaicData` is loaded

```
data(RailTrail, package = "mosaicData")
head(RailTrail)
```

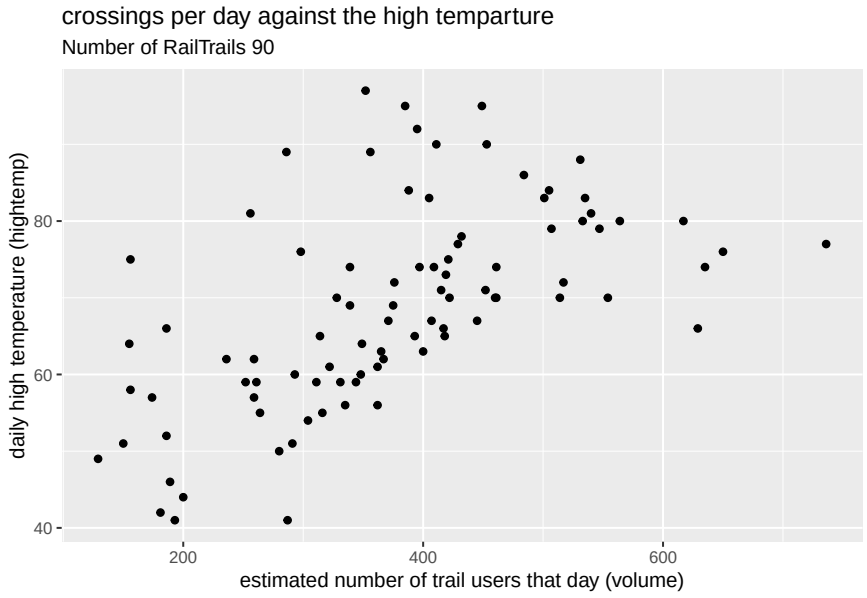
```
##   hightemp lowtemp avgtemp spring summer fall cloudcover precip volume weekday
## 1       83      50    66.5      0      1    0        7.6   0.00    501    TRUE
## 2       73      49    61.0      0      1    0        6.3   0.29    419    TRUE
## 3       74      52    63.0      1      0    0        7.5   0.32    397    TRUE
## 4       95      61    78.0      0      1    0        2.6   0.00    385   FALSE
## 5       44      52    48.0      1      0    0       10.0   0.14    200    TRUE
## 6       69      54    61.5      1      0    0        6.6   0.02    375    TRUE
##   dayType
## 1 weekday
## 2 weekday
## 3 weekday
## 4 weekend
## 5 weekday
## 6 weekday
```

3 Visualization

3.1 Scatterplot Volume / HighTemp

Scatterplot of the number of crossings per day `volume` against the high temparture that day

```
g <- ggplot(RailTrail, aes(x=volume, y=hightemp)) +
  geom_point() +
  xlab("estimated number of trail users that day (volume)") +
  ylab("daily high temperature (hightemp)") +
  labs(title = "crossings per day against the high temparture ", subtitle = paste("Number of RailTrails
g
```



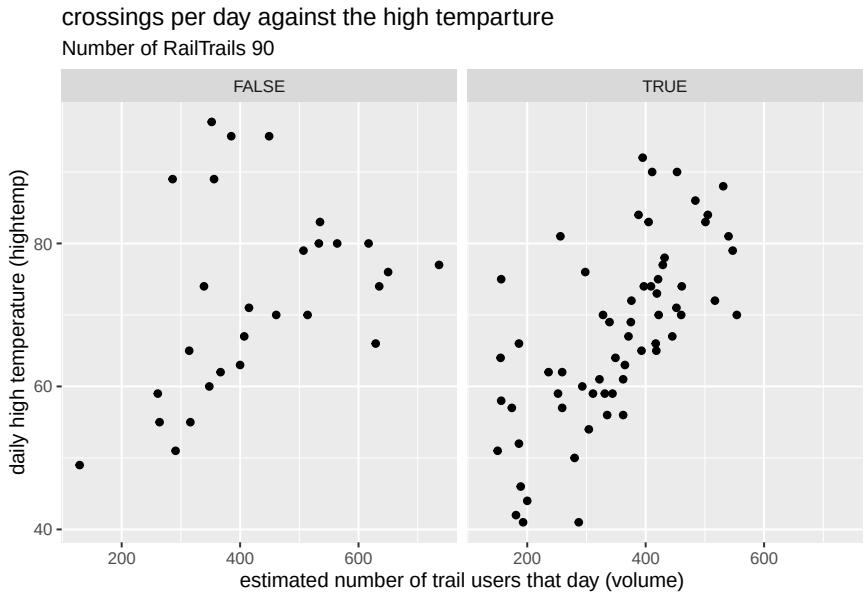
3.1.1 Interpretation

It can be seen that the higher the temperature, the higher the volume. However, when it gets too hot, the volume decreases again a little.

3.2 Facets

Seperate the plot into facets by `weekday`

```
g2 <- g + facet_wrap(~weekday)
g2
```



3.2.1 Interpretation

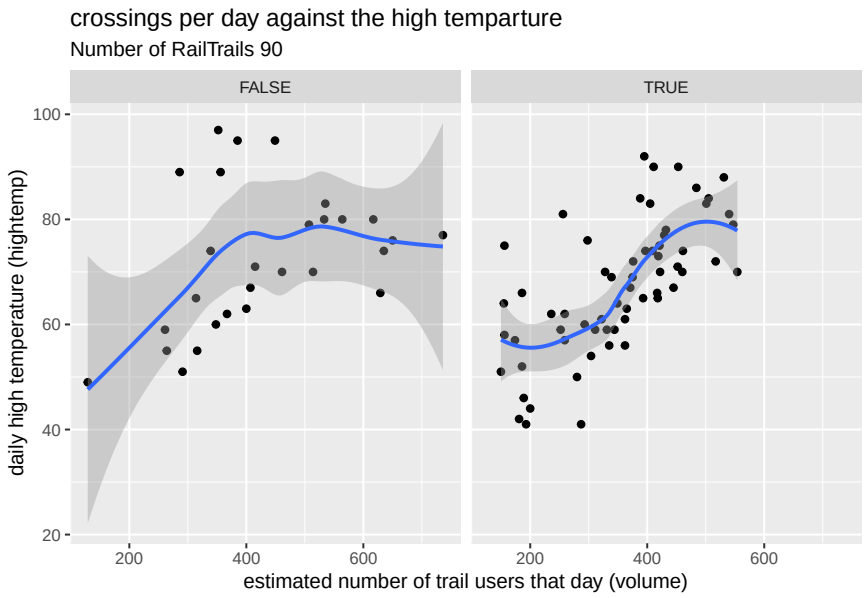
The Facet shows that the volume is slightly lower on weekdays than on weekends. Also that on weekdays there is a certain amount of low temperatures, which is not the case on weekends.

3.3 Regression Line

Add regression lines to the two facets

```
g3 <- g2 + geom_smooth()
g3
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



3.3.1 Interpretation

The regression line confirms our assumption of the scatterplot. With increasing temperature, the volume also increases, but weakens again from about 80°F onwards.