Rowery Waszyngton

Michał Syrkiewicz

Loading used libraries. Devtools used to load "R-package" my own small package made for this assignment. "R-package" needs to be present in project folder in order to work.

Reading and saving quarterly data from Washington bikes All needed csv-s in zip folder.

```
Q1 <- read.csv("2013Q1-capitalbikeshare-tripdata.csv")
Q2 <- read.csv("2013Q2-capitalbikeshare-tripdata.csv")
Q3 <- read.csv("2013Q3-capitalbikeshare-tripdata.csv")
Q4 <-read.csv("2013Q4-capitalbikeshare-tripdata.csv")
```

Extracting month and year from every dataset to group by day. Making two separate frames for "Members" and "Casual" users.

```
Q1 <- wyodrebnij(Q1,name="Start.date",from="%Y-%m-%d",what = c("%d","%m")
                 , newName = c("Day","Month"))
Q1 <- Q1 %>% group_by(Day,Month) %>% count(Member.type) %>%
  mutate(Data = paste("2013",Month,Day,sep="-"))
casual <- Q1[Q1$Member.type=="Casual",]
registere <<- Q1[Q1$Member.type=="Member",]
```

```
Q2 <- wyodrebnij(Q2,name="Start.date",from="%Y-%m-%d",what = c("%d","%m")
                 , newName = c("Day","Month"))
Q2 <- Q2 %>% group_by(Day,Month) %>% count(Member.type) %>%
  mutate(Data = paste("2013",Month,Day,sep="-"))
casual1 <- Q2[Q2$Member.type=="Casual",]
registere1 <- Q2[Q2$Member.type=="Member",]
```

```
Q3 <- wyodrebnij(Q3,name="Start.date",from="%Y-%m-%d",what = c("%d","%m")
                 ,newName = c("Day","Month"))
Q3 <- Q3 %>% group_by(Day,Month) %>% count(Member.type)  %>%
  mutate(Data = paste("2013",Month,Day,sep="-"))
casual2 <- Q3[Q3$Member.type=="Casual",]
registere2 <- Q3[Q3$Member.type=="Member",]
```

```
Q4 <- wyodrebnij(Q4,name="Start.date",from="%Y-%m-%d",what = c("%d","%m")
                 , newName = c("Day","Month"))
Q4 <- Q4 %>% group_by(Day,Month) %>% count(Member.type) %>%
  mutate(Data = paste("2013",Month,Day,sep="-"))
casual3 <- Q4[Q4$Member.type=="Casual",]
registere3 <- Q4[Q4$Member.type=="Member",]
```

Merging all data from year 2013

```
casuals<-rbind(casual,casual1,casual2,casual3)
registered<-rbind(registere,registere1,registere2,registere3)
```

Reading and saving weather data from Washington 2013

```
pogoda2013 <- read.csv("2013_DC.csv",skip=3)
```

Small aestethic changes and joining weather and bike data

```
colnames(registered)[4]<- "registered"
colnames(casuals)[4]<- "casual"
registered<- registered[,c(-3)]
casuals<- casuals[,c(-3)]
wynik<- inner_join(casuals, pogoda2013, join_by(Data == time))
wynik <- inner_join(wynik,registered, by=(join_by(Day,Month,Data)))
```

Looking for correlation in data

```
wynik <- wynik[,c(-1,-2)]
wynik<- wynik[,c(2,1,20,3:19)]
correlation(wynik, c(2:3), c(4:20), high=0.65)
```

```
## Sprawdzam 2 i 4
##        temperature_2m_max...C.
## casual                0.716257
## Sprawdzam 2 i 6
##        temperature_2m_mean...C.
## casual                0.6961752
## Sprawdzam 2 i 7
##        apparent_temperature_max...C.
## casual                     0.6939996
## Sprawdzam 2 i 8
##        apparent_temperature_min...C.
## casual                      0.657586
## Sprawdzam 2 i 9
##        apparent_temperature_mean...C.
## casual                      0.6839694
## Sprawdzam 2 i 19
##        shortwave_radiation_sum..MJ.m..
## casual                       0.6778133
## Sprawdzam 2 i 20
##        et0_fao_evapotranspiration..mm.
## casual                       0.7585045
## Sprawdzam 3 i 4
##            temperature_2m_max...C.
## registered               0.6598518
```

As you can see we have high correlation with evapotranspiration, temperature and shortwave radiation. I'll try making new columns with these columns

```
wynik <- wynik %>% mutate(Stosunki1 = temperature_2m_max...C./et0_fao_evapotranspiration..mm.,
                          Stosunki2 = temperature_2m_mean...C./et0_fao_evapotranspiration..mm.,
                          Stosunki3 = apparent_temperature_max...C./et0_fao_evapotranspiration..mm.,
                          Stosunki4 = shortwave_radiation_sum..MJ.m../et0_fao_evapotranspiration..mm.,
                          Inne = shortwave_radiation_sum..MJ.m..*apparent_temperature_max...C.,
                          Inne2 = shortwave_radiation_sum..MJ.m..*temperature_2m_max...C.,
                          Inne3 = shortwave_radiation_sum..MJ.m..*et0_fao_evapotranspiration..mm.)
correlation(wynik,c(2,3),c(4:27),high=0.7)
```

```
## Sprawdzam 2 i 4
##        temperature_2m_max...C.
## casual                0.716257
## Sprawdzam 2 i 20
##        et0_fao_evapotranspiration..mm.
## casual                        0.7585045
## Sprawdzam 2 i 25
##             Inne
## casual 0.7884978
## Sprawdzam 2 i 26
##            Inne2
## casual 0.7948381
## Sprawdzam 2 i 27
##           Inne3
## casual 0.731267
```

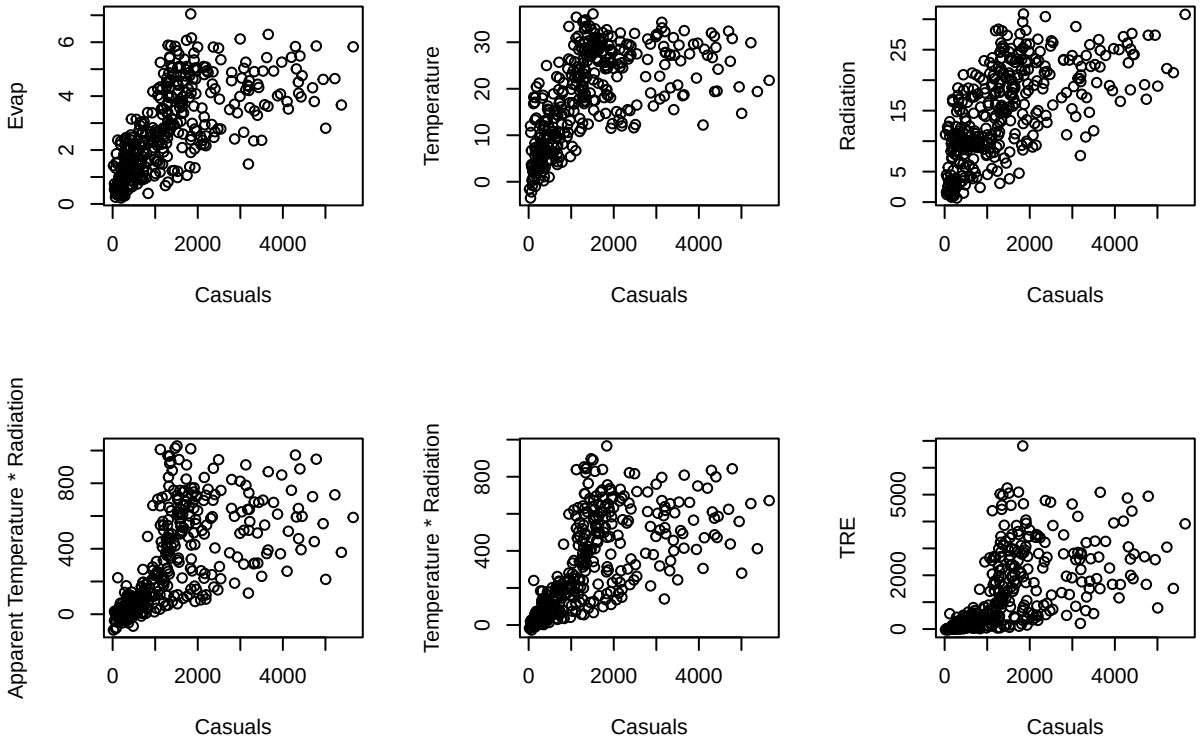
As you can see we have higher correlation with products of these columns trying one more time.

```
wynik <- wynik %>% mutate(Inne4 = Inne2 * et0_fao_evapotranspiration..mm.)
correlation(wynik,c(2,3),c(4:28),high=0.7)
```

```
## Sprawdzam 2 i 4
##        temperature_2m_max...C.
## casual                0.716257
## Sprawdzam 2 i 20
##        et0_fao_evapotranspiration..mm.
## casual                        0.7585045
## Sprawdzam 2 i 25
##             Inne
## casual 0.7884978
## Sprawdzam 2 i 26
##            Inne2
## casual 0.7948381
## Sprawdzam 2 i 27
##           Inne3
## casual 0.731267
## Sprawdzam 2 i 28
##            Inne4
## casual 0.7930774
```

Let's plot those columns in order to find more correlation inside data.

```
par(mfrow = c(2, 3))
plot(wynik$casual,wynik$et0_fao_evapotranspiration..mm., xlab = "Casuals", ylab="Evap")
plot(wynik$casual,wynik$temperature_2m_max...C., xlab = "Casuals", ylab="Temperature")
plot(wynik$casual,wynik$shortwave_radiation_sum..MJ.m.., xlab = "Casuals", ylab="Radiation")
plot(wynik$casual,wynik$Inne, xlab = "Casuals", ylab="Apparent Temperature * Radiation")
plot(wynik$casual,wynik$Inne2, xlab = "Casuals", ylab="Temperature * Radiation")
plot(wynik$casual,wynik$Inne4, xlab = "Casuals",ylab="TRE")
```



As we can see days with more than around 2400 casual bikers tend to be different than most day of the year. Let's eliminate them.

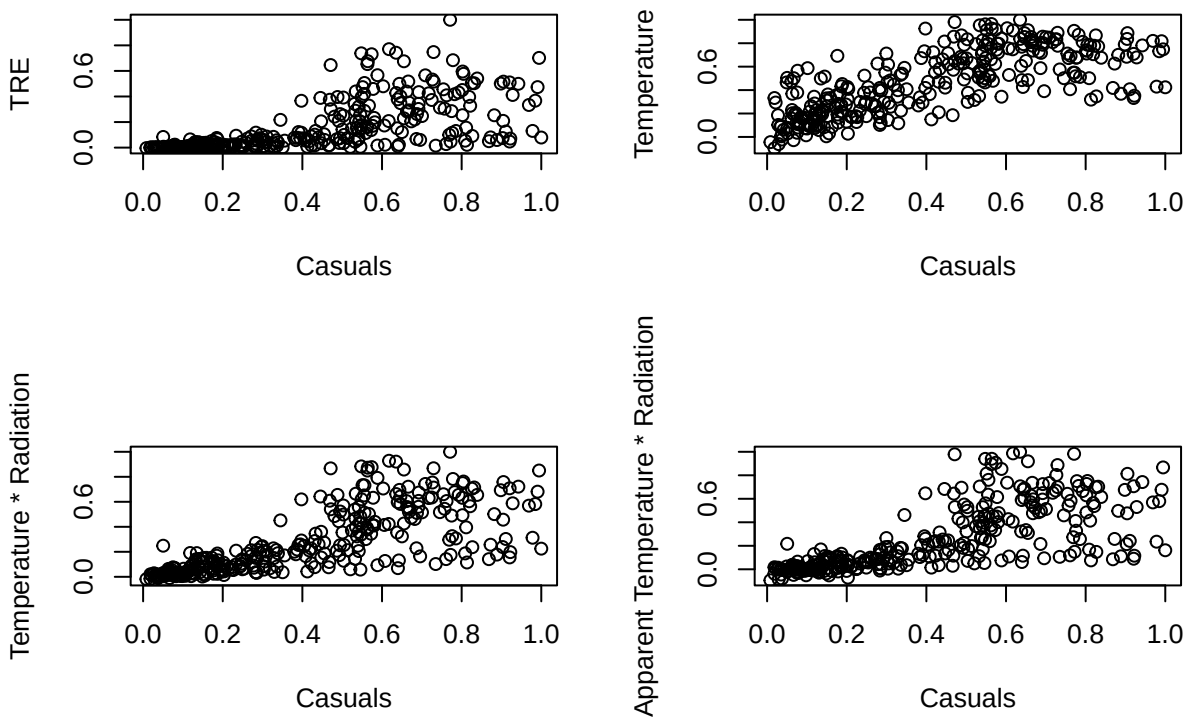
```
wynik1<-wynik[wynik$casual<2400,]
correlation(wynik1,c(2,3),c(4:28),high=0.75, method="spearman")
```

```
## Sprawdzam 2 i 4
##        temperature_2m_max...C.
## casual               0.7505447
## Sprawdzam 2 i 25
##             Inne
## casual 0.8082074
## Sprawdzam 2 i 26
##            Inne2
## casual 0.8111141
## Sprawdzam 2 i 28
##            Inne4
## casual 0.8028577
```

```
## Sprawdzam 3 i 25
##                  Inne
## registered 0.7831748
## Sprawdzam 3 i 26
##                Inne2
## registered 0.786002
## Sprawdzam 3 i 28
##                 Inne4
## registered 0.7751058
```

Let's normalize data in second data frame

```
wynik2 <- unormuj(wynik1,which = c(2:28),check=FALSE,decimals = 7)
par(mfrow = c(2, 2))
plot(wynik2$casual,wynik2$Inne4, xlab = "Casuals",ylab="TRE")
plot(wynik2$casual,wynik2$temperature_2m_max...C., xlab = "Casuals", ylab="Temperature")
plot(wynik2$casual,wynik2$Inne2, xlab = "Casuals", ylab="Temperature * Radiation")
plot(wynik2$casual,wynik2$Inne, xlab = "Casuals",ylab="Apparent Temperature * Radiation")
```



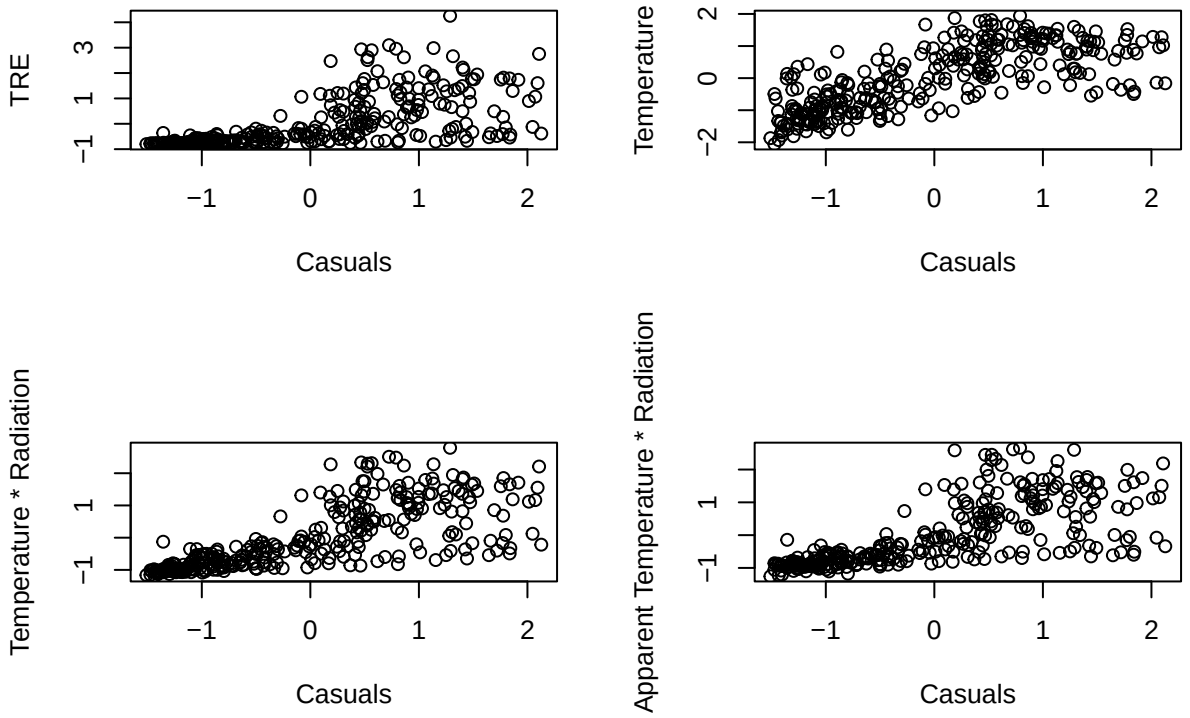

```
correlation(wynik1,c(2,3),c(4:28),high=0.75, method="spearman")
```

```
## Sprawdzam 2 i 4
##        temperature_2m_max...C.
## casual               0.7505447
## Sprawdzam 2 i 25
```

```
##             Inne
## casual 0.8082074
## Sprawdzam 2 i 26
##            Inne2
## casual 0.8111141
## Sprawdzam 2 i 28
##            Inne4
## casual 0.8028577
## Sprawdzam 3 i 25
##                 Inne
## registered 0.7831748
## Sprawdzam 3 i 26
##                Inne2
## registered 0.786002
## Sprawdzam 3 i 28
##                Inne4
## registered 0.7751058
```

Let's standarize data in 3 data frame

```
wynik3<-as.data.frame(sapply(wynik1[,-1], function(data) (data-mean(data))/sd(data)))
par(mfrow = c(2, 2))
plot(wynik3$casual,wynik3$Inne4, xlab = "Casuals",ylab="TRE")
plot(wynik3$casual,wynik3$temperature_2m_max...C., xlab = "Casuals", ylab="Temperature")
plot(wynik3$casual,wynik3$Inne2, xlab = "Casuals", ylab="Temperature * Radiation")
plot(wynik3$casual,wynik3$Inne, xlab = "Casuals",ylab="Apparent Temperature * Radiation")
```



```r
correlation(wynik1,c(2,3),c(4:28),high=0.75, method="spearman")
```

```
## Sprawdzam 2 i 4
##        temperature_2m_max...C.
## casual               0.7505447
## Sprawdzam 2 i 25
##             Inne
## casual 0.8082074
## Sprawdzam 2 i 26
##            Inne2
## casual 0.8111141
## Sprawdzam 2 i 28
##            Inne4
## casual 0.8028577
## Sprawdzam 3 i 25
##                 Inne
## registered 0.7831748
## Sprawdzam 3 i 26
##               Inne2
## registered 0.786002
## Sprawdzam 3 i 28
##                Inne4
## registered 0.7751058
```