Washington Bikes

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$Member.type=="Casual",]
registered <<- Q1[Q1$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$Member.type=="Casual",]
registered1 <- Q2[Q2$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$Member.type=="Casual",]
registered2 <- Q3[Q3$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$Member.type=="Casual",]
registered3 <- Q4[Q4$Member.type=="Member",]
```

Merging all data from year 2013

```
casuals<-rbind(casual,casual1,casual2,casual3)
registered<-rbind(registered,registered1,registered2,registered3)
head(casual)
```

```
## # A tibble: 6 x 5
## # Groups:   Day, Month [6]
##   Day   Month Member.type     n Data
##   <chr> <chr> <chr>       <int> <chr>
## 1 01    01    Casual        421 2013-01-01
## 2 01    02    Casual         84 2013-02-01
## 3 01    03    Casual        329 2013-03-01
## 4 02    01    Casual        284 2013-01-02
## 5 02    02    Casual        233 2013-02-02
## 6 02    03    Casual        674 2013-03-02
```

```
head(registered)
```

```
## # A tibble: 6 x 5
## # Groups:   Day, Month [6]
##   Day   Month Member.type     n Data
##   <chr> <chr> <chr>       <int> <chr>
## 1 01    01    Member       1543 2013-01-01
## 2 01    02    Member       3077 2013-02-01
## 3 01    03    Member       4557 2013-03-01
## 4 02    01    Member       3389 2013-01-02
## 5 02    02    Member       2217 2013-02-02
## 6 02    03    Member       3033 2013-03-02
```

Reading and saving weather data from Washington 2013

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

```
##         time temperature_2m_max...C. temperature_2m_min...C.
## 1 2013-01-01                    6.9                     1.5
## 2 2013-01-02                    2.7                    -3.1
## 3 2013-01-03                    3.5                    -3.4
## 4 2013-01-04                    5.9                    -3.3
## 5 2013-01-05                    6.2                    -1.0
## 6 2013-01-06                   10.2                    -0.2
##   temperature_2m_mean...C. apparent_temperature_max...C.
## 1                      3.8                           3.7
## 2                      0.4                          -1.2
## 3                     -0.4                          -0.7
## 4                      0.5                           0.6
## 5                      1.5                           3.0
## 6                      4.6                           6.5
##   apparent_temperature_min...C. apparent_temperature_mean...C.
## 1                          -2.3                            0.4
## 2                          -7.0                           -4.0
## 3                          -7.3                           -4.3
## 4                          -7.5                           -4.3
```

```
## 5                         -5.2                          -2.6
## 6                         -3.4                           1.2
##   daylight_duration..s. sunshine_duration..s. precipitation_sum..mm.
## 1              34207.39               492.90                    0.0
## 2              34250.91             30479.74                    0.0
## 3              34297.74             29260.52                    0.0
## 4              34347.81             30597.60                    0.0
## 5              34401.04             30424.26                    0.0
## 6              34457.38             30705.14                    0.1
##   rain_sum..mm. snowfall_sum..cm. precipitation_hours..h.
## 1           0.0                0                       0
## 2           0.0                0                       0
## 3           0.0                0                       0
## 4           0.0                0                       0
## 5           0.0                0                       0
## 6           0.1                0                       1
##   wind_speed_10m_max..km.h. wind_gusts_10m_max..km.h.
## 1                      11.3                      28.1
## 2                      14.8                      36.4
## 3                       9.8                      22.3
## 4                      18.6                      42.5
## 5                      16.3                      32.0
## 6                      12.3                      24.8
##   wind_direction_10m_dominant.... shortwave_radiation_sum..MJ.m..
## 1                             292                            3.47
## 2                             336                            9.82
## 3                             266                            8.42
## 4                             274                            9.70
## 5                             290                           10.14
## 6                             229                            9.81
##   et0_fao_evapotranspiration..mm.
## 1                            0.73
## 2                            1.18
## 3                            0.93
## 4                            1.33
## 5                            1.22
## 6                            1.28
```

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.

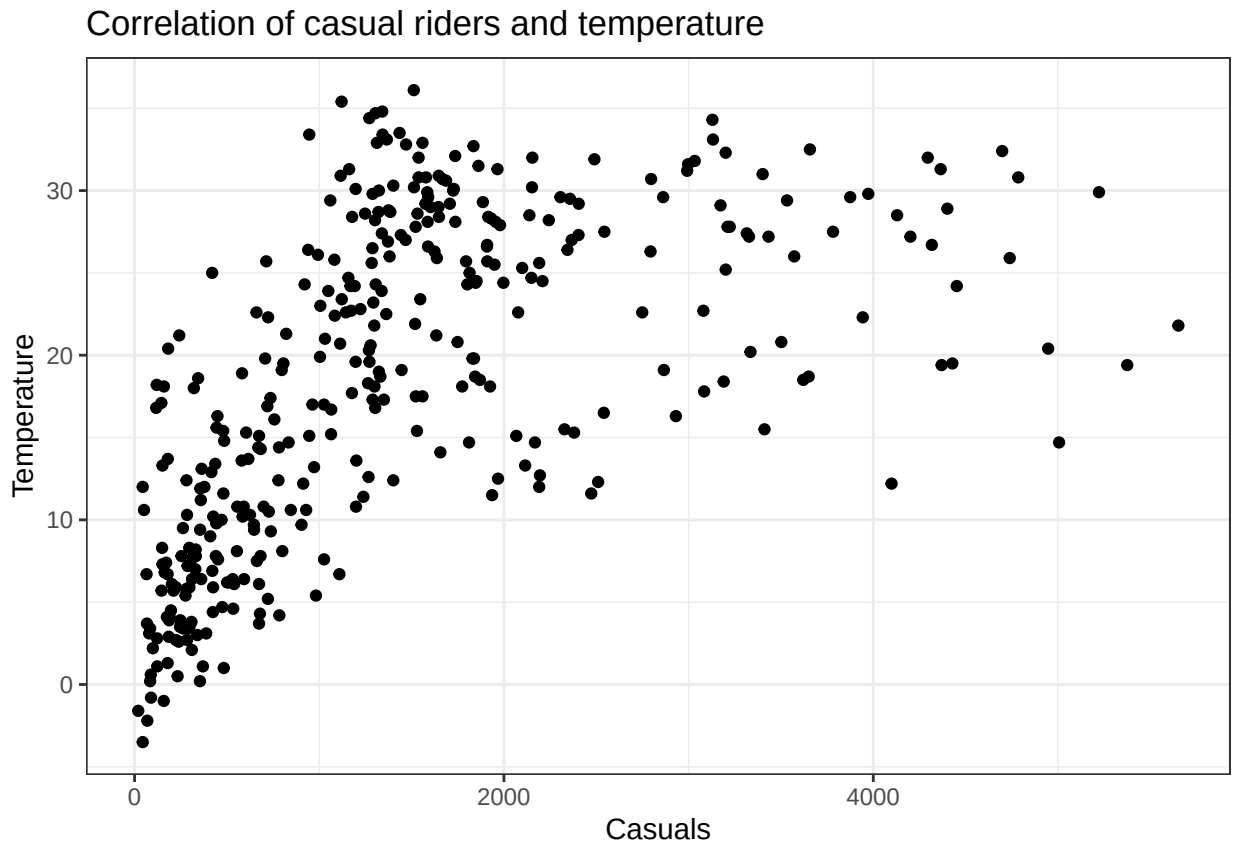
```
wynik <- wynik %>% mutate(Inne4 = Inne2 * et0_fao_evapotranspiration..mm.)
correlation(wynik,c(2,3),c(4:28),high=0.7)
```

```
## 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
```

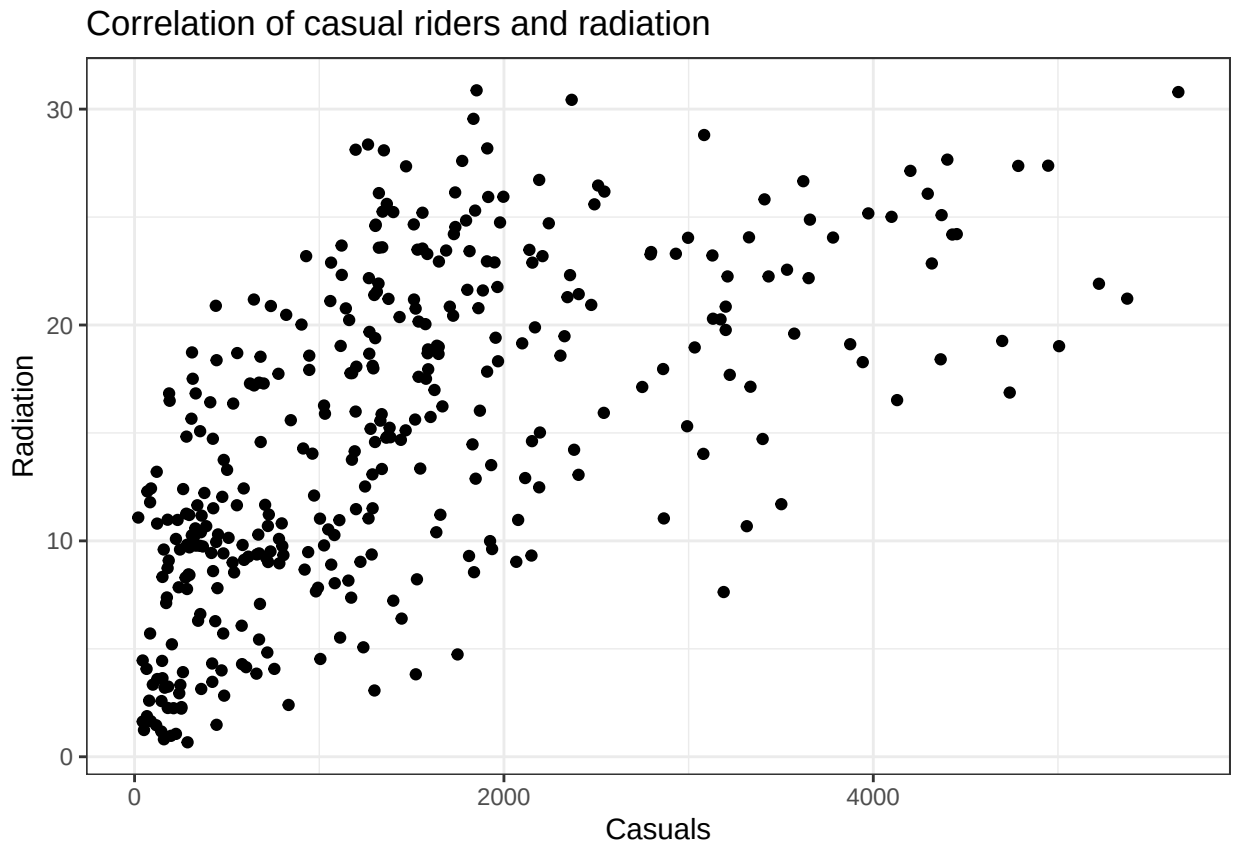
Let's plot those columns in order to find more correlation inside data.

```
p1 <- ggplot(wynik, aes(x = casual, y = temperature_2m_max...C.)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and temperature", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temperature") + # Label for y-axis
  theme_bw()
p2 <- ggplot(wynik, aes(x = casual, y = shortwave_radiation_sum..MJ.m..)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and radiation", # Add title
       x = "Casuals", # Label for x-axis
       y = "Radiation") + # Label for y-axis
  theme_bw()


p3 <- ggplot(wynik, aes(x = casual, y = et0_fao_evapotranspiration..mm.)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and evapotranspiration", # Add title
       x = "Casuals", # Label for x-axis
       y = "Evapotranspiration") + # Label for y-axis
  theme_bw()
p5 <- ggplot(wynik, aes(x = casual, y = Inne)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and Temp * Radiation", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temp * Radiation") + # Label for y-axis
  theme_bw()
p6 <- ggplot(wynik, aes(x = casual, y = Inne2)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and Temp * Radiation * Evap", # Add title
```

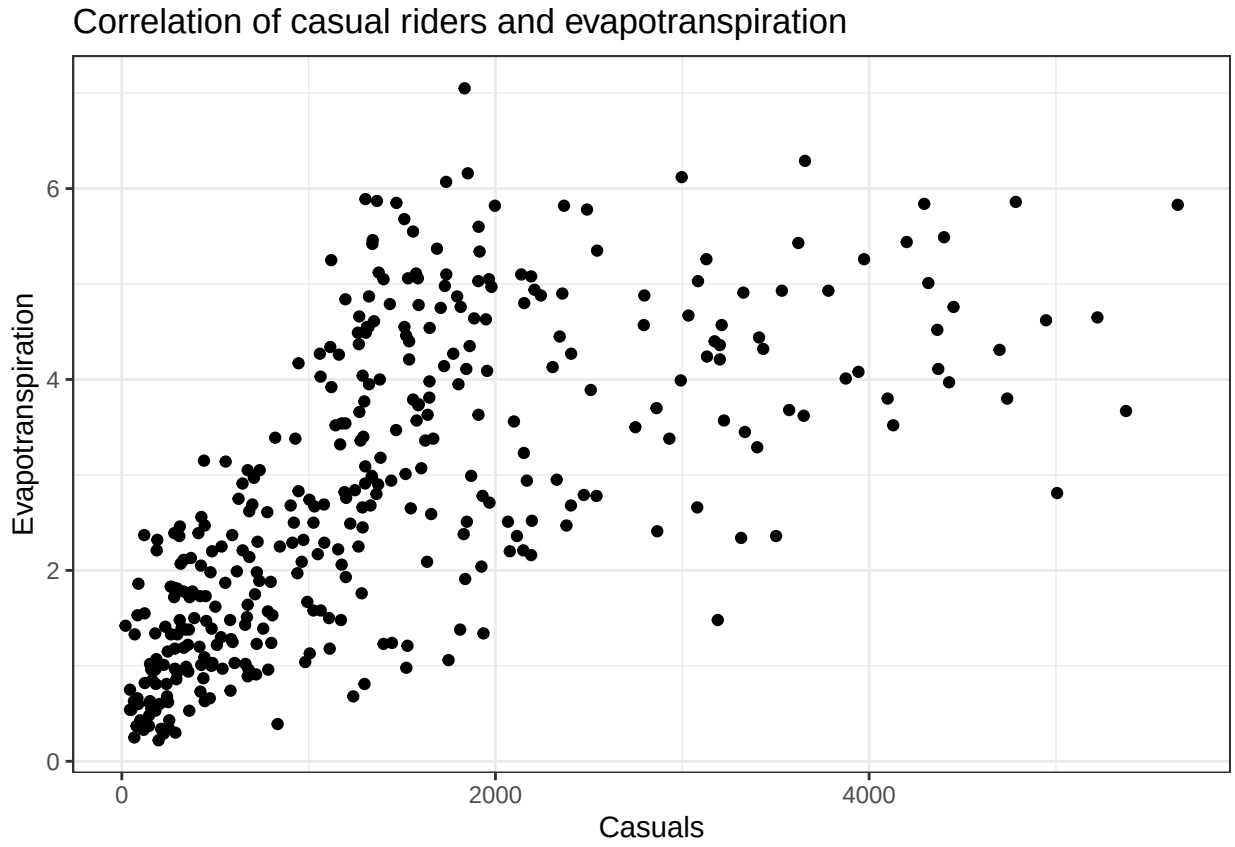
```
       x = "Casuals", # Label for x-axis
       y = "Temp * Radiation *Evap") + # Label for y-axis
  theme_bw()
p4 <- ggplot(wynik, aes(x = casual, y = temperature_2m_max...C., color = shortwave_radiation_sum..MJ.m.
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and temperature", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temperature") + # Label for y-axis
  theme_classic() +
  scale_color_gradient(low = "red", high = "blue")

#Będę wypisywał po kolei bo z użcyiem grida wykresy na siebie najeżdzają tworząc bardzo brzydki efekt
p1
```



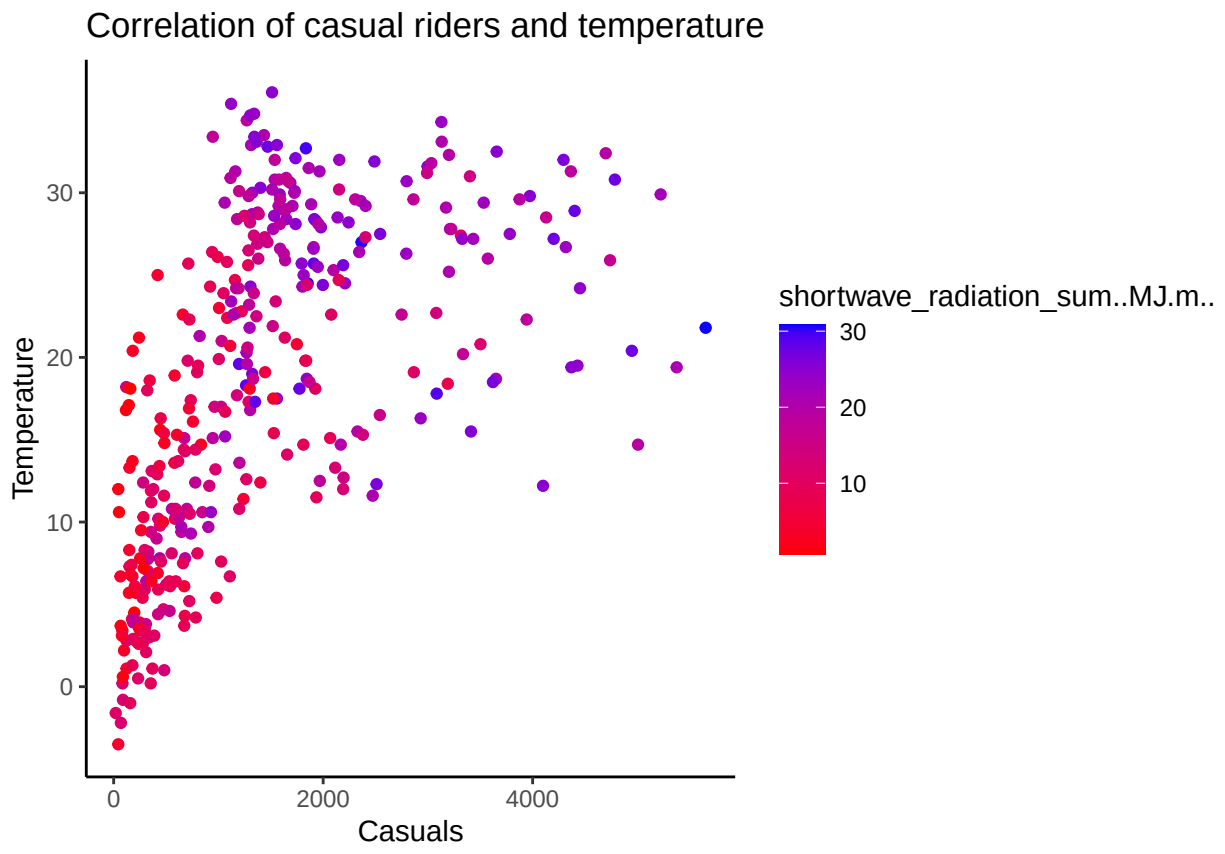

```
p2
```


Correlation of casual riders and radiation
Radiation
Casuals
0
10
20
30
0
2000
4000

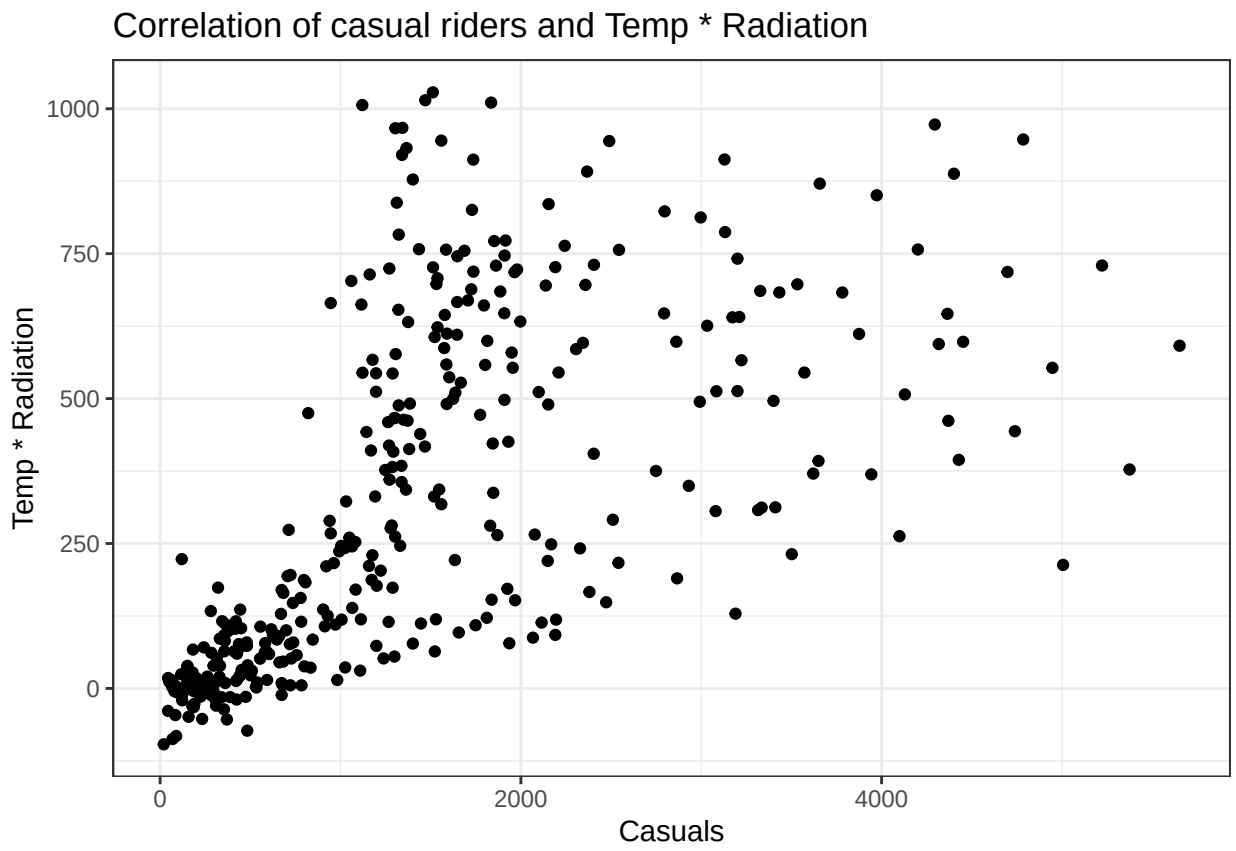
```
p3
```


Correlation of casual riders and evapotranspiration
Evapotranspiration
Casuals
0
2
4
6
0
2000
4000

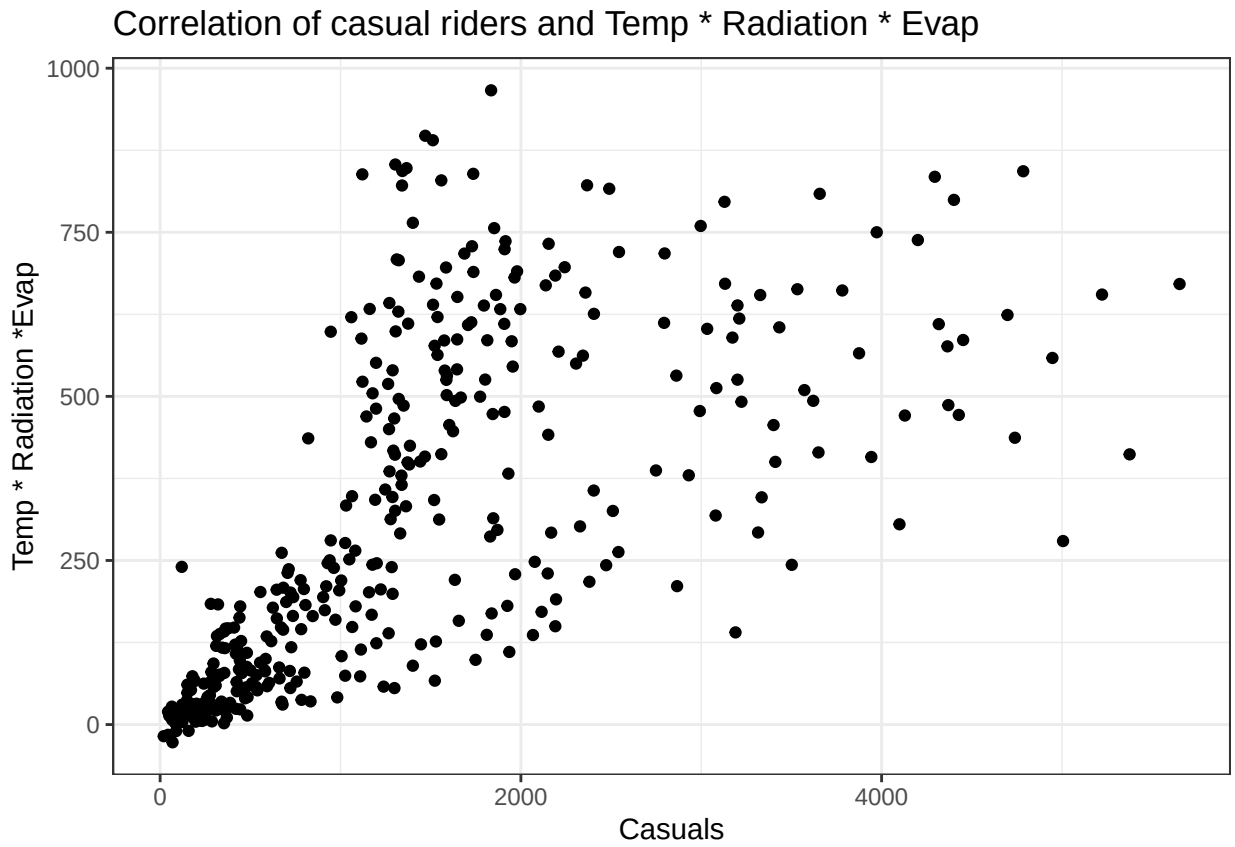
```
p4
```


Correlation of casual riders and temperature
Temperature
30
20
10
0
0
2000
4000
Casuals
shortwave_radiation_sum..MJ.m..
30
20
10

```
p5
```


Correlation of casual riders and Temp * Radiation
Temp * Radiation
1000
750
500
250
0
0
2000
4000
Casuals

```
p6
```



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)
p1 <- ggplot(wynik2, aes(x = casual, y = temperature_2m_max...C.)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and temperature (normed)", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temperature") + # Label for y-axis
  theme_bw()
p2 <- ggplot(wynik2, aes(x = casual, y = shortwave_radiation_sum..MJ.m..)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and radiation (normed)", # Add title
       x = "Casuals", # Label for x-axis
       y = "Radiation") + # Label for y-axis
  theme_bw()


p3 <- ggplot(wynik2, aes(x = casual, y = et0_fao_evapotranspiration..mm.)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and evapotranspiration (normed)", # Add title
       x = "Casuals", # Label for x-axis
       y = "Evapotranspiration") + # Label for y-axis
  theme_bw()
p5 <- ggplot(wynik2, aes(x = casual, y = Inne)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and Temp * Radiation (normed)", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temp * Radiation") + # Label for y-axis
  theme_bw()
p6 <- ggplot(wynik2, aes(x = casual, y = Inne2)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and Temp * Radiation * Evap (normed)", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temp * Radiation *Evap") + # Label for y-axis
  theme_bw()
p4 <- ggplot(wynik2, aes(x = casual, y = temperature_2m_max...C., color = shortwave_radiation_sum..MJ.m
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and temperature (normed)", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temperature") + # Label for y-axis
  theme_classic() +
  scale_color_gradient(low = "red", high = "blue")
#Nie wyświetlam są zapisane w pamięci ale wyglądają takk samo jak nieunormowane (zmienia się tylko prze
```

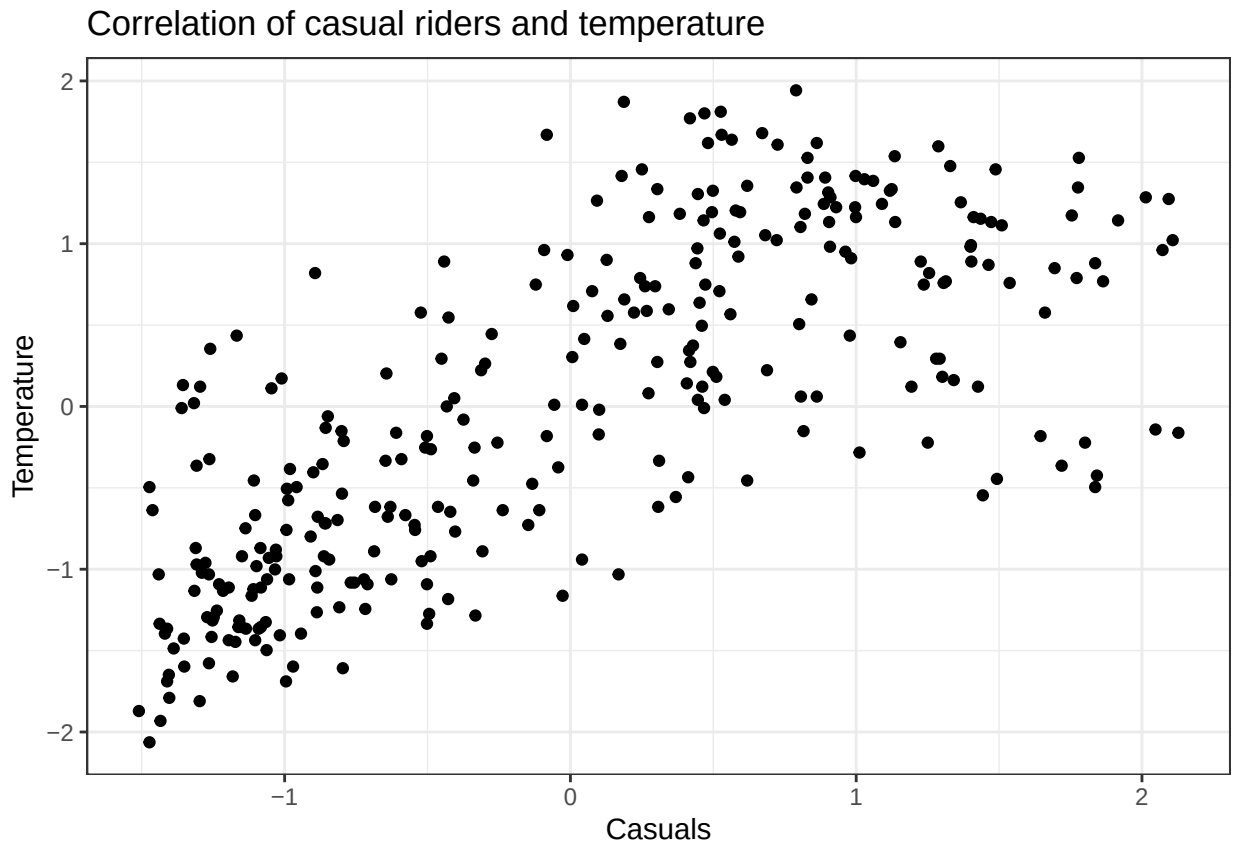
```
correlation(wynik1,c(2,3),c(4:28),high=0.75, method="spearman")
```

```
## Sprawdzam 2 i 4
##        temperature_2m_max...C.
## casual               0.7505447
## Sprawdzam 2 i 25
##             Inne
## casual 0.8082074
## Sprawdzam 2 i 26
```

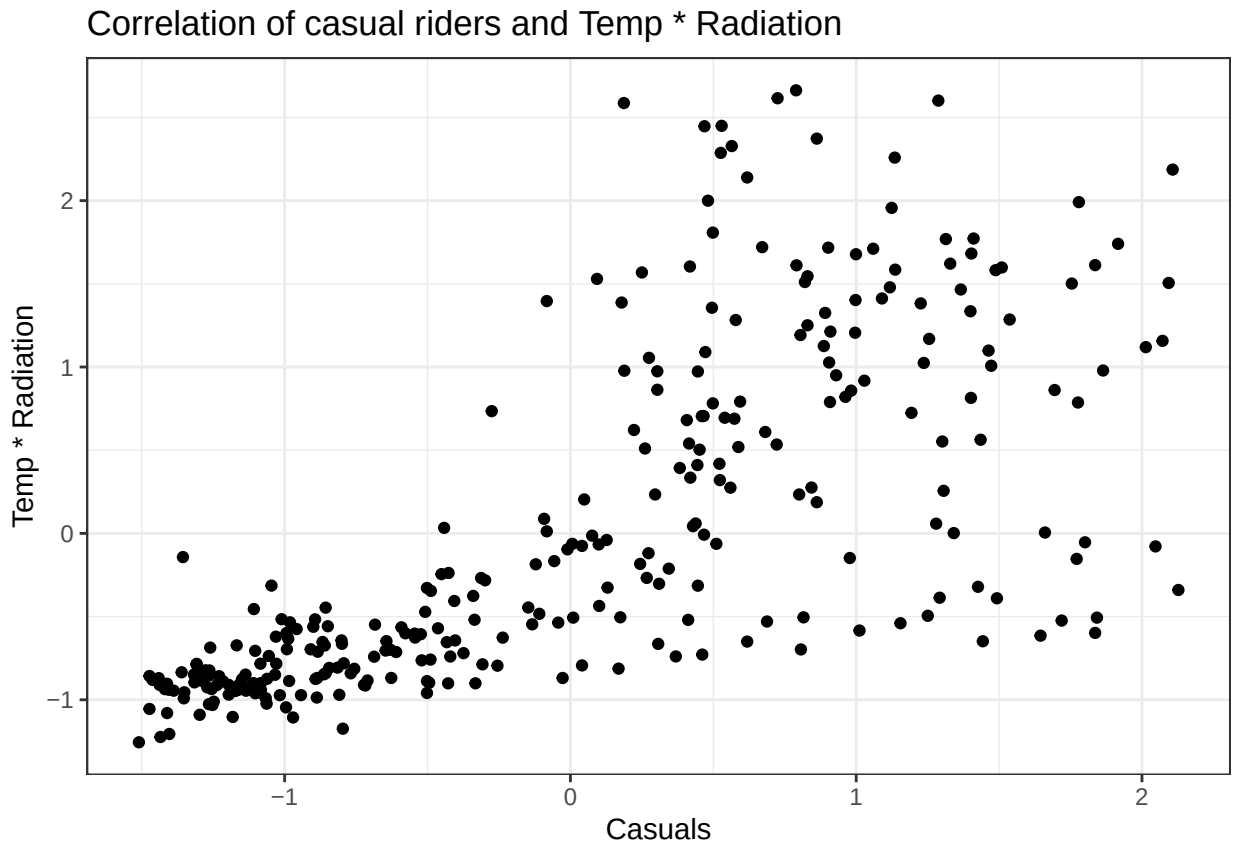
```
##            Inne2
## casual 0.8111141
## Sprawdzam 2 i 28
##            Inne4
## casual 0.8028577
## Sprawdzam 3 i 25
##                 Inne
## registered 0.7831748
## Sprawdzam 3 i 26
##               Inne2
## registered 0.786002
## Sprawdzam 3 i 28
##                Inne4
## registered 0.7751058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

Let's standarize data in 3 data frame

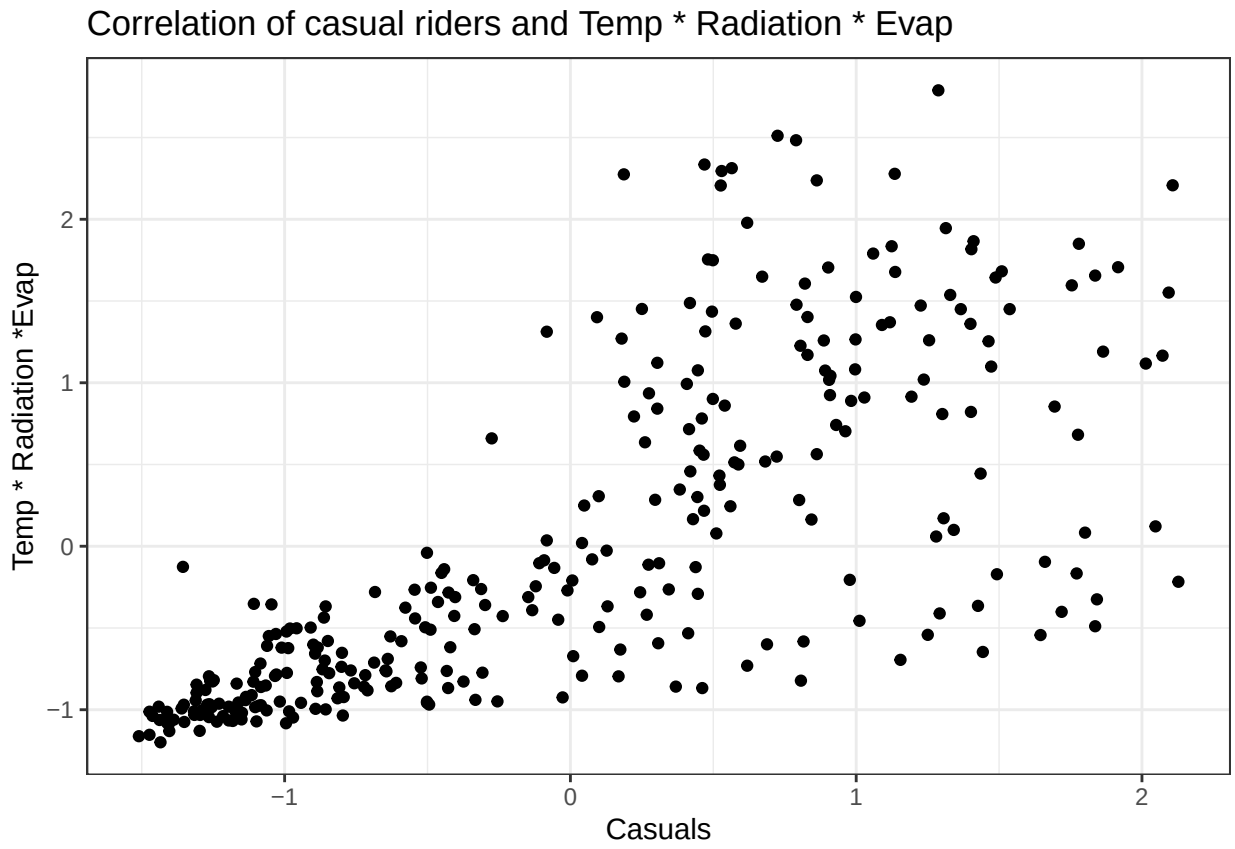
```
wynik3<-as.data.frame(sapply(wynik1[,-1], function(data) (data-mean(data))/sd(data)))
p1 <- ggplot(wynik3, aes(x = casual, y = temperature_2m_max...C.)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and temperature", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temperature") + # Label for y-axis
  theme_bw()
p2 <- ggplot(wynik3, aes(x = casual, y = Inne)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and Temp * Radiation", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temp * Radiation") + # Label for y-axis
  theme_bw()
p3 <- ggplot(wynik3, aes(x = casual, y = Inne2)) +
  geom_point() + # Add points
  labs(title = "Correlation of casual riders and Temp * Radiation * Evap", # Add title
       x = "Casuals", # Label for x-axis
       y = "Temp * Radiation *Evap") + # Label for y-axis
  theme_bw()
p1
```


Correlation of casual riders and temperature
Temperature
Casuals
2
1
0
−1
−2
−1
0
1
2

```
p2
```


Correlation of casual riders and Temp * Radiation
Temp * Radiation
Casuals
2
1
0
−1
−1
0
1
2

```
p3
```




```
correlation(wynik3,c(1,2),c(4:27),high=0.75, method="spearman")
```

```
## Sprawdzam 1 i 24
## [1] 0.8082074
## Sprawdzam 1 i 25
## [1] 0.8111141
## Sprawdzam 1 i 27
## [1] 0.8028577
## Sprawdzam 2 i 24
## [1] 0.7831748
## Sprawdzam 2 i 25
## [1] 0.786002
## Sprawdzam 2 i 27
## [1] 0.7751058
```