

Ruch rowerowy a pogoda

2026-01-03

```
library(sf)

## Linking to GEOS 3.13.1, GDAL 3.11.4, PROJ 9.7.0; sf_use_s2() is TRUE

library(ggplot2)
library(rmapshaper)
library(RColorBrewer)
library(dplyr)

##
## Dołączanie pakietu: 'dplyr'

## Następujące obiekty zostały zakryte z 'package:stats':
##
##      filter, lag

## Następujące obiekty zostały zakryte z 'package:base':
##
##      intersect, setdiff, setequal, union

library(readxl)

ruch <- read_excel(
  "Natężenie ruchu - zbiorcze dla Gdańska - dni, miesiące, lata.xlsx",
  sheet = "2019",
  range = "B1:AB366"
)

pogoda2019 <- read_excel(
  "Pogoda2019.xlsx"
)
```

Do analizy wpływu pogody na ruch rowerowy w Gdańsku zostały wykorzystane dane ze stacji pomiarowych w Gdańsku, przedstawiające liczbę przejazdów rowerowych w poszczególne dni w 2019 roku oraz dane pogodowe (<https://open-meteo.com/en/docs/historical-weather-api>) dla Gdańska, dla poszczególnych dni w 2019 roku. Dane przedstawiają informacje takie jak średnia temperatura (°C), suma opadów (mm), godziny opadów (h), czas nasłonecznienia (s), średnia prędkość wiatru (km/h), kod pogody (kod WMO), średnie ciśnienie (hPa).

#zsumowanie ruchu dla całego Gdańska:

```
ruch_total <- ruch %>%
  mutate(ruch_total = rowSums(across(-data), na.rm = TRUE)) %>% #sumuje
wszystko poza kolumną data
dplyr::select(data, ruch_total)
```

```
#łączenie danych ruch w całym mieście + pogoda
```

```
dane_miasto <- ruch_total %>%
```

```
  left_join(pogoda2019, by = "data")
```

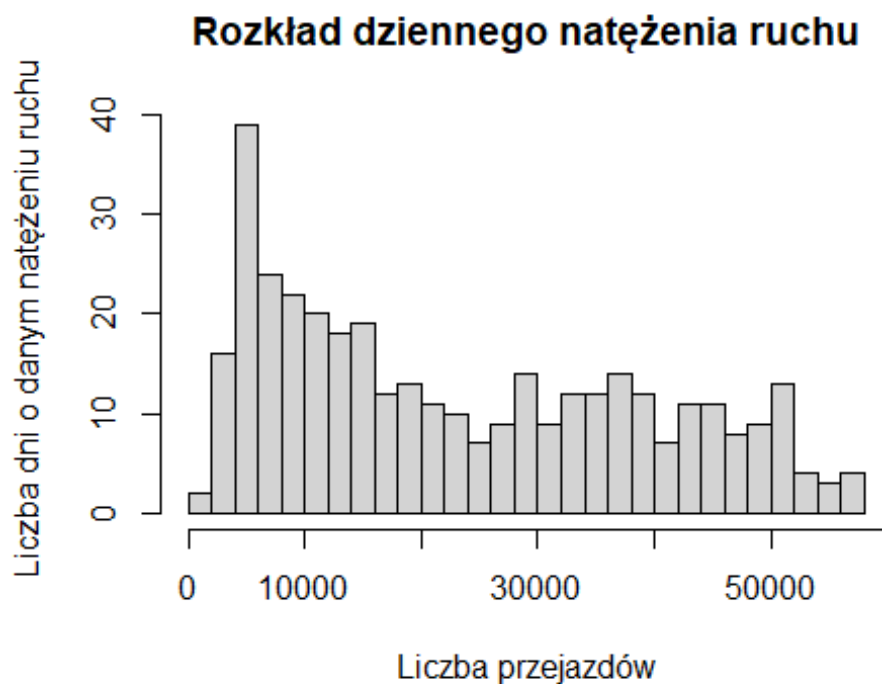
```
hist(dane_miasto$ruch_total,
```

```
  breaks = 30,
```

```
  main = "Rozkład dziennego natężenia ruchu",
```

```
  xlab = "Liczba przejazdów",
```

```
  ylab = "Liczba dni o danym natężeniu ruchu")
```



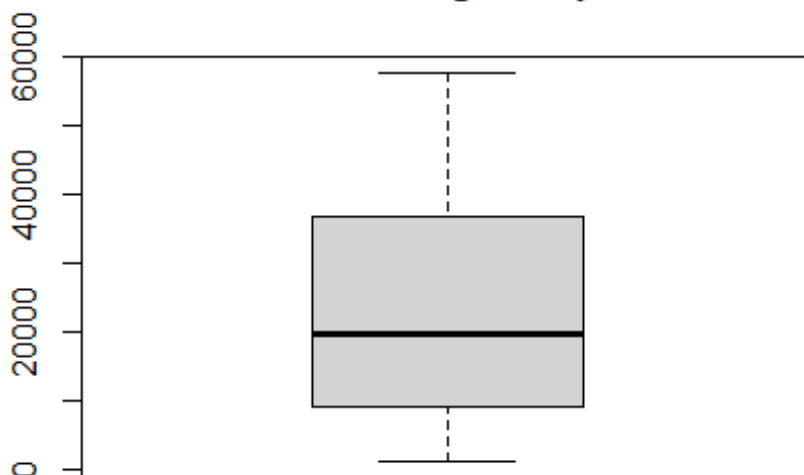
Rozkład dziennego

natężenia ruchu rowerowego w skali miasta charakteryzuje się asymetrią. Większość obserwacji koncentruje się w niższych wartościach liczby przejazdów, natomiast dni o bardzo wysokim natężeniu ruchu występują rzadziej. Szeroki zakres wartości wskazuje na silne zróżnicowanie sezonowe co sugeruje istotny wpływ warunków pogodowych na poziom aktywności rowerowej.

```
boxplot(dane_miasto$ruch_total,
```

```
  main = "Zmienność dziennego natężenia ruchu")
```

Zmienność dziennego natężenia ruchu



```
summary(dane_miasto$ruch_total)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1258	9277	19838	23305	36798	57755

Podstawowe statystyki opisowe wskazują na dużą zmienność dziennego natężenia ruchu rowerowego w skali miasta. Mediana liczby przejazdów wynosi powyżej 19 tys., podczas gdy średnia jest wyższa (23,3 tys.), co potwierdza prawostronną asymetrię rozkładu. Połowa obserwacji mieści się w szerokim przedziale od około 9,3 tys. do 36,8 tys. przejazdów dziennie.

```
plot(x = dane_miasto$ruch_total,  
     y = dane_miasto$`temperature_2m_mean (°C)`,  
     col="blue",  
     main = "Zależność między liczbą przejazdów a średnią temperaturą",  
     xlab = "Liczba przejazdów rowerowych",  
     ylab = "Średnia temperatura [°C]")
```

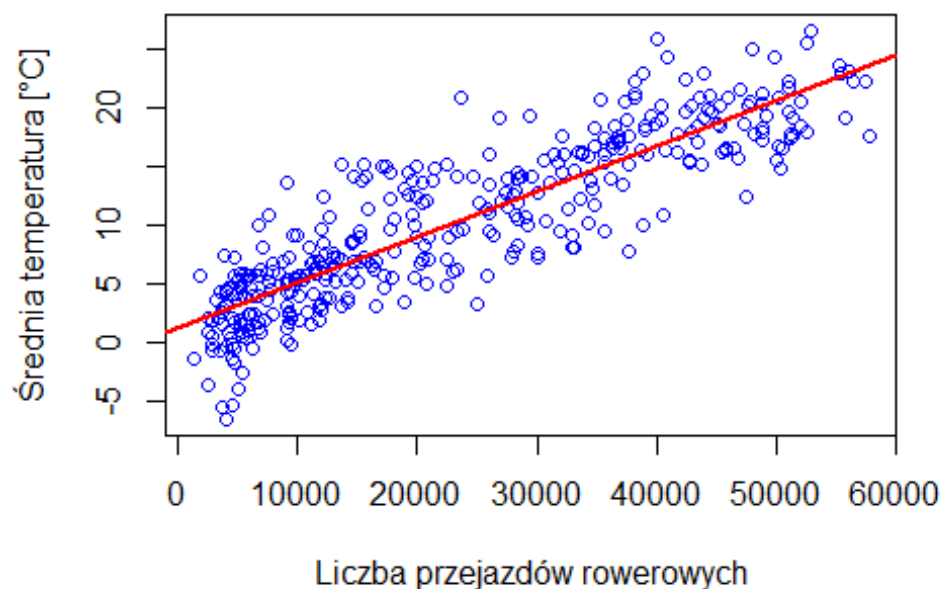
#model liniowy:

```
model_temp <- lm(`temperature_2m_mean (°C)` ~ ruch_total, data = dane_miasto)
```

#linia modelu

```
abline(model_temp, col = "red", lwd = 2)
```

ależność między liczbą przejazdów a średnią temper



Wykres rozrzutu

wraz z dopasowaną linią trendu wskazuje na wyraźną dodatnią zależność pomiędzy średnią temperaturą powietrza a dziennym natężeniem ruchu rowerowego. Wraz ze wzrostem temperatury obserwowany jest wzrost liczby przejazdów, przy czym rozrzut obserwacji zwiększa się dla wyższych wartości temperatury, co sugeruje wpływ dodatkowych czynników.

```
#summary(model_temp)
#wzrost liczby przejazdów o 1000 wiąże się średnio ze wzrostem temperatury o
ok. 0,39°C

cor(dane_miasto$ruch_total, dane_miasto$`temperature_2m_mean (°C)`, use =
"complete.obs")

## [1] 0.8787692
```

Współczynnik korelacji Pearsona pomiędzy średnią temperaturą powietrza a dziennym natężeniem ruchu rowerowego wynosi 0,88, co wskazuje na bardzo silną dodatnią zależność pomiędzy badanymi zmiennymi. Wynik ten jest zgodny z obserwowaną na wykresie rozrzutu rosnącą tendencją liczby przejazdów wraz ze wzrostem temperatury.

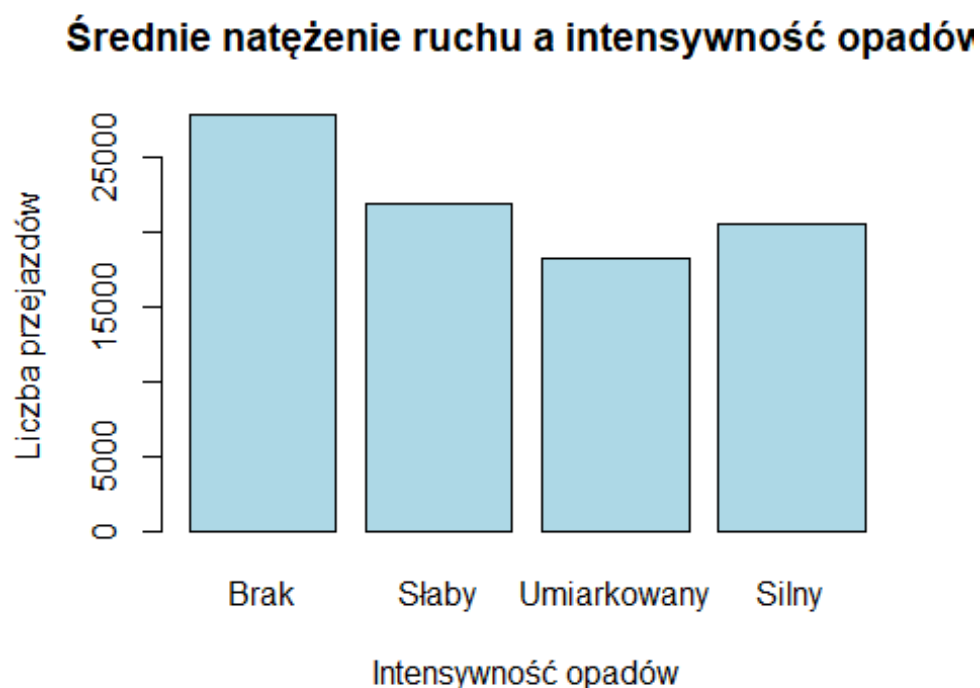
```
#kategorie deszczu:
dane_miasto$opad_cat <- cut(
  dane_miasto$`precipitation_sum (mm)`,
  breaks = c(-0.01, 0, 2, 5, Inf),
  labels = c("Brak", "Słaby", "Umiarkowany", "Silny")
)
table(dane_miasto$opad_cat)
```

```
##
##      Brak      Słaby Umiarkowany      Silny
##      133      134      62      36
```

Przeanalizowano dzienne natężenie ruchu rowerowego w 2019 roku oraz jego związek z występowaniem i intensywnością opadów atmosferycznych.

```
avg_cat <- dane_miasto %>%
  group_by(opad_cat) %>%
  summarise(avg_ruch = mean(ruch_total, na.rm = TRUE))

barplot(
  height = avg_cat$avg_ruch,
  names.arg = avg_cat$opad_cat,
  col = "lightblue",
  main = "Średnie natężenie ruchu a intensywność opadów",
  ylab = "Liczba przejazdów",
  xlab = "Intensywność opadów"
)
```



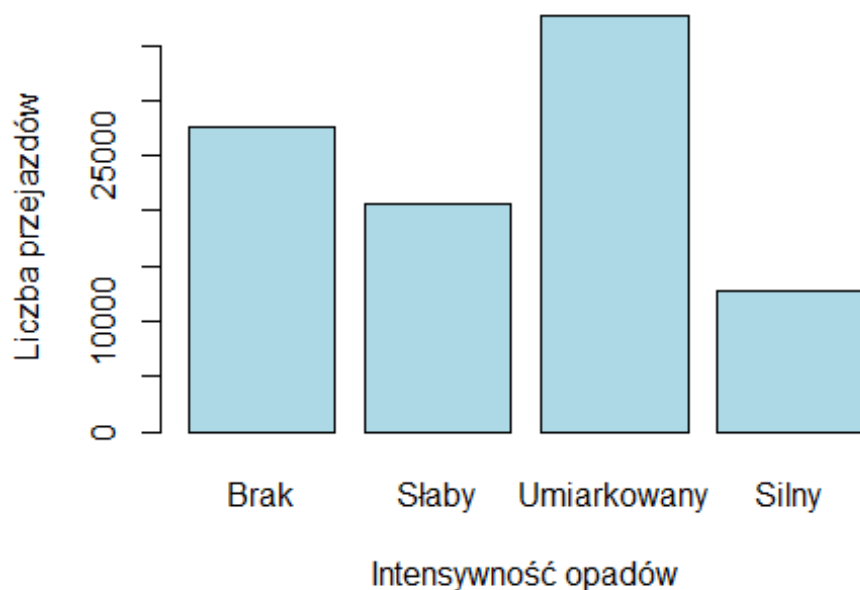
Wniosek: Analiza średniego natężenia ruchu w zależności od intensywności opadów wskazuje, że najwyższy poziom aktywności rowerowej występuje w dniach bez opadów. W dniach z opadami obserwowany jest wyraźnie niższy poziom ruchu, przy czym zależność pomiędzy intensywnością opadów a poziomem ruchu nie jest jednoznaczna.

..... Analiza obejmująca cały rok może w pewnym stopniu zaburzać obraz zależności pomiędzy opadami a natężeniem ruchu rowerowego, ze względu na uśrednianie wyników, pominięcie sezonowości. W celu ograniczenia tego efektu przeprowadzono dodatkową analizę dla krótszego, bardziej jednorodnego okresu - wiosny (miesiące marzec – maj 2019), kiedy warunki pogodowe są szczególnie zmienne.

```
avg_cat_april <- dane_miasto %>%
  filter(format(data, "%Y-%m") == "2019-04") %>%
  group_by(opad_cat) %>%
  summarise(avg_ruch = mean(ruch_total, na.rm = TRUE))
```

```
barplot(
  height = avg_cat_april$avg_ruch,
  names.arg = avg_cat_april$opad_cat,
  col = "lightblue",
  main = "Średnie natężenie ruchu a intensywność opadów w kwietniu 2019",
  ylab = "Liczba przejazdów",
  xlab = "Intensywność opadów"
)
```

Średnie natężenie ruchu a intensywność opadów w kwietniu 2019



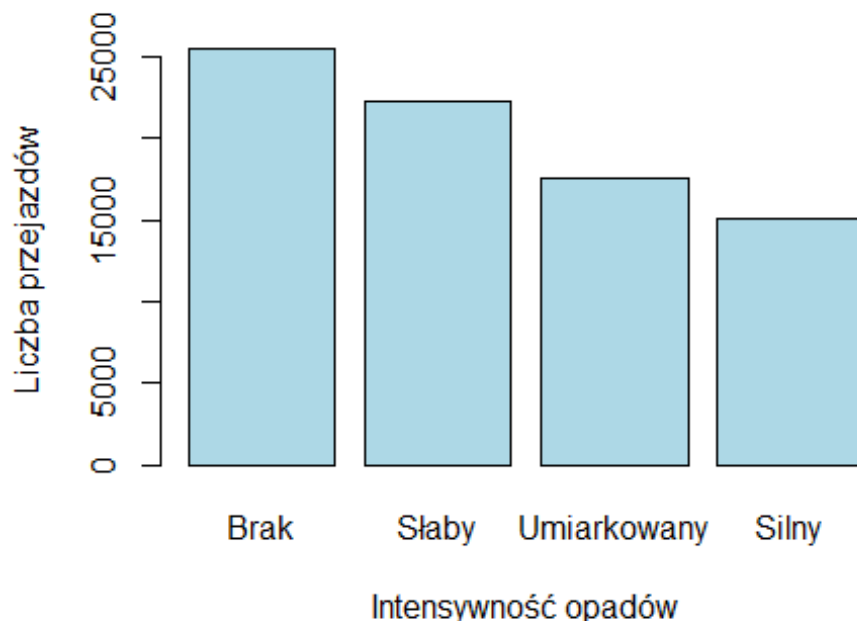
```
#dane_wiosna <- dane_miasto %>%
# filter(month(data) %in% 3:5)

avg_cat_spring <- dane_miasto %>%
  filter(as.integer(format(data, "%m")) %in% 3:5) %>%
```

```
group_by(opad_cat) %>%
  summarise(avg_ruch = mean(ruch_total, na.rm = TRUE))

barplot(
  height = avg_cat_spring$avg_ruch,
  names.arg = avg_cat_spring$opad_cat,
  col = "lightblue",
  main = "Średnie natężenie ruchu a intensywność opadów - wiosna 2019",
  ylab = "Liczba przejazdów",
  xlab = "Intensywność opadów"
)
```

dnie natężenie ruchu a intensywność opadów - wios



Analiza danych z

okresu marzec - maj 2019 wskazuje, że natężenie ruchu rowerowego jest najwyższe w dniach bez opadów atmosferycznych. Wraz ze wzrostem intensywności opadów obserwowany jest stopniowy spadek średniej liczby przejazdów, co potwierdza negatywny wpływ opadów na aktywność rowerową.

#.....PCA..... Dotychczasowa analiza wykazała, że pojedyncze elementy pogody wpływają na natężenie ruchu w różnym stopniu. W kolejnym kroku podjęto próbę ujęcia warunków pogodowych w sposób bardziej kompleksowy. Ponieważ zmienne meteorologiczne są ze sobą powiązane, w dalszej części pracy zastosowano analizę głównych składowych (PCA) w celu redukcji wymiarowości danych oraz identyfikacji głównych czynników opisujących warunki pogodowe.

```
colnames(dane_miasto)#czytamy dane..
```

```
## [1] "data" "ruch_total"
## [3] "temperature_2m_mean (°C)" "precipitation_sum (mm)"
## [5] "precipitation_hours (h)" "sunshine_duration (s)"
## [7] "wind_speed_10m_mean (km/h)" "weather_code (wmo code)"
## [9] "daylight_duration (s)" "wind_gusts_10m_mean (km/h)"
## [11] "dew_point_2m_mean (°C)" "pressure_msl_mean (hPa)"
## [13] "opad_cat"
```

#wybieram dane które nie mówią o tym samym..

```
pogoda_reduced <- dane_miasto %>%
  dplyr::select(
    `temperature_2m_mean (°C)` ,
    `precipitation_sum (mm)` ,
    `wind_speed_10m_mean (km/h)` ,
    `pressure_msl_mean (hPa)` ,
    `sunshine_duration (s)`
  ) %>%
  dplyr::rename(
    temp_srednia = `temperature_2m_mean (°C)` ,
    opady_mm = `precipitation_sum (mm)` ,
    wiatr_kmh = `wind_speed_10m_mean (km/h)` ,
    cisnienie_hpa = `pressure_msl_mean (hPa)` ,
    slonce_s = `sunshine_duration (s)`
  )
```

```
dane1 <- data.frame(pogoda_reduced)
```

```
Kor1 <- cor(dane1)
```

```
#sapply(dane1, class)
```

#macierz korelacji:

```
Kor1 <- cor(dane1[, sapply(dane1, is.numeric)],
            use = "complete.obs")
```

0.0-0.2 | bardzo słaba

0.2-0.4 | słaba

0.4-0.6 | umiarkowana

0.6-0.8 | silna |

0.8-1.0 | bardzo silna

```
Kor1
```

```
##           temp_srednia  opady_mm  wiatr_kmh  cisnienie_hpa
slonce_s
## temp_srednia      1.0000000  0.09116970 -0.28614421      0.0748493
0.6135441
## opady_mm          0.0911697  1.00000000  0.07223893     -0.2851087 -
0.2427722
## wiatr_kmh         -0.2861442  0.07223893  1.00000000     -0.1927567 -
0.2698891
## cisnienie_hpa     0.0748493 -0.28510874 -0.19275674      1.0000000
0.3477093
## slonce_s          0.6135441 -0.24277216 -0.26988910      0.3477093
1.0000000
```


Macierz korelacji wskazuje na umiarkowane zależności pomiędzy zmiennymi meteorologicznymi, bez istotnej współliniowości. Najsilniejszy związek obserwowany jest pomiędzy temperaturą a czasem nasłonecznienia.

#analiza głównych składowych:

```
wyniki1 <- princomp(dane1, cor=TRUE)
wyniki1
```

```
## Call:
## princomp(x = dane1, cor = TRUE)
##
## Standard deviations:
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## 1.4141785 1.1053146 0.9000761 0.8270164 0.5331846
##
## 5 variables and 365 observations.
```

```
summary(wyniki1)
```

```
## Importance of components:
##
##              Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## Standard deviation   1.4141785 1.1053146 0.9000761 0.8270164 0.5331846
## Proportion of Variance 0.3999801 0.2443441 0.1620274 0.1367912 0.05685716
## Cumulative Proportion 0.3999801 0.6443242 0.8063516 0.9431428 1.0000000
```

5 składowych: Proportion of Variance Comp.1 → 0.40 (40%) Comp.2 → 0.24 (24%) Comp.3 → 0.16 (16%) Comp.4 → 0.14 Comp.5 → 0.06

Cumulative Proportion PC1 + PC2 = 64% PC1 + PC2 + PC3 = 81%

Analiza głównych składowych wykazała, że pierwsza składowa wyjaśnia około 40% całkowitej zmienności danych, natomiast dwie pierwsze składowe łącznie około 64%. Uwzględnienie trzech pierwszych składowych pozwala zachować ponad 80% informacji zawartej w pierwotnych zmiennych.

Pierwsza składowa, wyjaśniająca około 40% całkowitej zmienności, jest silnie związana z temperaturą powietrza oraz czasem nasłonecznienia, co pozwala interpretować ją jako wymiar opisujący ciepłe i słoneczne warunki pogodowe.

Druga składowa, wyjaśniająca około 24% zmienności, jest powiązana głównie z opadami atmosferycznymi oraz prędkością wiatru.

Trzecia składowa wnosi dodatkową informację związaną m.in. z ciśnieniem atmosferycznym.

#wartosci własne wg Keizera

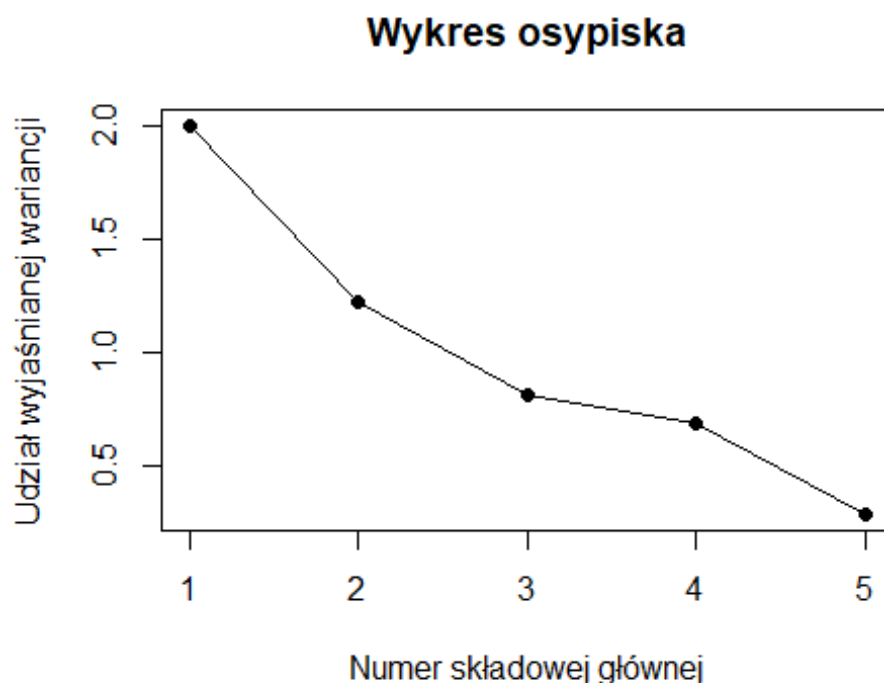
```
wyniki1$sdev^2
```

```
##   Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## 1.9999007 1.2217204 0.8101369 0.6839561 0.2842858
```

```
#Zachowujemy tylko te składowe, których wartość własna > 1
#Dlaczego? bo taka składowa wyjaśnia więcej wariancji niż pojedyncza zmienna
#czyli 1 i 2
```

Na podstawie kryterium Kaisera do dalszej analizy wybrano dwie składowe główne, których wartości własne przekraczały 1, wyjaśniające łącznie około 64% całkowitej wariancji danych.

```
#wykres osypiska
plot(wyniki1$sdev^2,
     pch = 16,
     xlab = "Numer składowej głównej",
     ylab = "Udział wyjaśnianej wariancji",
     main = "Wykres osypiska"
)
lines(wyniki1$sdev^2)
```



Na wykresie osypiska widoczny jest wyraźny spadek udziału wyjaśnianej wariancji po dwóch pierwszych składowych, co wskazuje, że to one najlepiej opisują strukturę danych.

```
#wagi głównych składowych:
wyniki1$loadings

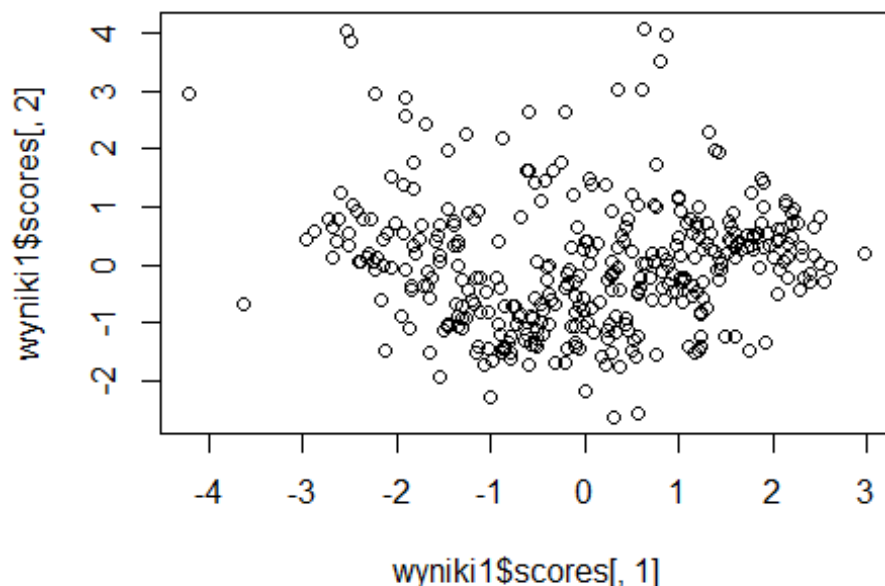
##
## Loadings:
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## temp_srednia  0.497  0.518  0.270      0.638
```

```
## opady_mm      -0.245  0.694 -0.148 -0.612 -0.249
## wiatr_kmh     -0.402 -0.128  0.876 -0.224
## cisnienie_hpa  0.397 -0.477          -0.754  0.192
## slonce_s      0.611          0.359          -0.699
##
##               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings      1.0   1.0   1.0   1.0   1.0
## Proportion Var    0.2   0.2   0.2   0.2   0.2
## Cumulative Var    0.2   0.4   0.6   0.8   1.0
```

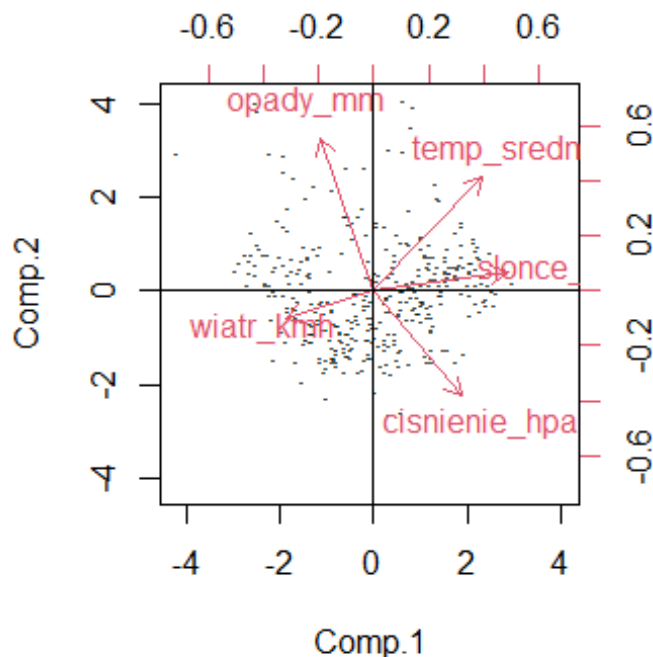
Na podstawie wartości ładunków głównych składowych możliwa była ich interpretacja. Pierwsza składowa główna jest związana z temperaturą oraz czasem nasłonecznienia, przy jednoczesnym ujemnym wpływie prędkości wiatru. Druga składowa jest zdominowana przez opady atmosferyczne oraz ujemną zależność z ciśnieniem. Trzecia składowa jest niemal jednoznacznie związana z prędkością wiatru.

Analiza PCA wykazała, że dwie pierwsze składowe główne wyjaśniają około 64% całkowitej wariancji danych meteorologicznych. Pierwsza składowa związana jest głównie z temperaturą i nasłonecznieniem, natomiast druga opisuje charakter warunków atmosferycznych, takich jak opady i wiatr. Wyniki PCA potwierdziły zasadność wyboru ograniczonego zestawu zmiennych do dalszego modelowania natężenia ruchu rowerowego.

```
#wykres
plot(wyniki1$scores[,1], wyniki1$scores[,2])
```



```
#text(wyniki1$scores[,1], wyniki1$scores[,2], labels=rownames(dane1),pos=2)
biplot(
  wyniki1,
  scale = 0,
  xlabs = rep(".", nrow(wyniki1$scores)) #powtarza "." tyle razy co jest
wierszy
)
abline(h = 0, v = 0)
```



```
#kropki = obserwacje
#im dluzsza strzałka tym większy wkład zmiennej w PCA
#Strzałki= zmienne pogodowe
#Osie:Comp.1 (PC1) - 40% wariacji,,Comp.2 (PC2) - 24% wariacji
#kierunek strzałki - z którą składową zmienna jest związana i czy dodatnio
czy ujemnie
#mały kąt - silna dodatnia korelacja, kąt 90 brak korelacji, 180 - ujemna
```

Dodatkowo wykres głównych składowych ilustruje sposób, w jaki poszczególne dni pogodowe rozmieszczają się względem dwóch dominujących wymiarów zmienności warunków atmosferycznych, a także pokazuje wpływ poszczególnych zmiennych meteorologicznych na ten układ.

Pierwsza składowa główna, wyjaśniająca około 40% całkowitej wariacji, jest silnie związana ze średnią temperaturą powietrza oraz czasem nasłonecznienia.

Druga składowa główna, wyjaśniająca około 24% wariancji, odzwierciedla przede wszystkim charakter warunków atmosferycznych, związany z sumą opadów oraz stabilnością baryczną.

Zauważono, że średnia temperatura powietrza silnie dodatnio koreluje z czasem nasłonecznienia, co potwierdza wcześniejsze obserwacje wynikające z analizy macierzy korelacji. Jednocześnie zmienne opisujące opady oraz prędkość wiatru wykazują ujemne powiązania z temperaturą i nasłonecznieniem.

Rozkład punktów reprezentujących poszczególne dni wskazuje, że większość obserwacji skupia się w pobliżu środka układu współrzędnych, co odpowiada warunkom pogodowym przeciętnym, natomiast dni o skrajnych warunkach atmosferycznych (bardzo ciepłe i słoneczne lub deszczowe i wietrzne) znajdują się na obrzeżach wykresu.

#.....drzewo

#ruch to zmienna ciągła więc drzewo regresyjne

```
library(rpart)
library(rpart.plot)
```

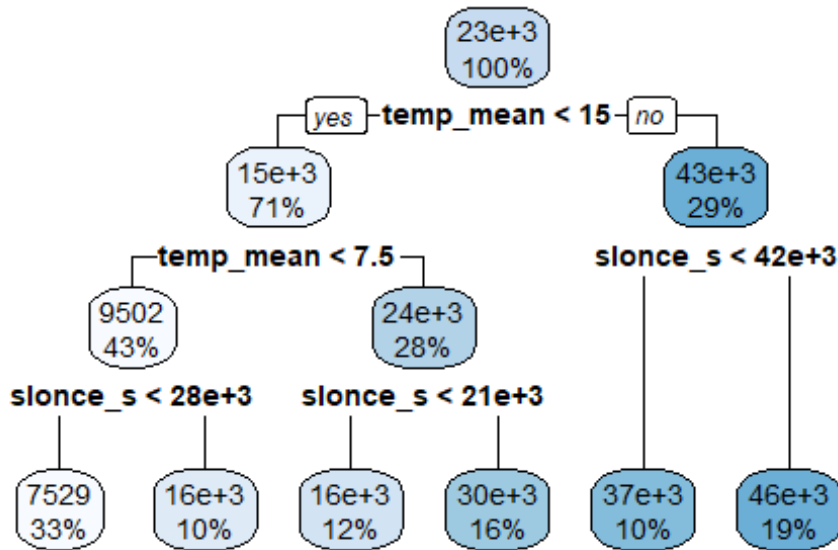
#zmiana nazw zmiennych:

```
dane_miasto$temp_mean <- dane_miasto$`temperature_2m_mean` (°C)`
dane_miasto$slonce_s <- dane_miasto$`sunshine_duration` (s)`
dane_miasto$opady_mm <- dane_miasto$`precipitation_sum` (mm)`
dane_miasto$wiatr_kmh <- dane_miasto$`wind_speed_10m_mean` (km/h)`
dane_miasto$cisnienie_hpa <- dane_miasto$`pressure_msl_mean` (hPa)`
```

```
drzewo_reg <- rpart(
  formula = ruch_total ~ temp_mean + slonce_s + opady_mm + wiatr_kmh +
  cisnienie_hpa,
  data = dane_miasto
)
```

```
rpart.plot(drzewo_reg,
  main = "Drzewo regresyjne: natężenie ruchu a pogoda")
```

Drzewo regresyjne: natężenie ruchu a pogoda



Drzewo regresyjne

wskazuje, że kluczowym czynnikiem różnicującym natężenie ruchu rowerowego jest temperatura powietrza. Pierwszy podział następuje przy progu około 15°C, powyżej którego obserwowany jest wyraźnie wyższy poziom ruchu. W gałęzi chłodniejszych dni dodatkowe znaczenie ma czas nasłonecznienia oraz próg około 7.5°C, który rozdziela dni o niskiej i umiarkowanej aktywności. Dla dni ciepłych ($\geq 15^\circ\text{C}$) nasłonecznienie nadal pozostaje istotnym czynnikiem – większa liczba godzin słońca wiąże się z wyższym przewidywanym natężeniem ruchu.

GLM?.....