

# Wind speed

Denisa Mensatorisova

## WIND-OBSERVATION speed rate

The rate of horizontal travel of air past a fixed point.

MIN: 0000 MAX: 0900 UNITS: meters per second

SCALING FACTOR: 10

DOM: A general domain comprised of the numeric characters (0-9).

9999 = Missing.

Atribút wind speed reprezentuje rýchlosť vetra v metroch za sekundu.

```
all_data <- read.csv(file= "../data/all.csv")
```

```
summary(all_data$WND_SPEED)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      0.000  0.000   1.000   1.716   2.600  26.000   18057
```

```
all_data['WND_SPEED'] %>% profiling_num()
```

```
##      variable      mean std_dev variation_coef p_01 p_05 p_25 p_50 p_75 p_95 p_99
## 1 WND_SPEED 1.715614 1.84761      1.076938    0    0    0    1  2.6  5.1    8
##      skewness kurtosis iqr range_98 range_80
## 1 1.576615 6.190994 2.6    [0, 8] [0, 4.1]
```

```
getmode(na.omit(all_data$WND_SPEED)) # modus
```

```
## [1] 0
```

## Centrálna poloha dát

Hodnota výberového mediánu je 1 a hodnota výberového priemeru 1.716. Keďže stredná hodnota (medián) je nižšia ako priemer, hodnotu priemeru mohli ovplyvniť výrazne vyššie hodnoty, prípadne extrémne vysoké hodnoty. Takže stredná hodnota rýchlosti vetra je 1 m/s.

Modus - najčastejšia hodnota je 0 m/s.

```
var(all_data$WND_SPEED, na.rm = T) # rozptyl
```

```
## [1] 3.413663
```

```
max(all_data$WND_SPEED, na.rm = T) - min(all_data$WND_SPEED, na.rm = T) # variacne rozpatie
```

```
## [1] 26
```

```
# Interquartile range and outliers
```

```
Q1 <- quantile(all_data$WND_SPEED, 0.25, na.rm = T) # 25% hodnot je mensich a 75% vacsich
```

```
Q3 <- quantile(all_data$WND_SPEED, 0.75, na.rm = T) # 75% hodnot je mensich a 25% vacsich
```

```
IQR <- IQR(all_data$WND_SPEED, na.rm = T) # interquartile range
```

```

IQR_dev <- IQR/2

# odlahle hodnoty
length(which(all_data$WND_SPEED < (Q1 - 1.5*IQR)))

## [1] 0

length(which(all_data$WND_SPEED > (Q3 + 1.5*IQR)))

## [1] 10253

# extreme hodnoty
length(which(all_data$WND_SPEED < (Q1 - 3*IQR)))

## [1] 0

length(which(all_data$WND_SPEED > (Q3 + 3*IQR)))

## [1] 561

```

## Variabilita

Výberový rozptyl je 3.413663.

Výberová smerodajná odchýlka je 1.8476. To znamená, že rýchlosť vetra sa pohybuje približne v rozsahu 1.8476 m/s okolo priemeru.

Medzikvartilová odchýlka ( $IQR/2$ ) je 1.3, teda hodnoty sú rozptýlené približne 1.3 okolo mediánu.

Variačné rozpätie je 26. Daná hodnota predstavuje rozdiel medzi maximálnou a minimálnou nameranou hodnotou rýchlosti vetra. Maximálna rýchlosť vetra je 26 m/s a minimálna 0 m/s.

Variačný koeficient rýchlosti vetra je 1.076938, čo je 108%. V porovnaní s ostatnými atribútmi má rýchlosť vetra vyššiu variabilitu, podobnú ako teplota.

## Asymetria

Hodnota šikmosti je kladná 1.58, to znamená, že väčšina hodnôt sa nachádza pod priemerom. Dáta sú zošimené a nejde o symetrické rozdelenie okolo strednej hodnoty.

Hodnota špicatosti je 6.2, je väčšia ako 3 teda hodnoty majú veľmi špicaté rozdelenie. To znamená, že početnosti hodnôt v súbore sú prevažne sústredené okolo strednej hodnoty.

## Boxplot

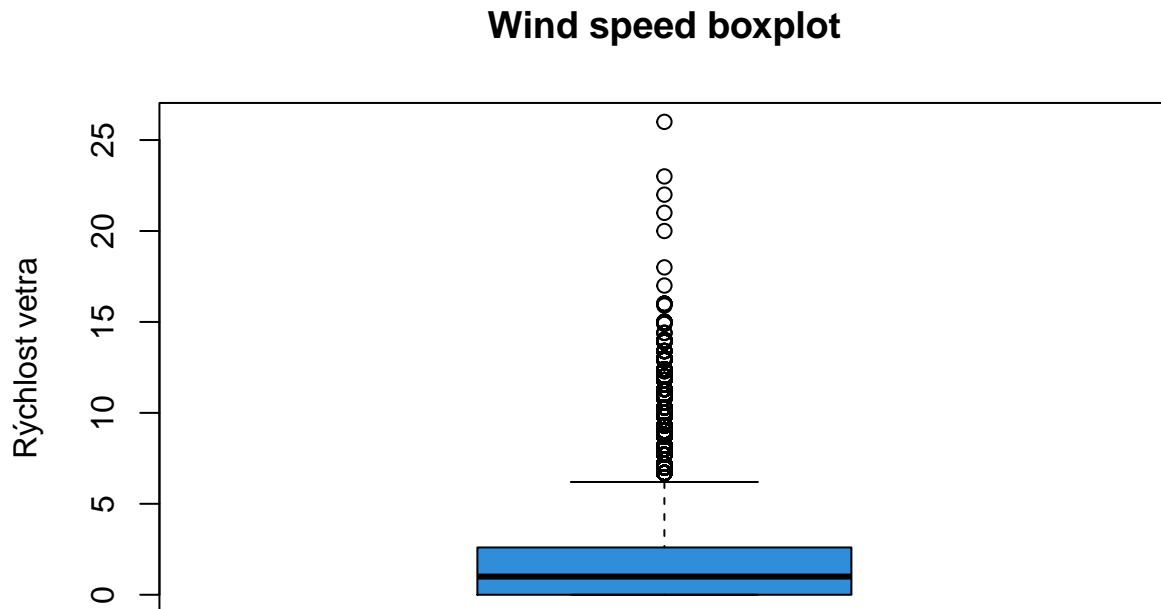
Hodnota 3.kvartilu je 2.6 a hodnota 1. kvartilu 0, teda medzikvartilové rozpätie ( $IQR$ ) je 2.6. Uprostred krabice je zvýraznený medián hrubou čiernou čiarou (1). Vidíme, že krabica boxplotu sa nachádza v nižších hodnotách, teda 50% hodnôt tvoria nízke hodnoty, avšak je tu aj niekoľko vysokých hodnôt.

Ďalej z boxplotu vidieť maximálnu a minimálnu hodnotu (vonkajšie hradby boxplotu). Maximálna hodnota (6.5) je vypočítaná ako 3.kvartil +  $1.5 * IQR$  (medzikvartilové rozpätie). Minimálna hodnota je vypočítaná ako 1.kvartil -  $1.5 * IQR$  (medzikvartilové rozpätie), v prípade rýchlosti je minimálna hodnota 0.

Všetky hodnoty nachádzajúce sa nad a pod maximálnou a minimálnou hodnotou môžeme považovať za odlahlé hodnoty. Počet odlahlých hodnôt nad maximálnou hodnotou je 10253, pod minimálnou sa nenachádzajú žiadne odlahlé hodnoty.

Nakoniec pre odlahlé hodnoty overíme či patria medzi extrémne. Horná hranica extrémnych hodnôt je vypočítaná ako 3.kvartil +  $3 * IQR$ . Dolná hranica extrémnych hodnôt je vypočítaná ako 1.kvartil -  $3 * IQR$ . V dátach sa nachádza 561 extrémne vysokých hodnôt, ktoré sú vyššie ako hodnota 3.kvartilu +  $3 * IQR$ , teda vyššie ako 10.4 m/s.

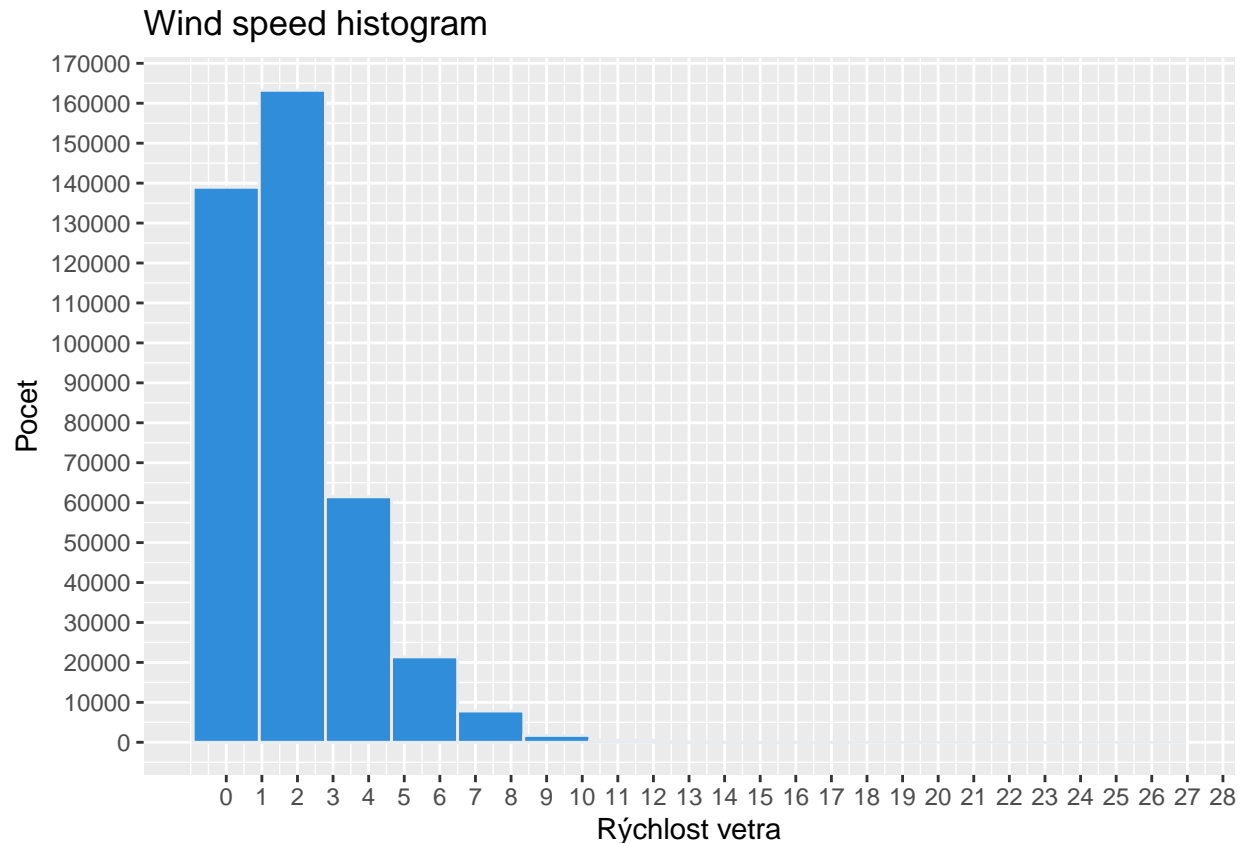
```
boxplot(all_data$WND_SPEED, col = "#308dd9", ylab = "Rýchlosť vetra", main = "Wind speed boxplot")
```



## Histogram

Už z histogramu vidíme, že rozdelenie hodnôt je výrazne vyšikmené doľava. Najpočetnejšie sú hodnoty okolo 0 - 1. S narastajúcou rýchlosťou vetra postupne početnosť hodnôt klesá.

```
ggplot(all_data, aes(x = WND_SPEED)) +
  geom_histogram(bins = 15, fill = "#308dd9", color = "#e9ecef") +
  labs(title = paste("Wind speed histogram")) +
  xlab("Rýchlosť vetra") +
  ylab("Počet") +
  scale_x_continuous(breaks = seq(0, 30, by = 1)) +
  scale_y_continuous(breaks = seq(0, 200000, by = 10000))
```



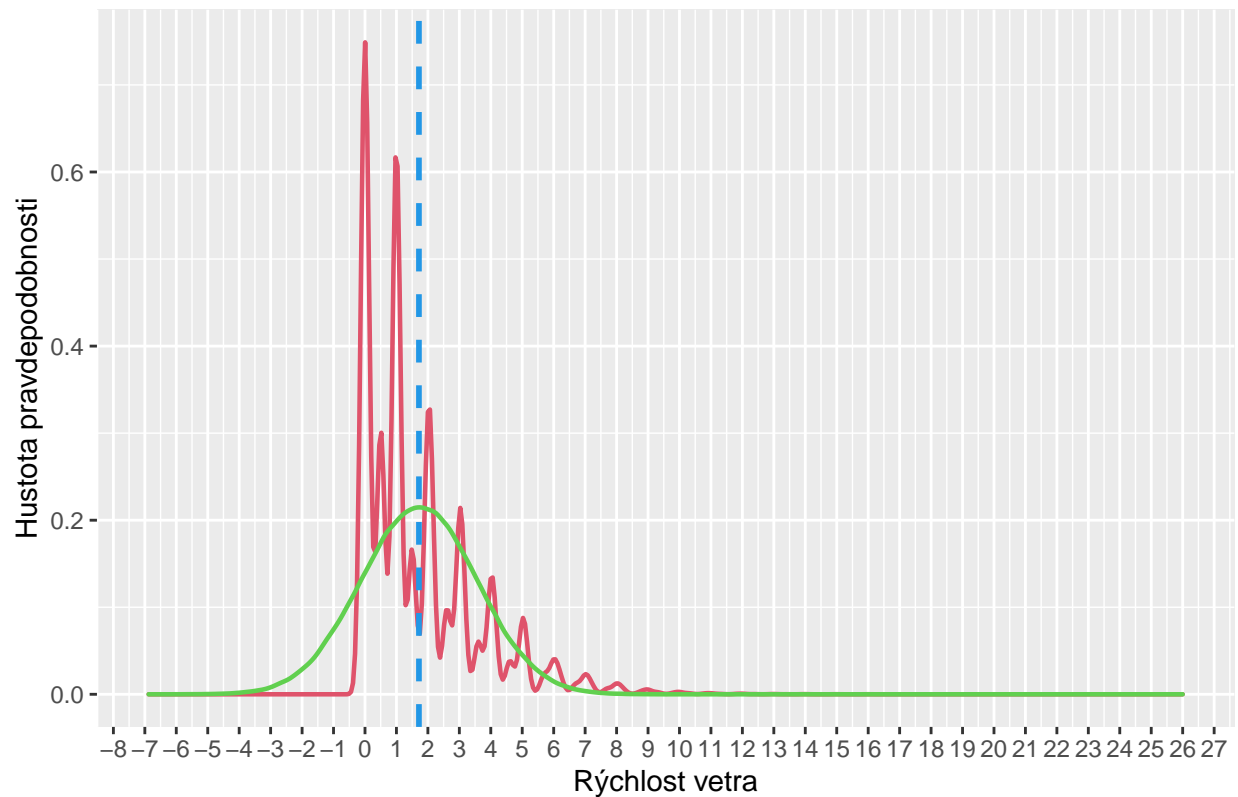
### Graf hustoty

Graf hustoty slúži na porovnanie priebehu hustoty pravdepodobnosti normálneho rozdelenia (zelená čiara) a odhadu hustoty vypočítaného z namerných hodnôt rýchlosti vetra (červená čiara). Čiary sa výrazne líšia, teda nejde o normálne rozdelenie. Modrá prerušovaná čiara predstavuje priemernú rýchlosť vetra. V dátach sa nachádza niekoľko vrcholov, najvyšší je v bode 0.

```
# density plot
# data z normalneho rozdelenia
data_norm <- data.frame(dens = c(rnorm(length(na.omit(all_data$WND_SPEED)), mean(all_data$WND_SPEED, na.rm = T), sd = 10))),

# porovnanie hodnot normalneho rozdelenia a WND_SPEED
ggplot(all_data, aes(x = WND_SPEED), color = 3) +
  geom_density(color = 2, size = 0.8) +
  geom_density(data_norm, mapping = aes(x = dens), color = 3, size = 0.8) +
  geom_vline(aes(xintercept = mean(WND_SPEED, na.rm = T)),
             color = 4, linetype = "dashed", size = 1) +
  scale_x_continuous(breaks = seq(-20, 50, by = 1)) +
  labs(title = paste("Odhad hustoty rýchlosti vetra")) +
  xlab("Rýchlosť vetra") +
  ylab("Hustota pravdepodobnosti")
```

## Odhad hustoty rýchlosti vetra



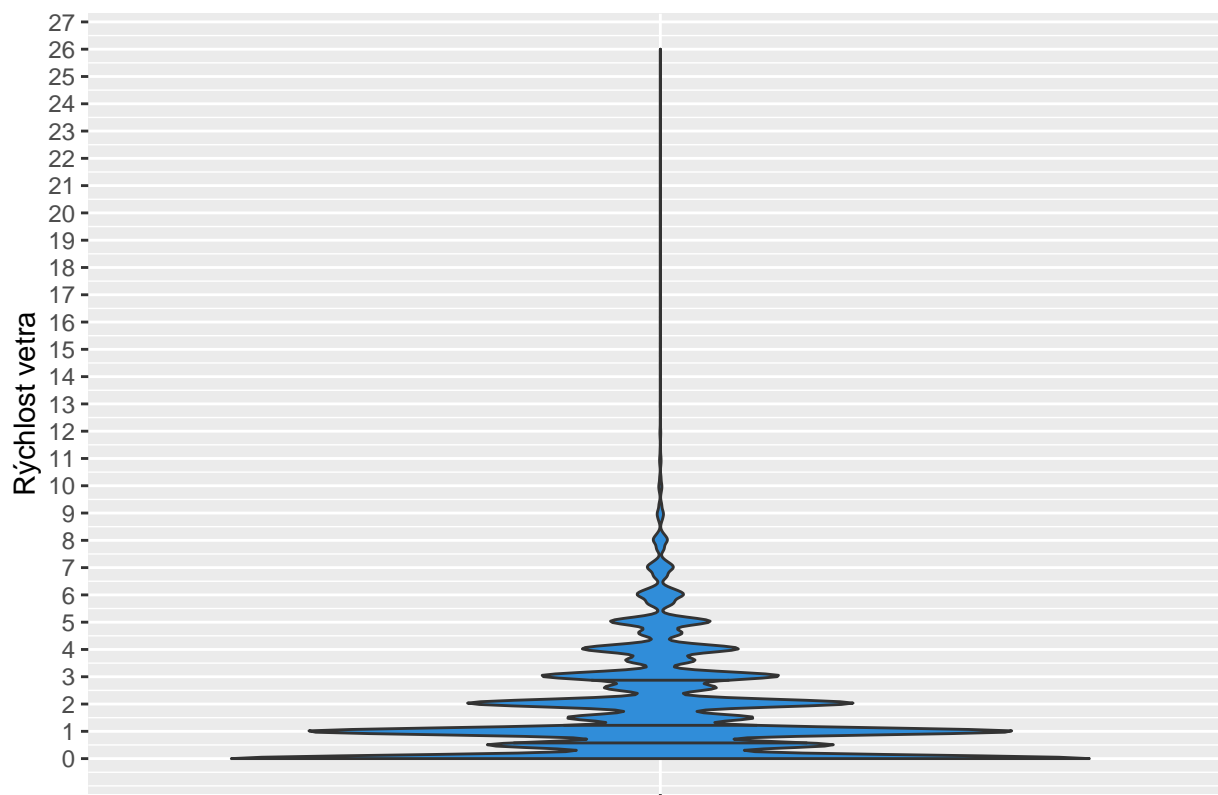
## Husľový graf

Husľový graf doplnený o hlavné kvantily zobrazuje rozdelenie hustoty, pričom aj podľa tohto grafu vidíme, že nejde o normálne rozdelenie. Dáta sú striedavo rozdelené, najviac v intervale od 0 po 8 ale vyskytujú sa v nich aj vysoké hodnoty (dlhá čiara).

```
df <- all_data %>%
  dplyr::select('WIND_SPEED') %>%
  tidyr::gather(key = 'label', value = 'speed')

ggplot(data = df, aes(factor(label), speed, fill = speed)) +
  geom_violin(draw_quantiles = c(0.25, 0.5, 0.75), fill = "#308dd9") +
  labs(title = paste("Husľový graf rýchlosti vetra"), y = "Rýchlosť vetra", fill = "speed") +
  theme(axis.title.x = element_blank()) +
  theme(axis.text.x = element_blank()) +
  scale_y_continuous(breaks = seq(0, 30, by = 1))
```

Huslový graf rýchlosti vetra

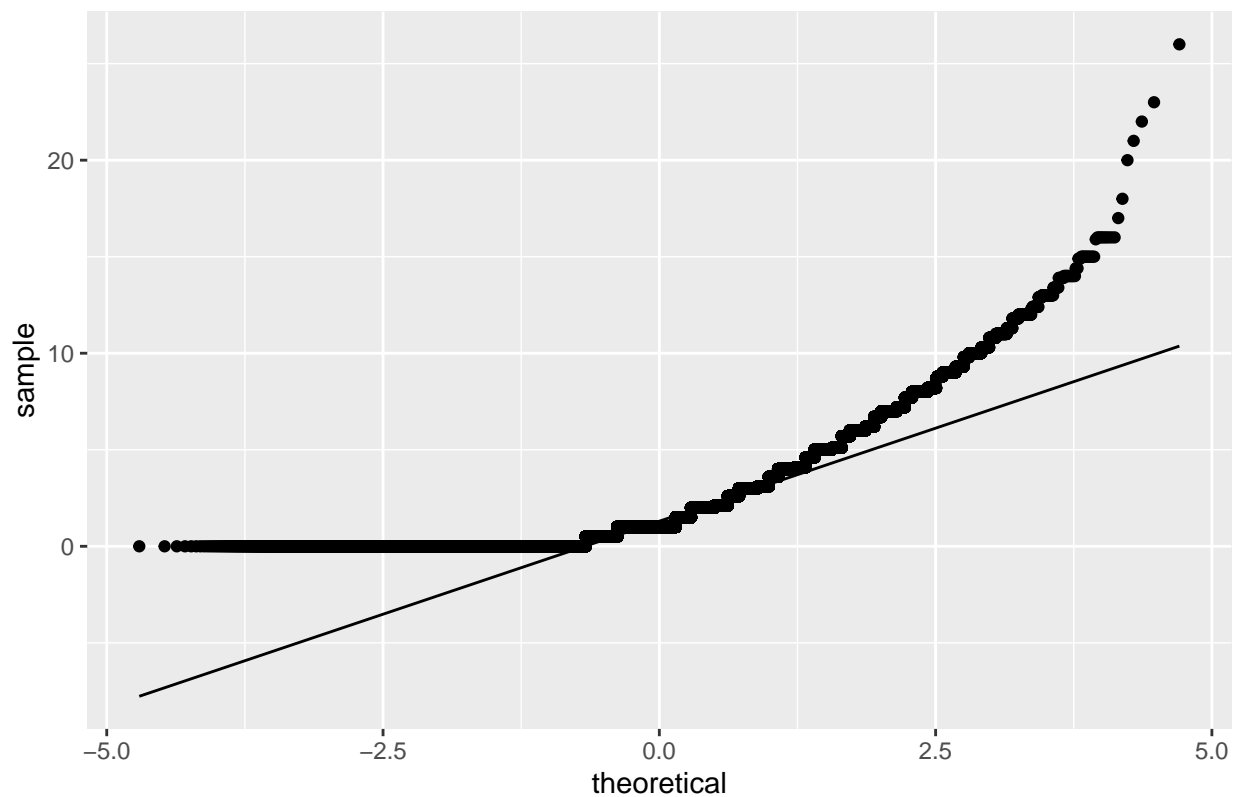


### Q-Q graf

Graf zobrazuje odchýlku empirického od teoretického normálneho rozdelenia. Empirické rozdelenie je v našom prípade rozdelenie nameraných hodnôt rýchlosti vetra. Keďže body sa výrazne odchyľujú od priamky normálneho rozdelenia, môžeme povedať, že rozdelenie hodnôt rýchlosti vetra nie je normálne.

```
ggplot(data = all_data, aes(sample = WND_SPEED)) +  
  stat_qq() +  
  stat_qq_line() +  
  labs(title = paste("Q-Q graf rýchlosti vetra"))
```

## Q–Q graf rýchlosti vetra



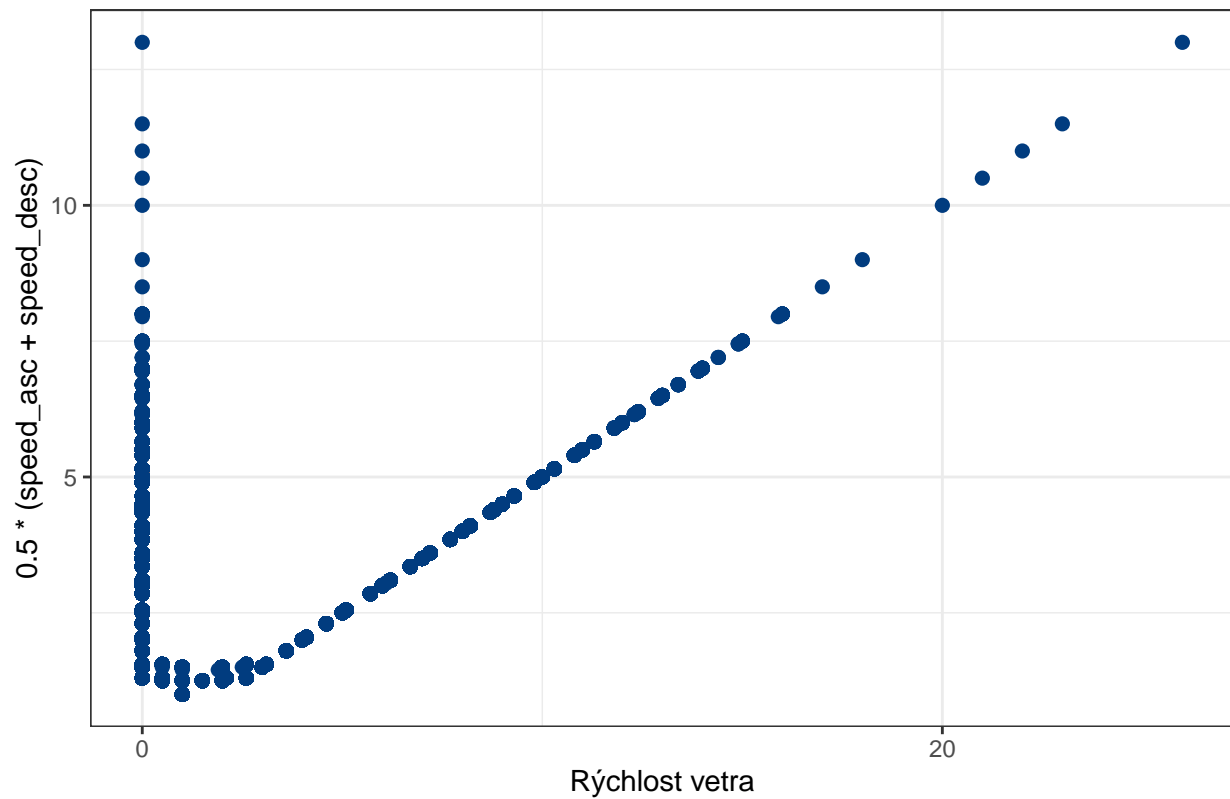
## Graf polosum

Z grafu vidíme, že hodnoty nie sú symetrické.

```
speed <- all_data$WIND_SPEED
speed_asc <- sort(speed, decreasing = FALSE)
speed_desc <- sort(speed, decreasing = TRUE)

ggplot(data.frame(speed_asc), aes(x = speed_asc, y = 0.5*(speed_asc + speed_desc))) +
  geom_point(size = 2, color = "#013c7f") +
  scale_x_continuous(breaks = seq(0, 400, by = 20)) +
  labs(title = "Graf polosum pre rýchlosť vetra", x = "Rýchlosť vetra") +
  theme_bw()
```

Graf polosum pre rýchlosť vetra



Časový graf rýchlosti vetra

```
all_data %>%
  dplyr::mutate(
    date = as_date(DATE)
  ) %>%
  dplyr::distinct(date, .keep_all=TRUE) %>%
  dplyr::select(date, WND_SPEED) %>%
  as_tsibble(
    index = date
  ) %>%
  autoplot(WND_SPEED) +
  labs(title = "Time graph of wind speed",
    y = "WND_SPEED", x = "Date")
```



Time graph of wind speed

