

Wind angle

Denisa Mensatorisova

WIND-OBSERVATION direction angle

The angle, measured in a clockwise direction, between true north and the direction from which the wind is blowing.

MIN: 001 MAX: 360 UNITS: Angular Degrees

SCALING FACTOR: 1

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

999 = Missing. If type code (below) = V, then 999 indicates variable wind direction.

Atribút wind angle reprezentuje uhol meraný v smere hodinových ručičiek medzi severom a smerom, z ktorého fúka vietor. Uhol je meraný v stupňoch teda ide o diskretný atribút, ktorý nadobúda hodnoty od 0 do 360°, meraný v desiatkach stupňov.

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

```
summary(all_data$WND_ANGLE)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      10.0    70.0   210.0   191.2   300.0   360.0   205906
```

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

```
##      variable      mean std_dev variation_coef p_01 p_05 p_25 p_50 p_75 p_95 p_99
## 1 WND_ANGLE 191.1615 119.3355      0.6242657   10   10   70  210  300  360  360
##      skewness kurtosis iqr  range_98  range_80
## 1 -0.1272726  1.65222 230 [10, 360] [20, 350]
```

```
freq_table <- table(all_data$WND_ANGLE)
```

```
angle_frequency <- as.data.frame(freq_table) %>% rename('Angle' = Var1)
```

```
angle_frequency[order(angle_frequency$Freq, decreasing = T),]
```

```
##      Angle  Freq
## 36      360 18095
## 1         10 13081
## 2         20 12359
## 35        350  8474
## 3          30  8008
## 34        340  7897
## 4          40  7875
## 22        220  7862
## 20        200  7451
## 21        210  6470
## 23        230  6133
## 24        240  5911
## 18        180  5758
```

```
## 33    330  5744
## 25    250  5663
## 27    270  5335
## 19    190  5285
## 5      50  5143
## 32    320  4920
## 26    260  4852
## 31    310  4672
## 6      60  4202
## 29    290  4195
## 13    130  4159
## 30    300  3954
## 12    120  3913
## 28    280  3836
## 11    110  3401
## 7      70  3383
## 9      90  2984
## 14    140  2934
## 16    160  2891
## 8      80  2806
## 17    170  2769
## 10    100  2532
## 15    150  2483
```

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

```
## [1] 360
```

Centrálna poloha dát

Hodnota výberového mediánu je 210 a hodnota výberového priemeru 191.2. Keďže stredná hodnota (medián) je vyššia ako priemer, hodnotu priemeru mohli ovplyvniť nejaké výrazne nižšie hodnoty. Teda stredná hodnota uhlu vetra je 210° a polovica hodnôt dosahuje hodnoty vyššie od mediánu a polovica hodnôt nižšie.

Modus, teda najčastejšia hodnota je 360. O niečo menej početné sú hodnoty 10 a 20, teda najčastejšie fúkal severný vietor. Naopak najmenej početné sú hodnoty 150, 170 a 100° teda južný vietor sa vyskytoval najmenej.

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

```
## [1] 14240.97
```

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

```
## [1] 350
```

```
# Interquartile range and outliers
```

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

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

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

```
IQR_dev <- IQR/2
```

```
# odlahle hodnoty
```

```
length(which(all_data$WND_ANGLE < (Q1 - 1.5*IQR)))
```

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

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

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

Variabilita

Výberový rozptyl je 14240.97, čo vzhľadom k tomu, že samotné hodnoty uhlu vetra sa pohybujú od 0 až do 360 nie je veľa.

Výberová smerodajná odchýlka je 119.3355. Teda hodnoty sú rozptýlené približne 119 stupňov okolo priemeru. Medzikvartilová odchýlka (IQR/2) je 115, teda hodnoty sú rozptýlené približne 115 stupňov okolo mediánu.

Variačné rozpätie je 350. Daná hodnota predstavuje rozdiel medzi maximálnou a minimálnou nameranou hodnotou uhlu vetra. Maximálny uhol je 360° a minimálny 10°.

Variačný koeficient uhlu vetra je 0.62, čo je 62%. V porovnaní s teplotou je približne o 40% menší. Teda uhol vetra má nižšiu variabilitu ako teplota.

Asymetria

Hodnota šikmosti je záporná ale blízka 0 (-0.1272726) teda hodnoty nie sú veľmi zošikmené a ide skôr o symetrické rozdelenie.

Hodnota špicatosti je 1.65222, je menšia ako 3 teda hodnoty majú ploché rozdelenie. To znamená, že v súbore sa nachádza viac hodnôt vzdialenejších od strednej hodnoty.

Boxplot

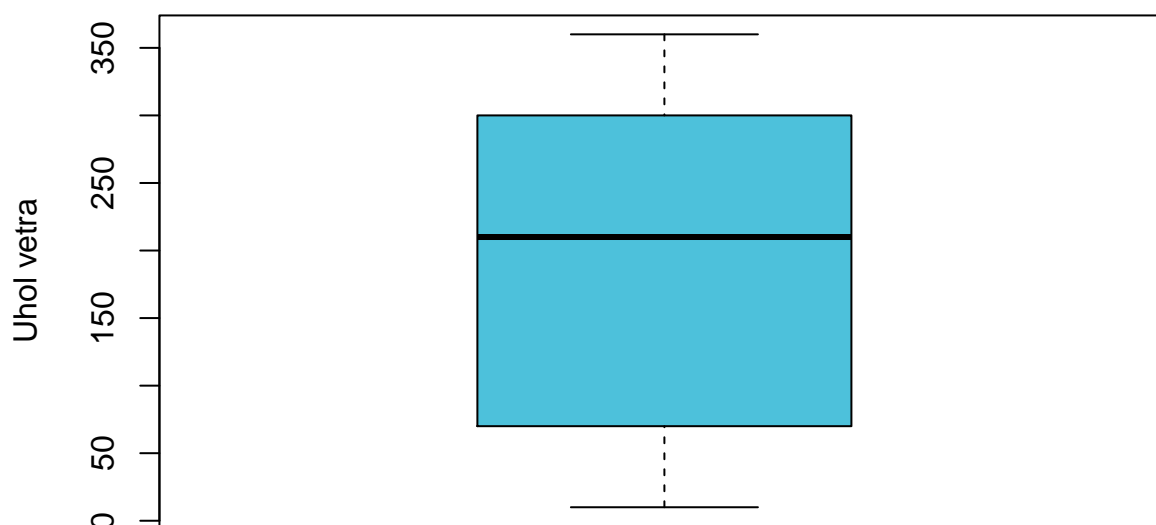
Krabica zobrazuje rozsah v ktorom sa nachádza 50% hodnôt - medzikvartilové rozpätie, čo je hodnota 3.kvartilu - hodnota 1. kvartilu. Hodnota 3.kvartilu je 300 a hodnota 1. kvartilu 70, teda medzikvartilové rozpätie (IQR) je $300 - 70 = 230$. Uprostred krabice je zvýraznený medián hrubou čiernou čiarou (210). Keďže sa nachádza nad polovicou krabice, podľa boxplotu sa zdá, že dáta sú zošikmené a teda nie sú rovnomerne rozptýlené okolo strednej hodnoty.

Ďalej z boxplotu vidieť maximálnu a minimálnu hodnotu (vonkajšie hradby boxplotu). Maximálna hodnota (360°) je vypočítaná ako $3.kvartil + 1.5 * IQR$ (medzikvartilové rozpätie). Minimálna hodnota (10°) je vypočítaná ako $1.kvartil - 1.5 * IQR$ (medzikvartilové rozpätie). V oboch prípadoch by vypočítané hodnoty boli vyššie, resp. nižšie ako rozsah stupňov (0-360) preto sú vonkajšie hradby boxplotu dané celkovou minimálnou a maximálnou hodnotou.

Všetky hodnoty nachádzajúce sa nad a pod maximálnou a minimálnou hodnotou môžeme považovať za odľahlé hodnoty. Takéto hodnoty sa však v uhle vetra nevyskytujú, preto ďalej neoverujeme výskyt extrémnych hodnôt.

```
boxplot(all_data$WND_ANGLE, col = "#4dc1db", ylab = "Uhol vetra", main = "Wind angle boxplot")
```

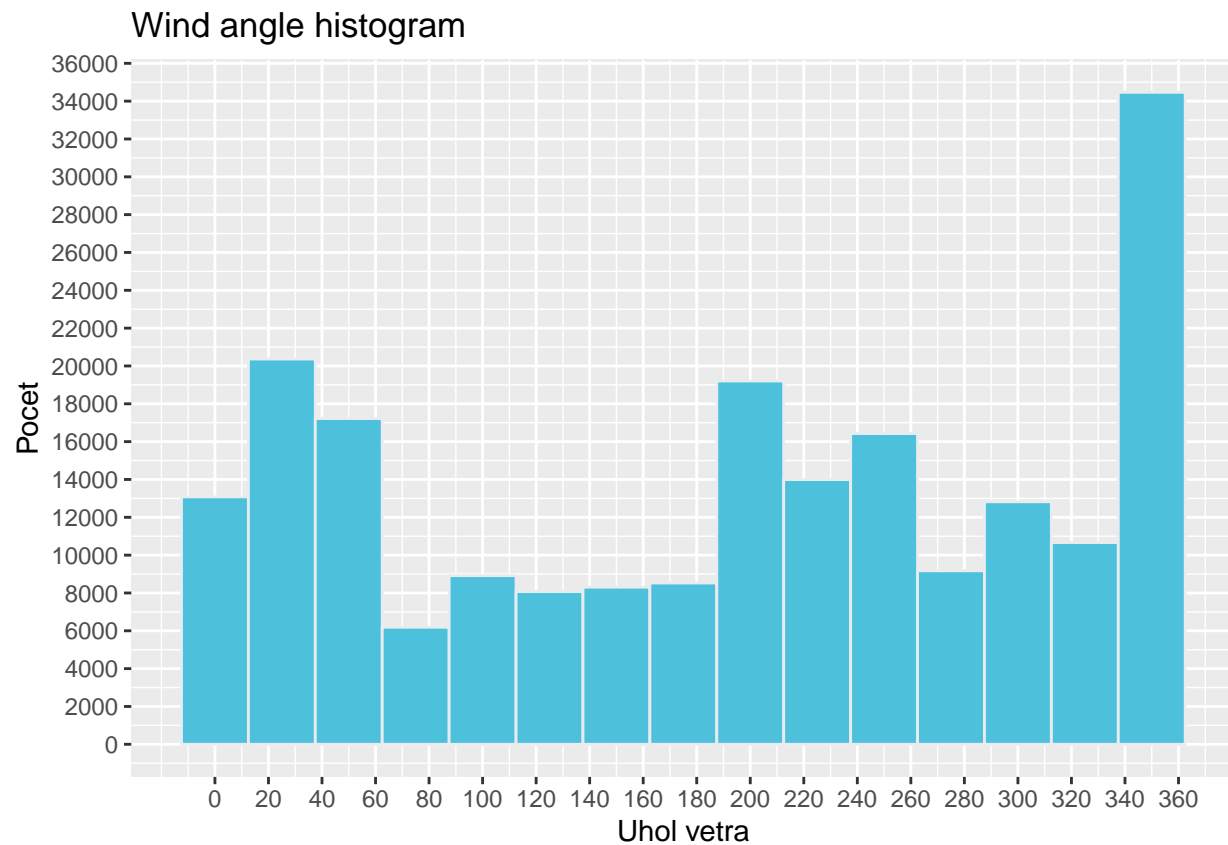
Wind angle boxplot



Histogram

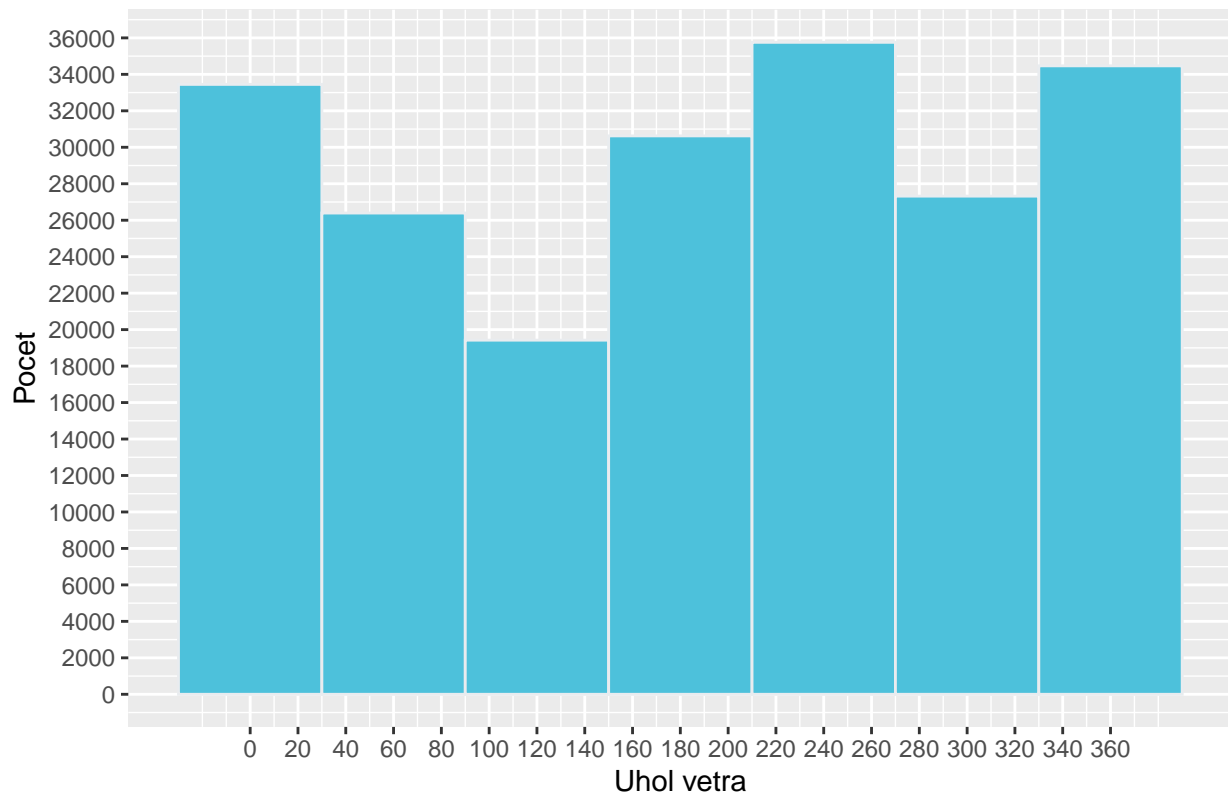
Už z histogramu vidíme, že rozdelenie hodnôt nie je normálne. Avšak problémovými bodmi môžu byť hraničné hodnoty 0 a 360 prípadne hodnoty blízke hraničným bodom, pretože uhol 360° je v podstate rovnaký ako uhol 0° a v smere vetra sa tento rozdiel neprejaví podobne ako aj hodnoty 0° a 10° alebo 360° a 350°.

```
ggplot(all_data, aes(x = WND_ANGLE)) +  
  geom_histogram(bins = 15, fill = "#4dc1db", color = "#e9ecef") +  
  labs(title = paste("Wind angle histogram")) +  
  xlab("Uhol vetra") +  
  ylab("Počet") +  
  scale_x_continuous(breaks = seq(0, 360, by = 20)) +  
  scale_y_continuous(breaks = seq(0, 50000, by = 2000))
```



```
ggplot(all_data, aes(x = WND_ANGLE)) +  
  geom_histogram(binwidth = 60, fill = "#4dc1db", color = "#e9ecef") +  
  labs(title = paste("Wind angle histogram")) +  
  xlab("Uhol vetra") +  
  ylab("Počet") +  
  scale_x_continuous(breaks = seq(0, 360, by = 20)) +  
  scale_y_continuous(breaks = seq(0, 50000, by = 2000))
```

Wind angle histogram



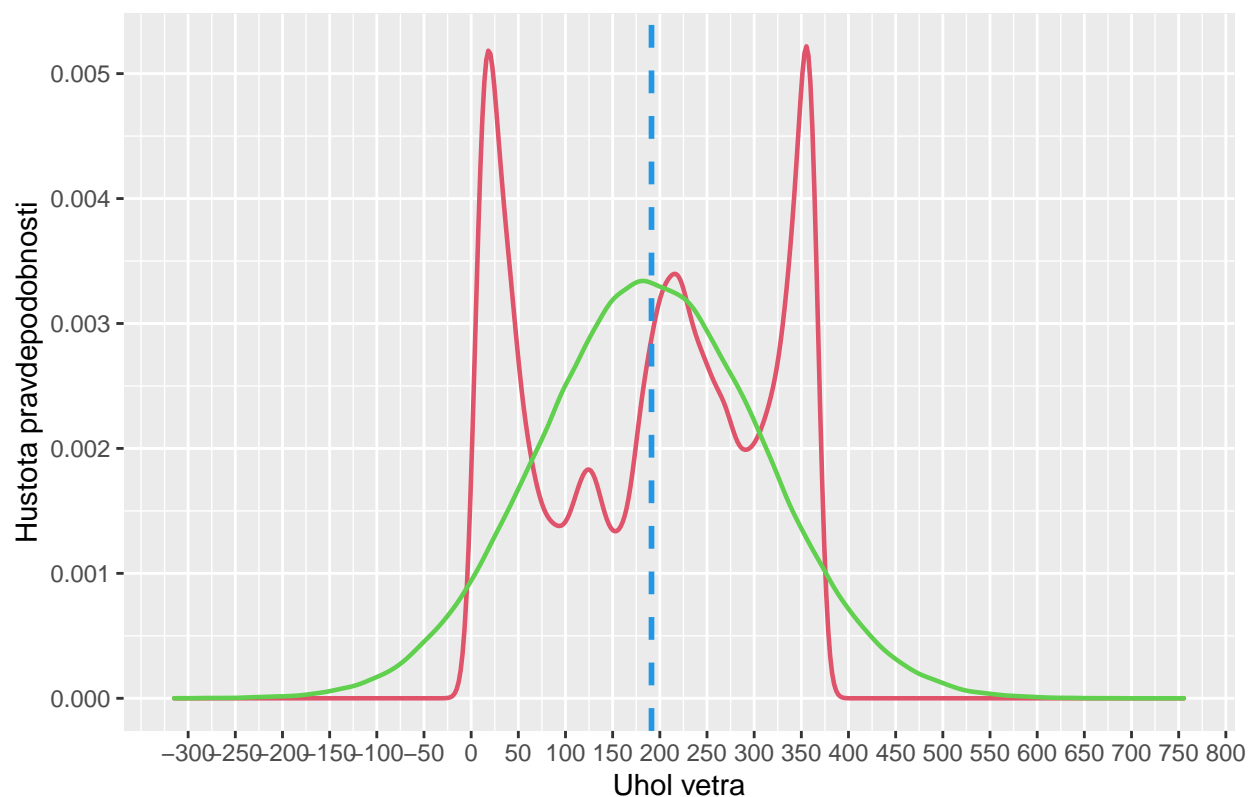
Graf hustoty

Graf hustoty slúži na porovnanie priebehu hustoty pravdepodobnosti normálneho rozdelenia (zelená čiara) a odhadu hustoty vypočítaného z namerných hodnôt uhlu vetra (červená čiara). Čiary nie sú rovnaké, teda nejde o normálne rozdelenie. Modrá prerušovaná čiara predstavuje priemer hodnôt uhlu vetra. V dátach sa nachádza niekoľko vrcholov pričom krajné hodnoty okolo 0 a 360 stupňov je možné spojiť, nakoľko uhol je cyklická veličina.

```
# density plot
# data z normalneho rozdelenia
data_norm <- data.frame(dens = c(rnorm(length(na.omit(all_data$WND_ANGLE)), mean(all_data$WND_ANGLE, na.rm = T), sd = 100))),

# porovnanie hodnot normalneho rozdelenia a wnd_angle
ggplot(all_data, aes(x = WND_ANGLE), 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_ANGLE, na.rm = T)),
             color = 4, linetype = "dashed", size=1) +
  scale_x_continuous(breaks = seq(-300, 800, by = 50)) +
  labs(title = paste("Odhad hustoty uhlu vetra")) +
  xlab("Uhol vetra") +
  ylab("Hustota pravdepodobnosti")
```

Odhad hustoty uhlu vetra



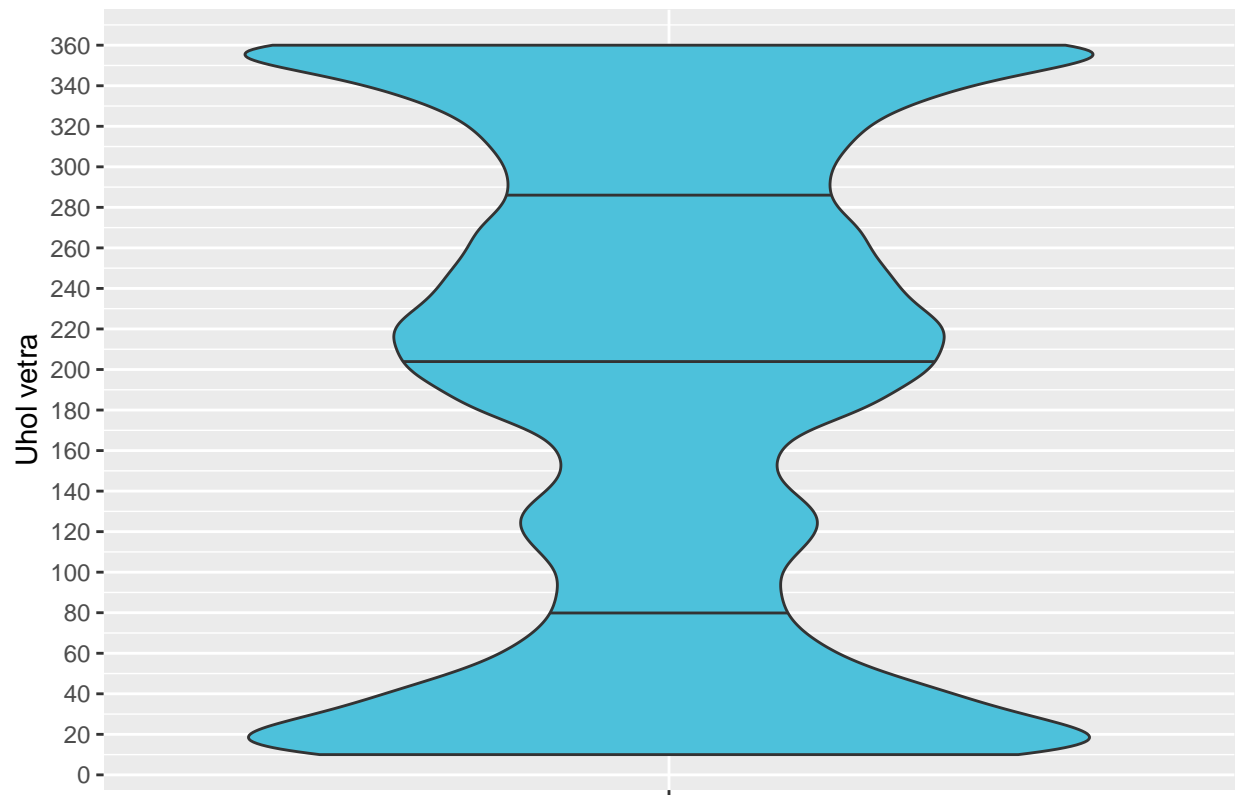
Husľový graf

Husľový graf doplnený o hlavné kvantily zobrazuje rozdelenie hustoty, pričom aj podľa tohto grafu vidíme, že nejde o normálne rozdelenie.

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

ggplot(data = df, aes(factor(label), angle, fill = angle)) +
  geom_violin(draw_quantiles = c(0.25, 0.5, 0.75), fill = "#4dc1db") +
  labs(title = paste("Husľový graf uhlu vetra"), y = "Uhol vetra", fill = "angle") +
  theme(axis.title.x = element_blank()) +
  theme(axis.text.x = element_blank()) +
  scale_y_continuous(breaks = seq(0, 400, by = 20))
```

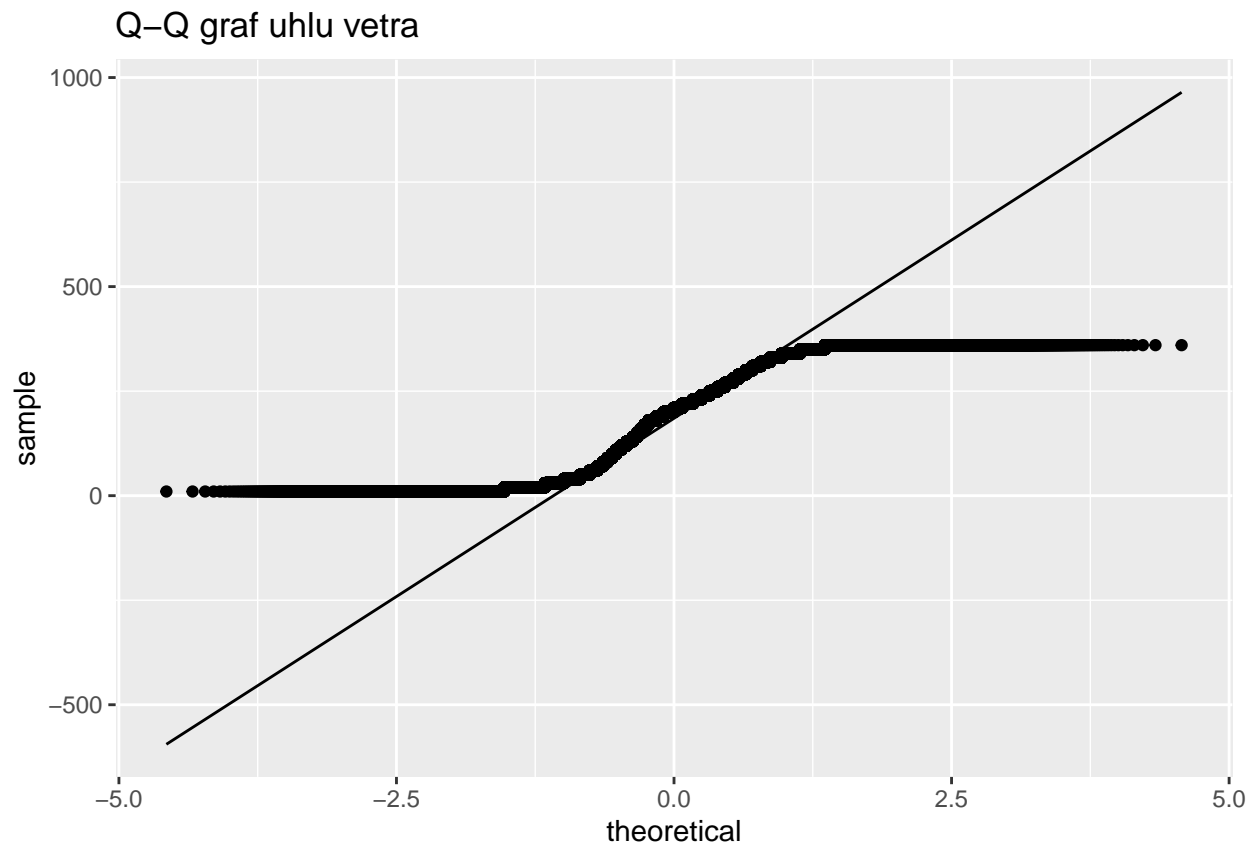
Huslový graf uhlu vetra



Q-Q graf

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

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

Graf polosum

Z grafu vidno, že hodnoty nie sú symetrické.

```
angle <- all_data$WIND_ANGLE
angle_asc <- sort(angle, decreasing = FALSE)
angle_desc <- sort(angle, decreasing = TRUE)

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

Graf polosum pre uhol vetra

