

DEW

Adam Stuller

DEW - rosny bod

```
all_data <- read.csv(file= "../data/all.csv")  
  
describe(all_data$DEW)  
  
## all_data$DEW  
##      n    missing distinct    Info     Mean      Gmd     .05     .10  
##  412837        499      486       1   4.574   8.841    -9.0    -5.8  
##    .25       .50       .75       .90      .95  
##    -1.0       5.0      11.0      14.5     16.0  
##  
## lowest : -29.5 -28.4 -28.0 -27.1 -27.0, highest:  21.4  21.5  21.8  22.0  23.0
```

Centralna poloha dat

Priemer je 4.5742, median 5. Tieto hodnoty sú si dost blízke a vidime, že data nie sú veľmi vychylene kvôli outlierom. Modus je ale 0, co je zaujímave.

```
getmode(na.omit(all_data$DEW)) %>%  
  print(cat("Modus: " ))  
  
## Modus: [1] 0  
  
median(all_data$DEW, na.rm = TRUE) %>%  
  print(cat("Median: "))  
  
## Median: [1] 5  
  
mean(all_data$DEW, na.rm = TRUE) %>%  
  print(cat("Mean: "))  
  
## Mean: [1] 4.5742
```

Variabilita

Vyberovy rozptyl je 60.56788, standardna odchylka 7.782537 a variancny koeficient 1.701398, cize 170,14 %. To je dosta velia a variancia je teda dosta velka.

Medzikvartilova odchylka je iba 6

Variacne rozpatie je 52.5

```
max_dew <- max(all_data$DEW, na.rm= TRUE)  
min_dew <- min(all_data$DEW, na.rm= TRUE)  
var_rozpatie <- max_dew - min_dew  
print(cat("Variacne rozpatie", var_rozpatie))
```

```

## Variacne rozpatie 52.5NULL

# Interquartile range
Q1 <- quantile(all_data$DEW, 0.25, na.rm = T) # 25% hodnot je mensich a 75% vacsich
Q3 <- quantile(all_data$DEW, 0.75, na.rm = T) # 75% hodnot je mensich a 25% vacsich

(IQR(all_data$DEW, na.rm = T) / 2) %>%# interquartile range
print(cat("Medzikvantilova odchýlka: "))

## Medzikvantilova odchýlka: [1] 6

var(all_data$DEW, na.rm = T) %>% print(cat("Rozptyl: "))# rozptyl

## Rozptyl: [1] 60.56788

EnvStats::cv(all_data$DEW, na.rm = T) %>% print(cat("Variacny koeficient: "))# variacny koeficient

## Variacny koeficient: [1] 1.701398

summary(all_data$DEW)

##      Min.   1st Qu.    Median     Mean  3rd Qu.    Max.    NA's
## -29.500 -1.000    5.000   4.574  11.000  23.000    499

all_data['DEW'] %>% profiling_num()

##      variable   mean std_dev variation_coef p_01 p_05 p_25 p_50 p_75 p_95 p_99
## 1       DEW 4.5742 7.782537     1.701398  -15   -9   -1     5   11   16   18
##      skewness kurtosis iqr  range_98   range_80
## 1 -0.3767112  2.6893  12 [-15, 18] [-5.8, 14.5]

```

Asimetria

Šikmost (skewness) je -0.3767112. Je teda zaporna a hovori, ze data su jemne nachylene doprava.

Špicatost (kurtosis) 2.6893 je kladna teda mierne spicatejsia ako normalne rozdelenie.

Histogram

Data sa podobaju na normalne rozdelenie, aj ked vrchol je trosku dlhsy a plytkejsi. Nahnutie su doprava a z tej strany su aj dost strme.

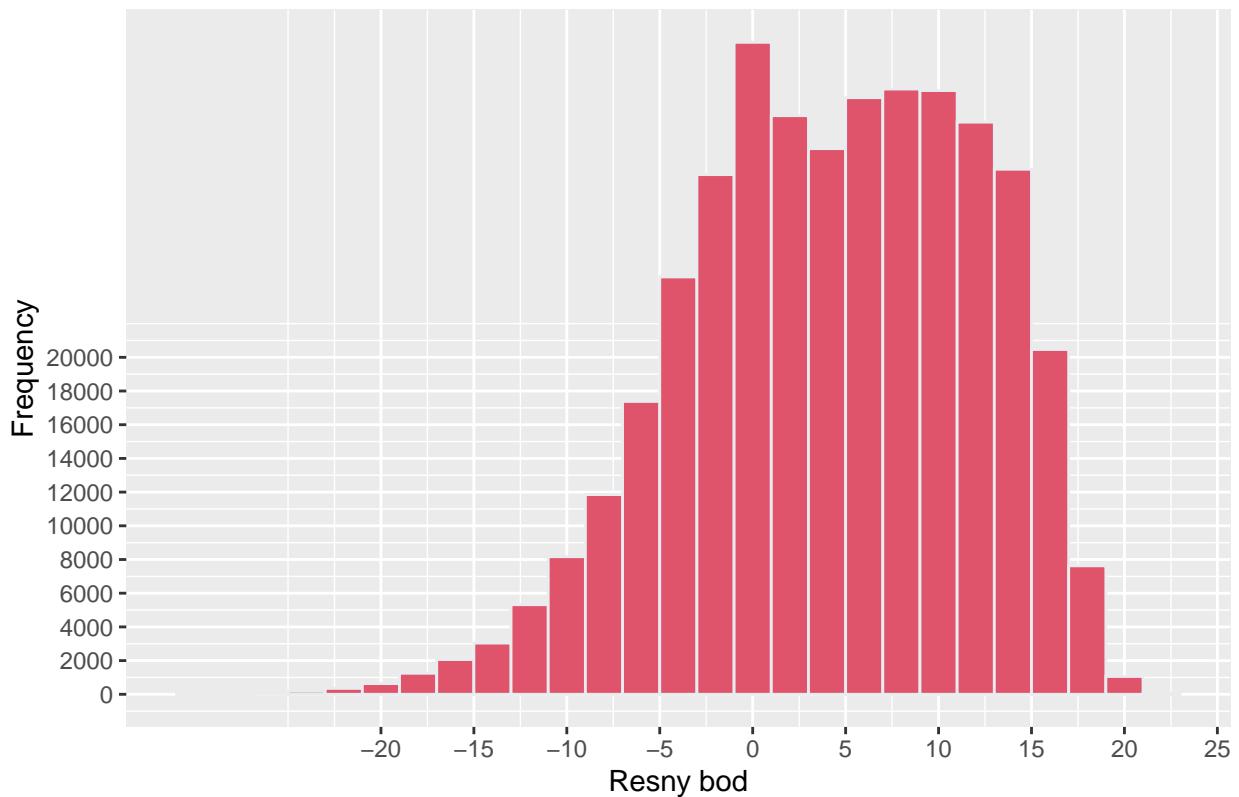
```

ggplot(all_data, aes(x=DEW)) +
  geom_histogram(bins = 40, binwidth = 2, fill="#e9ecf", color="#e9ecf") +
  labs(title = paste("Rosny bod - histogram")) +
  xlab("Resny bod") +
  ylab("Frequency") +
  scale_x_continuous(breaks = seq(-20, 40, by = 5)) +
  scale_y_continuous(breaks = seq(0, 20000, by = 2000))

## Warning: Removed 499 rows containing non-finite values (stat_bin).

```

Rosny bod – histogram

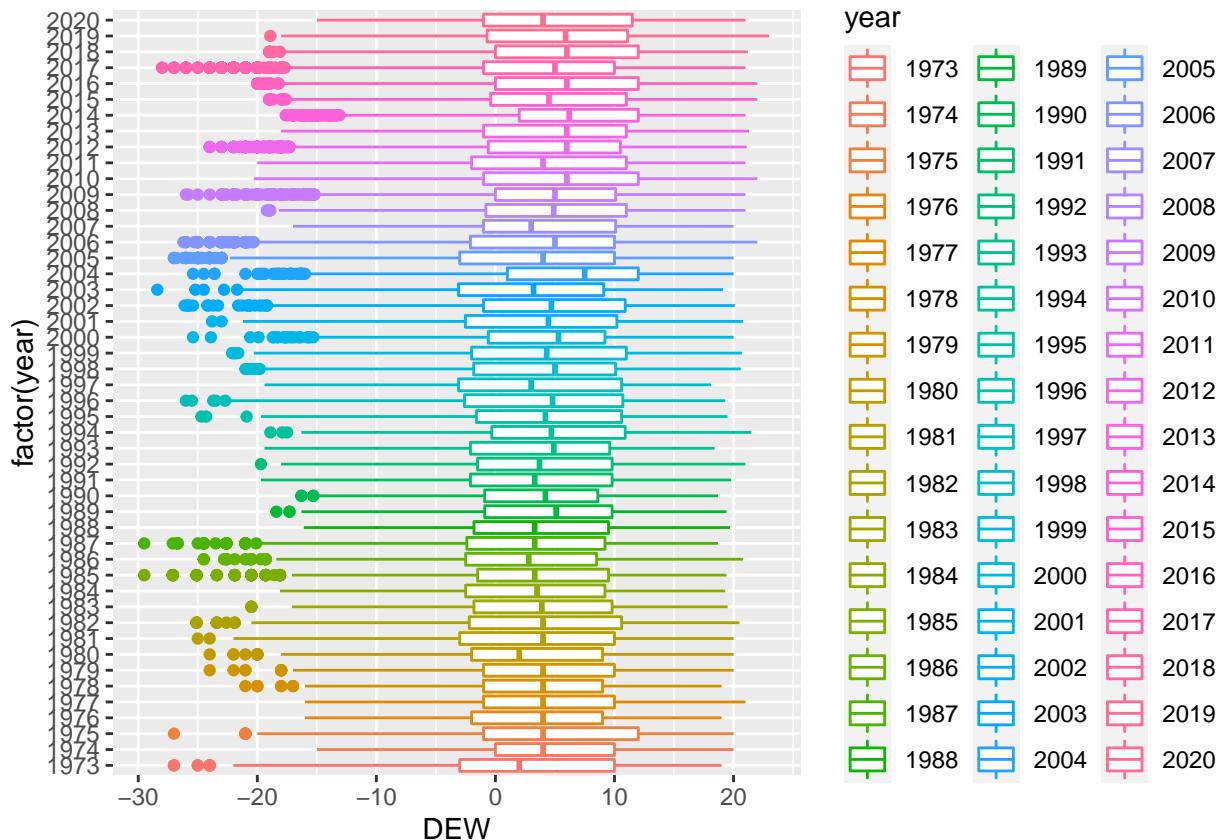


Boxplot

```
df <- all_data %>%
  dplyr::mutate(
    year = ymd_hms(DATE) %>%
      lubridate::year() %>%
      map_chr(~ as.character(.x))
  ) %>%
  dplyr::select(all_of(c('year', 'DEW')))

ggplot(data = df, aes(DEW, factor(year), colour=year)) +
  geom_boxplot()

## Warning: Removed 499 rows containing non-finite values (stat_boxplot).
```



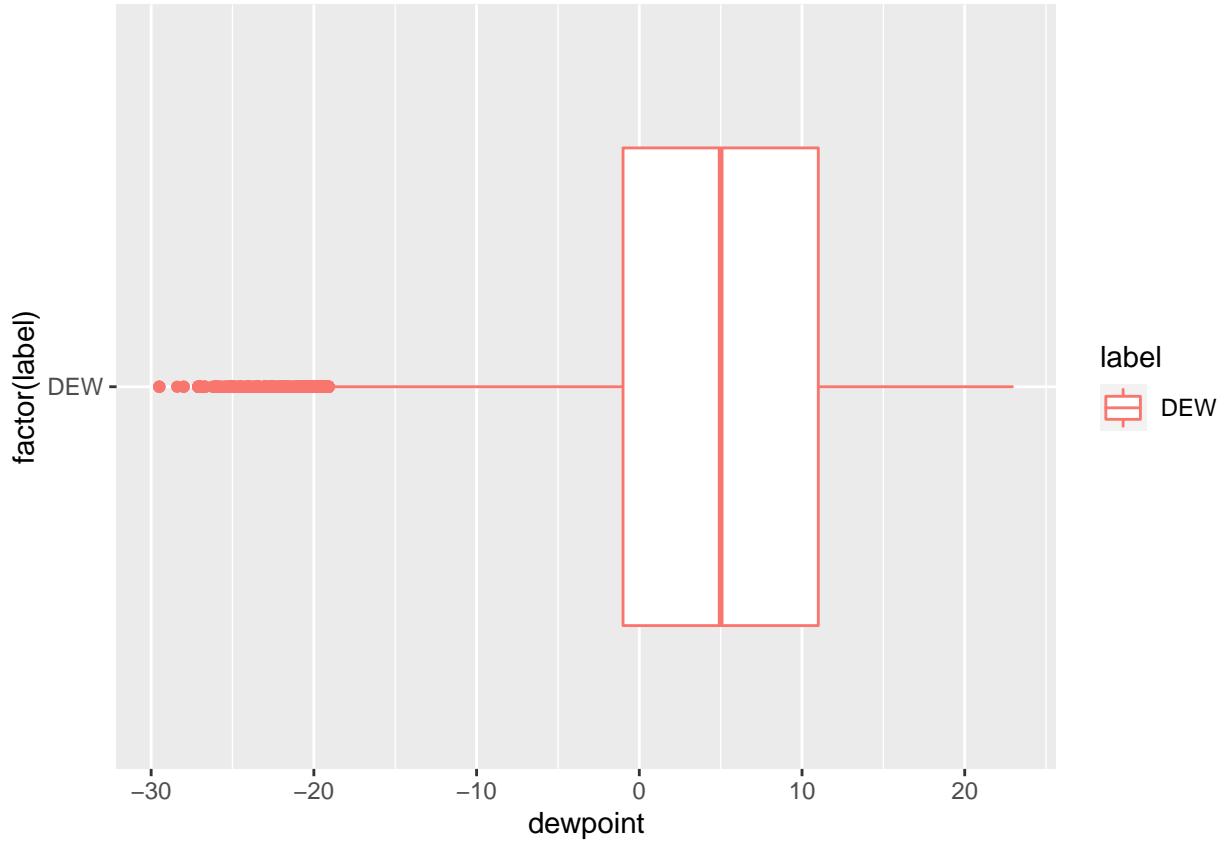
```

df <- all_data %>%
  dplyr::select('DEW') %>%
  tidyrr::gather(key='label', value = 'dewpoint')

ggplot(data = df, aes( dewpoint,factor(label), colour=label)) +
  geom_boxplot()

## Warning: Removed 499 rows containing non-finite values (stat_boxplot).

```

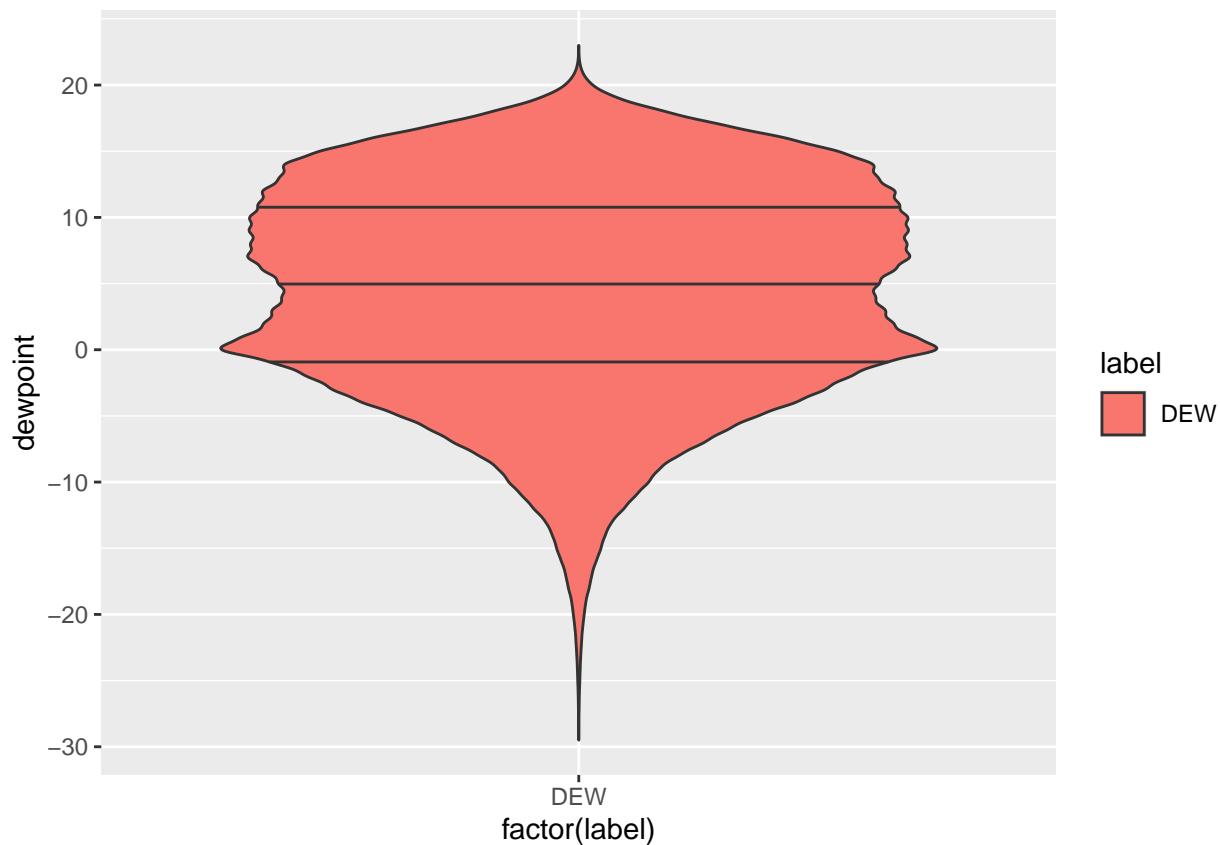


Violin

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

ggplot(data = df, aes(factor(label), dewpoint, fill=label)) +
  geom_violin(draw_quantiles=c(0.25, 0.5, 0.75))

## Warning: Removed 499 rows containing non-finite values (stat_ydensity).
```



Q-Q plot

Ako vidime aj tu, data sa mierne podobaju na normalne rozdelenie.

```
ggplot(data = all_data, aes(sample=DEW)) +  
  stat_qq() +  
  stat_qq_line()  
  
## Warning: Removed 499 rows containing non-finite values (stat_qq).  
## Warning: Removed 499 rows containing non-finite values (stat_qq_line).
```

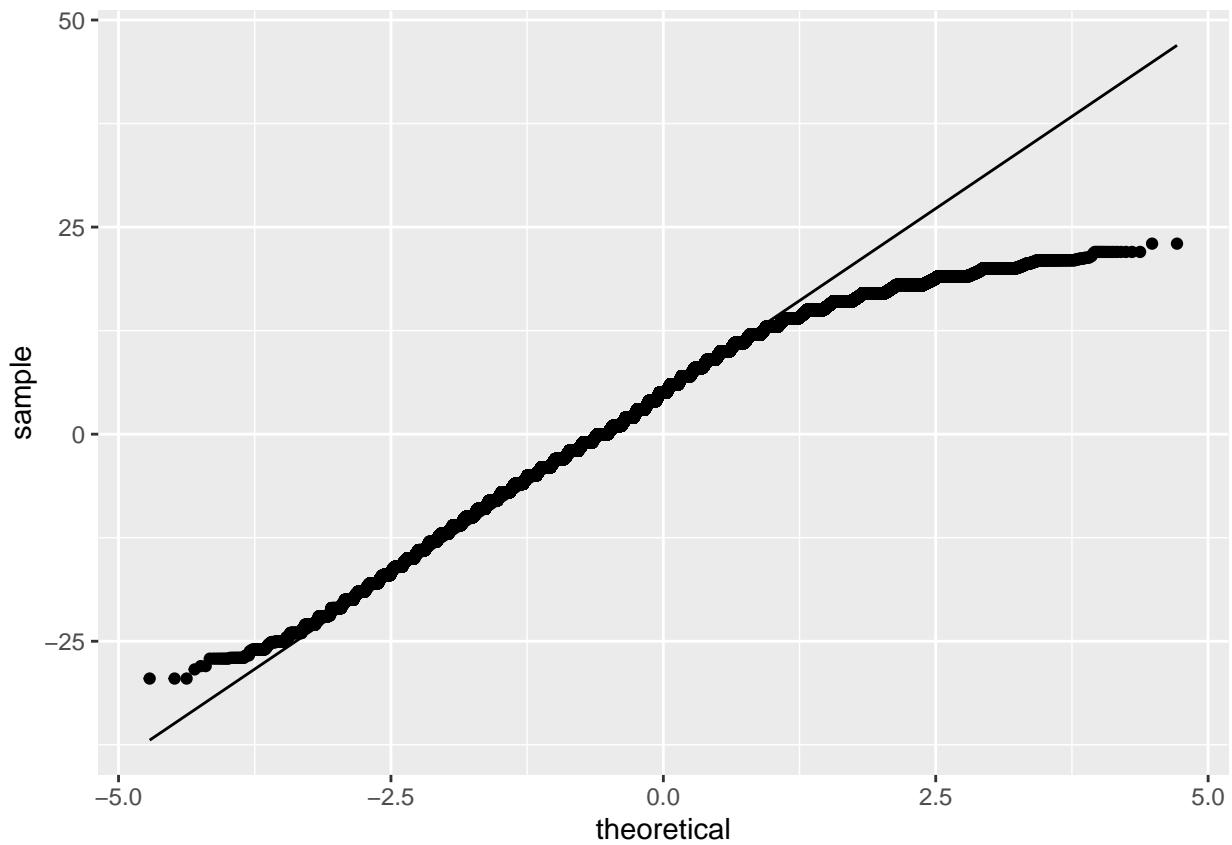
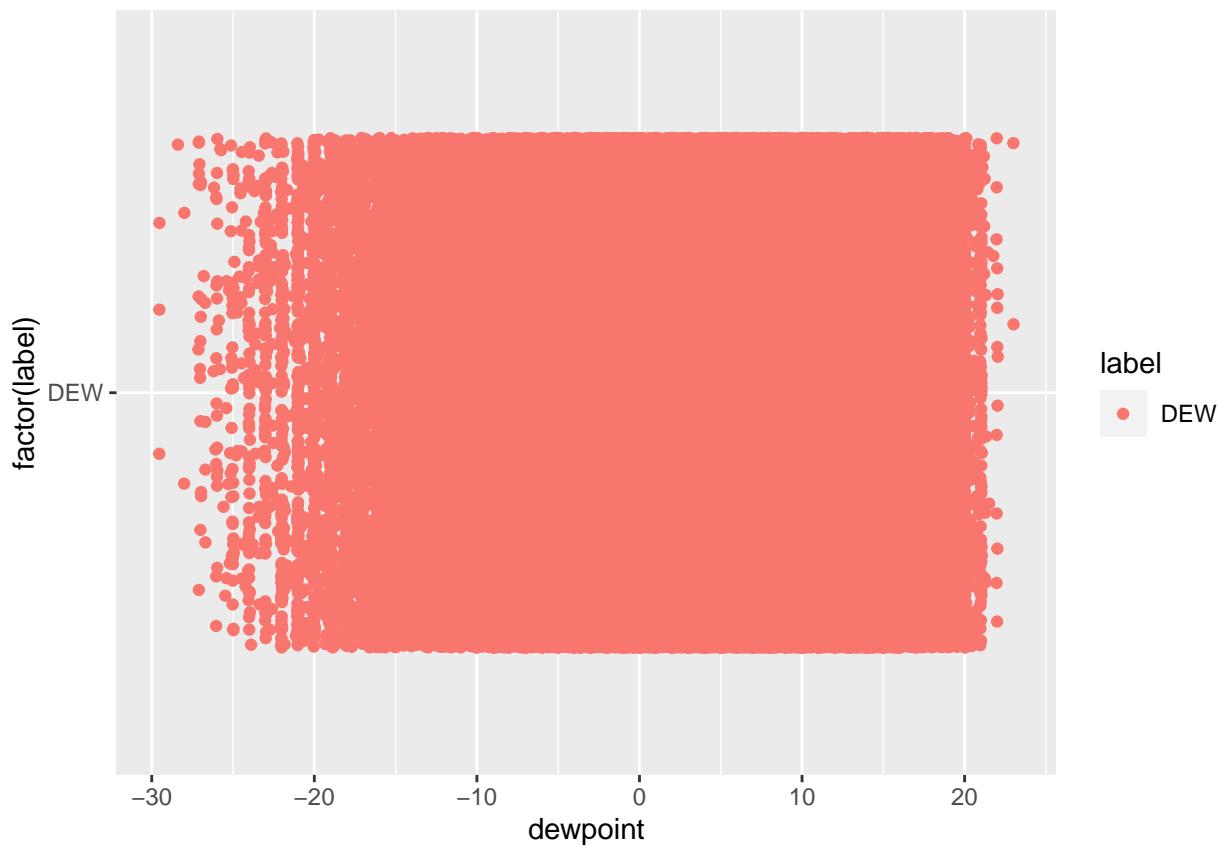


Diagram Rozptylenia

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

ggplot(data = df, aes( dewpoint,factor(label), colour=label)) +
  geom_jitter()

## Warning: Removed 499 rows containing missing values (geom_point).
```



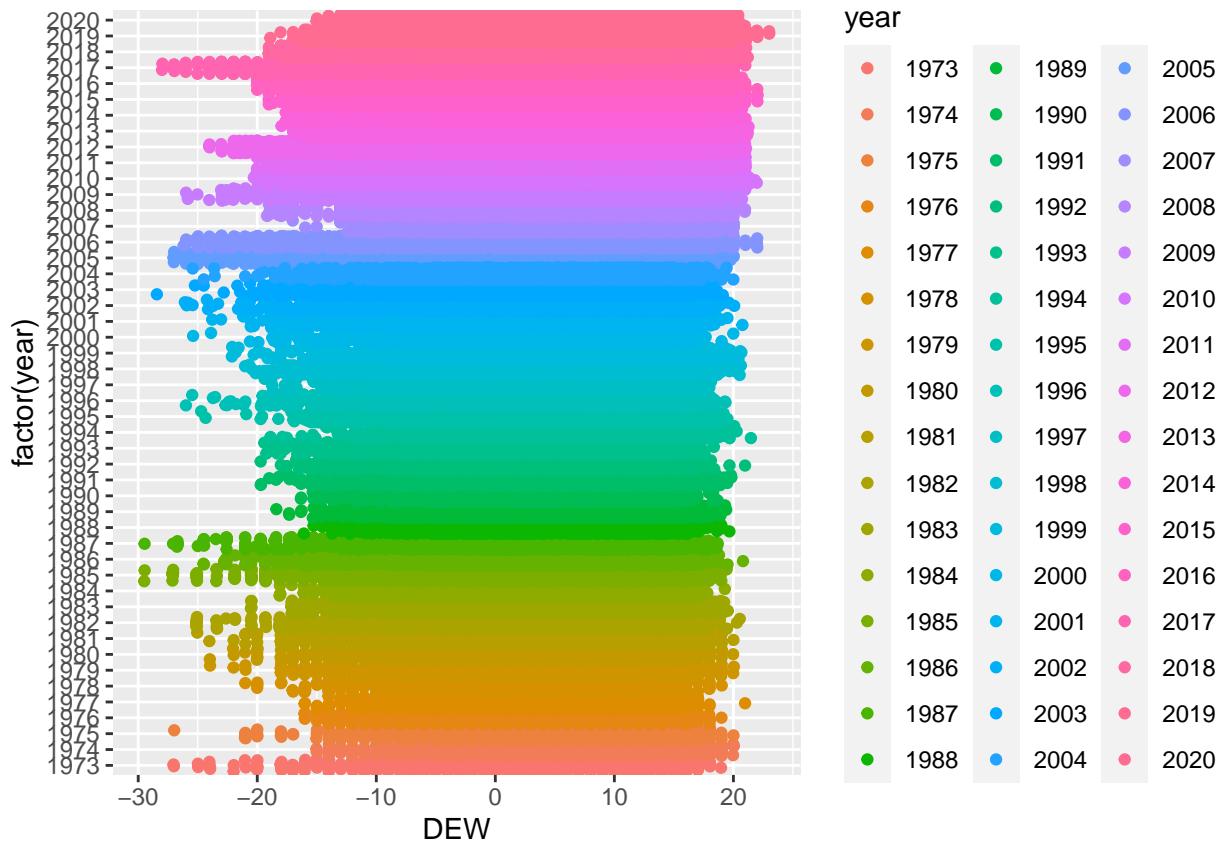
```

df <- all_data %>%
  dplyr::mutate(
    year = ymd_hms(DATE) %>%
      lubridate::year() %>%
      map_chr(~ as.character(.x))
  ) %>%
  dplyr::select(all_of(c('year', 'DEW')))

ggplot(data = df, aes( DEW,factor(year), colour=year)) +
  geom_jitter()

## Warning: Removed 499 rows containing missing values (geom_point).

```



Graf polosum

```

slp <- all_data$DEW
slp_asc <- sort(slp, decreasing = FALSE)
slp_desc <- sort(slp, decreasing = TRUE)

ggplot(data.frame(slp_asc), aes(x = slp_asc, y = 0.5*(slp_asc+slp_desc))) +
  geom_point(size = 2, color = 2) +
  scale_x_continuous(breaks = seq(-30, 40, by = 3)) +
  labs(title = "Graf polosum pre atmosfericky tlak", x = "Tlak") +
  theme_bw()

```

Graf polosum pre atmosfericky tlak

