SNOW_DEPTH

Adam Stuller

SNOW_DEPTH

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

Centralna poloha dat

Hodnota vyberoveho medianu je 1, modus je 0 a vyberovy priemer je 6.60858. Data teda nie su nahromadene centralne a ako neskor uvidime ani z nromalneho rozdelenia, co pri vyske snehu dava zmysel.

```
getmode(na.omit(all_data$SNOW_DEPTH)) %>%
  print(cat("Modus: " ))
```

```
## Modus: [1] 0
```

```
median(all_data$SNOW_DEPTH, na.rm = TRUE) %>%
  print(cat("Median: "))
```

```
## Median: [1] 1
```

```
mean(all_data$SNOW_DEPTH, na.rm = TRUE)%>%
  print(cat("Mean: "))
```

```
## Mean: [1] 6.60858
```

Variabilita

Vyberovy rozptyl je 102.5846 a variancny koeficient 1.532614, co hovori o relativne velkej variabilite dat.

Variacne rozpatie je 207, čo je ale sposobene vychýlenoou hodnootou z roku 1994, kedy z ničoho nič namerali 207 cm snehu, pričom chvílu pred tym bol iba 1cm a potom tiež. Predpokladáme, že je to chyba. Nahradzovať to ale veľmi nebudeme, keďže s týmito dátami, nebudeme veľa pracovať.

Medzi kvantilova odchylka je iba 5. To znamena, že gro dát sa nachádza na malom intervale.

```
var(all_data$SNOW_DEPTH, na.rm = TRUE) # rozptyl
```

```
## [1] 102.5846
```

```
EnvStats::cv(all_data$SNOW_DEPTH, na.rm = T) # variacny koeficient
```

```
## [1] 1.532614
```

```
max_snow <- max(all_data$SNOW_DEPTH, na.rm= TRUE)
min_snow <- min(all_data$SNOW_DEPTH, na.rm= TRUE)
var_rozpatie <- max_snow - min_snow
print(cat("Variacne rozpatie", var_rozpatie))
```

```
## Variacne rozpatie 207NULL
```

```r
# Interquartile range
Q1_slp <- quantile(all_data$SNOW_DEPTH, 0.25, na.rm = T) # 25% hodnot je mensich a 75% vacsich
Q3_slp <- quantile(all_data$SNOW_DEPTH, 0.75, na.rm = T) # 75% hodnot je mensich a 25% vacsich

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

```
## Medzikvantilova odchýlka: [1] 5
```

```r
summary(all_data$SNOW_DEPTH)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     0.0     0.0     1.0     6.6    10.0   207.0  408068
```

```r
all_data['SNOW_DEPTH'] %>% profiling_num()
```

```
##     variable    mean std_dev variation_coef p_01 p_05 p_25 p_50 p_75 p_95  p_99
## 1 SNOW_DEPTH 6.60858 10.1284       1.532614    0    0    0    1   10   28 37.33
##   skewness kurtosis iqr              range_98 range_80
## 1 3.188859 36.71543  10 [0, 37.3299999999999]  [0, 22]
```

Asymetria

Šikmost (skewness) je 3.188859. Je vyrazne kladna, teda rozdelenie je zasikmene do lava.

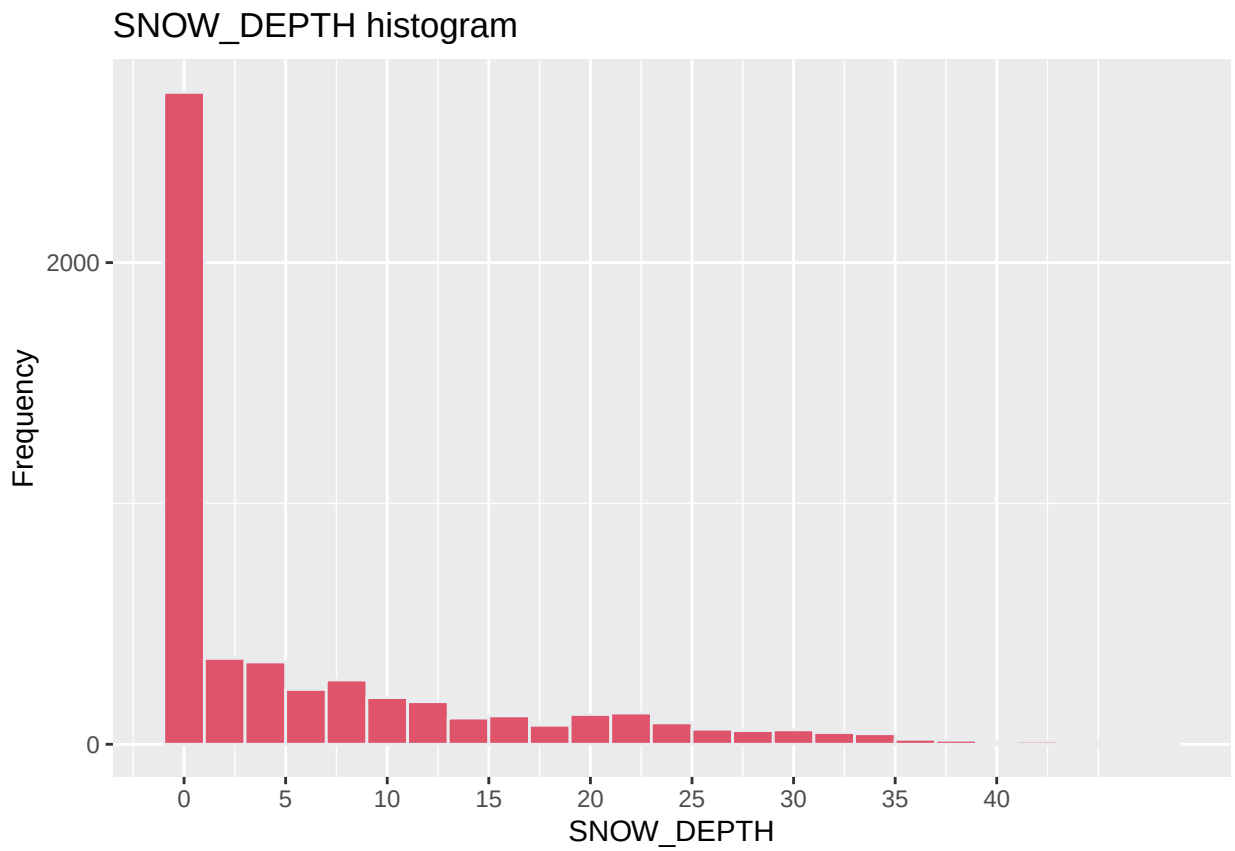
Špicatost (kurtosis) - 36.71543 je kladna a vysoka teda spicatejsia ako data z normálneho rozdelenia.

Histogram

Ako vidime rozdelenie snehu vzdialene pripomina exponencialne rozdelnenie

```r
df <- all_data %>%
  dplyr::filter(SNOW_DEPTH < 100)

ggplot(df, aes(x=SNOW_DEPTH)) +
  geom_histogram(bins = 40, binwidth = 2,fill="2", color="#e9ecef") +
  labs(title = paste("SNOW_DEPTH histogram")) +
  xlab("SNOW_DEPTH") +
  ylab("Frequency") +
  scale_x_continuous(breaks = seq(-20, 40, by = 5)) +
  scale_y_continuous(breaks = seq(0, 20000, by = 2000))
```



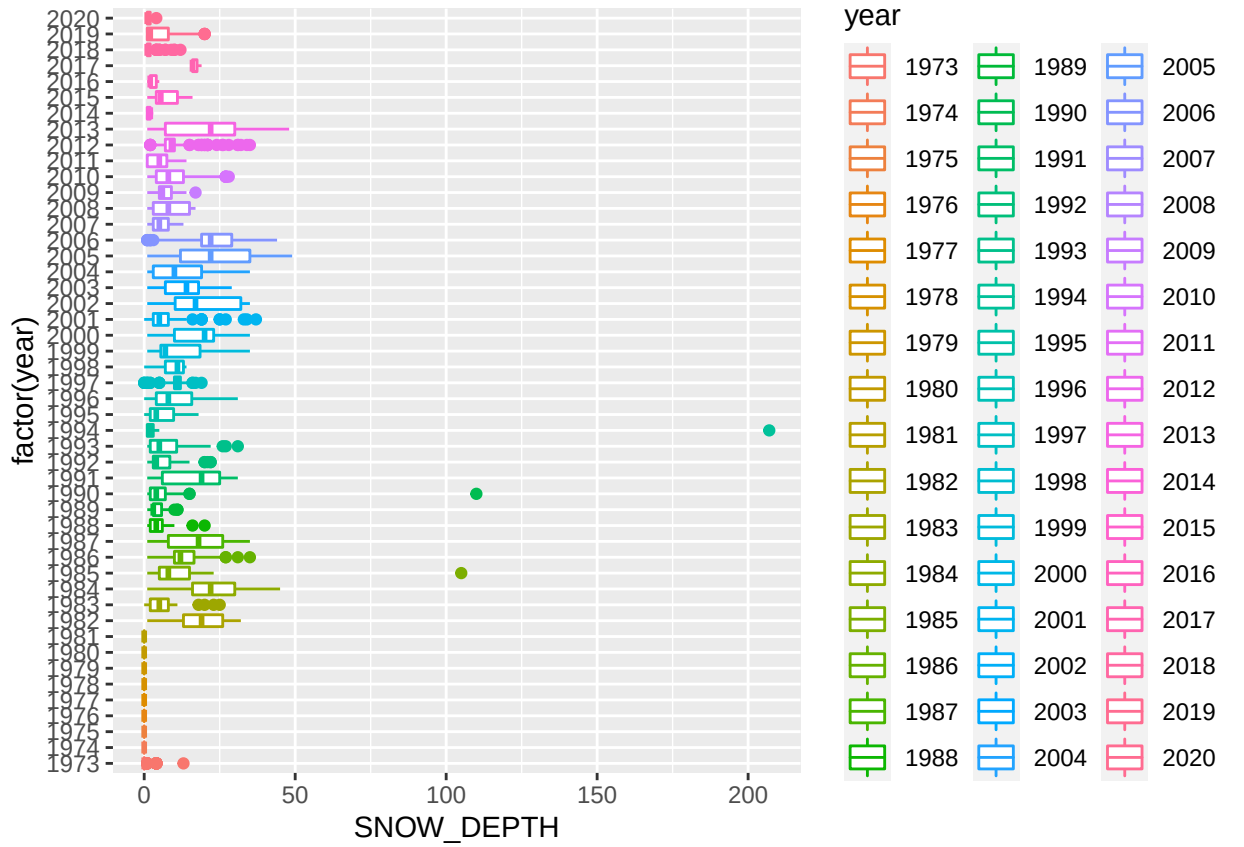
Boxplot

Na boxplote vidime aj vychylene hodnoty.

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

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

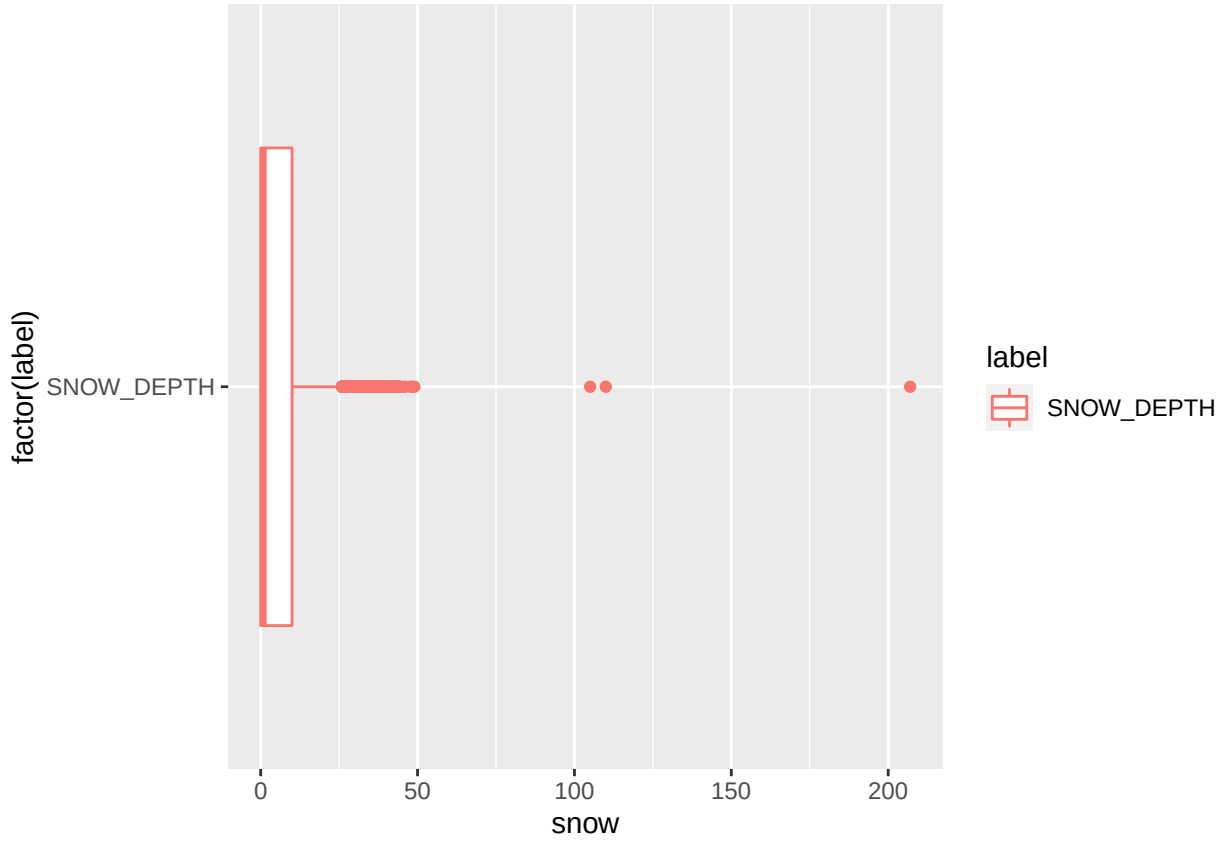
```
## Warning: Removed 408068 rows containing non-finite values (stat_boxplot).
```




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

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

```
## Warning: Removed 408068 rows containing non-finite values (stat_boxplot).
```



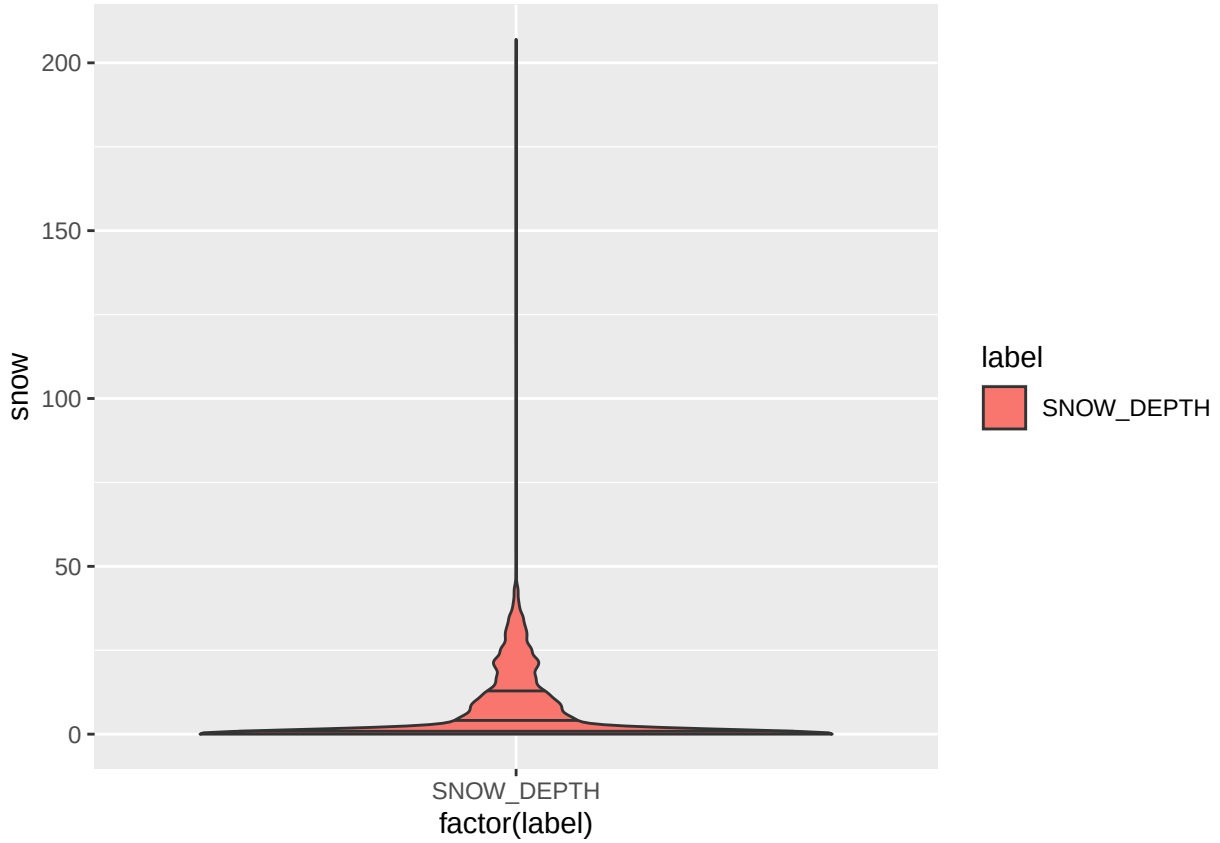
Violin

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

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

```
## Warning: Removed 408068 rows containing non-finite values (stat_ydensity).
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



Q-Q plot

```
ggplot(data = all_data, aes(sample=SNOW_DEPTH)) +
  stat_qq() +
  stat_qq_line()
```

```
## Warning: Removed 408068 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 408068 rows containing non-finite values (stat_qq_line).
```

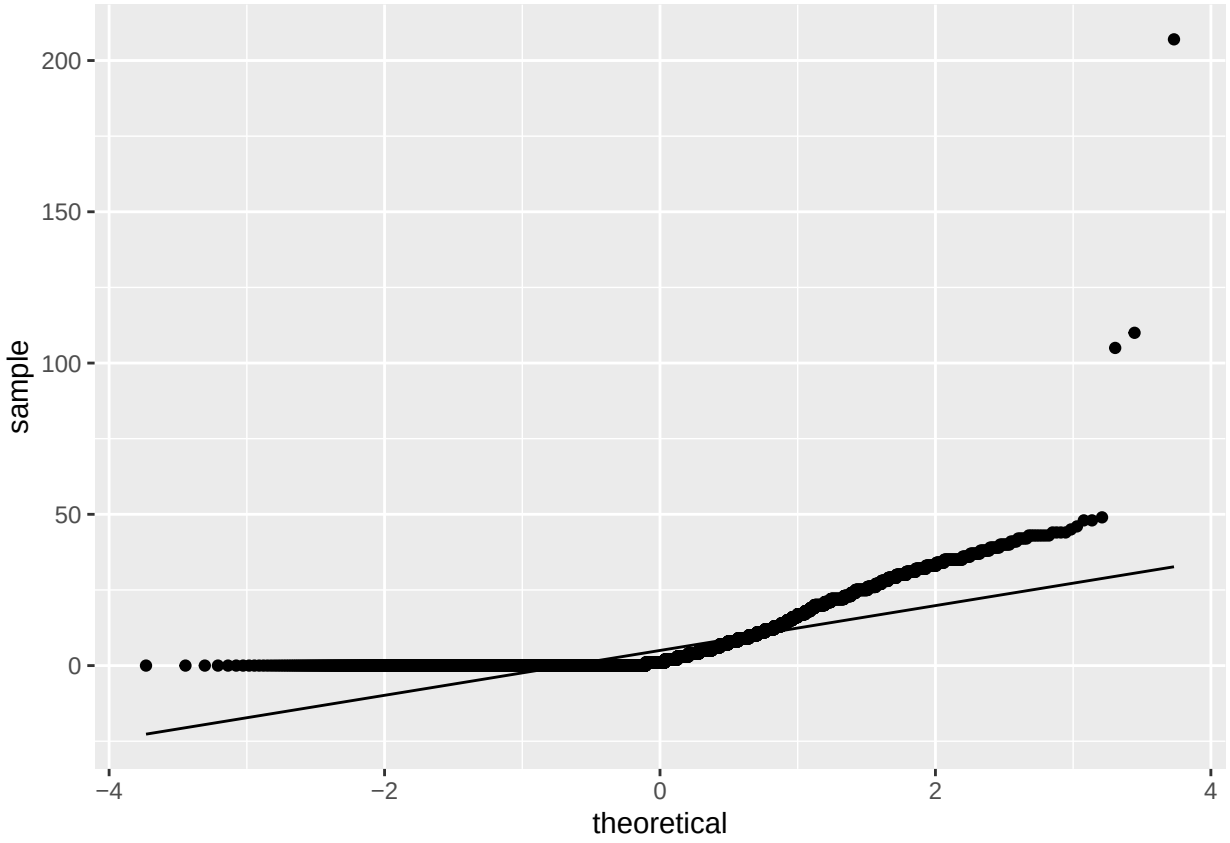
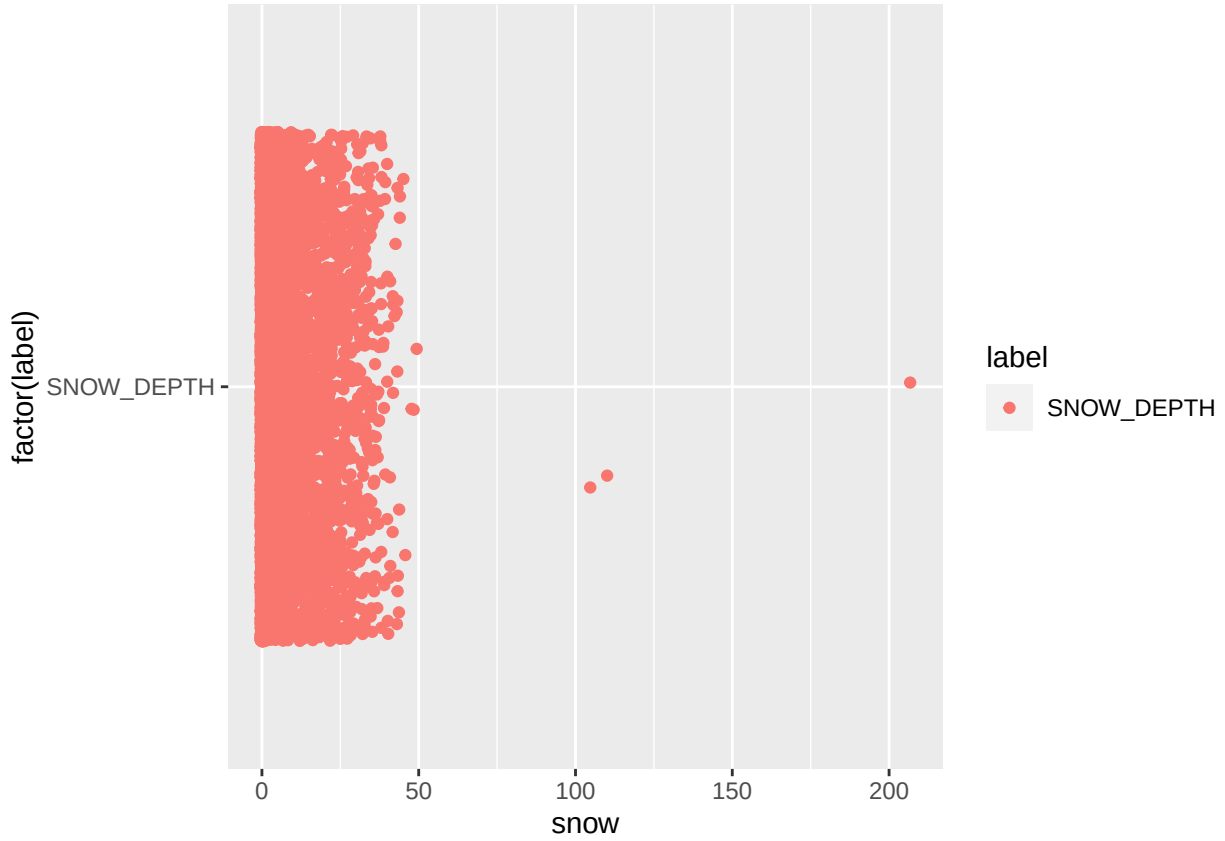


Diagram Rozptylenia

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

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

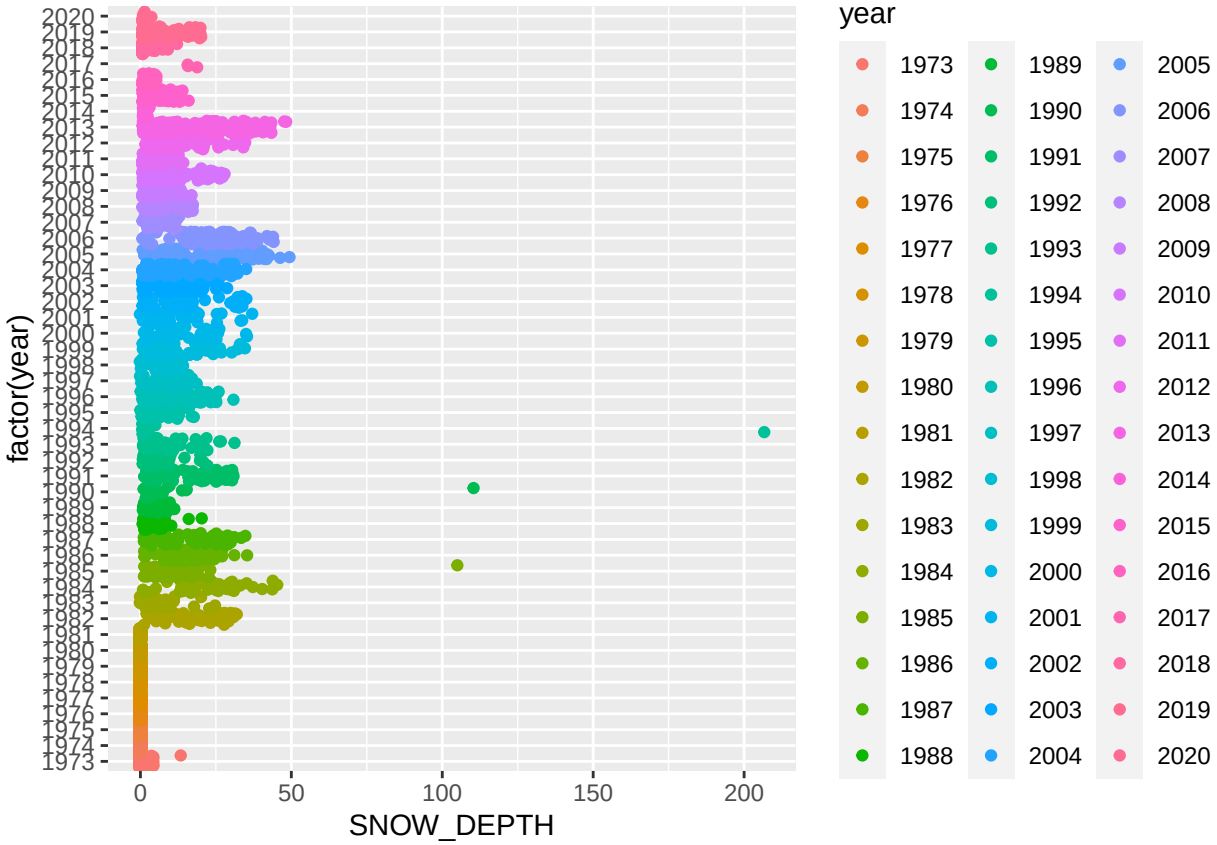
```
## Warning: Removed 408068 rows containing missing values (geom_point).
```




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

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

```
## Warning: Removed 408068 rows containing missing values (geom_point).
```


factor(year)
SNOW_DEPTH
0
50
100
150
200
year
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020