

ZGODNIE Z USTALENIEM PODJALEM SIE PROBY ROZWIAZANIA TYLKO PROBLEMU 1 I 3 Z UWAGI NA BRAK PARTNERA W ZESPOLE

!!!!

PROBLEM 1 Wybrałem Pszczynę, Pułtusk oraz Białowieżę.

```
df <- read.csv(file = 'data/k_d_07_2021.csv', header = FALSE)

cols <- c('STATION_ID', 'STATION_NAME', 'YEAR', 'MONTH', 'DAY',
          'T_MAX', 'STATUS_T_MAX', 'T_MIN', 'STATUS_T_MIN', 'T_AVG', 'STATUS_T_AVG',
          'T_MIN_GROUND', 'STATUS_T_MIN_GROUND', 'SUM_OF_PRECIPITATION', 'STATUS_SUM_OF_PRECIPITATION',
          'TYPE_OF_RAINFALL', 'SNOW_COVER_HEIGHT', 'STATUS_SNOW_COVER_HEIGHT')

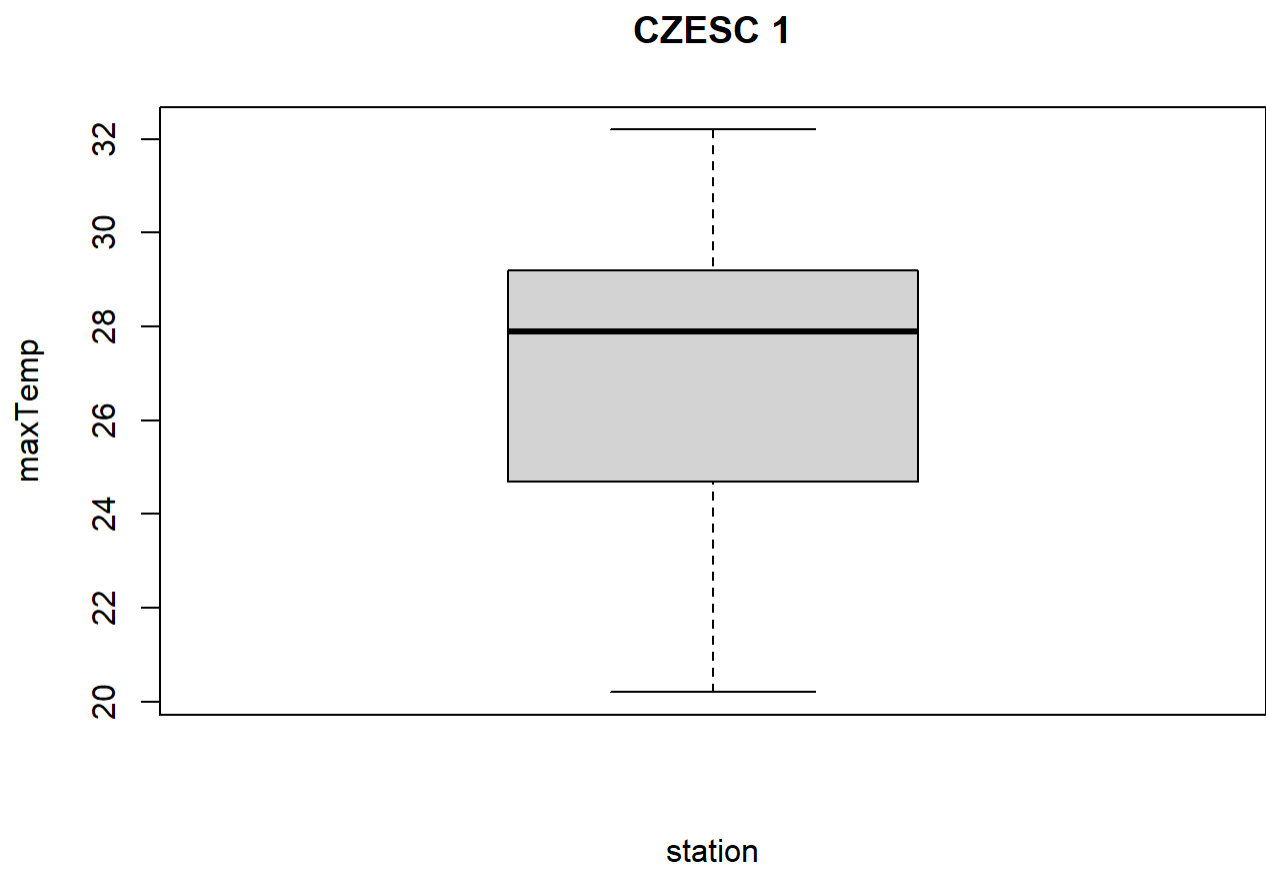
colnames(df) <- cols

stations = c('PSZCZYNA', 'PUŁTUSK', 'BIAŁOWIEŻA')
df <- df[df$STATION_NAME %in% stations, ]
str(df)

## 'data.frame':   31 obs. of  18 variables:
## $ STATION_ID      : int  249180010 249180010 249180010 249180010 249180010 249180010 249180010 249
180010 249180010 249180010 ...
## $ STATION_NAME    : chr   "PSZCZYNA" "PSZCZYNA" "PSZCZYNA" "PSZCZYNA" ...
## $ YEAR            : int   2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
## $ MONTH           : int    7 7 7 7 7 7 7 7 7 ...
## $ DAY             : int    1 2 3 4 5 6 7 8 9 10 ...
## $ T_MAX           : num   23.3 21.5 23.8 25.9 26.5 29.8 32.2 31.5 28.5 25.4 ...
## $ STATUS_T_MAX    : logi   NA NA NA NA NA NA ...
## $ T_MIN           : num   15.1 13.8 13.3 10.8 12 14.3 15.7 14.3 19.1 13.8 ...
## $ STATUS_T_MIN    : logi   NA NA NA NA NA NA ...
## $ T_AVG           : num   19 17.7 18.5 18.8 19.9 23.1 23.8 23.2 22.9 19.2 ...
## $ STATUS_T_AVG    : logi   NA NA NA NA NA NA ...
## $ T_MIN_GROUND    : num   14.2 13.7 12.8 8.3 9.7 10.8 12.2 13.9 17.1 12.3 ...
## $ STATUS_T_MIN_GROUND : int   NA NA NA NA NA NA NA NA ...
## $ SUM_OF_PRECIPITATION : num   1.7 0.4 0 0 0 0 0 0 0 ...
## $ STATUS_SUM_OF_PRECIPITATION: int   NA NA 9 9 9 9 9 9 9 ...
## $ TYPE_OF_RAINFALL  : chr    "w" "w" "" "" ...
## $ SNOW_COVER_HEIGHT : int    0 0 0 0 0 0 0 0 0 ...
## $ STATUS_SNOW_COVER_HEIGHT : int    9 9 9 9 9 9 9 9 9 ...

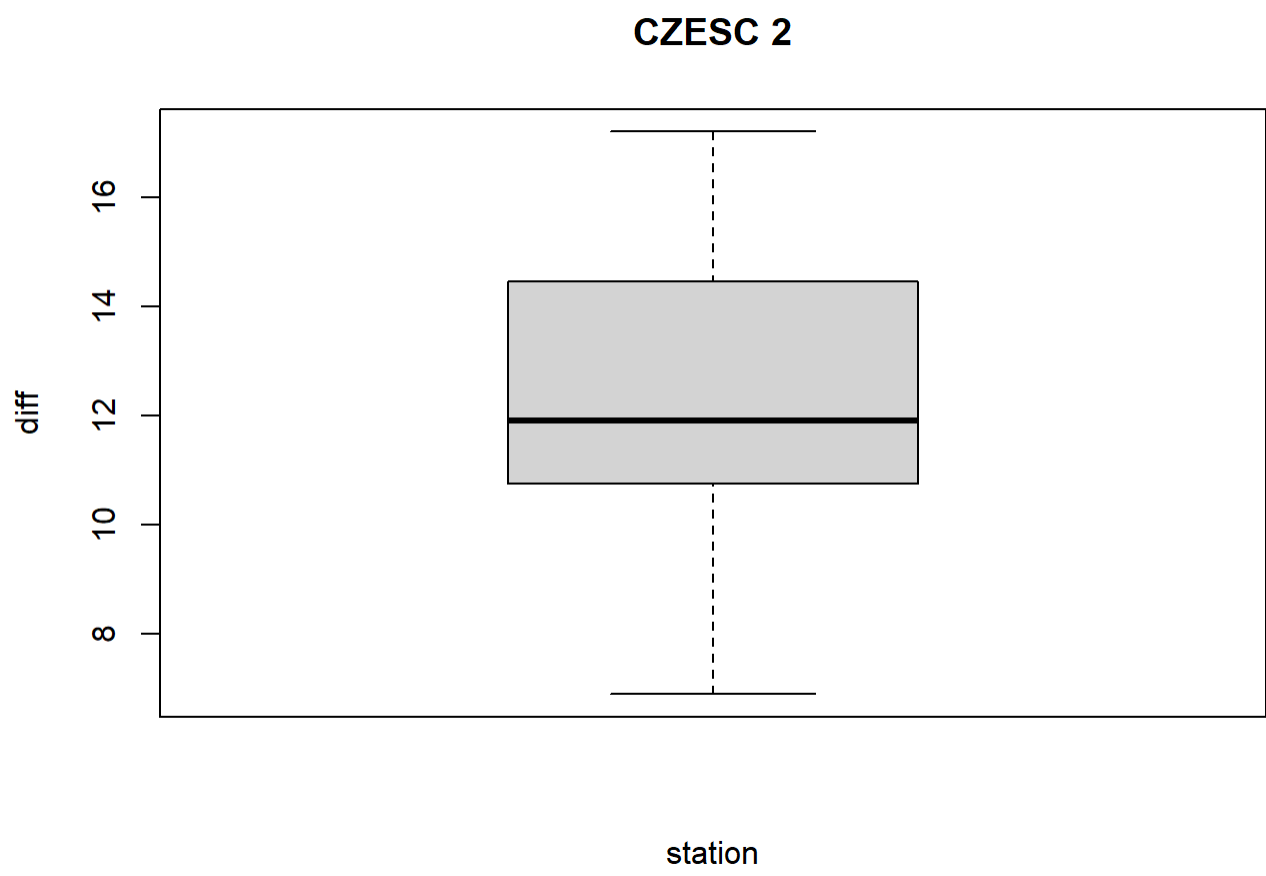
#selection
cut <- c('STATION_NAME', 'YEAR', 'MONTH', 'DAY', 'T_MAX', 'T_MIN', 'T_AVG')
df <- df[, cut]

boxplot(T_MAX~STATION_NAME, df, xlab="station", ylab="maxTemp", main="CZESC 1")
```



Najwyższe temperatury w ciągu dnia zostały zanotowane w Pułtusk. Jednak szysktie stacje posiadają nie roznia sie od siebie wiecej niz jeden stopnien Celsjusza. Sciany pudełka sa w przybliżeniu rozmiarowo podobne.

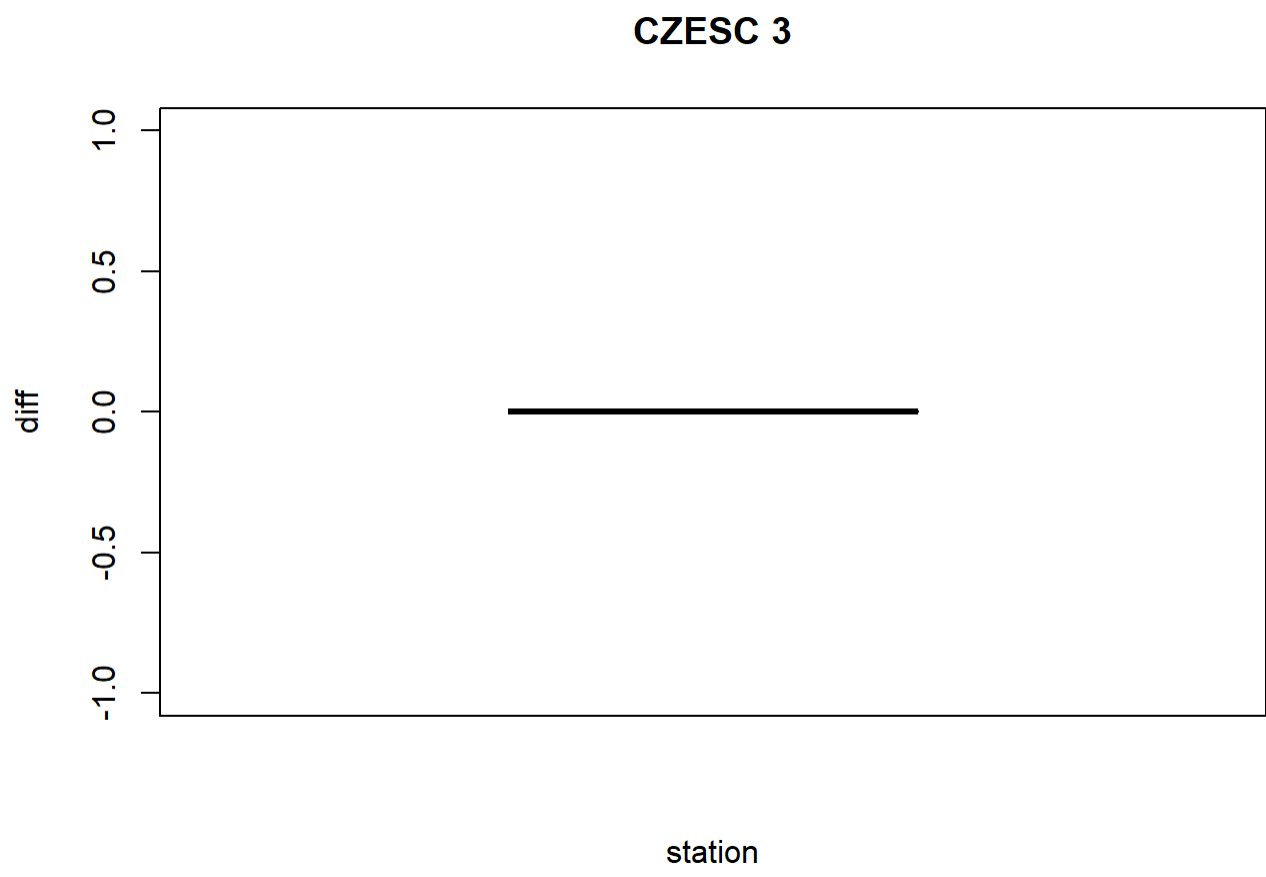
```
df$DAILY_T_DIFF <- df$T_MAX - df$T_MIN
boxplot(
  DAILY_T_DIFF ~ STATION_NAME, df, xlab="station",
  ylab="diff", main="CZESC 2"
)
```



Największe dobowe różnice sa w Pułtusk. Srednia różnic jest o kilka stopni Celsjusza wyzsza od pozostałych.

```
df$DATE <- as.Date(paste(df$YEAR, df$MONTH, df$DAY, sep = "-"), "%Y-%m-%d")
df <- subset(df, select = -c(YEAR, MONTH, DAY))
df <- split(df, with(df, interaction(STATION_NAME)), drop = TRUE)
result = data.frame()
for(station_df in df){
  sorted <- station_df[order(station_df$DATE), ]

  sorted$DIFF_MAX_T_BETWEEN_PREVIOUS_DAY <- sorted$T_MAX - lag(sorted$T_MAX)
  result <- rbind(result, sorted)
}
boxplot(DIFF_MAX_T_BETWEEN_PREVIOUS_DAY ~ STATION_NAME, result, xlab="station",
  ylab="diff", main="CZESC 3")
```



Roznica wacha sie w okolicy zera. Wariacja była najmniejsza w Białowieży, a największa w Pułtusk.

Szacowanie parametrów rozkładu normalnego:

```
data <- result[result$STATION_NAME == 'PUŁTUSK', ]
data <- data$DIFF_MAX_T_BETWEEN_PREVIOUS_DAY
datamean <- mean(data, na.rm = TRUE)
datasd <- sd(data, na.rm = TRUE)
# hist(data, breaks = 10, prob=TRUE, xlab="diff", main="CZESC4", xlim=c(-10,10))
plot.new()
curve(dnorm(x, mean=datamean, sd=datasd), add=TRUE, yaxt="n")
```