

# SAD - sprawozdanie 1

Łukasz Pietraszek, Marcel Kawski

30.11.2021

```
# load libraries

#install.packages("dplyr") # count()
library(dplyr)

#install.packages("zoo") # as.yearqtr()
library(zoo)

#install.packages("magrittr") # %>%
library(magrittr)
```

## Zadanie 1.

a)

i. Wybrane przez nas stacje to: **Marianowo II**, **Niedzica**, **Ceber**.

```
cities <- c('MARIANOWO II', 'NIEDZICA', 'CEBER')
data <- read.csv('/media/dell/Seagate/Studia/mgr/1_sem/SAD/projekty/projekt1/descriptive-statistics/data.csv',
                fill = TRUE, header = FALSE)

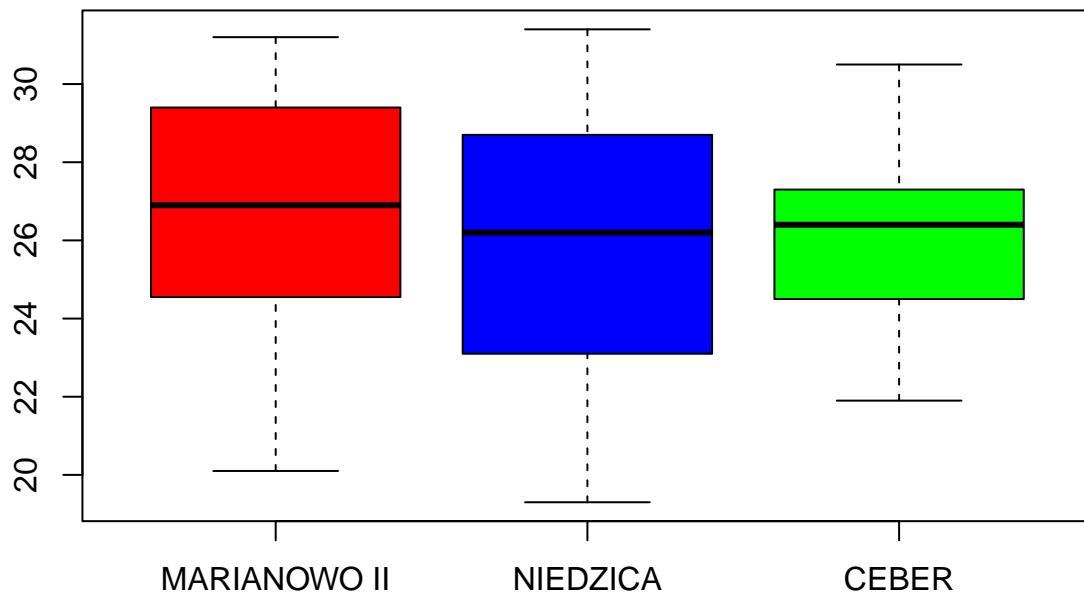
colnames(data) <- c('kod_stacji', 'nazwa_stacji', 'rok', 'miesiąc', 'dzień',
                    'maks_temp', 'stat_pom_tmax', 'min_temp', 'stat_pom_tmin',
                    'sr_temp', 'stat_pom_std', 'temp_min_gr', 'stat_pom_tmng',
                    'suma_opad', 'stat_pom_smdb', 'rodz_opad', 'wys_pokr_sn',
                    'stat_pom_pksn')

data1 <- data[data[, 'nazwa_stacji'] == cities[1],]
data2 <- data[data[, 'nazwa_stacji'] == cities[2],]
data3 <- data[data[, 'nazwa_stacji'] == cities[3],]
```

Wykresy pudełkowe maksymalnych dobowych temperatur dla wybranych miast:

```
colors = c('red', 'blue', 'green')

boxplot(data1[, 'maks_temp'], data2[, 'maks_temp'], data3[, 'maks_temp'],
        names = cities, col = colors)
```



**Komentarz:** Na wykresach widzimy, że najwyższą medianę maksymalnej dobowej temperatury ma miasto Marianowo II. Natomiast największy “rozstrzał” pomiędzy 1. i 3. kwartylem posiada miasto Ceber.

Mediany maksymalnych dobowych temperatur dla wybranych miast:

```
median1 <- median(data1[, 'maks_temp'])
median2 <- median(data2[, 'maks_temp'])
median3 <- median(data3[, 'maks_temp'])

cat(cities[1], ': ', median1, '\n')
```

```
## MARIANOWO II : 26.9
```

```
cat(cities[2], ': ', median2, '\n')
```

```
## NIEDZICA : 26.2
```

```
cat(cities[3], ': ', median3, '\n')
```

```
## CEBER : 26.4
```

**Komentarz:** Najwyższą medianę maksymalnej dobowej temperatury ma miasto Marianowo II. Zgadza się to z danymi zaprezentowanymi na wykresie pudełkowym.

Średnie arytmetyczne maksymalnych dobowych temperatur dla wybranych miast:

```
mean1 <- mean(data1[, 'maks_temp'])
mean2 <- mean(data2[, 'maks_temp'])
mean3 <- mean(data3[, 'maks_temp'])
```

```
cat(cities[1], ': ', mean1, '\n')
```

```
## MARIANOWO II : 26.95806
```

```
cat(cities[2], ': ', mean2, '\n')
```

```
## NIEDZICA : 25.68065
```

```
cat(cities[3], ': ', mean3, '\n')
```

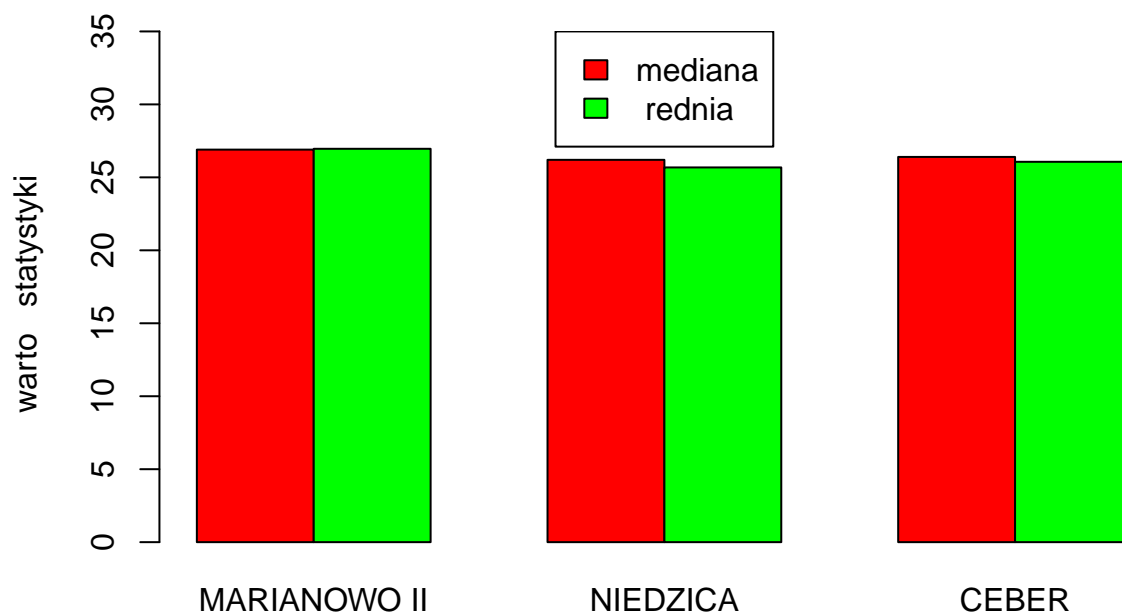
```
## CEBER : 26.06452
```

**Komentarz:** Najwyższą średnią arytmetyczną maksymalnej dobowej temperatury ma miasto Marianowo II. Zgadza się to z intuicyjną analizą wykresów pudełkowych.

Porównanie statystyk:

```
medians <- c(median1, median2, median3)
means <- c(mean1, mean2, mean3)
stats <- rbind(medians, means)

barplot(stats, names = cities, ylim = c(0, 35), ylab = 'wartość statystyki',
        col = c("red", "green"), beside=TRUE)
legend('top', c('mediana', 'średnia'), fill = c("red", "green"))
```



**Komentarz:** Wykres ten nie obrazuje dobrze różnic w statystykach. Najbardziej przydatnym wykresem okazuje się wykres pudełkowy.

Odchylenia standardowe maksymalnych dobowych temperatur dla wybranych miast:

```
sd1 <- sd(data1[, 'maks_temp'])
sd2 <- sd(data2[, 'maks_temp'])
sd3 <- sd(data3[, 'maks_temp'])

cat(cities[1], ': ', sd1, '\n')
```

```
## MARIANOWO II : 2.872255
```

```
cat(cities[2], ': ', sd2, '\n')
```

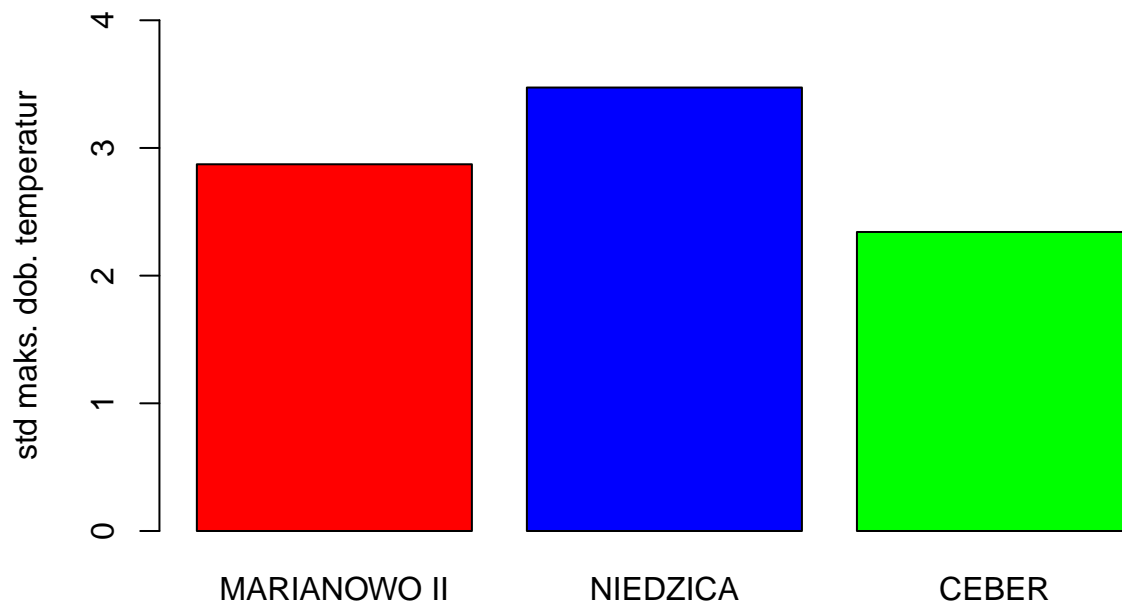
```
## NIEDZICA : 3.473175
```

```
cat(cities[3], ': ', sd3, '\n')
```

```
## CEBER : 2.341872
```

Porównanie odchyłeń standardowych:

```
sds <- c(sd1, sd2, sd3)
barplot(sds, names = cities, ylim = c(0, 4), ylab = 'std maks. dob. temperatur',
        col = colors)
```



**Komentarz:** Dane przedstawione na wykresie również pokrywają się z szerokościami pudełek na wykresach pudełkowych. W dalszych badaniach ten właśnie wykres będzie głównym sposobem wizualizacji danych.

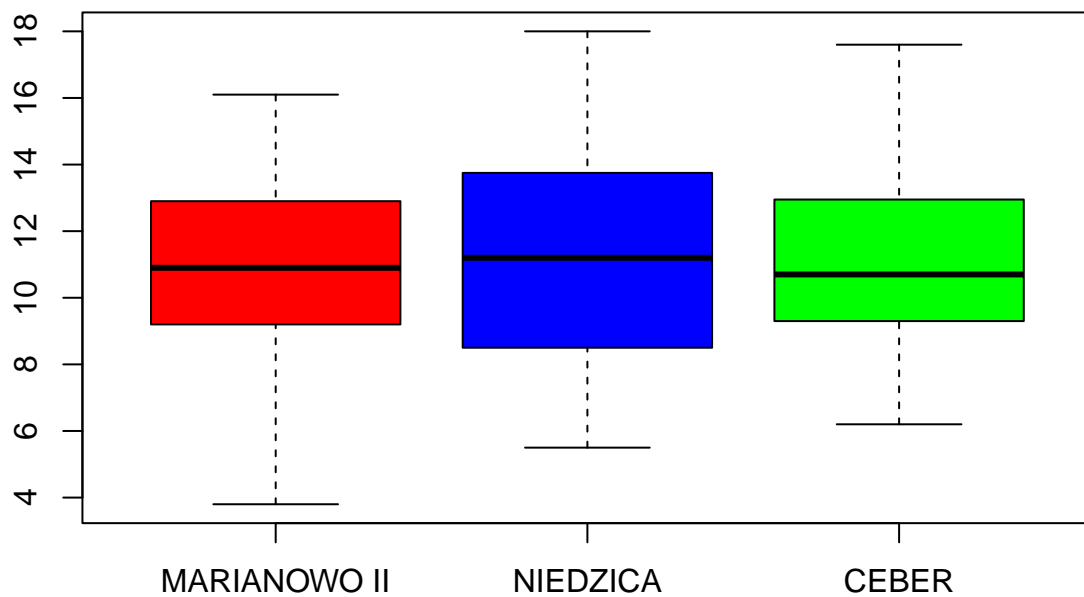
**Odpowiedź:** Najcieplejsze miasta według najwyższej dobowej temperatury dla każdej ze statystyk to kolejno: Marianowo II, Ceber, Niedzica.

ii. Dobowe amplitudy temperatur:

```
ampl1 = data1[, 'maks_temp'] - data1[, 'min_temp']
ampl2 = data2[, 'maks_temp'] - data2[, 'min_temp']
ampl3 = data3[, 'maks_temp'] - data3[, 'min_temp']
```

Wykresy pudełkowe dobowych amplitud temperatur:

```
boxplot(ampl1, ampl2, ampl3, names = cities, col = c('red', 'blue', 'green'))
```



Mediany dobowych amplitud temperatur dla wybranych miast:

```
median1 <- median(ampl1)
median2 <- median(ampl2)
median3 <- median(ampl3)

cat(cities[1], ': ', median1, '\n')
```

```
## MARIANOWO II : 10.9
```

```
cat(cities[2], ': ', median2, '\n')
```

```
## NIEDZICA : 11.2
```

```
cat(cities[3], ': ', median3, '\n')
```

```
## CEBER : 10.7
```

**Komentarz:** Kolejność miast według mediany dobowej amplitudy temperatur to: Niedzica, Marianowo II, Ceber.

Średnie dobowych amplitud temperatur dla wybranych miast:

```
mean1 <- mean(amp11)
mean2 <- mean(amp12)
mean3 <- mean(amp13)
```

```
cat(cities[1], ': ', mean1, '\n')
```

```
## MARIANOWO II : 10.92258
```

```
cat(cities[2], ': ', mean2, '\n')
```

```
## NIEDZICA : 11.33548
```

```
cat(cities[3], ': ', mean3, '\n')
```

```
## CEBER : 11.1129
```

**Komentarz:** Kolejność miast według średniej arytmetycznej dobowej amplitudy temperatur to: Niedzica, Ceber, Marianowo II.

**Odpowiedź:** Nie można jednoznacznie odpowiedzieć, w którym mieście dobową amplitudę temperatur jest najniższa (gdzie temperatura była najbardziej stabilna). Najniższą medianę dobowej temperatury osiągnęło miasto Ceber, natomiast najniższą średnią miasto Marianowo II.

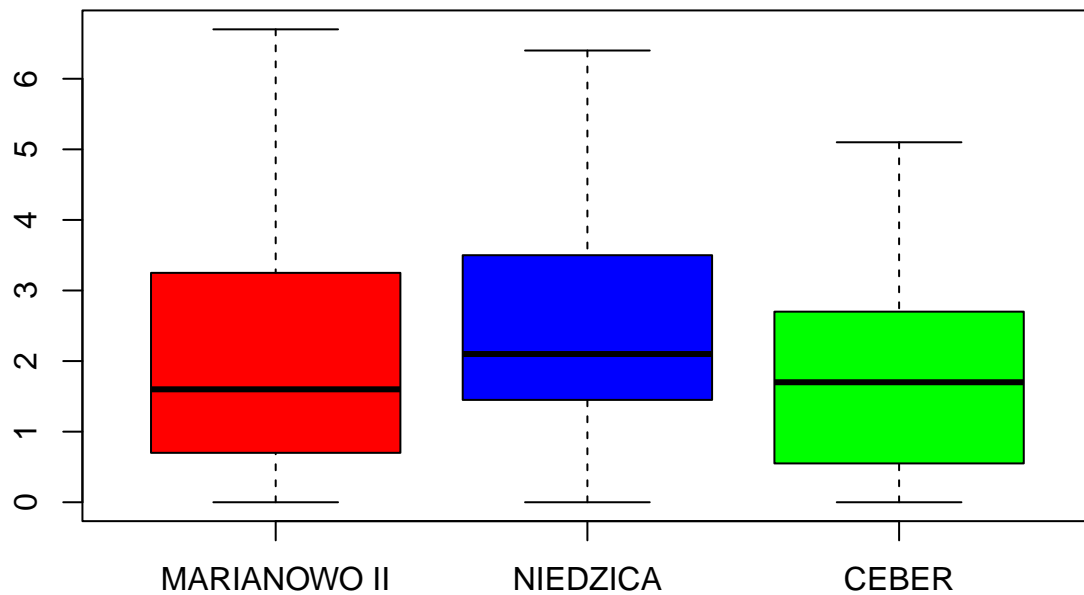
iii. Wartości bezwzględne zmian maksymalnej temperatury względem dnia poprzedniego w wybranych miastach:

```
get_daily_diffs <- function(data) {
  dds <- data %>% mutate(prev = lag(data[, 'maks_temp'], n=1),
                        diff = abs(data[, 'maks_temp'] - prev))
  dds[1, 'diff'] = 0
  return (dds[, 'diff'])
}

daily_diffs1 <- get_daily_diffs(data1)
daily_diffs2 <- get_daily_diffs(data2)
daily_diffs3 <- get_daily_diffs(data3)
```

Wykresy pudełkowe zmian maksymalnej temperatury względem dnia poprzedniego dla wybranych miast:

```
boxplot(daily_diffs1, daily_diffs2, daily_diffs3, names = cities,
        col = c('red', 'blue', 'green'))
```



Mediany zmian maksymalnej temperatury względem dnia poprzedniego dla wybranych miast:

```
median1 <- median(daily_diffs1)
median2 <- median(daily_diffs2)
median3 <- median(daily_diffs3)

cat(cities[1], ': ', median1, '\n')
```

```
## MARIANOWO II : 1.6
```

```
cat(cities[2], ': ', median2, '\n')
```

```
## NIEDZICA : 2.1
```

```
cat(cities[3], ': ', median3, '\n')
```

```
## CEBER : 1.7
```

**Komentarz:** Kolejność miast według zmian maksymalnej temperatury względem dnia poprzedniego to: Niedzica, Ceber, Marianowo II.

Średnie zmian maksymalnej temperatury względem dnia poprzedniego dla wybranych miast:



```
mean1 <- mean(daily_diffs1)
mean2 <- mean(daily_diffs2)
mean3 <- mean(daily_diffs3)
```

```
cat(cities[1], ': ', mean1, '\n')
```

```
## MARIANOWO II : 2.070968
```

```
cat(cities[2], ': ', mean2, '\n')
```

```
## NIEDZICA : 2.445161
```

```
cat(cities[3], ': ', mean3, '\n')
```

```
## CEBER : 1.916129
```

**Komentarz:** Kolejność miast według średniej arytmetycznej dobowej amplitudy temperatur to: Niedzica, Marianowo II, Ceber.

**Odpowiedź:** Miastem o największych wahanach maksymalnej dobowej temperatury z dnia na dzień jest Niedzica. Dla obydwu statystyk (mediana i średnia arytmetyczna) otrzymała najwyższe wartości. Z pozostałych miast Marianowo ma wyższą średnią, natomiast Ceber wyższą medianę. Stąd trudno jest powiedzieć, które miasto miało większe wahania maksymalnej temperatury z dnia na dzień.