

# Rattle - początki pracy

## Ćw. 1

### Instalacja Rattle

Ustawiamy serwer lustrzany CRAN przez polecenie menu **Pakiety/Ustaw serwer lustrzany CRAN**

Zazwyczaj do instalacji typowego pakietu w systemie Windows wystarczy wybranie z menu polecenia **Pakiety/Zainstaluj pakiet** lub za pomocą funkcji `install.packages()`:

```
install.packages("rattle")
```

W trakcie instalacji lub po niej mogą zostać zainstalowane pakiety będące zależnościami Rattle (np. Gtk+).

## Ćw. 2

### Start Rattle

Aby wystartować Rattle używamy polecenia:

```
library(rattle)
rattle()
```

Różne zakładki Rattle mogą wymagać do pracy dodatkowych pakietów, więc warto podczas pracy korzystać z połączenia z Internetem.

Aby wyjść z Rattle po prostu wciskamy przycisk Zakończ (Quit).

## Ćw. 3

### Wczytanie przykładowego zbioru danych weather

1. Klikamy na przycisk Wykonaj (Execute). Rattle zauważy brak danych i zaproponuje nam wykorzystanie wbudowanych danych.
2. Wybieramy Tak (Yes). Zbiór danych weather (pogoda), to prosty zbiór danych, który pozwoli nam przetestować zasady ED.
3. Dane są wczytane do programu.
4. W zakładce Data widzimy listę zmiennych i ich typy.

## Ćw. 4

### Proste podsumowania tekstowe

```
summary(weather[7:9])
```

Sunshine		WindGustDir		WindGustSpeed	
Min.	: 0.000	NW	: 73	Min.	:13.00
1st Qu.:	5.950	NNW	: 44	1st Qu.:	31.00
Median	: 8.600	E	: 37	Median	:39.00
Mean	: 7.909	WNW	: 35	Mean	:39.84
3rd Qu.:	10.500	ENE	: 30	3rd Qu.:	46.00
Max.	:13.600	(Other)	:144	Max.	:98.00
NA's	:3	NA's	: 3	NA's	:2

Wykonaj to ćwiczenie również w Rattle.

## Ćw. 5

## Podsumowywanie za pomocą Hmisc

Jeśli pakiet niedostępny, instalujemy go.

## Podsumowanie zmiennej Sunshine

```
library(Hmisc)
describe(weather[7])
weather[7]
```

1	Variables	366	Observations							
-----										
Sunshine										
	n missing	unique	Mean	.05	.10	.25	.50	.75	.90	
363	3	114	7.909	0.60	2.04	5.95	8.60	10.50	11.80	
	.95									
	12.60									
lowest : 0.0 0.1 0.2 0.3 0.4, highest: 13.1 13.2 13.3 13.5 13.6										

## Podsumowanie zmiennej WindGustDir

```
describe(weather[8])
weather[8]
```

1	Variables	366	Observations
-----			
WindGustDir			
	n missing	unique	
	363	3	16
	N	NNE	NE
	ENE	E	ESE
	SE	SSE	S
	SSW	SW	WSW
	W	WNW	NW
	NNW		
Frequency	21	8	16
	30	37	23
	12	12	22
	5	3	2
	20	35	73
	44		
%	6	2	4
	8	10	6
	3	3	6
	1	1	1
	6	10	20
	12		

Wykonaj to ćwiczenie również w Rattle.

## Ćw. 6

### Podsumowania numeryczne za pomocą fBasics

Jeśli pakiet nie jest zainstalowany, należy go zainstalować. Podsumowanie dla zmiennej Sunshine:

```
library(fBasics)
basicStats(weather$Sunshine)

X..weather.Sunshine
nobs          366.000000
NAs           3.000000
Minimum       0.000000
Maximum       13.600000
1. Quartile   5.950000
3. Quartile   10.500000
Mean          7.909366
Median        8.600000
Sum           2871.100000
SE Mean       0.182732
LCL Mean      7.550016
UCL Mean      8.268716
Variance      12.120962
Stdev         3.481517
Skewness      -0.723454
Kurtosis      -0.270625
```

### Skośność dla kilku zmiennych

```
skewness(weather[,c(7,9,12,13)], na.rm=TRUE)
```

```
      Sunshine WindGustSpeed WindSpeed9am WindSpeed3pm
-0.7234543    0.8361055    1.3601713    0.5912721
```

### Kurtoza dla powyższych zmiennych:

```
kurtosis(weather[,c(7,9,12,13)], na.rm=TRUE)
      Sunshine WindGustSpeed WindSpeed9am WindSpeed3pm
-0.2706248    1.4761027    1.4758254    0.1963276
```

Wykonaj to ćwiczenie również w Rattle.

## Ćw. 7

### Wykres słupkowy.

W Rattle, wykonaj wykres słupkowy (Bar plot) dla zmiennej WindGustDir z zestawu weather.

## Ćw. 8

### Wykres kropkowy.

W Rattle, wykonaj wykres kropkowy (Dot plot) dla zmiennej WindGustDir z zestawu weather.

## Ćw. 9

### Wykres mozaikowy.

W Rattle, wykonaj wykres mozaikowy (Mosaic) dla zmiennej WindGustDir z zestawu weather.

## Ćw. 10

### Wykres pudełkowy - praca domowa

Pierwszym krokiem jest wygenerowanie danych do wykresu. Następujący przykład tworzy jeden zestaw danych z dwoma kolumnami, jedna jest obserwacją z Humidity3pm i druga, określona przez zmienną o nazwie grp - grupa, do której obserwacja ta należy. Istnieją trzy grupy, dwa odpowiadające dwu wartościom zmiennej docelowej i trzecia obejmująca wszystkie obserwacje.

Stosujemy with(), co pozwala na odwoływanie do zmiennych w oryginalnym zbiorze danych bez konieczności podawania nazwy zestawu danych za każdym razem. Łączymy trzy obiekty data.frame (), stosując rbind (), aby wygenerować końcowy zestaw danych:

```
ds <- with(crs$dataset[crs$train,], rbind(data.frame(dat=Humidity3pm, grp="All"),
  data.frame(dat=Humidity3pm[RainTomorrow=="No"], grp="No"),
  data.frame(dat=Humidity3pm[RainTomorrow=="Yes"], grp="Yes")))
```

Teraz wyświetlamy wykres za pomocą funkcji boxplot() grupując dane dat wg zmiennej grp:

```
bp <- boxplot(formula=dat ~ grp, data=ds,
  col=rainbow_hcl(3),
  xlab="RainTomorrow", ylab="Humidity3pm",
  notch=TRUE)
```

Będziemy również oznaczać średnią na wykresie. Do tego wykorzystamy summaryBy() z pakietu doBy. Resztę załatwia użycie points() i pch=8:

```
library(doBy)
points(x=1:3, y=summaryBy(formula=dat ~ grp, data=ds,
  FUN=mean, na.rm=TRUE)$dat.mean, pch=8)
```

Następnie dodajemy dodatkowe opisy tekstowe identyfikujące medianę, zasięg międzykwartylowy:

```
for (i in seq(ncol(bp$stats)))
{
  text(x=i, y=bp$stats[,i] -
    0.02*(max(ds$dat, na.rm=TRUE) -
      min(ds$dat, na.rm=TRUE)),
    labels=bp$stats[,i])
}
```

Opisujemy również punkty oddalone za pomocą text() (zmniejszając trochę font):

```
text(x=bp$group+0.1, y=bp$out, labels=bp$out, cex=0.6)
```

Ostatecznie dodajemy tytuł i podtytuł do naszego wykresu (umieszczamy na nim datę i godzinę oraz nazwę użytkownika systemu):

```
title(main="Rozkład Humidity3pm",
  sub=paste("Rattle",
```

```
format(Sys.time(), "%Y-%b-%d %H:%M:%S"),  
Sys.info()["user"])
```

**Przyślij wykres pudełkowy dla Evaporation.**

## Ćw. 11

### Wykres pudełkowo-percentylowy z pakietu hmisc - praca domowa

Odmianą wykresu pudełkowego jest wykres pudełkowo-percentylowy. Wykres dostarcza więcej informacji o rozkładzie wartości, niż wykres pudełkowy. Jest on generowany przy użyciu `bpplot()` z pakietu `Hmisc`.

```
library(Hmisc)  
h3 <- weather$Humidity3pm  
hn <- h3[weather$RainTomorrow=="No"]  
hy <- h3[weather$RainTomorrow=="Yes"]  
ds <- list(h3, hn, hy)  
bpplot(ds, name=c("All", "No", "Yes"),  
       ylab="Humidity3pm", xlab="RainTomorrow")
```

Szerokość każdej ramki zależy od liczby obserwacji dla danego zakresu. Na wykresie widzimy również medianę i 25ty i 75ty percentyl.

**Przyślij wykres pudełkowo-percentylowy dla Sunshine.**