1)ÚKOL

1. **Základní statistika**

* Průměr 3. Sloupce (z) je:

mean(data1[,3])

[1] 5466.692 1

* Minimum 3. Sloupce (z) je:

min(data1[,3])

[1] 1053.606 2

* Maximum 3. Sloupce (z) je:

max(data1[,3])

[1] 25210.14 3

* Variace 3. Sloupce (z) je:

var(data1[,3])

[1] 10527006 4

* Lze udělat summary, kde se vypíše základní statistika ze 3. Sloupce

summary(data1[,3])

Min. 1st Qu. Median Mean 3rd Qu. Max.

1054 3342 4647 5467 6685 25210 6

Lze vidět, min, max, 1. Kvartil, medián, průměr, 3. Kvartil a maximum.

* Směrodatná odchylka 3. Sloupce:

sd(data1[,3])

[1] 3244.535 5

- je to to nejdůležitější číslo ve statistice

* Pro šikmost a špičatost musím zapnout knihovny e1071

library(e1071)

* Šikmost

skewness(data1[,3]) 7

[1] 1.97692

Pokud by byla šikmost 0, tak by nám značila rovnoměrné rozdělení vlevo i vpravo. Kladná šikmost nám značí, že vpravo od průměru se vyskytují odlehlejší hodnoty než vlevo od průměru.

* Špičatost

>kurtosis(data1[,3])

[1] 6.172273 8

Špičatost porovnané rozdělení s normálním rozdělením pravděpodobnosti. Normální rozdělení má špičatost 0.

Kladná špičatost značí, že většina hodnot náhodné veličiny leží blízko její [střední hodnoty](http://cs.wikipedia.org/wiki/St%C5%99edn%C3%AD_hodnota) a hlavní vliv na rozptyl mají málo pravděpodobné odlehlé hodnoty.

* medián

>median(data1[,3])

[1] 4647.324 9

Medián je prostřední hodnota v datech1, ve sloupci z

* IQR

Iqr je důležitá součást numerických charakteristik –mír variabilit

IQR(data1[,3])

[1] 3342.254 navíc

* Rozsah mezi nejvyšší a nejnižší hodnotou

>min <- min(data1[,3])

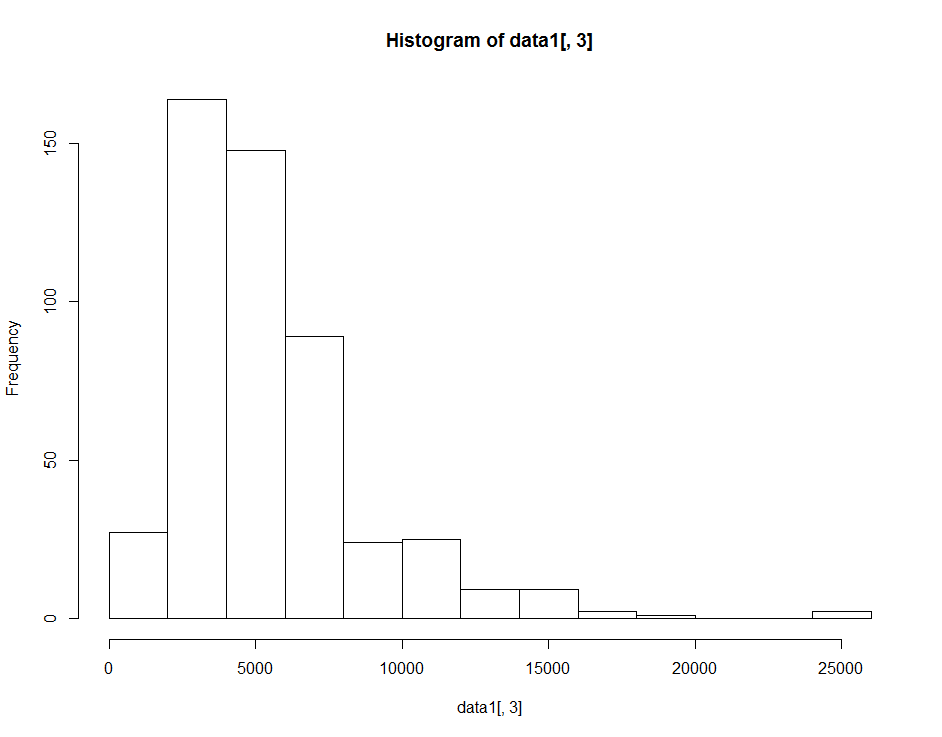
>max<- max(data1[,3])

>max-min

[1] 24156.54 10

* Histogram

hist(data1[,3]) 11



Histogram znázorňuje výskyt dat v intervalech, lze vidět největší výskyt okolo hodnoty 4000.

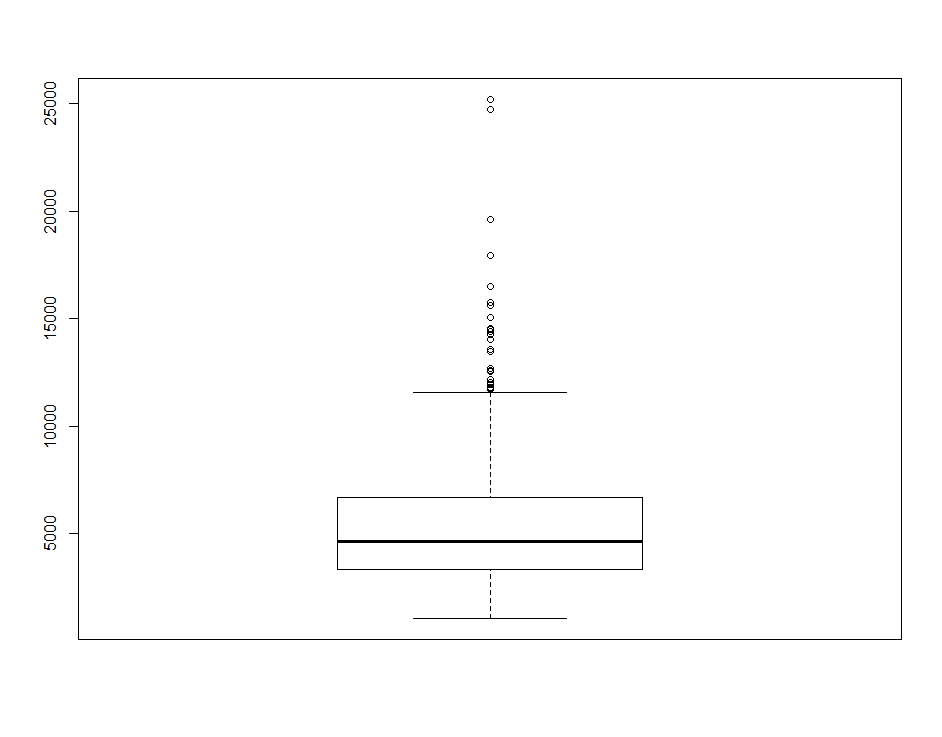
* Boxplot – whiskers plot

boxplot(data1[,3]) 12

Boxplot- znázorňuje nejnižší hodnotu spodní úsečkou a nejvyšší hodnotu a graficky vizualizuje numerické data pomocí kvartilů, prostřední černá úsečka nám vykresluje střední hodnotu – medián.

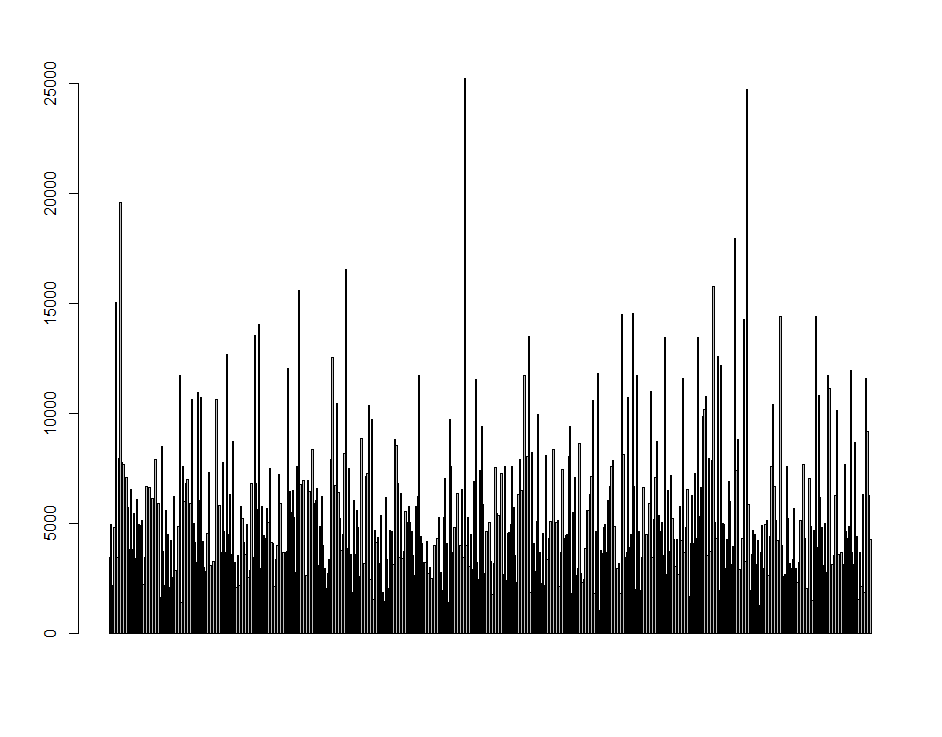
Spodní čára obdélníku je první kvartil a horní čára obdélníku je 3. Kvartil

Kolečka nad boxplotem nám znázorňují další hodnoty, jelikož jsou ve velkém rozmezí od sebe, min a max hodnoty.



* Barplot

barplot(data1[,3]) 13

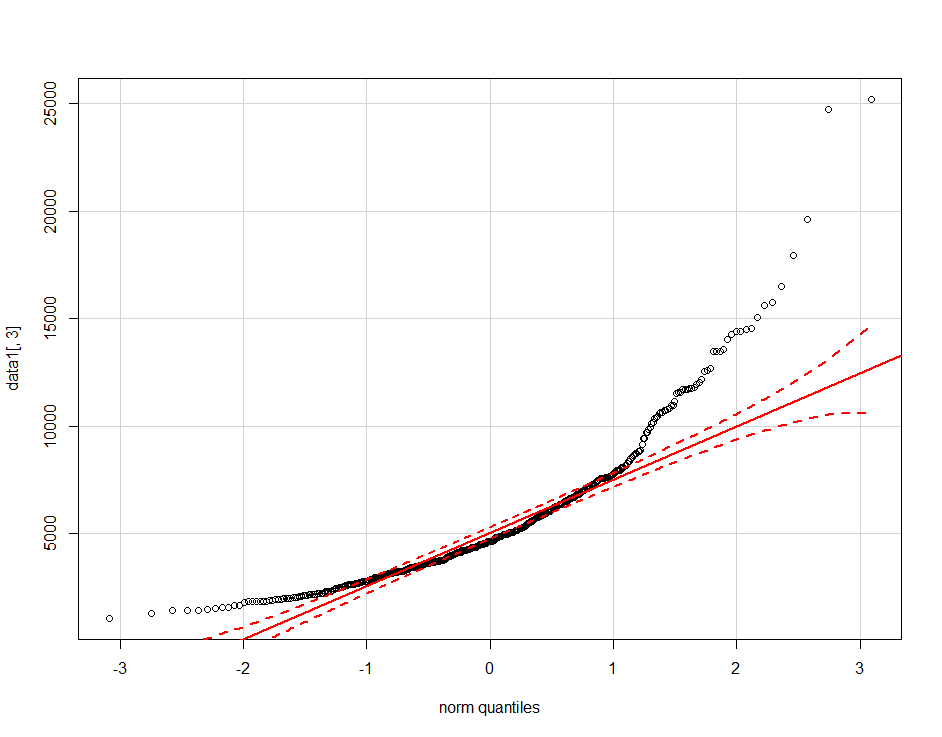


znázorňuje nám velikosti jednotlivých entit

* Qqplot 14

zapnutí knihovny car – bez toho nelze vytvořit qq plot

qqPlot(data1[,3]) - srovnání teoretických hodnot s teoretickým odhadem hodnot vybraného rozdělení.



2)ÚKOL

shapiro.test(data1[,3]) 15

Shapiro-Wilk normality test 16

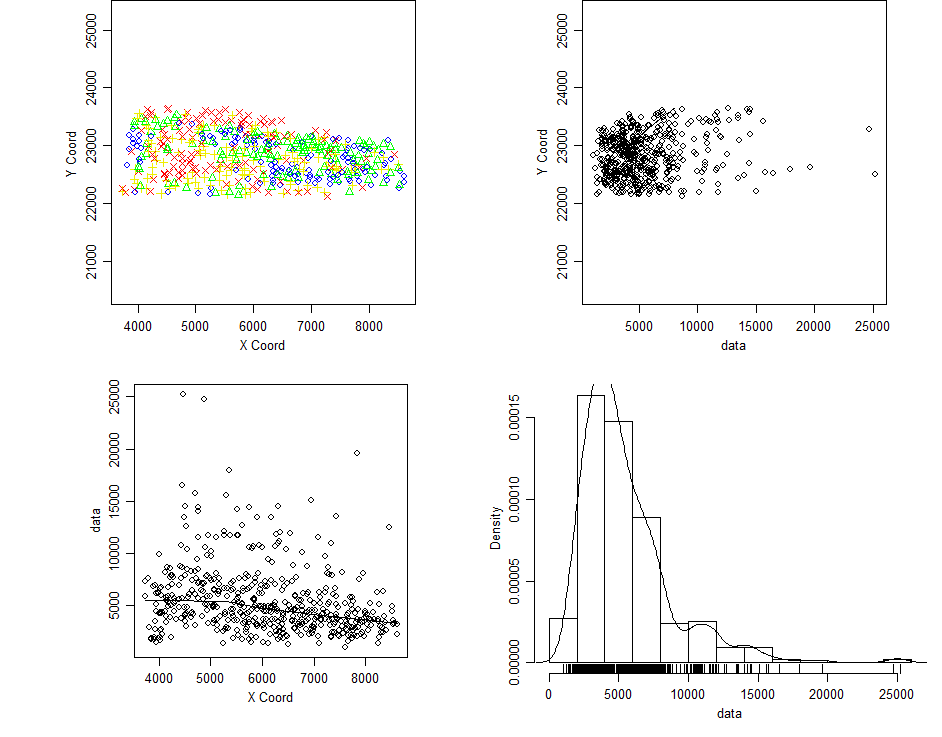
data: data1[, 3] 17

W = 0.8415, p-value < 2.2e-16 18

geodata<- as.geodata(data1, coords.col=1:2, data1.col=3)

> plot(geodata, lowess = T)

23



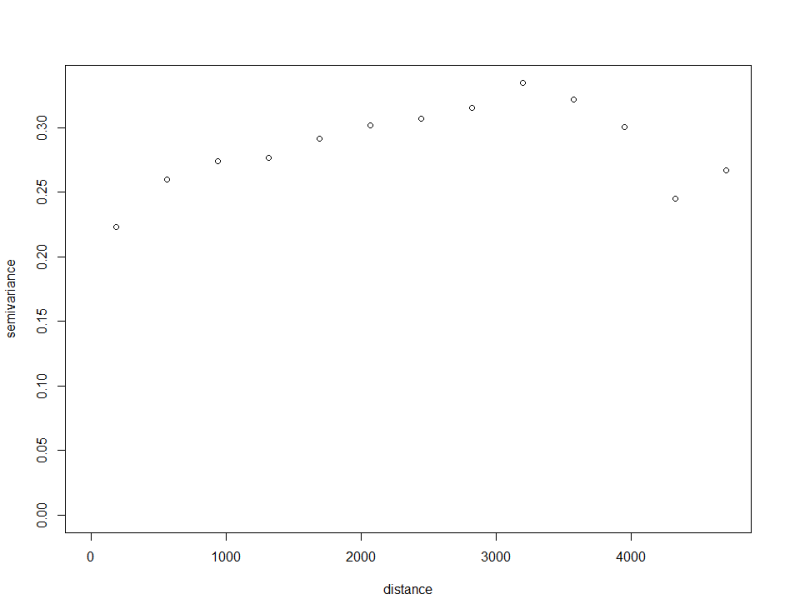
data1[,4] <- log(data1[,3]) 19

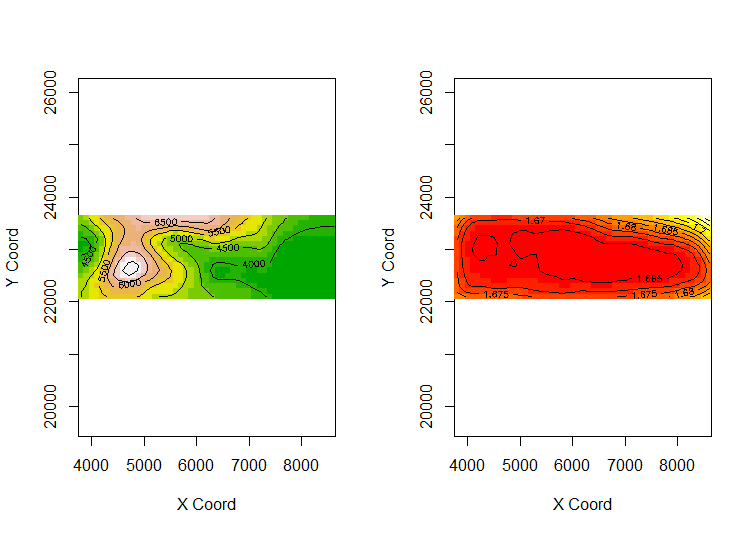
geodata<- as.geodata(data1, coords.col=1:2, data.col=4) 20

var<-variog(geodata) 21

plot(var) 22

24



25

vario.fit<- variofit(var, nugget = 0.25, max.dist = 3200, fix.nugget = TRUE) 26

lines(vario.fit,col="red") 27

summary(geodata$coords) 28

loci <- expand.grid(seq(3800,8600,b=100),seq(22100,23600,b=100)) 29

kc<- krige.conv(geodata,loc=loci,krige=krige.control(obj.model=vario.fit)) 30

par(mfrow = c(1,2)) - toto je aby se ti ty dva grafy nakreslili vedle sebe 31

image(kc, value = exp(kc$predict), col = terrain.colors(12)) 32

(kc,value = exp(kc$predict), nlev = 10, add=T) 33

image(kc, value = exp(sqrt(kc$krige.var))) 34

contour(kc,value = exp(sqrt(kc$krige.var)), add=T) 35

3)ÚKOL

> data1[,4] <- 0 36  
> a <- quantile(X, probs = c(0.7))   
> median(Y)   
[1] 19.2  
> b <- median(Y)

> a <- quantile(X, probs = c(0.7)) rozdělení osy x na 7 decil 37

> b <- median(Y) rozdělení osy Y podle mediánu 38

> data1[,4]<- ifelse(x<a[1]&z>b[1], 1, ifelse(x < a[1]&y<b[1], 1, 2)) 39

> data1[,4] <- as.factor(data1[,4]) 40

> anova <- aov(z~data1[,4], data = data1) 41

> plot(TukeyHSD(anova)) 42

4)ÚKOL

> attach(data2) 43

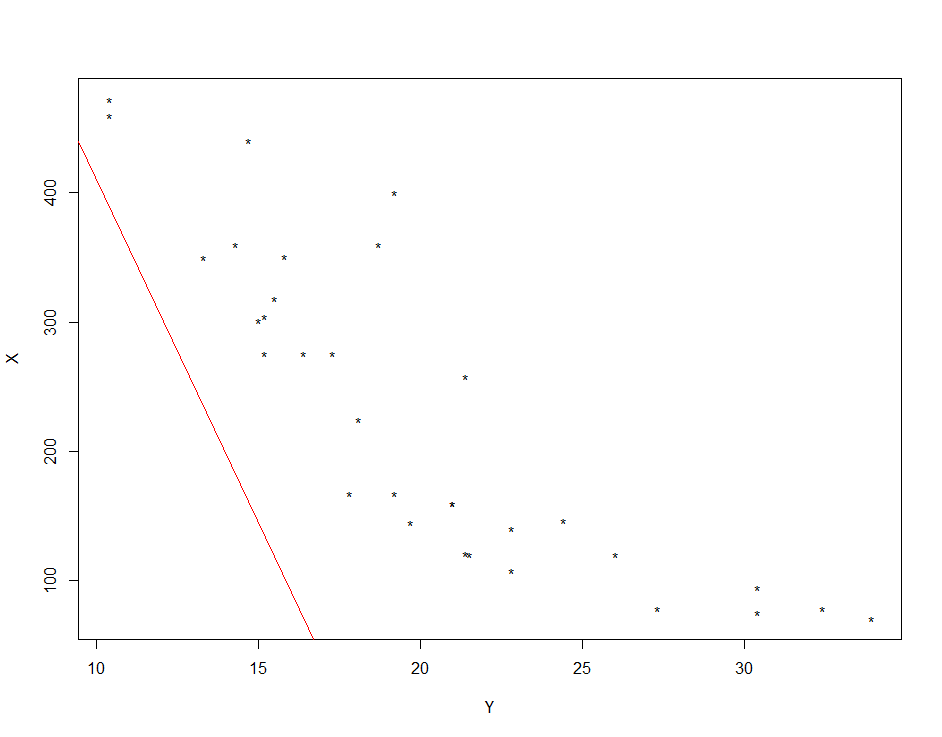
> cor(data2[,1], data2[,2]) 44

[1] -0.8475514 45

> kvm<- lm(X ~ Y + I(Y^2), data = data2)

> plot(X ~ Y, pch = '\*')

> abline(kvm, col = 'red')



navic

Regresní rovnice

X ~ Y + I(Y^2)