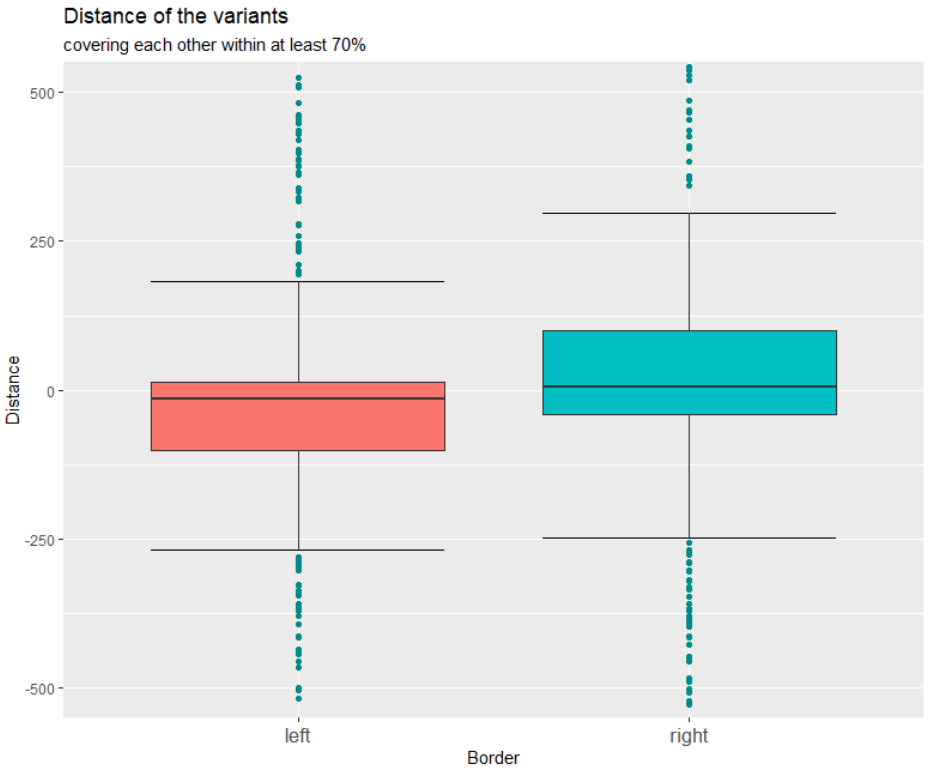
Marcel Buczkowski 145159

Bioconductor, raport 2

Zad 1.

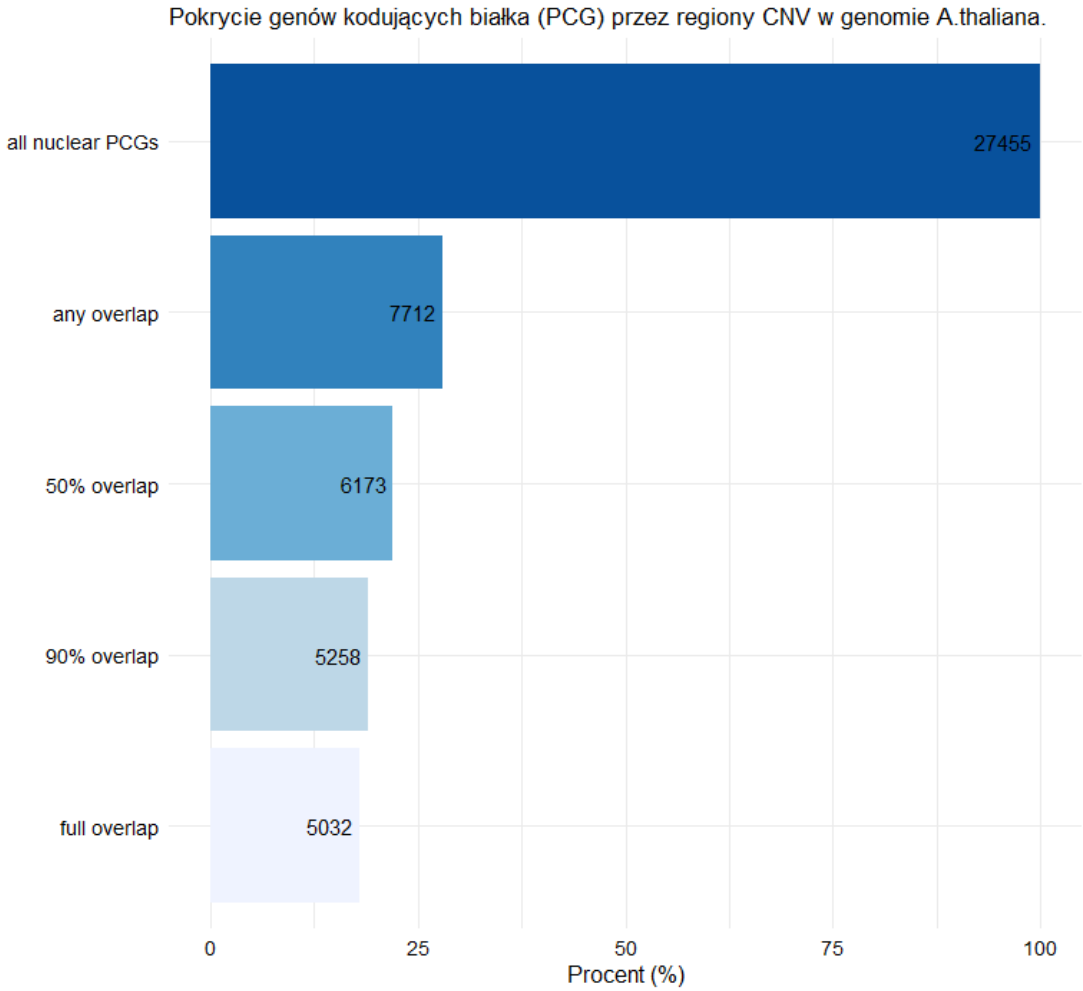
```r
library(ggplot2)
library(dplyr)
library(RColorBrewer)


###ZAD1###
data <- read.table("C:\\Users\\marce\\hakowanie\\programming_R\\bioconductor\\dane1.txt", header = TRUE, sep = "\t")
head(data)

ggplot(data, aes(x=Border, y = Distance, fill=Border)) + geom_boxplot() +
  stat_boxplot(geom = 'errorbar')+
  coord_cartesian(ylim = c(-500, 500))+
  theme(legend.position = "none", axis.text.x = element_text(size = 13)) +
  labs(title = "Distance of the variants", subtitle = 'covering each other within at least 70%') +
  geom_boxplot(outlier.color = "cyan4", outlier.shape = 16)

```



Marcel Buczkowski 145159

Zad 2

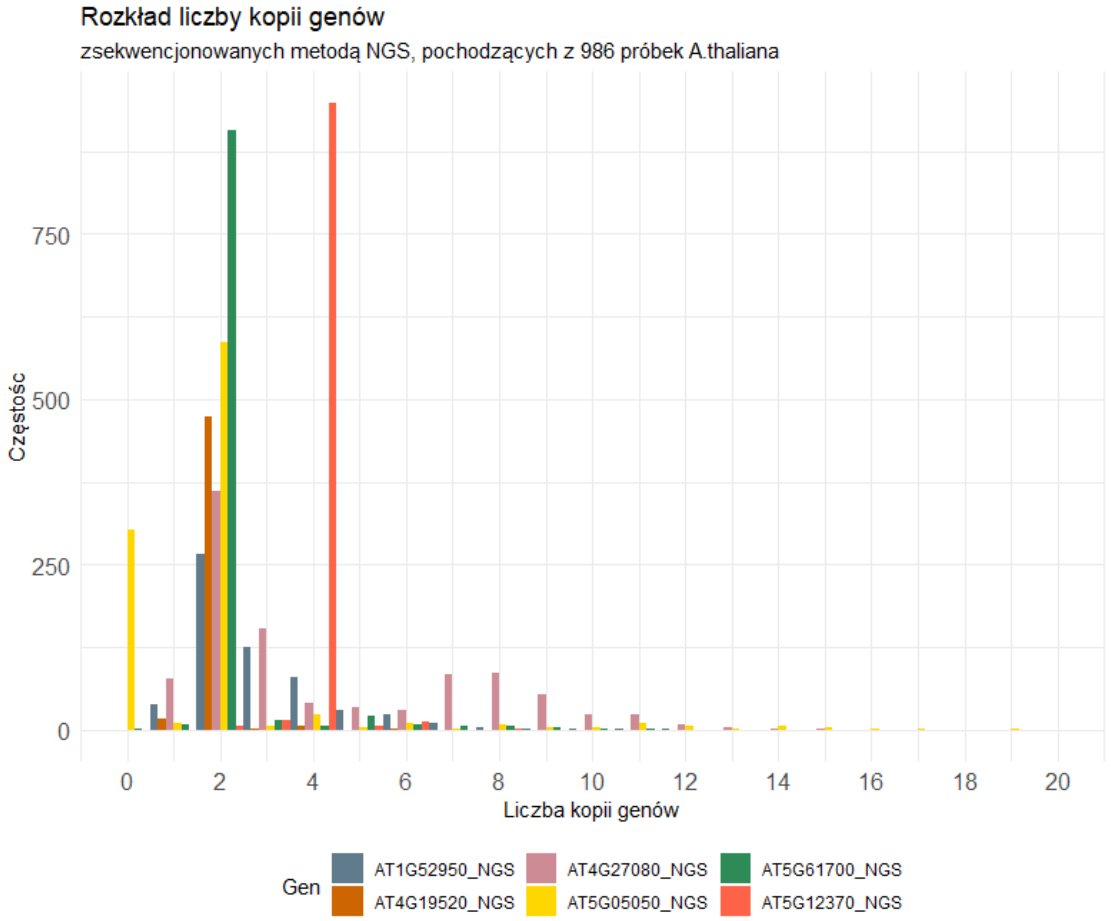
```
###ZAD2####
data_2 <- read.table("C:\\Users\\marce\\hakowanie\\programming_R\\bioconductor\\dane2.txt", header = TRUE, sep = "\t")
data_2$value <- as.numeric(data_2$value)
data_2$variable <- reorder(data_2$variable, data_2$value)
head(data_2)


ggplot(data_2, aes(x=variable, y=Procent, fill=variable)) + geom_bar(stat = "identity") +
  coord_flip() +
  ylab("Procent (%)") +
  xlab("") +
  theme_minimal() +
  labs(title = "Pokrycie genów kodujących białka (PCG) przez regiony CNV w genomie A.thaliana.") +
  theme(plot.title = element_text(size = 13), axis.text = element_text(size = 10.5), legend.position = "none") +
  theme(axis.text = element_text(color = "black"), axis.title.x = element_text(color = "black", size=12)) +
  scale_fill_brewer(palette = "Blues") + geom_text(aes(label = value), hjust = +1.15)
```

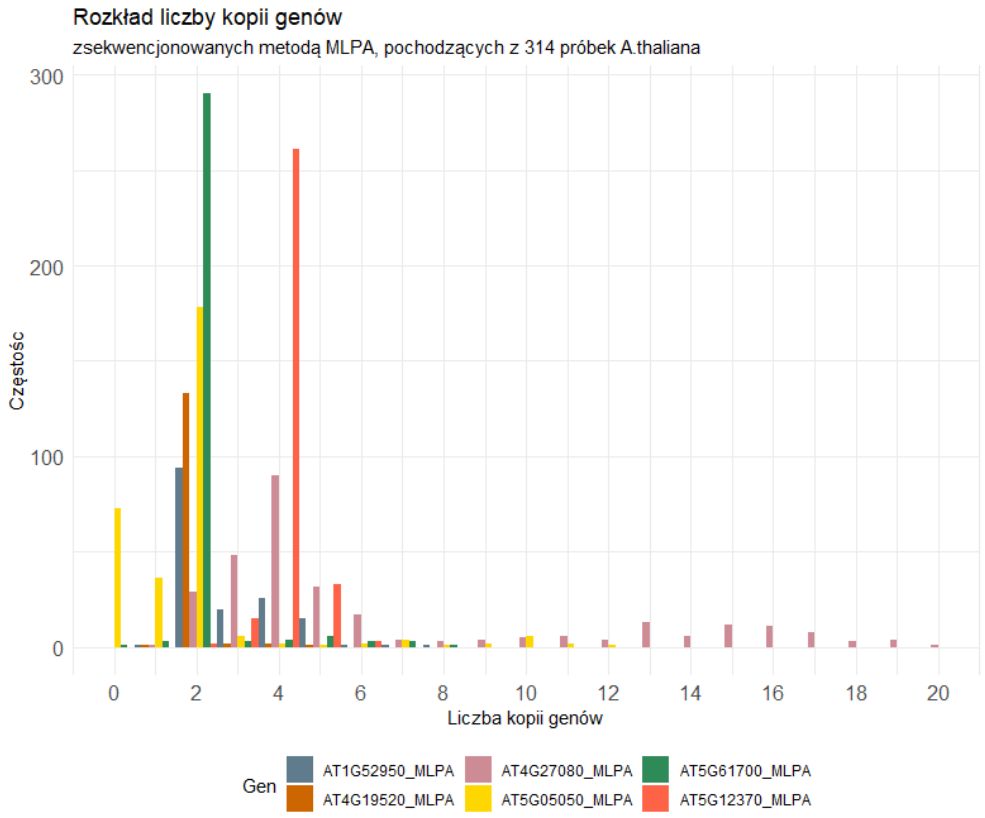


Marcel Buczkowski 145159

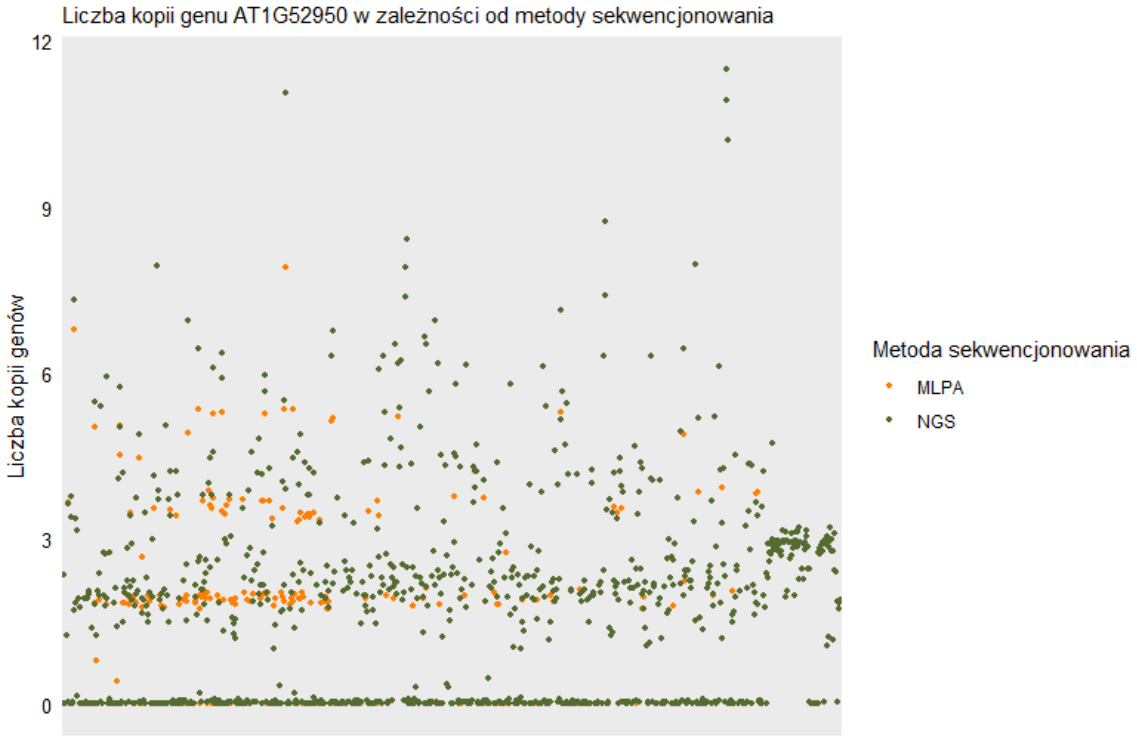
Zad 3

```
###ZAD3###
data_3 <- read.table("C:\\Users\\marce\\hakowanie\\programming_R\\bioconductor\\dane3.txt", header = TRUE, sep = "\t")

data_3_ngs <- data_3 %>% select(matches("NGS"))
data_3_mlpa <- data_3 %>% select(matches("MLPA"))

#pojedyncza kolumna
#ggplot(data_3_ngs, aes(x=data_3_ngs$AT1G52950_NGS)) + geom_histogram()


#wszytskie kolumny NGS
library(reshape2)
data_long <- melt(data_3_ngs)

ggplot(data_long, aes(x = value, fill = variable)) +
  geom_histogram(binwidth = 1, position = "dodge") +
  theme_minimal() +
  labs(title = "Rozkład liczby kopii genów", x = "Liczba kopii genów", y = "Częstośc",  subtitle = 'zsekwencjonowanych metodą NGS
  scale_fill_manual(values = c("lightskyblue4","darkorange3", "lightpink3", 'gold1', 'seagreen', 'tomato1')) +
  labs(fill = "Gen") +
  theme(axis.text = element_text(size = 12))+
  scale_x_continuous(limits = c(0,20), breaks = seq(0, 20, by = 2))+
  theme(legend.position = "bottom")


#wszytskie kolumny MLPA
data_long <- melt(data_3_mlpa)
ggplot(data_long, aes(x = value, fill = variable)) +
  geom_histogram(binwidth = 1, position = "dodge") +
  theme_minimal() +
  labs(title = "Rozkład liczby kopii genów", x = "Liczba kopii genów", y = "Częstośc",  subtitle = 'zsekwencjonowanych metodą MLP
  scale_fill_manual(values = c("lightskyblue4","darkorange3", "lightpink3", 'gold1', 'seagreen', 'tomato1')) +
  labs(fill = "Gen") +
  theme(axis.text = element_text(size = 12))+
  scale_x_continuous(limits = c(0,20), breaks = seq(0, 20, by = 2))+
  theme(legend.position = "bottom")
```






```r
#punktowy
colnames(data_3)
ggplot() +
geom_point(data = data_3, aes(x = Sample, y = AT1G52950_MLPA, color = "MLPA"), size = 1) +
geom_point(data = data_3, aes(x = Sample, y = AT1G52950_NGS, color = "NGS"), size = 1) +
theme_minimal() +
labs(title = "Liczba kopii genu AT1G52950 w zależności od metody sekwencjonowania",y = 'Liczba kopii genów', x = '') +
theme(axis.text = element_text(size = 10, color = 'black'), plot.title = element_text(size = 11)) +
scale_color_manual(values = c("MLPA" = "darkorange1", "NGS" = "darkolivegreen"), name = "Metoda sekwencjonowania")
```



```r
#boxplot z nałożonym scatterplotem
data_3_ngs <- data_3 %>% select(matches("NGS"))
data_long <- melt(data_3_ngs)
data_long['variable'] = 'NGS'
head(data_long)


data_3_mlpa <- data_3 %>% select(matches("MLPA"))
data_long_2 <- melt(data_3_mlpa)
data_long_2['variable'] = 'MLPA'
head(data_long_2)

combined_df <- rbind(data_long, data_long_2)


ggplot(combined_df, aes(x=variable, y = value,  fill=variable)) + geom_boxplot() +
  scale_fill_manual(values = alpha(c("blue", "blue"), alpha = 0.1)) +
  geom_point(position = position_jitter(width = 0.3), size = 1, color = "seagreen") +
  theme(legend.position = "none", axis.text.x = element_text(size = 13)) +
  labs(title = "Sekwencjonowanie wybranych genów A. Thaliana", subtitle = 'w zależności od wybranej metody', y = 'Liczba kopii g
  geom_boxplot(outlier.color = "darkslateblue", outlier.shape = 16)
```