Clinvar mutations in clients of hsp

Load the data

```
clinvar_path <- read.delim('../../body/1raw/clinvar_patho_missense.tsv', sep = ' ')
clinvar_ben <- read.delim('../../body/1raw/clinvar_benign_missense.tsv', sep = ' ')
clients_hsp90 <- read.delim('../../body/2derived/clients.ids.list.txt', header = F)
colnames(clients_hsp90) <- 'Gene'
nonclients_hsp90 <- read.delim('../../body/2derived/nonclients.ids.list.txt', header = F)
colnames(nonclients_hsp90) <- 'Gene'
```

Merge clinvar and clients information

```
clinvar_ben$hsp90_client <- ifelse(clinvar_ben$Gene %in% clients_hsp90$Gene, 1,0)
clinvar_path$hsp90_client <- ifelse(clinvar_path$Gene %in% clients_hsp90$Gene, 1,0)

clinvar_ben$Pathogenic <- 0
clinvar_path$Pathogenic <- 1

clinvar_clients <- rbind(clinvar_ben, clinvar_path)
clinvar_clients <- clinvar_clients[,c(8,12,16,17)]
```

Compare clients with all others genes - Fisher

```
clinvar_clients$client_text <- ifelse(clinvar_clients$hsp90_client, 'hsp90 clients', 'other genes')
clinvar_clients$Mutation_type_text <- ifelse(clinvar_clients$Pathogenic, 'pathogenic', 'benign')

pdf('../../body/4figures/Clinvar.mut.hsp.clients.vs.all.genes.mozaicplot.pdf')
mosaicplot(table(clinvar_clients$client_text, clinvar_clients$Mutation_type_text), ylab = 'Mutations',
           main = '', color = 'cyan3', cex.axis = 1.1)


knitr::kable(t(table(clinvar_clients$client_text, clinvar_clients$Mutation_type_text)))
```

	hsp90 clients	other genes
benign	636	45327
pathogenic	2507	41281

```
ft <- fisher.test(t(table(clinvar_clients$client_text, clinvar_clients$Mutation_type_text)))
print(ft)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  t(table(clinvar_clients$client_text, clinvar_clients$Mutation_type_text))
## p-value < 2.2e-16
```

```
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.2112255 0.2524291
## sample estimates:
## odds ratio
##  0.2310238
```

```
dev.off()
```

```
## pdf
##   2
```

Compare clients with control nonclients genes - Fisher

```
clinvar_control_noncl <- clinvar_clients[(clinvar_clients$Gene %in% clients_hsp90$Gene) | (clinvar_clien
clinvar_control_noncl$client_text <- ifelse(clinvar_control_noncl$hsp90_client, 'client', 'nonclient')
clinvar_control_noncl$Mutation_type_text <- ifelse(clinvar_control_noncl$Pathogenic, 'pathogenic', 'beni

pdf('../../body/4figures/Clinvar.mut.hsp.clients.vs.nonclients.mozaicplot.pdf')
mosaicplot(table(clinvar_control_noncl$client_text, clinvar_control_noncl$Mutation_type_text), ylab = 'M
           main = '', color = 'cyan3', cex.axis = 1.1)



knitr::kable(t(table(clinvar_control_noncl$client_text, clinvar_control_noncl$Mutation_type_text)))
```

	client	nonclient
benign	636	300
pathogenic	2507	288

```
ft <- fisher.test(t(table(clinvar_control_noncl$client_text, clinvar_control_noncl$Mutation_type_text)))
print(ft)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  t(table(clinvar_control_noncl$client_text, clinvar_control_noncl$Mutation_type_text))
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.2018740 0.2939869
## sample estimates:
## odds ratio
##  0.2436566
```

```
dev.off()
```

```
## pdf
##   2
```

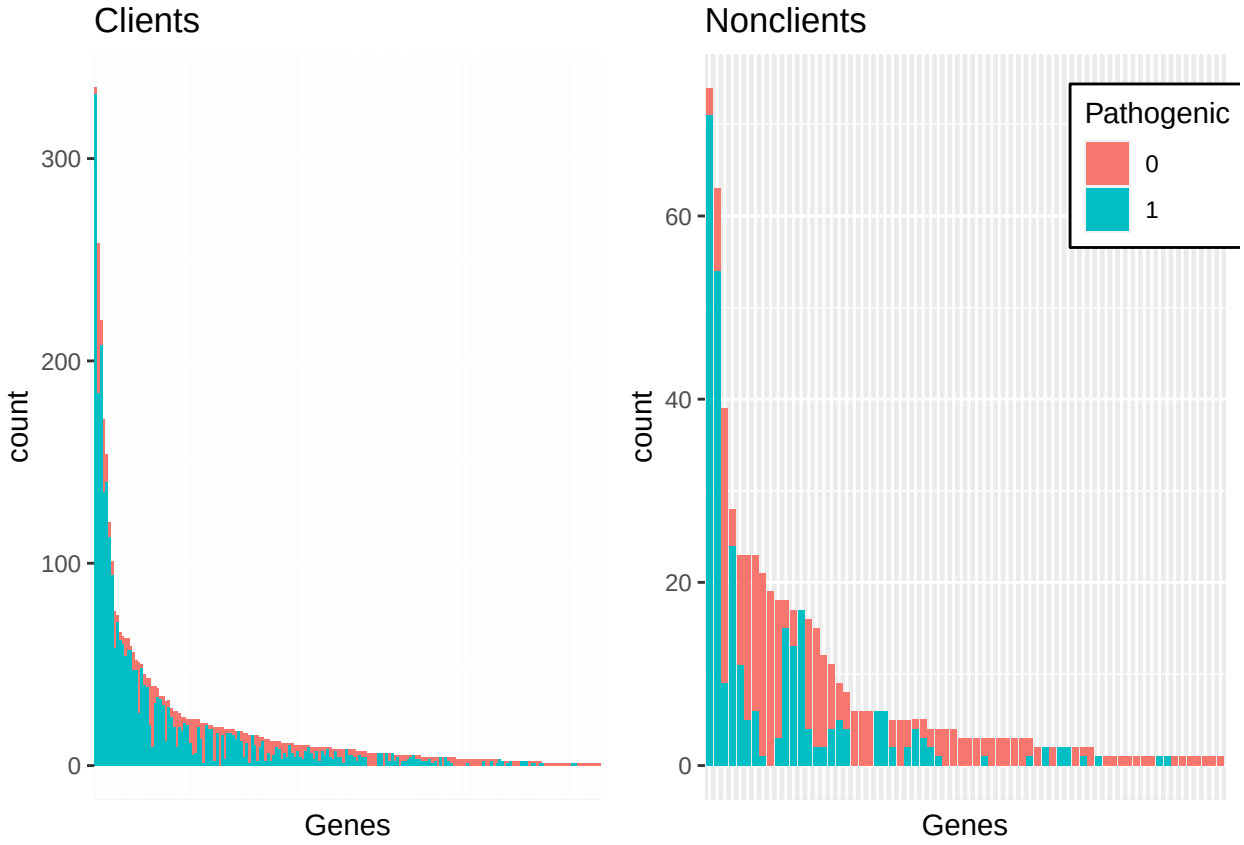
How many htere are motations in one gene?

```
require(gridExtra)
```

```
p1 <- ggplot(clinvar_control_noncl[,], aes(reorder(Gene, Gene, function(x)-length(x))))+
  geom_bar(aes(fill = as.factor(Pathogenic)))+
  #theme_bw()+
  xlab('Genes')+
  theme(axis.text.x=element_blank(), axis.ticks.x = element_blank(),
        legend.position = 'None')+
  ggtitle('Clients')

p2 <- ggplot(clinvar_control_noncl[clinvar_control_noncl$hsp90_client == 0,], aes(reorder(Gene, Gene, fu
  geom_bar(aes(fill = as.factor(Pathogenic)))+
  #theme_bw()+
  xlab('Genes')+
  theme(axis.text.x=element_blank(), axis.ticks.x = element_blank(), legend.position = c(0.87,0.85), leg
  ggtitle('Nonclients')+
  scale_fill_discrete(name = "Pathogenic")


grid.arrange(p1, p2, ncol=2)
```




```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## v purrr   0.3.4
```

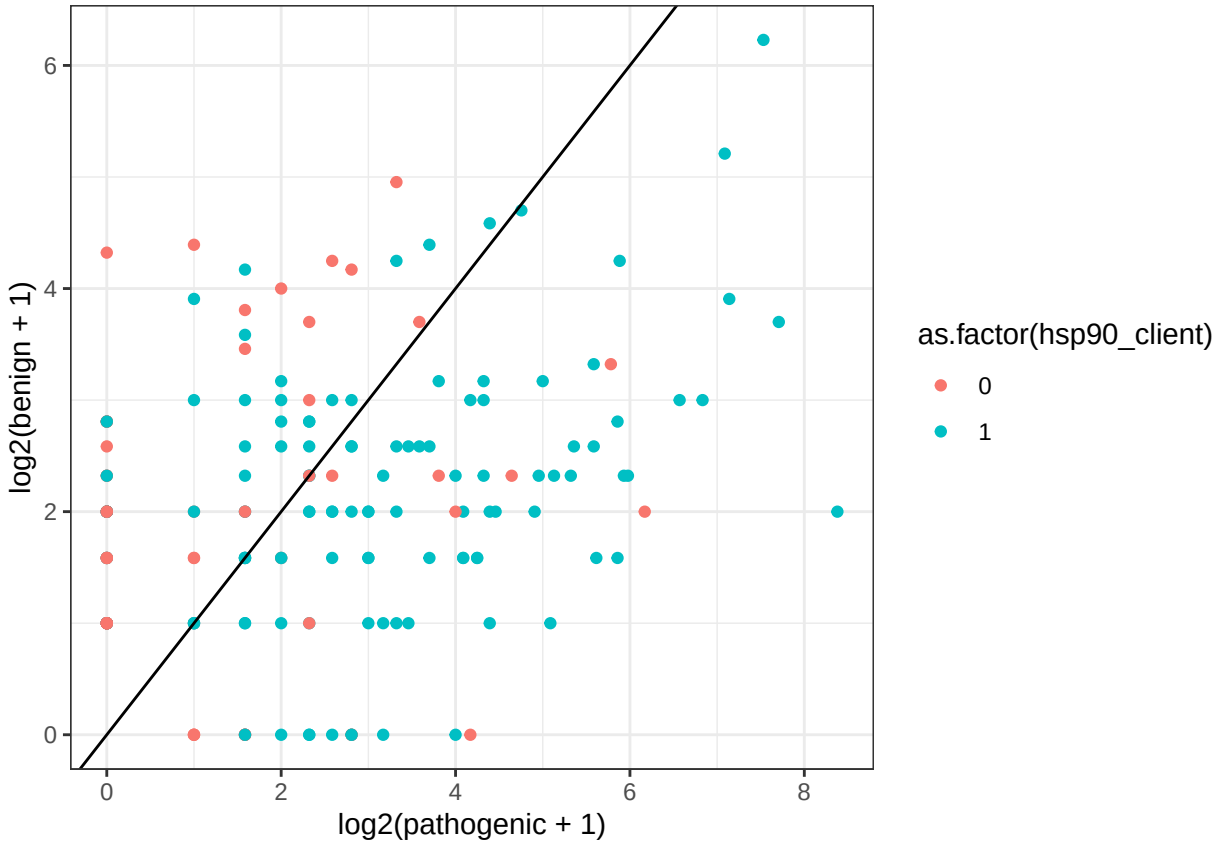
```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::combine() masks gridExtra::combine()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
count_path <- clinvar_control_noncl %>%
  count(Gene, Pathogenic)

count_path <- count_path %>%
  pivot_wider(names_from = 'Pathogenic', values_from = 'n')

count_path[is.na(count_path)] <- 0
colnames(count_path) <- c('Gene', 'pathogenic', 'benign')
count_path <- merge(count_path, clinvar_control_noncl[,c(1,3)], by = 'Gene')
count_path <- count_path[!duplicated(count_path),]

ggplot(count_path, aes(log2(pathogenic+1), log2(benign+1), color = as.factor(hsp90_client)))+
  geom_point()+
  geom_abline(intercept = 0, slope = 1)+
  theme_bw()
```



Look if the number of mutations and proportion of pathogenic per gene is different between clients and nonclients

```
clinvar_summary <- aggregate(clinvar_control_noncl$Pathogenic, list(clinvar_control_noncl$Gene), FUN = n
clinvar_summary <- merge(clinvar_summary, aggregate(clinvar_control_noncl$Pathogenic, list(clinvar_conti
clinvar_summary <- merge(clinvar_summary, aggregate(clinvar_control_noncl$Pathogenic, list(clinvar_conti
```

```
colnames(clinvar_summary) <- c('Gene', 'proportion_of_pat', 'number_of_mut', 'number_of_pat' )
clinvar_summary <- merge(clinvar_summary, clinvar_control_noncl[,c(1,5)][!duplicated(clinvar_control_no
clinvar_summary$num_of_ben <- clinvar_summary$number_of_mut - clinvar_summary$number_of_pat



p1 <- ggplot(clinvar_summary, aes(y = number_of_mut, x = client_text, fill = client_text))+
  geom_boxplot(outlier.shape = NA)+
  theme_bw()+
  theme(legend.position = 'None')+
  ylim(0,30)+
  ylab('Number of mutations per gene')+xlab('')+
  theme(axis.text = element_text(size=25), axis.title = element_text(size=22))

p2 <- ggplot(clinvar_summary, aes(y = number_of_pat, x = client_text, fill = client_text))+
  geom_boxplot(outlier.shape = NA)+
  theme_bw()+
  theme(legend.position = 'None')+
  ylim(0,22)+
  ylab('Number of pathogenic mutations per gene')+xlab('')+
  theme(axis.text = element_text(size=25), axis.title = element_text(size=22))

p3 <- ggplot(clinvar_summary, aes(y = num_of_ben, x = client_text, fill = client_text))+
  geom_boxplot(outlier.shape = NA)+
  theme_bw()+
  theme(legend.position = 'None')+
  ylim(0,10)+
  ylab('Number of benign mutations per gene')+xlab('')+
  theme(axis.text = element_text(size=25), axis.title = element_text(size=22))

p4 <- ggplot(clinvar_summary, aes(y = proportion_of_pat, x = client_text, fill = client_text))+
  geom_boxplot()+
  theme_bw()+
  theme(legend.position = 'None')+
  ylim(0,1)+
  ylab('Proportion of pathogenic mutations per gene')+xlab('')+
  theme(axis.text = element_text(size=25), axis.title = element_text(size=22))

pp <- plot_grid(p1,p2,p3,p4, labels = c('A','B','C','D'), ncol =2, nrow=2, label_size = 36)
```

```
## Warning: Removed 28 rows containing non-finite values (stat_boxplot).
```

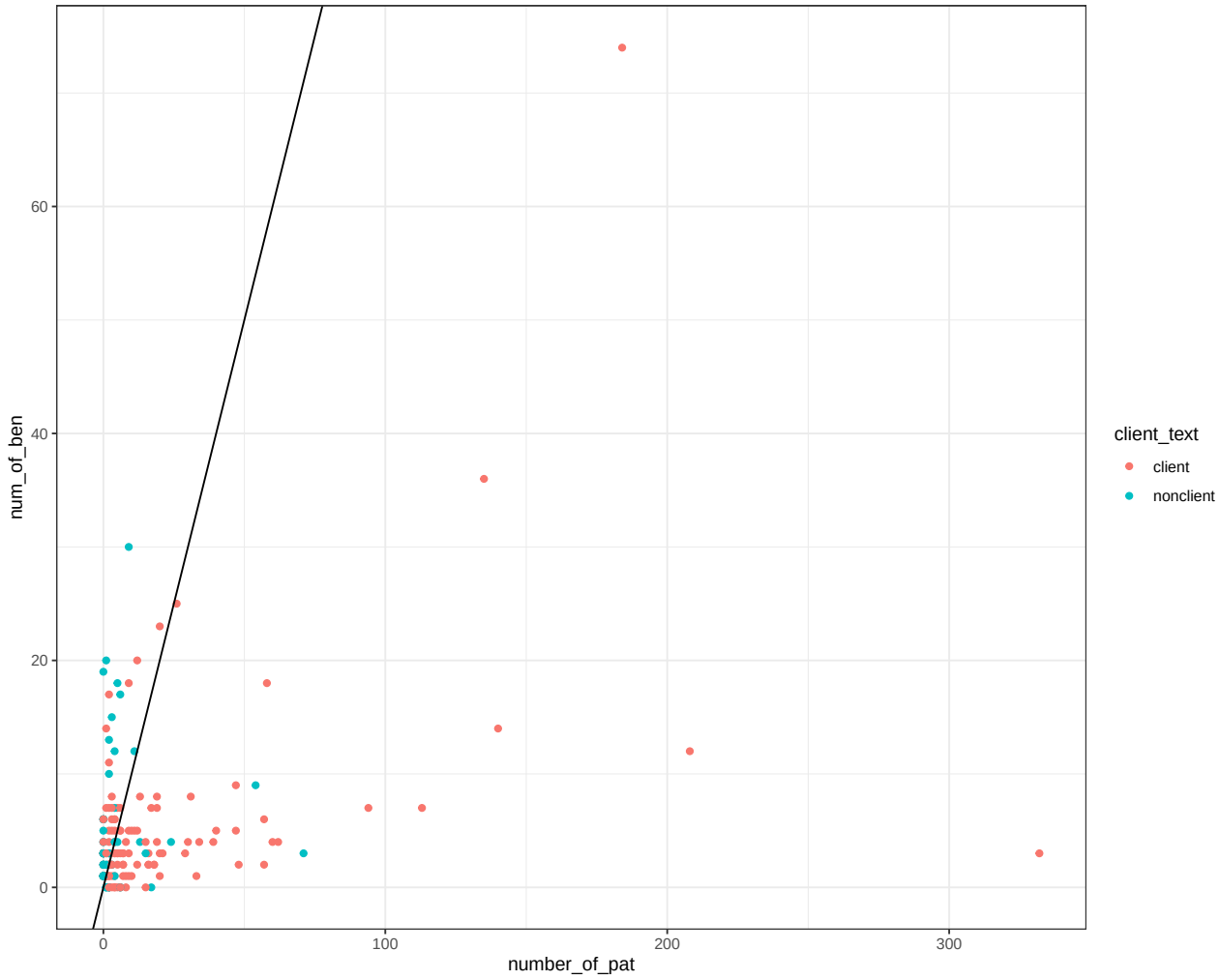
```
## Warning: Removed 26 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 21 rows containing non-finite values (stat_boxplot).
```

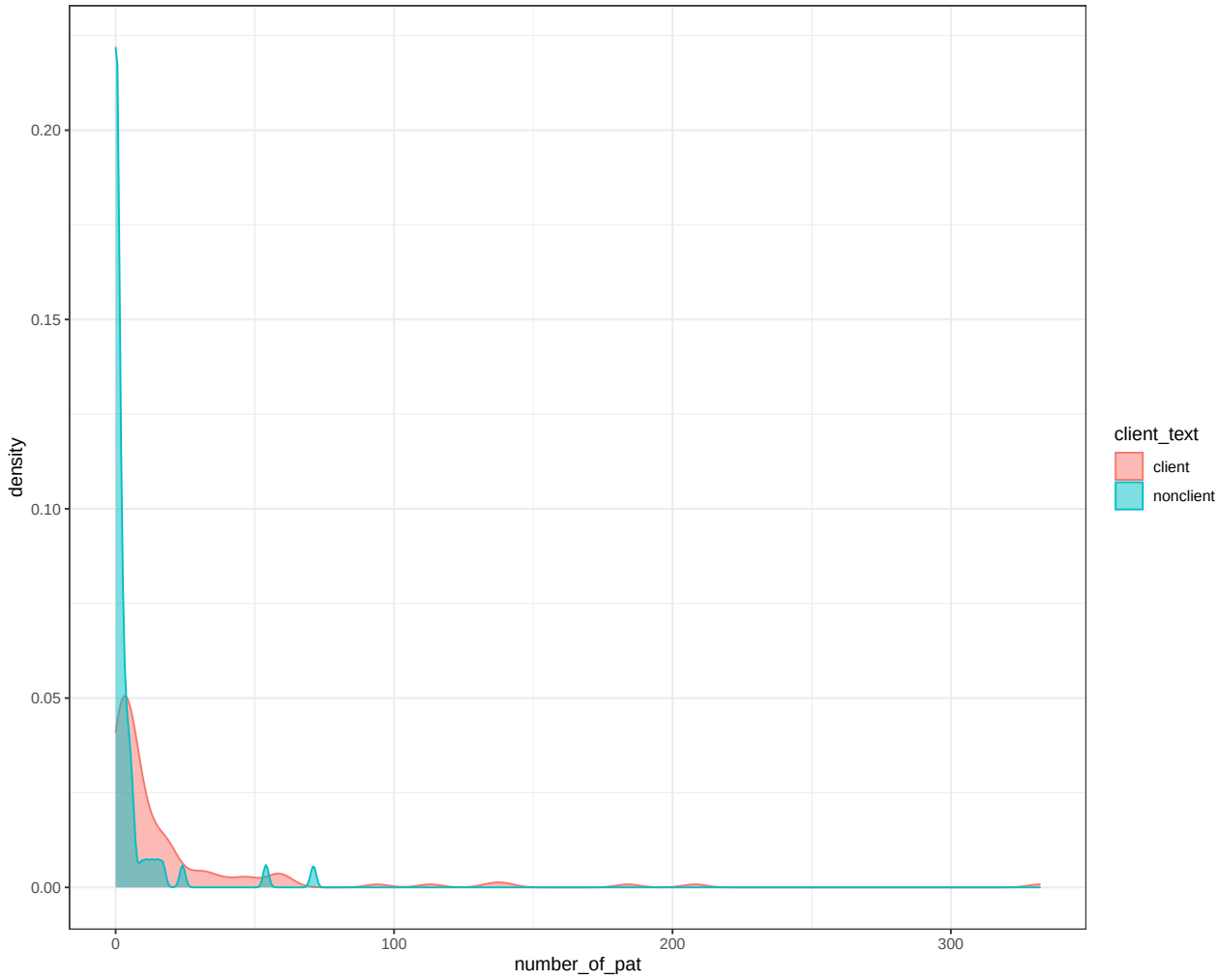
```
ggsave(pp, filename = '../../body/4figures/ClinVer.number.of.mut.per.gene.clients.vs.nonclients.pdf')
```

```
## Saving 10 x 8 in image
```

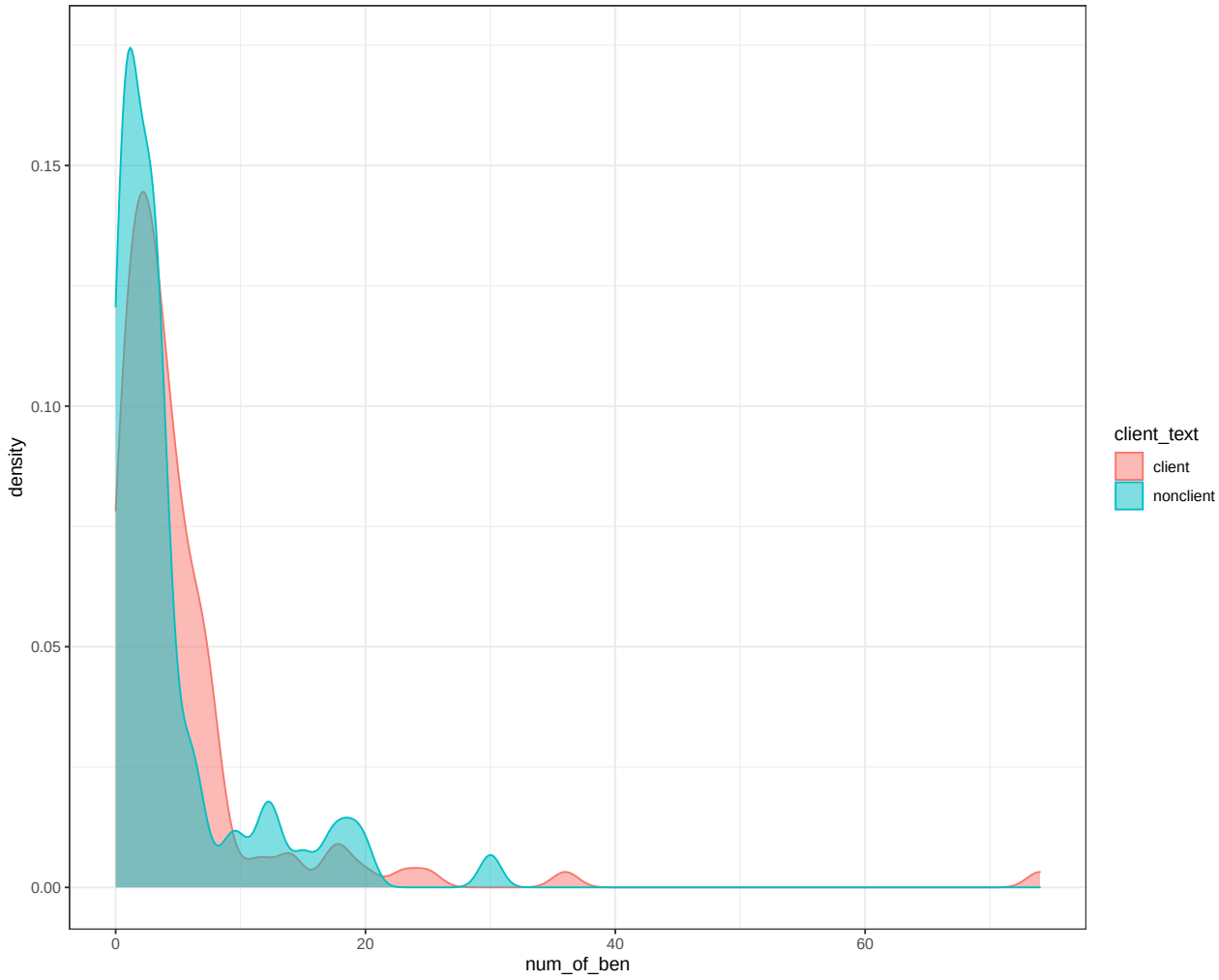
```
ggplot(clinvar_summary, aes(x = number_of_pat, y = num_of_ben, color = client_text))+
  geom_point()+
  geom_abline(intercept = 0, slope = 1)+
  theme_bw()
```




```
ggplot(clinvar_summary, aes(x = number_of_pat, fill = client_text, color = client_text))+
  geom_density(alpha = 0.5)+
  theme_bw()
```




```
ggplot(clinvar_summary, aes(x = num_of_ben, fill = client_text, color = client_text))+
  geom_density(alpha = 0.5)+
  theme_bw()
```


0.15
0.10
0.05
0.00
density
0
20
40
60
num_of_ben
client_text
client
nonclient