```r
library(xlsx)
```

```
## Loading required package: rJava
```

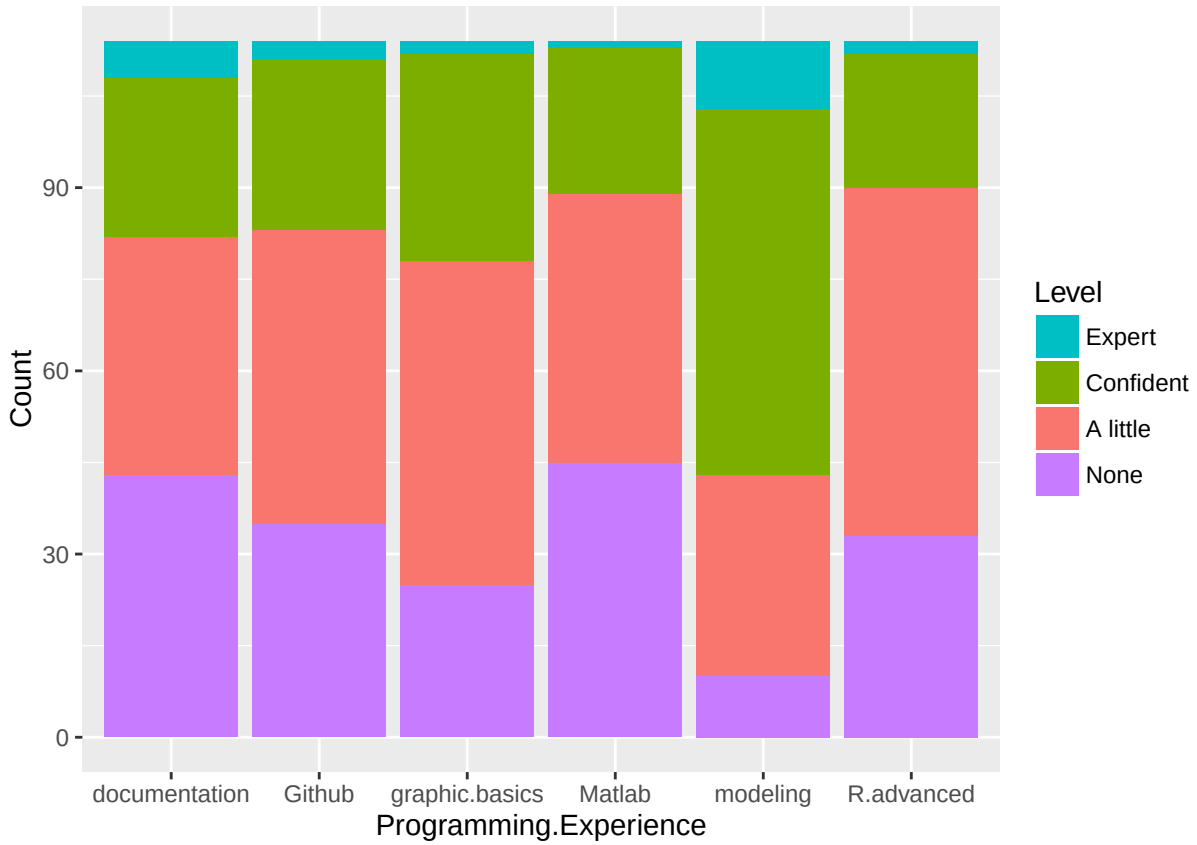
```
## Warning: package 'rJava' was built under R version 3.2.3
```

```
## Loading required package: xlsxjars
```

```r
data <- read.xlsx("data_clean.xlsx", 1)
library(reshape2)
program_data <- subset(data, select = c(2,6:11))
combined_program_data = melt(program_data, id.vars = "Program", variable.name = "Programming Experience
stat1 <- table(program_data$R..graphic.basics)
stat2 <- table(program_data$R..advanced)
stat3 <- table(program_data$Reproducible.documentation.with.R)
stat4 <- table(program_data$Matlab..data.manipulation..analysis..visualization.and.modeling)
stat5 <- table(program_data$Github)
stat6 <- table(program_data$R..data.manipulation.and.modeling)
library(plyr)
plot_data <- read.xlsx("plot_revised.xlsx", 1)
plot_data1 <- ddply(plot_data, "Programming.Experience", transform, label_y=cumsum(Count))
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

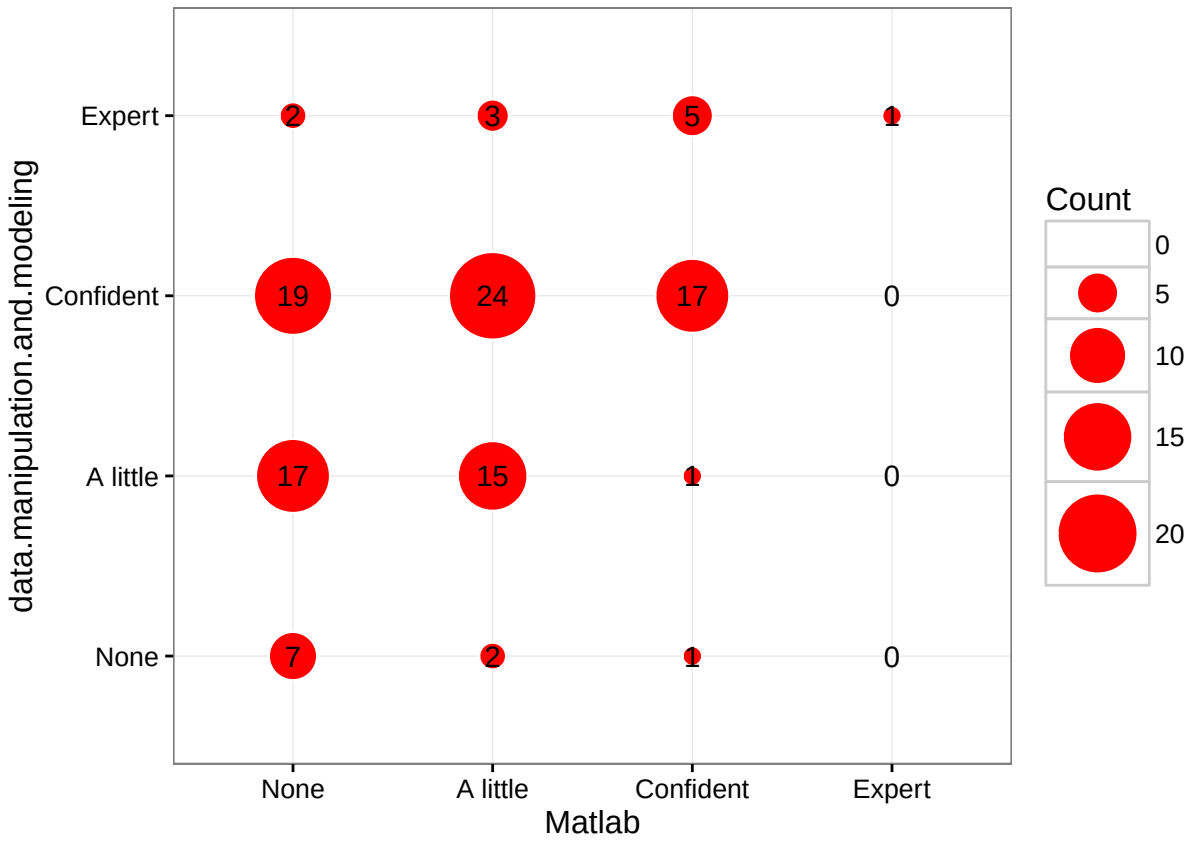
```r
bp <- ggplot(plot_data, aes(x=Programming.Experience, y=Count, fill = Level)) +
  geom_bar(stat = "identity") +
  guides(fill=guide_legend(reverse=TRUE))
bp + scale_fill_discrete(breaks=c("None","A little","Confident", "Expert"))
```



This plot showes the distribution of programming level among different skills Used R code to find the statistical data which is needed and then used ggplot to draw the graph.

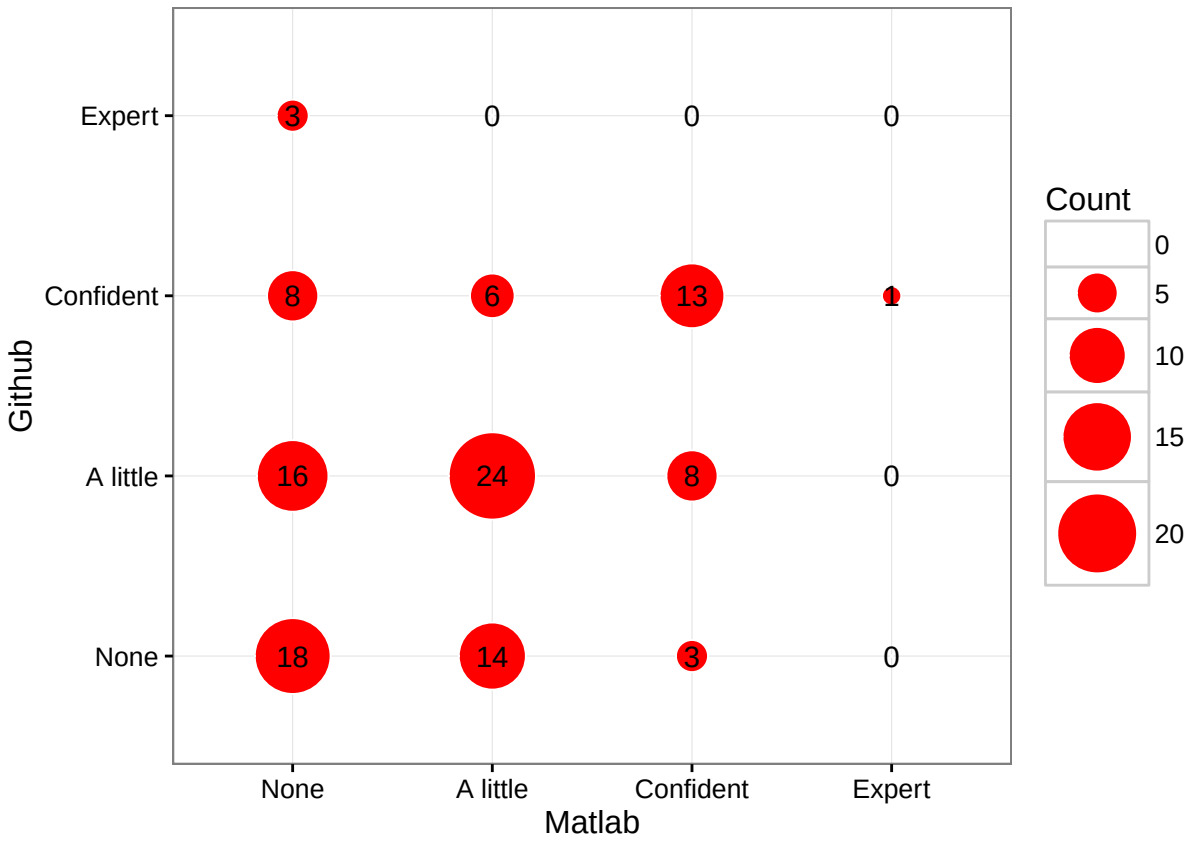
```
for(i in 1:dim(program_data)[2]){program_data[,i] = as.character(program_data[,i])}
programming_level <- c("None","A little","Confident", "Expert")
r_matlab_count = c()

for (level1 in programming_level){
  for (level2 in programming_level){
    count = program_data$Program[program_data$Matlab..data.manipulation..analysis..visualization.and.mo
    cnt = length(count)
    r_matlab_count = append(r_matlab_count,cnt)
  }
}
plot_data2 <- read.xlsx("Matlab&R_data_mani.xlsx", 1)
bp1 <- ggplot(plot_data2, aes(x=Matlab, y=data.manipulation.and.modeling, size=Count, label = Count),gu
  geom_point(colour="white", fill="red", shape=21)+ scale_size_area(max_size = 15)+
  geom_text(size=4)+
  theme_bw()
bp1 + scale_x_discrete(limits=c("None","A little","Confident", "Expert")) + scale_y_discrete(limits=c("
```

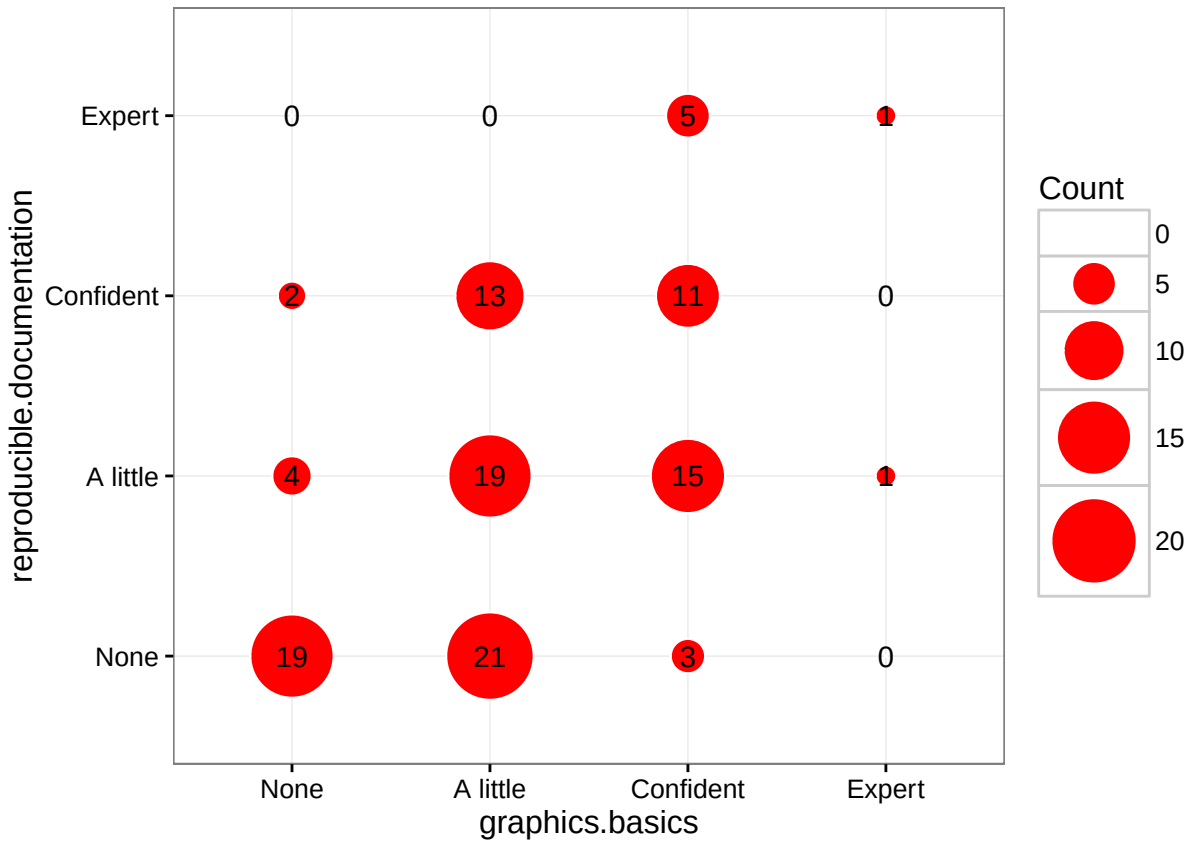


This graph showes the relationship between the skills of Matlab and data manipulation and modelling

```
matlab_github_count = c()
for (level1 in programming_level){
  for (level2 in programming_level){
    count = program_data$Program[program_data$Matlab..data.manipulation..analysis..visualization.and.mo
    cnt = length(count)
    matlab_github_count = append(matlab_github_count,cnt)
  }
}
plot_data3 <- read.xlsx("Matlab&Github.xlsx", 1)
bp1 <- ggplot(plot_data3, aes(x=Matlab, y=Github, size=Count, label = Count),guide=FALSE)+
  geom_point(colour="white", fill="red", shape=21)+ scale_size_area(max_size = 15)+
  geom_text(size=4)+
  theme_bw()
bp1 + scale_x_discrete(limits=c("None","A little","Confident", "Expert")) + scale_y_discrete(limits=c("
```




```
graphics_documentation_count = c()
for (level1 in programming_level){
  for (level2 in programming_level){
    count = program_data$Program[program_data$R..graphic.basics==level1 & program_data$Reproducible.docu
    cnt = length(count)
    graphics_documentation_count = append(graphics_documentation_count,cnt)
  }
}
plot_data4 <- read.xlsx("graphics&docu.xlsx", 1)
bp1 <- ggplot(plot_data4, aes(x=graphics.basics, y=reproducible.documentation, size=Count, label = Coun
  geom_point(colour="white", fill="red", shape=21)+ scale_size_area(max_size = 15)+
  geom_text(size=4)+
  theme_bw()
bp1 + scale_x_discrete(limits=c("None","A little","Confident", "Expert")) + scale_y_discrete(limits=c("
```


reproducible.documentation
Expert
Confident
A little
None
0
0
5
1
2
13
11
0
4
19
15
1
19
21
3
0
None
A little
Confident
Expert
graphics.basics
Count
0
5
10
15
20