Some Plots

Xuyan Xiao

2016/2/3

```
# setwd("C:\\Columbia Courses\\Visualization\\Project1")
library(xlsx)
```

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

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

```
survey = read.xlsx("Survey+Response.xlsx",sheetIndex = 1)

# change factor variables into char
for(i in 1:dim(survey)[2]){
  if(class(survey[,i])=="factor"){
    survey[,i] = as.character(survey[,i])
  }
}
# clear some of the answers
# unique(survey$Program)
survey$Program[survey$Program=="MSDS"]="IDSE (master)"
survey$Program[survey$Program=="Ms in ds"]="IDSE (master)"
survey$Program[survey$Program=="Data Science"]="IDSE (master)"
survey$Program[survey$Program=="QMSS"]="QMSS (master)"

# calculate the number of skills

numOfSkills = rowSums(survey[,12:31])
survey$numOfSkills=numOfSkills

library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

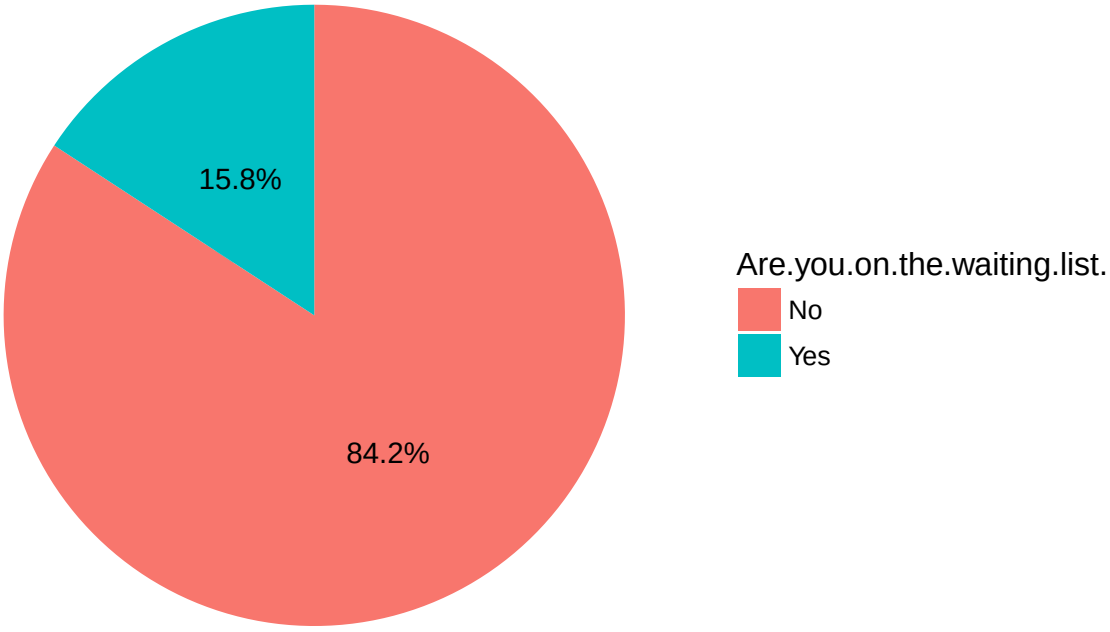
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(scales)
# Create blank theme
blank_theme = theme_minimal()+
```

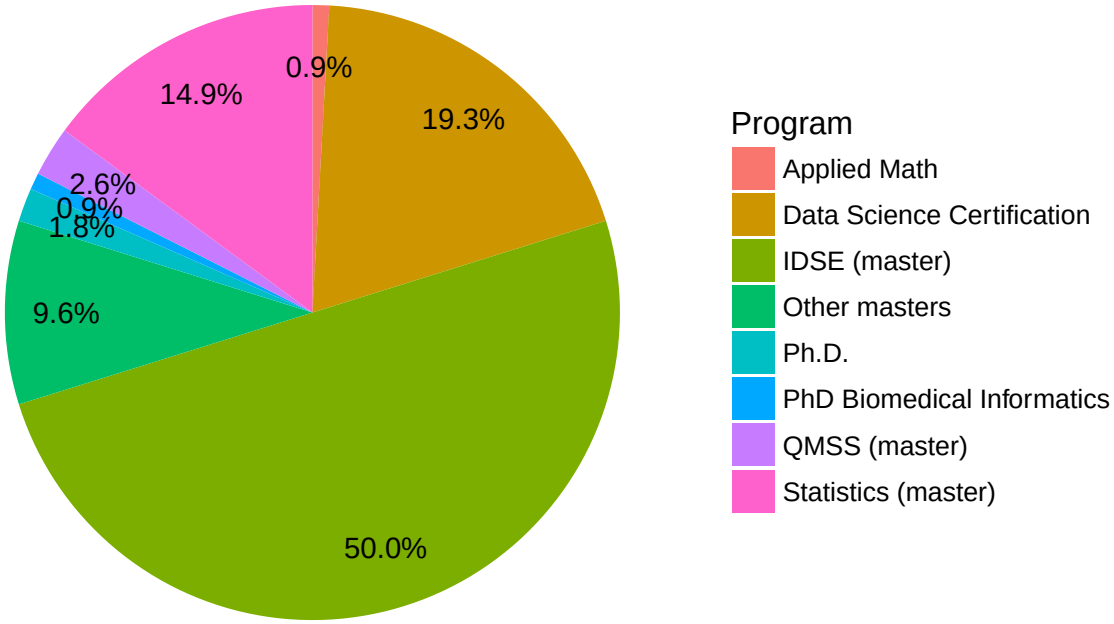
```
  theme(
    axis.title.x = element_blank(),
    axis.title.y = element_blank(),
    panel.border = element_blank(),
    panel.grid=element_blank(),
    axis.ticks = element_blank(),
    axis.text.x=element_blank(),
    axis.text.y=element_blank()
  )

# pie chart of the wl
wl_group = group_by(survey,Are.you.on.the.waiting.list.)
wl_sum = summarize(wl_group,num = n())
wl_sum$percent = percent(wl_sum$num/sum(wl_sum$num))
wl_bar = ggplot(wl_sum, aes(x ="Proportion",y=num, fill=Are.you.on.the.waiting.list.))+
  geom_bar(width = 1, stat = "identity")+ coord_polar("y", start=0)+blank_theme+
  geom_text(aes(y = num/2 + c(0, cumsum(num)[-length(num)]),label = percent))
wl_bar
```

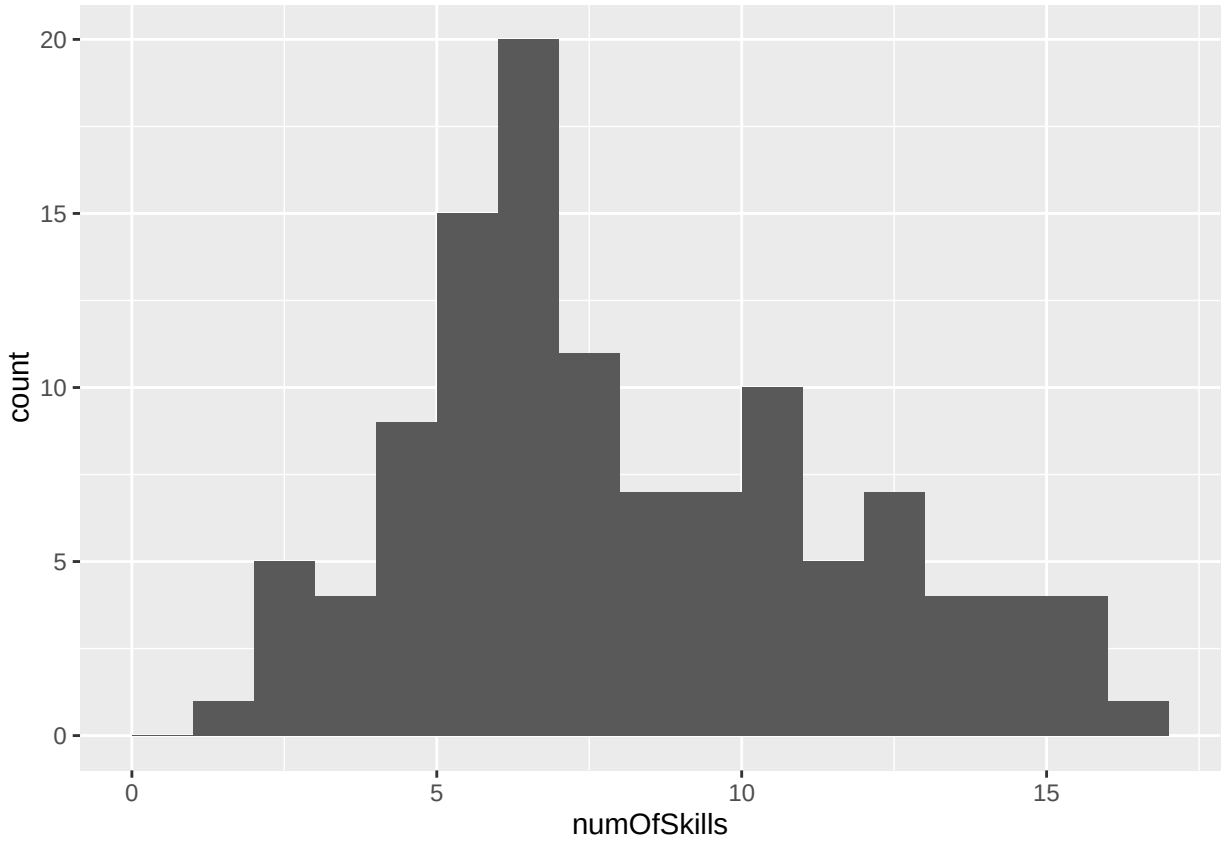



```
# pie chart of the programs
pr_group = group_by(survey,Program)
pr_sum = summarize(pr_group,num = n())
pr_sum$percent = percent(pr_sum$num/sum(pr_sum$num))
# pr_sum = pr_sum[order(-pr_sum$num),]
pr_bar = ggplot(pr_sum, aes(x ="Proportion",y=num, fill=Program))+
```

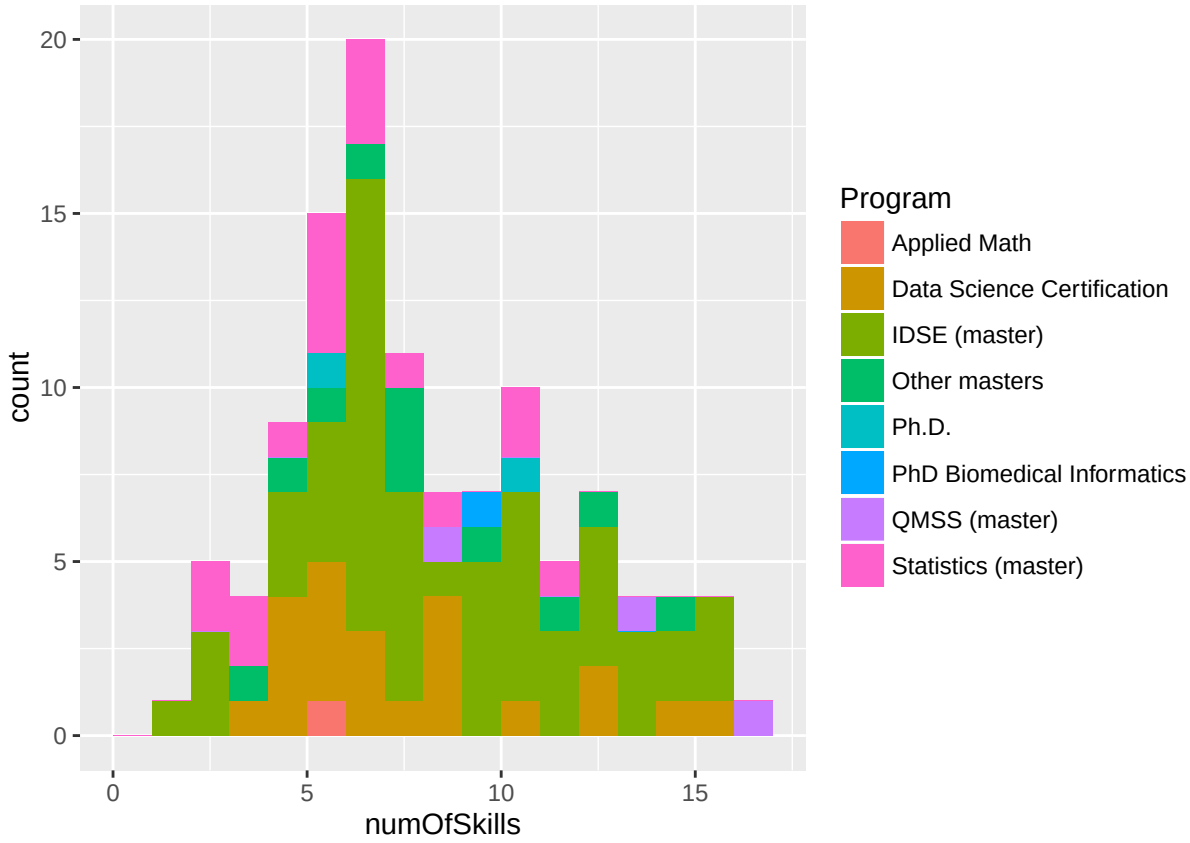
```
  geom_bar(width = 1, stat = "identity")+ coord_polar("y", start=0)+blank_theme+
  geom_text(aes(x = 1.3, y = num/2 + c(0, cumsum(num)[-length(num)]),label = percent))
pr_bar
```



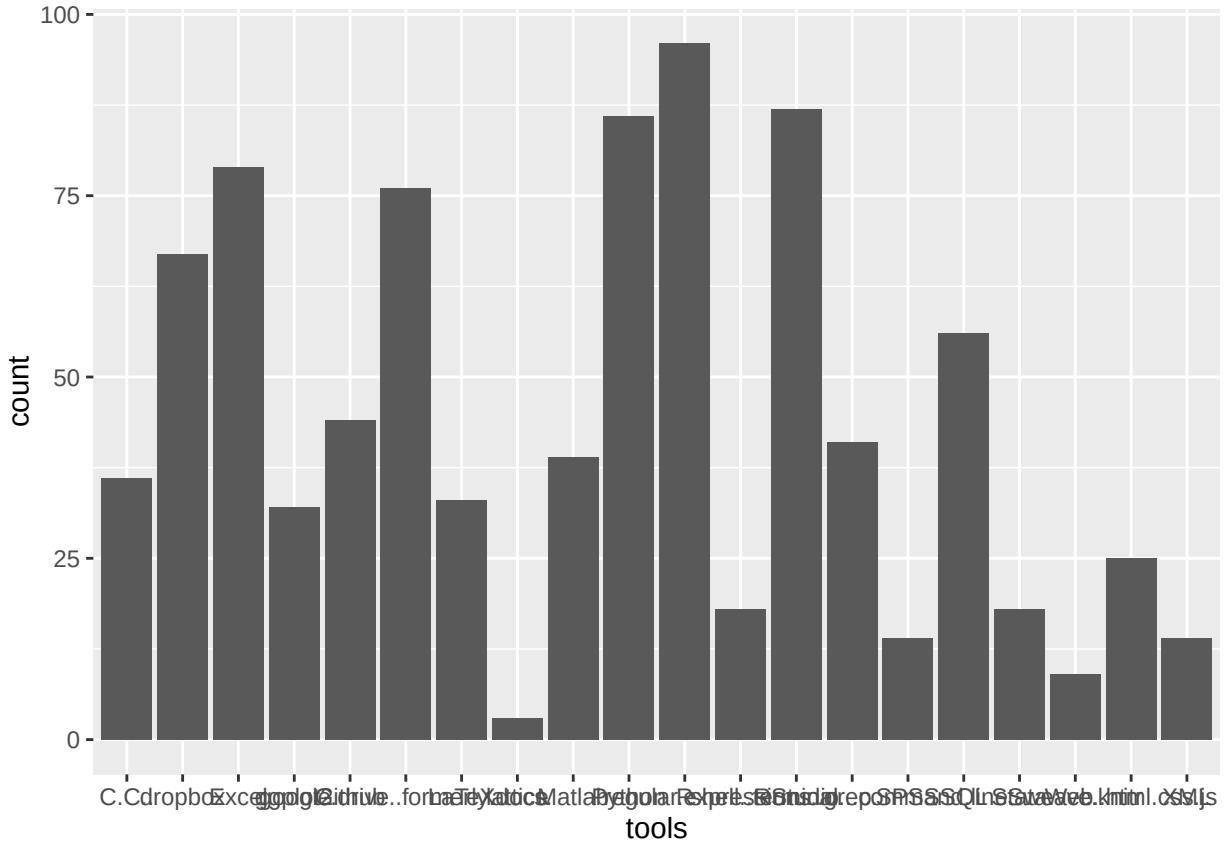

```
# number of tools
ggplot(survey, aes(numOfSkills)) +geom_histogram(binwidth=1)
```




```
ggplot(survey, aes(numOfSkills,fill=Program)) +geom_histogram(binwidth=1)
```




```
# Most popular tools
# names(survey)
tools = survey[,12:31]
toolCount = as.data.frame(colSums(tools))
names(toolCount) = "count"
toolCount$tools = row.names(toolCount)
row.names(toolCount) = seq(nrow(toolCount))
toolCount = toolCount[order(toolCount$count),]
ggplot(toolCount, aes(x = tools,weight = count)) +geom_bar()
```


100
75
50
25
0
count
tools