Skill_Confidence_graph

Sugandha

Wednesday, February 10, 2016

The graph Skill_Confidence_graph was written to analyse confidence level of skills:

- Programming and Analytical Experiences [R, graphic basics (base, lattice, grid etc.)]
- Programming and Analytical Experiences [R, advanced (multivariate data analysis, e.g. spatiotemporal data, visualization and modeling)]
- Programming and Analytical Experiences [Reproducible documentation with R (e.g. R Markdown)]
- Programming and Analytical Experiences [Matlab, data manipulation, analysis, visualization and modeling]
- Programming and Analytical Experiences [Github]

In order to make data quantitative, Confidence level were replaced with numerical value None - 0 , A little = 1, Confident - 2, Expert - 3

Programs with fewer students are combined as follow: Other Master : Other masters, QMSS, Applied Math

Other PhD : Phd , PhD Biomedical Informatics

So we have 5 programs to evaluate : DS Certification, IDSE masters, Other Masters, Other Phd, Statistics (Masters)

Following is the breakdown of code

```
source("src/tidydata.R")

cleanDF<-tidydata("raw/Survey+Response.xlsx")
```

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

```
df <- data.frame(cleanDF[,c(2,6,7,8,9,10)])

cols <- c("program","graphic","multivariate","Markdown","Matlab","Github")
colnames(df) <- cols



df[,c(2,3,4,5,6)] <- sapply(df[,c(2,3,4,5,6)],as.character)

# replace confidence level None = 0 , A little = 1, confident = 2, Expert = 3


df$graphic[df$graphic == "None"] <- 0
```

```
df$graphic[df$graphic == "A little"] <- 1
df$graphic[df$graphic == "Confident"] <- 2
df$graphic[df$graphic == "Expert"] <- 3

df$multivariate[df$multivariate == "None"] <- 0
df$multivariate[df$multivariate == "A little"] <- 1
df$multivariate[df$multivariate == "Confident"] <- 2
df$multivariate[df$multivariate == "Expert"] <- 3

df$Markdown[df$Markdown == "None"] <- 0
df$Markdown[df$Markdown == "A little"] <- 1
df$Markdown[df$Markdown == "Confident"] <- 2
df$Markdown[df$Markdown == "Expert"] <- 3

df$Matlab[df$Matlab == "None"] <- 0
df$Matlab[df$Matlab == "A little"] <- 1
df$Matlab[df$Matlab == "Confident"] <- 2
df$Matlab[df$Matlab == "Expert"] <- 3

df$Github[df$Github == "None"] <- 0
df$Github[df$Github == "A little"] <- 1
df$Github[df$Github == "Confident"] <- 2
df$Github[df$Github == "Expert"] <- 3
```

After making appropriate confidence numbers, mean is calculated across all 5 skills for different program

```
### calculate mean of confidence level ######

df_IDSE <- subset(df, program =="IDSE (master)")

mean_graph_idse <-mean (as.numeric(df_IDSE$graphic))
mean_multivariate_idse <-mean (as.numeric(df_IDSE$multivariate))
mean_Markdown_idse <-mean (as.numeric(df_IDSE$Markdown))
mean_Matlab_idse <-mean (as.numeric(df_IDSE$Matlab))
mean_Github_idse <-mean (as.numeric(df_IDSE$Github))

df_OTMS <- subset(df, program =="Other masters")

mean_graph_OTMS <-mean (as.numeric(df_OTMS$graphic))
mean_multivariate_OTMS <-mean (as.numeric(df_OTMS$multivariate))
mean_Markdown_OTMS <-mean (as.numeric(df_OTMS$Markdown))
mean_Matlab_OTMS <-mean (as.numeric(df_OTMS$Matlab))
mean_Github_OTMS <-mean (as.numeric(df_OTMS$Github))

df_DSCert <- subset(df, program =="DS Certification")

mean_graph_DSCert <-mean (as.numeric(df_DSCert$graphic))
mean_multivariate_DSCert <-mean (as.numeric(df_DSCert$multivariate))
mean_Markdown_DSCert <-mean (as.numeric(df_DSCert$Markdown))
mean_Matlab_DSCert <-mean (as.numeric(df_DSCert$Matlab))
mean_Github_DSCert <-mean (as.numeric(df_DSCert$Github))

df_stat <- subset(df, program =="Statistics (master)")
```

```
mean_graph_stat <-mean (as.numeric(df_stat$graphic))
mean_multivariate_stat <-mean (as.numeric(df_stat$multivariate))
mean_Markdown_stat <-mean (as.numeric(df_stat$Markdown))
mean_Matlab_stat <-mean (as.numeric(df_stat$Matlab))
mean_Github_stat <-mean (as.numeric(df_stat$Github))

df_phd <- subset(df, program =="Other PhD")

mean_graph_phd <-mean (as.numeric(df_phd$graphic))
mean_multivariate_phd <-mean (as.numeric(df_phd$multivariate))
mean_Markdown_phd <-mean (as.numeric(df_phd$Markdown))
mean_Matlab_phd <-mean (as.numeric(df_phd$Matlab))
mean_Github_phd <-mean (as.numeric(df_phd$Github))
```

The mean of the confidence level was then plotted as line graph

```
program<- c("IDSE (master)","Other masters","DS Certification","Statistics (master)","Other PhD")
mean_graph <- c(mean_graph_idse, mean_graph_OTMS,mean_graph_DSCert, mean_graph_stat, mean_graph_phd)
mean_multivariate <- c(mean_multivariate_idse, mean_multivariate_OTMS,mean_multivariate_DSCert, mean_mul
mean_Markdown <- c(mean_Markdown_idse, mean_Markdown_OTMS,mean_Markdown_DSCert, mean_Markdown_stat, mean
mean_Matlab <- c(mean_Matlab_idse, mean_Matlab_OTMS,mean_Matlab_DSCert, mean_Matlab_stat, mean_Matlab_pl

mean_Github <- c(mean_Github_idse, mean_Github_OTMS,mean_Github_DSCert, mean_Github_stat, mean_Github_pl

### Construct dataframe from the obtained means of skills across all programs ###

Program_exp <- data.frame(program, mean_graph, mean_multivariate,mean_Markdown , mean_Matlab,mean_Githul


Program_exp <- Program_exp[order(program),]

g_range <- range (0,3)

plot_colors <- c("blue","red","green","black","Cyan")

## create the plot ##
plot(c(1,5), c(0,3), type="n", xlab="Program", ylab="Average Confidence",axes=F, ann=F)


#### Labelling Axis ####
axis(1, at= Program_exp$program, labels = Program_exp$program )
axis(2, las=1, at= seq(0, max(g_range), .2))#0:max(g_range))
axis(4, las =0 ,at= c(0,1,2,3), labels = c("None","A Little","Confident","Expert"),cex= 2)
box()

## Title the axis
title(xlab= "Program", col.lab=rgb(0,0.5,0))
title(ylab= "Average Confidence", col.lab=rgb(0,0.5,0))

## plot line for all the skills ##

lines(Program_exp$mean_graph, type="o", pch=22, lty=2, col=plot_colors[1],lwd = 2)
```

```
lines(Program_exp$mean_multivariate, type="o", pch=22, lty=2, col=plot_colors[2],lwd = 2)

lines(Program_exp$mean_Markdown, type="o", pch=22, lty=2, col=plot_colors[3],lwd = 2)


lines(Program_exp$mean_Matlab, type="o", pch=22, lty=2, col=plot_colors[4],lwd = 2)


lines(Program_exp$mean_Github, type="o", pch=22, lty=2, col=plot_colors[5],lwd = 2)

## add legend #

legend("topright", c("Graphics","Multivariate","Markdown","Matlab","Github"),
       col=plot_colors, pch=22, lty=2,lwd = 2, bty= "n",cex=.6)
```