Lab\_1

Group 1

8/8/2019

## Lab 1

### Data Exploration

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.1.0 ✔ purrr 0.2.5  
## ✔ tibble 2.1.3 ✔ dplyr 0.7.8  
## ✔ tidyr 0.8.2 ✔ stringr 1.3.1  
## ✔ readr 1.2.1 ✔ forcats 0.3.0

## Warning: package 'tibble' was built under R version 3.5.2

## ── Conflicts ────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

df <- read\_csv("survey\_responses.csv", )

## Parsed with column specification:  
## cols(  
## Timestamp = col\_character(),  
## `1. I'm satisfied to apply my talents and expertise in my job.` = col\_character(),  
## `2. I have opportunities for professional development within the company.` = col\_character(),  
## `3. The job-related training my company offers is sufficient.` = col\_character(),  
## `4. Which type of training you've been provided by the company?` = col\_character(),  
## `5. Do you utilize the training skills and knowledge acquired through the training programme?` = col\_character(),  
## `6. Overall what do you feel about the training programme?` = col\_character(),  
## `7. How was this survey?` = col\_character()  
## )

#### Data Pre-processing

# remove columns that are not required and convert other characters to factors  
df\_clean <- df%>%select(-Timestamp)%>% unclass %>% data.frame  
str(df\_clean)

## 'data.frame': 12 obs. of 7 variables:  
## $ X1..I.m.satisfied.to.apply.my.talents.and.expertise.in.my.job. : Factor w/ 5 levels "Agree","Disagree",..: 4 3 3 1 1 5 1 1 2 3 ...  
## $ X2..I.have.opportunities.for.professional.development.within.the.company. : Factor w/ 5 levels "Agree","Disagree",..: 4 1 1 1 1 5 1 3 3 2 ...  
## $ X3..The.job.related.training.my.company.offers.is.sufficient. : Factor w/ 4 levels "Agree","Disagree",..: 1 2 1 1 1 3 3 3 3 2 ...  
## $ X4..Which.type.of.training.you.ve.been.provided.by.the.company. : Factor w/ 4 levels "E-learning (virtual classroom, online class etc)",..: 1 1 2 2 1 3 4 2 1 2 ...  
## $ X5..Do.you.utilize.the.training.skills.and.knowledge.acquired.through.the.training.programme.: Factor w/ 4 levels "Always","Never",..: 2 4 4 3 1 1 1 3 4 3 ...  
## $ X6..Overall.what.do.you.feel.about.the.training.programme. : Factor w/ 4 levels "Above Average",..: 1 2 2 4 1 4 2 2 2 2 ...  
## $ X7..How.was.this.survey. : Factor w/ 3 levels "Excellent","Fair",..: 1 1 2 3 3 1 3 2 3 3 ...

# Shorten the column\_names, also store the orig columns as key  
orig\_colnames <- colnames(df\_clean)%>% str\_replace\_all("[^[:alnum:]]", " ")  
new\_colnames <- c("Q1","Q2","Q3","Q4","Q5","Q6","Q7")  
colnames\_key <- data.frame(new\_colnames, orig\_colnames)  
colnames(df\_clean) <- new\_colnames

summary(df\_clean)

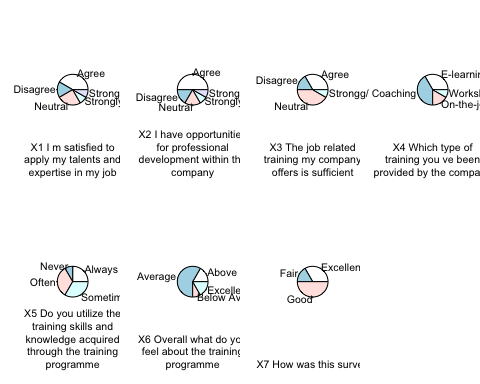
## Q1 Q2 Q3   
## Agree :5 Agree :6 Agree :4   
## Disagree :2 Disagree :2 Disagree :2   
## Neutral :3 Neutral :2 Neutral :5   
## Strongly Agree :1 Strongly Agree :1 Strongly disagree:1   
## Strongly disagree:1 Strongly disagree:1   
## Q4 Q5   
## E-learning (virtual classroom, online class etc):4 Always :3   
## Mentoring/ Shadowing/ Coaching :5 Never :1   
## On-the-job training :2 Often :4   
## Workshops :1 Sometimes:4   
##   
## Q6 Q7   
## Above Average:2 Excellent:4   
## Average :7 Fair :2   
## Below Average:1 Good :6   
## Excellent :2   
##

#### Plotting Charts

x1 <- df\_clean%>%group\_by(Q1)%>%count()  
x2 <- df\_clean%>%group\_by(Q2)%>%count()  
x3 <- df\_clean%>%group\_by(Q3)%>%count()  
x4 <- df\_clean%>%group\_by(Q4)%>%count()  
x5 <- df\_clean%>%group\_by(Q5)%>%count()  
x6 <- df\_clean%>%group\_by(Q6)%>%count()  
x7 <- df\_clean%>%group\_by(Q7)%>%count()

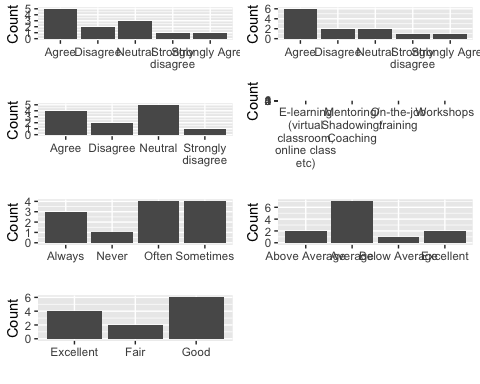
##### Pie Charts

par(mfrow = c(2, 4)) # Set up a 4 x 2 plotting space  
  
# Plot pie of x's  
pie(x = x1$n, labels = x1$Q1, xlab = paste(strwrap(orig\_colnames[1],width=25),collapse="\n"))  
pie(x = x2$n, labels = x2$Q2, xlab = paste(strwrap(orig\_colnames[2],width=25),collapse="\n"))  
pie(x = x3$n, labels = x3$Q3, xlab = paste(strwrap(orig\_colnames[3],width=25),collapse="\n"))  
pie(x = x4$n, labels = x4$Q4, xlab = paste(strwrap(orig\_colnames[4],width=25),collapse="\n"))  
pie(x = x5$n, labels = x5$Q5, xlab = paste(strwrap(orig\_colnames[5],width=25),collapse="\n"))  
pie(x = x6$n, labels = x6$Q6, xlab = paste(strwrap(orig\_colnames[6],width=25),collapse="\n"))  
pie(x = x7$n, labels = x7$Q7, xlab = paste(strwrap(orig\_colnames[7],width=25),collapse="\n"))



##### Bar Charts

par(mfrow = c(2, 4))  
b1 <- x1%>%ggplot(aes(x=str\_wrap(Q1,width = 16),y=n))+geom\_bar(stat="identity")+xlab("") + ylab("Count")  
b2 <- x2%>%ggplot(aes(x=str\_wrap(Q2,width = 16),y=n))+geom\_bar(stat="identity")+xlab("") + ylab("Count")  
b3 <- x3%>%ggplot(aes(x=str\_wrap(Q3,width = 16),y=n))+geom\_bar(stat="identity")+xlab("") + ylab("Count")  
b4 <- x4%>%ggplot(aes(x=str\_wrap(Q4,width = 16),y=n))+geom\_bar(stat="identity")+xlab("") + ylab("Count")  
b5 <- x5%>%ggplot(aes(x=str\_wrap(Q5,width = 16),y=n))+geom\_bar(stat="identity")+xlab("") + ylab("Count")  
b6 <- x6%>%ggplot(aes(x=str\_wrap(Q6,width = 16),y=n))+geom\_bar(stat="identity")+xlab("") + ylab("Count")  
b7 <- x7%>%ggplot(aes(x=str\_wrap(Q7,width = 16),y=n))+geom\_bar(stat="identity")+xlab("") + ylab("Count")  
grid.arrange(b1,b2,b3,b4,b5,b6,b7,ncol=2,nrow=4)



#### Analysis

##### Linear Regression

fit1<-glm(Q1~., data=df\_clean, family = "binomial")  
summary(fit1)

##   
## Call:  
## glm(formula = Q1 ~ ., family = "binomial", data = df\_clean)  
##   
## Deviance Residuals:   
## [1] 0 0 0 0 0 0 0 0 0 0 0 0  
##   
## Coefficients: (7 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -2.457e+01 1.310e+05 0 1  
## Q2Disagree 4.913e+01 2.620e+05 0 1  
## Q2Neutral 4.141e-14 2.620e+05 0 1  
## Q2Strongly Agree 4.913e+01 1.853e+05 0 1  
## Q2Strongly disagree 9.826e+01 3.209e+05 0 1  
## Q3Disagree 3.757e-10 2.620e+05 0 1  
## Q3Neutral -3.316e-15 1.853e+05 0 1  
## Q3Strongly disagree -1.100e-06 3.209e+05 0 1  
## Q4Mentoring/ Shadowing/ Coaching 7.091e-10 2.620e+05 0 1  
## Q4On-the-job training -4.913e+01 3.209e+05 0 1  
## Q4Workshops NA NA NA NA  
## Q5Never NA NA NA NA  
## Q5Often -7.092e-10 3.706e+05 0 1  
## Q5Sometimes 4.913e+01 2.620e+05 0 1  
## Q6Average NA NA NA NA  
## Q6Below Average NA NA NA NA  
## Q6Excellent NA NA NA NA  
## Q7Fair NA NA NA NA  
## Q7Good NA NA NA NA  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.6301e+01 on 11 degrees of freedom  
## Residual deviance: 5.1440e-10 on 0 degrees of freedom  
## AIC: 24  
##   
## Number of Fisher Scoring iterations: 23

fit2<-glm(Q7~., data=df\_clean, family = "binomial")  
summary(fit2)

##   
## Call:  
## glm(formula = Q7 ~ ., family = "binomial", data = df\_clean)  
##   
## Deviance Residuals:   
## [1] 0 0 0 0 0 0 0 0 0 0 0 0  
##   
## Coefficients: (9 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 2.457e+01 1.310e+05 0 1  
## Q1Disagree -1.202e-06 2.620e+05 0 1  
## Q1Neutral 5.006e-08 1.853e+05 0 1  
## Q1Strongly Agree -4.913e+01 1.853e+05 0 1  
## Q1Strongly disagree -5.822e-15 1.853e+05 0 1  
## Q2Disagree 4.913e+01 2.620e+05 0 1  
## Q2Neutral 1.002e-07 2.620e+05 0 1  
## Q2Strongly Agree NA NA NA NA  
## Q2Strongly disagree NA NA NA NA  
## Q3Disagree -4.913e+01 2.620e+05 0 1  
## Q3Neutral -5.076e-08 1.853e+05 0 1  
## Q3Strongly disagree -4.913e+01 3.209e+05 0 1  
## Q4Mentoring/ Shadowing/ Coaching -4.963e-08 1.853e+05 0 1  
## Q4On-the-job training -4.913e+01 1.853e+05 0 1  
## Q4Workshops NA NA NA NA  
## Q5Never NA NA NA NA  
## Q5Often NA NA NA NA  
## Q5Sometimes NA NA NA NA  
## Q6Average NA NA NA NA  
## Q6Below Average NA NA NA NA  
## Q6Excellent NA NA NA NA  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.5276e+01 on 11 degrees of freedom  
## Residual deviance: 5.1440e-10 on 0 degrees of freedom  
## AIC: 24  
##   
## Number of Fisher Scoring iterations: 23

None of the variable are significant predictors.