R Notebook For APS Survey Analysis

# Libraries

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.1.0 v dplyr 1.0.5  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# Reading datafile for year 2014 and 2020

aps\_2014 <- read.csv('2014-aps-employee-census-5-point-dataset.csv',na.strings = c(""," "))  
aps\_2020 <- read.csv('2020-aps-employee-census-dataset.csv', na.strings = c(""," "))

# DATA ACQUISITION - Common Questions

## Year 2014

# Fetching relevant columns from the dataset for analysis which are common in both the years

common\_questions\_2014 <- aps\_2014[c('ï..AS','q1','q2.','q6.','q36a', 'q36b','q36c', 'q36d',  
 'q36e','q36f','q36g','q70','q73','q18c','q18d','q18e','q18f','q18g',  
 'q19b','q22a','q22b','q22c','q22d','q22e','q22q','q22s')]

# Adding year variable to the dataset  
   
common\_questions\_2014$year <- '2014'

# rename column names  
  
colnames(common\_questions\_2014)<- c("AS",'Q1','Q2','Q3','Q4a','Q4b','Q4c','Q4d','Q4e','Q4f','Q4g','Q5',  
 'Q6','Q7a','Q7b','Q7c','Q7d','Q7e','Q8','Q9a','Q9b','Q9c','Q9d',  
 'Q9e','Q9f','Q9g','YEAR')

## Inconsistencies

# Checking unique values present for each of the categorical variable in the data

lapply(common\_questions\_2014, unique)

## $AS  
## [1] "Large (1,001 or more employees)" "Medium (251 to 1,000 employees)"  
## [3] "Small (Less than 250 employees)"  
##   
## $Q1  
## [1] "Male"   
## [2] NA   
## [3] "Female"   
## [4] "X (Indeterminate/Intersex/Unspecified)"  
##   
## $Q2  
## [1] "40 to 54 years" "Under 40 years" NA   
## [4] "55 years or older"  
##   
## $Q3  
## [1] "Trainee/Graduate/APS" "EL" NA   
## [4] "SES"   
##   
## $Q4a  
## [1] "Sometimes" "Rarely" "Often" NA "Never" "Always"   
##   
## $Q4b  
## [1] "Sometimes" "Rarely" "Always" "Often" NA "Never"   
##   
## $Q4c  
## [1] "Often" "Sometimes" "Always" "Rarely" NA "Never"   
##   
## $Q4d  
## [1] "Often" "Always" "Sometimes" NA "Rarely" "Never"   
##   
## $Q4e  
## [1] "Rarely" "Sometimes" NA "Often" "Never" "Always"   
##   
## $Q4f  
## [1] "Always" "Sometimes" "Often" NA "Never" "Rarely"   
##   
## $Q4g  
## [1] "Sometimes" "Rarely" "Never" "Always" "Often" NA   
##   
## $Q5  
## [1] "No" "Would prefer not to answer"  
## [3] "Not sure" NA   
## [5] "Yes"   
##   
## $Q6  
## [1] "No" "Not sure" NA "Yes"   
##   
## $Q7a  
## [1] "Agree" "Disagree"   
## [3] "Strongly agree" "Neither agree nor disagree"  
## [5] NA "Strongly disagree"   
##   
## $Q7b  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Disagree" NA   
## [5] "Strongly agree" "Strongly disagree"   
##   
## $Q7c  
## [1] "Neither agree nor disagree" "Strongly disagree"   
## [3] "Agree" "Disagree"   
## [5] NA "Strongly agree"   
##   
## $Q7d  
## [1] "Neither agree nor disagree" "Strongly disagree"   
## [3] "Agree" "Disagree"   
## [5] NA "Strongly agree"   
##   
## $Q7e  
## [1] "Neither agree nor disagree" "Strongly disagree"   
## [3] "Agree" NA   
## [5] "Strongly agree" "Disagree"   
##   
## $Q8  
## [1] "Agree" "Strongly agree"   
## [3] NA "Disagree"   
## [5] "Neither agree nor disagree" "Strongly disagree"   
##   
## $Q9a  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Strongly agree" "Disagree"   
## [5] NA "Strongly disagree"   
##   
## $Q9b  
## [1] "Neither agree nor disagree" "Agree"   
## [3] "Disagree" NA   
## [5] "Strongly agree" "Strongly disagree"   
##   
## $Q9c  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Strongly agree" NA   
## [5] "Strongly disagree" "Disagree"   
##   
## $Q9d  
## [1] "Disagree" "Agree"   
## [3] "Strongly disagree" "Neither agree nor disagree"  
## [5] "Strongly agree" NA   
##   
## $Q9e  
## [1] "Disagree" "Strongly disagree"   
## [3] "Agree" "Neither agree nor disagree"  
## [5] NA "Strongly agree"   
##   
## $Q9f  
## [1] "Neither agree nor disagree" NA   
## [3] "Agree" "Strongly agree"   
## [5] "Disagree" "Strongly disagree"   
##   
## $Q9g  
## [1] "Neither agree nor disagree" "Agree"   
## [3] "Disagree" NA   
## [5] "Strongly agree" "Strongly disagree"   
##   
## $YEAR  
## [1] "2014"

# Replacing the Indeterminate/Intersex/Unspecified in the Gender Variable to ‘X’

common\_questions\_2014['Q1'][common\_questions\_2014['Q1']=='X (Indeterminate/Intersex/Unspecified)']<-'X'  
common\_questions\_2014['Q6'][common\_questions\_2014['Q6'] == 'Not sure'] <- 'Not Sure'  
unique(common\_questions\_2014['Q1'])

## Q1  
## 1 Male  
## 4 <NA>  
## 6 Female  
## 641 X

## 

## Year 2020

#Extracting common questions from the year 2020

common\_questions\_2020 <- aps\_2020[ c('ï..AS','q1','q2.','q5.','q47a', 'q47b', 'q47c', 'q47d', 'q47e', 'q47f', 'q47g', 'q63', 'q61',"q17a","q17b","q17c","q17d","q17e", "q18b","q23a","q23l","q23c", "q23d","q23f","q23n", "q23g" )]

#Adding year variable  
common\_questions\_2020$year <- '2020'  
  
#Renaming column names

colnames(common\_questions\_2020)<- c("AS",'Q1','Q2','Q3','Q4a','Q4b','Q4c','Q4d','Q4e','Q4f','Q4g','Q5',  
 'Q6','Q7a','Q7b','Q7c','Q7d','Q7e','Q8','Q9a','Q9b','Q9c','Q9d','Q9e',  
 'Q9f','Q9g','YEAR')

### Inconsistencies

# Checking unique values for each of the categorical variable

lapply(common\_questions\_2020, unique)

## $AS  
## [1] "Large (1,001 or more employees)" "Medium (251 to 1,000 employees)"  
## [3] "Small (Less than 250 employees)"  
##   
## $Q1  
## [1] "Male"   
## [2] "Female"   
## [3] "Prefer not to say"   
## [4] "X (Indeterminate/Intersex/Unspecified)"  
##   
## $Q2  
## [1] "Under 40 years" "40 to 54 years" "55 years or older"  
## [4] NA   
##   
## $Q3  
## [1] "Trainee/Graduate/APS" "EL" "SES"   
## [4] NA   
##   
## $Q4a  
## [1] "Sometimes" "Rarely" "Always" "Often" NA "Never"   
##   
## $Q4b  
## [1] "Rarely" "Always" "Sometimes" "Never" NA "Often"   
##   
## $Q4c  
## [1] "Often" "Always" "Sometimes" NA "Rarely" "Never"   
##   
## $Q4d  
## [1] "Often" "Sometimes" "Always" NA "Rarely" "Never"   
##   
## $Q4e  
## [1] "Often" "Always" "Sometimes" NA "Rarely" "Never"   
##   
## $Q4f  
## [1] "Never" "Sometimes" "Rarely" "Always" NA "Often"   
##   
## $Q4g  
## [1] "Sometimes" "Often" "Rarely" "Always" NA "Never"   
##   
## $Q5  
## [1] "No" NA   
## [3] "Not sure" "Would prefer not to answer"  
## [5] "Yes"   
##   
## $Q6  
## [1] "No" NA "Yes" "Not Sure"  
##   
## $Q7a  
## [1] "Agree" "Strongly agree"   
## [3] "Disagree" "Strongly disagree"   
## [5] "Neither agree nor disagree" NA   
##   
## $Q7b  
## [1] "Agree" "Strongly agree"   
## [3] "Neither agree nor disagree" "Strongly disagree"   
## [5] "Disagree" NA   
##   
## $Q7c  
## [1] "Neither agree nor disagree" "Agree"   
## [3] "Disagree" "Strongly agree"   
## [5] "Strongly disagree" NA   
##   
## $Q7d  
## [1] "Agree" "Strongly disagree"   
## [3] "Strongly agree" "Disagree"   
## [5] "Neither agree nor disagree" NA   
##   
## $Q7e  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Strongly agree" "Disagree"   
## [5] "Strongly disagree" NA   
##   
## $Q8  
## [1] "Agree" "Strongly agree"   
## [3] "Neither agree nor disagree" "Disagree"   
## [5] NA "Strongly disagree"   
##   
## $Q9a  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Disagree" "Strongly disagree"   
## [5] "Strongly agree" NA   
##   
## $Q9b  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Disagree" "Strongly agree"   
## [5] NA "Strongly disagree"   
##   
## $Q9c  
## [1] "Strongly agree" "Agree"   
## [3] "Strongly disagree" "Neither agree nor disagree"  
## [5] "Disagree" NA   
##   
## $Q9d  
## [1] "Agree" "Disagree"   
## [3] "Strongly disagree" "Strongly agree"   
## [5] "Neither agree nor disagree" NA   
##   
## $Q9e  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Disagree" "Strongly agree"   
## [5] NA "Strongly disagree"   
##   
## $Q9f  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Strongly disagree" "Strongly agree"   
## [5] "Disagree" NA   
##   
## $Q9g  
## [1] "Strongly agree" "Neither agree nor disagree"  
## [3] "Agree" "Strongly disagree"   
## [5] "Disagree" NA   
##   
## $YEAR  
## [1] "2020"

In Q1 –> ‘Prefer not to say’ is converted to ‘X(Indeterminate/Intersex/Unspecified)’ to maintain consistency with 2014 Data

common\_questions\_2020['Q1'][common\_questions\_2020['Q1']=='Prefer not to say']<-'X'  
common\_questions\_2020['Q1'][common\_questions\_2020['Q1']=='X (Indeterminate/Intersex/Unspecified)']<-'X'  
unique(common\_questions\_2020['Q1'])

## Q1  
## 1 Male  
## 4 Female  
## 5 X

# MERGING

#Vertically merged the two datasets for the year 2014 and 2020

common\_questions\_merged <- rbind(common\_questions\_2014,common\_questions\_2020)

str(common\_questions\_merged)

## 'data.frame': 207477 obs. of 27 variables:  
## $ AS : chr "Large (1,001 or more employees)" "Large (1,001 or more employees)" "Large (1,001 or more employees)" "Large (1,001 or more employees)" ...  
## $ Q1 : chr "Male" "Male" "Male" NA ...  
## $ Q2 : chr "40 to 54 years" "Under 40 years" "Under 40 years" NA ...  
## $ Q3 : chr "Trainee/Graduate/APS" "Trainee/Graduate/APS" "Trainee/Graduate/APS" "Trainee/Graduate/APS" ...  
## $ Q4a : chr "Sometimes" "Rarely" "Sometimes" "Rarely" ...  
## $ Q4b : chr "Sometimes" "Rarely" "Always" "Sometimes" ...  
## $ Q4c : chr "Often" "Sometimes" "Always" "Often" ...  
## $ Q4d : chr "Often" "Often" "Always" "Often" ...  
## $ Q4e : chr "Rarely" "Sometimes" "Sometimes" "Rarely" ...  
## $ Q4f : chr "Always" "Sometimes" "Always" "Often" ...  
## $ Q4g : chr "Sometimes" "Rarely" "Never" "Sometimes" ...  
## $ Q5 : chr "No" "Would prefer not to answer" "No" "Not sure" ...  
## $ Q6 : chr "No" "Not Sure" "No" "No" ...  
## $ Q7a : chr "Agree" "Agree" "Agree" "Agree" ...  
## $ Q7b : chr "Agree" "Neither agree nor disagree" "Neither agree nor disagree" "Agree" ...  
## $ Q7c : chr "Neither agree nor disagree" "Neither agree nor disagree" "Strongly disagree" "Agree" ...  
## $ Q7d : chr "Neither agree nor disagree" "Neither agree nor disagree" "Strongly disagree" "Agree" ...  
## $ Q7e : chr "Neither agree nor disagree" "Neither agree nor disagree" "Strongly disagree" "Agree" ...  
## $ Q8 : chr "Agree" "Agree" "Agree" "Agree" ...  
## $ Q9a : chr "Agree" "Neither agree nor disagree" "Agree" "Strongly agree" ...  
## $ Q9b : chr "Neither agree nor disagree" "Neither agree nor disagree" "Agree" "Neither agree nor disagree" ...  
## $ Q9c : chr "Agree" "Neither agree nor disagree" "Strongly agree" "Strongly agree" ...  
## $ Q9d : chr "Disagree" "Disagree" "Disagree" "Disagree" ...  
## $ Q9e : chr "Disagree" "Strongly disagree" "Disagree" "Disagree" ...  
## $ Q9f : chr "Neither agree nor disagree" "Neither agree nor disagree" NA "Neither agree nor disagree" ...  
## $ Q9g : chr "Neither agree nor disagree" "Neither agree nor disagree" "Agree" "Neither agree nor disagree" ...  
## $ YEAR: chr "2014" "2014" "2014" "2014" ...

head(common\_questions\_merged)

## AS Q1 Q2 Q3  
## 1 Large (1,001 or more employees) Male 40 to 54 years Trainee/Graduate/APS  
## 2 Large (1,001 or more employees) Male Under 40 years Trainee/Graduate/APS  
## 3 Large (1,001 or more employees) Male Under 40 years Trainee/Graduate/APS  
## 4 Large (1,001 or more employees) <NA> <NA> Trainee/Graduate/APS  
## 5 Large (1,001 or more employees) Male 55 years or older Trainee/Graduate/APS  
## 6 Large (1,001 or more employees) Female Under 40 years Trainee/Graduate/APS  
## Q4a Q4b Q4c Q4d Q4e Q4f Q4g  
## 1 Sometimes Sometimes Often Often Rarely Always Sometimes  
## 2 Rarely Rarely Sometimes Often Sometimes Sometimes Rarely  
## 3 Sometimes Always Always Always Sometimes Always Never  
## 4 Rarely Sometimes Often Often Rarely Often Sometimes  
## 5 Rarely Sometimes Always Often Rarely Always Sometimes  
## 6 Sometimes Sometimes Sometimes Sometimes Sometimes Often Rarely  
## Q5 Q6 Q7a Q7b  
## 1 No No Agree Agree  
## 2 Would prefer not to answer Not Sure Agree Neither agree nor disagree  
## 3 No No Agree Neither agree nor disagree  
## 4 Not sure No Agree Agree  
## 5 No No Agree Agree  
## 6 No No Agree Disagree  
## Q7c Q7d  
## 1 Neither agree nor disagree Neither agree nor disagree  
## 2 Neither agree nor disagree Neither agree nor disagree  
## 3 Strongly disagree Strongly disagree  
## 4 Agree Agree  
## 5 Agree Agree  
## 6 Disagree Strongly disagree  
## Q7e Q8 Q9a  
## 1 Neither agree nor disagree Agree Agree  
## 2 Neither agree nor disagree Agree Neither agree nor disagree  
## 3 Strongly disagree Agree Agree  
## 4 Agree Agree Strongly agree  
## 5 Neither agree nor disagree Agree Agree  
## 6 Agree Agree Neither agree nor disagree  
## Q9b Q9c Q9d  
## 1 Neither agree nor disagree Agree Disagree  
## 2 Neither agree nor disagree Neither agree nor disagree Disagree  
## 3 Agree Strongly agree Disagree  
## 4 Neither agree nor disagree Strongly agree Disagree  
## 5 Neither agree nor disagree Neither agree nor disagree Agree  
## 6 Disagree Agree Strongly disagree  
## Q9e Q9f Q9g YEAR  
## 1 Disagree Neither agree nor disagree Neither agree nor disagree 2014  
## 2 Strongly disagree Neither agree nor disagree Neither agree nor disagree 2014  
## 3 Disagree <NA> Agree 2014  
## 4 Disagree Neither agree nor disagree Neither agree nor disagree 2014  
## 5 Disagree Neither agree nor disagree Agree 2014  
## 6 Strongly disagree Neither agree nor disagree Neither agree nor disagree 2014

# replacing Na values in Sex with X  
common\_questions\_merged$Q1[is.na(common\_questions\_merged$Q1)] <- 'X'  
  
# replacing Na values in Age with Unknown  
  
common\_questions\_merged$Q2[is.na(common\_questions\_merged$Q2)] <- 'Unknown'

# Function to replace the missing values with the most frequent value i.e. Mode

replace\_mode <- function(colname){  
   
unique\_vars <- unique(colname)  
mode <- unique\_vars[which.max(tabulate(match(colname, unique\_vars)))]  
colname[is.na(colname)] <- mode  
return(colname)  
}

# Applying the replace\_mode function to the merged, 2014 and 2020 dataset

common\_questions\_merged\_fillmode <- data.frame(sapply(common\_questions\_merged,replace\_mode))  
common\_questions\_2014\_fillmode <- data.frame(sapply(common\_questions\_2014,replace\_mode))  
common\_questions\_2020\_fillmode <- data.frame(sapply(common\_questions\_2020,replace\_mode))

# Verifying to see, there are no missing values in the data

lapply(common\_questions\_merged\_fillmode, unique)

## $AS  
## [1] "Large (1,001 or more employees)" "Medium (251 to 1,000 employees)"  
## [3] "Small (Less than 250 employees)"  
##   
## $Q1  
## [1] "Male" "X" "Female"  
##   
## $Q2  
## [1] "40 to 54 years" "Under 40 years" "Unknown"   
## [4] "55 years or older"  
##   
## $Q3  
## [1] "Trainee/Graduate/APS" "EL" "SES"   
##   
## $Q4a  
## [1] "Sometimes" "Rarely" "Often" "Never" "Always"   
##   
## $Q4b  
## [1] "Sometimes" "Rarely" "Always" "Often" "Never"   
##   
## $Q4c  
## [1] "Often" "Sometimes" "Always" "Rarely" "Never"   
##   
## $Q4d  
## [1] "Often" "Always" "Sometimes" "Rarely" "Never"   
##   
## $Q4e  
## [1] "Rarely" "Sometimes" "Often" "Never" "Always"   
##   
## $Q4f  
## [1] "Always" "Sometimes" "Often" "Never" "Rarely"   
##   
## $Q4g  
## [1] "Sometimes" "Rarely" "Never" "Always" "Often"   
##   
## $Q5  
## [1] "No" "Would prefer not to answer"  
## [3] "Not sure" "Yes"   
##   
## $Q6  
## [1] "No" "Not Sure" "Yes"   
##   
## $Q7a  
## [1] "Agree" "Disagree"   
## [3] "Strongly agree" "Neither agree nor disagree"  
## [5] "Strongly disagree"   
##   
## $Q7b  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Disagree" "Strongly agree"   
## [5] "Strongly disagree"   
##   
## $Q7c  
## [1] "Neither agree nor disagree" "Strongly disagree"   
## [3] "Agree" "Disagree"   
## [5] "Strongly agree"   
##   
## $Q7d  
## [1] "Neither agree nor disagree" "Strongly disagree"   
## [3] "Agree" "Disagree"   
## [5] "Strongly agree"   
##   
## $Q7e  
## [1] "Neither agree nor disagree" "Strongly disagree"   
## [3] "Agree" "Strongly agree"   
## [5] "Disagree"   
##   
## $Q8  
## [1] "Agree" "Strongly agree"   
## [3] "Disagree" "Neither agree nor disagree"  
## [5] "Strongly disagree"   
##   
## $Q9a  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Strongly agree" "Disagree"   
## [5] "Strongly disagree"   
##   
## $Q9b  
## [1] "Neither agree nor disagree" "Agree"   
## [3] "Disagree" "Strongly agree"   
## [5] "Strongly disagree"   
##   
## $Q9c  
## [1] "Agree" "Neither agree nor disagree"  
## [3] "Strongly agree" "Strongly disagree"   
## [5] "Disagree"   
##   
## $Q9d  
## [1] "Disagree" "Agree"   
## [3] "Strongly disagree" "Neither agree nor disagree"  
## [5] "Strongly agree"   
##   
## $Q9e  
## [1] "Disagree" "Strongly disagree"   
## [3] "Agree" "Neither agree nor disagree"  
## [5] "Strongly agree"   
##   
## $Q9f  
## [1] "Neither agree nor disagree" "Agree"   
## [3] "Strongly agree" "Disagree"   
## [5] "Strongly disagree"   
##   
## $Q9g  
## [1] "Neither agree nor disagree" "Agree"   
## [3] "Disagree" "Strongly agree"   
## [5] "Strongly disagree"   
##   
## $YEAR  
## [1] "2014" "2020"

# Exporting the cleaned datafiles

write.csv(common\_questions\_merged, 'common\_questions\_merged.csv')

write.csv(common\_questions\_2014, 'common\_questions\_2014.csv')  
write.csv(common\_questions\_2020, 'common\_questions\_2020.csv')

write.csv(common\_questions\_merged\_fillmode, 'common\_questions\_merged\_fillmode.csv')  
write.csv(common\_questions\_2014\_fillmode, 'common\_questions\_2014\_fillmode.csv')  
write.csv(common\_questions\_2020\_fillmode, 'common\_questions\_2020\_fillmode.csv')

library(rcompanion)

## Warning: package 'rcompanion' was built under R version 4.0.5

# Reading the dataset

survey\_df <- read.csv('common\_questions\_merged\_fillmode.csv')

# Recoding to convert categorical data to numerical

recode\_type1 <- function(vec){  
 return(recode(vec, 'Always' = 1,'Often' = 2, 'Sometimes' = 3,'Rarely' = 4,'Never' = 5))  
}  
recode\_type2 <- function(vec){  
 return(recode(vec, 'Strongly agree' = 1,'Agree' = 2, 'Neither agree nor disagree' = 3,'Disagree' = 4,'Strongly disagree' = 5))  
}  
survey\_df\_cde <- survey\_df  
survey\_df\_cde[,c(6,7,8,9,10,11,12)] <- sapply(survey\_df[,c(6,7,8,9,10,11,12)], recode\_type1)  
  
survey\_df\_cde[,c(15,16,17,18,19,20,21,22,23,24,25,26,27)] <- sapply(survey\_df[,c(15,16,17,18,19,20,21,22,23,24,25,26,27)], recode\_type2)

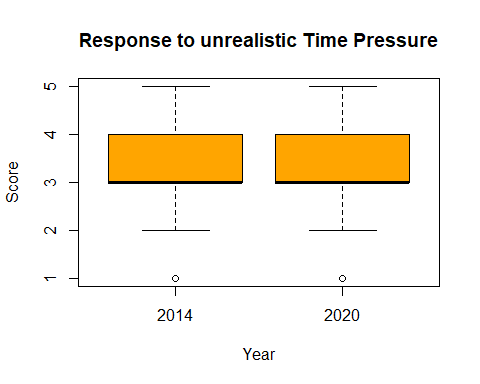
# Inferential Analysis

## Two Sample Wilcoxon Rank Test Unpaired

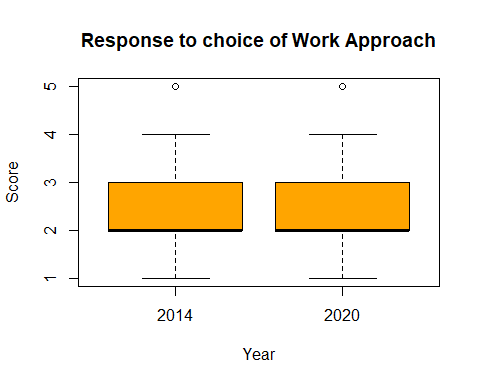
group\_by(survey\_df\_cde, YEAR) %>%  
 summarise(  
 count = n(),  
 median = median(Q7b, na.rm = TRUE),  
 IQR = IQR(Q4a, na.rm = TRUE)  
 )

## # A tibble: 2 x 4  
## YEAR count median IQR  
## <int> <int> <dbl> <dbl>  
## 1 2014 99392 2 1  
## 2 2020 108085 2 1

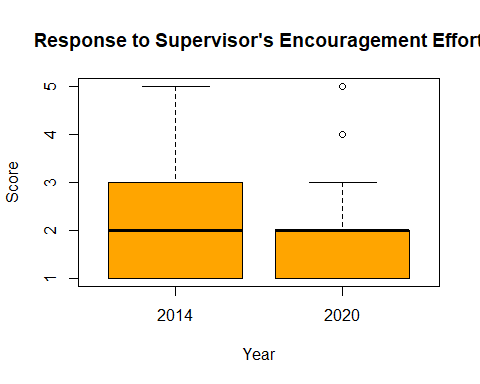
boxplot(survey\_df\_cde$Q4a ~ survey\_df\_cde$YEAR, col="orange", main="Response to unrealistic Time Pressure", ylab="Score", xlab="Year")



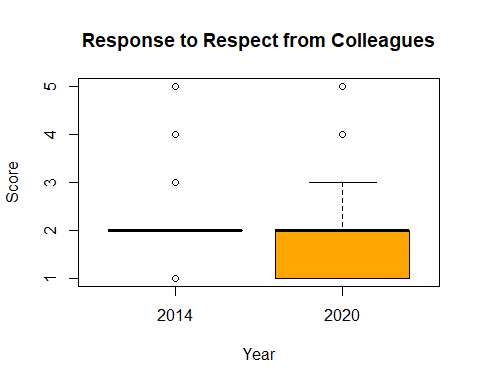
boxplot(survey\_df\_cde$Q4b ~ survey\_df\_cde$YEAR, col="orange", main="Response to choice of Work Approach", ylab="Score", xlab="Year")



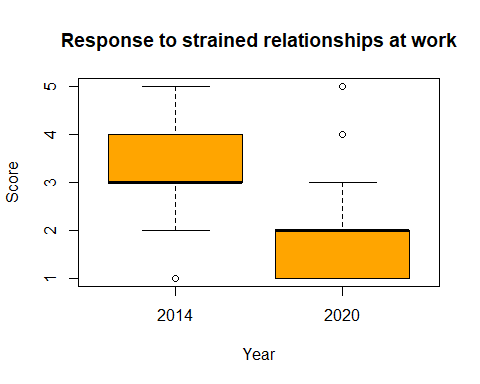
boxplot(survey\_df\_cde$Q4c ~ survey\_df\_cde$YEAR, col="orange", main="Response to Supervisor's Encouragement Effort", ylab="Score", xlab="Year")



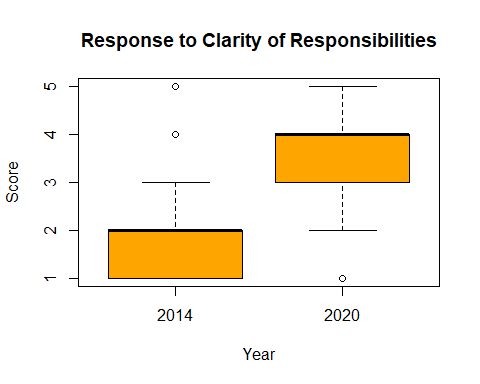
boxplot(survey\_df\_cde$Q4d ~ survey\_df\_cde$YEAR, col="orange", main="Response to Respect from Colleagues", ylab="Score", xlab="Year")



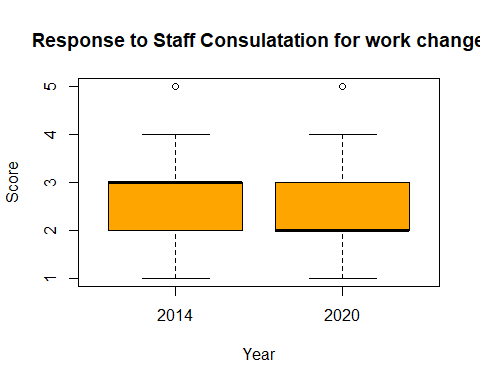
boxplot(survey\_df\_cde$Q4e ~ survey\_df\_cde$YEAR, col="orange", main="Response to strained relationships at work", ylab="Score", xlab="Year")



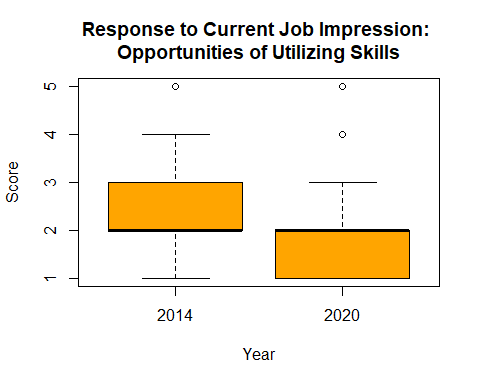
boxplot(survey\_df\_cde$Q4f ~ survey\_df\_cde$YEAR, col="orange", main="Response to Clarity of Responsibilities", ylab="Score", xlab="Year")



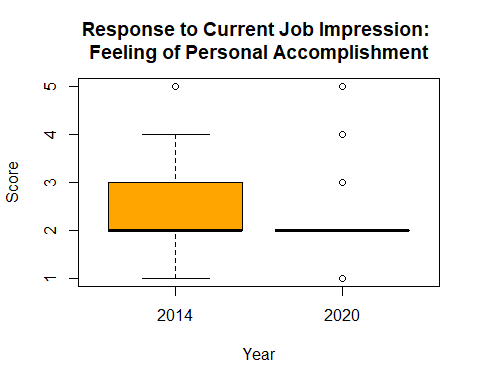
boxplot(survey\_df\_cde$Q4g ~ survey\_df\_cde$YEAR, col="orange", main="Response to Staff Consulatation for work change", ylab="Score", xlab="Year")



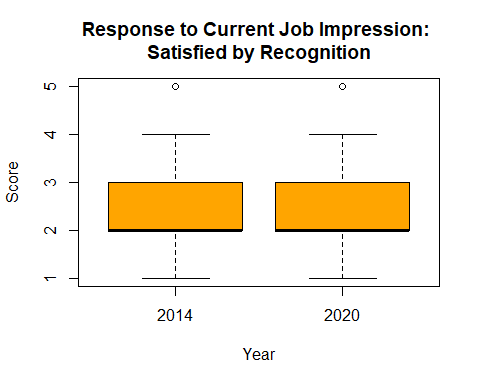
boxplot(survey\_df\_cde$Q7a ~ survey\_df\_cde$YEAR, col="orange", main="Response to Current Job Impression: \nOpportunities of Utilizing Skills", ylab="Score", xlab="Year")



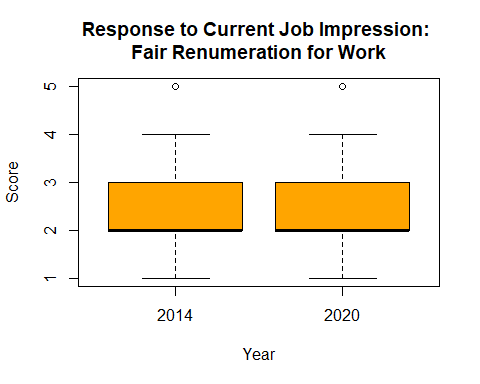
boxplot(survey\_df\_cde$Q7b ~ survey\_df\_cde$YEAR, col="orange", main="Response to Current Job Impression: \nFeeling of Personal Accomplishment", ylab="Score", xlab="Year")



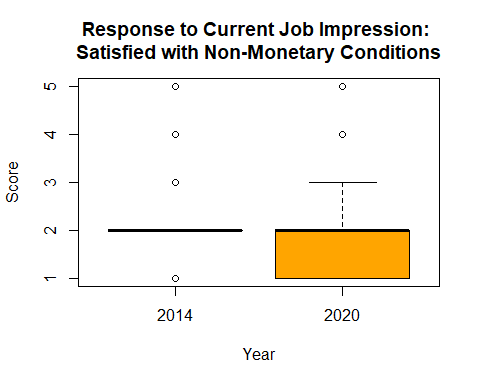
boxplot(survey\_df\_cde$Q7c ~ survey\_df\_cde$YEAR, col="orange", main="Response to Current Job Impression: \nSatisfied by Recognition", ylab="Score", xlab="Year")



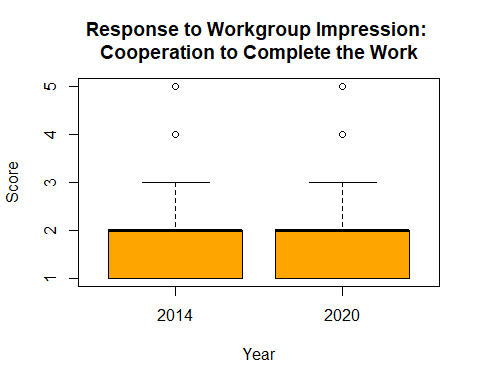
boxplot(survey\_df\_cde$Q7d ~ survey\_df\_cde$YEAR, col="orange", main="Response to Current Job Impression: \nFair Renumeration for Work", ylab="Score", xlab="Year")



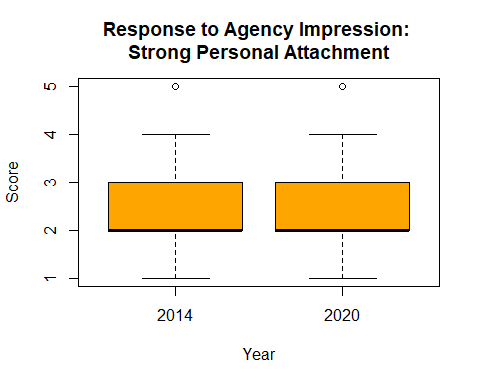
boxplot(survey\_df\_cde$Q7e ~ survey\_df\_cde$YEAR, col="orange", main="Response to Current Job Impression: \nSatisfied with Non-Monetary Conditions", ylab="Score", xlab="Year")



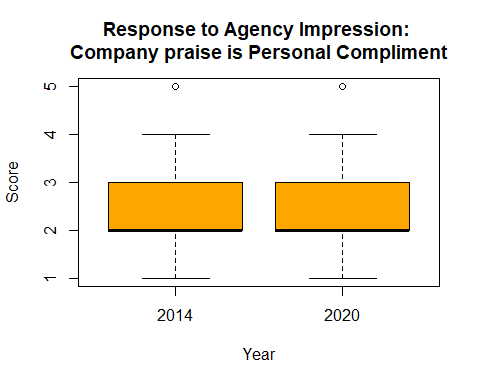
boxplot(survey\_df\_cde$Q8 ~ survey\_df\_cde$YEAR, col="orange", main="Response to Workgroup Impression: \nCooperation to Complete the Work", ylab="Score", xlab="Year")



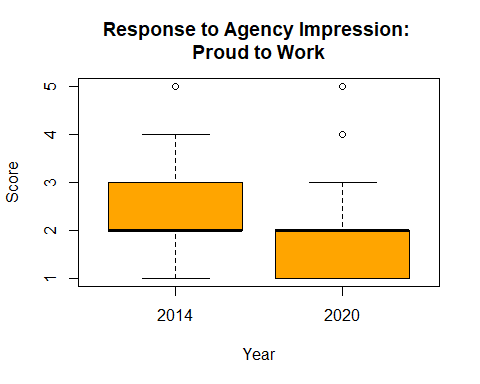
boxplot(survey\_df\_cde$Q9a ~ survey\_df\_cde$YEAR, col="orange", main="Response to Agency Impression: \nStrong Personal Attachment", ylab="Score", xlab="Year")



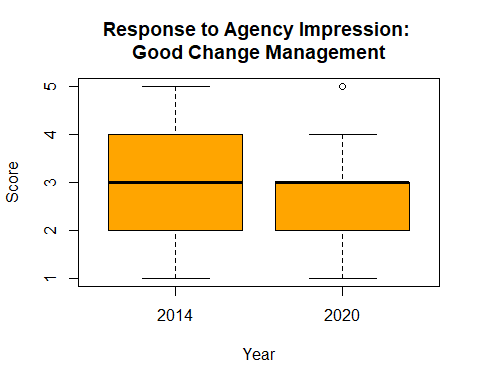
boxplot(survey\_df\_cde$Q9b ~ survey\_df\_cde$YEAR, col="orange", main="Response to Agency Impression: \nCompany praise is Personal Compliment", ylab="Score", xlab="Year")



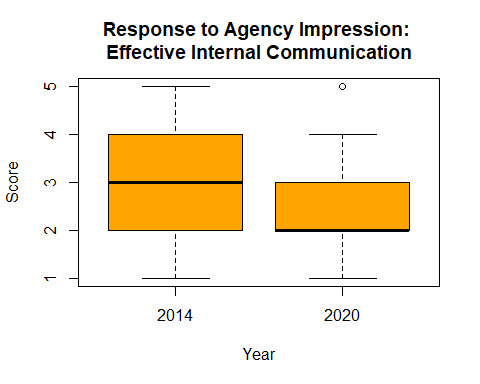
boxplot(survey\_df\_cde$Q9c ~ survey\_df\_cde$YEAR, col="orange", main="Response to Agency Impression: \nProud to Work", ylab="Score", xlab="Year")



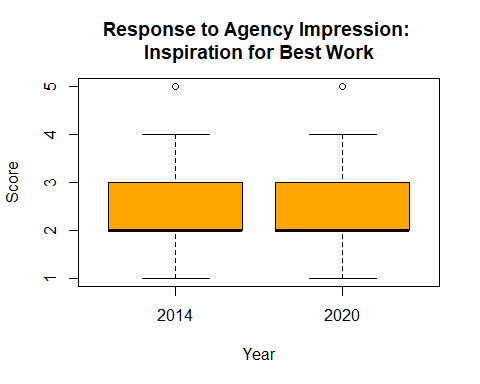
boxplot(survey\_df\_cde$Q9d ~ survey\_df\_cde$YEAR, col="orange", main="Response to Agency Impression: \nGood Change Management", ylab="Score", xlab="Year")



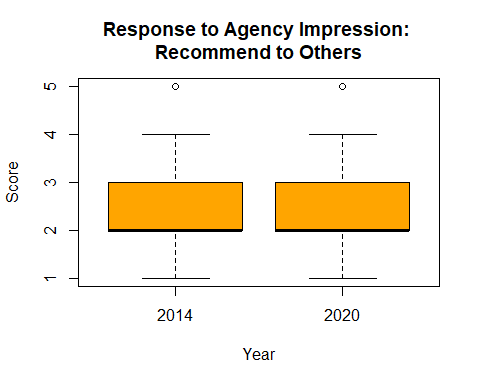
boxplot(survey\_df\_cde$Q9e ~ survey\_df\_cde$YEAR, col="orange", main="Response to Agency Impression: \nEffective Internal Communication", ylab="Score", xlab="Year")



boxplot(survey\_df\_cde$Q9f ~ survey\_df\_cde$YEAR, col="orange", main="Response to Agency Impression: \nInspiration for Best Work", ylab="Score", xlab="Year")



boxplot(survey\_df\_cde$Q9g ~ survey\_df\_cde$YEAR, col="orange", main="Response to Agency Impression: \nRecommend to Others", ylab="Score", xlab="Year")



# Running Wilcoxon test for all the numerical variable

wilcox\_all <- function(df){  
 print(wilcox.test(Q4a~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q4b~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q4c~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q4d~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q4e~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q4f~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q4g~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q7a~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q7b~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q7c~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q7d~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q7e~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q8~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q9a~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q9b~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q9c~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q9d~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q9e~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q9f~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
 print(wilcox.test(Q9g~YEAR, data = df, paired = FALSE, exact = FALSE, conf.int = TRUE, conf.level = 0.95, correct=FALSE))  
}  
  
wilcox\_all(survey\_df\_cde[,-c(1,2,3,4,5,13,14)])

##   
## Wilcoxon rank sum test  
##   
## data: Q4a by YEAR  
## W = 5328816114, p-value = 0.0008502  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## -4.524300e-05 -2.492267e-05  
## sample estimates:  
## difference in location   
## -3.054618e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q4b by YEAR  
## W = 5410533716, p-value = 0.001967  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 6.683580e-06 6.409354e-05  
## sample estimates:  
## difference in location   
## 4.526192e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q4c by YEAR  
## W = 6181725889, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 2.208900e-05 8.101977e-06  
## sample estimates:  
## difference in location   
## 5.26544e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q4d by YEAR  
## W = 6013796962, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 5.463852e-06 2.519658e-05  
## sample estimates:  
## difference in location   
## 2.105546e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q4e by YEAR  
## W = 9332253495, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 1.999967 1.999977  
## sample estimates:  
## difference in location   
## 1.999967   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q4f by YEAR  
## W = 1218442222, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## -1.999953 -1.999950  
## sample estimates:  
## difference in location   
## -1.999998   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q4g by YEAR  
## W = 5441779428, p-value = 5.836e-08  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 8.866490e-05 1.119056e-05  
## sample estimates:  
## difference in location   
## 3.330009e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q7a by YEAR  
## W = 6240879838, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 3.836304e-05 2.969371e-05  
## sample estimates:  
## difference in location   
## 6.869276e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q7b by YEAR  
## W = 6197707506, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 6.715917e-05 4.156719e-05  
## sample estimates:  
## difference in location   
## 4.282265e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q7c by YEAR  
## W = 5992816496, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 6.793402e-05 2.308333e-05  
## sample estimates:  
## difference in location   
## 2.135719e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q7d by YEAR  
## W = 5366866278, p-value = 0.7192  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## -3.862921e-05 4.019697e-05  
## sample estimates:  
## difference in location   
## -5.346912e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q7e by YEAR  
## W = 5364548988, p-value = 0.5785  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## -2.838618e-05 2.752715e-05  
## sample estimates:  
## difference in location   
## -6.536739e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q8 by YEAR  
## W = 5797881083, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 2.828544e-05 1.873142e-05  
## sample estimates:  
## difference in location   
## 1.262344e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q9a by YEAR  
## W = 5721169456, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 1.518958e-05 5.557266e-05  
## sample estimates:  
## difference in location   
## 1.488807e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q9b by YEAR  
## W = 5718528583, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 2.784127e-06 4.425335e-07  
## sample estimates:  
## difference in location   
## 4.364058e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q9c by YEAR  
## W = 6022353711, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 4.058663e-05 5.443843e-05  
## sample estimates:  
## difference in location   
## 3.311429e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q9d by YEAR  
## W = 6166387075, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 9.487654e-05 9.414594e-07  
## sample estimates:  
## difference in location   
## 8.933924e-06   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q9e by YEAR  
## W = 6.426e+09, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 3.178384e-05 4.709308e-05  
## sample estimates:  
## difference in location   
## 4.908655e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q9f by YEAR  
## W = 5845098434, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 3.863296e-05 2.705375e-05  
## sample estimates:  
## difference in location   
## 3.001181e-05   
##   
##   
## Wilcoxon rank sum test  
##   
## data: Q9g by YEAR  
## W = 6.115e+09, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## 1.254884e-06 5.765898e-05  
## sample estimates:  
## difference in location   
## 1.426567e-05

## Chi-Square Test

chi\_all <- function(df){  
 for (col in colnames(df)){  
 if (col == 'YEAR'){  
 next  
 }  
 cont\_table <- table(df[[col]], df$YEAR)  
 print(cont\_table)  
 print(chisq.test(cont\_table))  
 print(paste("Crammer's V : ", cramerV(cont\_table)))  
 print('=================================================')  
 }  
}

# Running Chi Square Test for the categorical variables

chi\_all(survey\_df\_cde[,c(2,3,4,5,13,14,28)])

##   
## 2014 2020  
## Large (1,001 or more employees) 86884 93586  
## Medium (251 to 1,000 employees) 8884 9998  
## Small (Less than 250 employees) 3624 4501  
##   
## Pearson's Chi-squared test  
##   
## data: cont\_table  
## X-squared = 45.128, df = 2, p-value = 1.587e-10  
##   
## [1] "Crammer's V : 0.01475"  
## [1] "================================================="  
##   
## 2014 2020  
## Female 56250 64494  
## Male 41047 39875  
## X 2095 3716  
##   
## Pearson's Chi-squared test  
##   
## data: cont\_table  
## X-squared = 668.98, df = 2, p-value < 2.2e-16  
##   
## [1] "Crammer's V : 0.05678"  
## [1] "================================================="  
##   
## 2014 2020  
## 40 to 54 years 42924 47638  
## 55 years or older 14492 18899  
## Under 40 years 37320 41544  
## Unknown 4656 4  
##   
## Pearson's Chi-squared test  
##   
## data: cont\_table  
## X-squared = 5342.4, df = 3, p-value < 2.2e-16  
##   
## [1] "Crammer's V : 0.1605"  
## [1] "================================================="  
##   
## 2014 2020  
## EL 29547 32755  
## SES 2102 2635  
## Trainee/Graduate/APS 67743 72695  
##   
## Pearson's Chi-squared test  
##   
## data: cont\_table  
## X-squared = 35.607, df = 2, p-value = 1.854e-08  
##   
## [1] "Crammer's V : 0.0131"  
## [1] "================================================="  
##   
## 2014 2020  
## No 89282 97624  
## Not sure 4733 4465  
## Would prefer not to answer 3002 2348  
## Yes 2375 3648  
##   
## Pearson's Chi-squared test  
##   
## data: cont\_table  
## X-squared = 365.55, df = 3, p-value < 2.2e-16  
##   
## [1] "Crammer's V : 0.04197"  
## [1] "================================================="  
##   
## 2014 2020  
## No 78661 88715  
## Not Sure 5607 7241  
## Yes 15124 12129  
##   
## Pearson's Chi-squared test  
##   
## data: cont\_table  
## X-squared = 778.02, df = 2, p-value < 2.2e-16  
##   
## [1] "Crammer's V : 0.06124"  
## [1] "================================================="