R- correlation analysis

# This is a collection of codes that we used to run correlation analyses

The feature of correlation analysis has been rolled out only to R so far and is to be transferred to tableau in a later stage of the project. Please find here the codes and approach used to analyse the data

## Required Packages

install.packages("dplyr")  
install.packages("ggplot2")  
  
library("data.table")  
library("readxl")  
library("dplyr")  
library("ggplot2")

## Load data into R

setwd("C:\\Users\\janho\\OneDrive\\TechChallenge\_Connection People\\Prototype\\Tableau Map\\clean datasets")  
getwd()  
setwd("C:\\Users\\janho\\OneDrive\\TechChallenge\_Connection People\\Prototype\\Tableau Map")  
data\_excel <- read\_xlsx("clean\_MR\_2016.xlsx")  
Men\_dt <- as.data.table(data\_excel)  
View(Men\_dt)

## Codes used to adjust data and make it usable

Men\_dt$`m number of trips away from home`[Men\_dt$`m number of trips away from home` == "None"] <- "0"  
# change to numeric creates some NAs   
Men\_dt$`m number of trips away from home` <- as.numeric(Men\_dt$`m number of trips away from home`)  
str(Men\_dt$`m number of trips away from home`)  
# remove NAs from the column  
Clean\_Men<- Men\_dt[!is.na(Men\_dt$`m number of trips away from home`), ]  
# check if NA in column  
sum(is.na(Men\_dt\_1\_rmNA$m\_number\_of\_trips\_away\_from\_home))

## Make “wealth urban rural” into quintiles and “education” into terciles

Clean\_Men$`m wealth urban rural`[Clean\_Men$`m wealth urban rural`== "Poorest"] <- "1"  
Clean\_Men$`m wealth urban rural`[Clean\_Men$`m wealth urban rural`== "Poorer"] <- "2"  
Clean\_Men$`m wealth urban rural`[Clean\_Men$`m wealth urban rural`== "Middle"] <- "3"  
Clean\_Men$`m wealth urban rural`[Clean\_Men$`m wealth urban rural`== "Richer"] <- "4"  
Clean\_Men$`m wealth urban rural`[Clean\_Men$`m wealth urban rural`== "Richest"] <- "5"  
Clean\_Men$`m wealth urban rural` <- as.numeric(Clean\_Men$`m wealth urban rural`)  
  
# Education  
Clean\_Men$`education level highest`[Clean\_Men$`education level highest`== "No education"] <- "0"  
Clean\_Men$`education level highest`[Clean\_Men$`education level highest`== "Primary"] <- "1"  
Clean\_Men$`education level highest`[Clean\_Men$`education level highest`== "Secondary"] <- "2"  
Clean\_Men$`education level highest`[Clean\_Men$`education level highest`== "Higher"] <- "3"

## Correlation tests

# Number of wives  
Clean\_Men$`number of wives` <- as.numeric(Clean\_Men$`number of wives`)  
  
cor.test(Clean\_Men$`m number of trips away from home`, Clean\_Men$`number of wives`  
 , method="pearson")  
View(Clean\_Men$`m # of trips away from home`)  
  
# Wealth  
cor.test(Clean\_Men$`m number of trips away from home`, Clean\_Men$`m wealth urban rural`  
 , method="pearson")  
# Education  
cor.test(Clean\_Men$`m number of trips away from home`, Clean\_Men$`education level highest`  
 , method="pearson")  
  
# m ownes telephone   
Clean\_Men$`m owns telephone`[Clean\_Men$`m owns telephone`== "No"] <- "0"  
Clean\_Men$`m owns telephone`[Clean\_Men$`m owns telephone`== "Yes"] <- "1"  
  
# change to numeric   
Clean\_Men$`m owns telephone` <- as.numeric(Clean\_Men$`m owns telephone`)  
  
# cor test  
cor.test(Clean\_Men$`m owns telephone`, Clean\_Men$m\_number\_of\_trips\_away\_from\_home  
 , method="pearson")

## Example Pattern plots

boxplot\_telephone <- ggplot(Clean\_Men, aes(factor(Clean\_Men$`m owns telephone`), Clean\_Men$`m number of trips away from home`)) + geom\_boxplot() + ylim(0,10) + labs(x= "Respondent owns a telephone", y="The number of trips away from home", title="Corrleation Analysis of Trips and Possession") + scale\_fill\_distiller(palette = "Blues")  
boxplot\_telephone  
  
  
p\_wife <- ggplot(Clean\_Men, aes(Clean\_Men$`m number of trips away from home`, Clean\_Men$`number of wives`)) + geom\_point() + labs(x="Number of trips away from home", y="number of wives", title="Correlation Analysis between number of wives and trips") + scale\_fill\_brewer(palette = "Blues")  
p\_wife

violinplot\_telephone <- ggplot(Clean\_Men, aes(factor(Clean\_Men$`m owns telephone`), Clean\_Men$`m number of trips away from home`)) + geom\_violin()+ ylim(0,10) + labs(x= "Respondent owns a telephone", y="The number of trips away from home", title="Corrleation Analysis of Trips and Possession")  
violinplot\_telephone

