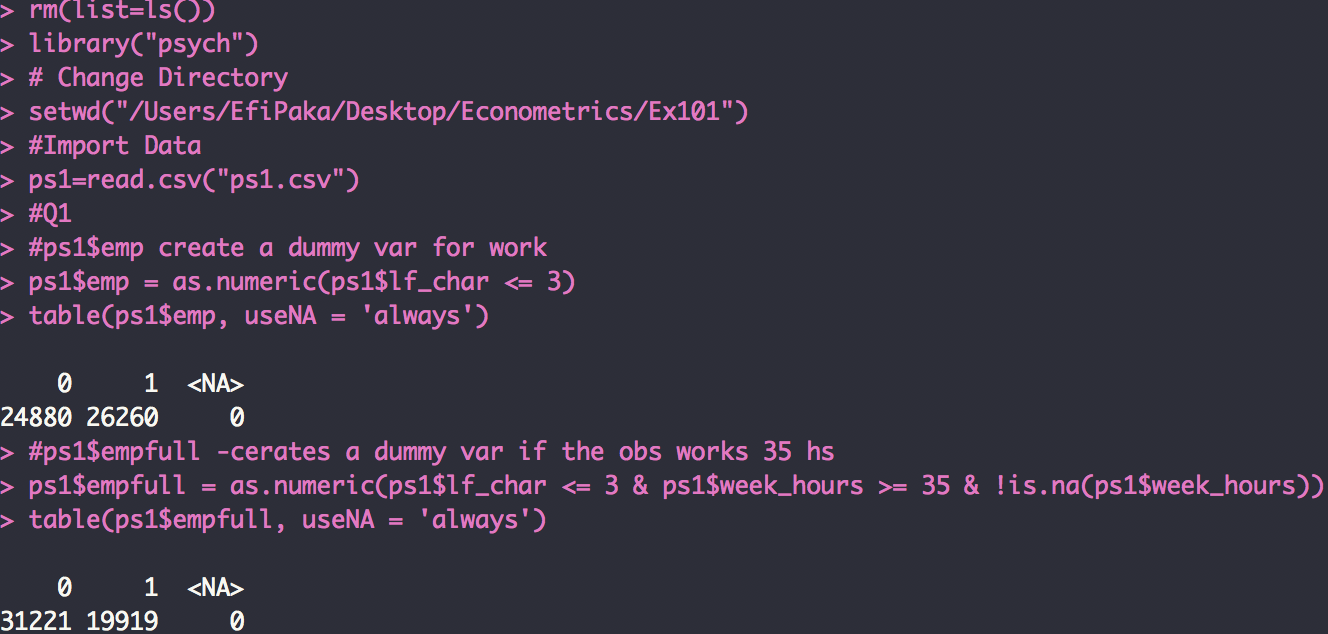
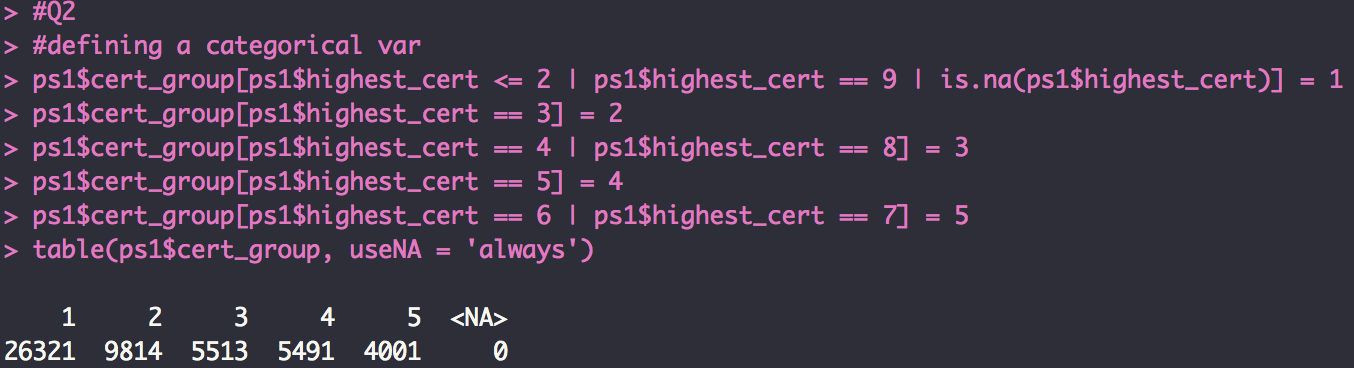
**תרגיל 1- יישומים אקונומטריים**

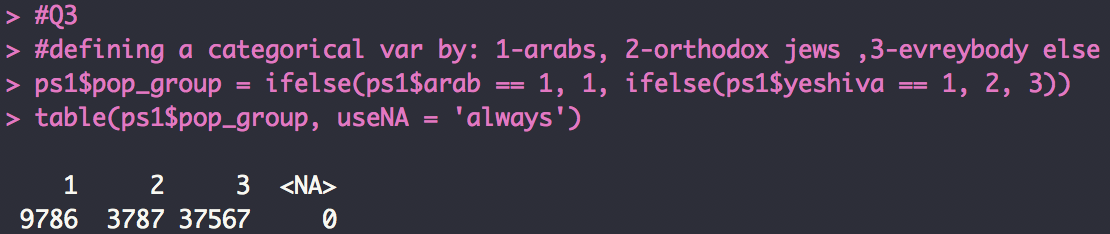
1. נקרא את קובץ ה-CSV ונקבל מסד נתונים עם 51,140 תצפיות ו-16 משתנים.
2. נגדיר משתנה דמי עבור עבודה – employment ע"פ המשתנה labore force (LF).
3. נגדיר גם משתנה דמי עבור עבודה – employment ע"פ המשתנה labore force (LF).



1. נגדיר משתנה קטגורי ל-5 קבוצות השכלה:

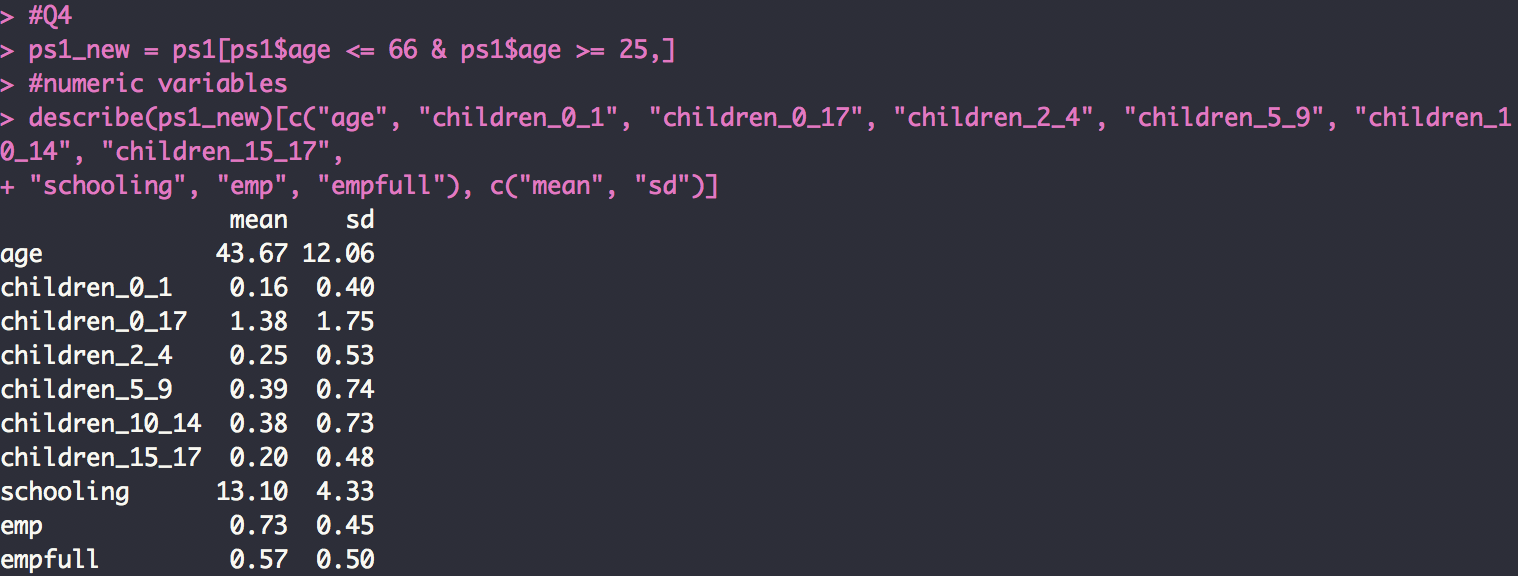


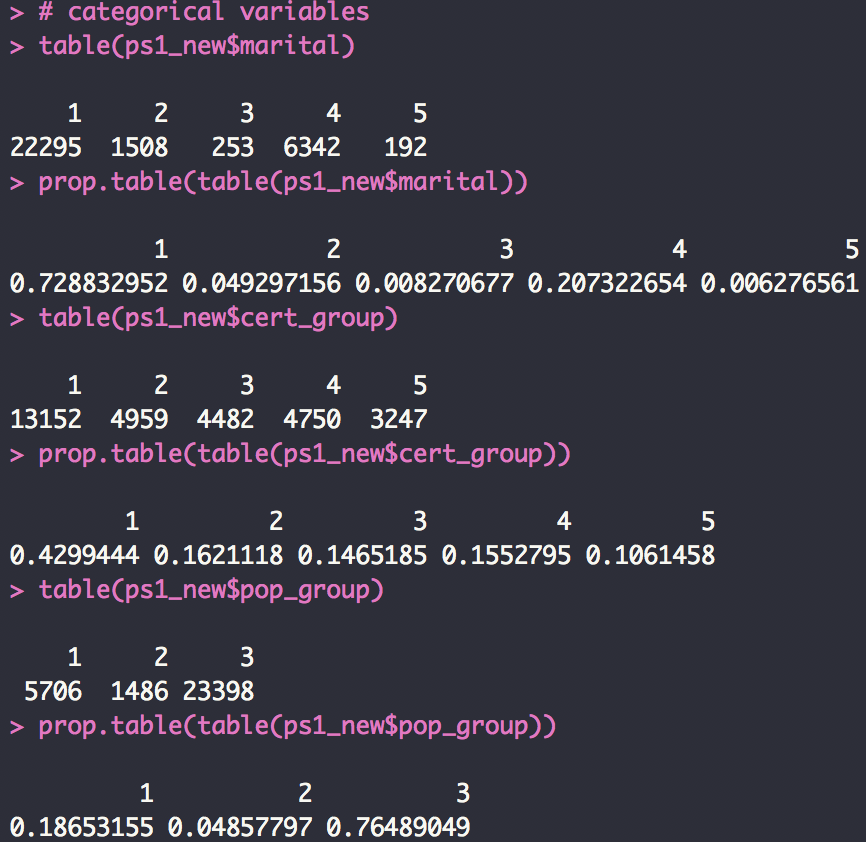
1. נסווג את הגברים במדגם ל-3 קבוצות ע"פ מגזר ונקבל:



1. נבנה מסד נתונים חדש המורכב מגילאי 25-66, שהם כפי שלמדנו מוגדרים כגילאי העבודה במשק:



* נבדוק תיאור סטטיסטי של המשתנים הכמותיים ונקבל:  
  
* נציג גם טבלת התפלגויות עבור המשתנים הקטגוריים ונקבל:



1. נציג את שיעור התעסוקה, ושיעור התעסוקה המלאה כפונקציה של הגיל בצורה גרפית:



1. נציג את שיעור התעסוקה כפונקציה של הגיל וההשכלה בצורה גרפית:



1. נציג את שיעור התעסוקה, ותעסוקה המלאה פונקציה של הגיל ומגזר בצורה גרפית:





1. הממצאים המרכזיים העולים מהסעיפים הם כדלהלן:

* מסעיף 5 עולה כי שיעור התעסוקה המלאה נמוך משיעור התעסוקה הכללי באוכלוסיה לכל גיל וגיל, ניתן לראות בנוסף כי ששיעור התעסוקה וששיעור התעסוקה במשרה מלאה עולים עד גיל 40 לערך ולאחר מכן נמצאים במגמת ירידה עד גיל הפרישה לגברים.

נוכל לראות כי ההפרש בין הגרפים משתמר לאורך כלל הגילאים, הם נעים בתיאום, כלומר ניתן להסיק כי שיעור המועסקים במשרה מלאה נותר קבוע ביחד לשיעור המועסקים הכללי.

* מסעיף 6 עולה שוב אותה מסקנה כי שיעור התעסוקה עולה עד גיל 40 ויורד עד גיל הפרישה.  
  קבוצת האוכלוסיה ללא תעודת הבגרות נמצאים בתחתית מבחינת שיעור התעסוקה בקרבם.  
  ניתן לראות שקבוצת האוכלוסיה בעלת התואר השני נמצאים בתחילה בשיעור תעסוקה נמוך ולאחר מכן נמצאים בחלקו העליון של שיעור התעסוקה באוכלוסיה, זאת משום שכנראה במצב התחילי הם טרם השלימו לימודיהם ואינם נמנים כחלק משוק העבודה ולאחר השלמת התואר שיעור התעסוקה בקרבם הוא הגבוה באוכלוסיה, כך גם בעלי תעודות הבגרות בלבד, ניתו להסיק, כי אלה, ככל הנראה, מוכנים לעבוד בכל עבודה ללא קשר לשכרם שכן בעלי תואר ראשון נדמה כי מתנהגים בצורה יותר בררנית ולכן שיעור התעסוקה בקרבם נראה מפתיע לכאורה ונמוך מבעלי ההשכלה הפחותה מהם.
* מסעיף 7 עולה כי במגזר החרדי ב-2 הגרפים בתחילת חייהם שיעור התעסוקה הוא הנמוך ביותר בקרבם, נוכל להסיק כי הדבר מתרחש מכיוון שעדיין בשלב זה הם אינם נשואים, לאחר מכן כאשר הם נישאים ומביאים ילדים הם מרגישים "עניים יותר" ולכן יוצאים לעבוד בשיעורים גבוהים יותר.  
  המגזר "האחר" גבוה יותר מ-2 הסקטורים האחרים, נוכל להסיק כי הדבר נובע עקב השכלתם.  
  נוכל לראות כי קבוצת ה"אחרים" מתנהגת בצורה יותר מסודרת וכצפוי ביחס למגזרים הפחות משכילים שמתנהגים בצורה יותר תנודתית בכל גיל.  
  כמו כן עבור כלל הגילאים שיעור התעסוקה יורד לקראת גיל הפרישה אליו הם מתכנסים.

**הסקריפט:**

install.packages("openxlsx")

library("openxlsx")

library('psych')

library(ggplot2)

install.packages("visualize")

library(visualize)

# Clear Objects

rm(list=ls())

#removing specific object

#rm( x )

# Change Directory

setwd("/Users/EfiPaka/Desktop/Econometrics/Ex101")

#Import Data

ps1=read.csv("ps1.csv")

#Q1

#ps1$emp -creates a dummy var for work

ps1$emp = as.numeric(ps1$lf\_char <= 3)

table(ps1$emp, useNA = 'always')

#ps1$empfull -cerates a dummy var if the obs works 35 hs

ps1$empfull = as.numeric(ps1$lf\_char <= 3 & ps1$week\_hours >= 35 & !is.na(ps1$week\_hours))

table(ps1$empfull, useNA = 'always')

#Q2

#defining a categorical var

ps1$cert\_group[ps1$highest\_cert <= 2 | ps1$highest\_cert == 9 | is.na(ps1$highest\_cert)] = 1

ps1$cert\_group[ps1$highest\_cert == 3] = 2

ps1$cert\_group[ps1$highest\_cert == 4 | ps1$highest\_cert == 8] = 3

ps1$cert\_group[ps1$highest\_cert == 5] = 4

ps1$cert\_group[ps1$highest\_cert == 6 | ps1$highest\_cert == 7] = 5

table(ps1$cert\_group, useNA = 'always')

#Q3

#defining a categorical var by: 1-arabs, 2-orthodox jews ,3-evreybody else

ps1$pop\_group = ifelse(ps1$arab == 1, 1, ifelse(ps1$yeshiva == 1, 2, 3))

table(ps1$pop\_group, useNA = 'always')

#Q4

ps1\_new = ps1[ps1$age <= 66 & ps1$age >= 25,]

describe(ps1\_new)

#numeric variables

describe(ps1\_new)[c("age", "children\_0\_1", "children\_0\_17", "children\_2\_4", "children\_5\_9",

"children\_10\_14", "children\_15\_17","schooling", "emp", "empfull"), c("mean", "sd")]

# categorical variables

table(ps1\_new$marital)

prop.table(table(ps1\_new$marital))

table(ps1\_new$cert\_group)

prop.table(table(ps1\_new$cert\_group))

table(ps1\_new$pop\_group)

prop.table(table(ps1\_new$pop\_group))

#Q5

#add a var of the avg employment by age

ps1\_new$avgEmpByAge = ave(ps1\_new$emp, ps1\_new$age)

#add a var of the avg full employment (>35 hs) rate by age

ps1\_new$avgFullEmpByAge = ave(ps1\_new$empfull, ps1\_new$age)

#order them out

ps1\_new = ps1\_new[order(ps1\_new$age), ]

avgFullEmpByAge = aggregate(ps1\_new[c("emp", "empfull")], by=list(age = ps1\_new$age), FUN=mean)

plot(ps1\_new$age, ps1\_new$avgEmpByAge, pch=18,type="l", ylim=c(0,1), main ="Employment Rate by Age" , xlab="Age", ylab="Employment Rate",lty=2,cex=0.5,col="blue")

lines(ps1\_new$age, ps1\_new$avgFullEmpByAge, pch=18,type="l", col="red")

# Add a legend

legend(48, 0.2, legend=c("employment rate", "full employment rate (>35 hs)"),

col=c("blue", "red"), lty=2:1, cex=0.8,text.font=4, bg='lightblue')

#Q6

AvgEmpAgeCert = aggregate(ps1\_new["emp"], by=list(age = ps1\_new$age, cert\_group = ps1\_new$cert\_group), FUN=mean)

AvgEmpAgeCert = reshape(AvgEmpAgeCert, v.names = "emp", idvar = "age", timevar = "cert\_group", direction = "wide")

plot(AvgEmpAgeCert$age, AvgEmpAgeCert$emp.1, type = "l", ylim = c(0.2,1),main ="Employment Rate by Age and by Education", xlab = "Age", ylab = "Rate", col= "darkslateblue", lwd=2)

lines(AvgEmpAgeCert$age, AvgEmpAgeCert$emp.2, type = "l", col= "firebrick1")

lines(AvgEmpAgeCert$age, AvgEmpAgeCert$emp.3, type = "l", col= "dodgerblue4")

lines(AvgEmpAgeCert$age, AvgEmpAgeCert$emp.4, type = "l", col= "forestgreen")

lines(AvgEmpAgeCert$age, AvgEmpAgeCert$emp.5, type = "l", col = "darkorange" ,lwd=1)

legend(55, 0.35, legend=c("Secondary school - ", "Matriculation", "Non academic","B.A", "M.A +"),col= c("darkslateblue", "firebrick1", "dodgerblue4", "forestgreen", "darkorange") ,lty=1,cex=0.5,bg='lightblue',text.font=4)

#Q7

week\_hour\_age <- aggregate(ps1\_new["empfull"], by = list(age = ps1\_new$age, hour\_full= ps1\_new$pop\_group), FUN = mean)

week\_hour\_age <-week\_hour\_age[order(week\_hour\_age$age),]

week\_hour\_age <- reshape(week\_hour\_age, v.names = "empfull", idvar = "age", timevar = "hour\_full", direction= "wide")

#create a graph

plot(week\_hour\_age$age, week\_hour\_age$empfull.1, type = "l",main ="Full Employment Rate by Age and by Pop.Sector" , ylim = c(0,1), xlab = "Age", ylab = "Rate",col= "darkslateblue",lwd=3)

lines(week\_hour\_age$age, week\_hour\_age$empfull.2 , type = "l", col= "firebrick1",lwd=3)

lines(week\_hour\_age$age, week\_hour\_age$empfull.3, type = "l", col= "darkorange",lwd=3)

legend("topright", legend=c("Arab sector", "Orthodox Jews", "Others"),col = c("darkslateblue", "firebrick1", "darkorange") ,lty=1,cex=0.8, bg='lightblue',lwd=3)

pop\_age\_cert <- aggregate(ps1\_new["emp"], by = list(age = ps1\_new$age, pop\_group= ps1\_new$pop\_group), FUN = mean)

pop\_age\_cert <- reshape(pop\_age\_cert, v.names = "emp", idvar = "age", timevar = "pop\_group", direction= "wide")

plot(pop\_age\_cert$age, pop\_age\_cert$emp.1, type = "l",main ="Employment Rate by Age and by Pop.Sector", ylim = c(0,1), xlab = "Age", ylab = "Rate",col= "darkslateblue", lwd=3)

lines(pop\_age\_cert$age, pop\_age\_cert$emp.2, type = "l", col= "firebrick1", lwd=2)

lines(pop\_age\_cert$age, pop\_age\_cert$emp.3, type = "l", col= "darkorange", lwd=2)

legend("topright", legend=c("Arab sector", "Orthodox Jews", "Others"),col = c("darkslateblue", "firebrick1", "darkorange") ,lty=1,cex=0.5,bg='lightblue',text.font=4)