9/11/2020

# loading and reading my researchdata  
  
data.gaain<-read.csv("researchdata/gaain/gaain.csv", header=T,sep = ',')  
  
data.mimic<-read.csv("researchdata/mimic/mimic.csv", header=T,sep = ',')  
  
data.inpc<-read.csv("researchdata/inpc/inpc.csv", header=T,sep = ',')  
  
#2. Import ps1.csv (Canvas ??? Files??? dataSets) into R and answer the following questions. (10 points)  
  
# \*\*\* All data are kept under respective folders into the working directory in R studio \*\*\*  
  
# Reading ps1.csv file  
  
data.ps1<-read.csv("datastore/ps1.csv", header=T,sep = ',')  
  
# installing package dplyr  
  
#install.packages("dplyr")  
  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

#a. Please use the glimpse() function to view the data. How is this function useful?  
# Using glimpse function  
glimpse(data.ps1)

## Rows: 30  
## Columns: 5  
## $ ID <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17...  
## $ gender <chr> "male", "male", "female", "female", "male", "male", "male...  
## $ education <int> 15, 16, 12, 8, 15, 15, 15, 12, 15, 12, 16, 8, 15, 15, 12,...  
## $ ethnicity <chr> "African American", "Asian American", "Caucasian", "Hispa...  
## $ salary <int> 55000, 40000, 20000, 20000, 45000, 30000, 35000, 20000, 2...

#b. Please show the statistics for only females. To do this, you need to determine which rows describe data from females (use a logical).  
#Pull out these rows and put them in a new variable.  
#This can/should be be done with 2-3 lines of code, by using logical indexing.   
  
# Working on gender variable  
ps1.gender<-data.ps1$gender  
  
# Working on ethnicity variable  
ps1.ethnicity<-data.ps1$ethnicity  
  
  
# Statistics for only females  
  
ps1.female.stat<-summary(data.ps1[which(ps1.gender=='female'),])  
ps1.female.stat

## ID gender education ethnicity   
## Min. : 3.00 Length:12 Min. : 8.00 Length:12   
## 1st Qu.: 8.75 Class :character 1st Qu.:12.00 Class :character   
## Median :12.50 Mode :character Median :13.50 Mode :character   
## Mean :14.33 Mean :13.33   
## 3rd Qu.:21.50 3rd Qu.:15.00   
## Max. :25.00 Max. :16.00   
## salary   
## Min. :16950   
## 1st Qu.:20000   
## Median :20000   
## Mean :23913   
## 3rd Qu.:26250   
## Max. :35000

#c. Please show the statistic for all Hispanic Females. Hint: use the & command  
  
# Statistic for all Hispanic Females  
  
ps1.hisp.female.stat<-summary(data.ps1[which(ps1.gender=='female' & ps1.ethnicity=='Hispanic'),])  
ps1.hisp.female.stat

## ID gender education ethnicity   
## Min. : 4.0 Length:2 Min. : 8 Length:2   
## 1st Qu.: 5.5 Class :character 1st Qu.: 9 Class :character   
## Median : 7.0 Mode :character Median :10 Mode :character   
## Mean : 7.0 Mean :10   
## 3rd Qu.: 8.5 3rd Qu.:11   
## Max. :10.0 Max. :12   
## salary   
## Min. :20000   
## 1st Qu.:20000   
## Median :20000   
## Mean :20000   
## 3rd Qu.:20000   
## Max. :20000