Arthur Harnisch Home Work #3

Arthur Harnisch

September 20, 2017

## Objective

The goal of this exercise is to subset a certain class of job titles (Equiptment Technician ) into a new data frames for each job level, then generate a boxplot showing the distribution of job title to salary.

## Setting the working directory

RStudio's default behavior when using the Knit button is to use the .Rmd file's directory as the working directory.This means that your directory path to things like the data files need to be set relative to the storage directory.

## Importing data

Next, I will import my data file of interest. Since the working directory used by RStudio (where this .Rmd file is located) is a sub-directory of my main directory 255E, I need to use the ../data/ syntax to get to my data folder.

calstateUniversity = read.csv("../data/california-state-university-2015.csv")

## Searching for the different classes

Since this data file contains numerous job titles, I want to find the different classes Equiptment Technician.

Jobs = levels(calstateUniversity$Job.Title)  
Jobs[grep(pattern = "EQUIPMENT TECHNICIAN" , x = Jobs)]

## [1] "EQUIPMENT TECHNICIAN I, ELECTRO-MECHANICAL"   
## [2] "EQUIPMENT TECHNICIAN I, ELECTRONIC"   
## [3] "EQUIPMENT TECHNICIAN I, MECHANICAL"   
## [4] "EQUIPMENT TECHNICIAN I, SPECIALIZED EQUIPMENT"   
## [5] "EQUIPMENT TECHNICIAN II, ELECTRO-MECHANICAL"   
## [6] "EQUIPMENT TECHNICIAN II, ELECTRONIC"   
## [7] "EQUIPMENT TECHNICIAN II, MECHANICAL"   
## [8] "EQUIPMENT TECHNICIAN II, SPECIALIZED EQUIPMENT"   
## [9] "EQUIPMENT TECHNICIAN III, ELECTRO-MECHANICAL"   
## [10] "EQUIPMENT TECHNICIAN III, ELECTRONIC"   
## [11] "EQUIPMENT TECHNICIAN III, MECHANICAL"   
## [12] "EQUIPMENT TECHNICIAN III, SPECIALIZED EQUIPMENT"

Jobs[grep(pattern = "ELECTRONIC" , x = Jobs)]

## [1] "EQUIPMENT TECHNICIAN I, ELECTRONIC"   
## [2] "EQUIPMENT TECHNICIAN II, ELECTRONIC"   
## [3] "EQUIPMENT TECHNICIAN III, ELECTRONIC"

## Subsetting the different classes

Here I am subsetting out the three classes of Equiptment Technician.

sjsu=calstateUniversity  
subone = subset(sjsu,calstateUniversity$Job.Title=="EQUIPMENT TECHNICIAN I, ELECTRONIC")   
subtwo = subset(sjsu,calstateUniversity$Job.Title=="EQUIPMENT TECHNICIAN II, ELECTRONIC")   
subthree = subset(sjsu,calstateUniversity$Job.Title=="EQUIPMENT TECHNICIAN III, ELECTRONIC")

## Merging the subsets

Then, I am merging the three subsets into one data set.

subs=rbind(subone,subtwo,subthree)

## Organizing the levels

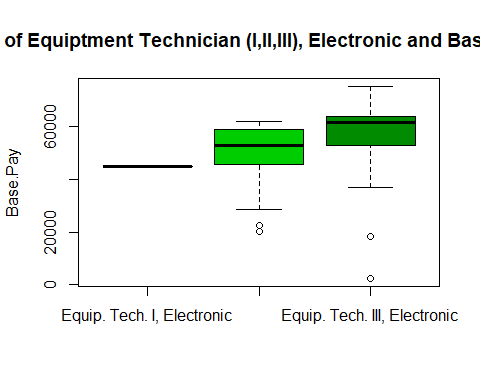
The data set now only has the job titles I am interesetd in, with three levels in the order that I want to have them displayed.

subs$Job.Title = factor(subs$Job.Title,   
 levels = c("EQUIPMENT TECHNICIAN I, ELECTRONIC",   
 "EQUIPMENT TECHNICIAN II, ELECTRONIC",  
 "EQUIPMENT TECHNICIAN III, ELECTRONIC" ))

## Including plots

Now we can make 3 boxplots, each representing the distribution of salary for each class. I abbreviated the name, specified the color for each class, and included axes titles.

boxplot(subone$Base.Pay, subtwo$Base.Pay, subthree$Base.Pay,names = c('Equip. Tech. I, Electronic',  
 'Equip. Tech. II, Electronic', 'Equip. Tech. III, Electronic'), ylab="Base.Pay",  
 main="Box Plots of Equiptment Technician (I,II,III), Electronic and Base Pay (salary)",   
 las=0, col = c("green2","green3","green4"))



Boxplot of CSU salary distribution of three different classes of Athletic trainer in 2015.

## Including R results

We can use the aggregate function to calculate the mean base pay of each level of Athletic Trainer.

mymeans = aggregate(Base.Pay~Job.Title, data = subs, FUN= mean)  
mymeans

## Job.Title Base.Pay  
## 1 EQUIPMENT TECHNICIAN I, ELECTRONIC 44995.08  
## 2 EQUIPMENT TECHNICIAN II, ELECTRONIC 48429.66  
## 3 EQUIPMENT TECHNICIAN III, ELECTRONIC 54917.49

The function kable can also be used to make a table outside of the data frame.

kable(mymeans, caption = "Mean base pay for each job level.")

Mean base pay for each job level.

|  |  |
| --- | --- |
| Job.Title | Base.Pay |
| EQUIPMENT TECHNICIAN I, ELECTRONIC | 44995.08 |
| EQUIPMENT TECHNICIAN II, ELECTRONIC | 48429.66 |
| EQUIPMENT TECHNICIAN III, ELECTRONIC | 54917.49 |