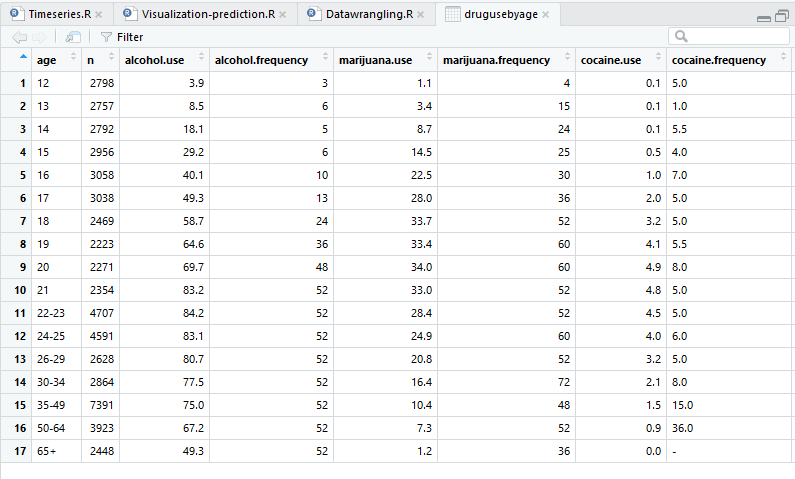
Name: Urvashibahen S.Parekh Student ID: 0699357

**Final Project**

Dataset Link:

<https://github.com/fivethirtyeight/data/tree/master/candy-power-ranking>

Dataset used: drug-use-by-age



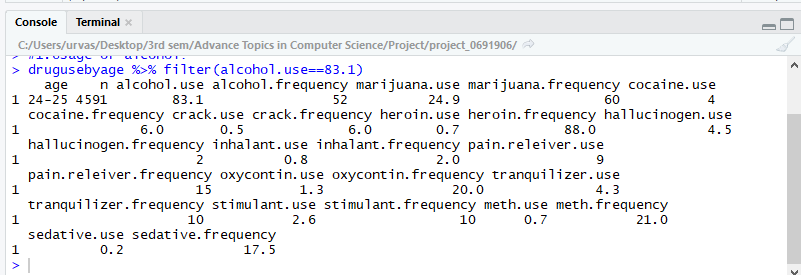
1 Datawrangling

Ans: drugusebyage <- read.csv("C:/Users/urvas/Desktop/3rd sem/Advance Topics in Computer Science/Project/project\_0699357/drug-use-by-age.csv")

library(dplyr)

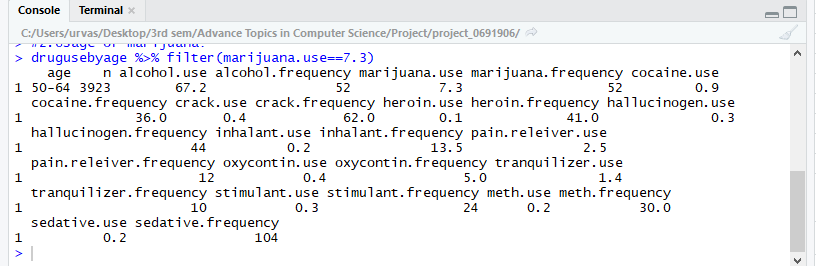
#1.Usage of alcohol?

drugusebyage %>% filter(alcohol.use==83.1)



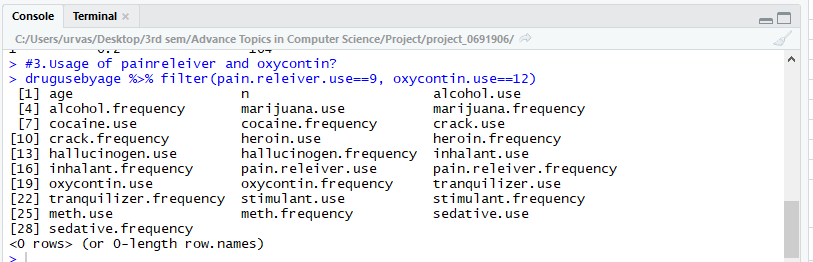
#2.Usage of marijuana?

drugusebyage %>% filter(marijuana.use==7.3)

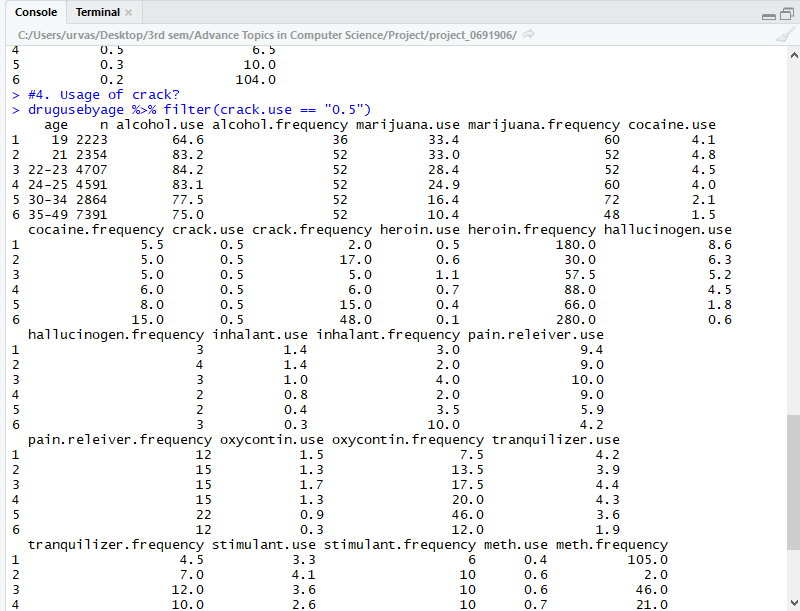


#3.Usage of painreleiver and oxycontin?

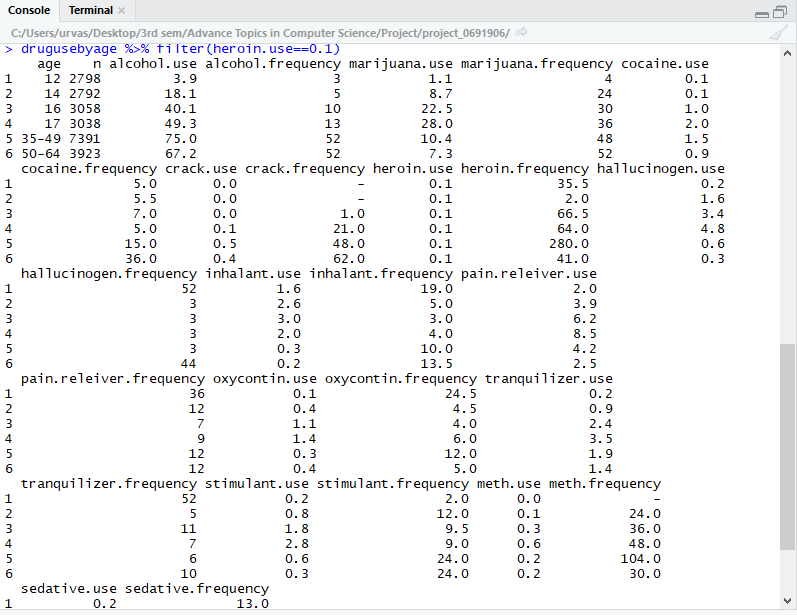
drugusebyage %>% filter(pain.releiver.use==9, oxycontin.use==12)



#4. Usage of crack?

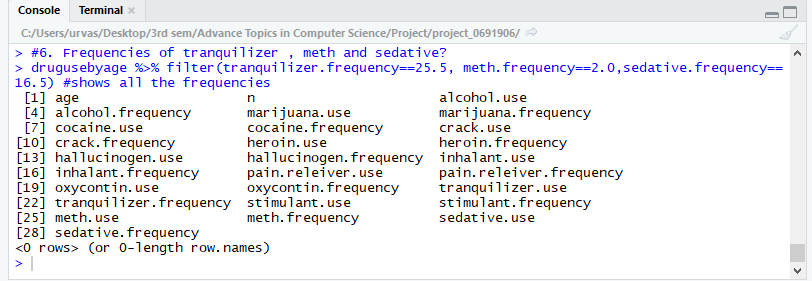
drugusebyage %>% filter(crack.use == "0.5") 

#5. Usage of heroin?

drugusebyage %>% filter(heroin.use==0.1) 

#6. Frequencies of tranquilizer , meth and sedative?

drugusebyage %>% filter(tranquilizer.frequency==25.5, meth.frequency==2.0,sedative.frequency==16.5) #shows all the frequencies



drugusebyage <- read.csv("C:/Users/urvas/Desktop/3rd sem/Advance Topics in Computer Science/Project/project\_0699357/drug-use-by-age.csv")

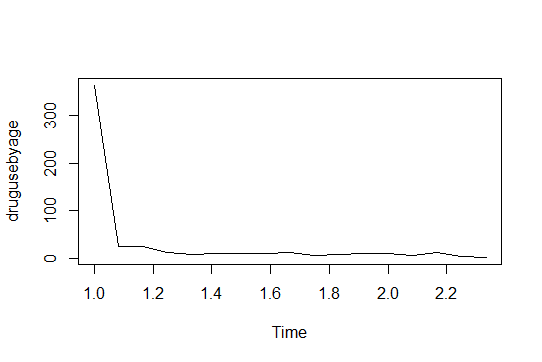
2. Timeseries

Ans: #1.Plotting the time series of the column stimulant.frequency by converting into a time series

drugusebyage <- rev(drugusebyage$stimulant.frequency)

drugusebyage <- ts(drugusebyage, frequency = 12)

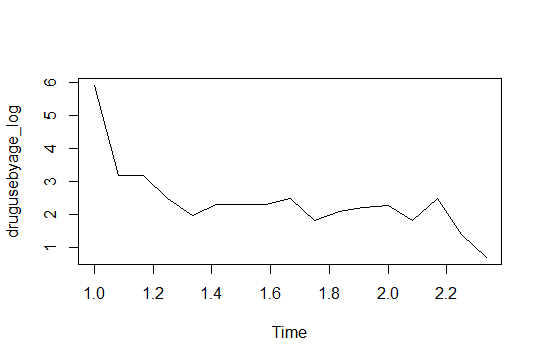
plot.ts(drugusebyage)



#2. log additive model to the time series.

drugusebyage\_log <- log(drugusebyage)

plot.ts(drugusebyage\_log)



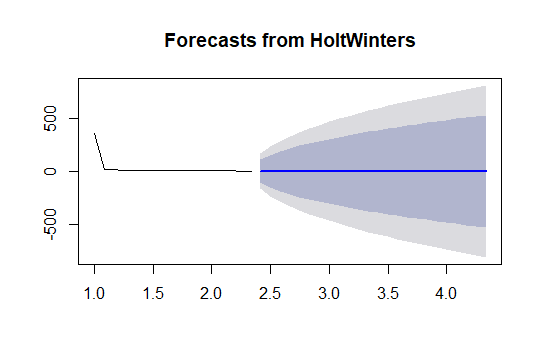
#3.Smoothing the drugsusebyage using the Holtwinter() method and plotting the result.

library("forecast")

drugusebyage\_Holtw <- HoltWinters(drugusebyage, beta = FALSE, gamma = FALSE)

plot(drugusebyage\_Holtw)

plot(forecast(drugusebyage\_Holtw))



3. Visualization and Prediction

library(readr)

library(dplyr)

library(texreg)

library(lazyeval)

library(ggplot2)

library(curl)

library(Zelig)

library(ggrepel)

library(HistData)

library(tidyverse)

library(ggthemes)

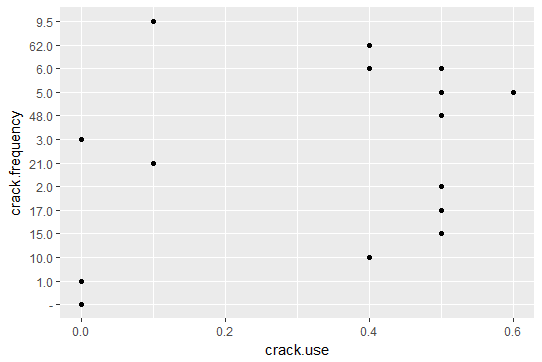
# scatterplot representing crackdrug with respect to usage and frequecy

drugusebyage <- read.csv("C:/Users/urvas/Desktop/3rd sem/Advance Topics in Computer Science/Project/project\_0699357/drug-use-by-age.csv")

head(drugusebyage)

ggplot(data = drugusebyage, aes(x = crack.use, y = crack.frequency)) +

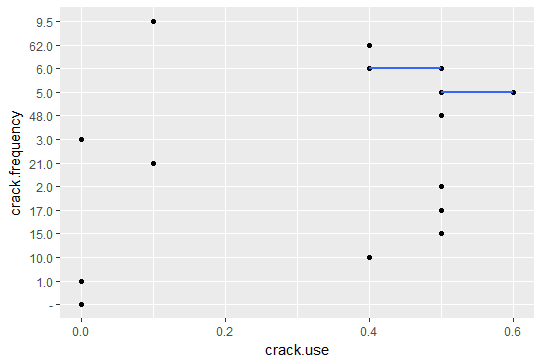
geom\_point()



ggplot(data = drugusebyage, aes(x = crack.use, y = crack.frequency)) +

geom\_point() +

geom\_smooth(method = "lm")

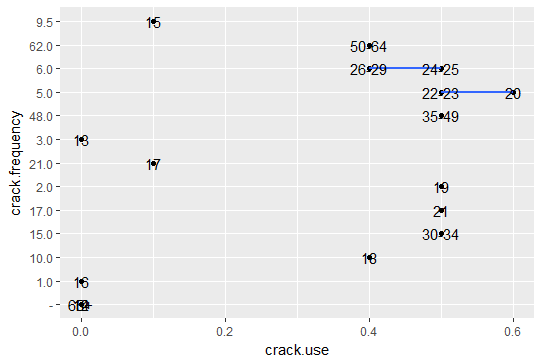


ggplot(data = drugusebyage, aes(x = crack.use, y = crack.frequency,label = age)) +

geom\_point() +

geom\_smooth(method = "lm") +

geom\_text()



ggplot(data = drugusebyage, aes(x = crack.use, y = crack.frequency,label = age)) +

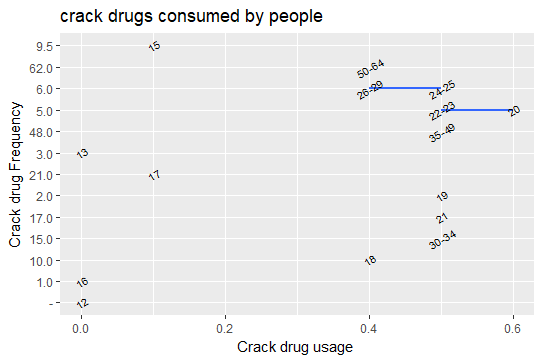
geom\_smooth(method = "lm") + geom\_text(check\_overlap = T,angle = 32,

size = 2.9) +

labs(title = "crack drugs consumed by people",

x = "Crack drug usage ",

y = "Crack drug Frequency" )



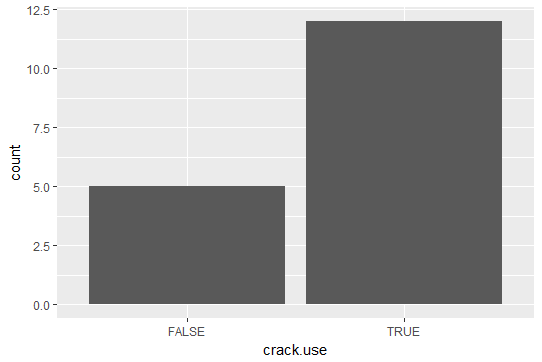
drugusebyage <- drugusebyage %>% select(2:10)

drugusebyage[] <- lapply(drugusebyage, as.logical)

# make bar plot

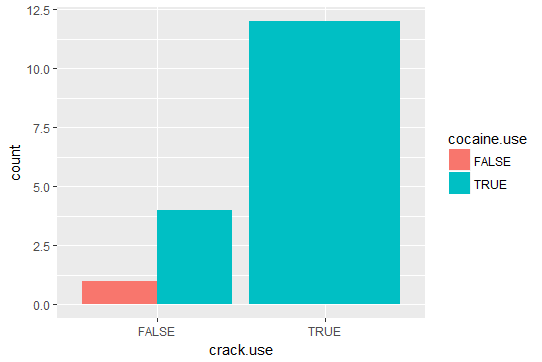
ggplot(drugusebyage, aes(x = crack.use)) +

geom\_bar()



ggplot(drugusebyage, aes(x = crack.use, fill = cocaine.use)) +

geom\_bar(position = "dodge")



ggplot(drugusebyage, aes(x = crack.use,fill = cocaine.use)) +geom\_bar(position = "dodge", size = 2) + facet\_wrap(c("cocaine.use")) + scale\_fill\_manual(values=c("#BBBBBB", "#E69F00")) +

labs(title = "Usage of cocaine highre then crack ", # title

x = "crack or cocaine ?",

y = "Count of people") + theme(legend.position = c(0.9, 0.9), strip.background = element\_blank(),strip.text.x = element\_blank())

