NAME:Sruthi.RSofiya.P

ROLLNO:17F240 17L244

CASE STUDY:

**Attribute Information:**

1. When one of our apologies apologizes when our discussions go in a bad direction, the issue does not extend.   
2. I know we can ignore our differences, even if things get hard sometimes.   
3. When we need it, we can take our discussions with my wife from the beginning and correct it.   
4. When I argue with my wife, it will eventually work for me to contact him.   
5. The time I spent with my wife is special for us.   
6. We don't have time at home as partners.   
7. We are like two strangers who share the same environment at home rather than family.   
8. I enjoy our holidays with my wife.   
9. I enjoy traveling with my wife.   
10. My wife and most of our goals are common.   
11. I think that one day in the future, when I look back, I see that my wife and I are in harmony with each other.   
12. My wife and I have similar values in terms of personal freedom.   
13. My husband and I have similar entertainment.   
14. Most of our goals for people (children, friends, etc.) are the same.   
15. Our dreams of living with my wife are similar and harmonious   
16. We're compatible with my wife about what love should be   
17. We share the same views with my wife about being happy in your life   
18. My wife and I have similar ideas about how marriage should be   
19. My wife and I have similar ideas about how roles should be in marriage   
20. My wife and I have similar values in trust   
21. I know exactly what my wife likes.   
22. I know how my wife wants to be taken care of when she's sick.   
23. I know my wife's favorite food.   
24. I can tell you what kind of stress my wife is facing in her life.   
25. I have knowledge of my wife's inner world.   
26. I know my wife's basic concerns.   
27. I know what my wife's current sources of stress are.   
28. I know my wife's hopes and wishes.   
29. I know my wife very well.   
30. I know my wife's friends and their social relationships.   
31. I feel aggressive when I argue with my wife.   
32. When discussing with my wife, I usually use expressions such as â€œyoualwaysâ€œ or â€œyouneverâ€.   
33. I can use negative statements about my wife's personality during our discussions.   
34. I can use offensive expressions during our discussions.   
35. I can insult our discussions.   
36. I can be humiliating when we argue.   
37. My argument with my wife is not calm.   
38. I hate my wife's way of bringing it up.   
39. Fights often occur suddenly.   
40. We're just starting a fight before I know what's going on.   
41. When I talk to my wife about something, my calm suddenly breaks.   
42. When I argue with my wife, it only snaps in and I don't say a word.   
43. I'm mostly thirsty to calm the environment a little bit.   
44. Sometimes I think it's good for me to leave home for a while.   
45. I'd rather stay silent than argue with my wife.   
46. Even if I'm right in the argument, I'm thirsty not to upset the other side.   
47. When I argue with my wife, I remain silent because I am afraid of not being able to control my anger.   
48. I feel right in our discussions.   
49. I have nothing to do with what I've been accused of.   
50. I'm not actually the one who's guilty about what I'm accused of.   
51. I'm not the one who's wrong about problems at home.   
52. I wouldn't hesitate to tell her about my wife's inadequacy.   
53. When I discuss it, I remind her of my wife's inadequate issues.   
54. I'm not afraid to tell her about my wife's incompetence.

**DATA representation:**

4:agree 100%

3:agree 75%

2:agree 50%

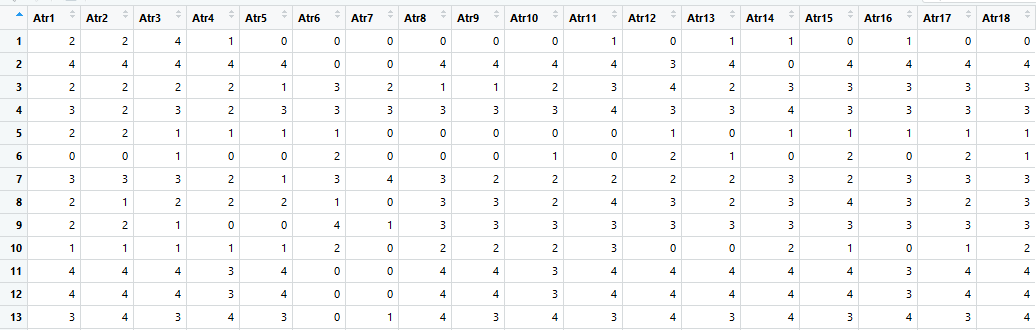
1:agree 25%

0:agree 0%

**Import data:**

>d<-read.csv(file="divorce.csv",sep=";")

>View(d)



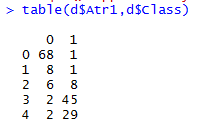
**EXPLORATORY DATA ANALYSIS:**

**Classification:**

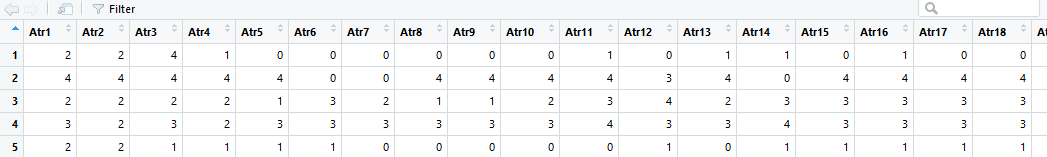
Numerical

>summary(d)

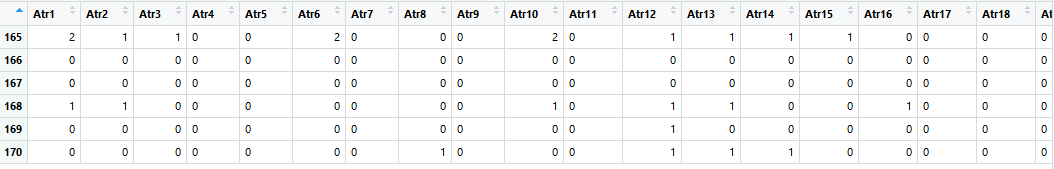
>table(d$Atr1,d$Class)



>View(head(d,n=5))



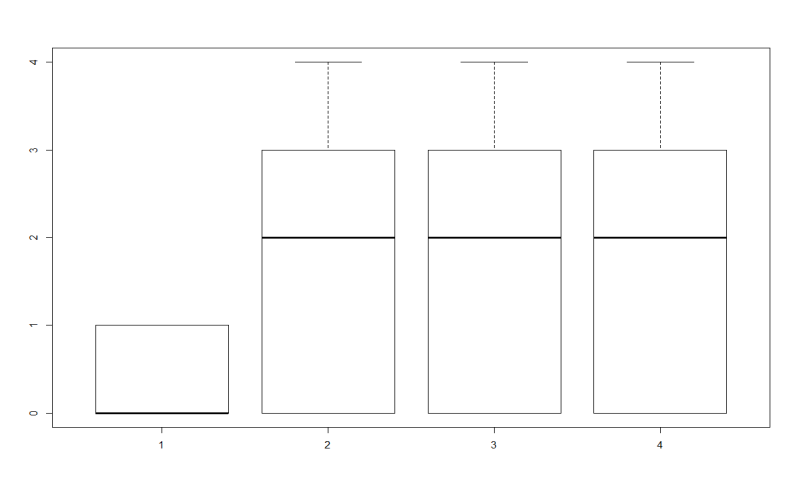
>View(tail(d))







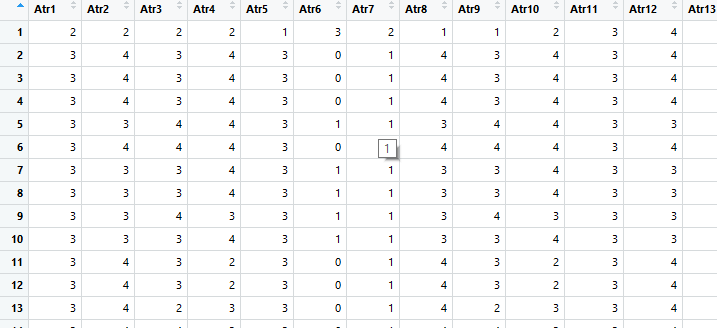
>boxplot(d$Class,d$Atr1,d$Atr2,d$Atr3)



>library(dplyr)

>d<-d%>%filter(d$Atr29>=3)%>%arrange(Atr29)%>%group\_by(Class)

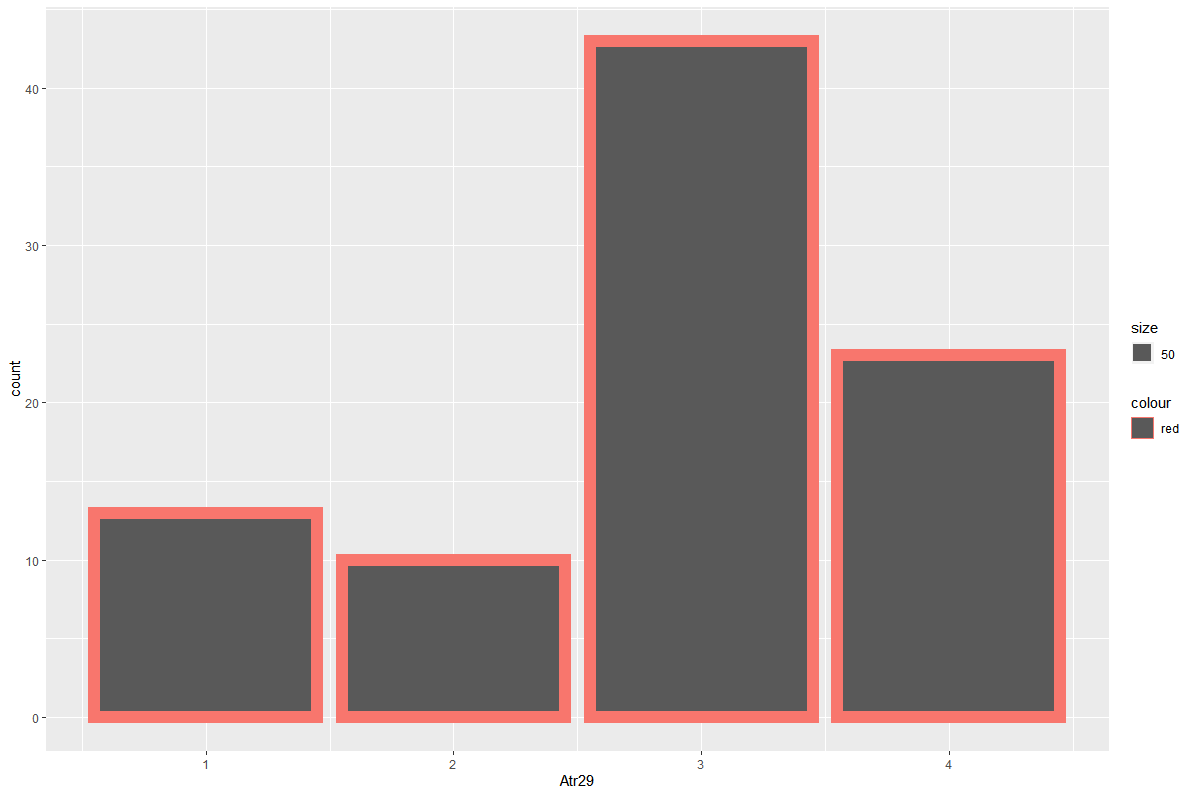
>View(d)



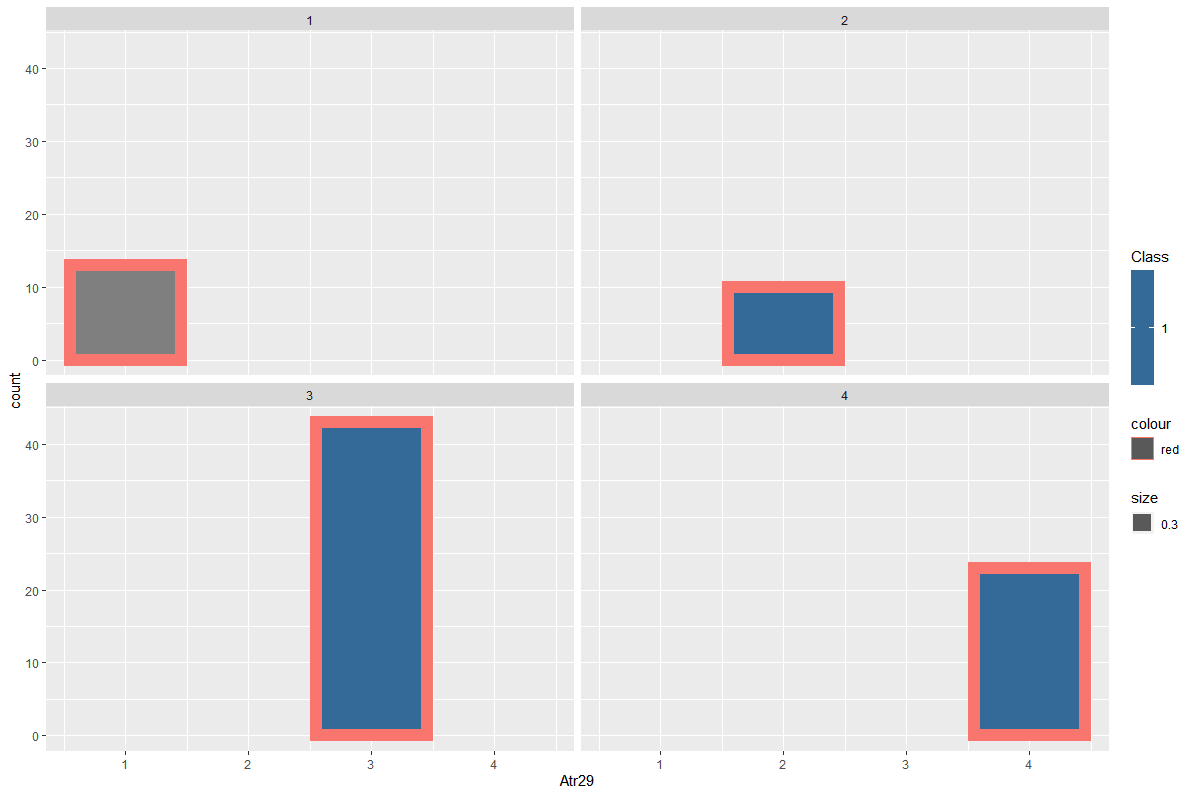
**#visualise**

>library(ggplot2)

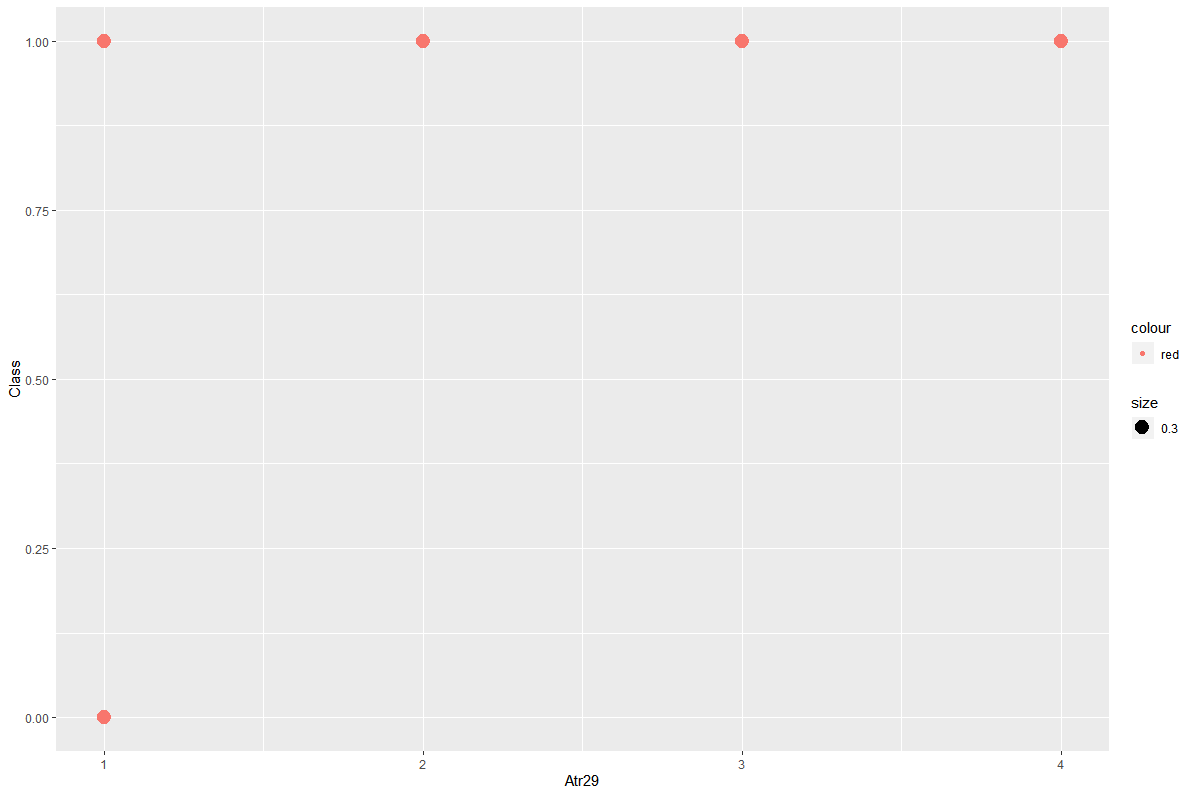
>ggplot(d,aes(x=Atr29,color='red',size=5))+geom\_bar(position="dodge")



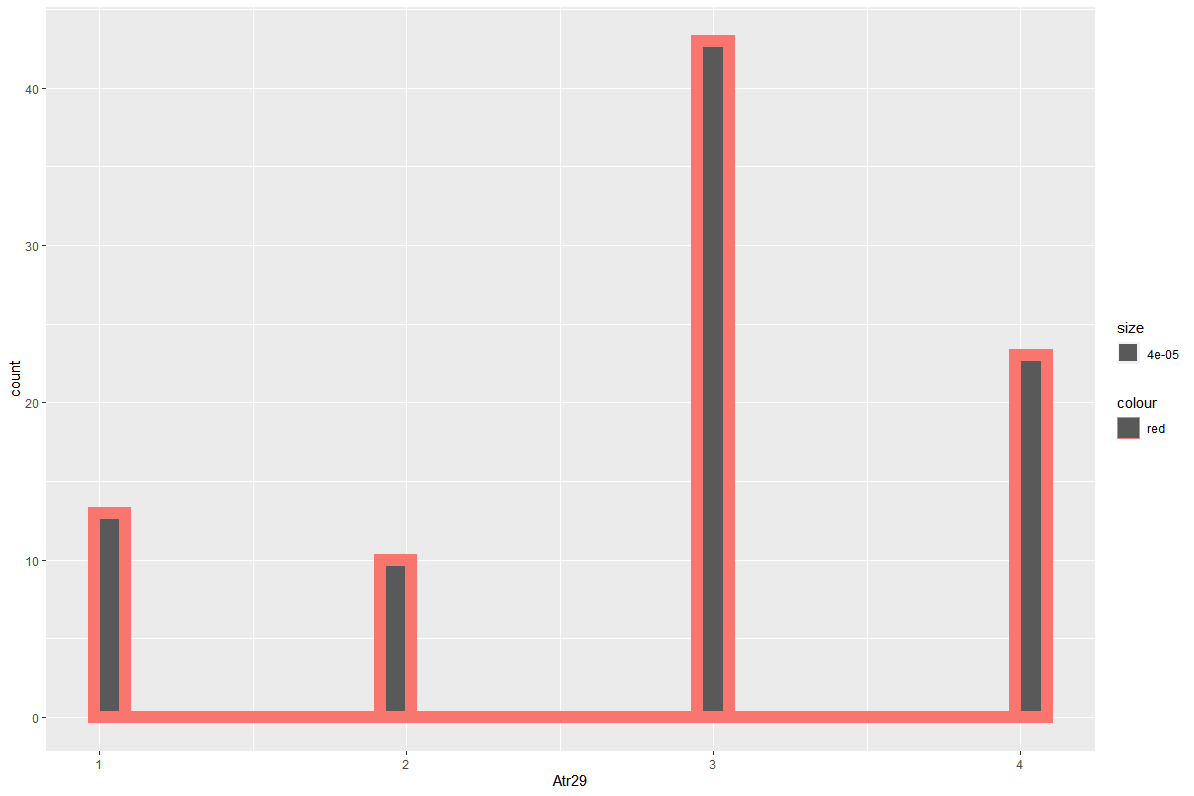
>d%>%ggplot()+geom\_bar(aes(x=Atr29,fill=Class,color='red',size=.3))+facet\_wrap(~Atr29)



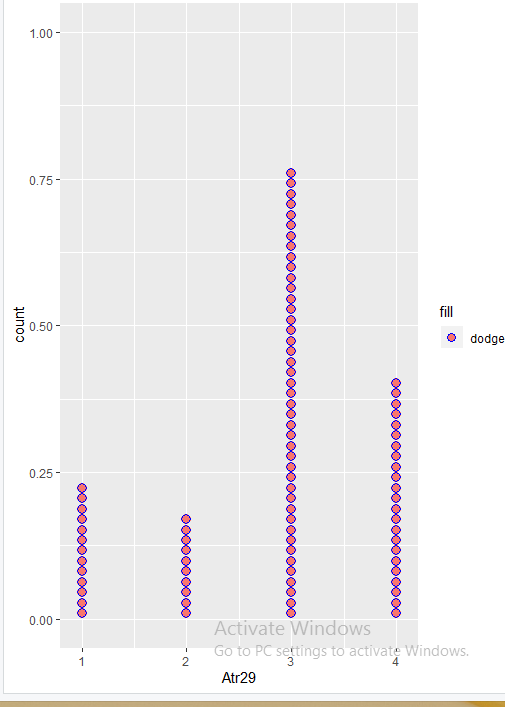
>d%>%ggplot()+geom\_point(aes(x=Atr29,y=Class,color='red',size=.3))



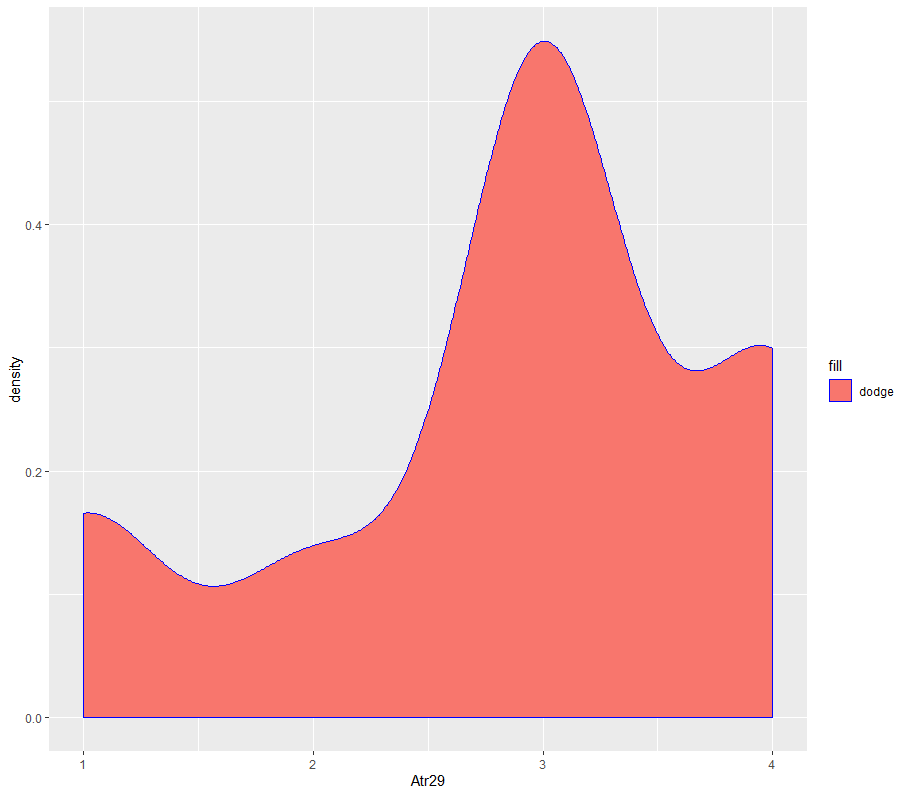
>d%>%ggplot()+geom\_histogram(aes(x=Atr29,fill=Class,color='red',size=0.00004))



ggplot(d , aes(x = Atr29 , fill = "dodge" ))+geom\_dotplot(dotsize = 1.0 , color = "blue")



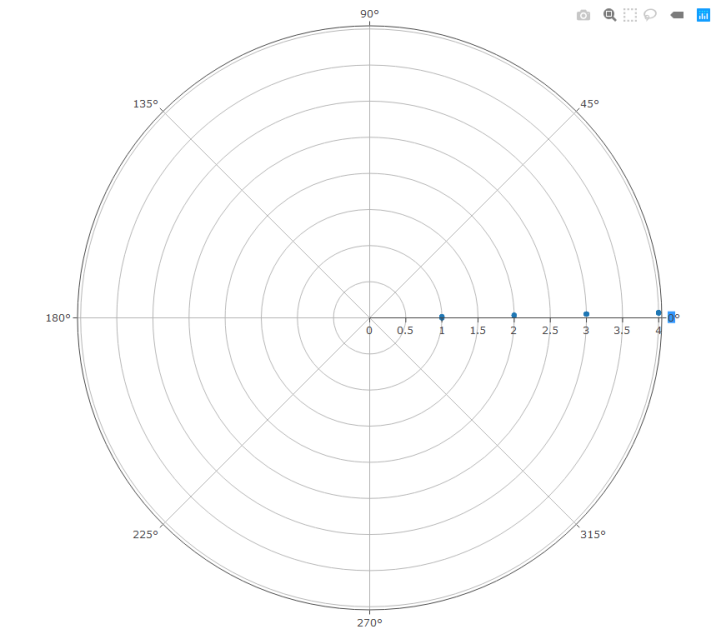
ggplot(d , aes(x = Atr29 , fill = "dodge" , color = "pink"))+geom\_density(alpha = 1.0 , color = "blue")



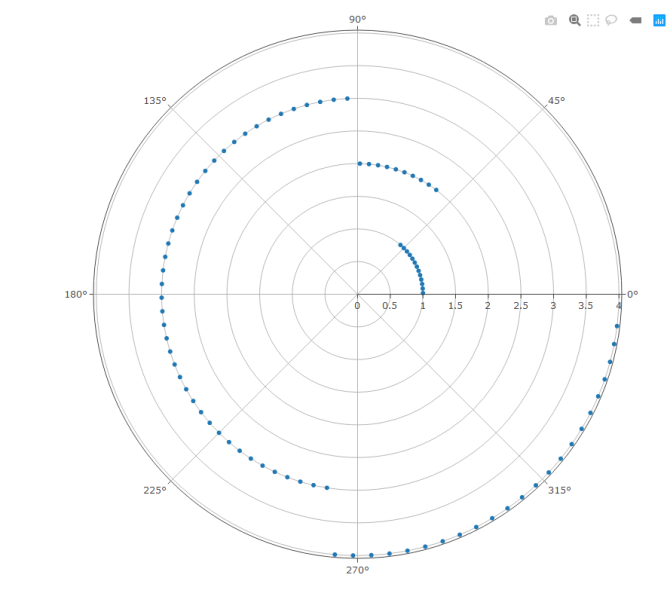
**#Plotly**

library(plotly)

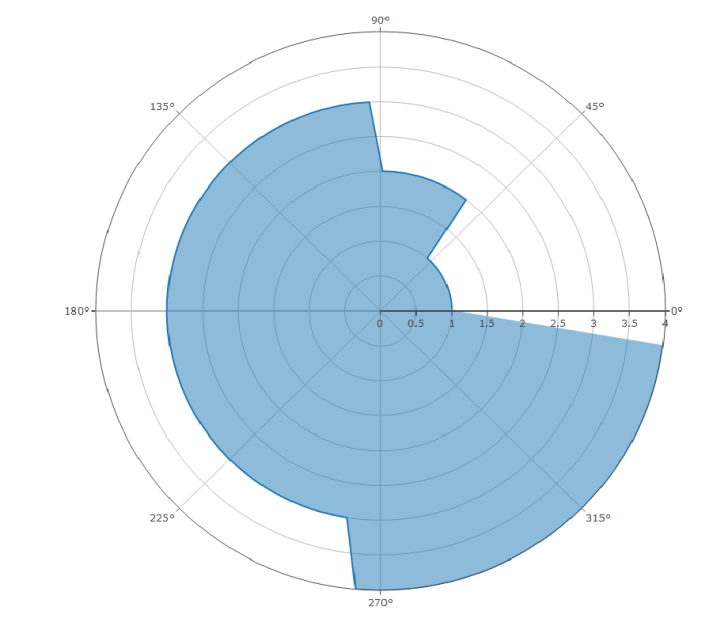
plot\_ly(type = "scatterpolar" , r=d$Atr29 ,theta =d$Class , mode = "markers")



plot\_ly(type = 'scatterpolar' , r=d$Atr29 ,theta =seq(from = 1 , to = 356 , by = 4) , mode = "markers")

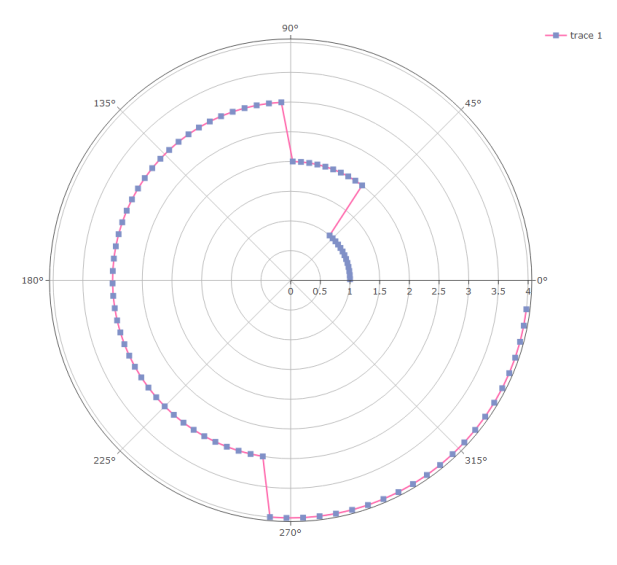


plot\_ly(type = 'scatterpolar' , r=d$Atr29 ,theta =seq(from = 1 , to = 356 , by = 4) , fill = “toself” , mode = "lines")



p<-plot\_ly(type = 'scatterpolar' ,mode = "lines+markers") %>% add\_trace(r=d$Atr29 ,theta =seq(from = 1 , to = 356 , by = 4) ,line = list(color = "#ff66ab") , marker = list(color="#8090c7",symbol='square',size=8) )

p



**#take care of missing data(replacing the null values with mean):**

We don’t have any missing data.

**#encoding categorical data (changing the character into numeric):**

Our data is numerical so we don’t need encoding.

**#splitting the dataset into the training set and testing set:**

library(caTools)

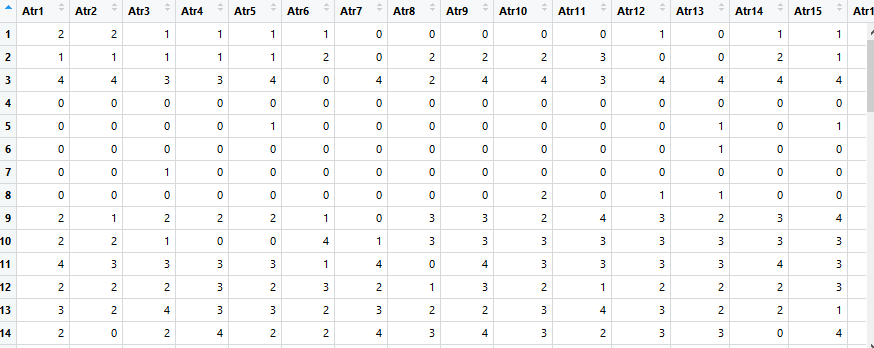
split=sample.split(d$Atr29,SplitRatio=3/5)

split

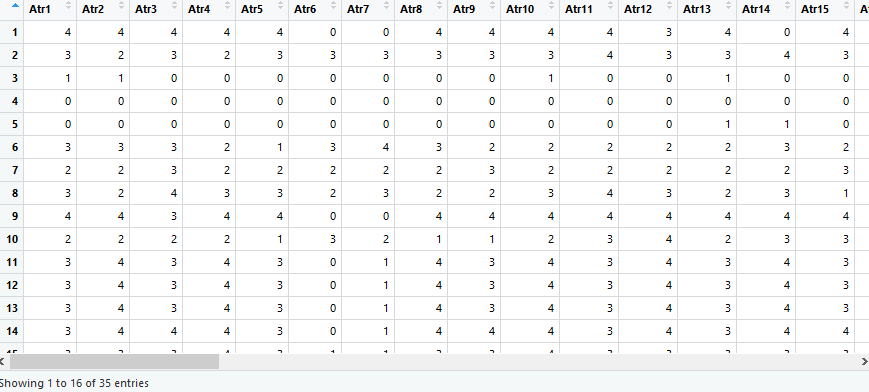
training\_set=subset(d,split==TRUE)

testing\_set=subset(d,split==FALSE)

View(training\_set)



View(testing\_set)



dim(testing\_set)



dim(training\_set)



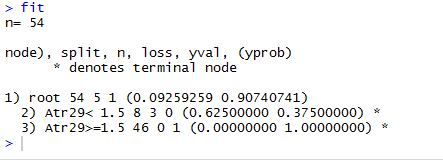
**Build a model**

library(rpart)

library(rpart.plot)

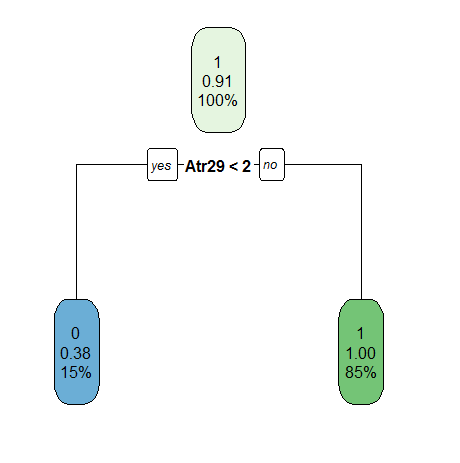
fit<- rpart(formula = Class~Atr29 ,data = training\_set ,method = 'class')

fit



**Plot Decesion Tree**

rpart.plot(fit)



**#feature scaling**

**#scale is used for only numeric**

**# converting the larger distance value into a smaller for future purpose by usingscale(normalization process is taken)..(equally distributed value)**

Already the values are smaller so feature scaling is not necessary.

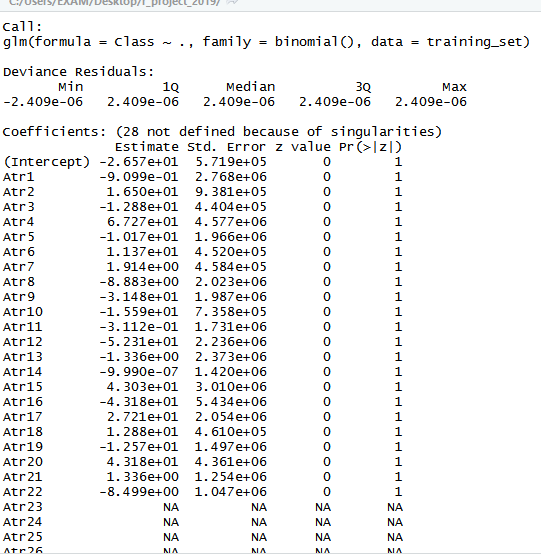
**#fitting logistic regression to training set**

classifier<- glm(formula = Class ~ .,

family = quasibinomial(),

data = training\_set)

View(summary(classifier))



**#predicting the test set Result**

prob\_pred<- predict(classifier,

type='response',

newdata=testing\_set)

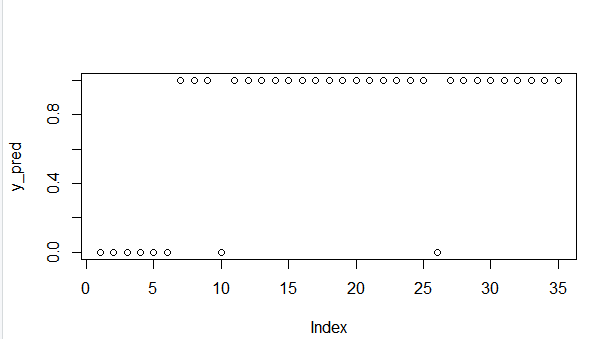
?predict

y\_pred<- ifelse(prob\_pred>0.5,1,0)

y\_pred



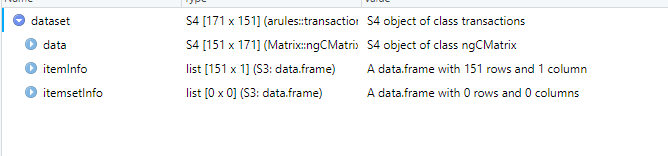
plot(y\_pred)



**Apriori**

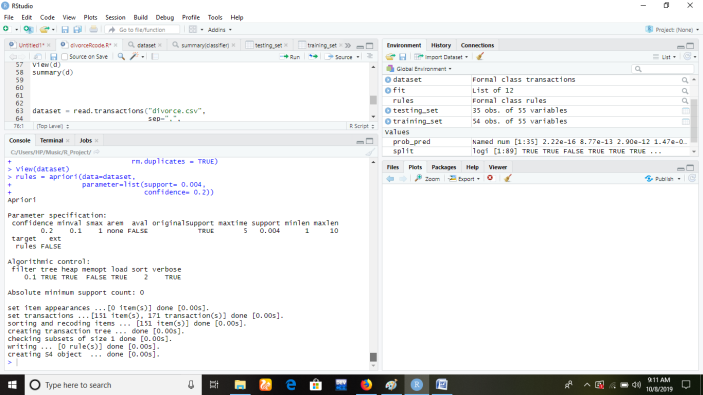
dataset = read.transactions("divorce.csv", sep="," , rm.duplicates = TRUE)

View(dataset)



**Training Apriori on the dataset**

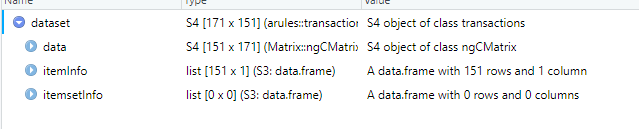
rules = apriori(data=dataset , parameter=list(support= 0.004 , confidence= 0.2))



**Visualising the result**

inspect(sort(rules, by='lift')[1:10])

View(dataset)



**K-Means Clustering**

**Using elbow method to find optimal cluster**

library(cluster)

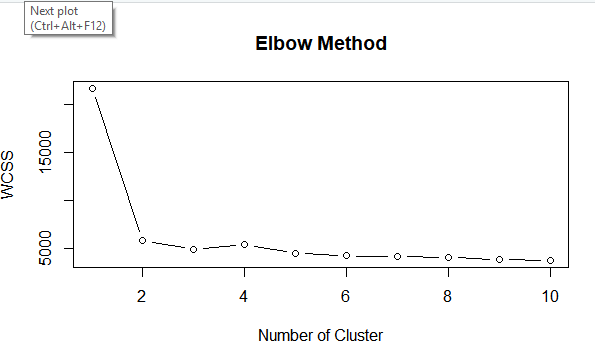
set.seed(12)

wcss = vector()

for(i in 1:10)

wcss[i]=sum(kmeans(d, i)$withinss)

plot(1:10, wcss, type = 'b', main="Elbow Method", xlab = 'Number of Cluster', ylab = "WCSS")



**Fitting K-means Dataset**

set.seed(24)

kmeans = kmeans(x= dataset, centers = 5)

y\_kmeans = kmeans$cluster

**Visualize**

clusplot(dataset, y\_kmeans, lines = 0,

# shade = T,

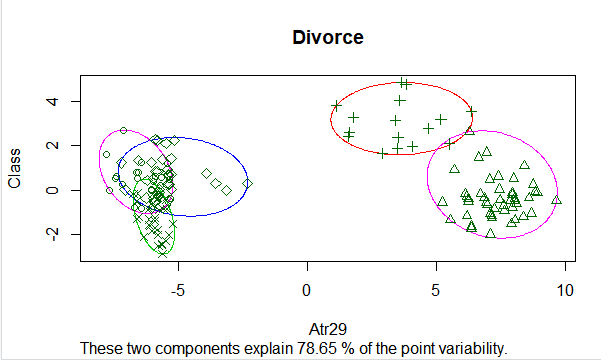
color = T,

#labels = 2,

main="Cluster of Customer",

xlab = 'Annual Income',

ylab = 'Spending Scores')



library(plotrix)

**ANOVA**

**Step 1**

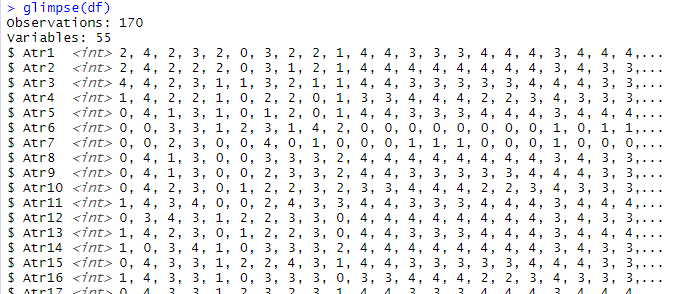
df <- read.csv(PATH) %>% select(-X) %>% mutate(poison =factor(poison, ordered=T))

View(df)

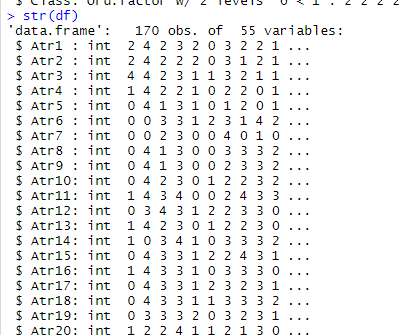
levels(df)



glimpse(df)

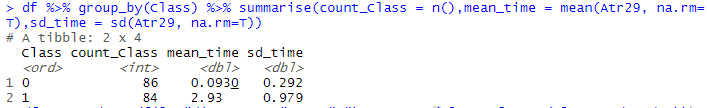


str(df)



**Step 2 count, mean, sd**

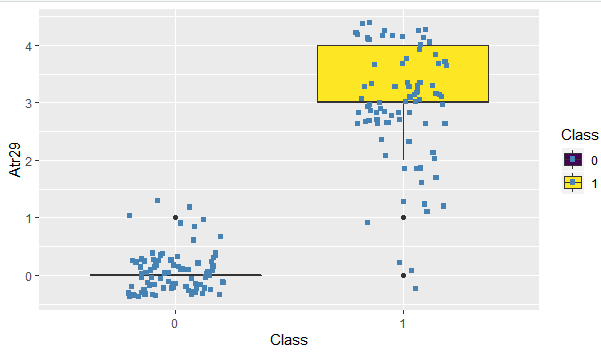
df %>% group\_by(Class) %>% summarise(count\_Class = n(),mean\_time = mean(Atr29, na.rm=T),sd\_time = sd(Atr29, na.rm=T))



**Step 3 Box Plot**

library(ggplot2)

ggplot(df, aes(x=Class, y=Atr29, fill=Class))+ geom\_boxplot()+ geom\_jitter(shape=15 , color = 'steelblue')

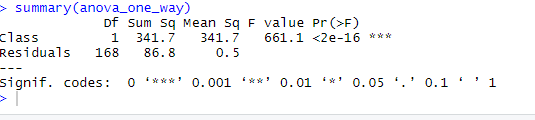
****

**Step 4 : one-way anova**

anova\_one\_way=aov(formula = Atr29~Class, data = df)

summary(anova\_one\_way)

TukeyHSD(anova\_one\_way)

****

**Step 5 : Two-way anova**

anova\_two\_way =aov(formula = Atr29~Class+Atr1, data = df)

summary(anova\_two\_way)

