library(readr)  
library(ggplot2)  
library(tidyr)  
library(plyr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

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

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

library(lattice)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.2 ✔ tibble 3.2.1  
## ✔ purrr 1.0.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::arrange() masks plyr::arrange()  
## ✖ purrr::compact() masks plyr::compact()  
## ✖ dplyr::count() masks plyr::count()  
## ✖ dplyr::desc() masks plyr::desc()  
## ✖ dplyr::failwith() masks plyr::failwith()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::id() masks plyr::id()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ dplyr::mutate() masks plyr::mutate()  
## ✖ dplyr::rename() masks plyr::rename()  
## ✖ dplyr::summarise() masks plyr::summarise()  
## ✖ dplyr::summarize() masks plyr::summarize()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(Hmisc)

##   
## Attaching package: 'Hmisc'  
##   
## The following objects are masked from 'package:dplyr':  
##   
## src, summarize  
##   
## The following objects are masked from 'package:plyr':  
##   
## is.discrete, summarize  
##   
## The following objects are masked from 'package:base':  
##   
## format.pval, units

df <- read\_csv("loan\_test.csv")

## Rows: 367 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (6): Gender, Married, Dependents, Education, Self\_Employed, Area  
## dbl (5): Applicant\_Income, Coapplicant\_Income, Loan\_Amount, Term, Credit\_His...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

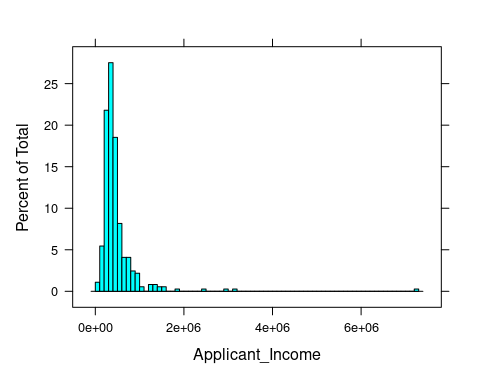
summary(df)

## Gender Married Dependents Education   
## Length:367 Length:367 Length:367 Length:367   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Self\_Employed Applicant\_Income Coapplicant\_Income Loan\_Amount   
## Length:367 Min. : 0 Min. : 0 Min. : 0   
## Class :character 1st Qu.: 286400 1st Qu.: 0 1st Qu.:10000000   
## Mode :character Median : 378600 Median : 102500 Median :12500000   
## Mean : 480560 Mean : 156958 Mean :13427793   
## 3rd Qu.: 506000 3rd Qu.: 243050 3rd Qu.:15750000   
## Max. :7252900 Max. :2400000 Max. :55000000   
##   
## Term Credit\_History Area   
## Min. : 6.0 Min. :0.0000 Length:367   
## 1st Qu.:360.0 1st Qu.:1.0000 Class :character   
## Median :360.0 Median :1.0000 Mode :character   
## Mean :342.5 Mean :0.8254   
## 3rd Qu.:360.0 3rd Qu.:1.0000   
## Max. :480.0 Max. :1.0000   
## NA's :6 NA's :29

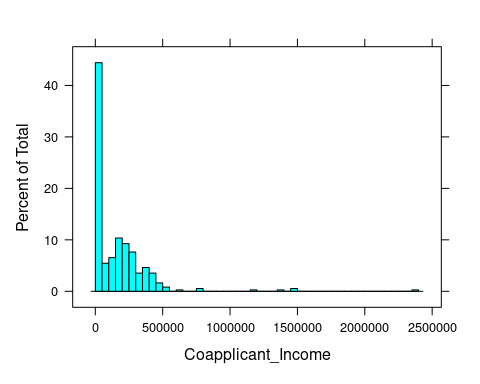
n\_distinct(df$Area)

## [1] 3

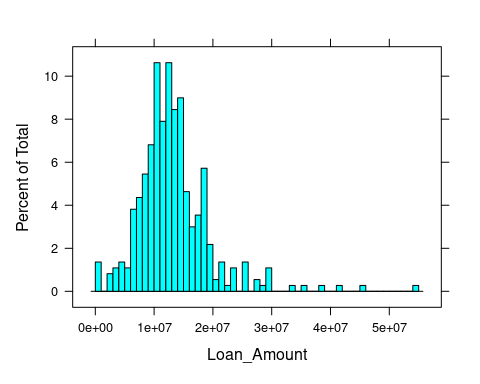
df$income<-(df$Applicant\_Income+df$Coapplicant\_Income)  
  
#Histogram  
histogram(~Applicant\_Income, data=df, breaks = 100)



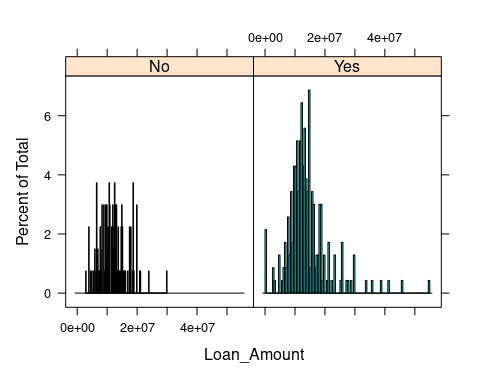
histogram(~Coapplicant\_Income, data=df, breaks = 50)



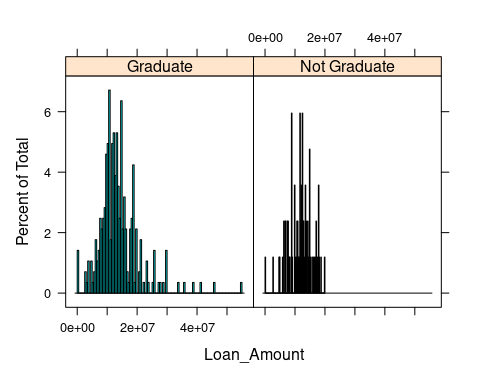
histogram(~Loan\_Amount, data=df, breaks = 50)



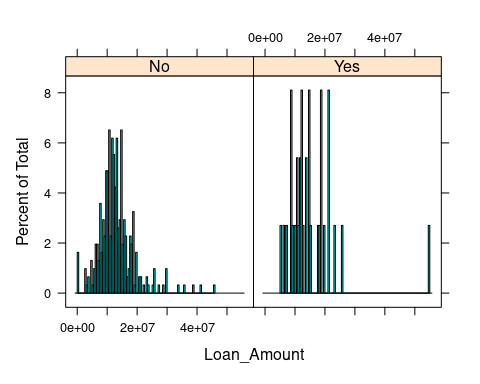
histogram(~Loan\_Amount|Married,data = df,breaks = 100)



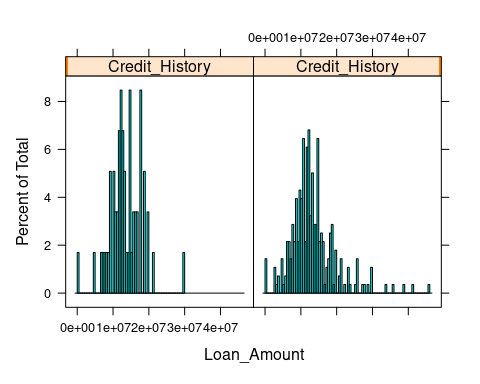
histogram(~Loan\_Amount|Education,data = df,breaks = 100)



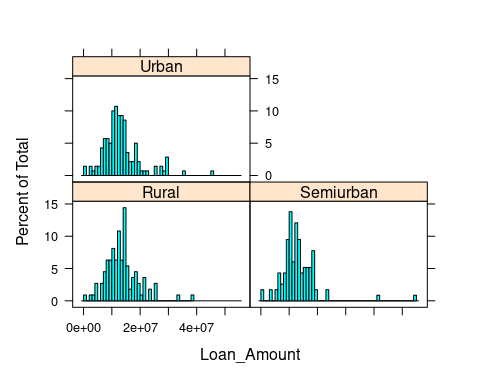
histogram(~Loan\_Amount|Self\_Employed,data = df,breaks = 100)



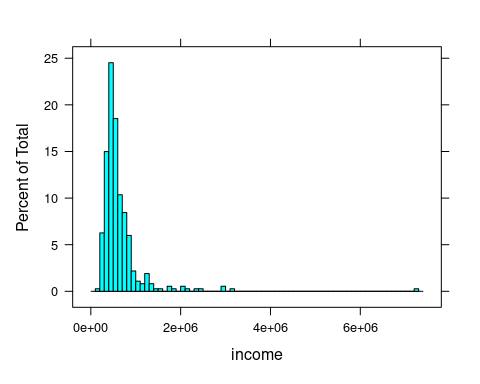
histogram(~Loan\_Amount|Credit\_History,data = df,breaks = 100)



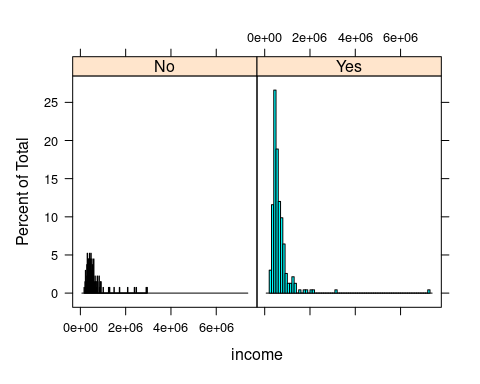
histogram(~Loan\_Amount|Area,data = df,breaks = 50)



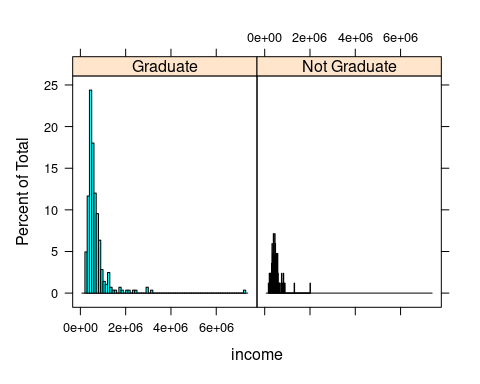
histogram(~income,data = df,breaks = 100)



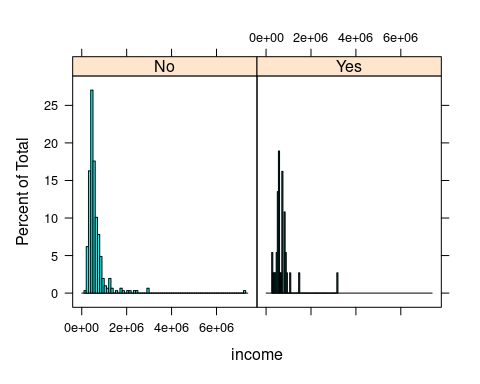
histogram(~income|Married,data = df,breaks = 100)



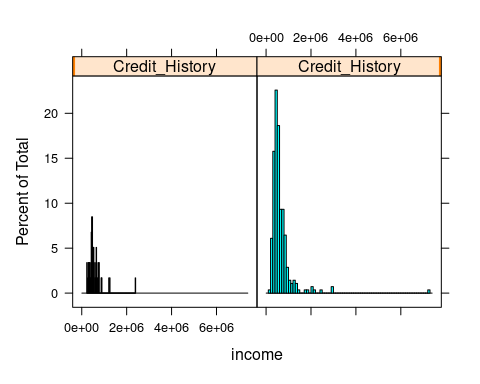
histogram(~income|Education,data = df,breaks = 100)



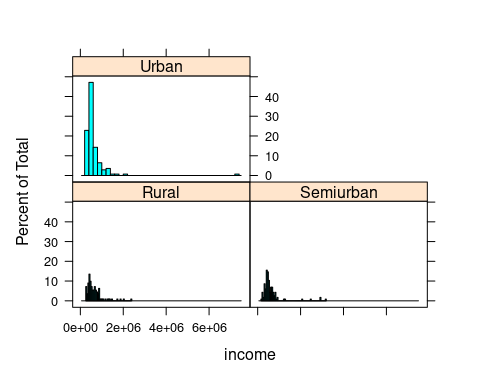
histogram(~income|Self\_Employed,data = df,breaks = 100)



histogram(~income|Credit\_History,data = df,breaks = 100)



histogram(~income|Area,data = df,breaks = 50)



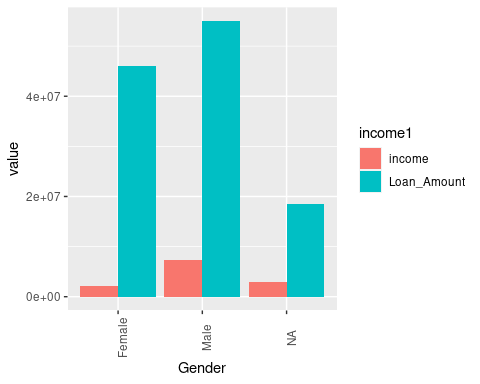
#subset  
urban=subset(df, Area=='Urban')  
semiurban=subset(df, Area=='Semiurban')  
rural=subset(df, Area=='Rural')  
female=subset(df,Gender=="Female")  
male=subset(df,Gender=='male')  
  
summary(urban)

## Gender Married Dependents Education   
## Length:140 Length:140 Length:140 Length:140   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Self\_Employed Applicant\_Income Coapplicant\_Income Loan\_Amount   
## Length:140 Min. : 114100 Min. : 0 Min. : 0   
## Class :character 1st Qu.: 290950 1st Qu.: 0 1st Qu.: 9975000   
## Mode :character Median : 381000 Median : 101250 Median :12500000   
## Mean : 503891 Mean : 136890 Mean :13427857   
## 3rd Qu.: 559450 3rd Qu.: 216600 3rd Qu.:15050000   
## Max. :7252900 Max. :1166600 Max. :46000000   
##   
## Term Credit\_History Area income   
## Min. : 60.0 Min. :0.0000 Length:140 Min. : 208300   
## 1st Qu.:360.0 1st Qu.:1.0000 Class :character 1st Qu.: 415600   
## Median :360.0 Median :1.0000 Mode :character Median : 529850   
## Mean :344.1 Mean :0.8615 Mean : 640781   
## 3rd Qu.:360.0 3rd Qu.:1.0000 3rd Qu.: 674525   
## Max. :480.0 Max. :1.0000 Max. :7252900   
## NA's :3 NA's :10

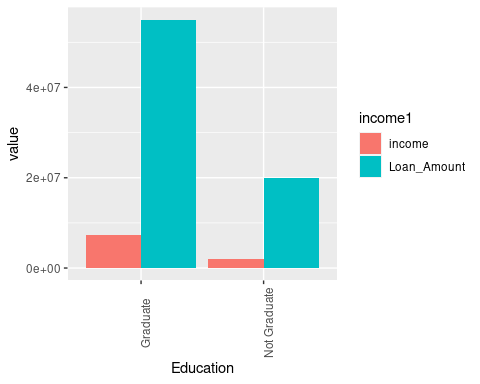
range(female$income)

## [1] 176000 2166600

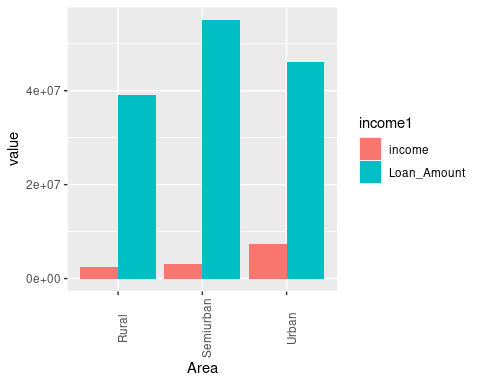
#bar chart  
df\_income <- gather(df, key ='income1', value = "value", 'income','Loan\_Amount')  
  
ggplot(data = df\_income, aes(x = Gender, y = value, fill = income1)) +  
 geom\_bar(stat = "identity", position = position\_dodge())+theme(axis.text.x = element\_text(angle = 90))



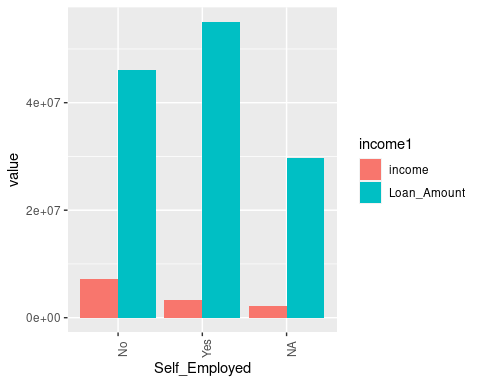
ggplot(data = df\_income, aes(x =Education, y = value, fill = income1)) +  
 geom\_bar(stat = "identity", position = position\_dodge())+theme(axis.text.x = element\_text(angle = 90))



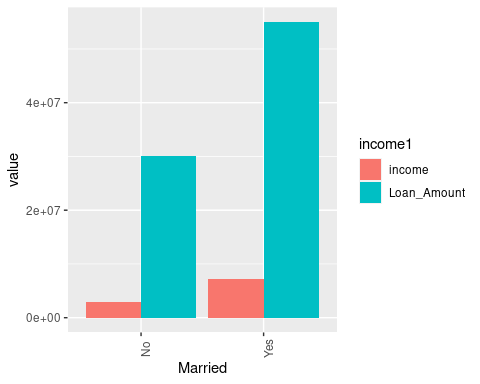
ggplot(data = df\_income, aes(x =Area, y = value, fill = income1)) +  
 geom\_bar(stat = "identity", position = position\_dodge())+theme(axis.text.x = element\_text(angle = 90))



ggplot(data = df\_income, aes(x =Self\_Employed, y = value, fill = income1)) +  
 geom\_bar(stat = "identity", position = position\_dodge())+theme(axis.text.x = element\_text(angle = 90))

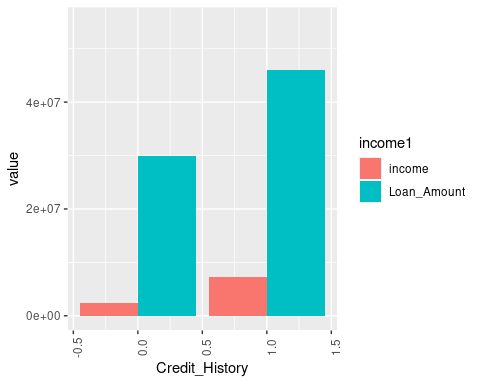


ggplot(data = df\_income, aes(x =Married, y = value, fill = income1)) +  
 geom\_bar(stat = "identity", position = position\_dodge())+theme(axis.text.x = element\_text(angle = 90))

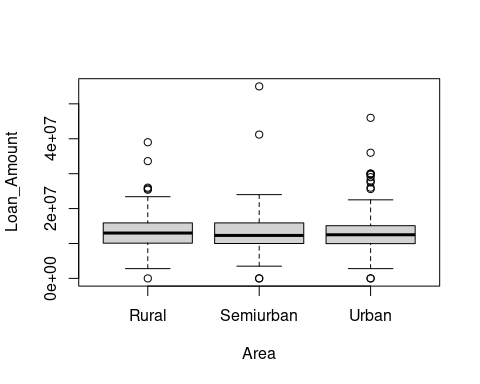


ggplot(data = df\_income, aes(x =Credit\_History, y = value, fill = income1)) +  
 geom\_bar(stat = "identity", position = position\_dodge())+theme(axis.text.x = element\_text(angle = 90))

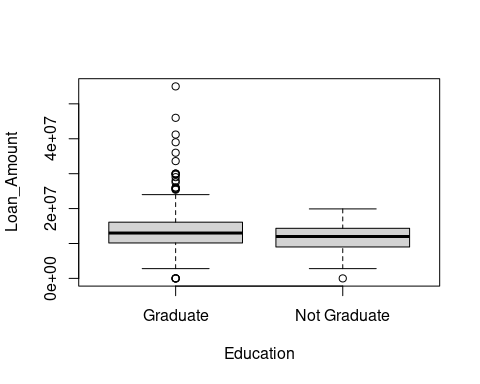
## Warning: Removed 58 rows containing missing values (`geom\_bar()`).



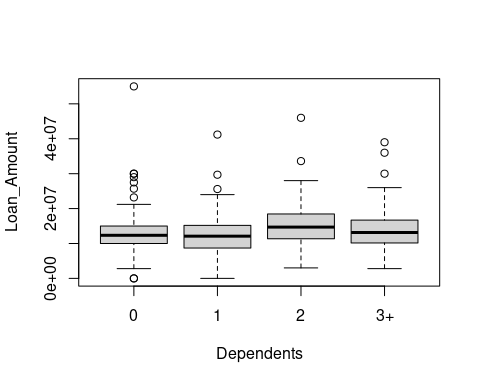
#boxplot  
library(ggplot2)  
boxplot(Loan\_Amount~Area,data = df,title="loan vs area")



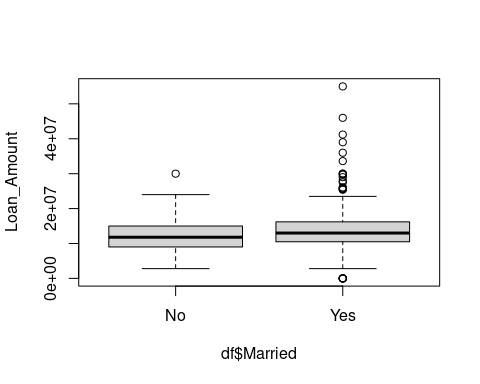
boxplot(Loan\_Amount~Education,data = df,title="loan vs Education")



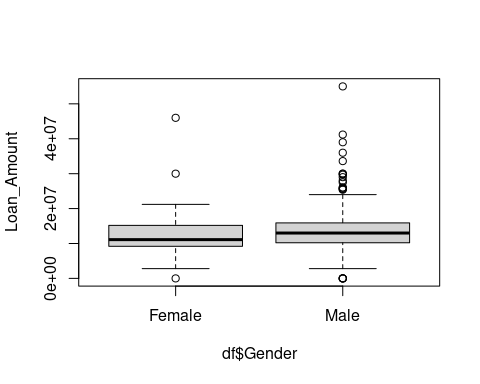
boxplot(Loan\_Amount~Dependents,data = df,title="loan vs area")



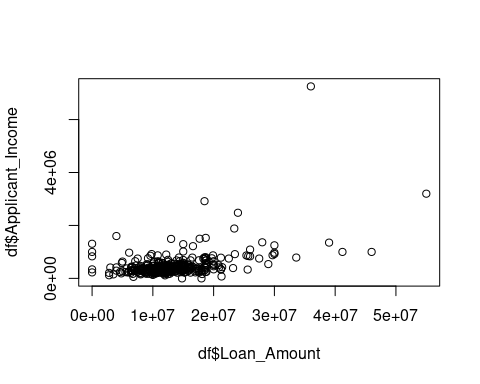
boxplot(Loan\_Amount~df$Married,data = df,title="loan vs area")



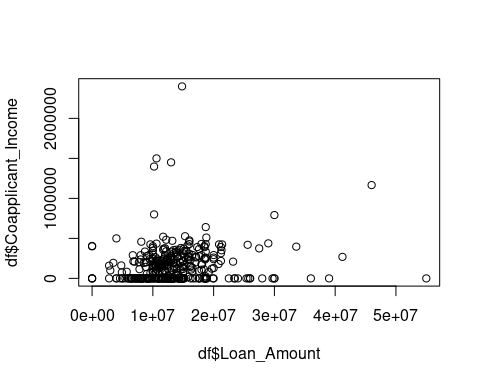
boxplot(Loan\_Amount~df$Gender,data = df,title="loan vs area")



#scatterplot  
  
plot(df$Loan\_Amount,df$Applicant\_Income)



plot(df$Loan\_Amount,df$Coapplicant\_Income)



plot(x=df$Loan\_Amount,y=df$income)

