

Loan_prediction_problem_v1.R

jas

Thu Sep 08 21:32:07 2016

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.5

getwd()

## [1] "C:/Users/jas/Dropbox/Personal/New folder/Analytics Vidhya"

dir("C:/Users/ghumanja/Desktop/New folder/Analytics vidhya")

## character(0)

train <- read.csv('train_home loan.csv', na.strings=c("", "NA"))
head(train)

##   Loan_ID Gender Married Dependents Education Self_Employed
## 1 LP001002  Male      No           0 Graduate             No
## 2 LP001003  Male     Yes           1 Graduate             No
## 3 LP001005  Male     Yes           0 Graduate             Yes
## 4 LP001006  Male     Yes           0 Not Graduate          No
## 5 LP001008  Male     No            0 Graduate             No
## 6 LP001011  Male     Yes           2 Graduate             Yes
##   ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term
## 1             5849              0         NA             360
## 2             4583             1508         128             360
## 3             3000              0          66             360
## 4             2583             2358         120             360
## 5             6000              0         141             360
## 6             5417             4196         267             360
##   Credit_History Property_Area Loan_Status
## 1              1         Urban           Y
## 2              1         Rural           N
## 3              1         Urban           Y
## 4              1         Urban           Y
## 5              1         Urban           Y
## 6              1         Urban           Y

dim(train)

## [1] 614  13

typeof(train)

## [1] "list"
```

```
str(train)
```

```
## 'data.frame':      614 obs. of  13 variables:
## $ Loan_ID          : Factor w/ 614 levels "LP001002","LP001003",...: 1 2 3
4 5 6 7 8 9 10 ...
## $ Gender           : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2
2 2 ...
## $ Married          : Factor w/ 2 levels "No","Yes": 1 2 2 2 1 2 2 2 2 2
...
## $ Dependents       : Factor w/ 4 levels "0","1","2","3+": 1 2 1 1 1 3 1 4
3 2 ...
## $ Education        : Factor w/ 2 levels "Graduate","Not Graduate": 1 1 1
2 1 1 2 1 1 1 ...
## $ Self_Employed    : Factor w/ 2 levels "No","Yes": 1 1 2 1 1 2 1 1 1 1
...
## $ ApplicantIncome  : int   5849 4583 3000 2583 6000 5417 2333 3036 4006
12841 ...
## $ CoapplicantIncome: num    0 1508 0 2358 0 ...
## $ LoanAmount       : int   NA 128 66 120 141 267 95 158 168 349 ...
## $ Loan_Amount_Term : int   360 360 360 360 360 360 360 360 360 360 ...
## $ Credit_History   : int    1 1 1 1 1 1 1 0 1 1 ...
## $ Property_Area    : Factor w/ 3 levels "Rural","Semiurban",...: 3 1 3 3 3
3 3 2 3 2 ...
## $ Loan_Status      : Factor w/ 2 levels "N","Y": 2 1 2 2 2 2 2 1 2 1 ...
```

```
table(is.na(train))
```

```
##
## FALSE TRUE
## 7833 149
```

```
colSums(is.na(train))
```

##	Loan_ID	Gender	Married	Dependents
##	0	13	3	15
##	Education	Self_Employed	ApplicantIncome	CoapplicantIncome
##	0	32	0	0
##	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area
##	22	14	50	0
##	Loan_Status			
##	0			

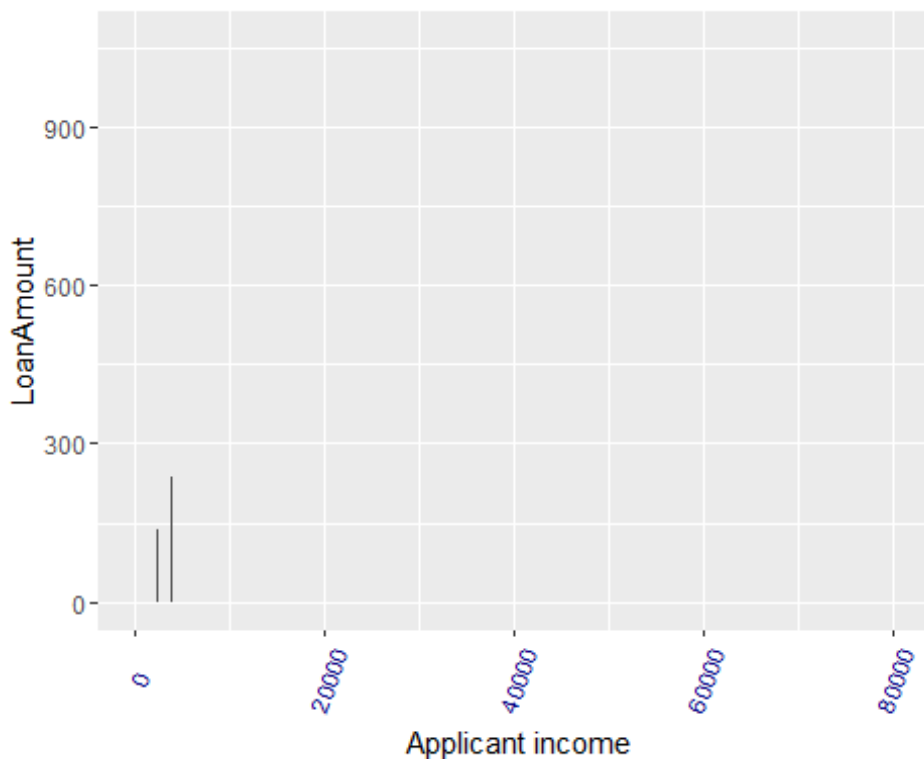
```
summary(train)
```

##	Loan_ID	Gender	Married	Dependents	Education
##	LP001002: 1	Female:112	No :213	0 :345	Graduate :480
##	LP001003: 1	Male :489	Yes :398	1 :102	Not Graduate:134
##	LP001005: 1	NA's : 13	NA's: 3	2 :101	
##	LP001006: 1			3+ : 51	
##	LP001008: 1			NA's: 15	
##	LP001011: 1				

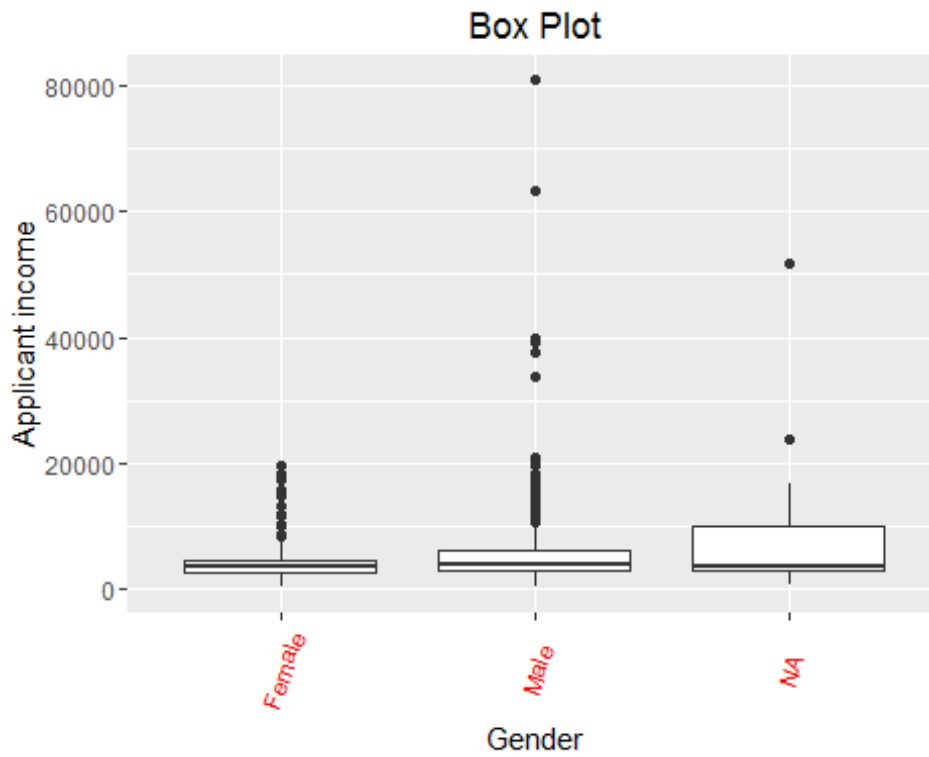
```
## (Other) :608
## Self_Employed ApplicantIncome CoapplicantIncome LoanAmount
## No :500 Min. : 150 Min. : 0 Min. : 9.0
## Yes : 82 1st Qu.: 2878 1st Qu.: 0 1st Qu.:100.0
## NA's: 32 Median : 3812 Median : 1188 Median :128.0
## Mean : 5403 Mean : 1621 Mean :146.4
## 3rd Qu.: 5795 3rd Qu.: 2297 3rd Qu.:168.0
## Max. :81000 Max. :41667 Max. :700.0
## NA's :22
## Loan_Amount_Term Credit_History Property_Area Loan_Status
## Min. : 12 Min. :0.0000 Rural :179 N:192
## 1st Qu.:360 1st Qu.:1.0000 Semiurban:233 Y:422
## Median :360 Median :1.0000 Urban :202
## Mean :342 Mean :0.8422
## 3rd Qu.:360 3rd Qu.:1.0000
## Max. :480 Max. :1.0000
## NA's :14 NA's :50
```

```
ggplot(data=train, aes(ApplicantIncome,LoanAmount))+geom_bar( stat =
"identity") +theme(axis.text.x = element_text(angle = 70, vjust = 0.5, color
= "navy")) + xlab("Applicant income")
```

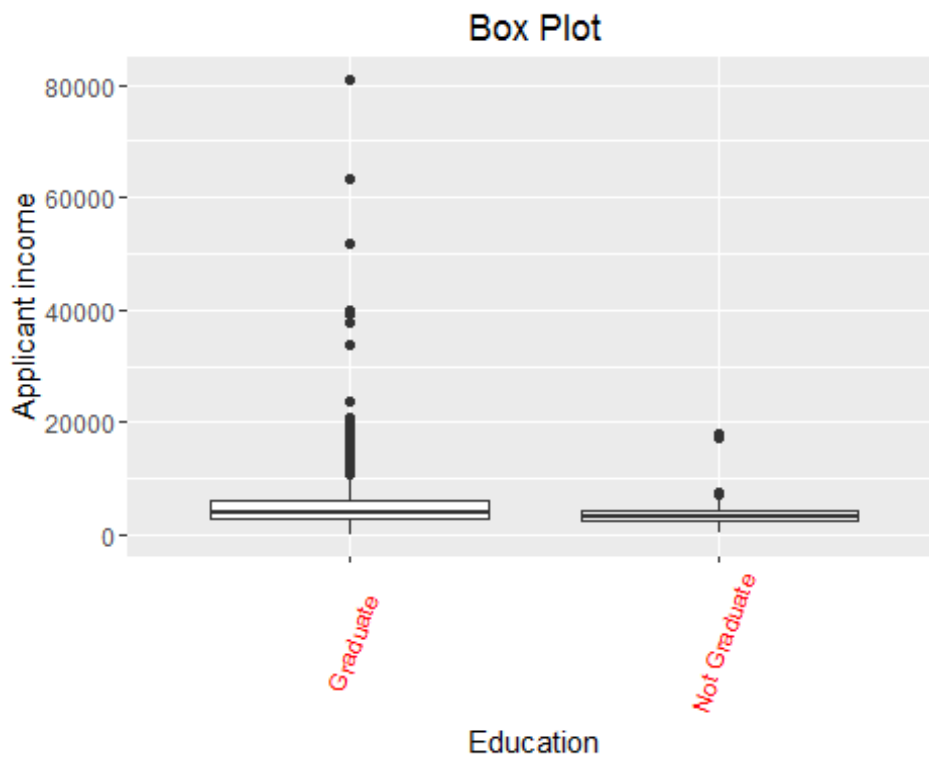
```
## Warning: Removed 22 rows containing missing values (position_stack).
```



```
ggplot(data=train, aes(Gender,ApplicantIncome))+geom_boxplot() +ggtitle("Box
Plot") + theme(axis.text.x = element_text(angle = 70, vjust = 0.5, color =
"red")) + ylab("Applicant income")+xlab("Gender")
```



```
ggplot(data=train, aes(Education,ApplicantIncome))+geom_boxplot()+
ggtitle("Box Plot") + theme(axis.text.x = element_text(angle = 70, vjust =
0.5, color = "red")) + ylab("Applicant income")+xlab("Education")
```



```
summary(train$Loan_Status)
```

```
##    N    Y
```

```
## 192 422
```

```
table(train$Credit_History)
```

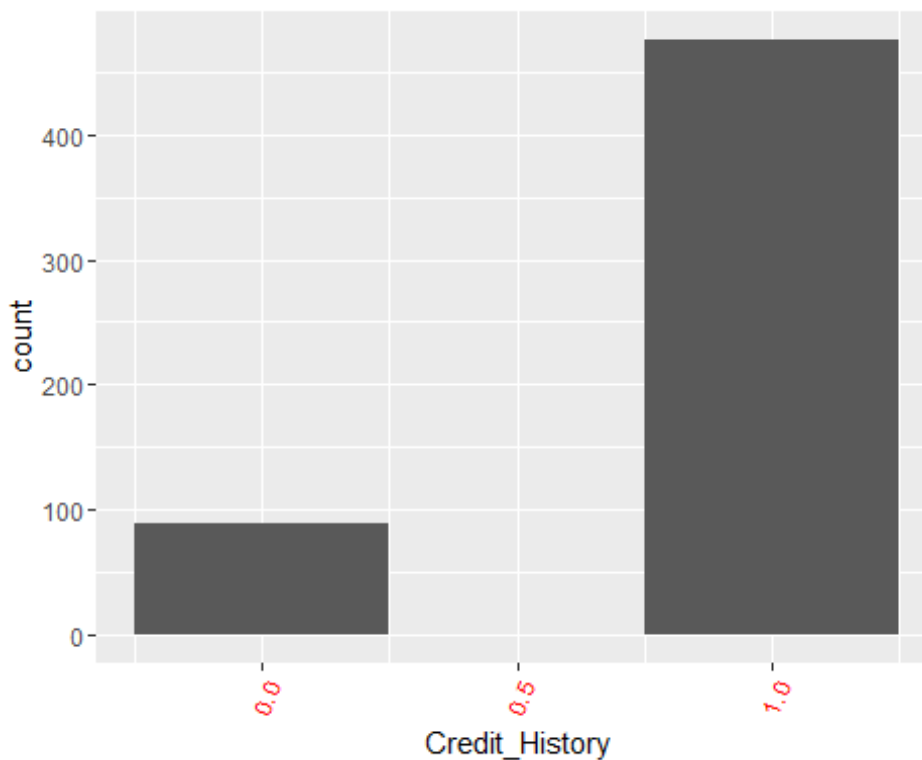
```
##
```

```
##    0    1
```

```
##   89 475
```

```
ggplot(train, aes(Credit_History))+geom_bar(width = 0.5) +theme(axis.text.x =  
element_text(angle = 70, vjust = 0.5, color = "red"))
```

```
## Warning: Removed 50 rows containing non-finite values (stat_count).
```

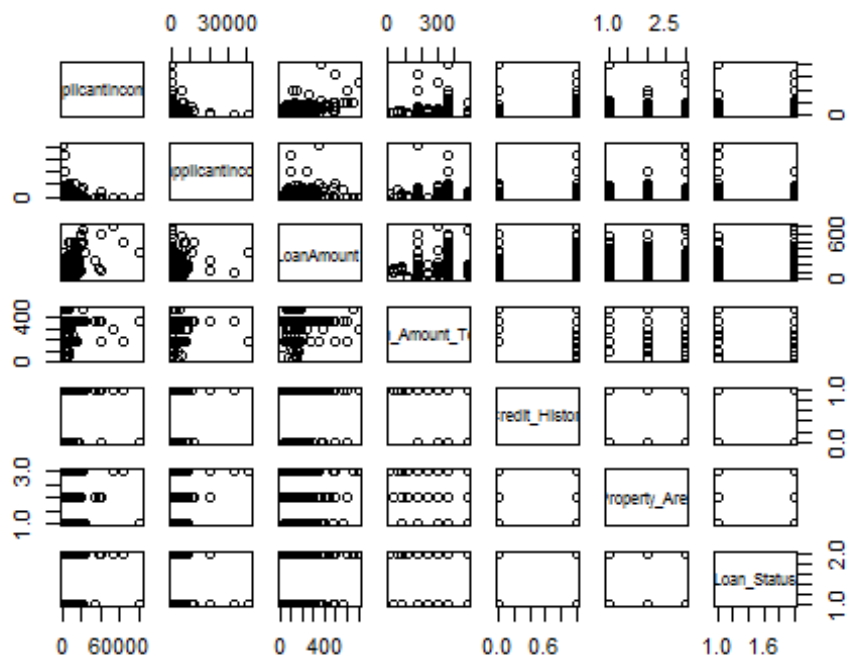


```
head(train)
```

```
##   Loan_ID Gender Married Dependents Education Self_Employed
## 1 LP001002  Male    No         0      Graduate          No
## 2 LP001003  Male   Yes         1      Graduate          No
## 3 LP001005  Male   Yes         0      Graduate          Yes
## 4 LP001006  Male   Yes         0 Not Graduate          No
## 5 LP001008  Male   No         0      Graduate          No
## 6 LP001011  Male   Yes         2      Graduate          Yes
## ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term
## 1          5849              0         NA          360
## 2          4583             1508        128          360
```

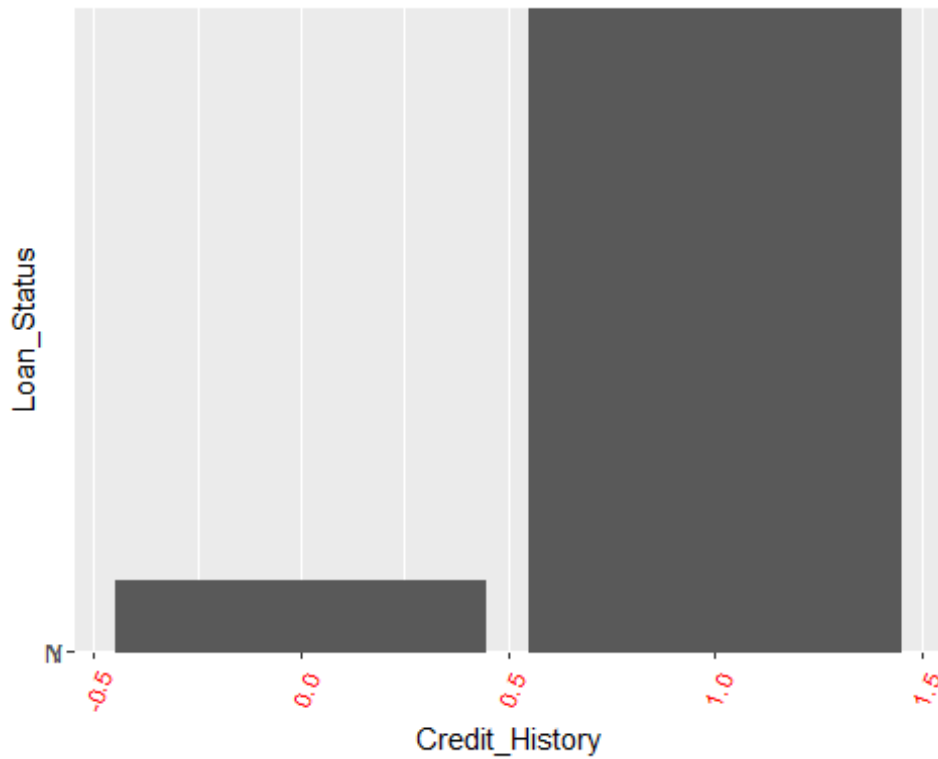
```
## 3          3000          0          66          360
## 4          2583        2358         120          360
## 5          6000          0         141          360
## 6          5417        4196         267          360
##   Credit_History Property_Area Loan_Status
## 1              1         Urban           Y
## 2              1         Rural           N
## 3              1         Urban           Y
## 4              1         Urban           Y
## 5              1         Urban           Y
## 6              1         Urban           Y
```

```
pairs(train[,7:13])
```



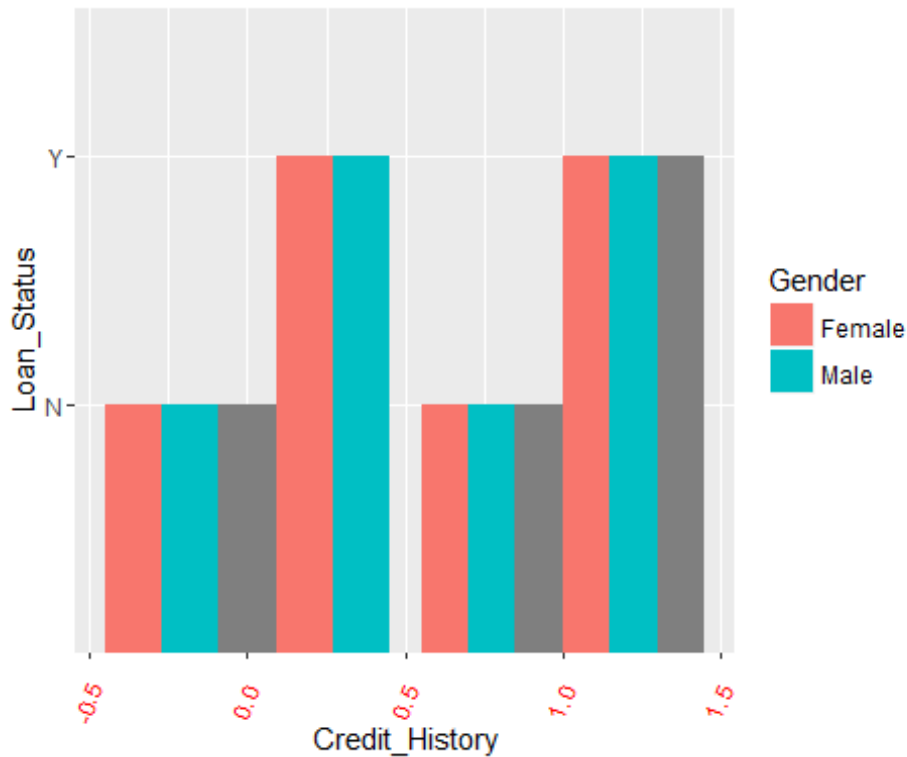
```
ggplot(train, aes(Credit_History,Loan_Status))+geom_bar(stat = "identity")
+theme(axis.text.x = element_text(angle = 70, vjust = 0.5, color = "red"))
```

```
## Warning: Removed 50 rows containing missing values (position_stack).
```



```
ggplot(train, aes(Credit_History, Loan_Status, fill=Gender))+geom_bar(position = "dodge", stat = "identity") +theme(axis.text.x = element_text(angle = 70, vjust = 0.2, color = "red"))
```

```
## Warning: Removed 50 rows containing missing values (geom_bar).
```



```
table(is.na(train))
```

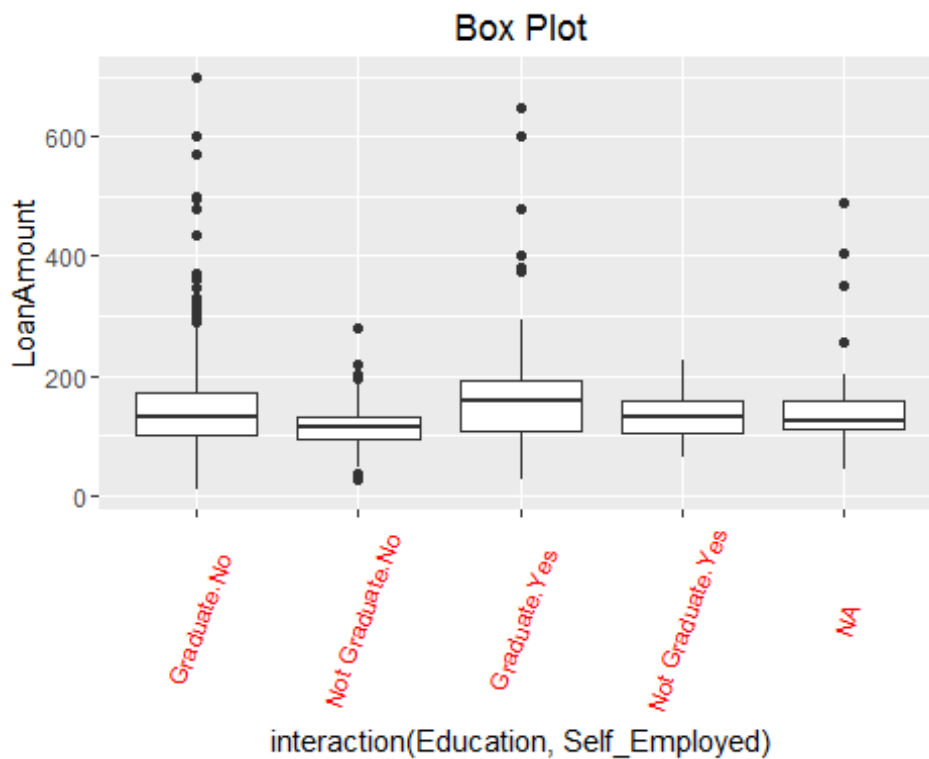
```
##
## FALSE TRUE
## 7833 149
```

```
colSums(is.na(train))
```

```
##      Loan_ID      Gender      Married      Dependents
##           0          13           3           15
##      Education  Self_Employed ApplicantIncome CoapplicantIncome
##           0             32           0             0
##      LoanAmount Loan_Amount_Term  Credit_History  Property_Area
##          22             14           50             0
##      Loan_Status
##           0
```

```
ggplot(data=train,
aes(interaction(Education,Self_Employed),LoanAmount))+geom_boxplot()
+ggtitle("Box Plot") + theme(axis.text.x = element_text(angle = 70, vjust =
0.5, color = "red"))
```

```
## Warning: Removed 22 rows containing non-finite values (stat_boxplot).
```

```
table(train$Self_Employed)
```

```
##
```

```
## No Yes
```

```
## 500 82
```

```
#Data imputation by mean, median etc#
```

```
#train$Self_Employed[which(is.na(train$Self_Employed))] <- "No"#
```

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 3.2.5
```

```
## Loading required package: Rcpp
```

```
## Warning: package 'Rcpp' was built under R version 3.2.5
```

```
## mice 2.25 2015-11-09
```

```
md.pattern(train)
```

```
##      Loan_ID Education ApplicantIncome CoapplicantIncome Property_Area
## 480         1         1              1              1          1
## 12          1         1              1              1          1
## 10          1         1              1              1          1
## 25          1         1              1              1          1
## 19          1         1              1              1          1
## 12          1         1              1              1          1
## 43          1         1              1              1          1
```

```

##      2      1      1      1      1      1      1
##      1      1      1      1      1      1      1
##      1      1      1      1      1      1      1
##      1      1      1      1      1      1      1
##      1      1      1      1      1      1      1
##      5      1      1      1      1      1      1
##      1      1      1      1      1      1      1
##      1      1      1      1      1      1      1
##      0      0      0      0      0      0      0
##      Loan_Status Married Gender Loan_Amount_Term Dependents LoanAmount
## 480      1      1      1      1      1      1
## 12      1      1      0      1      1      1
## 10      1      1      1      1      0      1
## 25      1      1      1      1      1      1
## 19      1      1      1      1      1      0
## 12      1      1      1      0      1      1
## 43      1      1      1      1      1      1
## 2      1      0      1      1      0      1
## 1      1      1      1      1      0      0
## 1      1      1      1      0      0      1
## 1      1      1      1      0      1      1
## 1      1      1      0      1      1      1
## 5      1      1      1      1      1      1
## 1      1      0      1      1      0      0
## 1      1      1      1      1      1      0
##      0      3     13      14      15      22
##      Self_Employed Credit_History
## 480      1      1      0
## 12      1      1      1
## 10      1      1      1
## 25      0      1      1
## 19      1      1      1
## 12      1      1      1
## 43      1      0      1
## 2      1      1      2
## 1      1      1      2
## 1      1      1      2
## 1      0      1      2
## 1      1      0      2
## 5      0      0      2
## 1      1      1      3
## 1      0      0      3
##      32      50 149

```

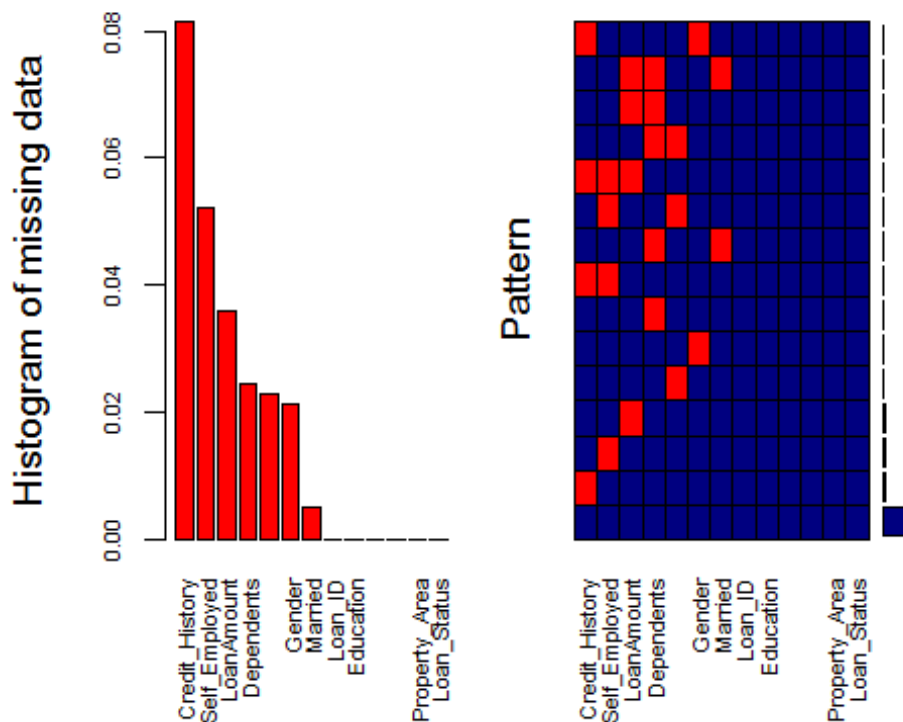
```
library(VIM)
```

```
## Warning: package 'VIM' was built under R version 3.2.5
```

```
## Loading required package: colorspace
```

```
## Warning: package 'colorspace' was built under R version 3.2.5
```

```
## Loading required package: grid
## Loading required package: data.table
## Warning: package 'data.table' was built under R version 3.2.5
## VIM is ready to use.
## Since version 4.0.0 the GUI is in its own package VIMGUI.
##
## Please use the package to use the new (and old) GUI.
## Suggestions and bug-reports can be submitted at:
## https://github.com/alexkova/VIM/issues
##
## Attaching package: 'VIM'
## The following object is masked from 'package:datasets':
##
## sleep
#Plot missing values and check#
aggr_plot <- aggr(train, col=c('navyblue','red'), numbers=TRUE,
sortVars=TRUE, labels=names(train), cex.axis=.7, gap=3, ylab=c("Histogram of
missing data","Pattern"))
## Warning in plot.aggr(res, ...): not enough horizontal space to display
## frequencies
```



```
##
## Variables sorted by number of missings:
##      Variable      Count
##      Credit_History 0.081433225
##      Self_Employed 0.052117264
##      LoanAmount 0.035830619
##      Dependents 0.024429967
##      Loan_Amount_Term 0.022801303
##      Gender 0.021172638
##      Married 0.004885993
##      Loan_ID 0.000000000
##      Education 0.000000000
##      ApplicantIncome 0.000000000
##      CoapplicantIncome 0.000000000
##      Property_Area 0.000000000
##      Loan_Status 0.000000000
```

```
labels(train)
```

```
## [[1]]
## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11"
## [12] "12" "13" "14" "15" "16" "17" "18" "19" "20" "21" "22"
## [23] "23" "24" "25" "26" "27" "28" "29" "30" "31" "32" "33"
## [34] "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44"
## [45] "45" "46" "47" "48" "49" "50" "51" "52" "53" "54" "55"
## [56] "56" "57" "58" "59" "60" "61" "62" "63" "64" "65" "66"
## [67] "67" "68" "69" "70" "71" "72" "73" "74" "75" "76" "77"
## [78] "78" "79" "80" "81" "82" "83" "84" "85" "86" "87" "88"
## [89] "89" "90" "91" "92" "93" "94" "95" "96" "97" "98" "99"
## [100] "100" "101" "102" "103" "104" "105" "106" "107" "108" "109" "110"
## [111] "111" "112" "113" "114" "115" "116" "117" "118" "119" "120" "121"
## [122] "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143"
## [144] "144" "145" "146" "147" "148" "149" "150" "151" "152" "153" "154"
## [155] "155" "156" "157" "158" "159" "160" "161" "162" "163" "164" "165"
## [166] "166" "167" "168" "169" "170" "171" "172" "173" "174" "175" "176"
## [177] "177" "178" "179" "180" "181" "182" "183" "184" "185" "186" "187"
## [188] "188" "189" "190" "191" "192" "193" "194" "195" "196" "197" "198"
## [199] "199" "200" "201" "202" "203" "204" "205" "206" "207" "208" "209"
## [210] "210" "211" "212" "213" "214" "215" "216" "217" "218" "219" "220"
## [221] "221" "222" "223" "224" "225" "226" "227" "228" "229" "230" "231"
## [232] "232" "233" "234" "235" "236" "237" "238" "239" "240" "241" "242"
## [243] "243" "244" "245" "246" "247" "248" "249" "250" "251" "252" "253"
## [254] "254" "255" "256" "257" "258" "259" "260" "261" "262" "263" "264"
## [265] "265" "266" "267" "268" "269" "270" "271" "272" "273" "274" "275"
## [276] "276" "277" "278" "279" "280" "281" "282" "283" "284" "285" "286"
## [287] "287" "288" "289" "290" "291" "292" "293" "294" "295" "296" "297"
## [298] "298" "299" "300" "301" "302" "303" "304" "305" "306" "307" "308"
## [309] "309" "310" "311" "312" "313" "314" "315" "316" "317" "318" "319"
## [320] "320" "321" "322" "323" "324" "325" "326" "327" "328" "329" "330"
```

```
## [331] "331" "332" "333" "334" "335" "336" "337" "338" "339" "340" "341"
## [342] "342" "343" "344" "345" "346" "347" "348" "349" "350" "351" "352"
## [353] "353" "354" "355" "356" "357" "358" "359" "360" "361" "362" "363"
## [364] "364" "365" "366" "367" "368" "369" "370" "371" "372" "373" "374"
## [375] "375" "376" "377" "378" "379" "380" "381" "382" "383" "384" "385"
## [386] "386" "387" "388" "389" "390" "391" "392" "393" "394" "395" "396"
## [397] "397" "398" "399" "400" "401" "402" "403" "404" "405" "406" "407"
## [408] "408" "409" "410" "411" "412" "413" "414" "415" "416" "417" "418"
## [419] "419" "420" "421" "422" "423" "424" "425" "426" "427" "428" "429"
## [430] "430" "431" "432" "433" "434" "435" "436" "437" "438" "439" "440"
## [441] "441" "442" "443" "444" "445" "446" "447" "448" "449" "450" "451"
## [452] "452" "453" "454" "455" "456" "457" "458" "459" "460" "461" "462"
## [463] "463" "464" "465" "466" "467" "468" "469" "470" "471" "472" "473"
## [474] "474" "475" "476" "477" "478" "479" "480" "481" "482" "483" "484"
## [485] "485" "486" "487" "488" "489" "490" "491" "492" "493" "494" "495"
## [496] "496" "497" "498" "499" "500" "501" "502" "503" "504" "505" "506"
## [507] "507" "508" "509" "510" "511" "512" "513" "514" "515" "516" "517"
## [518] "518" "519" "520" "521" "522" "523" "524" "525" "526" "527" "528"
## [529] "529" "530" "531" "532" "533" "534" "535" "536" "537" "538" "539"
## [540] "540" "541" "542" "543" "544" "545" "546" "547" "548" "549" "550"
## [551] "551" "552" "553" "554" "555" "556" "557" "558" "559" "560" "561"
## [562] "562" "563" "564" "565" "566" "567" "568" "569" "570" "571" "572"
## [573] "573" "574" "575" "576" "577" "578" "579" "580" "581" "582" "583"
## [584] "584" "585" "586" "587" "588" "589" "590" "591" "592" "593" "594"
## [595] "595" "596" "597" "598" "599" "600" "601" "602" "603" "604" "605"
## [606] "606" "607" "608" "609" "610" "611" "612" "613" "614"
##
## [[2]]
## [1] "Loan_ID"          "Gender"            "Married"
## [4] "Dependents"       "Education"         "Self_Employed"
## [7] "ApplicantIncome"  "CoapplicantIncome" "LoanAmount"
## [10] "Loan_Amount_Term" "Credit_History"   "Property_Area"
## [13] "Loan_Status"
```

#https://rpubs.com/corey_sparks/63681 . follow this link. Good#

```
train1 <- mice(data = train[,c("Gender", "Married",
"Dependents", "Education", "Self_Employed", "ApplicantIncome", "CoapplicantIncome",
"LoanAmount", "Loan_Amount_Term", "Credit_History", "Property_Area", "Loan_Status")], m=5, maxit=50, seed = 500)
```

```
##
## iter imp variable
## 1 1 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 1 2 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 1 3 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 1 4 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
```

[illegible]

	6	5	Gender	Married	Dependents	Self_Employed	LoanAmount
##	7	1	Gender	Married	Dependents	Self_Employed	LoanAmount
##	7	2	Gender	Married	Dependents	Self_Employed	LoanAmount
##	7	3	Gender	Married	Dependents	Self_Employed	LoanAmount
##	7	4	Gender	Married	Dependents	Self_Employed	LoanAmount
##	7	5	Gender	Married	Dependents	Self_Employed	LoanAmount
##	8	1	Gender	Married	Dependents	Self_Employed	LoanAmount
##	8	2	Gender	Married	Dependents	Self_Employed	LoanAmount
##	8	3	Gender	Married	Dependents	Self_Employed	LoanAmount
##	8	4	Gender	Married	Dependents	Self_Employed	LoanAmount
##	8	5	Gender	Married	Dependents	Self_Employed	LoanAmount
##	9	1	Gender	Married	Dependents	Self_Employed	LoanAmount
##	9	2	Gender	Married	Dependents	Self_Employed	LoanAmount
##	9	3	Gender	Married	Dependents	Self_Employed	LoanAmount
##	9	4	Gender	Married	Dependents	Self_Employed	LoanAmount
##	9	5	Gender	Married	Dependents	Self_Employed	LoanAmount
##	10	1	Gender	Married	Dependents	Self_Employed	LoanAmount
##	10	2	Gender	Married	Dependents	Self_Employed	LoanAmount
##	10	3	Gender	Married	Dependents	Self_Employed	LoanAmount
##	10	4	Gender	Married	Dependents	Self_Employed	LoanAmount
##	10	5	Gender	Married	Dependents	Self_Employed	LoanAmount
##	11	1	Gender	Married	Dependents	Self_Employed	LoanAmount
##	11	2	Gender	Married	Dependents	Self_Employed	LoanAmount
##	11	3	Gender	Married	Dependents	Self_Employed	LoanAmount
##	11	4	Gender	Married	Dependents	Self_Employed	LoanAmount
##	11	5	Gender	Married	Dependents	Self_Employed	LoanAmount

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

## 46 5 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 47 1 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 47 2 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 47 3 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 47 4 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 47 5 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 48 1 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 48 2 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 48 3 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 48 4 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 48 5 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 49 1 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 49 2 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 49 3 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 49 4 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 49 5 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 50 1 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 50 2 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 50 3 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 50 4 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History
## 50 5 Gender Married Dependents Self_Employed LoanAmount
Loan_Amount_Term Credit_History

```

```
summary(train1)
```

```
## Multiply imputed data set
```

```
## Call:
```

```
## mice(data = train[, c("Gender", "Married", "Dependents", "Education",
##   "Self_Employed", "ApplicantIncome", "CoapplicantIncome",
##   "LoanAmount", "Loan_Amount_Term", "Credit_History", "Property_Area",
```

```

##      "Loan_Status")]], m = 5, maxit = 50, seed = 500)
## Number of multiple imputations: 5
## Missing cells per column:
##      Gender      Married      Dependents      Education
##      13           3           15           0
##      Self_Employed ApplicantIncome CoapplicantIncome LoanAmount
##      32           0           0           22
##      Loan_Amount_Term Credit_History Property_Area Loan_Status
##      14           50           0           0
## Imputation methods:
##      Gender      Married      Dependents      Education
##      "logreg"     "logreg"     "polyreg"     ""
##      Self_Employed ApplicantIncome CoapplicantIncome LoanAmount
##      "logreg"     ""           ""           "pmm"
##      Loan_Amount_Term Credit_History Property_Area Loan_Status
##      "pmm"       "pmm"       ""           ""
## VisitSequence:
##      Gender      Married      Dependents      Self_Employed
##      1           2           3           5
##      LoanAmount Loan_Amount_Term Credit_History
##      8           9           10
## PredictorMatrix:
##      Gender Married Dependents Education Self_Employed
## Gender      0      1      1      1      1
## Married      1      0      1      1      1
## Dependents   1      1      0      1      1
## Education     0      0      0      0      0
## Self_Employed 1      1      1      1      0
## ApplicantIncome 0      0      0      0      0
## CoapplicantIncome 0      0      0      0      0
## LoanAmount     1      1      1      1      1
## Loan_Amount_Term 1      1      1      1      1
## Credit_History  1      1      1      1      1
## Property_Area   0      0      0      0      0
## Loan_Status     0      0      0      0      0
##      ApplicantIncome CoapplicantIncome LoanAmount
## Gender      1      1      1
## Married      1      1      1
## Dependents   1      1      1
## Education     0      0      0
## Self_Employed 1      1      1
## ApplicantIncome 0      0      0
## CoapplicantIncome 0      0      0
## LoanAmount     1      1      0
## Loan_Amount_Term 1      1      1
## Credit_History  1      1      1
## Property_Area   0      0      0
## Loan_Status     0      0      0
##      Loan_Amount_Term Credit_History Property_Area
## Gender      1      1      1

```



```
## Married 1 1 1
## Dependents 1 1 1
## Education 0 0 0
## Self_Employed 1 1 1
## ApplicantIncome 0 0 0
## CoapplicantIncome 0 0 0
## LoanAmount 1 1 1
## Loan_Amount_Term 0 1 1
## Credit_History 1 0 1
## Property_Area 0 0 0
## Loan_Status 0 0 0
```

```
## Loan_Status
## Gender 1
## Married 1
## Dependents 1
## Education 0
## Self_Employed 1
## ApplicantIncome 0
## CoapplicantIncome 0
## LoanAmount 1
## Loan_Amount_Term 1
## Credit_History 1
## Property_Area 0
## Loan_Status 0
## Random generator seed value: 500
```

#This is to see the variability in the 5 different imputations for each outcome#

```
names(train)
```

```
## [1] "Loan_ID" "Gender" "Married"
## [4] "Dependents" "Education" "Self_Employed"
## [7] "ApplicantIncome" "CoapplicantIncome" "LoanAmount"
## [10] "Loan_Amount_Term" "Credit_History" "Property_Area"
## [13] "Loan_Status"
```

```
train2 <- with(data=train1
,exp=lm(Loan_Status~Gender+Married+Dependents+Education+Self_Employed+ApplicantIncome+CoapplicantIncome+LoanAmount+Loan_Amount_Term+Credit_History+Property_Area))
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a
## factor response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a
## factor response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```

## Warning in model.response(mf, "numeric"): using type = "numeric" with a
## factor response will be ignored

## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors

## Warning in model.response(mf, "numeric"): using type = "numeric" with a
## factor response will be ignored

## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors

## Warning in model.response(mf, "numeric"): using type = "numeric" with a
## factor response will be ignored

## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors

train2

## call :
## with.mids(data = train1, expr = lm(Loan_Status ~ Gender + Married +
##   Dependents + Education + Self_Employed + ApplicantIncome +
##   CoapplicantIncome + LoanAmount + Loan_Amount_Term + Credit_History +
##   Property_Area))
##
## call1 :
## mice(data = train[, c("Gender", "Married", "Dependents", "Education",
##   "Self_Employed", "ApplicantIncome", "CoapplicantIncome",
##   "LoanAmount", "Loan_Amount_Term", "Credit_History", "Property_Area",
##   "Loan_Status")], m = 5, maxit = 50, seed = 500)
##
## nmis :
##           Gender           Married           Dependents           Education
##           13              3              15              0
##   Self_Employed ApplicantIncome CoapplicantIncome           LoanAmount
##           32              0              0              22
##   Loan_Amount_Term Credit_History           Property_Area           Loan_Status
##           14              50              0              0
##
## analyses :
## [[1]]
##
## Call:
## lm(formula = Loan_Status ~ Gender + Married + Dependents + Education +
##   Self_Employed + ApplicantIncome + CoapplicantIncome + LoanAmount +
##   Loan_Amount_Term + Credit_History + Property_Area)
##
## Coefficients:
##   (Intercept)           Gender2           Married2
##   1.074e+00      5.169e-04      9.230e-02
##   Dependents2      Dependents3      Dependents4
##   -6.223e-02      2.546e-02      -1.649e-03
##   Education2      Self_Employed2      ApplicantIncome

```

```

##          -4.773e-02          1.647e-02          9.584e-07
## CoapplicantIncome      LoanAmount      Loan_Amount_Term
##          -9.241e-06          -2.444e-04          -1.381e-04
##      Credit_History      Property_Area2      Property_Area3
##          7.127e-01          1.255e-01          3.108e-02
##
##
## [[2]]
##
## Call:
## lm(formula = Loan_Status ~ Gender + Married + Dependents + Education +
##      Self_Employed + ApplicantIncome + CoapplicantIncome + LoanAmount +
##      Loan_Amount_Term + Credit_History + Property_Area)
##
## Coefficients:
##      (Intercept)          Gender2          Married2
##      1.110e+00          -3.430e-03          8.810e-02
##      Dependents2      Dependents3      Dependents4
##      -6.417e-02          2.507e-02          -5.733e-03
##      Education2      Self_Employed2      ApplicantIncome
##      -6.153e-02          -2.457e-03          1.914e-06
## CoapplicantIncome      LoanAmount      Loan_Amount_Term
##      -9.059e-06          -3.522e-04          -1.549e-04
##      Credit_History      Property_Area2      Property_Area3
##      6.963e-01          1.311e-01          3.609e-02
##
##
## [[3]]
##
## Call:
## lm(formula = Loan_Status ~ Gender + Married + Dependents + Education +
##      Self_Employed + ApplicantIncome + CoapplicantIncome + LoanAmount +
##      Loan_Amount_Term + Credit_History + Property_Area)
##
## Coefficients:
##      (Intercept)          Gender2          Married2
##      1.063e+00          4.929e-03          9.237e-02
##      Dependents2      Dependents3      Dependents4
##      -6.197e-02          2.607e-02          6.639e-03
##      Education2      Self_Employed2      ApplicantIncome
##      -5.172e-02          2.928e-02          6.681e-07
## CoapplicantIncome      LoanAmount      Loan_Amount_Term
##      -1.084e-05          -2.287e-04          -1.527e-04
##      Credit_History      Property_Area2      Property_Area3
##      7.121e-01          1.377e-01          4.836e-02
##
##
## [[4]]
##
## Call:

```

[illegible]

```

print(train3)

## Call: pool(object = train2)
##
## Pooled coefficients:
##      (Intercept)      Gender2      Married2      Dependents2
##      1.081481e+00      2.353873e-03      9.263618e-02      -6.298524e-02
##      Dependents3      Dependents4      Education2      Self_Employed2
##      2.729063e-02      9.838884e-04      -5.816442e-02      8.793945e-03
##      ApplicantIncome CoapplicantIncome      LoanAmount      Loan_Amount_Term
##      1.608077e-06      -9.333187e-06      -3.271587e-04      -1.375821e-04
##      Credit_History      Property_Area2      Property_Area3
##      7.097754e-01      1.309825e-01      3.523838e-02
##
## Fraction of information about the coefficients missing due to nonresponse:
##      (Intercept)      Gender2      Married2      Dependents2
##      NA      NA      NA      NA
##      Dependents3      Dependents4      Education2      Self_Employed2
##      NA      NA      NA      NA
##      ApplicantIncome CoapplicantIncome      LoanAmount      Loan_Amount_Term
##      NA      NA      NA      NA
##      Credit_History      Property_Area2      Property_Area3
##      NA      NA      NA

```