Predicting insurance purchase for Indian farmers STAT 471/571/701, Fall 2018

Contents

troduction
Background
Goal of the study
The data
Characteristics of the Data Set
Description of variables
esearch approach
Analyses of dataset
EDA on variables in the dataset
Split dataset into test and training
Model Building

Introduction

Background

Goal of the study

The data

Characteristics of the Data Set

Description of variables

Research approach

Analyses of dataset

```
data.main<- read.csv("insurance_dataset.csv", header = T)</pre>
str(data.main)
                   46568 obs. of 87 variables:
## 'data.frame':
## $ Customer.main.type
                                                     : int 88831059883 ...
## $ Customer.Subtype
                                                            33 37 37 9 40 23 39 33 33 11 ...
## $ Number.of.houses
                                                            1 1 1 1 1 1 2 1 1 2 ...
                                                     : int
## $ Avg.size.household
                                                            3 2 2 3 4 2 3 2 2 3 ...
                                                     : int
                                                            2 2 2 3 2 1 2 3 4 3 ...
## $ Avg.age
## $ Roman.catholic
                                                            0 1 0 2 1 0 2 0 0 3 ...
                                                            5 4 4 3 4 5 2 7 1 5 ...
   $ Protestant
                                                     : int 1 1 2 2 1 0 0 0 3 0 ...
## $ Other.religion
## $ No.religion
                                                     : int
                                                            3 4 4 4 4 5 5 2 6 2 ...
                                                            7 6 3 5 7 0 7 7 6 7 ...
## $ Married
## $ Living.together
                                                     : int 0 2 2 2 1 6 2 2 0 0 ...
                                                     : int 2 2 4 2 2 3 0 0 3 2 ...
## $ Other.relation
## $ Singles
                                                     : int 1 0 4 2 2 3 0 0 3 2 ...
```

```
$ Household.without.children
                                                         2 4 4 3 4 5 3 5 3 2 ...
   $ Household.with.children
                                                   : int
                                                         6524426436...
                                                         1003500000...
## $ High.level.education
                                                   : int
## $ Medium.level.education
                                                         2554454314...
                                                   : int
##
   $ Lower.level.education
                                                   : int
                                                         7 4 4 2 0 4 5 6 8 5 ...
##
   $ High.status
                                                   : int
                                                         1 0 0 4 0 2 0 2 1 2 ...
##
  $ Entrepreneur
                                                         0000500010...
                                                         1000400000...
##
   $ Farmer
                                                   : int
   $ Middle.management
                                                   : int
                                                         2 5 7 3 0 4 4 2 1 3 ...
##
   $ Skilled.labourers
                                                   : int
                                                         5 0 0 1 0 2 1 5 8 3 ...
   $ Unskilled.labourers
                                                         2 4 2 2 0 2 5 2 1 3 ...
##
   $ Social.class.A
                                                         1 0 0 3 9 2 0 2 1 1 ...
                                                   : int
   $ Social.class.B1
                                                         1 2 5 2 0 2 1 1 1 2 ...
##
                                                   : int
                                                         2 3 0 1 0 2 4 2 0 1 ...
##
  $ Social.class.B2
                                                   : int
   $ Social.class.C
                                                   : int
                                                         6 5 4 4 0 4 5 5 8 4 ...
##
   $ Social.class.D
                                                   : int
                                                         1 0 0 0 0 2 0 2 1 2 ...
##
   $ Rented.house
                                                   : int
                                                         1 2 7 5 4 9 6 0 9 0 ...
                                                         8724503909...
##
   $ Home.owners
                                                   : int
##
  $ X1.car
                                                   : int
                                                         8779658456 ...
   $ X2.cars
                                                         0 1 0 0 2 3 0 4 2 1 ...
##
                                                   : int
                                                   : int
##
   $ No.car
                                                         1 2 2 0 1 3 1 2 3 2 ...
##
  $ National.Health.Service
                                                   : int
                                                         8 6 9 7 5 9 9 6 7 6 ...
   $ Private.health.insurance
                                                         1 3 0 2 4 0 0 3 2 3 ...
##
                                                   : int
                                                         0 2 4 1 0 5 4 2 7 2 ...
   $ Income...30
                                                   : int
                                                   : int
                                                         4 0 5 5 0 2 3 5 2 3 ...
##
   $ Income.30.45.000
   $ Income.45.75.000
                                                   : int
                                                         5 5 0 3 9 3 3 3 1 3 ...
##
   $ Income.75.122.000
                                                   : int
                                                         0 2 0 0 0 0 0 0 0 1 ...
   $ Income..123.000
                                                         0 0 0 0 0 0 0 0 0 0 ...
##
                                                   : int
##
                                                   : int
                                                         4 5 3 4 6 3 3 3 2 4 ...
  $ Average.income
                                                         3 4 4 4 3 3 5 3 3 7 ...
## $ Purchasing.power.class
                                                   : int
                                                         55446888 30947608 52223731 76547439 100719
##
   $ pr_num
                                                   : int
   $ Contribution.private.third.party.insurance
                                                   : int
                                                         0 2 2 0 0 0 0 0 0 2 ...
   $ Contribution.third.party.insurance..firms.
                                                         0 0 0 0 0 0 0 0 0 0 ...
                                                   : int
## $ Contribution.third.party.insurane..agriculture. : int
                                                         0 0 0 0 0 0 0 0 0 0 ...
                                                         6 0 6 6 0 6 6 0 5 0 ...
## $ Contribution.car.policies
                                                   : int
## $ Contribution.delivery.van.policies
                                                   : int
                                                         0000000000...
## $ Contribution.motorcycle.scooter.policies
                                                   : int
                                                         0000000000...
## $ Contribution.lorry.policies
                                                   : int
                                                         0000000000...
##
   $ Contribution.trailer.policies
                                                   : int
                                                         0 0 0 0 0 0 0 0 0 0 ...
                                                         0 0 0 0 0 0 0 0 0 0 ...
##
   $ Contribution.tractor.policies
                                                   : int
## $ Contribution.agricultural.machines.policies
                                                   : int
                                                         0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.moped.policies
                                                   : int 000000300...
   $ Contribution.life.insurances
                                                   : int
                                                         0000000000...
## $ Contribution.private.accident.insurance.policies: int
                                                         0 0 0 0 0 0 0 0 0 0 ...
  $ Contribution.family.accidents.insurance.policies: int
                                                         0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.disability.insurance.policies : int
##
                                                         0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.fire.policies
                                                         5 2 2 2 6 0 0 0 0 3 ...
                                                   : int
                                                         0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.surfboard.policies
                                                  : int
## $ Contribution.boat.policies
                                                   : int
                                                         0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.bicycle.policies
                                                         0 0 0 0 0 0 0 0 0 0 ...
                                                   : int
## $ Contribution.property.insurance.policies
                                                         0 0 0 0 0 0 0 0 0 0 ...
                                                  : int
## $ Contribution.social.security.insurance.policies : int 0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.private.3rd.party.insurance
                                                 : int 021000001...
## $ No..of.3rd.party.insurance..firms.
                                                  : int 0000000000...
```

```
$ No..of.3rd.party.insurane..agriculture.
                                                 : int
                                                       0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.car.policies
                                                       1011011010...
                                                 : int
## $ No..of.delivery.van.policies
                                                       0000000000...
                                                : int
## $ No..of.motorcycle.scooter.policies
                                                       0000000000...
                                                 : int
## $ No..of.lorry.policies
                                                 : int
                                                       0000000000...
## $ No..of.trailer.policies
                                                       0 0 0 0 0 0 0 0 0 0 ...
                                                 : int
## $ No..of.tractor.policies
                                                       0000000000...
                                                 : int
## $ No..of.agricultural.machines.policies
                                                       0000000000...
                                                 : int
## $ No..of.moped.policies
                                                 : int
                                                       0 0 0 0 0 0 0 1 0 0 ...
## $ No..of.life.insurances
                                                 : int
                                                       0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.private.accident.insurance.policies
                                                 : int
                                                       0 0 0 0 0 0 0 0 0 0 ...
                                                       0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.family.accidents.insurance.policies
                                                 : int
## $ No..of.disability.insurance.policies
                                                       0 0 0 0 0 0 0 0 0 0 ...
                                                 : int
## $ No..of.fire.policies
                                                       1 1 1 1 1 0 0 0 0 1 ...
                                                 : int
## $ No..of.surfboard.policies
                                                 : int
                                                       0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.boat.policies
                                                 : int
                                                       0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.bicycle.policies
                                                 : int 0000000000...
## $ No..of.property.insurance.policies
                                                 : int 0000000000...
## $ No..of.social.security.insurance.policies
                                                 : int 0000000000...
                                                 : int 0000000000...
## $ No..of.mobile.home.policies
#find out if dataset has missing values
sum(is.na(data.main))
```

[1] 2891271

```
#analyze which columns, rows have missing values
```

#it seems there are empty rows at the bottom of the dataset, so these can simply be deleted
data <- na.omit(data.main) #there are empty rows in the dataset, to omit these data
str(data)</pre>

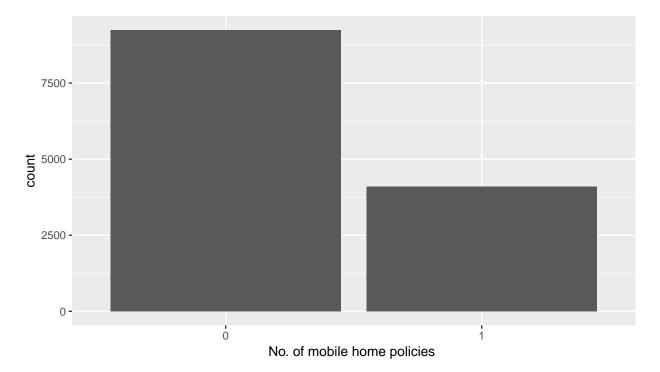
```
## 'data.frame':
                  13335 obs. of 87 variables:
## $ Customer.main.type
                                                        8 8 8 3 10 5 9 8 8 3 ...
                                                  : int
                                                        33 37 37 9 40 23 39 33 33 11 ...
## $ Customer.Subtype
                                                  : int
## $ Number.of.houses
                                                        1 1 1 1 1 1 2 1 1 2 ...
                                                  : int
## $ Avg.size.household
                                                  : int
                                                        3 2 2 3 4 2 3 2 2 3 ...
                                                        2 2 2 3 2 1 2 3 4 3 ...
## $ Avg.age
                                                  : int
   $ Roman.catholic
                                                        0 1 0 2 1 0 2 0 0 3 ...
                                                 : int
                                                 : int 5 4 4 3 4 5 2 7 1 5 ...
## $ Protestant
## $ Other.religion
                                                 : int 1 1 2 2 1 0 0 0 3 0 ...
                                                        3 4 4 4 4 5 5 2 6 2 ...
## $ No.religion
                                                 : int
   $ Married
                                                 : int
                                                        7635707767...
                                                 : int 0 2 2 2 1 6 2 2 0 0 ...
## $ Living.together
                                                        2 2 4 2 2 3 0 0 3 2 ...
## $ Other.relation
                                                 : int
                                                 : int 1042230032...
## $ Singles
## $ Household.without.children
                                                 : int
                                                        2 4 4 3 4 5 3 5 3 2 ...
## $ Household.with.children
                                                 : int 6524426436...
## $ High.level.education
                                                 : int 1003500000...
## $ Medium.level.education
                                                        2 5 5 4 4 5 4 3 1 4 ...
                                                 : int
## $ Lower.level.education
                                                 : int 7 4 4 2 0 4 5 6 8 5 ...
## $ High.status
                                                 : int 1004020212...
                                                 : int 0000500010...
## $ Entrepreneur
## $ Farmer
                                                  : int 100040000 ...
                                                 : int 2573044213 ...
## $ Middle.management
```

```
$ Skilled.labourers
                                                        5 0 0 1 0 2 1 5 8 3 ...
   $ Unskilled.labourers
                                                  : int
                                                        2 4 2 2 0 2 5 2 1 3 ...
                                                        1003920211...
##
   $ Social.class.A
  $ Social.class.B1
                                                        1 2 5 2 0 2 1 1 1 2 ...
##
                                                  : int
##
   $ Social.class.B2
                                                  : int
                                                        2 3 0 1 0 2 4 2 0 1 ...
##
   $ Social.class.C
                                                  : int
                                                        6 5 4 4 0 4 5 5 8 4 ...
   $ Social.class.D
                                                        1000020212...
##
   $ Rented.house
##
                                                  : int
                                                        1 2 7 5 4 9 6 0 9 0 ...
##
   $ Home.owners
                                                  : int
                                                        8724503909...
##
   $ X1.car
                                                  : int
                                                        8779658456 ...
##
  $ X2.cars
                                                        0 1 0 0 2 3 0 4 2 1 ...
##
   $ No.car
                                                        1 2 2 0 1 3 1 2 3 2 ...
                                                  : int
   $ National.Health.Service
                                                        8697599676...
                                                  : int
##
  $ Private.health.insurance
                                                        1 3 0 2 4 0 0 3 2 3 ...
                                                  : int
   $ Income...30
                                                  : int
                                                        0 2 4 1 0 5 4 2 7 2 ...
                                                        4 0 5 5 0 2 3 5 2 3 ...
##
   $ Income.30.45.000
                                                  : int
##
   $ Income.45.75.000
                                                  : int
                                                        5 5 0 3 9 3 3 3 1 3 ...
## $ Income.75.122.000
                                                        0 2 0 0 0 0 0 0 0 1 ...
                                                  : int
## $ Income..123.000
                                                        0000000000...
                                                        4 5 3 4 6 3 3 3 2 4 ...
## $ Average.income
                                                  : int
## $ Purchasing.power.class
                                                  : int
                                                        3 4 4 4 3 3 5 3 3 7 ...
                                                        55446888 30947608 52223731 76547439 100719
## $ Contribution.private.third.party.insurance
                                                        0 2 2 0 0 0 0 0 0 2 ...
                                                 : int
   $ Contribution.third.party.insurance..firms.
                                                  : int
                                                        0000000000...
   $ Contribution.third.party.insurane..agriculture. : int 0 0 0 0 0 0 0 0 0 0 ...
  $ Contribution.car.policies
                                                 : int
                                                        6066066050 ...
##
   $ Contribution.delivery.van.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.motorcycle.scooter.policies
                                                        0 0 0 0 0 0 0 0 0 0 ...
                                                  : int
## $ Contribution.lorry.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.trailer.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.tractor.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.agricultural.machines.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.moped.policies
                                                        0 0 0 0 0 0 0 3 0 0 ...
                                                  : int
## $ Contribution.life.insurances
                                                  : int 0000000000...
## $ Contribution.private.accident.insurance.policies: int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.family.accidents.insurance.policies: int
                                                        0000000000...
## $ Contribution.disability.insurance.policies : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.fire.policies
                                                  : int
                                                        5 2 2 2 6 0 0 0 0 3 ...
   $ Contribution.surfboard.policies
                                                  : int
                                                        0000000000...
                                                        0000000000...
##
   $ Contribution.boat.policies
                                                 : int
## $ Contribution.bicycle.policies
                                                 : int 0000000000...
## $ Contribution.property.insurance.policies
                                                 : int
                                                        0000000000...
   $ Contribution.social.security.insurance.policies : int
                                                        0000000000...
##
   $ No..of.private.3rd.party.insurance
                                                        0 2 1 0 0 0 0 0 0 1 ...
                                                : int
  $ No..of.3rd.party.insurance..firms.
                                                        0 0 0 0 0 0 0 0 0 0 ...
                                                  : int
   $ No..of.3rd.party.insurane..agriculture.
                                                        0 0 0 0 0 0 0 0 0 0 ...
##
                                                 : int
   $ No..of.car.policies
                                                        1 0 1 1 0 1 1 0 1 0 ...
                                                  : int
##
   $ No..of.delivery.van.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.motorcycle.scooter.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.lorry.policies
                                                        0 0 0 0 0 0 0 0 0 0 ...
                                                  : int
## $ No..of.trailer.policies
                                                 : int
                                                        0000000000...
## $ No..of.tractor.policies
                                                 : int 0000000000...
## $ No..of.agricultural.machines.policies
                                                : int 0000000000...
                                                  : int 000000100...
## $ No..of.moped.policies
```

```
$ No..of.life.insurances
                                                          0 0 0 0 0 0 0 0 0 0 ...
                                                   : int
## $ No..of.private.accident.insurance.policies
                                                          0000000000...
                                                   : int
## $ No..of.family.accidents.insurance.policies
                                                   : int
                                                          0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.disability.insurance.policies
                                                          0 0 0 0 0 0 0 0 0 0 ...
                                                   : int
## $ No..of.fire.policies
                                                   : int
                                                          1 1 1 1 1 0 0 0 0 1 ...
## $ No..of.surfboard.policies
                                                          0 0 0 0 0 0 0 0 0 0 ...
                                                   : int
## $ No..of.boat.policies
                                                          0 0 0 0 0 0 0 0 0 0 ...
                                                   : int
## $ No..of.bicycle.policies
                                                   : int
                                                          0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.property.insurance.policies
                                                   : int
                                                          0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.social.security.insurance.policies
                                                   : int 0000000000...
## $ No..of.mobile.home.policies
                                                   : int 0000000000...
## - attr(*, "na.action")= 'omit' Named int 13336 13337 13338 13339 13340 13341 13342 13343 13344 133
   ..- attr(*, "names")= chr "13336" "13337" "13338" "13339" ...
View (data)
```

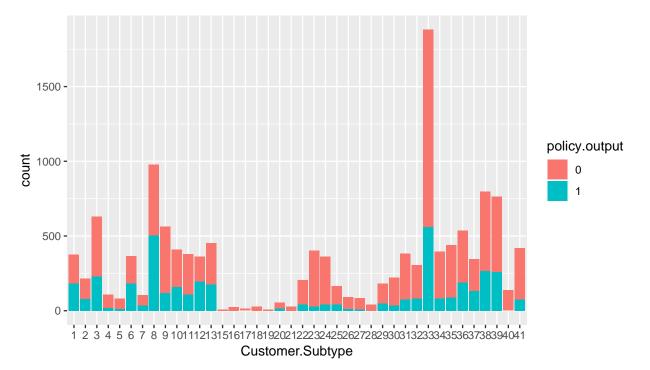
EDA on variables in the dataset

```
#RESPONSE VARIABLE - convert the response variable to a factor and rename it to "policy.output" for con
data <- na.omit(data.main)
data <- data %>% rename(policy.output = No..of.mobile.home.policies )
data <- data %>% mutate(policy.output = as.factor(policy.output))
ggplot(data,aes(x=policy.output)) + geom_bar() + labs(x="No. of mobile home policies ")
```



#There is about a 2/3rd split in response variable i.e. 1/3rd of the data population has a mobile polic #to determine which variables should be considered ion our model, we plot each variable and see if there

```
# Var 44 (pr_num) is ignored for this analysis as it is an accounting or identification variable, and
# ANALYZING VARIABLE 182 - The Customer types
head(data$Customer.main.type)
## [1] 8 8 8 3 10 5
head(data$Customer.Subtype)
## [1] 33 37 37 9 40 23
#These varaibles are clearly factors, which have been found from previous demographic and psychographic
data$Customer.main.type <- as.factor(data$Customer.main.type)</pre>
data$Customer.Subtype <- as.factor(data$Customer.Subtype)</pre>
#Plot these varaibles, check if they have any correlation with the output varaible, and check if there
plot<-ggplot(data,aes(x=Customer.main.type, fill= policy.output))</pre>
plot<-plot + geom bar()</pre>
plot<-plot + labs(x="Customer Main Type")</pre>
#There is reasonable variation, all levels are represented. This variable should be left as is
plot<-ggplot(data,aes(x=Customer.Subtype, fill= policy.output))</pre>
plot<-plot + geom_bar()</pre>
plot
```

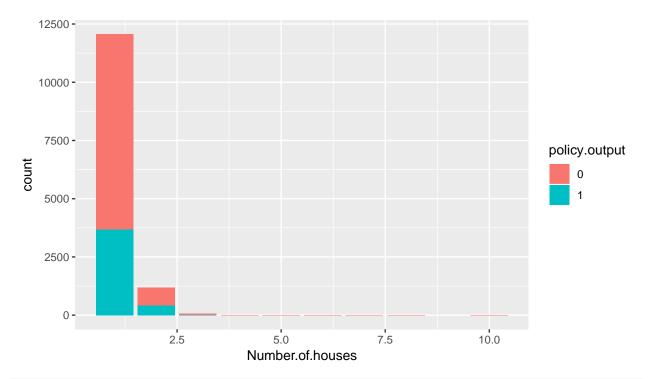


#There is reasonable variation, all levels are represented. This variable should be left as is

#Analyzing var 3- number of houses
head(data\$Number.of.houses)

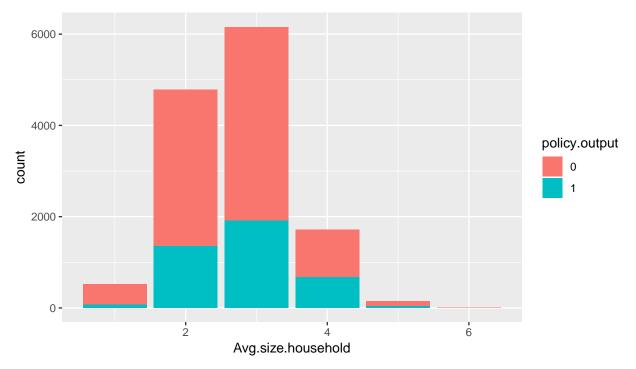
[1] 1 1 1 1 1 1

```
#These are integer values
plot<-ggplot(data,aes(x=Number.of.houses, fill= policy.output))
plot<-plot + geom_bar()
plot</pre>
```



#Data is highly skewed, and could benefit from a log transformation, but since the number is so low (10 #Analyzing var 4- avg size household head(data\$Avg.size.household)

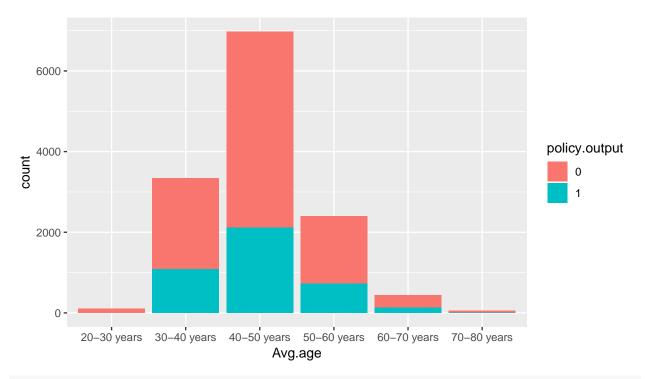
```
## [1] 3 2 2 3 4 2
#These are integer values
plot<-ggplot(data,aes(x=Avg.size.household, fill= policy.output))
plot<-plot + geom_bar()
plot</pre>
```



```
#Data is normal and has significant variation, leave the variable as is
#Analyzing var 5- avg age
head(data$Avg.age)
```

[1] 2 2 2 3 2 1

```
#PLOT data
plot<-ggplot(data,aes(x=Avg.age, fill= policy.output))
plot<-plot + geom_bar()
plot</pre>
```



#Data is normal and has significant variation, and is not interpretable. We can move on data\$Roman.catholic

$[1] \ 0 \ 1 \ 0 \ 2 \ 1 \ 0 \ 2 \ 0 \ 0 \ 3 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 2 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 2 \ 1 \ 0$ [35] 0 2 0 0 0 0 1 1 0 2 1 0 2 0 0 0 1 0 0 1 0 0 0 1 0 0 4 0 0 0 4 2 0 ## ## ## [103] 3 0 1 2 0 0 2 0 2 0 0 1 0 3 1 0 0 0 0 1 0 2 2 0 1 0 0 1 1 0 1 4 1 1 [137] 0 2 0 0 0 0 1 0 0 2 1 0 5 0 0 2 0 2 2 1 0 0 1 9 0 0 0 2 0 0 2 2 1 0 ## ## ## [205] 3 5 0 1 0 0 0 1 1 0 1 0 1 0 0 2 0 0 0 0 2 2 0 0 2 1 2 0 1 0 0 0 1 0 [239] 0 0 9 1 0 0 5 0 1 0 1 0 2 0 1 0 1 0 2 3 0 1 1 1 1 0 0 1 1 0 0 0 ## [273] 1 3 0 1 2 1 0 1 0 2 1 0 1 0 0 1 0 3 0 1 1 0 2 0 2 2 0 2 0 1 0 0 1 0 ## ## [307] 0 0 0 2 1 0 2 0 0 2 0 0 1 0 0 0 2 1 0 3 0 4 0 1 3 1 1 0 0 0 0 1 0 [341] 0 0 0 1 0 0 1 1 0 2 0 2 0 0 0 1 0 0 0 1 4 2 0 2 2 0 0 1 1 0 2 0 1 1 ## ## ## [409] 0 1 2 3 0 1 0 0 0 0 0 0 0 2 1 1 0 0 1 0 1 1 2 1 0 2 1 0 0 1 1 0 0 0 ## ## ## [511] 0 0 2 0 0 4 2 0 1 1 0 0 2 1 1 1 0 0 0 1 0 1 0 0 1 0 0 0 2 2 0 0 1 0 ## ## [613] 2 0 3 1 1 0 1 0 0 0 0 1 0 1 0 1 2 3 0 0 0 0 0 1 0 1 0 0 1 1 2 0 1 0 ## ## [647] 1 0 0 0 2 1 3 0 0 2 0 1 0 0 1 4 0 1 0 0 0 1 0 0 0 0 2 0 1 0 0 0 1 ## [681] 4 0 4 3 0 1 1 1 2 2 0 0 1 0 0 0 1 0 1 0 0 2 1 0 0 1 1 0 0 1 0 0 1 0 ## ## ## [783] 1 0 1 0 0 0 0 0 0 1 2 0 0 1 1 0 0 0 1 0 2 0 0 0 2 0 0 0 1 5 0 0 2 0 ## ## ## [885] 0 0 0 1 0 1 2 0 1 1 2 0 0 0 0 0 3 0 0 0 2 0 0 0 0 0 0 0 1 0 2 0 1 0

data\$Income..123.000

[239] 0 0 2 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ## ## ##

[375] 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 2 4 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 ## [409] 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 0 ## ## [477] 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 0 0 2 0 0 0 ## ## [545] 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ## ## ## ## ## ## ## ## ## ## ## ## ## ## [1021] 0 1 0 0 0 0 0 0 1 0 2 0 4 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 ## ## ## ## ## $\begin{smallmatrix} 1225 \end{smallmatrix}] \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{$ ## [1259] 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 ## $\begin{smallmatrix} 1293 \end{smallmatrix}] \hspace{.1cm} 0 \hspace{.1cm} 2 \hspace{.1cm} 2 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 3 \hspace{.1cm} 0 \hspace{$ ## ## ## ## ## $\begin{smallmatrix} 1463 \end{smallmatrix}] \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{$ ## ## ## ## ## ## ## ## ## [1769] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 4 0 0 0 2 0 2 0 0 0 0 1 0 0 0 0 ## ## ## ## [1905] 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 2 0 1 0 0 0 3 0 0 0 0 0 0 ## [1939] 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 ## ## ## ## ## ##

[2211] 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 0 1 ## ## [2279] 3 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 ## ## ## ## ## ## ## ## ## ## ## $\begin{picture}(2653) \hline (2653) \hline (26$ ## ## ## ## ## ## ## [2857] 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 ## ## [2959] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 3 2 0 0 0 0 1 0 0 0 1 0 ## [2993] 1 0 0 0 1 0 0 0 0 1 1 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 ## ## [3027] 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 1 0 0 0 1 0 0 0 1 0 0 0 ## ## ## ## ## ## ## ## ## ## ## [3401] 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 ## ## ## ## ## [3605] 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 ## ## [3673] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 ## [3707] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 ## [3741] 0 0 0 0 0 0 0 0 2 1 0 0 0 0 0 0 0 4 0 0 1 0 0 0 0 1 0 0 ## ## ## ## [3843] 1 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 ## ## ## [3979] 0 0 0 1 0 1 0 0 0 0 0 1 4 0 0 2 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0

```
## [13227] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 2
## [13329] 0 0 0 0 0 0 0
#Variable 6-41 are variables with similar factor values (% of society). All of them are clearly factors
for (i in which(colnames(data) == "Roman.catholic"): which(colnames(data) == "Income..123.000")){
 data[,i] <- factor(data[,i],</pre>
                  levels=c(0:9),
                  labels=c("0\%",
                           "1-10%",
                           "11-23%",
                           "24-36%",
                           "37-49%",
                           "50-62%",
                           "63-75%",
                           "76-88%",
                           "89-99%",
                           "100%"))
}
str(data)
                  13335 obs. of 87 variables:
## 'data.frame':
## $ Customer.main.type
                                                   : Factor w/ 10 levels "1", "2", "3", "4", ...: 8 8 8 3
## $ Customer.Subtype
                                                   : Factor w/ 40 levels "1", "2", "3", "4", ...: 32 36 3
                                                   : int 1 1 1 1 1 1 2 1 1 2 ...
## $ Number.of.houses
                                                   : int 3 2 2 3 4 2 3 2 2 3 ...
## $ Avg.size.household
                                                   : Factor w/ 6 levels "20-30 years",..: 2 2 2 3 2
## $ Avg.age
## $ Roman.catholic
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 6
## $ Protestant
## $ Other.religion
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 4
## $ No.religion
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 8
## $ Married
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
## $ Living.together
## $ Other.relation
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
## $ Singles
                                                   : Factor w/ 10 levels "0%","1-10%","11-23%",...: 3
## $ Household.without.children
   $ Household.with.children
                                                   : Factor w/ 10 levels "0%","1-10%","11-23%",...: 7
##
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
## $ High.level.education
## $ Medium.level.education
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
                                                   : Factor w/ 10 levels "0%","1-10%","11-23%",...: 8
## $ Lower.level.education
## $ High.status
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%",...: 2
## $ Entrepreneur
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
## $ Farmer
                                                   : Factor w/ 10 levels "0%","1-10%","11-23%",...: 3
## $ Middle.management
## $ Skilled.labourers
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 6
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
## $ Unskilled.labourers
## $ Social.class.A
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
   $ Social.class.B1
## $ Social.class.B2
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
## $ Social.class.C
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 7
## $ Social.class.D
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
## $ Rented.house
                                                   : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 9
## $ Home.owners
```

```
$ X1.car
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 9
##
   $ X2.cars
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
##
   $ No.car
   $ National.Health.Service
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 9
##
   $ Private.health.insurance
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
##
   $ Income...30
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
   $ Income.30.45.000
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 5
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 6
   $ Income.45.75.000
##
   $ Income.75.122.000
                                                    : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
                                                    : Factor w/ 10 levels "0%","1-10%","11-23%",...: 1
##
   $ Income..123.000
   $ Average.income
                                                    : int 4534633324 ...
                                                           3 4 4 4 3 3 5 3 3 7 ...
##
   $ Purchasing.power.class
                                                    : int
                                                           55446888 30947608 52223731 76547439 100719
##
   $ pr num
                                                    : int
## $ Contribution.private.third.party.insurance
                                                    : int
                                                           0 2 2 0 0 0 0 0 0 2 ...
## $ Contribution.third.party.insurance..firms.
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.third.party.insurane..agriculture. : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.car.policies
                                                           6 0 6 6 0 6 6 0 5 0 ...
                                                    : int
## $ Contribution.delivery.van.policies
                                                           0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.motorcycle.scooter.policies
                                                           0000000000...
                                                    : int
   $ Contribution.lorry.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.trailer.policies
                                                    : int
                                                           0000000000...
## $ Contribution.tractor.policies
                                                           0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.agricultural.machines.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.moped.policies
                                                           0 0 0 0 0 0 0 3 0 0 ...
                                                    : int
## $ Contribution.life.insurances
                                                    : int 0000000000...
## $ Contribution.private.accident.insurance.policies: int
                                                           0000000000...
   $ Contribution.family.accidents.insurance.policies: int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.disability.insurance.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
                                                           5 2 2 2 6 0 0 0 0 3 ...
## $ Contribution.fire.policies
                                                    : int
## $ Contribution.surfboard.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
##
   $ Contribution.boat.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.bicycle.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.property.insurance.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.social.security.insurance.policies : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.private.3rd.party.insurance
                                                    : int
                                                           0 2 1 0 0 0 0 0 0 1 ...
   $ No..of.3rd.party.insurance..firms.
                                                           0000000000...
                                                    : int
## $ No..of.3rd.party.insurane..agriculture.
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.car.policies
                                                    : int
                                                           1 0 1 1 0 1 1 0 1 0 ...
   $ No..of.delivery.van.policies
                                                           0 0 0 0 0 0 0 0 0 0 ...
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
##
   $ No..of.motorcycle.scooter.policies
                                                    : int
## $ No..of.lorry.policies
                                                           0 0 0 0 0 0 0 0 0 0 ...
##
   $ No..of.trailer.policies
                                                    : int
                                                           0000000000...
   $ No..of.tractor.policies
                                                           0000000000...
                                                    : int
## $ No..of.agricultural.machines.policies
                                                           0 0 0 0 0 0 0 0 0 0 ...
                                                    : int
   $ No..of.moped.policies
                                                           0 0 0 0 0 0 0 1 0 0 ...
                                                    : int
   $ No..of.life.insurances
##
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.private.accident.insurance.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.family.accidents.insurance.policies
                                                           0 0 0 0 0 0 0 0 0 0 ...
                                                    : int
## $ No..of.disability.insurance.policies
                                                    : int
                                                           0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.fire.policies
                                                    : int
                                                           1 1 1 1 1 0 0 0 0 1 ...
## $ No..of.surfboard.policies
                                                    : int
                                                           0000000000...
                                                    : int 0000000000...
## $ No..of.boat.policies
## $ No..of.bicycle.policies
                                                    : int 0000000000...
## $ No..of.property.insurance.policies
                                                    : int 0000000000...
```

```
## $ No..of.social.security.insurance.policies
                                                       : int 0000000000...
                                                       : Factor w/ 2 levels "0", "1": 1 1 1 1 1 1 1 1 1 1
## $ policy.output
#Variables Average.income and Purchasing.power.class are written as percentages as per the dataset, how
for (i in which(colnames(data) == "Average.income"): which(colnames(data) == "Purchasing.power.class"))
  {
  data[,i] <- factor(data[,i],</pre>
                   levels=c(0:8),
                   labels=c("RS 0",
                             "Rs 1 - 49",
                             "RS 50 - 99",
                             "Rs 100 - 199",
                             "RS 200 - 499",
                             "RS 500 - 999"
                             "RS 1000 - 4999",
                             "RS 5000 - 9999",
                             "RS 10000 - 19999"
                             ))
}
str(data)
                    13335 obs. of 87 variables:
## 'data.frame':
                                                        : Factor w/ 10 levels "1", "2", "3", "4",...: 8 8 8 3
## $ Customer.main.type
                                                        : Factor w/ 40 levels "1","2","3","4",...: 32 36 3
##
   $ Customer.Subtype
## $ Number.of.houses
                                                        : int 1 1 1 1 1 1 2 1 1 2 ...
## $ Avg.size.household
                                                        : int 3 2 2 3 4 2 3 2 2 3 ...
                                                        : Factor w/ 6 levels "20-30 years",..: 2 2 2 3 2
## $ Avg.age
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
## $ Roman.catholic
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 6
## $ Protestant
## $ Other.religion
                                                        : Factor w/ 10 levels "0%","1-10%","11-23%",...: 2
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 4
## $ No.religion
## $ Married
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 8
## $ Living.together
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
## $ Other.relation
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
## $ Singles
## $ Household.without.children
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
## $ Household.with.children
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 7
                                                        : Factor w/ 10 levels "0%","1-10%","11-23%",...: 2
## $ High.level.education
    $ Medium.level.education
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%",...: 3
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 8
## $ Lower.level.education
## $ High.status
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
## $ Entrepreneur
## $ Farmer
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%",...: 2
## $ Middle.management
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 6
## $ Skilled.labourers
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
## $ Unskilled.labourers
## $ Social.class.A
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
## $ Social.class.B1
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 3
## $ Social.class.B2
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 7
   $ Social.class.C
##
## $ Social.class.D
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
##
  $ Rented.house
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 9
## $ Home.owners
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 9
##
   $ X1.car
## $ X2.cars
                                                        : Factor w/ 10 levels "0%", "1-10%", "11-23%",...: 1
```

```
##
   $ No.car
                                                      : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
   $ National.Health.Service
                                                      : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 9
   $ Private.health.insurance
                                                     : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 2
                                                      : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
##
   $ Income...30
##
   $ Income.30.45.000
                                                      : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 5
##
   $ Income.45.75.000
                                                      : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 6
   $ Income.75.122.000
                                                      : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
                                                      : Factor w/ 10 levels "0%", "1-10%", "11-23%", ...: 1
##
   $ Income..123.000
##
   $ Average.income
                                                      : Factor w/ 9 levels "RS 0", "Rs 1 - 49",..: 5 6 4
                                                      : Factor w/ 9 levels "RS 0", "Rs 1 - 49", ...: 4 5 5
##
   $ Purchasing.power.class
## $ pr_num
                                                            55446888 30947608 52223731 76547439 100719
##
   $ Contribution.private.third.party.insurance
                                                            0 2 2 0 0 0 0 0 0 2 ...
                                                      : int
   $ Contribution.third.party.insurance..firms.
                                                       int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.third.party.insurane..agriculture. : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.car.policies
                                                      : int
                                                            6 0 6 6 0 6 6 0 5 0 ...
##
   $ Contribution.delivery.van.policies
                                                       int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.motorcycle.scooter.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
  $ Contribution.lorry.policies
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.trailer.policies
                                                            0000000000...
                                                      : int
   $ Contribution.tractor.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.agricultural.machines.policies
                                                      : int
                                                            0000000000...
## $ Contribution.moped.policies
                                                            0 0 0 0 0 0 0 3 0 0 ...
                                                      : int
   $ Contribution.life.insurances
                                                            0 0 0 0 0 0 0 0 0 0 ...
##
                                                      : int
   $ Contribution.private.accident.insurance.policies: int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.family.accidents.insurance.policies: int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.disability.insurance.policies
                                                    : int
                                                            0000000000...
##
   $ Contribution.fire.policies
                                                            5 2 2 2 6 0 0 0 0 3 ...
                                                      : int
   $ Contribution.surfboard.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.boat.policies
                                                            0 0 0 0 0 0 0 0 0 0 ...
                                                      : int
## $ Contribution.bicycle.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
##
   $ Contribution.property.insurance.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.social.security.insurance.policies : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.private.3rd.party.insurance
                                                    : int
                                                            0 2 1 0 0 0 0 0 0 1 ...
  $ No..of.3rd.party.insurance..firms.
                                                            0 0 0 0 0 0 0 0 0 0 ...
                                                      : int
##
   $ No..of.3rd.party.insurane..agriculture.
                                                            0 0 0 0 0 0 0 0 0 0 ...
                                                      : int
  $ No..of.car.policies
                                                      : int
                                                            1011011010...
##
  $ No..of.delivery.van.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.motorcycle.scooter.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.lorry.policies
                                                            0 0 0 0 0 0 0 0 0 0 ...
##
                                                      : int
   $ No..of.trailer.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
##
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.tractor.policies
##
   $ No..of.agricultural.machines.policies
                                                            0000000000...
                                                      : int
   $ No..of.moped.policies
                                                      : int
                                                            0 0 0 0 0 0 0 1 0 0 ...
   $ No..of.life.insurances
##
                                                            0 0 0 0 0 0 0 0 0 0 ...
                                                      : int
   $ No..of.private.accident.insurance.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.family.accidents.insurance.policies
                                                            0 0 0 0 0 0 0 0 0 0 ...
##
                                                      : int
##
   $ No..of.disability.insurance.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.fire.policies
                                                      : int
                                                            1 1 1 1 1 0 0 0 0 1 ...
## $ No..of.surfboard.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.boat.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.bicycle.policies
                                                      : int
                                                            0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.property.insurance.policies
                                                      : int 0000000000...
## $ No..of.social.security.insurance.policies
                                                      : int 0000000000...
                                                      : Factor w/ 2 levels "0", "1": 1 1 1 1 1 1 1 1 1 1
## $ policy.output
```

```
#If any entry was not in the 0-9 range, it must be coded incorrectly, and would appear as NA now. Let u sum(is.na(data))
```

[1] 50

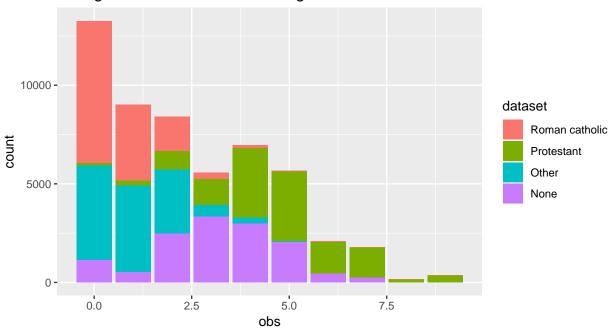
```
#50 NA varaibles, which are not coded properly, so can be removed.
data <- na.omit(data.main)
```

#Certain demographic and behavioral factors are another great place to explore.

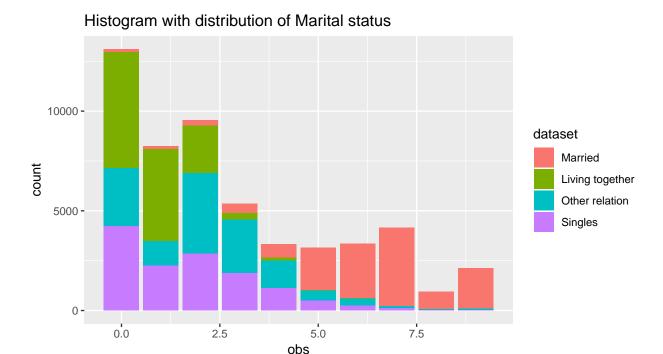
#Among the demographic factors, we thought religion, marital status, level of education, occupation, and

#Among the behavioral factors, such as contribution to and number of other insurance, may be influentia

Histogram with distribution of Religion

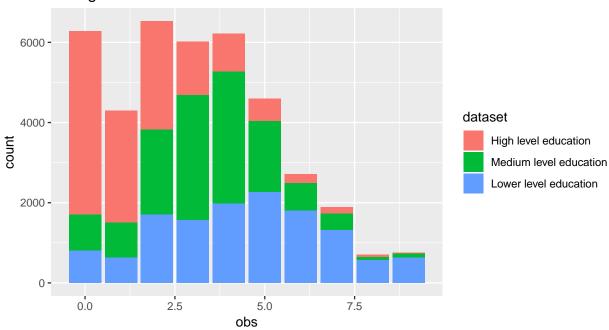


#We can see there is significant variation between each type of religion, and therefore these variables

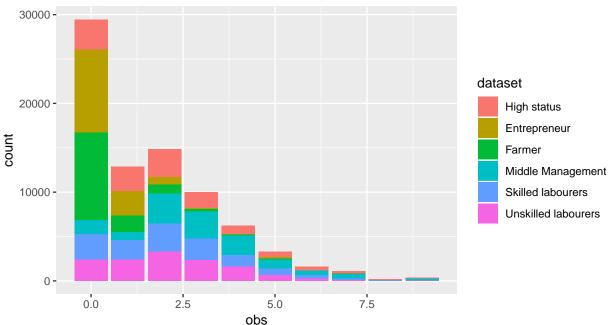


#We can see there is significant variation between each type of marital status, and therefore these var

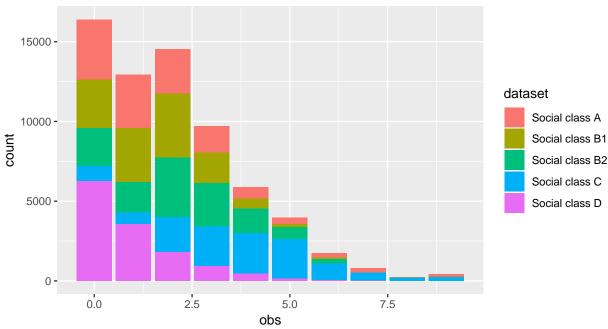


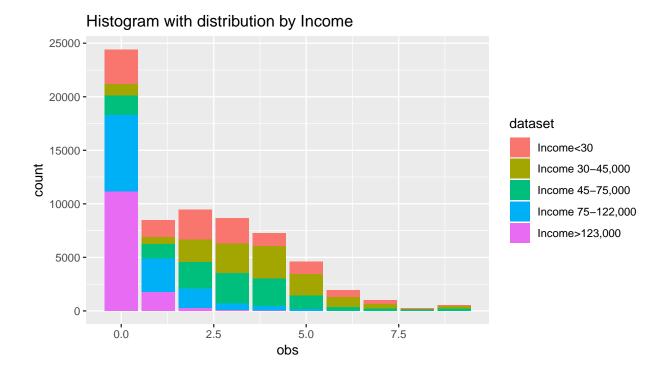












Split dataset into test and training

```
#split the training and test data sets, so we can proceed with analysis on training data set, and check
N <- length(data$pr_num)</pre>
set.seed(10)
index.train <- sample(N, 0.8*N)</pre>
data.train <- data[index.train,] #dim(data.train)</pre>
data.test <- data[-index.train,]</pre>
dim(data.train)
## [1] 10668
                87
dim(data.test)
## [1] 2667
str(data.train)
## 'data.frame':
                    10668 obs. of 87 variables:
##
   $ Customer.main.type
                                                              1538821681...
                                                       : int
   $ Customer.Subtype
                                                              3 23 11 36 33 7 4 26 33 1 ...
                                                              1 1 1 1 1 1 1 1 1 1 ...
##
   $ Number.of.houses
                                                         int
   $ Avg.size.household
                                                              3 2 3 2 3 3 2 2 3 2 ...
##
                                                         int
                                                              3 2 3 4 3 2 4 4 2 4 ...
##
  $ Avg.age
                                                         int
##
   $ Roman.catholic
                                                         int
                                                              1 0 1 2 0 1 0 0 0 0 ...
                                                              5 2 4 4 5 2 5 7 5 2 ...
   $ Protestant
##
                                                         int
                                                              2 1 1 2 1 1 2 0 0 0 ...
##
   $ Other.religion
                                                         int
                                                              3 7 4 2 4 5 3 2 4 7 ...
## $ No.religion
                                                         int
## $ Married
                                                         int
                                                              5 3 6 7 7 7 5 4 7 6 ...
##
   $ Living.together
                                                         int
                                                              1500211000...
##
   $ Other.relation
                                                              4 2 3 2 1 1 4 5 2 3 ...
                                                         int
                                                       : int 1 2 2 2 0 1 3 4 1 5 ...
  $ Singles
```

```
$ Household.without.children
                                                          4 6 2 4 5 3 4 5 0 3 ...
   $ Household.with.children
                                                    : int
                                                          4 2 5 3 4 5 3 1 8 2 ...
                                                          3 0 1 1 2 3 2 1 0 3 ...
  $ High.level.education
## $ Medium.level.education
                                                          5 4 4 5 3 5 4 3 0 4 ...
                                                    : int
##
   $ Lower.level.education
                                                    : int
                                                          1554424592...
##
   $ High.status
                                                    : int
                                                          2021433237...
##
  $ Entrepreneur
                                                          1010011000...
##
   $ Farmer
                                                          0 0 1 0 0 0 1 0 2 0 ...
                                                    : int
   $ Middle.management
                                                     int
                                                          6 4 3 3 1 4 3 4 3 0 ...
##
   $ Skilled.labourers
                                                          1 2 3 1 4 2 2 1 1 2 ...
                                                    : int
   $ Unskilled.labourers
                                                    : int
                                                          1 4 2 4 2 2 2 3 3 0 ...
   $ Social.class.A
                                                    : int
                                                          3 0 2 1 3 2 2 1 0 4 ...
##
   $ Social.class.B1
                                                          2 2 2 1 0 3 2 3 2 0 ...
                                                     int
                                                          4 3 2 4 2 3 2 2 7 4 ...
##
  $ Social.class.B2
                                                    : int
##
   $ Social.class.C
                                                    : int
                                                          1544424211...
                                                          1 0 1 0 1 1 1 3 0 0 ...
##
   $ Social.class.D
                                                    : int
##
   $ Rented.house
                                                    : int
                                                          1 6 3 2 3 3 4 8 9 0 ...
##
   $ Home.owners
                                                    : int
                                                          8 3 6 7 6 6 5 1 0 9 ...
##
  $ X1.car
                                                    : int 5565576259 ...
                                                          1 1 1 3 4 1 1 1 3 0 ...
##
   $ X2.cars
                                                    : int
##
   $ No.car
                                                    : int
                                                          3 4 2 2 1 2 3 7 2 0 ...
##
  $ National.Health.Service
                                                          5 9 6 5 6 5 6 8 5 2 ...
   $ Private.health.insurance
                                                          4 0 3 4 4 4 3 1 4 7 ...
##
                                                    : int
   $ Income...30
                                                    : int
                                                          1 2 3 2 1 2 4 4 1 2 ...
##
   $ Income.30.45.000
                                                   : int
                                                          4 6 2 5 5 3 2 4 6 2 ...
   $ Income.45.75.000
                                                    : int
                                                          2 2 3 3 2 4 3 2 3 5 ...
##
   $ Income.75.122.000
                                                    : int
                                                          4 0 1 0 0 2 1 0 0 0 ...
   $ Income..123.000
                                                          1 0 1 0 2 1 0 0 0 0 ...
                                                    : int
##
                                                    : int
                                                          6 3 4 4 6 5 3 3 4 4 ...
  $ Average.income
                                                          6 3 6 3 3 6 6 1 3 8 ...
## $ Purchasing.power.class
                                                    : int
                                                          38915379 47056194 69337147 54625419 184956
##
   $ pr_num
                                                    : int
   $ Contribution.private.third.party.insurance
                                                    : int
                                                          0 2 2 0 0 2 0 2 0 2 ...
                                                          0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.third.party.insurance..firms.
                                                    : int
## $ Contribution.third.party.insurane..agriculture. : int
                                                          0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.car.policies
                                                    : int
                                                          0660066060...
## $ Contribution.delivery.van.policies
                                                    : int
                                                          0000000000...
## $ Contribution.motorcycle.scooter.policies
                                                   : int
                                                          0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.lorry.policies
                                                    : int
                                                          0000000000...
   $ Contribution.trailer.policies
                                                    : int
                                                          0 0 0 0 2 0 0 0 0 0 ...
                                                          0 0 0 0 0 0 0 0 0 0 ...
##
   $ Contribution.tractor.policies
                                                    : int
## $ Contribution.agricultural.machines.policies
                                                          0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.moped.policies
                                                    : int
                                                          0030000000...
   $ Contribution.life.insurances
                                                          0000000000...
                                                    : int
## $ Contribution.private.accident.insurance.policies: int
                                                          0 0 0 0 0 0 0 0 0 0 ...
  $ Contribution.family.accidents.insurance.policies: int
                                                          0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.disability.insurance.policies : int
##
                                                          0 0 0 0 0 0 0 0 0 0 ...
   $ Contribution.fire.policies
                                                    : int
                                                          3 2 4 0 0 4 0 2 0 0 ...
## $ Contribution.surfboard.policies
                                                          0 0 0 0 0 0 0 0 0 0 ...
                                                  : int
## $ Contribution.boat.policies
                                                   : int
                                                          0 0 0 0 0 0 0 0 0 0 ...
                                                          0 0 0 0 0 0 0 0 0 0 ...
## $ Contribution.bicycle.policies
                                                   : int
## $ Contribution.property.insurance.policies
                                                          0 0 0 0 0 0 0 0 0 0 ...
                                                  : int
## $ Contribution.social.security.insurance.policies : int 0 0 0 0 0 3 0 0 0 0 ...
## $ No..of.private.3rd.party.insurance
                                                  : int 0 1 1 0 0 1 0 1 0 1 ...
                                                   : int 0000000000...
## $ No..of.3rd.party.insurance..firms.
```

```
$ No..of.3rd.party.insurane..agriculture.
                                                  : int 0000000000...
## $ No..of.car.policies
                                                  : int 0 1 1 0 0 1 1 0 1 0 ...
## $ No..of.delivery.van.policies
                                                  : int
                                                        0000000000...
## $ No..of.motorcycle.scooter.policies
                                                  : int 0000000000...
## $ No..of.lorry.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.trailer.policies
                                                  : int
                                                        0 0 0 0 1 0 0 0 0 0 ...
## $ No..of.tractor.policies
                                                        0000000000...
                                                  : int
                                                  : int
## $ No..of.agricultural.machines.policies
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.moped.policies
                                                  : int
                                                        0 0 1 0 0 0 0 0 0 0 ...
## $ No..of.life.insurances
                                                        0 0 0 0 0 0 0 0 0 0 ...
                                                  : int
## $ No..of.private.accident.insurance.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.family.accidents.insurance.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
   $ No..of.disability.insurance.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.fire.policies
                                                  : int 1 1 1 0 0 1 0 1 0 0 ...
## $ No..of.surfboard.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.boat.policies
                                                  : int
                                                        0 0 0 0 0 0 0 0 0 0 ...
## $ No..of.bicycle.policies
                                                  : int 0000000000...
## $ No..of.property.insurance.policies
                                                 : int 0000000000...
                                                 : int 0000010000...
## $ No..of.social.security.insurance.policies
                                                 : int 000100000...
## $ No..of.mobile.home.policies
## - attr(*, "na.action")= 'omit' Named int 13336 13337 13338 13339 13340 13341 13342 13343 13344 133
    ..- attr(*, "names")= chr "13336" "13337" "13338" "13339" ...
```

Model Building