Aarushi_project

Reading the Data

```
datapath<-"/Users/archana/Desktop"
Businesses<-read.csv(file=paste(datapath,"final.data_v4.csv",sep="/"),header =
TRUE,sep=",")
nrow(Businesses)
```

```
## [1] 28969
```

Removing businesses from zipcodes not available in the demographic data

```
Businesses<-Businesses[!(Businesses$zipcode %in% c(60635,60666,60707,60827)),]
nrow(Businesses)
```

```
## [1] 28710
```

Putting variables in the right format

```
Businesses$Business <- as.character(Businesses$Business)
Businesses$Address <- as.character(Businesses$Address)
Businesses$Under.5.Years <- as.integer(Businesses$Under.5.Years)
```

How many businesses are active post 1st Jan 2010 (Demographic data only available as a 5 year estimate from 2010 to 2014)

```
Businesses$Start.Date <- as.Date(Businesses$Start.Date,"%m/%d/%y")
Businesses$End.Date <- as.Date(Businesses$End.Date,"%m/%d/%y")
total.2010<-Businesses[Businesses$End.Date >= "2010-01-01",] # no of businesses
with end date>= jan 2010
nrow(total.2010)
```

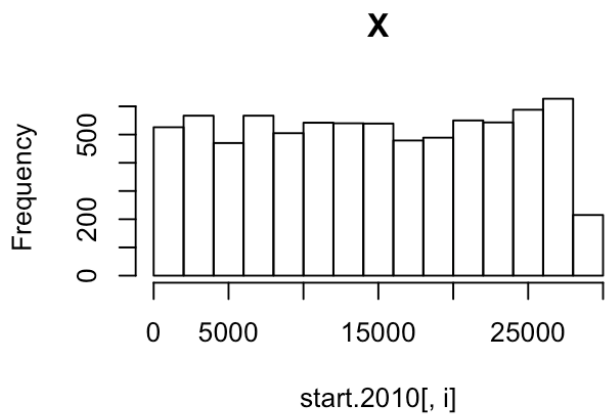
```
## [1] 18694
```

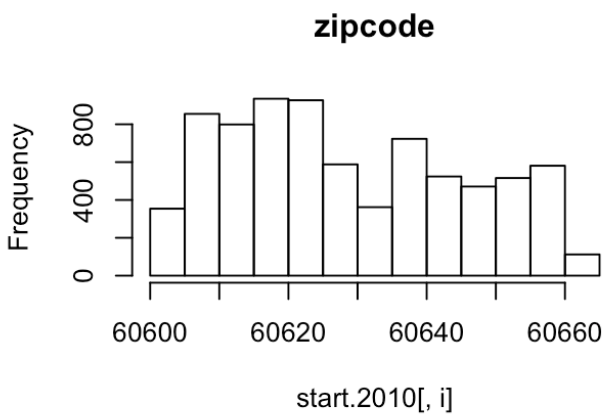
```
#active.2010 <- Businesses[Businesses$Active == "Y" &
                           #Businesses$End.Date >= "2010-01-01",]
#How many businesses started after jan'2010
start.2010<-Businesses[Businesses$Start.Date >= "2010-01-01",]
nrow(start.2010)
```

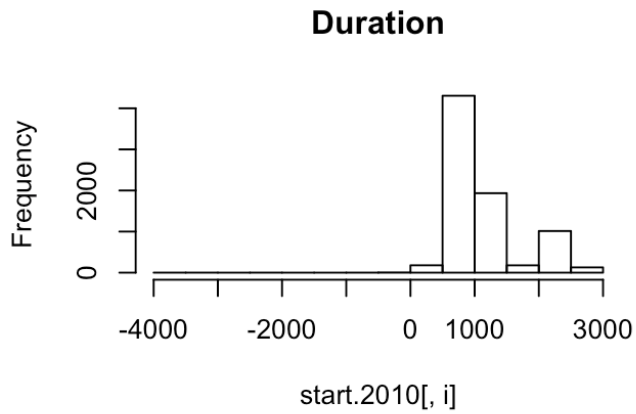
```
## [1] 7747
```

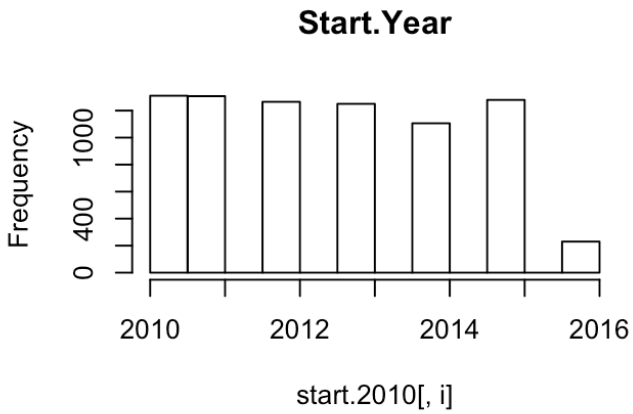
Histograms for continuous variables

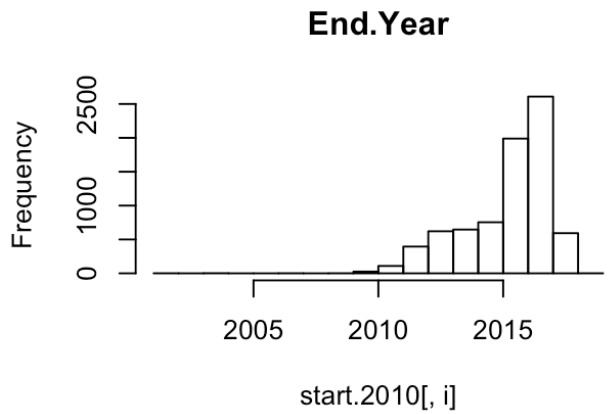
```
par(mfrow=c(2,2))
for (i in 1:ncol(start.2010)) {
  if (is.integer(start.2010[,i])){
    hist(start.2010[,i],main = colnames(start.2010)[i])
  }
}
```

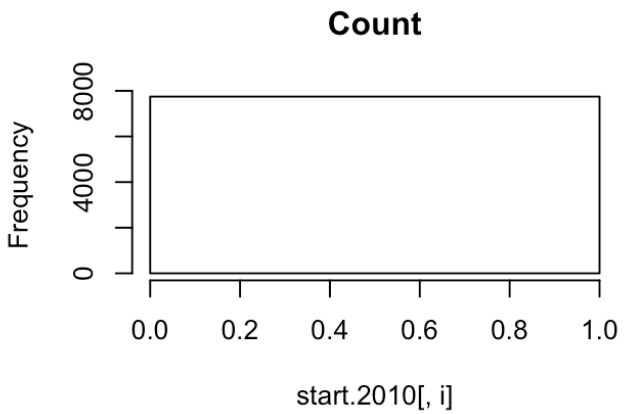

X
Frequency
0
200
500
0
5000
15000
25000
start.2010[, i]

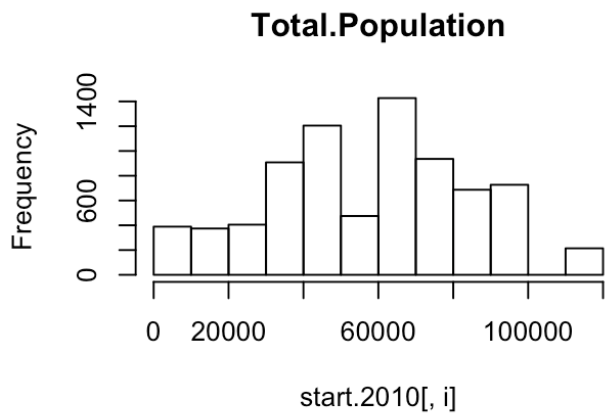

zipcode
Frequency
0
400
800
60600
60620
60640
60660
start.2010[, i]

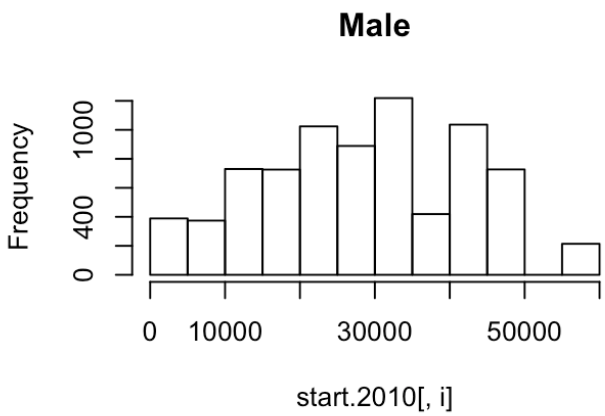

Duration
Frequency
0
2000
-4000
-2000
0
1000
3000
start.2010[, i]


Start.Year
Frequency
0
400
1000
2010
2012
2014
2016
start.2010[, i]

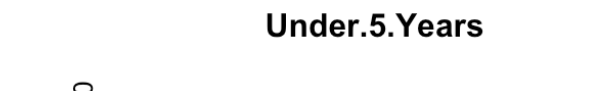

End.Year
Frequency
0
1000
2500
2005
2010
2015
start.2010[, i]

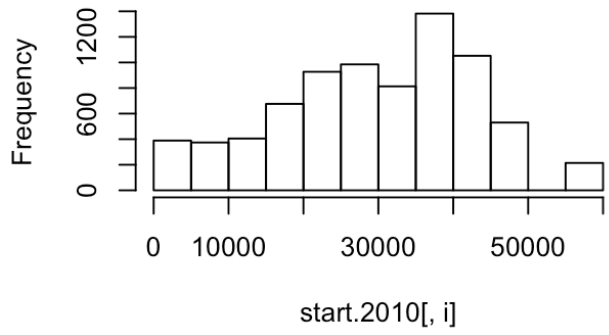

Count
Frequency
0
4000
8000
0.0
0.2
0.4
0.6
0.8
1.0
start.2010[, i]

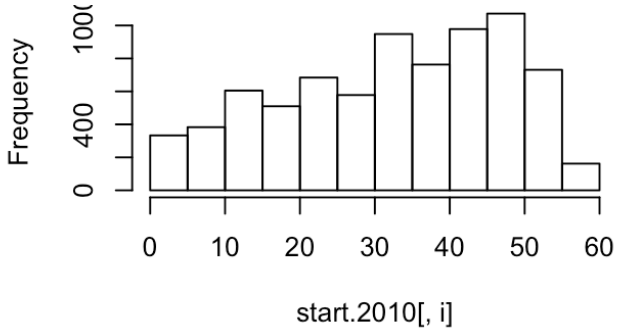

Total.Population
Frequency
0
600
1400
0
20000
60000
100000
start.2010[, i]

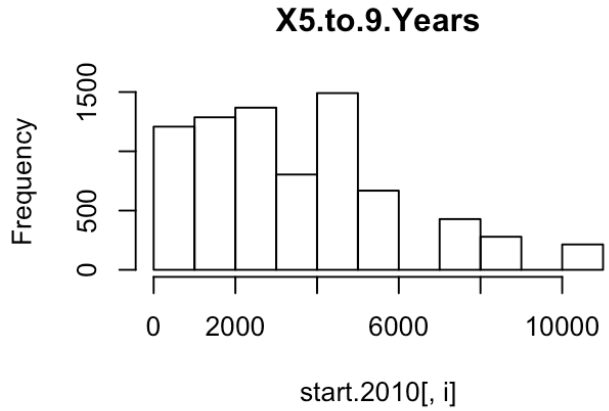

Male
Frequency
0
400
1000
0
10000
30000
50000
start.2010[, i]

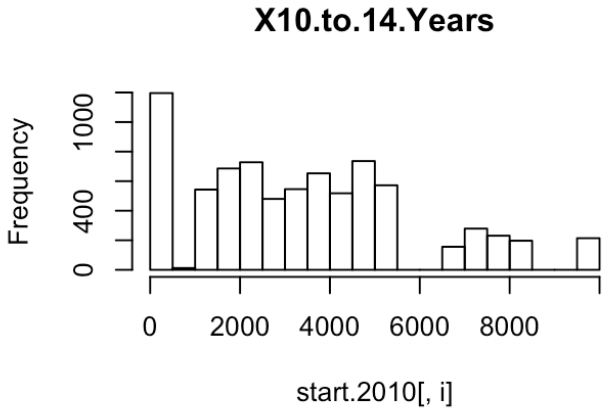

Female

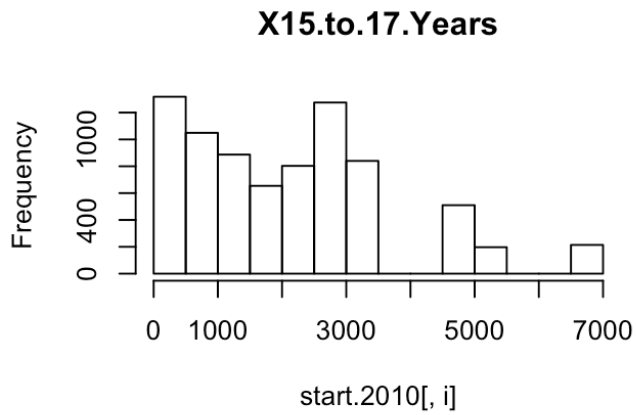

Under.5.Years

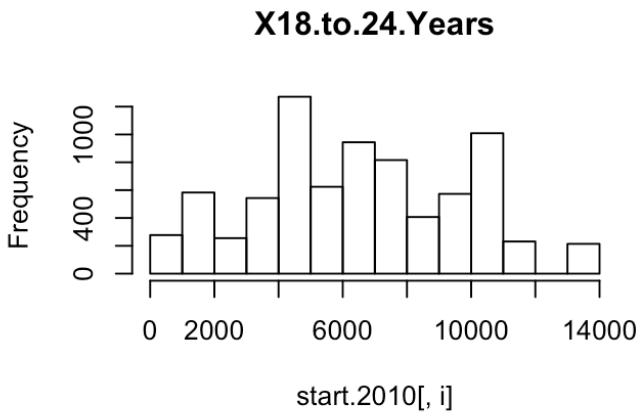

Frequency
0
600
1200
0
10000
30000
50000
start.2010[, i]

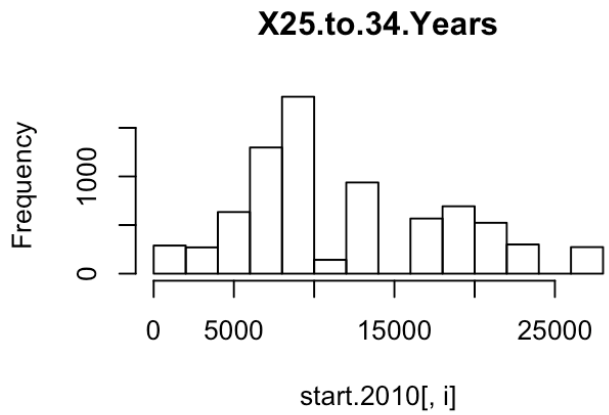

Frequency
0
400
0
10
20
30
40
50
60
start.2010[, i]

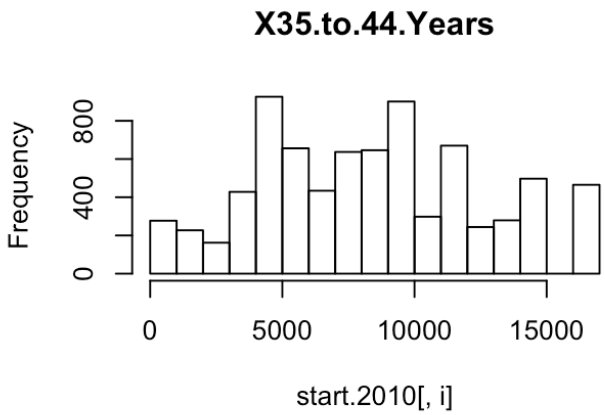

X5.to.9.Years
Frequency
0
500
1500
0
2000
6000
10000
start.2010[, i]

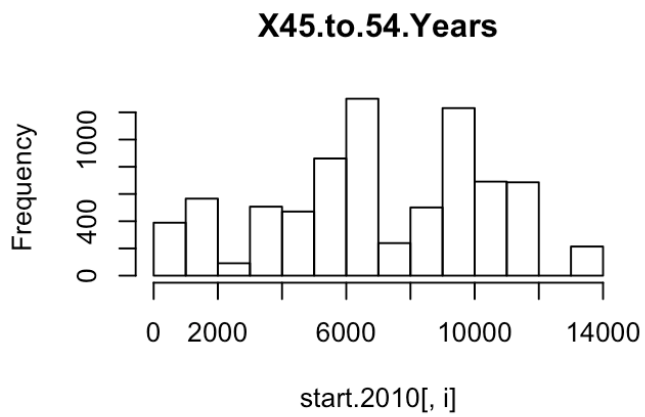

X10.to.14.Years
Frequency
0
400
1000
0
2000
4000
6000
8000
start.2010[, i]

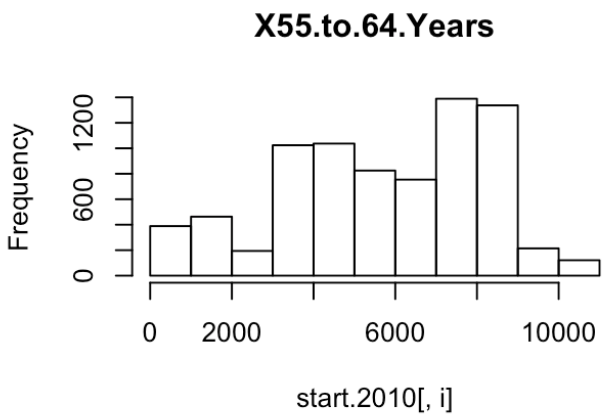

X15.to.17.Years
Frequency
0
400
1000
0
1000
3000
5000
7000
start.2010[, i]

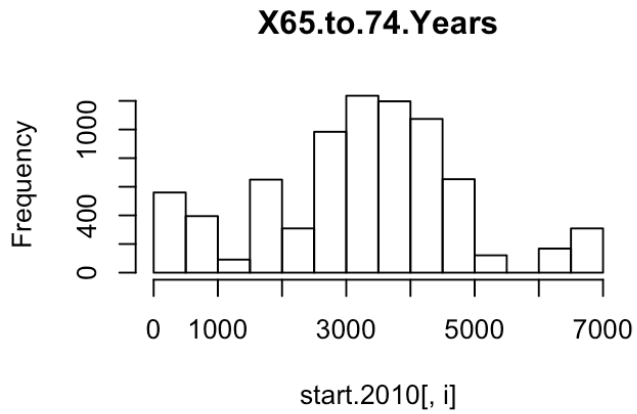

X18.to.24.Years
Frequency
0
400
1000
0
2000
6000
10000
14000
start.2010[, i]

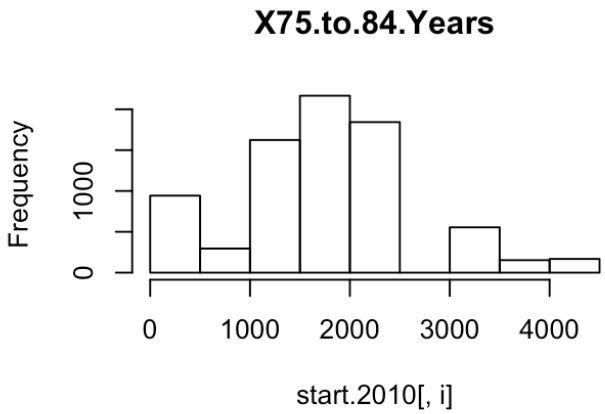

X25.to.34.Years
Frequency
0
1000
0
5000
15000
25000
start.2010[, i]

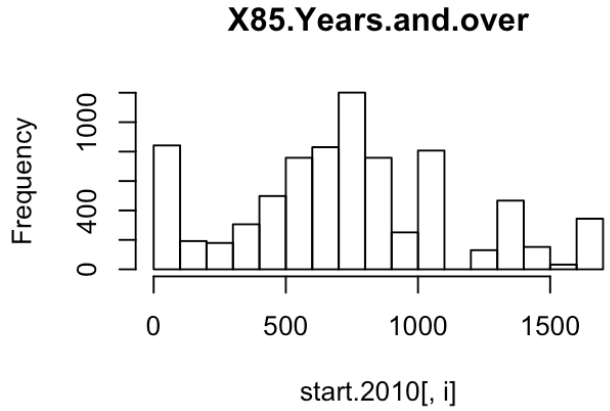

X35.to.44.Years
Frequency
0
400
800
0
5000
10000
15000
start.2010[, i]

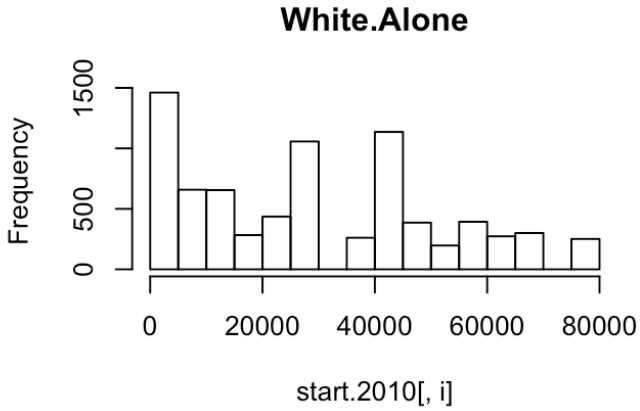

X45.to.54.Years
Frequency
0
400
1000
0
2000
6000
10000
14000
start.2010[, i]

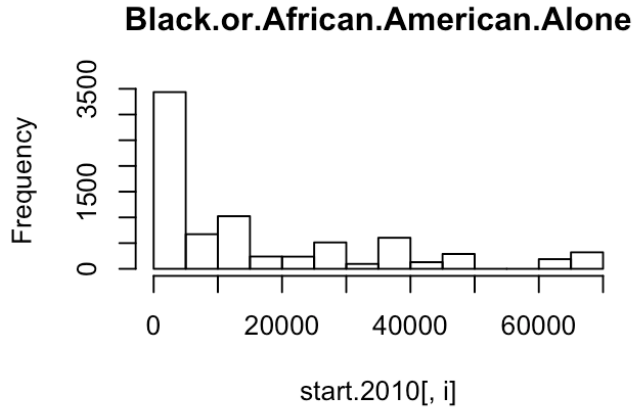

X55.to.64.Years
Frequency
0
600
1200
0
2000
6000
10000
start.2010[, i]

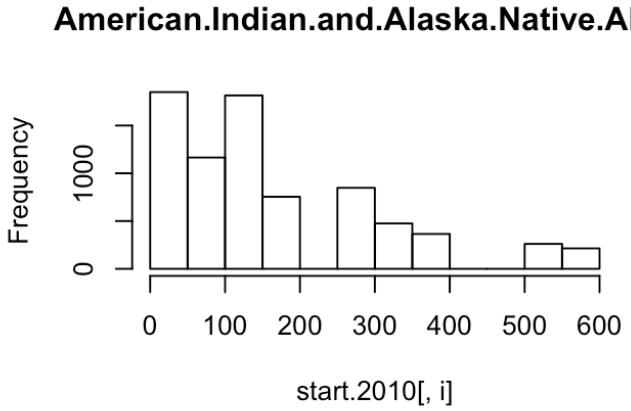

X65.to.74.Years
Frequency
0
400
1000
0
1000
3000
5000
7000
start.2010[, i]

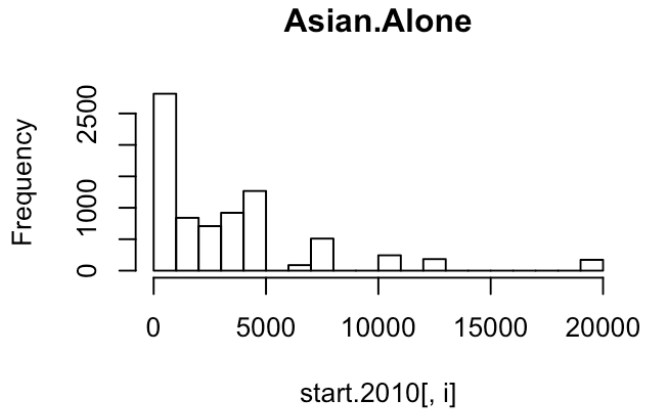

X75.to.84.Years
Frequency
0
1000
0
1000
2000
3000
4000
start.2010[, i]

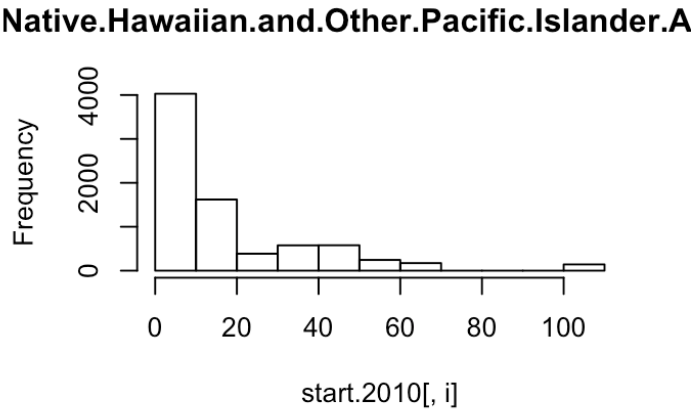

X85.Years.and.over
Frequency
0
400
1000
0
500
1000
1500
start.2010[, i]

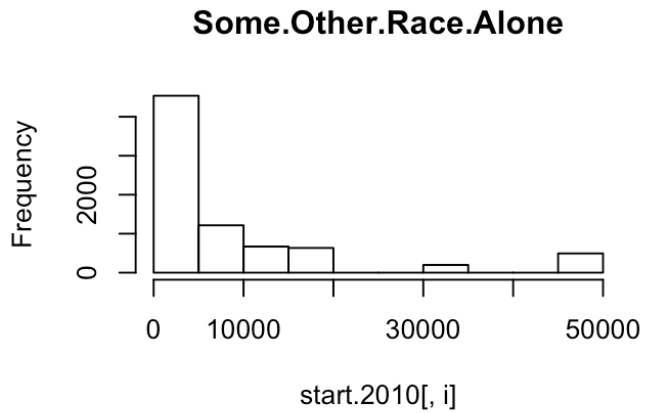

White.Alone
Frequency
0
500
1500
0
20000
40000
60000
80000
start.2010[, i]

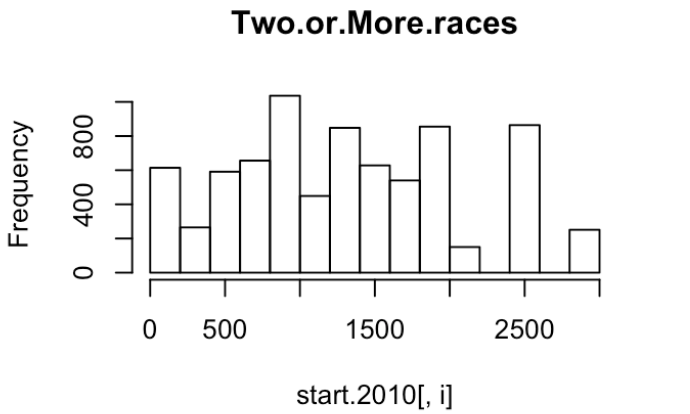

Black.or.African.American.Alone
Frequency
0
1500
3500
0
20000
40000
60000
start.2010[, i]

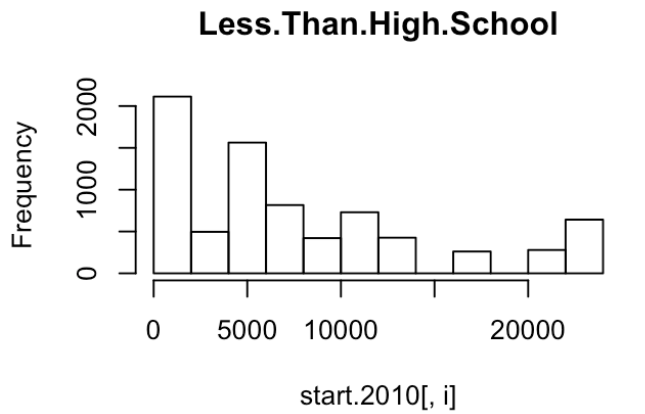

American.Indian.and.Alaska.Native.Alone
Frequency
0
1000
0
100
200
300
400
500
600
start.2010[, i]

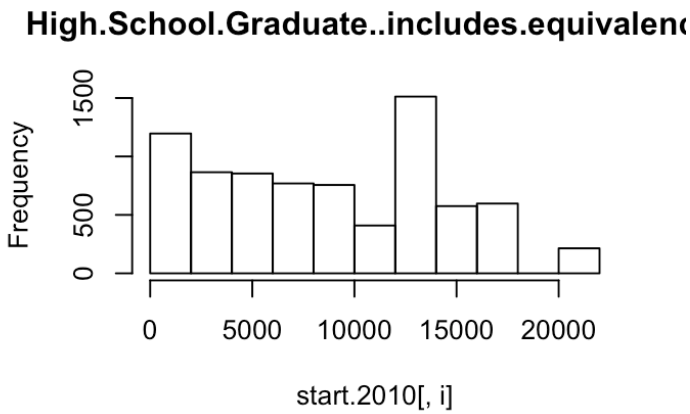

Asian.Alone
Frequency
0
1000
2500
0
5000
10000
15000
20000
start.2010[, i]

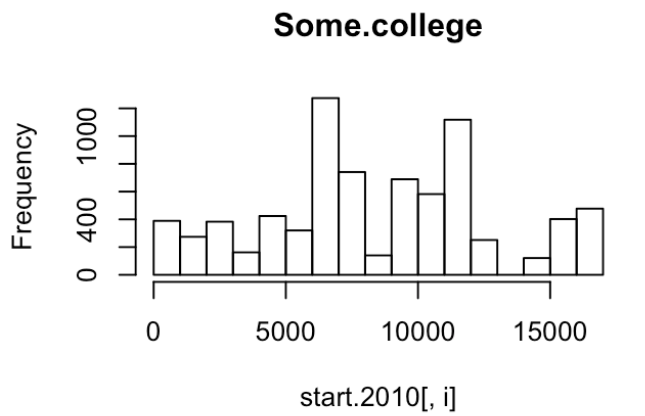

Native.Hawaiian.and.Other.Pacific.Islander.A
Frequency
0
2000
4000
0
20
40
60
80
100
start.2010[, i]


Some.Other.Race.Alone
Frequency
0
2000
0
10000
30000
50000
start.2010[, i]

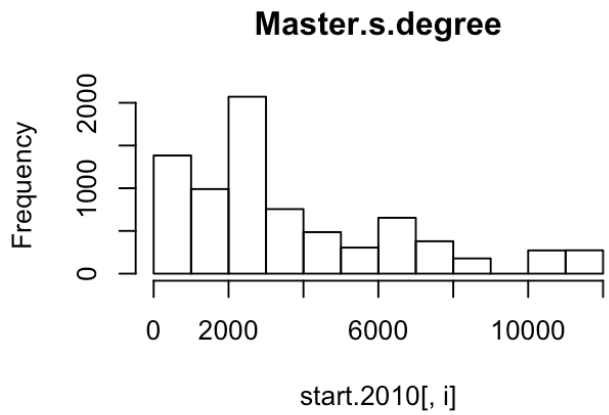

Two.or.More.races
Frequency
0
400
800
0
500
1500
2500
start.2010[, i]

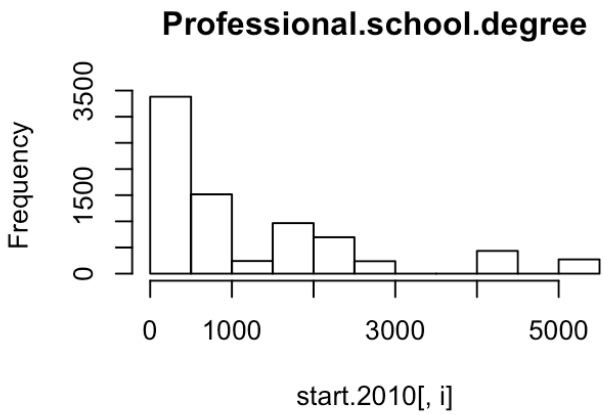

Less.Than.High.School
Frequency
0
1000
2000
0
5000
10000
20000
start.2010[, i]

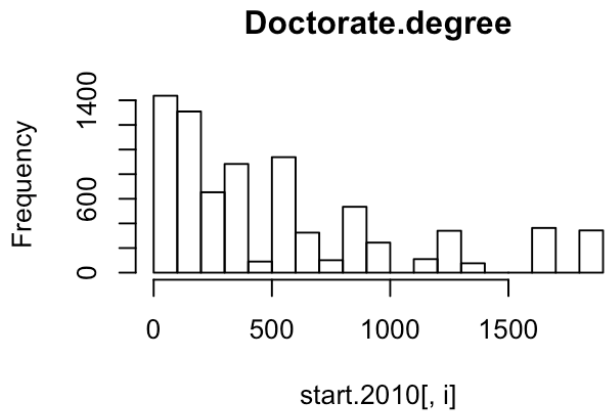

High.School.Graduate..includes.equivalenc
Frequency
0
500
1500
0
5000
10000
15000
20000
start.2010[, i]

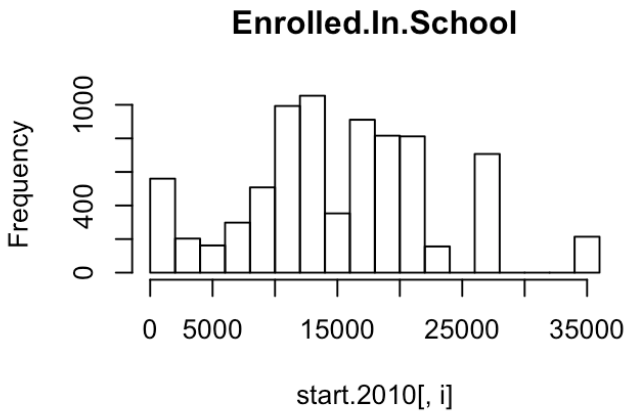

Some.college
Frequency
0
400
1000
0
5000
10000
15000
start.2010[, i]

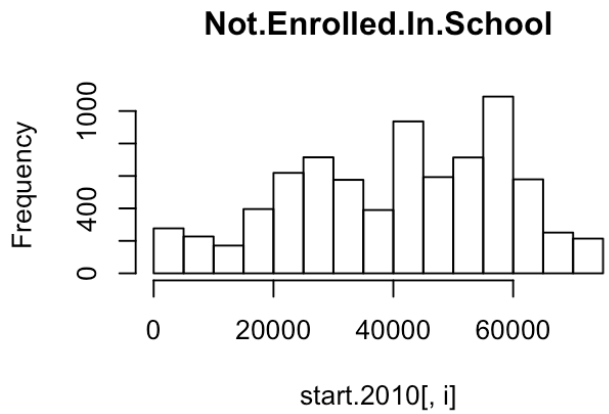

Bachelor.s.degree
Frequency
0
1000
0
5000
15000
25000
start.2010[, i]


Master.s.degree
Frequency
0
1000
2000
0
2000
6000
10000
start.2010[, i]

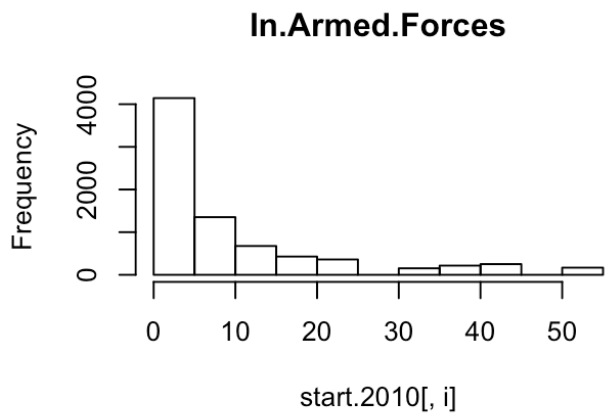

Professional.school.degree
Frequency
0
1500
3500
0
1000
3000
5000
start.2010[, i]

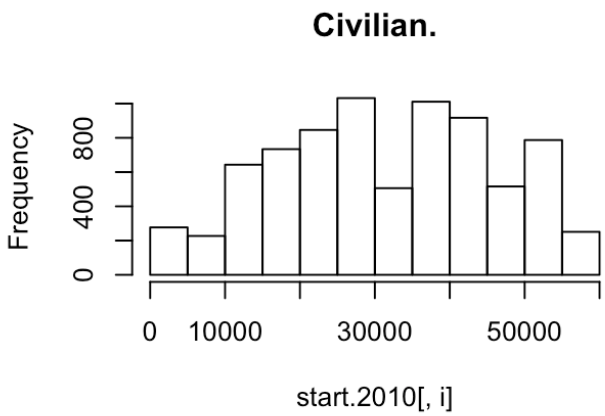

Doctorate.degree
Frequency
0
600
1400
0
500
1000
1500
start.2010[, i]

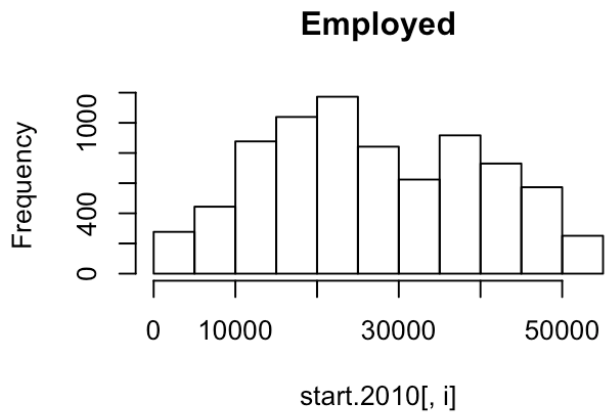

Enrolled.In.School
Frequency
0
400
1000
0
5000
15000
25000
35000
start.2010[, i]

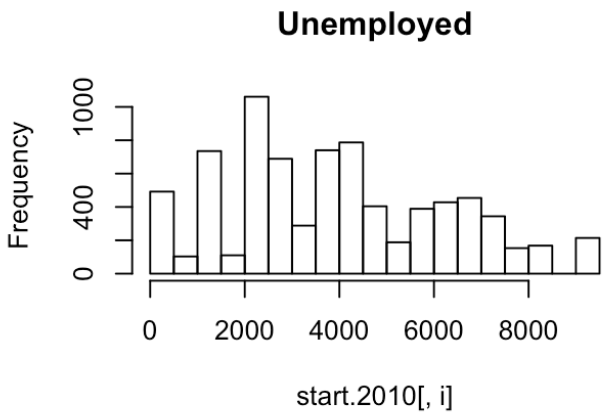

Not.Enrolled.In.School
Frequency
0
400
1000
0
20000
40000
60000
start.2010[, i]


In.labor.force.
Frequency
0
400
800
0
10000
30000
50000
start.2010[, i]

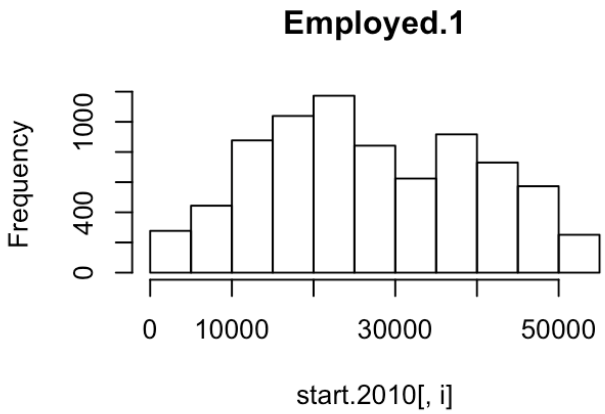

In.Armed.Forces
Frequency
0
2000
4000
0
10
20
30
40
50
start.2010[, i]

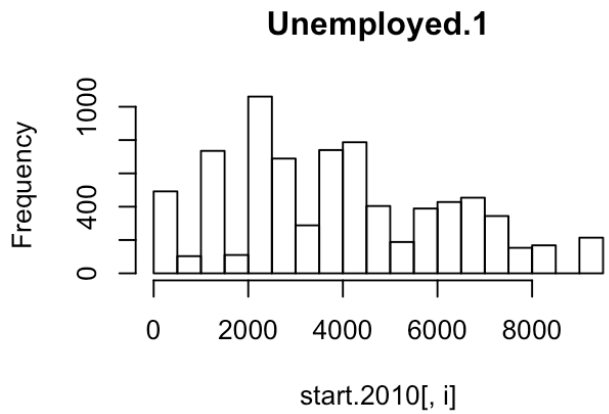

Civilian.
Frequency
0
400
800
0
10000
30000
50000
start.2010[, i]

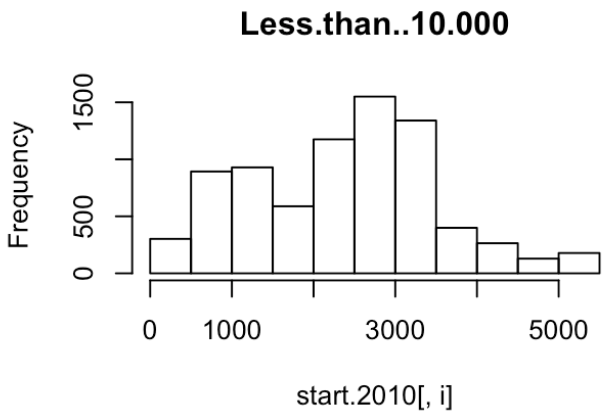

Employed
Frequency
0
400
1000
0
10000
30000
50000
start.2010[, i]

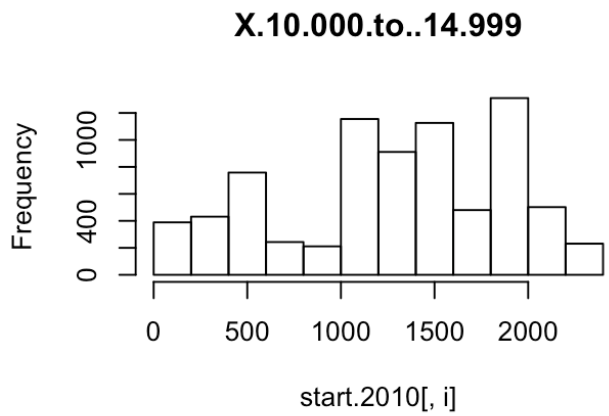

Unemployed
Frequency
0
400
1000
0
2000
4000
6000
8000
start.2010[, i]

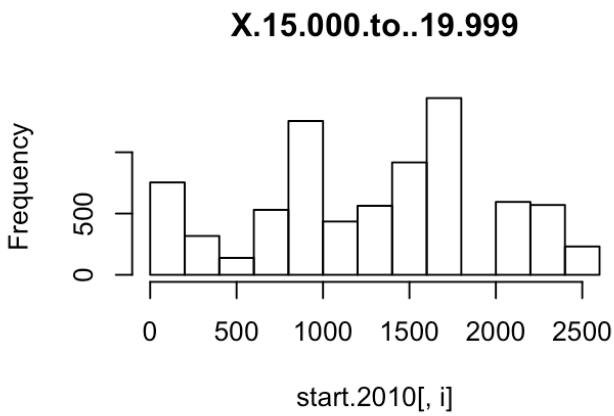

Not.In.labor.force
Frequency
0
400
800
0
5000
15000
25000
start.2010[, i]

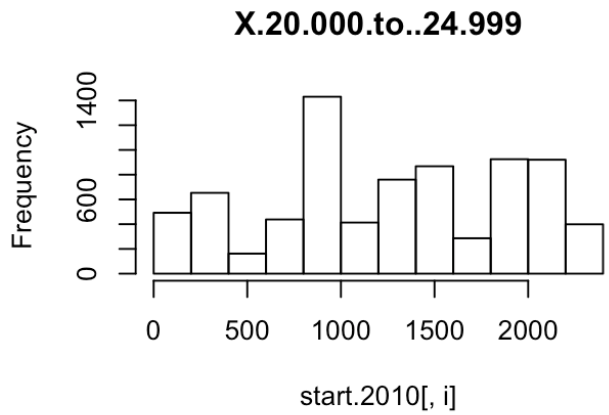

Employed.1
Frequency
0
400
1000
0
10000
30000
50000
start.2010[, i]

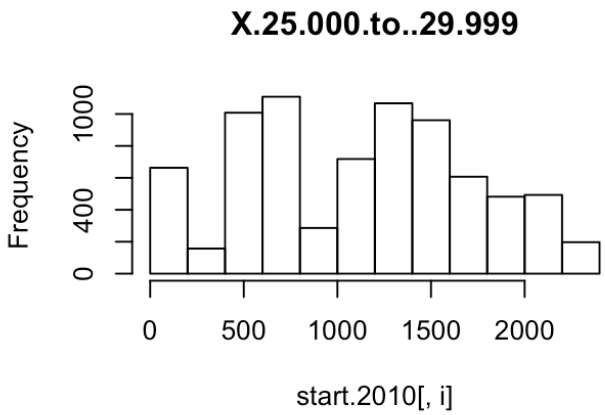

Unemployed.1
Frequency
0
400
1000
0
2000
4000
6000
8000
start.2010[, i]

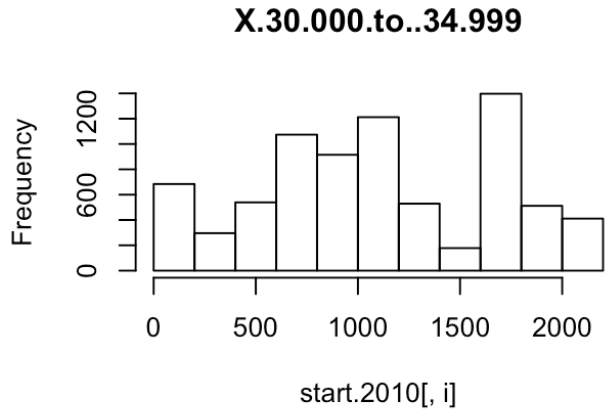

Less.than..10.000
Frequency
0
500
1500
0
1000
3000
5000
start.2010[, i]

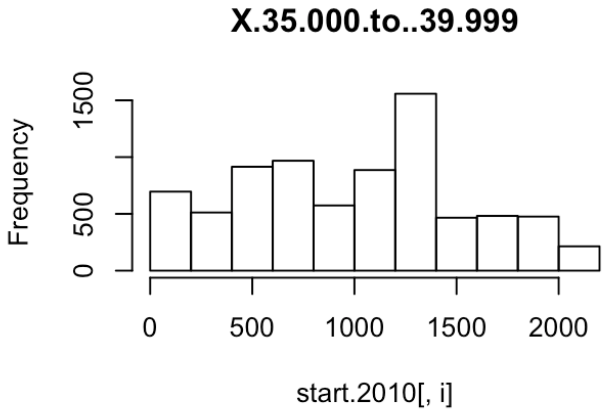

X.10.000.to..14.999
Frequency
0
400
1000
0
500
1000
1500
2000
start.2010[, i]

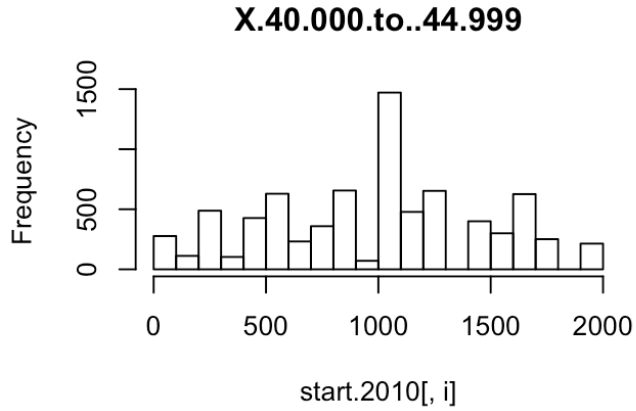

X.15.000.to..19.999
Frequency
0
500
0
500
1000
1500
2000
2500
start.2010[, i]

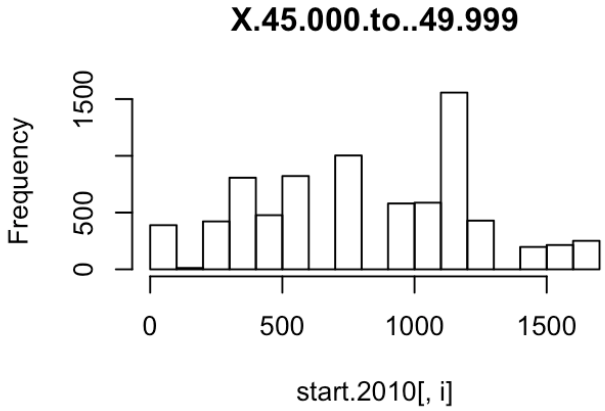

X.20.000.to..24.999
Frequency
0
600
1400
0
500
1000
1500
2000
start.2010[, i]

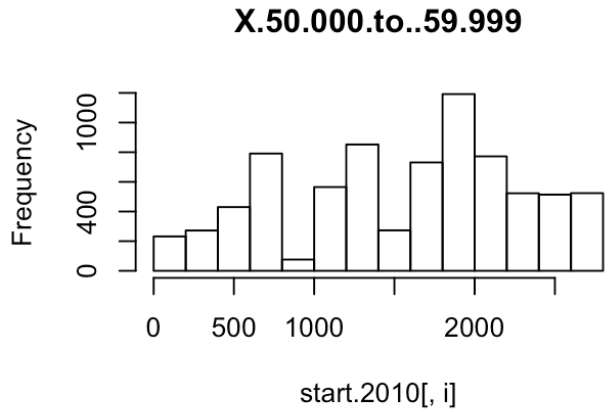

X.25.000.to..29.999
Frequency
0
400
1000
0
500
1000
1500
2000
start.2010[, i]

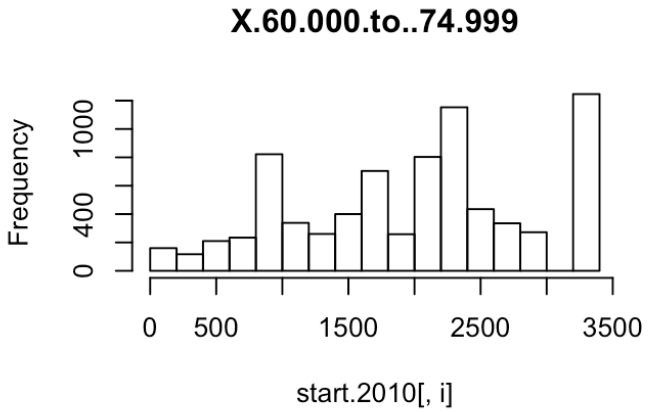

X.30.000.to..34.999
Frequency
0
600
1200
0
500
1000
1500
2000
start.2010[, i]

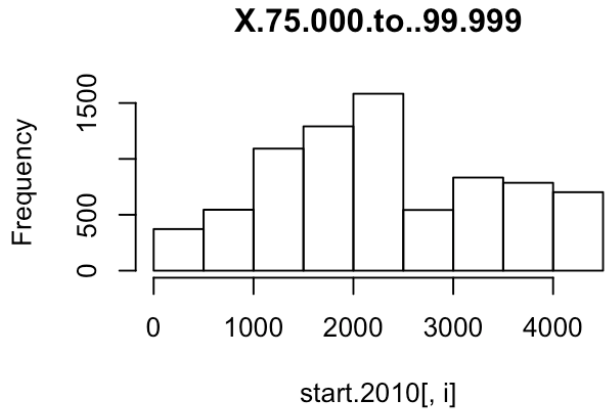

X.35.000.to..39.999
Frequency
0
500
1500
0
500
1000
1500
2000
start.2010[, i]

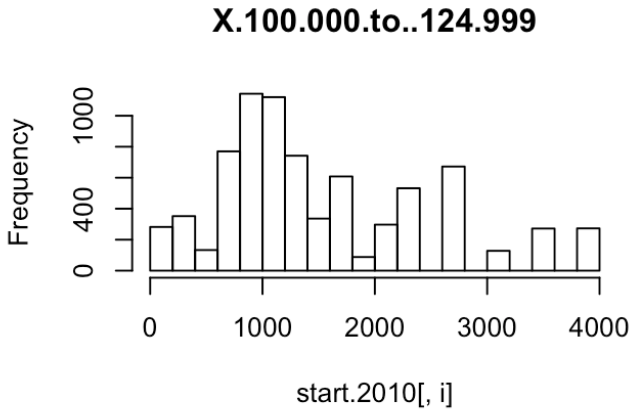

X.40.000.to..44.999
Frequency
0
500
1500
0
500
1000
1500
2000
start.2010[, i]

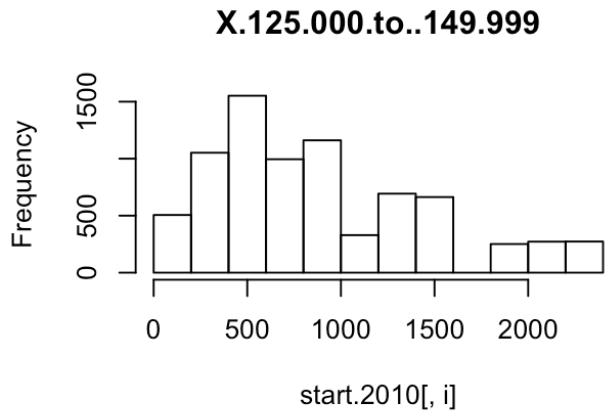

X.45.000.to..49.999
Frequency
0
500
1500
0
500
1000
1500
start.2010[, i]

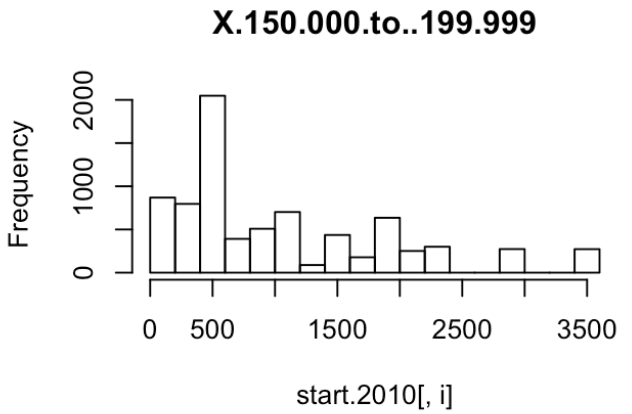

X.50.000.to..59.999
Frequency
0
400
1000
0
500
1000
2000
start.2010[, i]

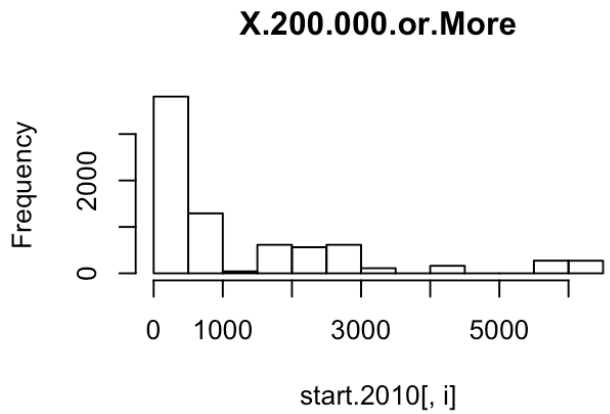

X.60.000.to..74.999
Frequency
0
400
1000
0
500
1500
2500
3500
start.2010[, i]

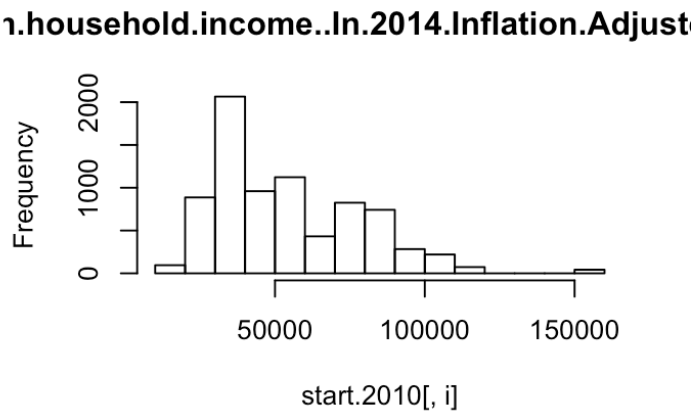

X.75.000.to..99.999
Frequency
0
500
1500
0
1000
2000
3000
4000
start.2010[, i]

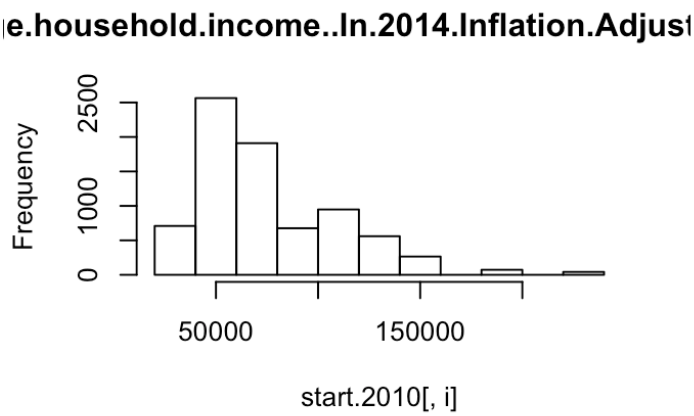

X.100.000.to..124.999
Frequency
0
400
1000
0
1000
2000
3000
4000
start.2010[, i]

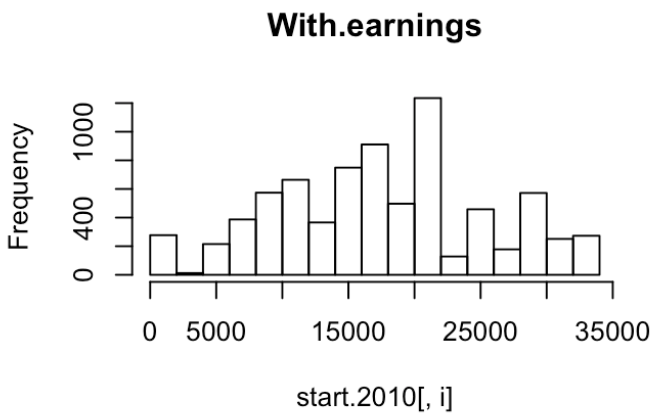

X.125.000.to..149.999
Frequency
0
500
1500
0
500
1000
1500
2000
start.2010[, i]

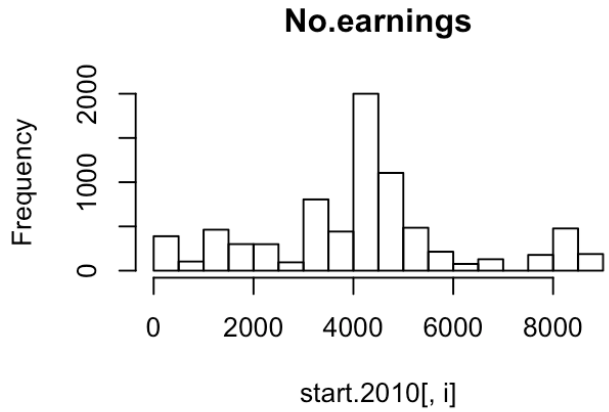

X.150.000.to..199.999
Frequency
0
1000
2000
0
500
1500
2500
3500
start.2010[, i]

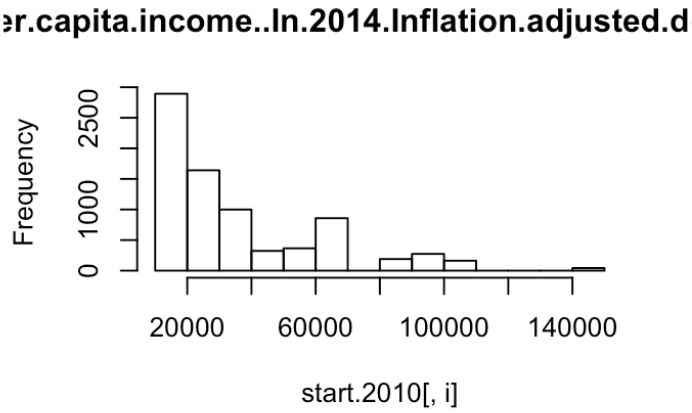

X.200.000.or.More
Frequency
0
2000
0
1000
3000
5000
start.2010[, i]

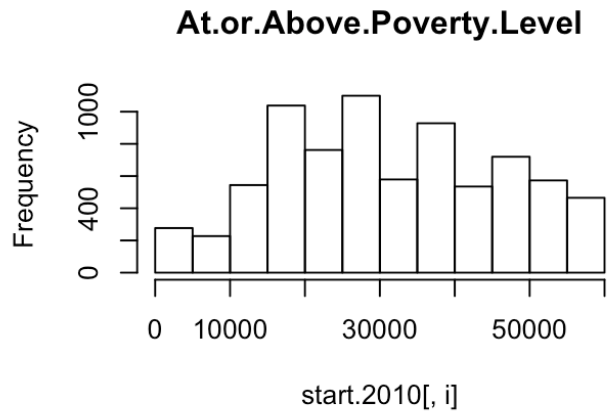

ı.household.income..In.2014.Inflation.Adjust
Frequency
0
1000
2000
50000
100000
150000
start.2010[, i]


e.household.income..In.2014.Inflation.Adjust
Frequency
0
1000
2500
50000
150000
start.2010[, i]


With.earnings
Frequency
0
400
1000
0
5000
15000
25000
35000
start.2010[, i]

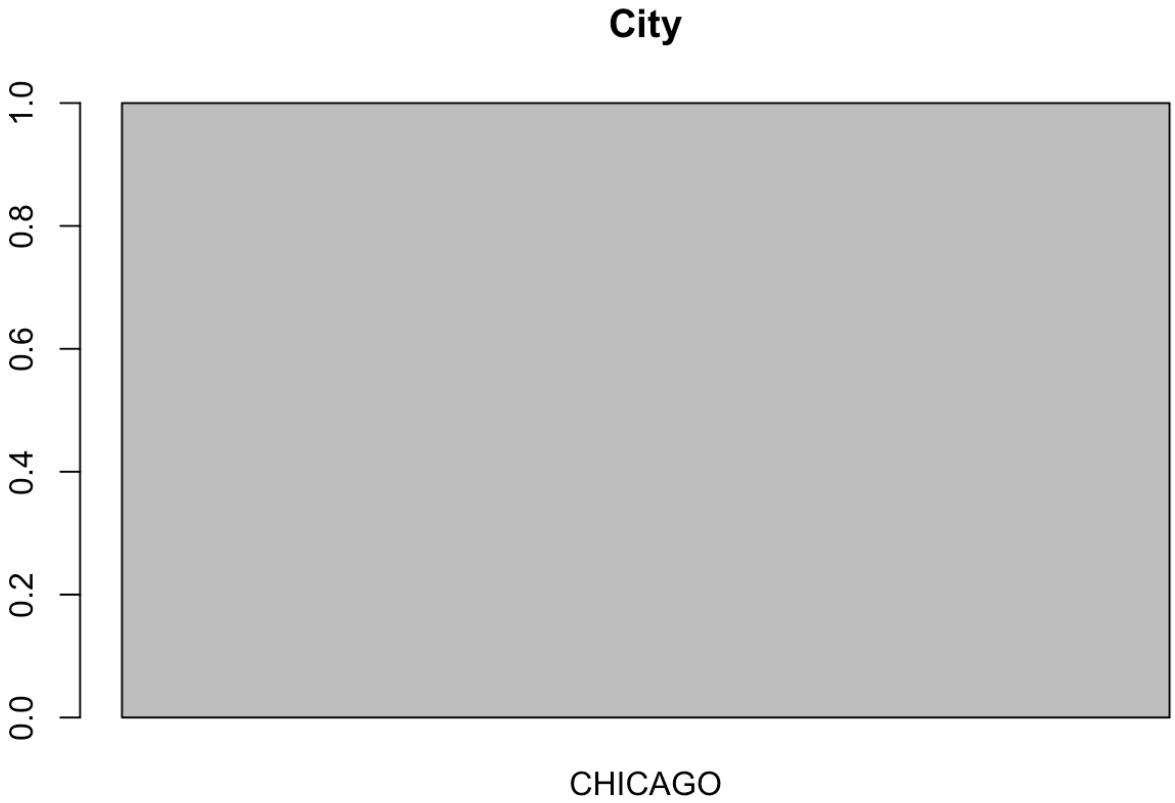


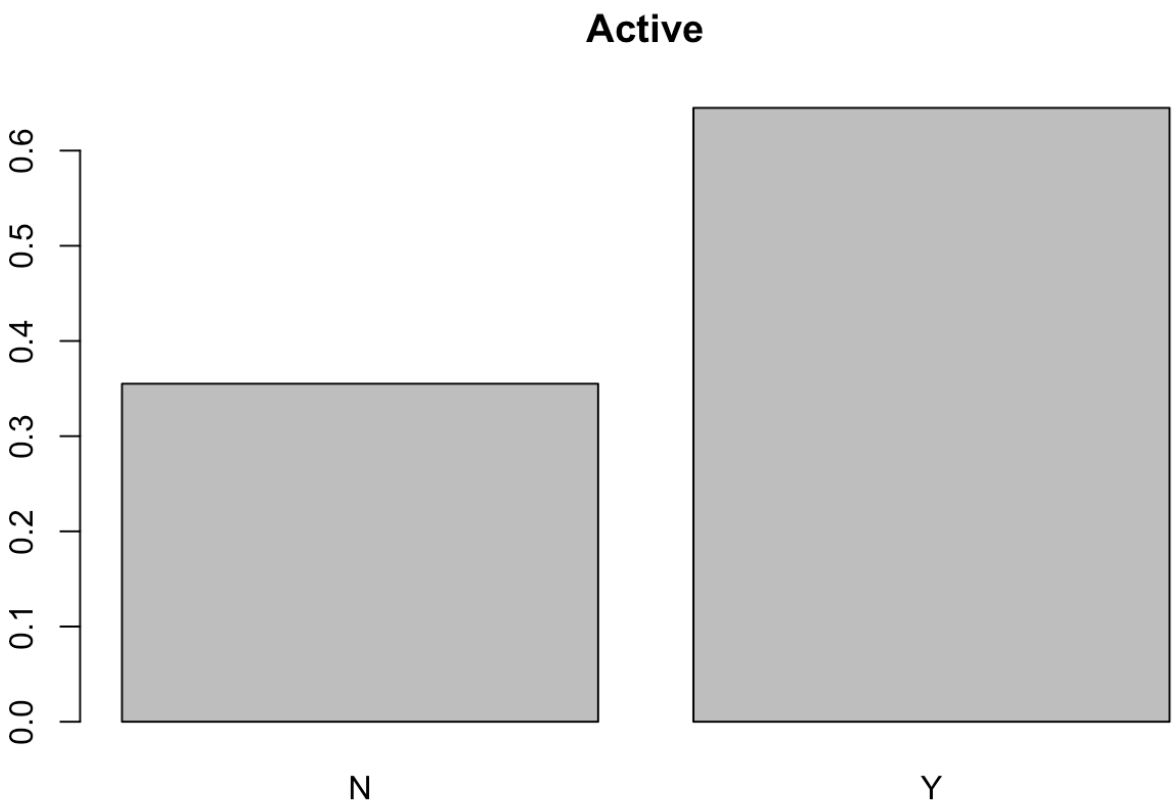




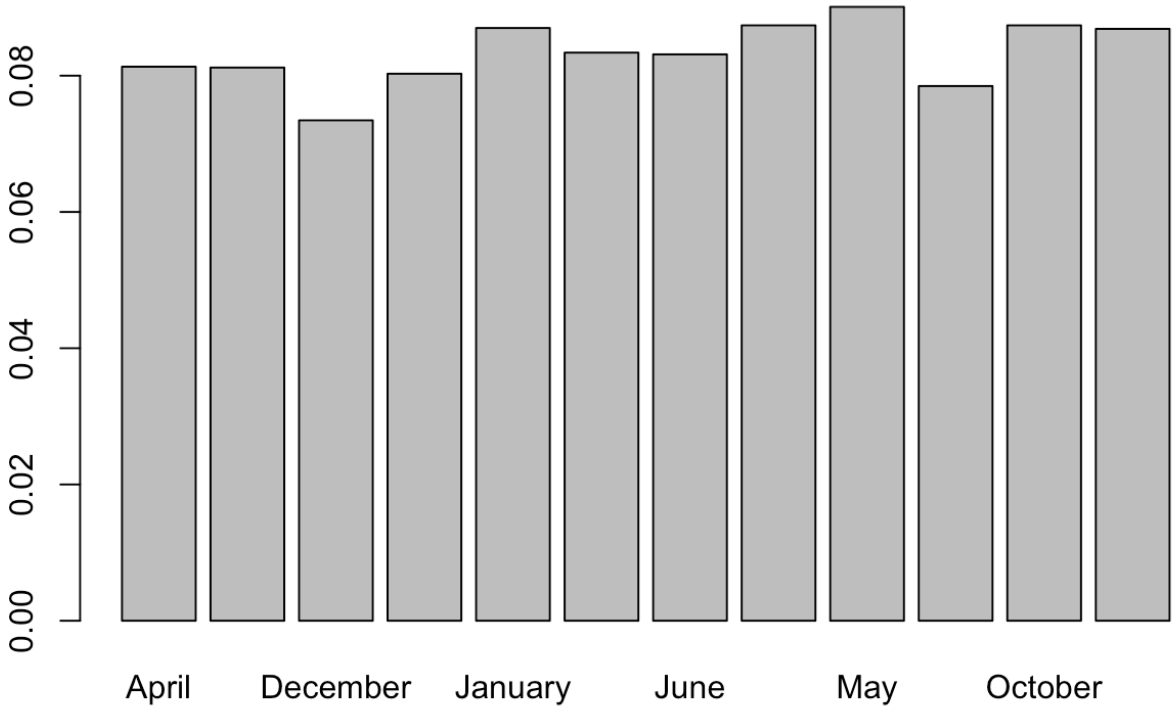
Frequency plots for categorical variables

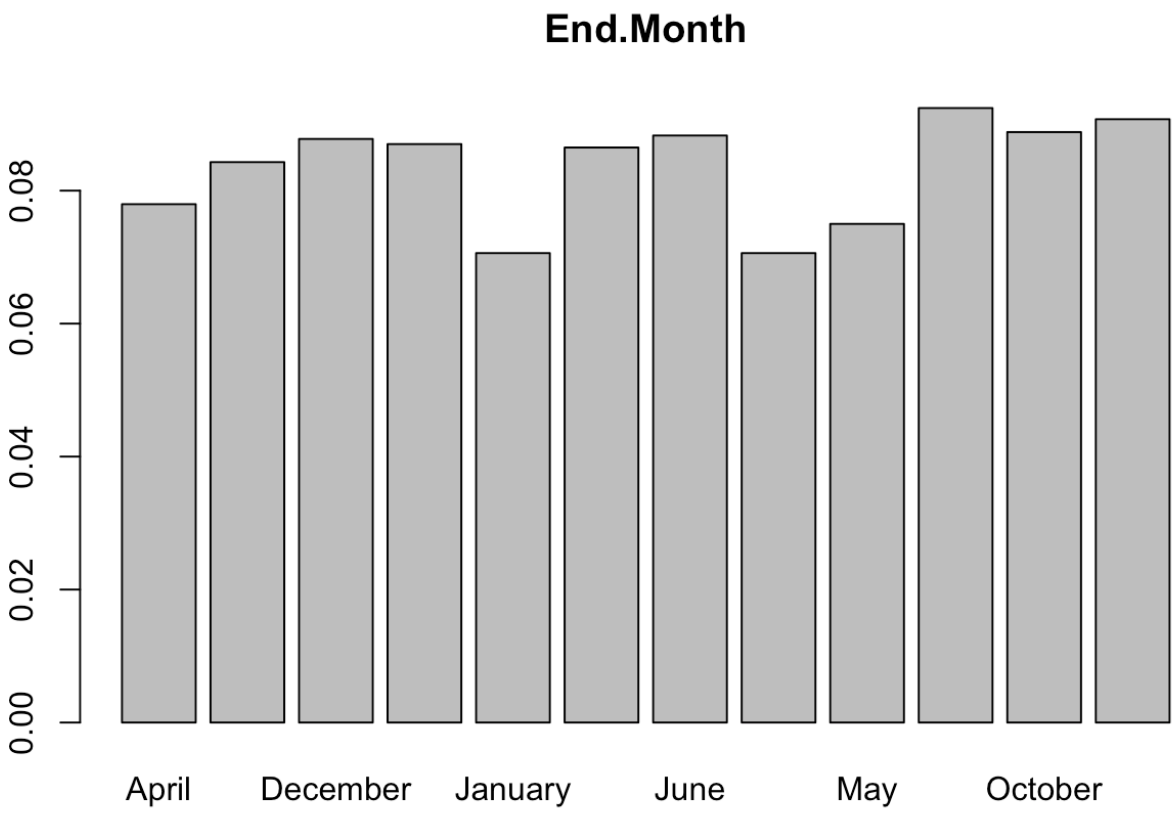
```
for (i in 1:ncol(start.2010)){
  if (is.factor(start.2010[,i])){
    barplot(prop.table(table(start.2010[,i])),main=colnames(start.2010)[i])
  }
}
```





Start.Month





Correlation Matrix

```
#temp<- start.2010[-c(1,3:7,9:10,12,14)]
#print(correlation.matrix<-cor(temp))
#correlation.matrix
```

```
#Converting Zipcode to Factor variable
start.2010$zipcode <- as.factor(start.2010$zipcode)

names(start.2010)
```

```
##  [1] "X"
##  [2] "zipcode"
##  [3] "Business"
##  [4] "Address"
##  [5] "City"
##  [6] "Start.Date"
##  [7] "End.Date"
##  [8] "Duration"
##  [9] "Active"
## [10] "Start.Month"
## [11] "Start.Year"
## [12] "End.Month"
## [13] "End.Year"
## [14] "Count"
## [15] "Total.Population"
## [16] "Male"
## [17] "Female"
## [18] "Under.5.Years"
## [19] "X5.to.9.Years"
## [20] "X10.to.14.Years"
## [21] "X15.to.17.Years"
## [22] "X18.to.24.Years"
## [23] "X25.to.34.Years"
## [24] "X35.to.44.Years"
## [25] "X45.to.54.Years"
## [26] "X55.to.64.Years"
## [27] "X65.to.74.Years"
## [28] "X75.to.84.Years"
## [29] "X85.Years.and.over"
## [30] "White.Alone"
## [31] "Black.or.African.American.Alone"
## [32] "American.Indian.and.Alaska.Native.Alone"
## [33] "Asian.Alone"
## [34] "Native.Hawaiian.and.Other.Pacific.Islander.Alone"
## [35] "Some.Other.Race.Alone"
## [36] "Two.or.More.races"
## [37] "Average.Household.Size"
## [38] "Less.Than.High.School"
## [39] "High.School.Graduate..includes.equivalency."
## [40] "Some.college"
## [41] "Bachelor.s.degree"
## [42] "Master.s.degree"
## [43] "Professional.school.degree"
## [44] "Doctorate.degree"
## [45] "Enrolled.In.School"
## [46] "Not.Enrolled.In.School"
## [47] "In.labor.force."
## [48] "In.Armed.Forces"
## [49] "Civilian."
## [50] "Employed"
## [51] "Unemployed"
## [52] "Not.In.labor.force"
## [53] "Employed.1"
```

```
## [54] "Unemployed.1"
## [55] "Less.than..10.000"
## [56] "X.10.000.to..14.999"
## [57] "X.15.000.to..19.999"
## [58] "X.20.000.to..24.999"
## [59] "X.25.000.to..29.999"
## [60] "X.30.000.to..34.999"
## [61] "X.35.000.to..39.999"
## [62] "X.40.000.to..44.999"
## [63] "X.45.000.to..49.999"
## [64] "X.50.000.to..59.999"
## [65] "X.60.000.to..74.999"
## [66] "X.75.000.to..99.999"
## [67] "X.100.000.to..124.999"
## [68] "X.125.000.to..149.999"
## [69] "X.150.000.to..199.999"
## [70] "X.200.000.or.More"
## [71] "Median.household.income..In.2014.Inflation.Adjusted.Dollars."
## [72] "Average.household.income..In.2014.Inflation.Adjusted.Dollars."
## [73] "With.earnings"
## [74] "No.earnings"
## [75] "Per.capita.income..In.2014.Inflation.adjusted.dollars."
## [76] "Living.in.Poverty"
## [77] "At.or.Above.Poverty.Level"
```

```
length(unique(start.2010$zipcode))
```

```
## [1] 56
```

```
#creating 1/0 flag
start.2010$Active_flag=1*(start.2010$Active=="Y")

#train[,-c(1,3,4,5,6,7,10,11,12,13,14)]

#start.2010$mnth_yr <- as.factor(do.call(paste, c(start.2010[c("Start.Month","S
tart.Year")], sep = "_")))

#Dummy coding for Random Forest since it does not work for more than 32 categor
ies in 1 variable
xfactors.model.matrix <- model.matrix(~start.2010[,c(2)],start.2010)[,-1]
X<- data.frame(start.2010[,-c(1,3,4,5,6,7,10,11,12,13,14)],model.matrix(~start.
2010[,c(2)],start.2010)[,-1])

#Converting the dummy coded variables to factor
for (i in 67:122){
  X[,i] <- as.factor(X[,i])
}

# Reference level Zipcode: 60601

smp_size <- floor(0.7 * nrow(X))

## set the seed to make your partition reproductible
set.seed(123)
train <- sample(seq_len(nrow(X)), size = smp_size)
test<- X[-train,]
train <- X[train,]
train <- data.frame(train)
test <- data.frame(test)
```

```
#install.packages("randomForest")
library(randomForest)
```

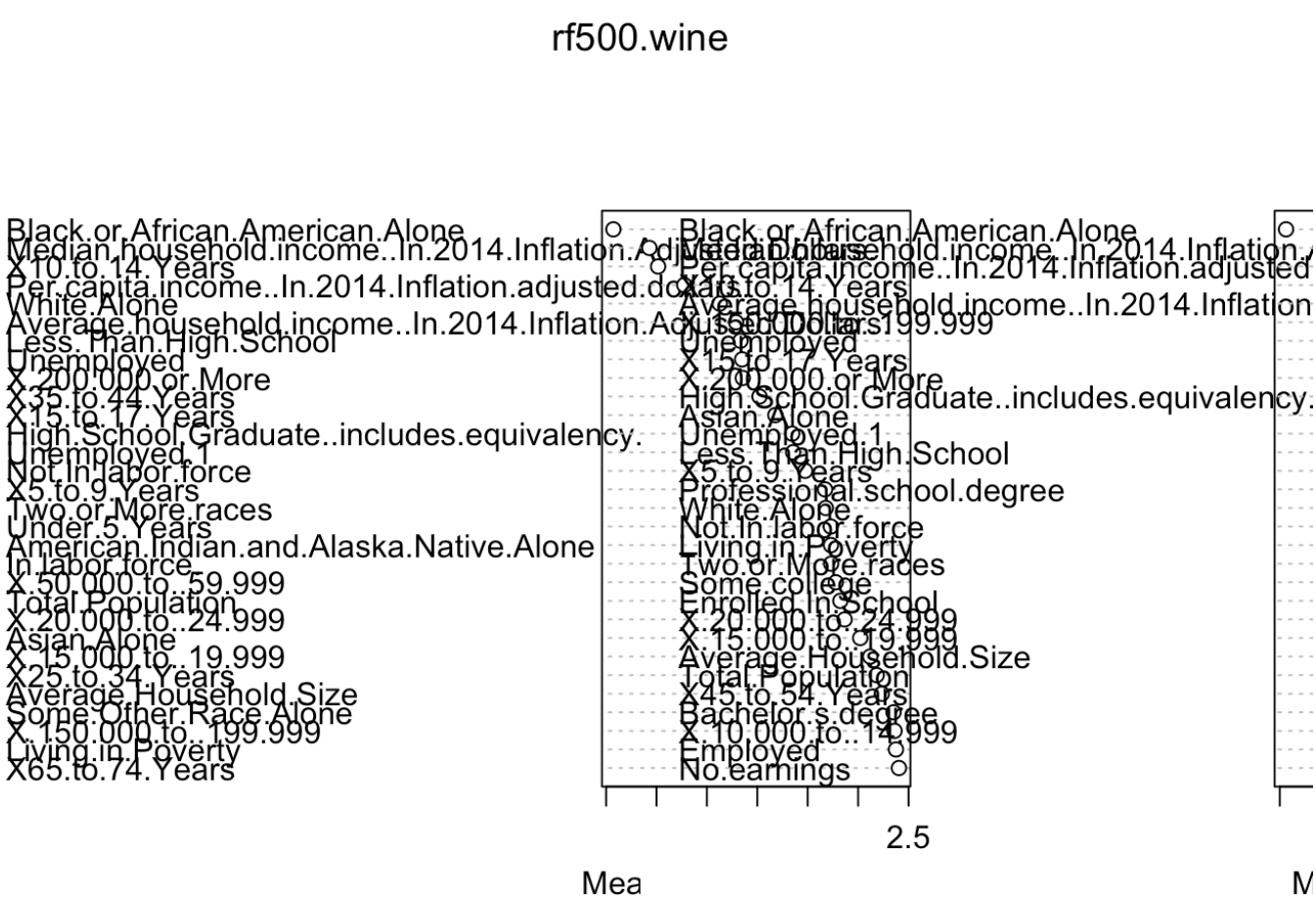
```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
set.seed(123)
start = proc.time()
rf500.wine = randomForest(Active~., data=train[,-c(1,2,67)], importance = T, nt
ree = 500)
total.time = proc.time()-start

pred1 <- predict(rf500.wine, test[,-c(1,2,67)])
mse.rf.500 = mean((1*(pred1=="Y")-1*(test$Active=="Y") )^2)

mse.rf.500
```

```
## [1] 0.356129
```

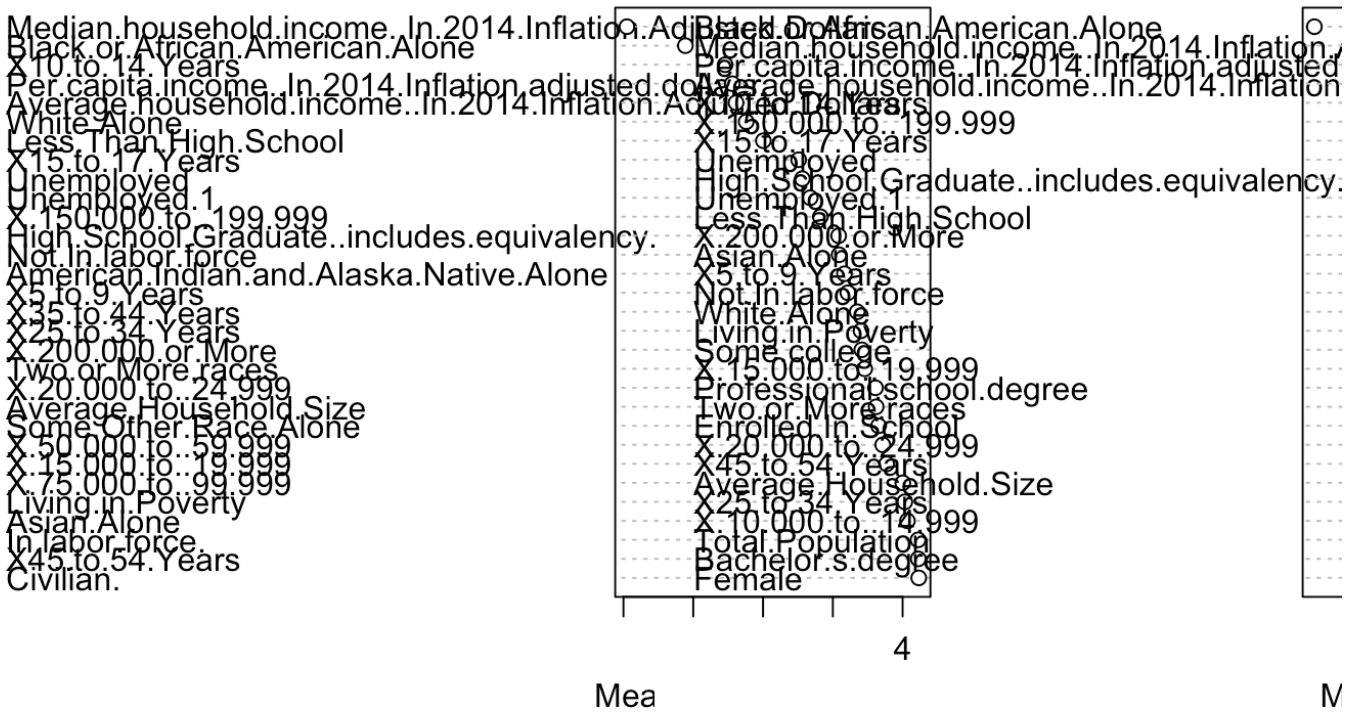
```
out.rf.500<- cbind(mse=mse.rf.500,t(total.time[1:3]))


varImpPlot(rf500.wine)
```




```
set.seed(123)
start = proc.time()
rf1000.wine = randomForest(Active~., data=train[,-c(1,2,67)], importance = T,nt
ree = 1000)
total.time = proc.time()-start

pred2 <- predict(rf1000.wine, test)
mse.rf.1000 = mean((1*(pred2=="Y")-1*(test$Active=="Y") )^2)
mse.rf.1000
```

```
## [1] 0.3535484
```

```
out.rf.1000<- cbind(mse=mse.rf.1000,t(total.time[1:3]))
varImpPlot(rf1000.wine)
```




```
set.seed(123)
start = proc.time()
rf2000.wine = randomForest(Active~., data=train[,-c(1,2,67)], importance = T,nt
ree = 2000)
total.time = proc.time()-start

pred3 <- predict(rf2000.wine, test)
mse.rf.2000 = mean((1*(pred3=="Y")-1*(test$Active=="Y") )^2)
mse.rf.2000
```

```
## [1] 0.3535484
```

```
out.rf.2000<- cbind(mse=mse.rf.2000,t(total.time[1:3]))
```

Boosting

```
#install.packages("gbm")
library(gbm)
```

```
## Loading required package: survival
```

```
## Loading required package: lattice
```

```
## Loading required package: splines
```

```
## Loading required package: parallel
```

```
## Loaded gbm 2.1.1
```

```
#train.b<- cbind(train[,-2],Active_flag=1*(train$Active=="Y"))
#test.b<- cbind(test[,-2],Active_flag=1*(test$Active=="Y"))

start = proc.time()
boost.wine1 = gbm(Active_flag~., train[,-c(1,2,3)], distribution =
'bernoulli',n.trees = 5000,interaction.depth = 1)
total.time = proc.time()-start

pred4 = round(predict(boost.wine1,newdata = test[,-c(1,2,3)], distribution = "b
ernoulli", n.trees = 5000, interaction.depth = 1,type = "response"))
mse.boost1 = mean((pred4 - test$Active_flag)^2)
```

```
## Warning in Ops.factor(pred4, test$Active_flag): '-' not meaningful for
## factors
```

```
mse.boost1
```

```
## [1] NA
```

```
out.boost1<- cbind(mse=mse.boost1,t(total.time[1:3]))


start = proc.time()
boost.wine3 = gbm(Active_flag~., train[,-c(1,2,3)], distribution =
'bernoulli',n.trees = 5000,interaction.depth = 3)
total.time = proc.time()-start

pred5 = round(predict(boost.wine3,newdata =test[,-c(1,2,3)], distribution = 'be
rnoulli', n.trees = 5000, interaction.depth = 3,type = "response"))
mse.boost3 = mean((pred5 - test$Active_flag)^2)
```

```
## Warning in Ops.factor(pred5, test$Active_flag): '-' not meaningful for
## factors
```

```
mse.boost3
```

```
## [1] NA
```

```
out.boost3 <- cbind(mse=mse.boost3,t(total.time[1:3]))
```

```
#Logistic Model
#install.packages("relaimpo")
#suppressMessages(library(relaimpo))

library(MASS)
# T-test on numeric variables
multi.ttest <- lapply(train[,c(4:66)], function(x) t.test(x ~ train[,67]))

options(scipen=999)
p <- data.frame(matrix(ncol = 1, nrow = 63))

for (i in 1:length(multi.ttest)){
  p[i,1] <-multi.ttest[[i]]$p.value
  rownames(p) <- names(multi.ttest[])
}

p$var_name <- rownames(p)
# variables of interest with p-value<0.05
p[p$matrix.ncol...1..nrow...63.< 0.05,]$var_name
```

```
##  [1] "Total.Population"
##  [2] "Male"
##  [3] "Female"
##  [4] "X5.to.9.Years"
##  [5] "X10.to.14.Years"
##  [6] "X15.to.17.Years"
##  [7] "X18.to.24.Years"
##  [8] "X35.to.44.Years"
##  [9] "X45.to.54.Years"
## [10] "X55.to.64.Years"
## [11] "X65.to.74.Years"
## [12] "X75.to.84.Years"
## [13] "X85.Years.and.over"
## [14] "White.Alone"
## [15] "Black.or.African.American.Alone"
## [16] "American.Indian.and.Alaska.Native.Alone"
## [17] "Asian.Alone"
## [18] "Some.Other.Race.Alone"
## [19] "Average.Household.Size"
## [20] "Less.Than.High.School"
## [21] "High.School.Graduate..includes.equivalency."
## [22] "Some.college"
## [23] "Bachelor.s.degree"
## [24] "Master.s.degree"
## [25] "Professional.school.degree"
## [26] "Doctorate.degree"
## [27] "Enrolled.In.School"
## [28] "Not.Enrolled.In.School"
## [29] "Unemployed"
## [30] "Not.In.labor.force"
## [31] "Unemployed.1"
## [32] "Less.than..10.000"
## [33] "X.10.000.to..14.999"
## [34] "X.15.000.to..19.999"
## [35] "X.20.000.to..24.999"
## [36] "X.25.000.to..29.999"
## [37] "X.30.000.to..34.999"
## [38] "X.35.000.to..39.999"
## [39] "X.40.000.to..44.999"
## [40] "X.45.000.to..49.999"
## [41] "X.50.000.to..59.999"
## [42] "X.75.000.to..99.999"
## [43] "X.100.000.to..124.999"
## [44] "X.125.000.to..149.999"
## [45] "X.150.000.to..199.999"
## [46] "X.200.000.or.More"
## [47] "Median.household.income..In.2014.Inflation.Adjusted.Dollars."
## [48] "Average.household.income..In.2014.Inflation.Adjusted.Dollars."
## [49] "No.earnings"
## [50] "Per.capita.income..In.2014.Inflation.adjusted.dollars."
## [51] "Living.in.Poverty"
```

```
#options(scipen=999)
#p.ch <- data.frame(matrix(ncol = 1, nrow = 55))
#for(i in 1:55){
#tbl = table(train[,i+66],train[,66])
#chisq <- chisq.test(tbl)
#p.ch[i,1] <- chisq$p.value
##rownames(p.ch) <- names(train)[i+66]
#tbl <- NULL
#chisq <- NULL
#}
```

```
#PCA on the variables selected
xPCA<-prcomp(train[,c(4:66)],scale=T)
names(xPCA)
```

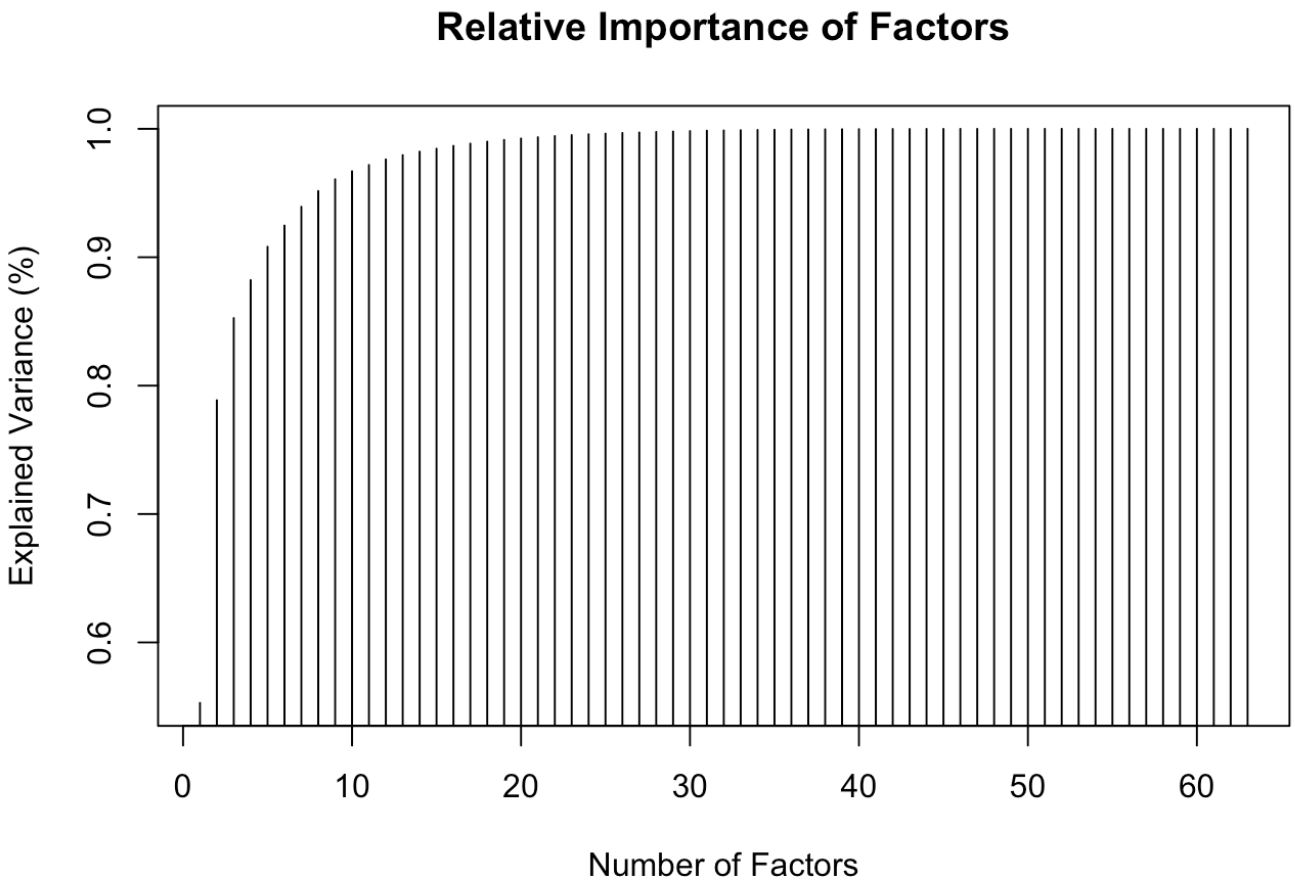
```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

```
summary(xPCA,loadings=T)
```

```
## Warning: In summary.prcomp(xPCA, loadings = T) :
##  extra argument 'loadings' will be disregarded
```

```
## Importance of components:
##                            PC1    PC2     PC3     PC4     PC5    PC6
## Standard deviation     5.9019 3.8540 2.00742 1.36529 1.27925 1.0196
## Proportion of Variance 0.5529 0.2358 0.06396 0.02959 0.02598 0.0165
## Cumulative Proportion  0.5529 0.7887 0.85262 0.88221 0.90819 0.9247
##                            PC7     PC8     PC9    PC10    PC11    PC12
## Standard deviation     0.96086 0.87751 0.75986 0.62805 0.55255 0.52557
## Proportion of Variance 0.01465 0.01222 0.00916 0.00626 0.00485 0.00438
## Cumulative Proportion  0.93934 0.95157 0.96073 0.96699 0.97184 0.97622
##                           PC13    PC14    PC15    PC16    PC17    PC18
## Standard deviation     0.45863 0.41440 0.37924 0.36991 0.33588 0.30991
## Proportion of Variance 0.00334 0.00273 0.00228 0.00217 0.00179 0.00152
## Cumulative Proportion  0.97956 0.98229 0.98457 0.98674 0.98853 0.99006
##                           PC19   PC20    PC21    PC22    PC23    PC24
## Standard deviation     0.28948 0.2630 0.25280 0.23448 0.21983 0.20187
## Proportion of Variance 0.00133 0.0011 0.00101 0.00087 0.00077 0.00065
## Cumulative Proportion  0.99139 0.9925 0.99350 0.99437 0.99514 0.99579
##                           PC25    PC26    PC27    PC28    PC29    PC30
## Standard deviation     0.1777 0.17486 0.16277 0.15510 0.14853 0.14508
## Proportion of Variance 0.0005 0.00049 0.00042 0.00038 0.00035 0.00033
## Cumulative Proportion  0.9963 0.99677 0.99719 0.99757 0.99792 0.99826
##                           PC31    PC32   PC33    PC34    PC35    PC36
## Standard deviation     0.13511 0.11481 0.1136 0.10253 0.09980 0.09419
## Proportion of Variance 0.00029 0.00021 0.0002 0.00017 0.00016 0.00014
## Cumulative Proportion  0.99855 0.99876 0.9990 0.99913 0.99929 0.99943
##                           PC37    PC38    PC39    PC40    PC41    PC42
## Standard deviation     0.08791 0.07565 0.07057 0.06710 0.06011 0.05030
## Proportion of Variance 0.00012 0.00009 0.00008 0.00007 0.00006 0.00004
## Cumulative Proportion  0.99955 0.99964 0.99972 0.99979 0.99985 0.99989
##                           PC43    PC44    PC45    PC46    PC47    PC48
## Standard deviation     0.04449 0.03841 0.03225 0.03175 0.02202 0.02094
## Proportion of Variance 0.00003 0.00002 0.00002 0.00002 0.00001 0.00001
## Cumulative Proportion  0.99992 0.99994 0.99996 0.99998 0.99998 0.99999
##                           PC49    PC50     PC51     PC52     PC53     PC54
## Standard deviation     0.01628 0.01483 0.005914 0.004713 0.001979 0.000341
## Proportion of Variance 0.00000 0.00000 0.000000 0.000000 0.000000 0.000000
## Cumulative Proportion  1.00000 1.00000 1.000000 1.000000 1.000000 1.000000
##                              PC55                 PC56                  PC57
## Standard deviation     0.00004896 0.00000000000002016 0.000000000000006319
## Proportion of Variance 0.00000000 0.00000000000000000 0.000000000000000000
## Cumulative Proportion  1.00000000 1.00000000000000000 1.000000000000000000
##                                        PC58                  PC59
## Standard deviation     0.000000000000004858 0.00000000000003825
## Proportion of Variance 0.000000000000000000 0.00000000000000000
## Cumulative Proportion  1.000000000000000000 1.00000000000000000
##                                        PC60                  PC61
## Standard deviation     0.000000000000002267 0.00000000000006207
## Proportion of Variance 0.000000000000000000 0.00000000000000000
## Cumulative Proportion  1.000000000000000000 1.00000000000000000
##                                         PC62                  PC63
## Standard deviation     0.0000000000000005754 0.0000000000000005754
## Proportion of Variance 0.0000000000000000000 0.0000000000000000000
## Cumulative Proportion  1.0000000000000000000 1.0000000000000000000
```

```
plot(summary(xPCA)$importance[3,],type="h",main="Relative Importance of Factor
s",
     xlab="Number of Factors",ylab="Explained Variance (%)")
```




```
nFactors<-5
factorLoadings<-xPCA$rotation[,1:nFactors]
factorScores<-as.matrix(train[,c(4:66)])%*%xPCA$rotation[,1:nFactors]
dim(factorScores)
```

```
## [1] 5422    5
```

```
zeroLoading<-xPCA$center

# building dataframe with zipcode with factor scores
factors10Data<-data.frame(Y=train[,c(67)],as.factor(train[,c(1)]),factorScores)
#as.factor(train[,c(1)]),
m10.PCA<-glm(Y~.,family=binomial(link='logit'),data=factors10Data)

summary(m10.PCA)
```

```
##
## Call:
## glm(formula = Y ~ ., family = binomial(link = "logit"), data = factors10Dat
a)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0393  -1.2373   0.7832   0.9520   1.2887
##
## Coefficients: (5 not defined because of singularities)
##                               Estimate Std. Error z value   Pr(>|z|)
## (Intercept)                    1.18270    0.26222   4.510 0.00000647 ***
## as.factor.train...c.1...60602  0.20360    0.52640   0.387   0.698920
## as.factor.train...c.1...60603 -0.26640    0.41080  -0.648   0.516665
## as.factor.train...c.1...60604 -0.18417    0.51405  -0.358   0.720145
## as.factor.train...c.1...60605  0.26422    0.41437   0.638   0.523700
## as.factor.train...c.1...60606  0.28364    0.38834   0.730   0.465154
## as.factor.train...c.1...60607  0.18621    0.33469   0.556   0.577963
## as.factor.train...c.1...60608 -0.54404    0.30317  -1.794   0.072734 .
## as.factor.train...c.1...60609 -0.62641    0.32458  -1.930   0.053616 .
## as.factor.train...c.1...60610 -0.10178    0.37316  -0.273   0.785039
## as.factor.train...c.1...60611  0.59859    0.37655   1.590   0.111904
## as.factor.train...c.1...60612 -1.44052    0.32175  -4.477 0.00000756 ***
## as.factor.train...c.1...60613 -0.31770    0.35773  -0.888   0.374490
## as.factor.train...c.1...60614 -0.26264    0.30846  -0.851   0.394519
## as.factor.train...c.1...60615 -0.15819    0.40722  -0.388   0.697673
## as.factor.train...c.1...60616 -0.48955    0.32457  -1.508   0.131479
## as.factor.train...c.1...60617 -0.62675    0.33031  -1.897   0.057771 .
## as.factor.train...c.1...60618 -0.44429    0.31005  -1.433   0.151862
## as.factor.train...c.1...60619 -1.41563    0.31327  -4.519 0.00000622 ***
## as.factor.train...c.1...60620 -1.09861    0.32005  -3.433   0.000598 ***
## as.factor.train...c.1...60621 -1.12385    0.35726  -3.146   0.001657 **
## as.factor.train...c.1...60622 -0.21988    0.31884  -0.690   0.490425
## as.factor.train...c.1...60623 -0.97335    0.30589  -3.182   0.001463 **
## as.factor.train...c.1...60624 -1.18270    0.33796  -3.500   0.000466 ***
## as.factor.train...c.1...60625 -0.49806    0.30701  -1.622   0.104744
## as.factor.train...c.1...60626 -0.62308    0.33528  -1.858   0.063113 .
## as.factor.train...c.1...60628 -1.06937    0.32652  -3.275   0.001057 **
## as.factor.train...c.1...60629 -1.12713    0.31074  -3.627   0.000286 ***
## as.factor.train...c.1...60630  0.13298    0.42902   0.310   0.756588
## as.factor.train...c.1...60631 -0.23823    0.51689  -0.461   0.644870
## as.factor.train...c.1...60632 -0.63718    0.31580  -2.018   0.043627 *
## as.factor.train...c.1...60633  0.07007    0.84357   0.083   0.933804
## as.factor.train...c.1...60634 -0.31770    0.35773  -0.888   0.374490
## as.factor.train...c.1...60636 -1.20290    0.33041  -3.641   0.000272 ***
## as.factor.train...c.1...60637 -1.14014    0.39231  -2.906   0.003658 **
## as.factor.train...c.1...60638  0.32138    0.47068   0.683   0.494730
## as.factor.train...c.1...60639 -0.82486    0.29752  -2.772   0.005564 **
## as.factor.train...c.1...60640 -0.35437    0.32645  -1.086   0.277686
## as.factor.train...c.1...60641 -0.66648    0.33454  -1.992   0.046346 *
## as.factor.train...c.1...60642  0.24134    0.40522   0.596   0.551457
## as.factor.train...c.1...60643 -0.91241    0.35994  -2.535   0.011249 *
## as.factor.train...c.1...60644 -1.11952    0.33303  -3.362   0.000775 ***
```

```
## as.factor.train...c.1...60645 -0.90506    0.37291  -2.427   0.015224 *
## as.factor.train...c.1...60646 -0.72316    0.45247  -1.598   0.109983
## as.factor.train...c.1...60647 -0.57532    0.29987  -1.919   0.055041 .
## as.factor.train...c.1...60649 -1.15800    0.34373  -3.369   0.000755 ***
## as.factor.train...c.1...60651 -1.04293    0.31610  -3.299   0.000969 ***
## as.factor.train...c.1...60652 -0.81497    0.40346  -2.020   0.043387 *
## as.factor.train...c.1...60653 -1.18270    0.43175  -2.739   0.006157 **
## as.factor.train...c.1...60654  0.37065    0.35190   1.053   0.292202
## as.factor.train...c.1...60655 -0.08408    0.57916  -0.145   0.884568
## as.factor.train...c.1...60656  0.76321    0.80012   0.954   0.340144
## as.factor.train...c.1...60657 -0.40042    0.30571  -1.310   0.190269
## as.factor.train...c.1...60659 -0.77723    0.31751  -2.448   0.014369 *
## as.factor.train...c.1...60660 -0.69434    0.35850  -1.937   0.052772 .
## as.factor.train...c.1...60661 -0.23823    0.36728  -0.649   0.516574
## PC1                                NA         NA      NA         NA
## PC2                                NA         NA      NA         NA
## PC3                                NA         NA      NA         NA
## PC4                                NA         NA      NA         NA
## PC5                                NA         NA      NA         NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7067.1  on 5421  degrees of freedom
## Residual deviance: 6795.9  on 5366  degrees of freedom
## AIC: 6907.9
##
## Number of Fisher Scoring iterations: 4
```

```
# When run with zipcode as one of the variables, PC`s come out to be insignific
ant.
#hence removing zipcode and running the code again

# Building a dataframe without zipcode and only PC`s
factors10Data<-data.frame(Y=train[,c(67)],factorScores)
#as.factor(train[,c(1)]),
m10.PCA<-glm(Y~.,family=binomial(link='logit'),data=factors10Data)

summary(m10.PCA)
```

```
##
## Call:
## glm(formula = Y ~ ., family = binomial(link = "logit"), data = factors10Dat
a)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0348  -1.2591   0.7947   0.9438   1.2066
##
## Coefficients:
##                 Estimate   Std. Error z value Pr(>|z|)
## (Intercept)  0.298652442  0.152161523   1.963   0.0497 *
## PC1          0.000004148  0.000001988   2.086   0.0369 *
## PC2         -0.000005617  0.000002405  -2.335   0.0195 *
## PC3          0.000005817  0.000003824   1.521   0.1282
## PC4          0.000014857  0.000006929   2.144   0.0320 *
## PC5         -0.000017180  0.000007623  -2.254   0.0242 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7067.1  on 5421  degrees of freedom
## Residual deviance: 6876.1  on 5416  degrees of freedom
## AIC: 6888.1
##
## Number of Fisher Scoring iterations: 4
```

```
#PC`s are significant but don`t contribute anything to predicted probabilities.
# Estimates are zero to five decimals
```

```
#Standardize the data
#online.numeric <- apply(X[,c(4:66)],2,function(z) z/sd(z))

#online.numeric<- data.frame(online.numeric)

#online.norm <- data.frame(online.numeric,X[,c(1,67)])

#smp_size <- floor(0.7 * nrow(X))

## set the seed to make your partition reproductible
#set.seed(123)
#train <- sample(seq_len(nrow(X)), size = smp_size)
#test.lasso<- online.norm[-train,]
#train.lasso <- online.norm[train,]
#train.lasso <- data.frame(train.lasso)
#test.lasso <- data.frame(test.lasso)
```

```
#Lasso Regression Model
#http://stats.stackexchange.com/questions/72251/an-example-lasso-regression-usi
ng-glmnet-for-binary-outcome
#install.packages("glmnet")
#library(glmnet)

#xfactors.model.matrix <- model.matrix(~.,train[,c(46:59)])[,-1]
#X<- as.matrix(data.frame(train.lasso[,c(1:63)]))
#Y<- train.lasso[,65]

#lasso.online<- glmnet(x=X,y=as.factor(Y),alpha=1,nlambda=100,lambda.min.ratio=
0.0001,standardize=F,family='binomial')

#names(lasso.online)
#head(cbind(DF=lasso.online$df,DevExpected=lasso.online$dev.ratio,Lambda=lasso.
online$lambda))

#coef(lasso.online)[,10]
#Parameter df returned by glmnet() is the number of active predictors at the co
rresponding level of lambda

#plot(lasso.online$lambda,lasso.online$dev.ratio,type="l")
# explained deviance (%) grows as ???? becomes smaller

#plot(lasso.online)

#set.seed(15)
#cv.out=cv.glmnet(x=X,y=Y,alpha=1,family='binomial')
#plot(cv.out)
#bestlam =cv.out$lambda.min

#lasso.pred=predict(lasso.online,s=bestlam,
                    #newx=data.matrix(data.frame(test.lasso[,c(1:63)])),family
='binomial')
#testOutput<-test.lasso[,65]

#(MSE.lasso<-mean((testOutput - lasso.pred )^2))

#options(scipen=999)
#lasso.coef=predict(lasso.online,type="coefficients",s=bestlam)
#lasso.coef

#use the best ?? to fit the model to the entire sample
#out=glmnet(x=data.matrix(data.frame(online.norm[,c(1:44)],model.matrix(~.,onli
ne.norm[,c(46:59)])[,-1])),y=as.vector(online.norm[,45]),alpha=1,lambda=bestla
m,standardize=F)


# Lasso tried but failed since the prediction were not probabilities
```

```
#Unsuccessful in variable selection, going ahead with variables selected by Ran
dom Forest

#install.packages("Hmisc")
#library(Hmisc)
#vclus <- varclus(data.matrix(train[,c(3:66)]))
#head(vclus)

importance(rf1000.wine)
```

```
##                                                          N
## Total.Population                                  0.79536806
## Male                                              0.46784942
## Female                                           -0.06813982
## Under.5.Years                                    -0.62695813
## X5.to.9.Years                                    -0.91709938
## X10.to.14.Years                                   1.73805363
## X15.to.17.Years                                   3.61496060
## X18.to.24.Years                                   0.35426574
## X25.to.34.Years                                   2.30241904
## X35.to.44.Years                                  -1.52287676
## X45.to.54.Years                                  -1.53092972
## X55.to.64.Years                                   0.70325439
## X65.to.74.Years                                  -2.33868473
## X75.to.84.Years                                  -1.01360587
## X85.Years.and.over                                0.52969962
## White.Alone                                       8.71022219
## Black.or.African.American.Alone                  12.67656318
## American.Indian.and.Alaska.Native.Alone           2.38027909
## Asian.Alone                                       7.73798795
## Native.Hawaiian.and.Other.Pacific.Islander.Alone -2.95211331
## Some.Other.Race.Alone                             0.15581298
## Two.or.More.races                                 4.09276251
## Average.Household.Size                           -2.80596500
## Less.Than.High.School                            -0.75475780
## High.School.Graduate..includes.equivalency.       0.78885138
## Some.college                                     -0.19480530
## Bachelor.s.degree                                 0.21124712
## Master.s.degree                                  -2.22474001
## Professional.school.degree                        3.82713734
## Doctorate.degree                                 -2.09292218
## Enrolled.In.School                               -2.62127403
## Not.Enrolled.In.School                           -0.09202545
## In.labor.force.                                  -2.67568017
## In.Armed.Forces                                  -1.00600396
## Civilian.                                        -0.23210159
## Employed                                          2.39058291
## Unemployed                                        1.83769339
## Not.In.labor.force                                0.20537654
## Employed.1                                        2.08723281
## Unemployed.1                                      1.41654891
## Less.than..10.000                                 3.19856404
## X.10.000.to..14.999                               1.11598735
## X.15.000.to..19.999                               0.46872761
## X.20.000.to..24.999                              -1.20854077
## X.25.000.to..29.999                              -2.70263498
## X.30.000.to..34.999                              -2.44991183
## X.35.000.to..39.999                              -1.95000894
## X.40.000.to..44.999                              -0.77678521
## X.45.000.to..49.999                              -0.71095416
## X.50.000.to..59.999                              -2.30637705
## X.60.000.to..74.999                              -0.22683042
## X.75.000.to..99.999                               0.53129221
```

```
## X.100.000.to..124.999                                          0.09905136
## X.125.000.to..149.999                                          0.59002002
## X.150.000.to..199.999                                          6.13965303
## X.200.000.or.More                                              0.79654402
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.   6.13977247
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.  3.34828849
## With.earnings                                                  0.27924514
## No.earnings                                                    3.73783033
## Per.capita.income..In.2014.Inflation.adjusted.dollars.         4.47130613
## Living.in.Poverty                                             -0.17045271
## At.or.Above.Poverty.Level                                      0.79710428
## start.2010...c.2..60602                                        0.00000000
## start.2010...c.2..60603                                        0.00000000
## start.2010...c.2..60604                                        0.00000000
## start.2010...c.2..60605                                       -1.00050038
## start.2010...c.2..60606                                        0.00000000
## start.2010...c.2..60607                                        0.00000000
## start.2010...c.2..60608                                        1.51795628
## start.2010...c.2..60609                                       -0.92857651
## start.2010...c.2..60610                                        0.00000000
## start.2010...c.2..60611                                        0.00000000
## start.2010...c.2..60612                                        2.42152093
## start.2010...c.2..60613                                        0.00000000
## start.2010...c.2..60614                                        0.00000000
## start.2010...c.2..60615                                       -1.00050038
## start.2010...c.2..60616                                        0.00000000
## start.2010...c.2..60617                                       -1.19993916
## start.2010...c.2..60618                                        0.00000000
## start.2010...c.2..60619                                        4.05507604
## start.2010...c.2..60620                                       -0.43663782
## start.2010...c.2..60621                                        0.00000000
## start.2010...c.2..60622                                        0.00000000
## start.2010...c.2..60623                                       -0.34697986
## start.2010...c.2..60624                                       -1.04201694
## start.2010...c.2..60625                                        0.00000000
## start.2010...c.2..60626                                        0.00000000
## start.2010...c.2..60628                                       -0.18296247
## start.2010...c.2..60629                                        1.92543548
## start.2010...c.2..60630                                        0.00000000
## start.2010...c.2..60631                                        0.00000000
## start.2010...c.2..60632                                        1.00050038
## start.2010...c.2..60633                                        1.00050038
## start.2010...c.2..60634                                        0.00000000
## start.2010...c.2..60636                                        0.59049884
## start.2010...c.2..60637                                        0.48334910
## start.2010...c.2..60638                                        1.00050038
## start.2010...c.2..60639                                       -1.00050038
## start.2010...c.2..60640                                        0.00000000
## start.2010...c.2..60641                                        0.00000000
## start.2010...c.2..60642                                        0.00000000
## start.2010...c.2..60643                                        0.73041023
## start.2010...c.2..60644                                        0.42998749
## start.2010...c.2..60645                                        2.34994854
## start.2010...c.2..60646                                        3.19378673
```

```
## start.2010...c.2..60647                                    0.00000000
## start.2010...c.2..60649                                    0.24052273
## start.2010...c.2..60651                                   -0.35479729
## start.2010...c.2..60652                                    0.12122223
## start.2010...c.2..60653                                   -0.03513719
## start.2010...c.2..60654                                    0.00000000
## start.2010...c.2..60655                                    0.00000000
## start.2010...c.2..60656                                    0.00000000
## start.2010...c.2..60657                                    0.00000000
## start.2010...c.2..60659                                   -1.00050038
## start.2010...c.2..60660                                    0.00000000
## start.2010...c.2..60661                                    0.00000000
##                                                                     Y
## Total.Population                                           1.88274595
## Male                                                       0.77737382
## Female                                                     1.90922675
## Under.5.Years                                              1.84348171
## X5.to.9.Years                                              3.55675328
## X10.to.14.Years                                            4.45994648
## X15.to.17.Years                                            1.89047138
## X18.to.24.Years                                            0.67802607
## X25.to.34.Years                                            0.47428983
## X35.to.44.Years                                            3.25365311
## X45.to.54.Years                                            3.31296171
## X55.to.64.Years                                            1.09238090
## X65.to.74.Years                                            3.04121043
## X75.to.84.Years                                            2.55329995
## X85.Years.and.over                                         1.17564866
## White.Alone                                               -4.49256044
## Black.or.African.American.Alone                           -7.39742692
## American.Indian.and.Alaska.Native.Alone                    0.96494198
## Asian.Alone                                               -4.84112007
## Native.Hawaiian.and.Other.Pacific.Islander.Alone           2.02728713
## Some.Other.Race.Alone                                      2.31704307
## Two.or.More.races                                         -0.12423000
## Average.Household.Size                                     4.16668133
## Less.Than.High.School                                      4.58895425
## High.School.Graduate..includes.equivalency.                2.81508834
## Some.college                                               1.54540156
## Bachelor.s.degree                                          0.63163978
## Master.s.degree                                            2.90578477
## Professional.school.degree                                -3.12724839
## Doctorate.degree                                           2.51906620
## Enrolled.In.School                                         3.34984463
## Not.Enrolled.In.School                                     1.22645855
## In.labor.force.                                            3.50406360
## In.Armed.Forces                                            0.97578470
## Civilian.                                                  2.15355626
## Employed                                                  -0.23247452
## Unemployed                                                 3.05831443
## Not.In.labor.force                                         3.81382586
## Employed.1                                                 0.26421334
## Unemployed.1                                               3.09264892
## Less.than..10.000                                         -1.18506259
```

```
## X.10.000.to..14.999                                                 1.41741375
## X.15.000.to..19.999                                                 2.18423813
## X.20.000.to..24.999                                                 3.62830735
## X.25.000.to..29.999                                                 3.34020391
## X.30.000.to..34.999                                                 2.85735720
## X.35.000.to..39.999                                                 2.96790162
## X.40.000.to..44.999                                                 1.43692437
## X.45.000.to..49.999                                                 1.62426174
## X.50.000.to..59.999                                                 3.68828902
## X.60.000.to..74.999                                                 1.87863100
## X.75.000.to..99.999                                                 1.59787164
## X.100.000.to..124.999                                               0.81061446
## X.125.000.to..149.999                                               0.88706550
## X.150.000.to..199.999                                              -2.08550222
## X.200.000.or.More                                                   1.62228285
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.        1.23964495
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.       2.33537055
## With.earnings                                                       1.63439979
## No.earnings                                                        -1.19904260
## Per.capita.income..In.2014.Inflation.adjusted.dollars.              1.39003078
## Living.in.Poverty                                                   2.69211182
## At.or.Above.Poverty.Level                                           0.59343114
## start.2010...c.2..60602                                             0.00000000
## start.2010...c.2..60603                                             0.00000000
## start.2010...c.2..60604                                             0.00000000
## start.2010...c.2..60605                                             1.00050038
## start.2010...c.2..60606                                             0.00000000
## start.2010...c.2..60607                                             0.00000000
## start.2010...c.2..60608                                            -1.52415698
## start.2010...c.2..60609                                             1.15489284
## start.2010...c.2..60610                                             0.00000000
## start.2010...c.2..60611                                             0.00000000
## start.2010...c.2..60612                                            -2.04030574
## start.2010...c.2..60613                                             0.00000000
## start.2010...c.2..60614                                             0.00000000
## start.2010...c.2..60615                                             1.00050038
## start.2010...c.2..60616                                            -1.00050038
## start.2010...c.2..60617                                             1.35329729
## start.2010...c.2..60618                                             0.00000000
## start.2010...c.2..60619                                            -3.86938708
## start.2010...c.2..60620                                             0.38644970
## start.2010...c.2..60621                                             0.00000000
## start.2010...c.2..60622                                             0.00000000
## start.2010...c.2..60623                                            -0.11630114
## start.2010...c.2..60624                                             0.64914938
## start.2010...c.2..60625                                             0.00000000
## start.2010...c.2..60626                                             0.00000000
## start.2010...c.2..60628                                            -0.45594714
## start.2010...c.2..60629                                            -2.00219070
## start.2010...c.2..60630                                             0.00000000
## start.2010...c.2..60631                                            -1.00050038
## start.2010...c.2..60632                                            -1.46759483
## start.2010...c.2..60633                                            -1.40817079
## start.2010...c.2..60634                                             0.00000000
```

```
## start.2010...c.2..60636                                    -1.32351123
## start.2010...c.2..60637                                    -0.62269989
## start.2010...c.2..60638                                     0.64627788
## start.2010...c.2..60639                                     1.00050038
## start.2010...c.2..60640                                     0.00000000
## start.2010...c.2..60641                                     0.00000000
## start.2010...c.2..60642                                     0.00000000
## start.2010...c.2..60643                                    -0.78023123
## start.2010...c.2..60644                                    -1.43140868
## start.2010...c.2..60645                                    -2.70297618
## start.2010...c.2..60646                                    -3.53869119
## start.2010...c.2..60647                                     0.00000000
## start.2010...c.2..60649                                    -0.56022902
## start.2010...c.2..60651                                     0.04382267
## start.2010...c.2..60652                                    -0.77237956
## start.2010...c.2..60653                                    -0.73927219
## start.2010...c.2..60654                                     0.00000000
## start.2010...c.2..60655                                     0.00000000
## start.2010...c.2..60656                                     0.00000000
## start.2010...c.2..60657                                     0.00000000
## start.2010...c.2..60659                                     1.00050038
## start.2010...c.2..60660                                     0.00000000
## start.2010...c.2..60661                                     0.00000000
##                                                          MeanDecreaseAc
curacy
## Total.Population                                                    3.6
639046
## Male                                                                1.9
589065
## Female                                                              2.7
717761
## Under.5.Years                                                       3.0
743199
## X5.to.9.Years                                                       4.7
845599
## X10.to.14.Years                                                     6.5
428676
## X15.to.17.Years                                                     5.4
899083
## X18.to.24.Years                                                     2.0
815382
## X25.to.34.Years                                                     4.5
958319
## X35.to.44.Years                                                     4.6
502799
## X45.to.54.Years                                                     3.7
759522
## X55.to.64.Years                                                     2.3
706505
## X65.to.74.Years                                                     3.0
114289
## X75.to.84.Years                                                     3.0
860241
## X85.Years.and.over                                                  2.4
```

```
885597
## White.Alone                                                    6.2
410407
## Black.or.African.American.Alone                                7.1
126405
## American.Indian.and.Alaska.Native.Alone                        4.8
332904
## Asian.Alone                                                    3.9
314172
## Native.Hawaiian.and.Other.Pacific.Islander.Alone              -1.0
337235
## Some.Other.Race.Alone                                          4.3
444201
## Two.or.More.races                                              4.5
438373
## Average.Household.Size                                         4.3
796880
## Less.Than.High.School                                          5.9
952808
## High.School.Graduate..includes.equivalency.                    4.9
165666
## Some.college                                                   2.5
047632
## Bachelor.s.degree                                              2.5
574122
## Master.s.degree                                                2.9
517225
## Professional.school.degree                                     1.1
021281
## Doctorate.degree                                               1.9
464386
## Enrolled.In.School                                             2.9
852234
## Not.Enrolled.In.School                                         2.4
740162
## In.labor.force.                                                3.7
786102
## In.Armed.Forces                                                0.5
382592
## Civilian.                                                      3.7
690561
## Employed                                                       3.1
397833
## Unemployed                                                     5.4
373969
## Not.In.labor.force                                             4.9
096608
## Employed.1                                                     3.3
480207
## Unemployed.1                                                   5.3
239571
## Less.than..10.000                                              2.4
183206
## X.10.000.to..14.999                                            3.1
```

```
067732
## X.15.000.to..19.999                                                    4.2
292166
## X.20.000.to..24.999                                                    4.3
889856
## X.25.000.to..29.999                                                    3.3
418073
## X.30.000.to..34.999                                                    2.4
609546
## X.35.000.to..39.999                                                    3.0
794083
## X.40.000.to..44.999                                                    1.6
484714
## X.45.000.to..49.999                                                    2.5
003914
## X.50.000.to..59.999                                                    4.2
849671
## X.60.000.to..74.999                                                    3.3
325604
## X.75.000.to..99.999                                                    3.9
972537
## X.100.000.to..124.999                                                  1.9
076343
## X.125.000.to..149.999                                                  2.8
063369
## X.150.000.to..199.999                                                  5.1
786141
## X.200.000.or.More                                                      4.5
846191
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.            7.9
508231
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.           6.3
864169
## With.earnings                                                          3.6
954246
## No.earnings                                                            3.3
756754
## Per.capita.income..In.2014.Inflation.adjusted.dollars.                 6.4
406841
## Living.in.Poverty                                                      3.9
690667
## At.or.Above.Poverty.Level                                              2.5
390151
## start.2010...c.2..60602                                                0.0
000000
## start.2010...c.2..60603                                                0.0
000000
## start.2010...c.2..60604                                                0.0
000000
## start.2010...c.2..60605                                                1.0
005004
## start.2010...c.2..60606                                                0.0
000000
## start.2010...c.2..60607                                                0.0
```

```
000000
## start.2010...c.2..60608                                    0.5
825157
## start.2010...c.2..60609                                    1.0
198997
## start.2010...c.2..60610                                    0.0
000000
## start.2010...c.2..60611                                    0.0
000000
## start.2010...c.2..60612                                    1.6
708022
## start.2010...c.2..60613                                    0.0
000000
## start.2010...c.2..60614                                    0.0
000000
## start.2010...c.2..60615                                    1.0
005004
## start.2010...c.2..60616                                   -1.0
005004
## start.2010...c.2..60617                                    1.1
166137
## start.2010...c.2..60618                                    0.0
000000
## start.2010...c.2..60619                                    2.7
897535
## start.2010...c.2..60620                                    0.1
387340
## start.2010...c.2..60621                                    0.0
000000
## start.2010...c.2..60622                                    0.0
000000
## start.2010...c.2..60623                                   -1.4
565028
## start.2010...c.2..60624                                   -1.7
180769
## start.2010...c.2..60625                                    0.0
000000
## start.2010...c.2..60626                                    0.0
000000
## start.2010...c.2..60628                                   -1.4
894574
## start.2010...c.2..60629                                   -1.6
714047
## start.2010...c.2..60630                                    0.0
000000
## start.2010...c.2..60631                                   -1.0
005004
## start.2010...c.2..60632                                   -1.6
029813
## start.2010...c.2..60633                                   -1.4
155379
## start.2010...c.2..60634                                    0.0
000000
## start.2010...c.2..60636                                   -1.5
```

```
690143
## start.2010...c.2..60637                                          -0.4
791193
## start.2010...c.2..60638                                           0.8
686130
## start.2010...c.2..60639                                           1.0
005004
## start.2010...c.2..60640                                           0.0
000000
## start.2010...c.2..60641                                           0.0
000000
## start.2010...c.2..60642                                           0.0
000000
## start.2010...c.2..60643                                          -0.6
588948
## start.2010...c.2..60644                                          -2.0
264015
## start.2010...c.2..60645                                          -2.4
191178
## start.2010...c.2..60646                                          -3.4
037646
## start.2010...c.2..60647                                           0.0
000000
## start.2010...c.2..60649                                          -1.6
477942
## start.2010...c.2..60651                                          -0.5
790217
## start.2010...c.2..60652                                          -1.1
449816
## start.2010...c.2..60653                                          -0.9
041655
## start.2010...c.2..60654                                           0.0
000000
## start.2010...c.2..60655                                           0.0
000000
## start.2010...c.2..60656                                           0.0
000000
## start.2010...c.2..60657                                           0.0
000000
## start.2010...c.2..60659                                           1.0
005004
## start.2010...c.2..60660                                           0.0
000000
## start.2010...c.2..60661                                           0.0
000000
##                                                    MeanDecreaseGi
ni
## Total.Population                                     1.7002787070
05
## Male                                                 1.0358055264
45
## Female                                               1.6204010188
58
## Under.5.Years                                        0.8767282873
```

```
81
## X5.to.9.Years                                                3.1726411800
22
## X10.to.14.Years                                              5.5197333948
16
## X15.to.17.Years                                              4.3797345914
33
## X18.to.24.Years                                              1.1270225364
07
## X25.to.34.Years                                              1.7534408640
01
## X35.to.44.Years                                              1.1870264752
54
## X45.to.54.Years                                              2.0810709958
53
## X55.to.64.Years                                              0.9083814708
73
## X65.to.74.Years                                              0.9054737408
94
## X75.to.84.Years                                              0.9074286742
50
## X85.Years.and.over                                           0.8453433419
68
## White.Alone                                                  2.9800310545
49
## Black.or.African.American.Alone                              9.5229333456
64
## American.Indian.and.Alaska.Native.Alone                      1.2569897632
32
## Asian.Alone                                                  3.1736654820
09
## Native.Hawaiian.and.Other.Pacific.Islander.Alone             0.3519001162
72
## Some.Other.Race.Alone                                        1.2658543481
84
## Two.or.More.races                                            2.4545691439
65
## Average.Household.Size                                       1.8343269657
48
## Less.Than.High.School                                        3.4966153869
44
## High.School.Graduate..includes.equivalency.                  3.7588514875
83
## Some.college                                                 2.7209035906
57
## Bachelor.s.degree                                            1.6618918590
25
## Master.s.degree                                              1.1527931270
52
## Professional.school.degree                                   2.4793002186
93
## Doctorate.degree                                             1.1777412309
69
## Enrolled.In.School                                           2.1338189925
```

```
48
## Not.Enrolled.In.School                                          0.8471502886
25
## In.labor.force.                                                 0.8790403017
74
## In.Armed.Forces                                                 0.2786161370
42
## Civilian.                                                       0.9860196865
04
## Employed                                                        1.3452568131
40
## Unemployed                                                      4.1939124945
96
## Not.In.labor.force                                              3.0017586724
96
## Employed.1                                                      1.3759448794
76
## Unemployed.1                                                    3.5527757788
17
## Less.than..10.000                                               1.4614861312
19
## X.10.000.to..14.999                                             1.7489059779
01
## X.15.000.to..19.999                                             2.7096279523
82
## X.20.000.to..24.999                                             2.1088572011
03
## X.25.000.to..29.999                                             1.2812459522
90
## X.30.000.to..34.999                                             0.9286275582
18
## X.35.000.to..39.999                                             1.1420880728
31
## X.40.000.to..44.999                                             0.9508456506
83
## X.45.000.to..49.999                                             1.3511409899
60
## X.50.000.to..59.999                                             1.2706181063
74
## X.60.000.to..74.999                                             1.0827137071
04
## X.75.000.to..99.999                                             1.4515050022
99
## X.100.000.to..124.999                                           1.5287159089
14
## X.125.000.to..149.999                                           1.5749072284
85
## X.150.000.to..199.999                                           5.1223626524
63
## X.200.000.or.More                                               3.4394521223
70
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.    8.1261277000
50
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.   5.7455014360
```

```
33
## With.earnings                                                    1.1357852890
63
## No.earnings                                                      1.4594009576
42
## Per.capita.income..In.2014.Inflation.adjusted.dollars.           7.2479913091
60
## Living.in.Poverty                                                2.8330910517
47
## At.or.Above.Poverty.Level                                        1.0852106126
74
## start.2010...c.2..60602                                          0.0086964485
37
## start.2010...c.2..60603                                          0.0328032351
78
## start.2010...c.2..60604                                          0.0020762071
67
## start.2010...c.2..60605                                          0.0058697790
35
## start.2010...c.2..60606                                          0.0193076574
30
## start.2010...c.2..60607                                          0.0040654088
92
## start.2010...c.2..60608                                          0.0262487980
54
## start.2010...c.2..60609                                          0.0416456011
80
## start.2010...c.2..60610                                          0.0099751228
14
## start.2010...c.2..60611                                          0.0149477887
53
## start.2010...c.2..60612                                          0.0961440983
93
## start.2010...c.2..60613                                          0.0120411760
58
## start.2010...c.2..60614                                          0.0037874162
11
## start.2010...c.2..60615                                          0.0294968689
07
## start.2010...c.2..60616                                          0.0120023652
94
## start.2010...c.2..60617                                          0.0312679568
24
## start.2010...c.2..60618                                          0.0038527868
35
## start.2010...c.2..60619                                          0.1505159885
87
## start.2010...c.2..60620                                          0.0044081651
77
## start.2010...c.2..60621                                          0.0000018746
96
## start.2010...c.2..60622                                          0.0057017091
14
## start.2010...c.2..60623                                          0.0138291411
```

```
16
## start.2010...c.2..60624                                    0.0089199387
31
## start.2010...c.2..60625                                    0.0133638192
80
## start.2010...c.2..60626                                    0.0073504890
36
## start.2010...c.2..60628                                    0.0125321313
36
## start.2010...c.2..60629                                    0.0292227373
04
## start.2010...c.2..60630                                    0.0540701403
76
## start.2010...c.2..60631                                    0.0097496532
54
## start.2010...c.2..60632                                    0.0186668183
71
## start.2010...c.2..60633                                    0.0047768732
67
## start.2010...c.2..60634                                    0.0299701590
80
## start.2010...c.2..60636                                    0.0066809170
50
## start.2010...c.2..60637                                    0.0105017680
92
## start.2010...c.2..60638                                    0.1071127870
77
## start.2010...c.2..60639                                    0.0090831365
14
## start.2010...c.2..60640                                    0.0014666755
97
## start.2010...c.2..60641                                    0.0114924682
29
## start.2010...c.2..60642                                    0.0071168892
13
## start.2010...c.2..60643                                    0.0405725492
82
## start.2010...c.2..60644                                    0.0225732797
27
## start.2010...c.2..60645                                    0.0463742988
49
## start.2010...c.2..60646                                    0.1362389683
08
## start.2010...c.2..60647                                    0.0072234370
81
## start.2010...c.2..60649                                    0.0116387867
03
## start.2010...c.2..60651                                    0.0051294101
58
## start.2010...c.2..60652                                    0.0216873104
90
## start.2010...c.2..60653                                    0.0053928353
65
## start.2010...c.2..60654                                    0.0140125120
```

```
11
## start.2010...c.2..60655                                    0.0065490260
03
## start.2010...c.2..60656                                    0.0361892733
32
## start.2010...c.2..60657                                    0.0032770779
15
## start.2010...c.2..60659                                    0.0180174259
62
## start.2010...c.2..60660                                    0.0209088246
65
## start.2010...c.2..60661                                    0.1087966491
05
```

```
str(train)
```

```
## 'data.frame':    5422 obs. of  122 variables:
##  $ zipcode                                                    : Factor w/
56 levels "60601","60602",..: 17 44 22 49 53 5 27 50 28 24 ...
##  $ Duration                                                   : int  739 1
461 742 755 2200 1481 2710 905 748 720 ...
##  $ Active                                                     : Factor w/
2 levels "N","Y": 2 2 1 2 2 2 2 2 2 2 ...
##  $ Total.Population                                           : int  82685
28367 54607 31038 69444 25938 69921 16244 115013 39706 ...
##  $ Male                                                       : int  38246
13492 27887 12982 34116 12743 30621 7822 55840 18476 ...
##  $ Female                                                     : int  44439
14875 26720 18056 35328 13195 39300 8422 59173 21230 ...
##  $ Under.5.Years                                              : int  44 9
27 11 31 4 36 51 55 26 ...
##  $ X5.to.9.Years                                              : int  5874
1940 2444 2354 1923 728 4941 125 10612 3634 ...
##  $ X10.to.14.Years                                            : int  6882
1638 2096 2366 1174 198 5078 82 9681 3731 ...
##  $ X15.to.17.Years                                            : int  3402
831 1145 1381 572 242 3031 14 6501 2110 ...
##  $ X18.to.24.Years                                            : int  8191
1521 4761 2899 9406 4088 7336 1269 13287 4918 ...
##  $ X25.to.34.Years                                            : int  9768
2837 19352 3847 27025 7485 8130 7095 18031 5328 ...
##  $ X35.to.44.Years                                            : int  10781
3789 9567 4753 9706 4285 8283 3023 16711 4362 ...
##  $ X45.to.54.Years                                            : int  11220
4126 5061 3923 6687 2548 9233 1839 13394 5082 ...
##  $ X55.to.64.Years                                            : int  8980
3653 3195 3093 4370 2715 8430 1457 9515 3226 ...
##  $ X65.to.74.Years                                            : int  6853
2814 1648 2244 2682 1458 6501 460 4938 2045 ...
##  $ X75.to.84.Years                                            : int  3317
1842 1115 1268 1415 698 3834 121 2133 1257 ...
##  $ X85.Years.and.over                                         : int  1638
1251 512 664 651 90 1039 30 1081 331 ...
##  $ White.Alone                                                : int  29693
23749 40374 1542 60182 15282 2557 12437 40073 1004 ...
##  $ Black.or.African.American.Alone                            : int  45057
184 4542 28443 1840 5108 65974 745 25258 37607 ...
##  $ American.Indian.and.Alaska.Native.Alone                    : int  186 0
180 17 101 9 32 100 576 47 ...
##  $ Asian.Alone                                                : int  190 2
896 1939 293 4630 4291 211 1890 745 63 ...
##  $ Native.Hawaiian.and.Other.Pacific.Islander.Alone           : int  23 0
0 18 17 12 0 0 0 0 ...
##  $ Some.Other.Race.Alone                                      : int  6317
737 5695 281 825 428 602 555 46385 798 ...
##  $ Two.or.More.races                                          : int  1219
801 1877 444 1849 808 545 517 1976 187 ...
##  $ Average.Household.Size                                     : num  3 2.6
2.3 2.2 1.9 1.7 2.9 1.5 3.7 3.3 ...
```

```
##  $ Less.Than.High.School                                 : int  11867
1685 4973 3425 1137 814 7717 204 22590 5985 ...
##  $ High.School.Graduate..includes.equivalency.           : int  15611
4097 4307 4584 2939 1208 13241 580 21922 8056 ...
##  $ Some.college                                          : int  16957
4443 6378 6106 6280 2736 16606 1484 15806 6209 ...
##  $ Bachelor.s.degree                                     : int  4861
6210 15879 3205 24691 6783 4978 6786 4144 998 ...
##  $ Master.s.degree                                       : int  2679
2665 5814 1852 11489 3694 2425 2777 1062 337 ...
##  $ Professional.school.degree                            : int  345 6
80 2247 385 4186 2363 232 1640 138 29 ...
##  $ Doctorate.degree                                      : int  237 5
32 852 235 1814 1681 251 554 141 17 ...
##  $ Enrolled.In.School                                    : int  23837
6816 11271 9781 13369 6490 19891 1863 35984 13102 ...
##  $ Not.Enrolled.In.School                                : int  55632
20505 40804 20033 53311 18456 47686 13850 73975 24458 ...
##  $ In.labor.force.                                       : int  36845
13998 36802 14395 51957 17785 29665 13229 53511 12478 ...
##  $ In.Armed.Forces                                       : int  23 0
0 0 14 0 32 0 7 0 ...
##  $ Civilian.                                             : int  36822
13998 36802 14395 51943 17785 29633 13229 53504 12478 ...
##  $ Employed                                              : int  29326
12852 34289 11066 49823 16893 22035 12227 44060 9757 ...
##  $ Unemployed                                            : int  7496
1146 2513 3329 2120 892 7598 1002 9444 2721 ...
##  $ Not.In.labor.force                                    : int  26161
8295 9197 9208 10339 5618 25097 2065 29853 15500 ...
##  $ Employed.1                                            : int  29326
12852 34289 11066 49823 16893 22035 12227 44060 9757 ...
##  $ Unemployed.1                                          : int  7496
1146 2513 3329 2120 892 7598 1002 9444 2721 ...
##  $ Less.than..10.000                                     : int  3291
696 1428 3479 2487 1210 3457 906 2903 3162 ...
##  $ X.10.000.to..14.999                                   : int  2001
469 1015 1537 1041 574 1875 392 1730 1187 ...
##  $ X.15.000.to..19.999                                   : int  2034
557 723 1092 934 158 1662 184 2265 1171 ...
##  $ X.20.000.to..24.999                                   : int  2156
419 1052 813 968 297 1608 290 2063 836 ...
##  $ X.25.000.to..29.999                                   : int  1555
523 775 529 1258 576 1398 52 2044 783 ...
##  $ X.30.000.to..34.999                                   : int  1659
418 747 646 939 440 1350 260 2112 775 ...
##  $ X.35.000.to..39.999                                   : int  1418
301 567 672 1180 270 1265 355 2193 450 ...
##  $ X.40.000.to..44.999                                   : int  1639
410 1170 516 1653 449 1055 213 1903 623 ...
##  $ X.45.000.to..49.999                                   : int  1291
311 749 366 1296 285 1114 250 1552 372 ...
##  $ X.50.000.to..59.999                                   : int  1842
518 1691 757 2611 767 2119 556 2544 677 ...
```

```
##  $ X.60.000.to..74.999                                          : int  2475
926 2327 834 3312 1141 2322 908 2638 736 ...
##  $ X.75.000.to..99.999                                          : int  2746
1116 2833 797 4264 1600 2241 1479 3668 831 ...
##  $ X.100.000.to..124.999                                        : int  1558
982 2085 798 3860 1504 1176 1163 1609 181 ...
##  $ X.125.000.to..149.999                                        : int  814 8
65 1575 209 2396 945 537 740 856 104 ...
##  $ X.150.000.to..199.999                                        : int  798 9
45 1941 460 2886 1156 304 1028 470 90 ...
##  $ X.200.000.or.More                                            : int  305 1
378 2623 433 5593 2087 166 2219 142 29 ...
##  $ Median.household.income..In.2014.Inflation.Adjusted.Dollars. : int  38825
72086 71019 25343 78796 82777 37038 92106 40095 22835 ...
##  $ Average.household.income..In.2014.Inflation.Adjusted.Dollars.: int  51014
100940 99320 47453 116146 115185 46308 138233 49797 32885 ...
##  $ With.earnings                                                : int  19396
7750 20856 9057 32636 11675 15566 9851 24975 7189 ...
##  $ No.earnings                                                  : int  8186
3084 2445 4881 4042 1784 8083 1144 5717 4818 ...
##  $ Per.capita.income..In.2014.Inflation.adjusted.dollars.       : int  18183
39861 44186 22162 63402 60988 17007 94773 14295 10888 ...
##  $ Living.in.Poverty                                            : int  10778
687 5109 5888 6705 2310 11633 1364 15453 9661 ...
##  $ At.or.Above.Poverty.Level                                    : int  38134
15201 36705 12438 50360 16280 29461 13319 55465 13115 ...
##  $ Active_flag                                                  : Factor w/
2 levels "0","1": 2 2 1 2 2 2 2 2 2 2 ...
##  $ start.2010...c.2..60602                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60603                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60604                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60605                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 2 1 1 1 1 ...
##  $ start.2010...c.2..60606                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60607                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60608                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60609                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60610                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60611                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60612                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60613                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60614                                      : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
##  $ start.2010...c.2..60615                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60616                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60617                                    : Factor w/
2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60618                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60619                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60620                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60621                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60622                                    : Factor w/
2 levels "0","1": 1 1 2 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60623                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60624                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...
##  $ start.2010...c.2..60625                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60626                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60628                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 2 1 1 1 ...
##  $ start.2010...c.2..60629                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 2 1 ...
##  $ start.2010...c.2..60630                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60631                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60632                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60633                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ start.2010...c.2..60634                                    : Factor w/
2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##   [list output truncated]
```

```
 model <- glm(Active_flag~.,family=binomial(link='logit'),data=train[,c(1,8,9,1
0,12,13,19,20,21,27,28,36,40,42,60,61,64,67)])

summary(model)
```

```
##
## Call:
## glm(formula = Active_flag ~ ., family = binomial(link = "logit"),
##     data = train[, c(1, 8, 9, 10, 12, 13, 19, 20, 21, 27, 28,
##         36, 40, 42, 60, 61, 64, 67)])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0393  -1.2373   0.7832   0.9520   1.2887
##
## Coefficients: (16 not defined because of singularities)
##                                                      Estimate
## (Intercept)                                           1.18270
## zipcode60602                                          0.20360
## zipcode60603                                         -0.26640
## zipcode60604                                         -0.18417
## zipcode60605                                          0.26422
## zipcode60606                                          0.28364
## zipcode60607                                          0.18621
## zipcode60608                                         -0.54404
## zipcode60609                                         -0.62641
## zipcode60610                                         -0.10178
## zipcode60611                                          0.59859
## zipcode60612                                         -1.44052
## zipcode60613                                         -0.31770
## zipcode60614                                         -0.26264
## zipcode60615                                         -0.15819
## zipcode60616                                         -0.48955
## zipcode60617                                         -0.62675
## zipcode60618                                         -0.44429
## zipcode60619                                         -1.41563
## zipcode60620                                         -1.09861
## zipcode60621                                         -1.12385
## zipcode60622                                         -0.21988
## zipcode60623                                         -0.97335
## zipcode60624                                         -1.18270
## zipcode60625                                         -0.49806
## zipcode60626                                         -0.62308
## zipcode60628                                         -1.06937
## zipcode60629                                         -1.12713
## zipcode60630                                          0.13298
## zipcode60631                                         -0.23823
## zipcode60632                                         -0.63718
## zipcode60633                                          0.07007
## zipcode60634                                         -0.31770
## zipcode60636                                         -1.20290
## zipcode60637                                         -1.14014
## zipcode60638                                          0.32138
## zipcode60639                                         -0.82486
## zipcode60640                                         -0.35437
## zipcode60641                                         -0.66648
## zipcode60642                                          0.24134
## zipcode60643                                         -0.91241
```

```
## zipcode60644                                                    -1.11952
## zipcode60645                                                    -0.90506
## zipcode60646                                                    -0.72316
## zipcode60647                                                    -0.57532
## zipcode60649                                                    -1.15800
## zipcode60651                                                    -1.04293
## zipcode60652                                                    -0.81497
## zipcode60653                                                    -1.18270
## zipcode60654                                                     0.37065
## zipcode60655                                                    -0.08408
## zipcode60656                                                     0.76321
## zipcode60657                                                    -0.40042
## zipcode60659                                                    -0.77723
## zipcode60660                                                    -0.69434
## zipcode60661                                                    -0.23823
## X5.to.9.Years                                                         NA
## X10.to.14.Years                                                       NA
## X15.to.17.Years                                                       NA
## X25.to.34.Years                                                       NA
## X35.to.44.Years                                                       NA
## White.Alone                                                           NA
## Black.or.African.American.Alone                                       NA
## American.Indian.and.Alaska.Native.Alone                               NA
## Less.Than.High.School                                                 NA
## High.School.Graduate..includes.equivalency.                           NA
## In.labor.force.                                                       NA
## Unemployed                                                            NA
## Employed.1                                                            NA
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.          NA
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.         NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.                NA
##                                                                 Std. Error
## (Intercept)                                                        0.26222
## zipcode60602                                                       0.52640
## zipcode60603                                                       0.41080
## zipcode60604                                                       0.51405
## zipcode60605                                                       0.41437
## zipcode60606                                                       0.38834
## zipcode60607                                                       0.33469
## zipcode60608                                                       0.30317
## zipcode60609                                                       0.32458
## zipcode60610                                                       0.37316
## zipcode60611                                                       0.37655
## zipcode60612                                                       0.32175
## zipcode60613                                                       0.35773
## zipcode60614                                                       0.30846
## zipcode60615                                                       0.40722
## zipcode60616                                                       0.32457
## zipcode60617                                                       0.33031
## zipcode60618                                                       0.31005
## zipcode60619                                                       0.31327
## zipcode60620                                                       0.32005
## zipcode60621                                                       0.35726
## zipcode60622                                                       0.31884
```

```
## zipcode60623                                                    0.30589
## zipcode60624                                                    0.33796
## zipcode60625                                                    0.30701
## zipcode60626                                                    0.33528
## zipcode60628                                                    0.32652
## zipcode60629                                                    0.31074
## zipcode60630                                                    0.42902
## zipcode60631                                                    0.51689
## zipcode60632                                                    0.31580
## zipcode60633                                                    0.84357
## zipcode60634                                                    0.35773
## zipcode60636                                                    0.33041
## zipcode60637                                                    0.39231
## zipcode60638                                                    0.47068
## zipcode60639                                                    0.29752
## zipcode60640                                                    0.32645
## zipcode60641                                                    0.33454
## zipcode60642                                                    0.40522
## zipcode60643                                                    0.35994
## zipcode60644                                                    0.33303
## zipcode60645                                                    0.37291
## zipcode60646                                                    0.45247
## zipcode60647                                                    0.29987
## zipcode60649                                                    0.34373
## zipcode60651                                                    0.31610
## zipcode60652                                                    0.40346
## zipcode60653                                                    0.43175
## zipcode60654                                                    0.35190
## zipcode60655                                                    0.57916
## zipcode60656                                                    0.80012
## zipcode60657                                                    0.30571
## zipcode60659                                                    0.31751
## zipcode60660                                                    0.35850
## zipcode60661                                                    0.36728
## X5.to.9.Years                                                       NA
## X10.to.14.Years                                                     NA
## X15.to.17.Years                                                     NA
## X25.to.34.Years                                                     NA
## X35.to.44.Years                                                     NA
## White.Alone                                                         NA
## Black.or.African.American.Alone                                     NA
## American.Indian.and.Alaska.Native.Alone                             NA
## Less.Than.High.School                                               NA
## High.School.Graduate..includes.equivalency.                         NA
## In.labor.force.                                                     NA
## Unemployed                                                          NA
## Employed.1                                                          NA
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.        NA
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.       NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.              NA
##                                                                z value
## (Intercept)                                                      4.510
## zipcode60602                                                     0.387
## zipcode60603                                                    -0.648
```

```
## zipcode60604                                                    -0.358
## zipcode60605                                                     0.638
## zipcode60606                                                     0.730
## zipcode60607                                                     0.556
## zipcode60608                                                    -1.794
## zipcode60609                                                    -1.930
## zipcode60610                                                    -0.273
## zipcode60611                                                     1.590
## zipcode60612                                                    -4.477
## zipcode60613                                                    -0.888
## zipcode60614                                                    -0.851
## zipcode60615                                                    -0.388
## zipcode60616                                                    -1.508
## zipcode60617                                                    -1.897
## zipcode60618                                                    -1.433
## zipcode60619                                                    -4.519
## zipcode60620                                                    -3.433
## zipcode60621                                                    -3.146
## zipcode60622                                                    -0.690
## zipcode60623                                                    -3.182
## zipcode60624                                                    -3.500
## zipcode60625                                                    -1.622
## zipcode60626                                                    -1.858
## zipcode60628                                                    -3.275
## zipcode60629                                                    -3.627
## zipcode60630                                                     0.310
## zipcode60631                                                    -0.461
## zipcode60632                                                    -2.018
## zipcode60633                                                     0.083
## zipcode60634                                                    -0.888
## zipcode60636                                                    -3.641
## zipcode60637                                                    -2.906
## zipcode60638                                                     0.683
## zipcode60639                                                    -2.772
## zipcode60640                                                    -1.086
## zipcode60641                                                    -1.992
## zipcode60642                                                     0.596
## zipcode60643                                                    -2.535
## zipcode60644                                                    -3.362
## zipcode60645                                                    -2.427
## zipcode60646                                                    -1.598
## zipcode60647                                                    -1.919
## zipcode60649                                                    -3.369
## zipcode60651                                                    -3.299
## zipcode60652                                                    -2.020
## zipcode60653                                                    -2.739
## zipcode60654                                                     1.053
## zipcode60655                                                    -0.145
## zipcode60656                                                     0.954
## zipcode60657                                                    -1.310
## zipcode60659                                                    -2.448
## zipcode60660                                                    -1.937
## zipcode60661                                                    -0.649
## X5.to.9.Years                                                       NA
```

```
## X10.to.14.Years                                                    NA
## X15.to.17.Years                                                    NA
## X25.to.34.Years                                                    NA
## X35.to.44.Years                                                    NA
## White.Alone                                                        NA
## Black.or.African.American.Alone                                    NA
## American.Indian.and.Alaska.Native.Alone                            NA
## Less.Than.High.School                                              NA
## High.School.Graduate..includes.equivalency.                        NA
## In.labor.force.                                                    NA
## Unemployed                                                         NA
## Employed.1                                                         NA
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.       NA
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.      NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.             NA
##                                                                 Pr(>|z|)
## (Intercept)                                                  0.00000647
## zipcode60602                                                 0.698920
## zipcode60603                                                 0.516665
## zipcode60604                                                 0.720145
## zipcode60605                                                 0.523700
## zipcode60606                                                 0.465154
## zipcode60607                                                 0.577963
## zipcode60608                                                 0.072734
## zipcode60609                                                 0.053616
## zipcode60610                                                 0.785039
## zipcode60611                                                 0.111904
## zipcode60612                                                 0.00000756
## zipcode60613                                                 0.374490
## zipcode60614                                                 0.394519
## zipcode60615                                                 0.697673
## zipcode60616                                                 0.131479
## zipcode60617                                                 0.057771
## zipcode60618                                                 0.151862
## zipcode60619                                                 0.00000622
## zipcode60620                                                 0.000598
## zipcode60621                                                 0.001657
## zipcode60622                                                 0.490425
## zipcode60623                                                 0.001463
## zipcode60624                                                 0.000466
## zipcode60625                                                 0.104744
## zipcode60626                                                 0.063113
## zipcode60628                                                 0.001057
## zipcode60629                                                 0.000286
## zipcode60630                                                 0.756588
## zipcode60631                                                 0.644870
## zipcode60632                                                 0.043627
## zipcode60633                                                 0.933804
## zipcode60634                                                 0.374490
## zipcode60636                                                 0.000272
## zipcode60637                                                 0.003658
## zipcode60638                                                 0.494730
## zipcode60639                                                 0.005564
## zipcode60640                                                 0.277686
```

```
## zipcode60641                                                    0.046346
## zipcode60642                                                    0.551457
## zipcode60643                                                    0.011249
## zipcode60644                                                    0.000775
## zipcode60645                                                    0.015224
## zipcode60646                                                    0.109983
## zipcode60647                                                    0.055041
## zipcode60649                                                    0.000755
## zipcode60651                                                    0.000969
## zipcode60652                                                    0.043387
## zipcode60653                                                    0.006157
## zipcode60654                                                    0.292202
## zipcode60655                                                    0.884568
## zipcode60656                                                    0.340144
## zipcode60657                                                    0.190269
## zipcode60659                                                    0.014369
## zipcode60660                                                    0.052772
## zipcode60661                                                    0.516574
## X5.to.9.Years                                                         NA
## X10.to.14.Years                                                       NA
## X15.to.17.Years                                                       NA
## X25.to.34.Years                                                       NA
## X35.to.44.Years                                                       NA
## White.Alone                                                           NA
## Black.or.African.American.Alone                                       NA
## American.Indian.and.Alaska.Native.Alone                               NA
## Less.Than.High.School                                                 NA
## High.School.Graduate..includes.equivalency.                           NA
## In.labor.force.                                                       NA
## Unemployed                                                            NA
## Employed.1                                                            NA
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.          NA
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.         NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.                NA
##
## (Intercept)                                                    ***
## zipcode60602
## zipcode60603
## zipcode60604
## zipcode60605
## zipcode60606
## zipcode60607
## zipcode60608                                                   .
## zipcode60609                                                   .
## zipcode60610
## zipcode60611
## zipcode60612                                                   ***
## zipcode60613
## zipcode60614
## zipcode60615
## zipcode60616
## zipcode60617                                                   .
## zipcode60618
## zipcode60619                                                   ***
```

```
## zipcode60620                                                    ***
## zipcode60621                                                    **
## zipcode60622
## zipcode60623                                                    **
## zipcode60624                                                    ***
## zipcode60625
## zipcode60626                                                    .
## zipcode60628                                                    **
## zipcode60629                                                    ***
## zipcode60630
## zipcode60631
## zipcode60632                                                    *
## zipcode60633
## zipcode60634
## zipcode60636                                                    ***
## zipcode60637                                                    **
## zipcode60638
## zipcode60639                                                    **
## zipcode60640
## zipcode60641                                                    *
## zipcode60642
## zipcode60643                                                    *
## zipcode60644                                                    ***
## zipcode60645                                                    *
## zipcode60646
## zipcode60647                                                    .
## zipcode60649                                                    ***
## zipcode60651                                                    ***
## zipcode60652                                                    *
## zipcode60653                                                    **
## zipcode60654
## zipcode60655
## zipcode60656
## zipcode60657
## zipcode60659                                                    *
## zipcode60660                                                    .
## zipcode60661
## X5.to.9.Years
## X10.to.14.Years
## X15.to.17.Years
## X25.to.34.Years
## X35.to.44.Years
## White.Alone
## Black.or.African.American.Alone
## American.Indian.and.Alaska.Native.Alone
## Less.Than.High.School
## High.School.Graduate..includes.equivalency.
## In.labor.force.
## Unemployed
## Employed.1
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.
## Per.capita.income..In.2014.Inflation.adjusted.dollars.
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7067.1  on 5421  degrees of freedom
## Residual deviance: 6795.9  on 5366  degrees of freedom
## AIC: 6907.9
##
## Number of Fisher Scoring iterations: 4
```