Univariate Analysis

Venukanan Subenthiran

2022-11-09

Understanding the Datafram as a Whole

Provides baseline for creating a more details data dictionary.

```
str(fraudTotal.db)
```

```
## 'data.frame':    1852394 obs. of  23 variables:
##  $ X                    : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ trans_date_trans_time: POSIXct, format: "2019-01-01 00:00:18" "2019-01-01 00:00:44" ...
##  $ cc_num               : num  2.70e+15 6.30e+11 3.89e+13 3.53e+15 3.76e+14 ...
##  $ merchant             : Factor w/ 693 levels "fraud_Abbott-Rogahn",..: 516 244 390 364 298
608 534 108 251 565 ...
##  $ category             : Factor w/ 14 levels "entertainment",..: 9 5 1 3 10 3 4 3 10 5 ...
##  $ amt                  : num  4.97 107.23 220.11 45 41.96 ...
##  $ first                : Factor w/ 355 levels "Aaron","Adam",..: 165 313 117 166 340 165 202
315 147 242 ...
##  $ last                 : Factor w/ 486 levels "Abbott","Adams",..: 19 162 387 469 154 85 365
473 73 4 ...
##  $ gender               : Factor w/ 2 levels "F","M": 1 1 2 2 2 1 1 2 1 1 ...
##  $ street               : Factor w/ 999 levels "000 Jennifer Mills",..: 577 440 611 946 423 4
78 897 229 697 218 ...
##  $ city                 : Factor w/ 906 levels "Achille","Acworth",..: 533 620 475 85 218 225
355 238 481 150 ...
##  $ state                : Factor w/ 51 levels "AK","AL","AR",..: 28 48 14 27 46 39 17 46 39 4
3 ...
##  $ zip                  : int  28654 99160 83252 59632 24433 18917 67851 22824 15665 37040
...
##  $ lat                  : num  36.1 48.9 42.2 46.2 38.4 ...
##  $ long                 : num  -81.2 -118.2 -112.3 -112.1 -79.5 ...
##  $ city_pop             : int  3495 149 4154 1939 99 2158 2691 6018 1472 151785 ...
##  $ job                  : Factor w/ 497 levels "Academic librarian",..: 373 432 309 331 117 4
83 30 128 378 332 ...
##  $ dob                  : Date, format: "1988-03-09" "1978-06-21" ...
##  $ trans_num            : Factor w/ 1852394 levels "00000ecad06b03d3a8d34b4e30b5ce3b",..: 803
27 227463 1169031 777910 1186867 177885 954104 789719 1824660 430621 ...
##  $ unix_time            : int  1325376018 1325376044 1325376051 1325376076 1325376186 1325376
248 1325376282 1325376308 1325376318 1325376361 ...
##  $ merch_lat            : num  36 49.2 43.2 47 38.7 ...
##  $ merch_long           : num  -82 -118.2 -112.2 -112.6 -78.6 ...
##  $ is_fraud             : int  0 0 0 0 0 0 0 0 0 0 ...
```

Unvirariate Analysis of trans_date_trans_time

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$trans_date_trans_time))
```

```
## [1] 0
```

Converting Character to DateTime class

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

```
## Loading required package: timechange
```

```
## 
## Attaching package: 'lubridate'
```

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

```
fraudTotal.db$trans_date_trans_time <- ymd_hms(fraudTotal.db$trans_date_trans_time)
```

Summary of trans_date_trans_time column

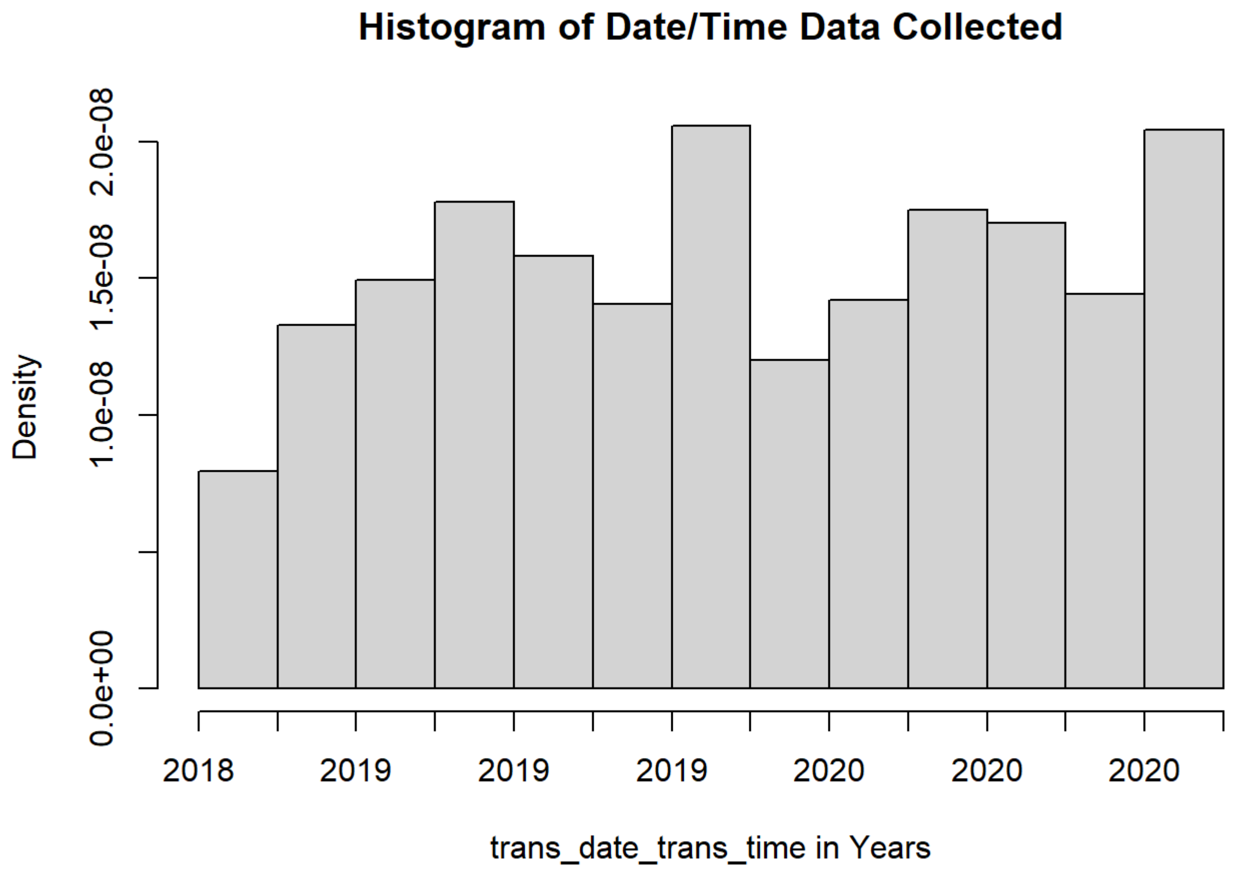
```
summary(fraudTotal.db$trans_date_trans_time)
```

```
##                        Min.                    1st Qu. 
## "2019-01-01 00:00:18.0000" "2019-07-23 04:13:43.7500" 
##                      Median                       Mean 
## "2020-01-02 01:15:31.0000" "2020-01-20 21:31:46.8018" 
##                     3rd Qu.                       Max. 
## "2020-07-23 12:11:25.2500" "2020-12-31 23:59:34.0000"
```

Histogram of trans_date_trans_time variable

```
hist(fraudTotal.db$trans_date_trans_time, breaks = 10 , main = "Histogram of Date/Time Data Coll
ected", xlab = "trans_date_trans_time in Years")
```

```
## Warning in breaks[-1L] + breaks[-nB]: NAs produced by integer overflow
```



Univariate Analysis of cc_num

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$cc_num))
```

```
## [1] 0
```

Summary of cc_num column

```
options(scipen = 999)
summary(fraudTotal.db$cc_num)
```

```
##                Min.             1st Qu.              Median                Mean 
##         60416207185     180042946491150    3521417320836166  417386038393710464 
##             3rd Qu.                Max. 
##    4642255475285942 4992346398065154048
```

Find the Standard Deviation and Variance of cc_num variable

```
sd(fraudTotal.db$cc_num)
```

```
## [1] 1309115265318735104
```

```
var(fraudTotal.db$cc_num)
```

```
## [1] 171378277789054226544482860488642682
```

Frequency of cc_num values

```
table_cc_num <- table(fraudTotal.db$cc_num)
head(table_cc_num)
```

```
## 
## 60416207185 60422928733 60423098130 60427851591 60487002085 60490596305 
##        2196        2200         738         743         735        1465
```

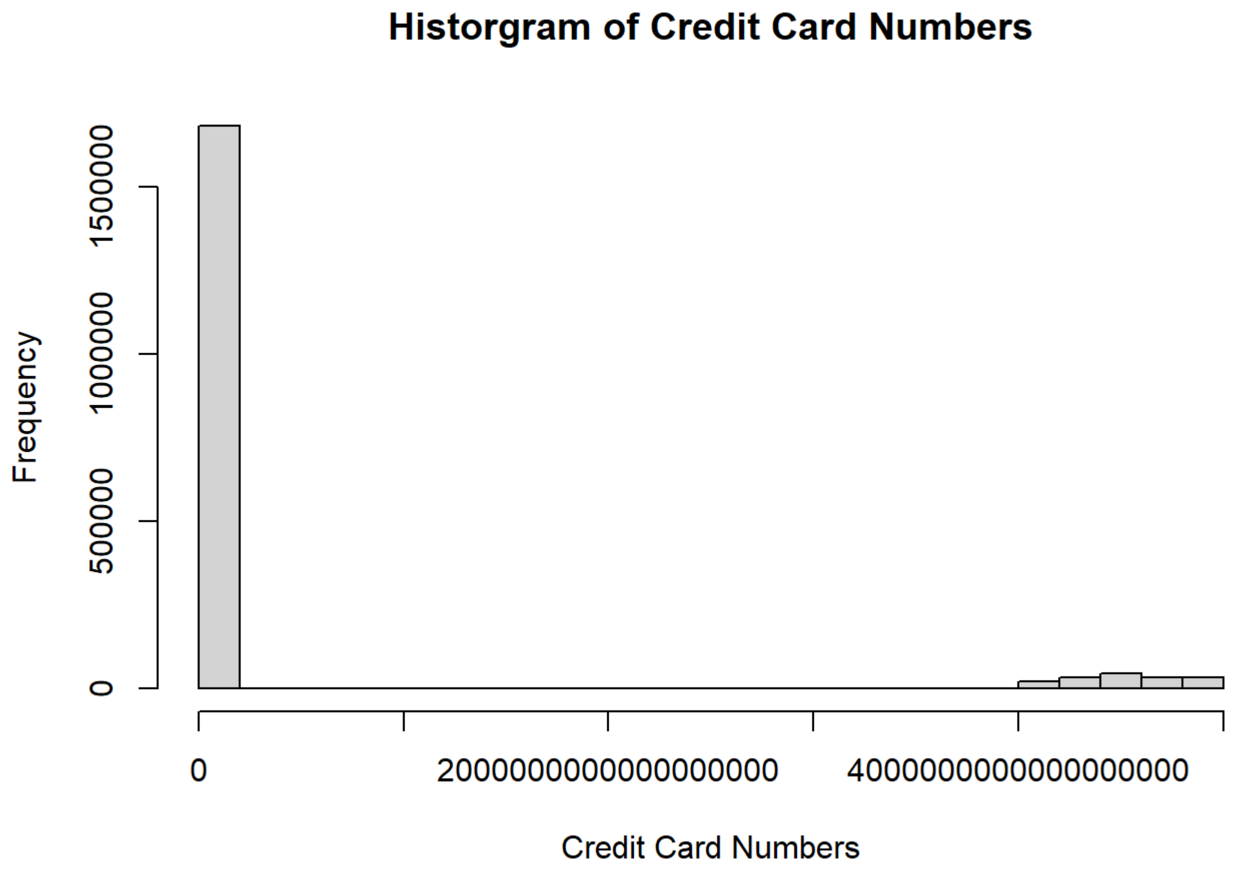
Unique Values of cc_num

```
head(unique(fraudTotal.db$cc_num))
```

```
## [1] 2703186189652095     630423337322   38859492057661 3534093764340240
## [5]  375534208663984 4767265376804500
```

Histogram of cc_num

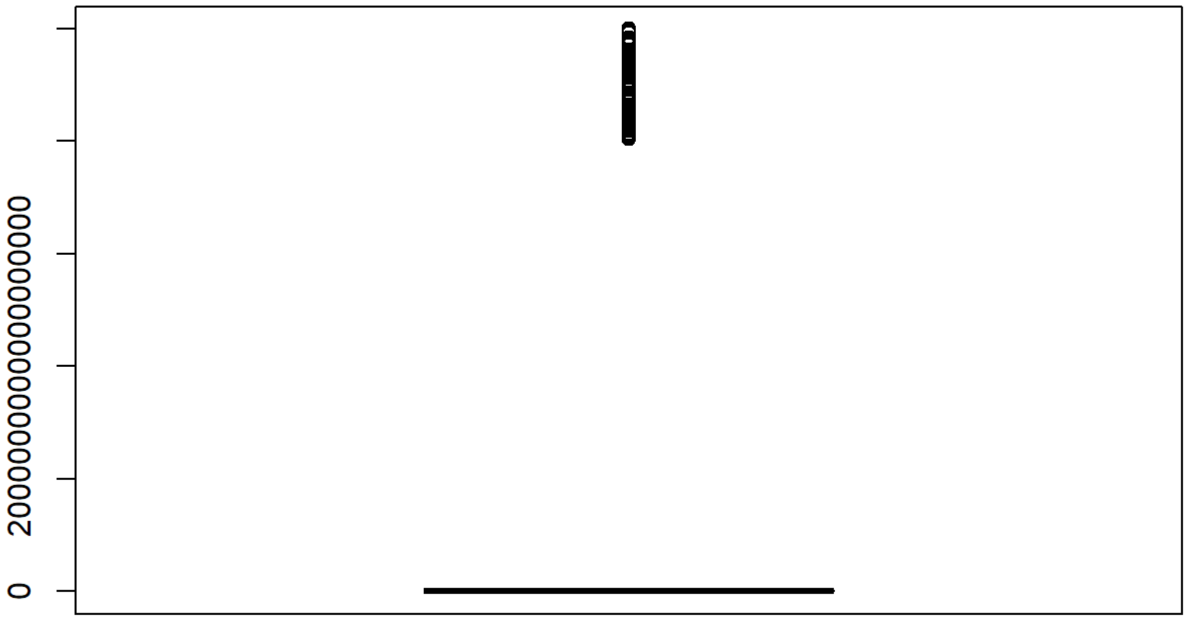
```
hist(fraudTotal.db$cc_num, main = "Historgram of Credit Card Numbers", xlab = "Credit Card Numbe
rs")
```



###

Boxplot of cc_num variable

```
boxplot(fraudTotal.db$cc_num)
```



Univariate Analysis of merchant

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$merchant))
```

```
## [1] 0
```

Summary of merchant column

```
summary(fraudTotal.db$merchant)
```

```
##                           fraud_Kilback LLC
##                                        6262
##                           fraud_Cormier LLC
##                                        5246
##                            fraud_Schumm PLC
##                                        5195
##                              fraud_Kuhn LLC
##                                        5031
##                             fraud_Boyer PLC
##                                        4999
##                         fraud_Dickinson Ltd
##                                        4953
##                             fraud_Emard Inc
##                                        3867
##                       fraud_Cummerata-Jones
##                                        3860
##                        fraud_Corwin-Collins
##                                        3853
##                       fraud_Rodriguez Group
##                                        3843
##                             fraud_Kling Inc
##                                        3841
##                      fraud_Erdman-Kertzmann
##                                        3839
##                     fraud_Parisian and Sons
##                                        3839
##                            fraud_Huels-Hahn
##                                        3835
##            fraud_Stroman, Hudson and Erdman
##                                        3829
##                             fraud_Kutch LLC
##                                        3828
##            fraud_Jenkins, Hauck and Friesen
##                                        3817
##                       fraud_Prohaska-Murray
##                                        3809
##                fraud_Olson, Becker and Koch
##                                        3806
##         fraud_Eichmann, Bogan and Rodriguez
##                                        3798
## fraud_Christiansen, Goyette and Schamberger
##                                        3794
##         fraud_Greenholt, Jacobi and Gleason
##                                        3794
##                     fraud_Bartoletti-Wunsch
##                                        3793
##        fraud_Connelly, Reichert and Fritsch
##                                        3788
##                           fraud_Mraz-Herzog
##                                        3788
##                             fraud_Berge LLC
##                                        3786
##              fraud_Streich, Hansen and Veum
##                                        3785
##                             fraud_Bins-Rice
##                                        3784
##                       fraud_Brekke and Sons
##                                        3781
##                         fraud_Friesen-Stamm
##                                        3774
##                          fraud_Torp-Labadie
##                                        3769
##                   fraud_Ledner-Pfannerstill
##                                        3764
##           fraud_Raynor, Reinger and Hagenes
##                                        3763
##                         fraud_Koss and Sons
##                                        3758
##                           fraud_Schmitt Inc
##                                        3747
##        fraud_Tillman, Dickinson and Labadie
##                                        3746
##           fraud_Schaefer, McGlynn and Bosco
##                                        3742
##                          fraud_Bernhard Inc
##                                        3741
##          fraud_Kutch, Hermiston and Farrell
##                                        3725
##                    fraud_Conroy-Cruickshank
##                                        3722
##                          fraud_Cummings LLC
##                                        3721
##                fraud_Zieme, Bode and Dooley
##                                        3720
##                          fraud_Luettgen PLC
##                                        3719
##                            fraud_Sporer Inc
##                                        3719
##                           fraud_Huels-Nolan
##                                        3714
##                fraud_Lind, Huel and McClure
##                                        3714
##         fraud_Robel, Cummerata and Prosacco
##                                        3701
##                            fraud_Harris Inc
##                                        3700
##                           fraud_Kuvalis Ltd
##                                        3700
##               fraud_Reilly, Heaney and Cole
##                                        3698
##              fraud_Raynor, Feest and Miller
##                                        3673
##        fraud_Schaefer, Maggio and Daugherty
##                                        3671
##                      fraud_Pacocha-O'Reilly
##                                        3650
##                        fraud_Heller-Langosh
##                                        3648
##                             fraud_Marks Inc
##                                        3643
##                       fraud_Friesen-D'Amore
##                                        3640
##                            fraud_Harber Inc
##                                        3640
##                     fraud_Hackett-Lueilwitz
##                                        3626
##                      fraud_Eichmann-Kilback
##                                        3616
##         fraud_Denesik, Powlowski and Pouros
##                                        3611
##               fraud_Lockman, West and Runte
##                                        3607
##              fraud_O'Reilly, Mohr and Purdy
##                                        3605
##                        fraud_Murray-Smitham
##                                        3603
##                          fraud_Medhurst Inc
##                                        3600
##                      fraud_Goodwin-Nitzsche
##                                        3598
##                          fraud_Bauch-Raynor
##                                        3597
##                    fraud_Altenwerth-Kilback
##                                        3594
##          fraud_Schiller, Blanda and Johnson
##                                        3585
##                         fraud_Gulgowski LLC
##                                        3584
##                             fraud_Terry Ltd
##                                        3583
##           fraud_Schoen, Kuphal and Nitzsche
##                                        3581
##           fraud_Goldner, Kovacek and Abbott
##                                        3580
##                           fraud_Lockman Ltd
##                                        3580
##          fraud_O'Connell, Botsford and Hand
##                                        3578
##                     fraud_Botsford and Sons
##                                        3576
##                        fraud_Kiehn-Emmerich
##                                        3574
##                            fraud_Renner Ltd
##                                        3570
##                        fraud_White and Sons
##                                        3570
##                               fraud_Cole PLC 
##                                         3562 
##                        fraud_Kutch-Wilderman 
##                                         3562 
##                        fraud_Quitzon-Goyette 
##                                         3562 
##           fraud_Osinski, Ledner and Leuschke 
##                                         3559 
##             fraud_Schumm, Bauch and Ondricka 
##                                         3559 
##                        fraud_Deckow-O'Conner 
##                                         3558 
##                            fraud_Pollich LLC 
##                                         3558 
##                         fraud_Gislason Group 
##                                         3556 
##                        fraud_Connelly-Carter 
##                                         3555 
##                           fraud_Hudson-Ratke 
##                                         3555 
##                fraud_Casper, Hand and Zulauf 
##                                         3553 
##               fraud_Huel, Hammes and Witting 
##                                         3553 
##        fraud_Bahringer, Bergnaum and Quitzon 
##                                         3552 
##                            fraud_Bradtke PLC 
##                                         3551 
##                           fraud_Lynch-Wisozk 
##                                         3550 
##                         fraud_Kutch and Sons 
##                                         3547 
##                           fraud_Rau and Sons 
##                                         3546 
##                              fraud_Kunze Inc 
##                                         3535 
##                   fraud_Schamberger-O'Keefe 
##                                         3535 
##                     fraud_Gaylord-Powlowski 
##                                         3534 
##                           fraud_Miller-Hauck 
##                                         3533 
##                                      (Other) 
##                                      1479036
```

Check to see all Unique Values

```
head(unique(fraudTotal.db$merchant))
```

```
## [1] fraud_Rippin, Kub and Mann         fraud_Heller, Gutmann and Zieme
## [3] fraud_Lind-Buckridge               fraud_Kutch, Hermiston and Farrell
## [5] fraud_Keeling-Crist                fraud_Stroman, Hudson and Erdman
## 693 Levels: fraud_Abbott-Rogahn ... fraud_Zulauf LLC
```

```
table_merchant <- table(fraudTotal.db$merchant)
head(table_merchant)
```

```
## 
##              fraud_Abbott-Rogahn             fraud_Abbott-Steuber 
##                             2647                             2529 
##         fraud_Abernathy and Sons                fraud_Abshire PLC 
##                             2513                             2733 
##              fraud_Adams-Barrows fraud_Adams, Kovacek and Kuhlman 
##                             2535                             1354
```

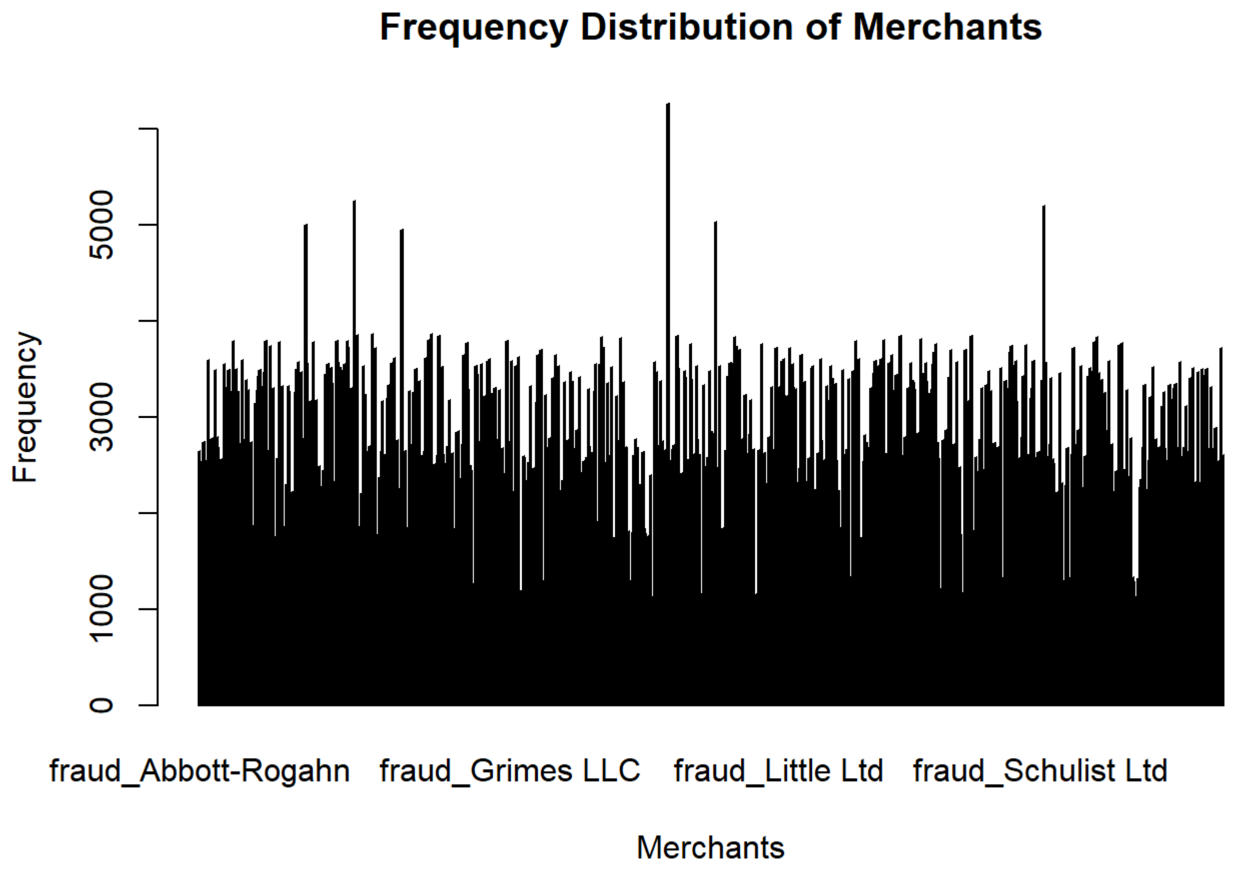
Convert Characater Class to a Factor Class

```
fraudTotal.db$merchant <- as.factor(fraudTotal.db$merchant)
class(fraudTotal.db$merchant)
```

```
## [1] "factor"
```

Frequency Distribution of trans_date_trans_time variable

```
barplot(table(fraudTotal.db$merchant), main = "Frequency Distribution of Merchants", xlab = "Mer
chants", ylab = "Frequency")
```



Univariate Analysis of category

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$category))
```

```
## [1] 0
```

Summary of category Column

```
summary(fraudTotal.db$category)
```

```
##   entertainment     food_dining  gas_transport    grocery_net    grocery_pos
##          134118          130729         188029          64878         176191
## health_fitness            home      kids_pets       misc_net       misc_pos
##          122553          175460         161727          90654         114229
##   personal_care    shopping_net   shopping_pos         travel
##          130085          139322         166463          57956
```

```
class(fraudTotal.db$category)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$category <- as.factor(fraudTotal.db$category)
```

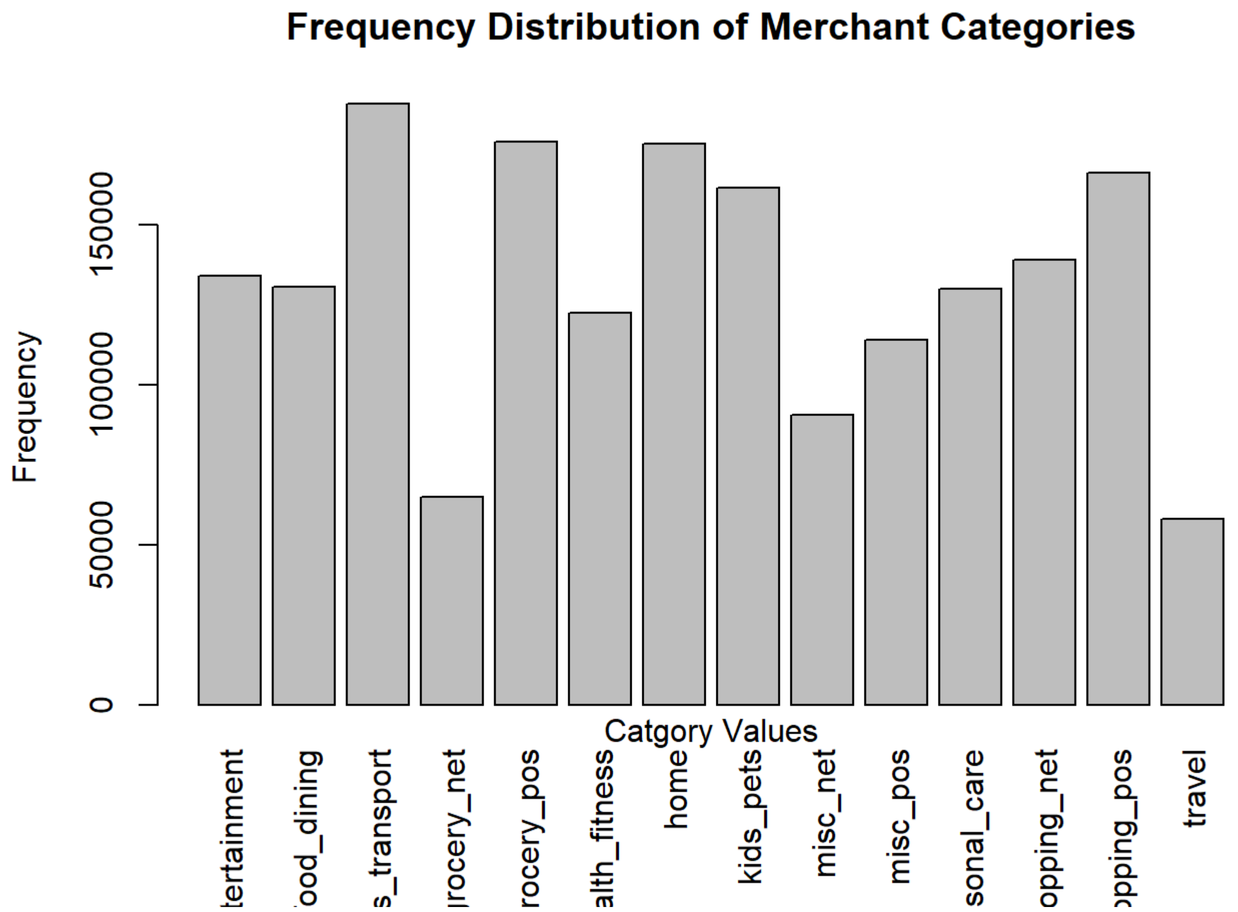
Frequency of category values

```
table(fraudTotal.db$category)
```

```
## 
##  entertainment    food_dining  gas_transport    grocery_net    grocery_pos 
##         134118         130729         188029          64878         176191 
## health_fitness           home      kids_pets       misc_net       misc_pos 
##         122553         175460         161727          90654         114229 
##  personal_care   shopping_net   shopping_pos         travel 
##         130085         139322         166463          57956
```

Frequency Distribution of category

```
barplot(table(fraudTotal.db$category), las = 3, main = "Frequency Distribution of Merchant Categ
ories", xlab = "", ylab = "Frequency")
mtext("Catgory Values", side = 1)
```



Univariate Analysis of amt

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$amt))
```

```
## [1] 0
```

Summary of amt Column

```
summary(fraudTotal.db$amt)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##     1.00     9.64    47.45    70.06    83.10 28948.90
```

```
class(fraudTotal.db$amt)
```

```
## [1] "numeric"
```

Find the Standard Deviation and Variance of amt Column

```
sd(fraudTotal.db$amt)
```

```
## [1] 159.254
```

```
var(fraudTotal.db$amt)
```

```
## [1] 25361.83
```

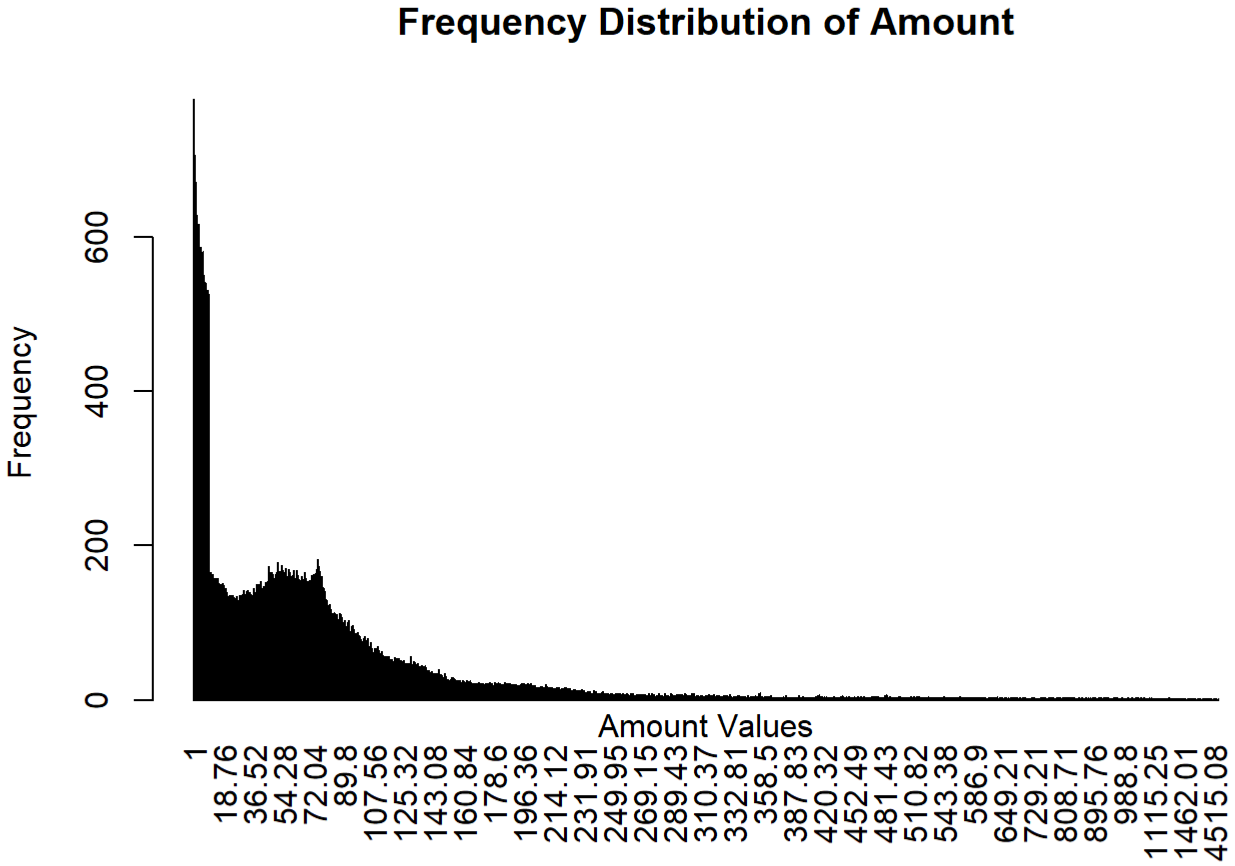
Frequency of amt Column

```
table_amt <- table(fraudTotal.db$amt)
head(table_amt)
```

```
## 
##    1 1.01 1.02 1.03 1.04 1.05 
##  332  735  736  726  744  721
```

Frequency Distribution of amt Column

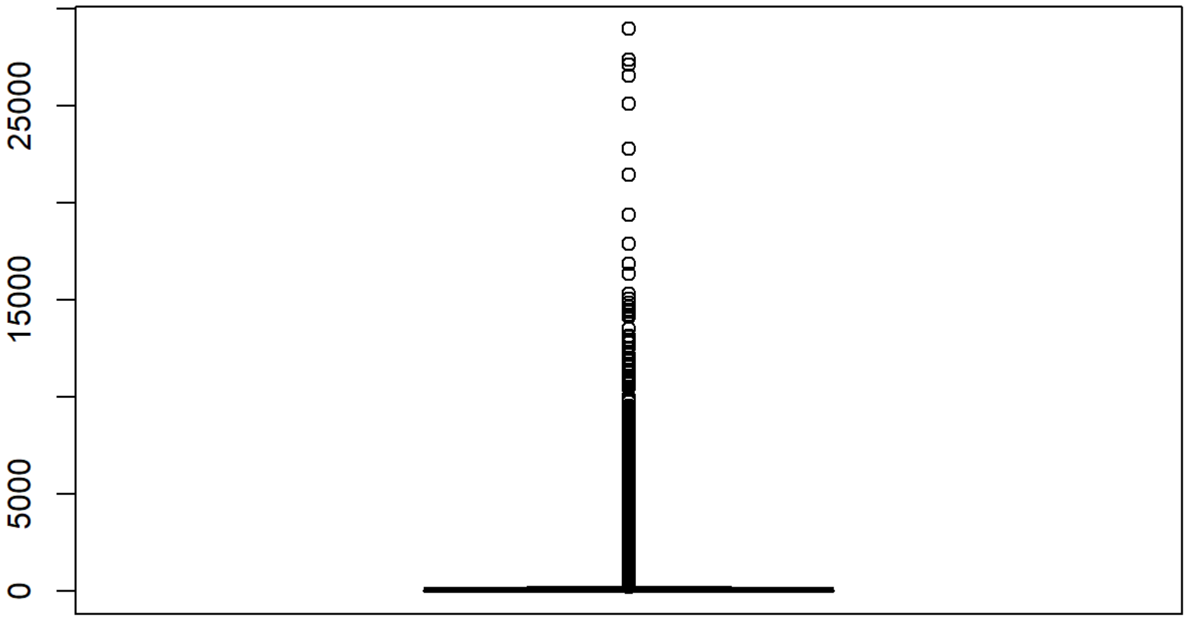
```
barplot(table(fraudTotal.db$amt), las = 3, main = "Frequency Distribution of Amount", xlab = "",
ylab = "Frequency")
mtext("Amount Values", side = 1)
```



```
##
```

Boxplot of amt Column

```
boxplot(fraudTotal.db$amt)
```



```
# does not look good.
# displying the number of outliers existing with the data??
```

Univariate Analysis of first

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$first))
```

```
## [1] 0
```

Summary of first Column

```
summary(fraudTotal.db$first)
```

```
## Christopher      Robert     Jessica       David     Michael       James
##       38112       30743       29236       28564       28539       28496
##    Jennifer        John        Mary     William    Margaret      Joseph
##       24181       23445       23424       23396       21886       21187
##        Lisa      Daniel      Amanda      Ashley     Jeffrey    Michelle
##       19782       19750       19062       19001       18309       18263
##      Samuel    Kimberly      Steven     Kenneth   Stephanie     Melissa
##       17542       16808       16807       16800       15365       14651
##       Susan      Lauren        Adam   Christine      Nathan  Jacqueline
##       14623       14593       13916       13912       13894       13192
##       Scott      Angela     Charles       Sarah     Rebecca       Jason
##       13170       13164       13162       13162       13129       12446
##       Linda     Barbara     Matthew      Monica        Mark      Rachel
##       12439       12404       11707       11699       10989       10986
##      Thomas      Justin      Jeremy        Lori    Danielle      Andrew
##       10986       10974       10271       10240       10235       10228
##       Kayla       Karen     Vincent        Dawn        Gina       Tyler
##       10220        9538        9518        9515        9505        9498
##      Sharon       Amber    Benjamin      Alicia      Joshua     Shannon
##        9496        9495        8795        8784        8770        8770
##       Laura       Tammy      Teresa        Sara     Richard       Larry
##        8768        8762        8754        8749        8081        8064
##    Kathleen   Elizabeth     Allison        Gary     Crystal         Ana
##        8045        8039        8035        8034        8031        8021
##        Ryan    Patricia       Jacob       Jamie       Jared       Stacy
##        7340        7332        7320        7309        7307        7307
##     Sabrina       Janet        Juan    Nicholas       Aaron        Alan
##        7306        7290        6605        6597        6589        6589
##     Gregory     Theresa       Megan        Jodi   Mackenzie       Donna
##        6588        6587        6583        6581        6574        6570
##    Kristina        Tara     Patrick        Kyle       Kevin       Bryan
##        6570        6559        5879        5874        5868        5867
##       Brian     Brianna       Maria     (Other)
##        5865        5863        5860      633658
```

```
class(fraudTotal.db$first)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$first <- as.factor(fraudTotal.db$first)
```

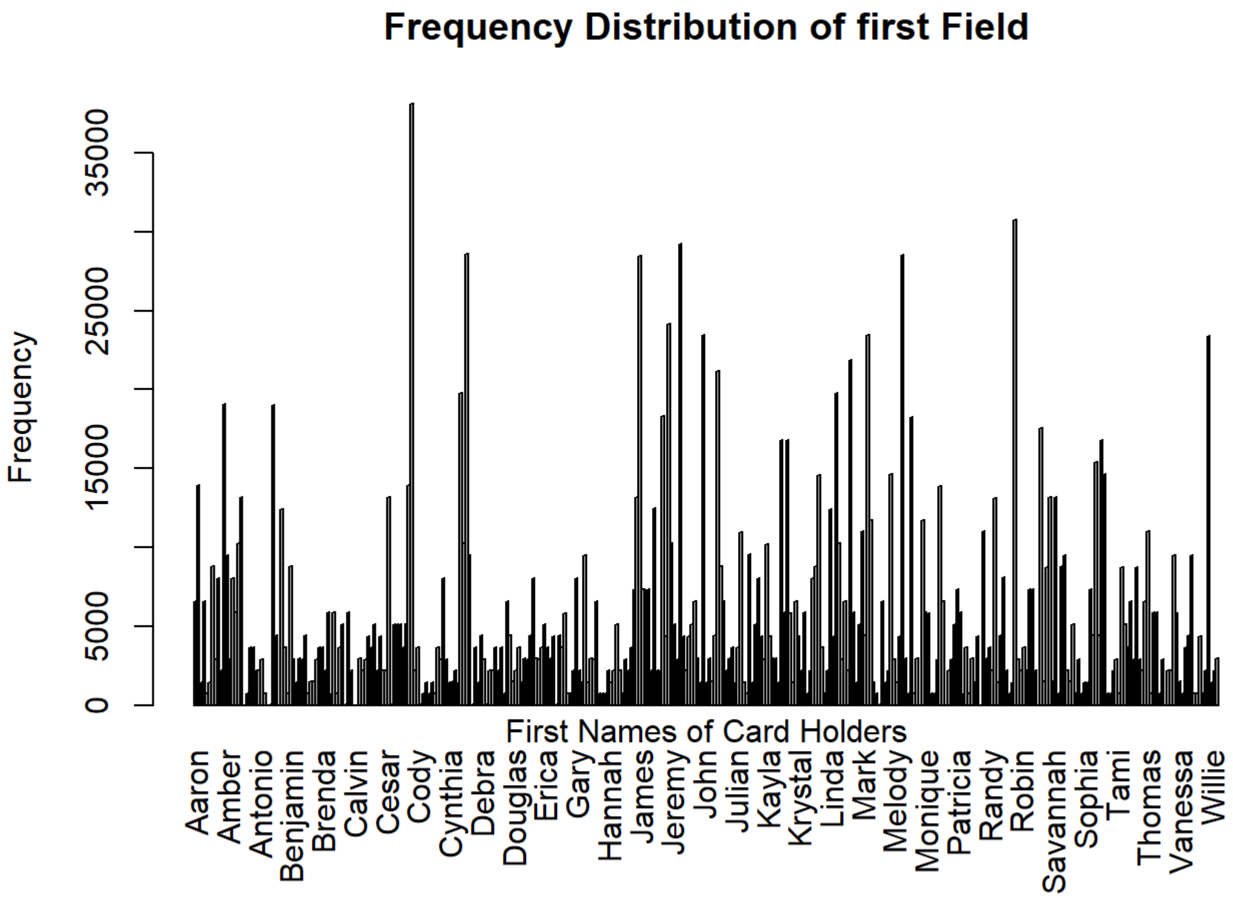
Frequency of first

```
table_first <- table(fraudTotal.db$first)
head(table_first)
```

```
##
##    Aaron     Adam Adriana     Alan     Alex    Alice
##     6589    13916     1465     6589      741     1468
```

Frequency Distribution of first

```
barplot(table(fraudTotal.db$first), las = 3, main = "Frequency Distribution of first Field", xla
b = "", ylab = "Frequency")
mtext("First Names of Card Holders", side = 1)
```



Univariate Analysis of last

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$last))
```

```
## [1] 0
```

Summary of last Column

```
summary(fraudTotal.db$last)
```

```
##      Smith   Williams      Davis    Johnson  Rodriguez   Martinez      Jones 
##      40940      33661      31434      28590      24879      21246      19825 
##      Lewis     Miller   Gonzalez     Martin       Lowe       Bell      Perez 
##      18293      16821      16809      16065      16056      15353      13881 
##     Garcia   Robinson     Bishop     Thomas      Clark    Mendoza      Allen 
##      13221      13188      13173      12479      12428      12426      11744 
##     Foster     Taylor   Anderson      Gomez     Tucker    Sanders      Brown 
##      11712      11708      11702      11700      11679      11665      11005 
##  Patterson      White    Sanchez     Harris    Lambert     Mendez  Hernandez 
##      10962      10268      10255      10225      10213      10198       9533 
##   Campbell     Flores     Fuller    Jenkins   Johnston   Thompson    Roberts 
##       9520       9515       9495       9494       9483       8787       8783 
##      Myers    Walters     Murphy Washington     Moreno    Ramirez   Richards 
##       8780       8774       8770       8770       8767       8758       8051 
##     Torres     Murray     Powell      Lopez      Johns    Spencer      Evans 
##       8036       8031       7333       7325       7316       7315       7313 
##     Briggs     Brooks     Howard     Hughes      Payne     Fisher       Wood 
##       7303       7300       6588       6583       6582       6578       6578 
##   Mckinney     Gamble     Howell    Whitney     Curtis      Ayala       Cruz 
##       6576       6572       6572       6569       6568       6567       6567 
##    Edwards     Rivera   Stephens     Grimes      Vance     Jordan      Cohen 
##       5872       5864       5859       5858       5858       5854       5851 
##    Gregory     Wright       Hall     Hudson    Stewart     Morgan       Ward 
##       5851       5851       5850       5848       5848       5840       5839 
##  Carpenter      Mckee     Wilson     Walker       Rice    Russell     Wagner 
##       5834       5133       5132       5131       5128       5127       5124 
##  Gallagher       Lane    Mcmahon      Stark    Stevens Villarreal        Gay 
##       5122       5122       5122       5122       5122       5122       5121 
##     Joseph    (Other) 
##       5121     865612
```

```
class(fraudTotal.db$last)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$last <- as.factor(fraudTotal.db$last)
```

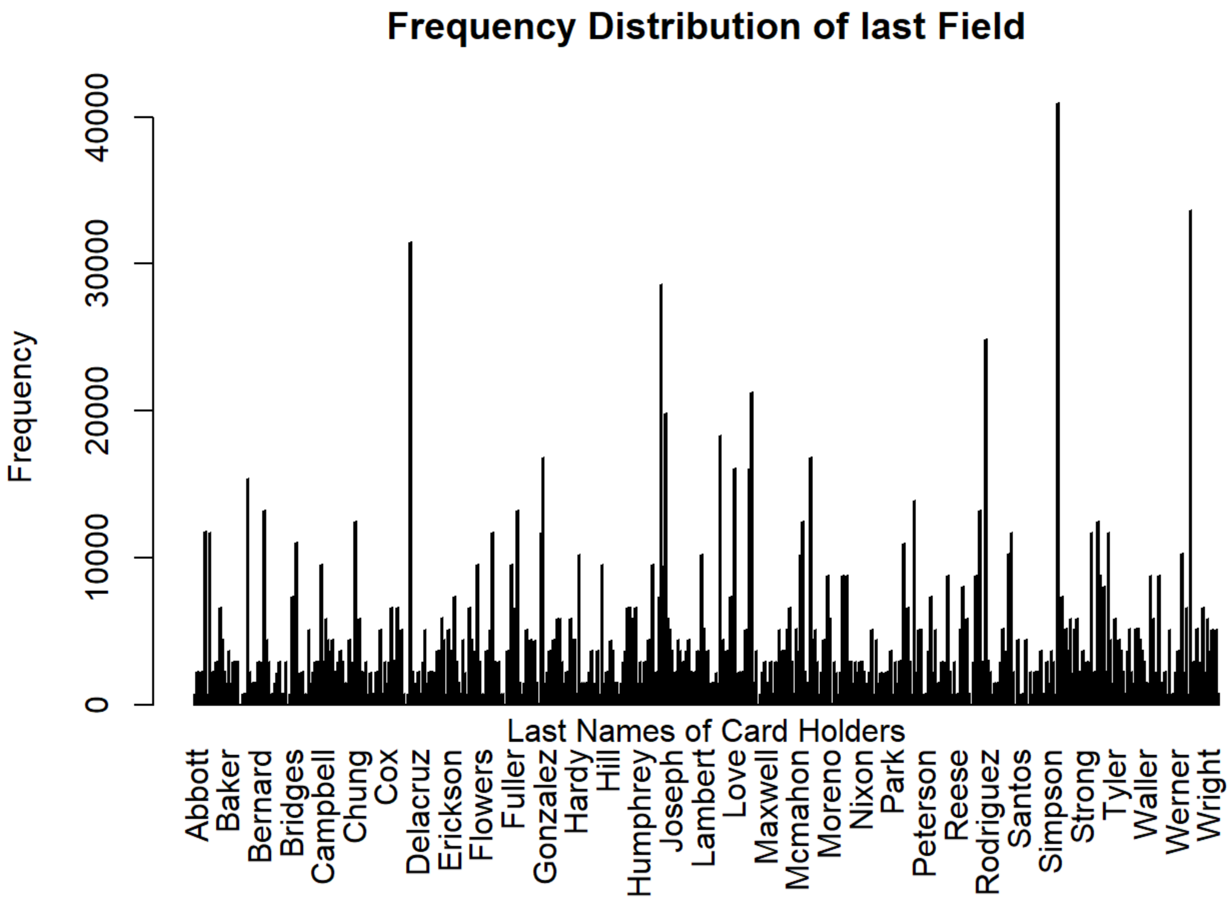
Frequency of last

```
table_last <- table(fraudTotal.db$last)
head(table_last)
```

```
## 
##    Abbott     Adams    Adkins   Aguilar Alexander     Allen 
##       736      2211       752      2202       740     11744
```

Frequency Distribution of last

```
barplot(table(fraudTotal.db$last), las = 3, main = "Frequency Distribution of last Field", xlab
 = "", ylab = "Frequency")
mtext("Last Names of Card Holders", side = 1)
```



Univariate Analysis of gender

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$gender))
```

```
## [1] 0
```

Summary of gender Column

```
summary(fraudTotal.db$gender)
```

```
##       F       M 
## 1014749  837645
```

```
class(fraudTotal.db$gender)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$gender <- as.factor(fraudTotal.db$gender)
```

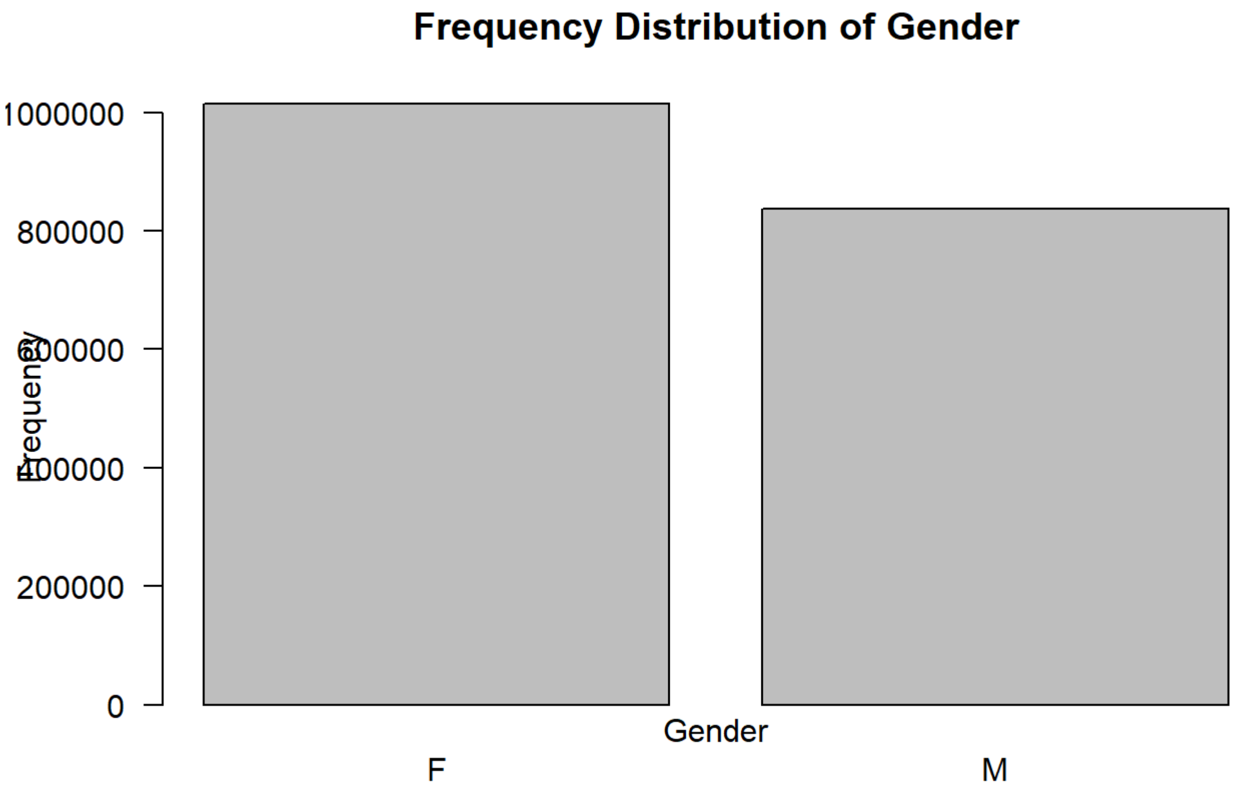
Frequency of gender

```
table(fraudTotal.db$gender)
```

```
## 
##       F       M 
## 1014749  837645
```

Frequency Distribution of gender

```
barplot(table(fraudTotal.db$gender), las = 1, main = "Frequency Distribution of Gender", xlab = "", ylab = "Frequency")
mtext("Gender", side = 1)
```



Univariate Analysis of street

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$street))
```

```
## [1] 0
```

Summary of street Column

```
summary(fraudTotal.db$street)
```

```
##                   444 Robert Mews                  908 Brooks Brook 
##                              4392                              4392 
##               03512 Jackson Ports            320 Nicholson Orchard 
##                              4386                              4386 
##           5796 Lee Coves Apt. 286             2924 Bobby Trafficway 
##                              4386                              4385 
##               40624 Rebecca Spurs         574 David Locks Suite 207 
##                              4385                              4384 
##       6114 Adams Harbor Suite 096                6983 Carrillo Isle 
##                              4384                              4384 
##               864 Reynolds Plains    29606 Martinez Views Suite 653 
##                              4384                              4383 
## 8172 Robertson Parkways Suite 072                   2481 Mills Lock 
##                              4383                              4382 
##        6033 Young Track Suite 804              0925 Lang Extensions 
##                              4382                              4381 
##                7202 Jeffrey Mills                   1652 James Mews 
##                              4381                              4380 
##                 4038 Smith Avenue     4664 Sanchez Common Suite 930 
##                              4380                              4380 
##             7618 Gonzales Mission        899 Michele View Suite 960 
##                              4380                              4380 
##      19838 Tonya Prairie Apt. 947                 26544 Andrea Glen 
##                              4379                              4379 
##               17666 David Valleys                 27479 Reeves Dale 
##                              4378                              4378 
##                372 Jeffrey Course     43235 Mckenzie Views Apt. 837 
##                              4378                              4377 
##                   516 Brown Parks        2870 Bean Terrace Apt. 756 
##                              4377                              4376 
##              4293 Ramirez Squares                 03030 White Lakes 
##                              4376                              4375 
##    06959 Stephen Branch Suite 246                23843 Scott Island 
##                              4375                              4375 
##              3379 Williams Common        0069 Robin Brooks Apt. 695 
##                              4375                              4374 
##       47029 Jimmy Tunnel Apt. 106         561 Little Plain Apt. 738 
##                              4374                              4374 
##           597 Jenny Ford Apt. 543         854 Walker Dale Suite 488 
##                              4374                              4374 
##          6296 John Keys Suite 858        50872 Alex Plain Suite 088 
##                              4373                              4372 
##       72966 Shannon Pass Apt. 391                    08236 Kim Hill 
##                              4372                              4371 
##                  742 Oneill Shore        5395 Colon Burgs Suite 037 
##                              4371                              4369 
##         594 Berry Lights Apt. 392    72269 Elizabeth Field Apt. 132 
##                              4369                              4366 
##          11014 Chad Lake Apt. 573                8030 Beck Motorway 
##                              4365                              4364 
##             3531 Hamilton Highway      43039 Riley Greens Suite 393 
##                              4362                              4362 
##                   7952 Karen Pike        9486 Joel Common Suite 554
##                              4357                              3664
##     2807 Parker Station Suite 080               572 Davis Mountains
##                              3661                              3661
##                   350 Stacy Glens       117 Natasha Vista Suite 936
##                              3660                              3658
##                269 Sanchez Rapids     7600 Stephen Course Suite 031
##                              3657                              3657
##        31472 Cody Place Suite 740                  428 Morgan River
##                              3656                              3656
##           1166 Castillo Mountains        250 Benjamin Hill Apt. 026
##                              3655                              3655
##         3522 Park Wells Suite 528        4130 Tiffany Glen Apt. 562
##                              3655                              3655
##     838 Franklin Prairie Apt. 902                  982 Melissa Lock
##                              3655                              3655
##              16285 Jessica Lights         1898 Parker Fork Apt. 057
##                              3654                              3654
##        2838 White Fields Apt. 473                3283 James Station
##                              3654                              3654
##         537 Rice Square Suite 040               576 House Crossroad
##                              3654                              3654
##      3310 Davidson Spurs Apt. 107               57256 Raymond Ports
##                              3653                              3653
##        622 Bradley Knoll Apt. 758            767 Adam Mill Apt. 115
##                              3653                              3653
##            911 Sabrina Trafficway          319 Wendy Fort Suite 179
##                              3653                              3652
##             329 Michael Extension     382 Williams Stream Suite 197
##                              3652                              3652
##                  821 Solis Points                  861 Karen Common
##                              3652                              3652
##                000 Jennifer Mills     01892 Patricia Vista Apt. 828
##                              3651                              3651
##        144 Evans Islands Apt. 683          830 Myers Plaza Apt. 384
##                              3651                              3651
##               094 Owens Underpass               3603 Mitchell Court
##                              3650                              3650
##      5939 Garcia Forges Suite 297      7118 Jessica Unions Apt. 789
##                              3650                              3650
##      79472 Stevens Trace Apt. 120         87665 Karen Mill Apt. 586
##                              3650                              3650
##              9333 Valentine Point               98897 Bennett Lodge
##                              3650                              3650
##       3645 Atkins Island Apt. 238          6602 Ortiz Pine Apt. 179
##                              3649                              3649
##       6911 Nicholas Keys Apt. 237                           (Other)
##                              3649                           1452353
```

```
class(fraudTotal.db$street)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$street <- as.factor(fraudTotal.db$street)
```

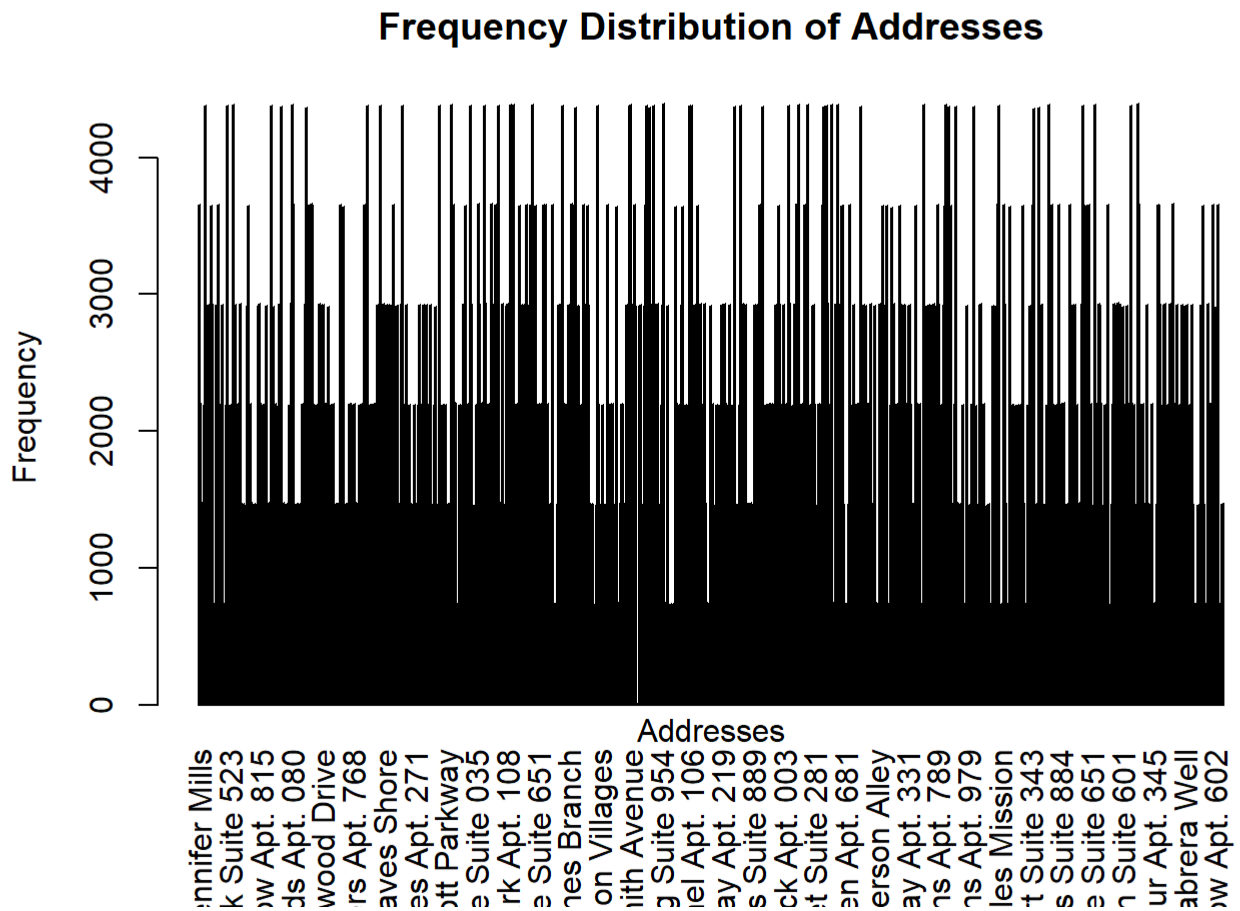
Frequency of Street

```
table_street <- table(fraudTotal.db$street)
head(table_street)
```

```
## 
##          000 Jennifer Mills          0005 Morrison Land
##                        3651                        2196
##        00315 Ashley Valleys 00378 Sarah Burgs Suite 106
##                        1475                          11
##            0043 Henry Plaza            005 Cody Estates
##                        1468                        2192
```

Frequency Distribution of street

```
barplot(table(fraudTotal.db$street), las = 3, main = "Frequency Distribution of Addresses", xlab
= "", ylab = "Frequency")
mtext("Addresses", side = 1)
```



Univariate Analysis of city

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$city))
```

```
## [1] 0
```

Summary of city Column

```
summary(fraudTotal.db$city)
```

```
##        Birmingham       San Antonio             Utica           Phoenix
##              8040              7312              7309              7297
##          Meridian            Warren            Conway         Cleveland
##              7289              6584              6574              6572
##            Thomas           Houston           Arcadia            Naples
##              6571              5865              5850              5849
##           Brandon            Fulton      Indianapolis           Burbank
##              5844              5841              5838              5831
##            Dallas        Washington           Detroit            Hudson
##              5141              5130              5124              5123
##          Lakeland         Allentown     Fort Washakie            Lahoma
##              5120              5119              5116              5116
##      Philadelphia           Andrews        Huntsville            Orient
##              5113              5107              5103              5093
##            Topeka             Tulsa      Clarks Mills             Lomax
##              4401              4400              4392              4392
##           Gadsden              Reno          Thompson      Walnut Ridge
##              4387              4386              4386              4386
##           De Witt           Sebring         Cottekill     Edisto Island
##              4385              4385              4384              4384
## Kingsford Heights           Norwalk             Uledi         Hinesburg
##              4384              4384              4384              4383
##          Superior       East Canaan        Plainfield           Shields
##              4383              4382              4382              4381
##           Bradley        Centerview          Hinckley             Jones
##              4380              4380              4380              4380
##          Goodrich       Rocky Mount        Morrisdale          Sun City
##              4379              4379              4378              4378
##        Sutherland       Whaleyville        Wilmington        Manistique
##              4378              4378              4378              4377
##            Norman          Westport            Ranier         Grandview
##              4377              4377              4376              4375
##         Littleton             Thida           Bowdoin           Elberta
##              4375              4375              4374              4374
##           Newhall       Tupper Lake           Wetmore Pembroke Township
##              4374              4374              4374              4373
##       Baton Rouge           Bauxite          Florence            Thrall
##              4372              4372              4371              4369
##       Heart Butte          Moorhead              Roma           De Soto
##              4365              4364              4362              4357
##     New York City            Camden         San Diego          Fenelton
##              3680              3678              3664              3663
##         Meadville           Diamond      Lake Jackson       Stanchfield
##              3662              3661              3661              3661
##           Spencer          Glendale           Leonard       Springfield
##              3660              3659              3658              3658
##         Elizabeth         Red River        Kensington           Wichita
##              3657              3657              3656              3656
##            Bagley          Key West            Mobile           (Other)
##              3655              3655              3655           1389261
```

```
class(fraudTotal.db$city)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$city <- as.factor(fraudTotal.db$city)
```

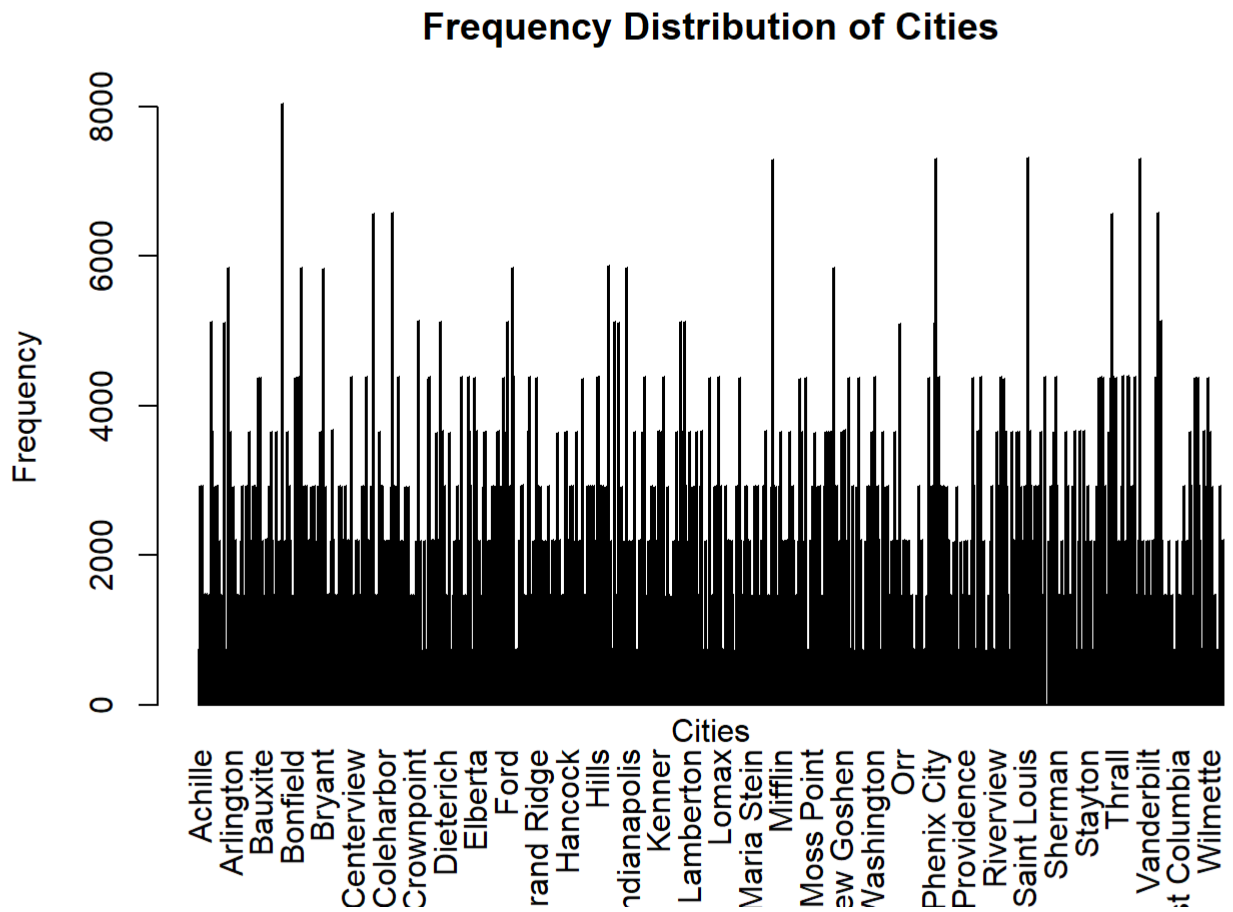
Frequency of city

```
table_city <- table(fraudTotal.db$city)
head(table_city)
```

```
## 
## Achille Acworth   Adams   Afton   Akron  Albany 
##     740    2925     739    2932     733    1479
```

Frequency Distribution of city

```
barplot(table(fraudTotal.db$city), las = 3, main = "Frequency Distribution of Cities", xlab = ""
, ylab = "Frequency")
mtext("Cities", side = 1)
```



Univariate Analysis of state

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$state))
```

```
## [1] 0
```

Summary of state Column

```
summary(fraudTotal.db$state)
```

```
##     AK     AL     AR     AZ     CA     CO     CT     DC     DE     FL     GA 
##   2963  58521  44611  15362  80495  19766  10979   5130      9  60775  37340 
##     HI     IA     ID     IL     IN     KS     KY     LA     MA     MD     ME 
##   3649  38804   8035  62212  39539  32939  40981  29953  17562  37345  23433 
##     MI     MN     MO     MS     MT     NC     ND     NE     NH     NJ     NM 
##  65825  45433  54904  30021  16806  43134  21183  34425  11727  35131  23427 
##     NV     NY     OH     OK     OR     PA     RI     SC     SD     TN     TX 
##   8058 119419  66627  38050  26408 114173    745  41731  17574  24913 135269 
##     UT     VA     VT     WA     WI     WV     WY 
##  15357  41756  16812  27040  41738  36529  27776
```

```
class(fraudTotal.db$state)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$state <- as.factor(fraudTotal.db$state)
```

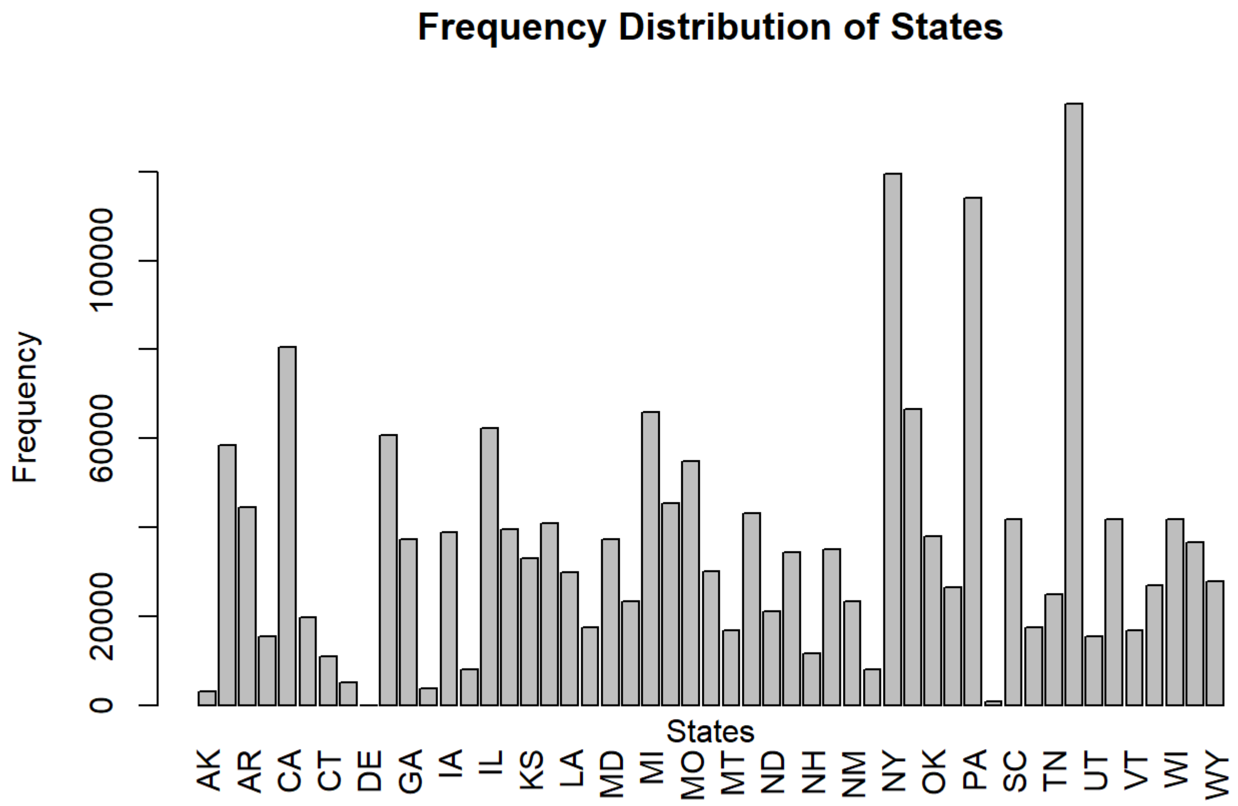
Frequency of state

```
table(fraudTotal.db$state)
```

```
## 
##     AK     AL     AR     AZ     CA     CO     CT     DC     DE     FL     GA 
##   2963  58521  44611  15362  80495  19766  10979   5130      9  60775  37340 
##     HI     IA     ID     IL     IN     KS     KY     LA     MA     MD     ME 
##   3649  38804   8035  62212  39539  32939  40981  29953  17562  37345  23433 
##     MI     MN     MO     MS     MT     NC     ND     NE     NH     NJ     NM 
##  65825  45433  54904  30021  16806  43134  21183  34425  11727  35131  23427 
##     NV     NY     OH     OK     OR     PA     RI     SC     SD     TN     TX 
##   8058 119419  66627  38050  26408 114173    745  41731  17574  24913 135269 
##     UT     VA     VT     WA     WI     WV     WY 
##  15357  41756  16812  27040  41738  36529  27776
```

Frequency Distribution of state

```
barplot(table(fraudTotal.db$state), las = 3, main = "Frequency Distribution of States", xlab = 
"", ylab = "Frequency")
mtext("States", side = 1)
```



Univariate Analysis of zip

###Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$zip))
```

```
## [1] 0
```

Summary of zip Column

```
summary(fraudTotal.db$zip)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1257   26237   48174   48813   72042   99921
```

```
class(fraudTotal.db$zip)
```

```
## [1] "integer"
```

Find the Standard Deviation and Variance of zip variable

```
sd(fraudTotal.db$zip)
```

```
## [1] 26881.85
```

```
var(fraudTotal.db$zip)
```

```
## [1] 722633643
```

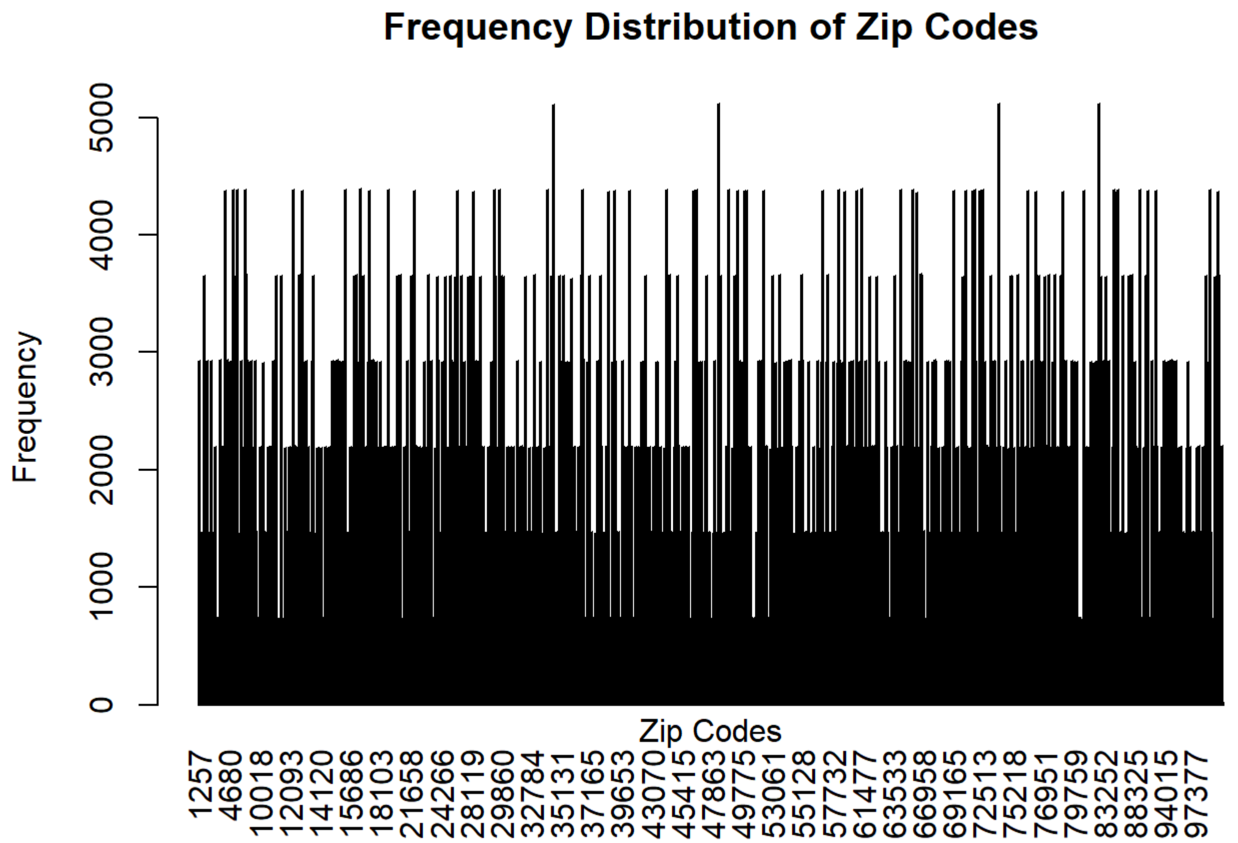
Frequency of zip

```
table_zip <- table(fraudTotal.db$zip)
head(table_zip)
```

```
## 
## 1257 1330 1535 1545 1612 1843 
## 2923 1466  734 1468  738 3652
```

Frequency Distribution of zip

```
barplot(table(fraudTotal.db$zip), las = 3, main = "Frequency Distribution of Zip Codes", xlab = "", ylab = "Frequency")
mtext("Zip Codes", side = 1)
```



Univariate Analysis of lat

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$lat))
```

```
## [1] 0
```

Summary of lat Column

```
summary(fraudTotal.db$lat)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   20.03   34.67   39.35   38.54   41.94   66.69
```

```
class(fraudTotal.db$lat)
```

```
## [1] "numeric"
```

Find the Standard Deviation and Variance of lat variable

```
sd(fraudTotal.db$lat)
```

```
## [1] 5.07147
```

```
var(fraudTotal.db$lat)
```

```
## [1] 25.71981
```

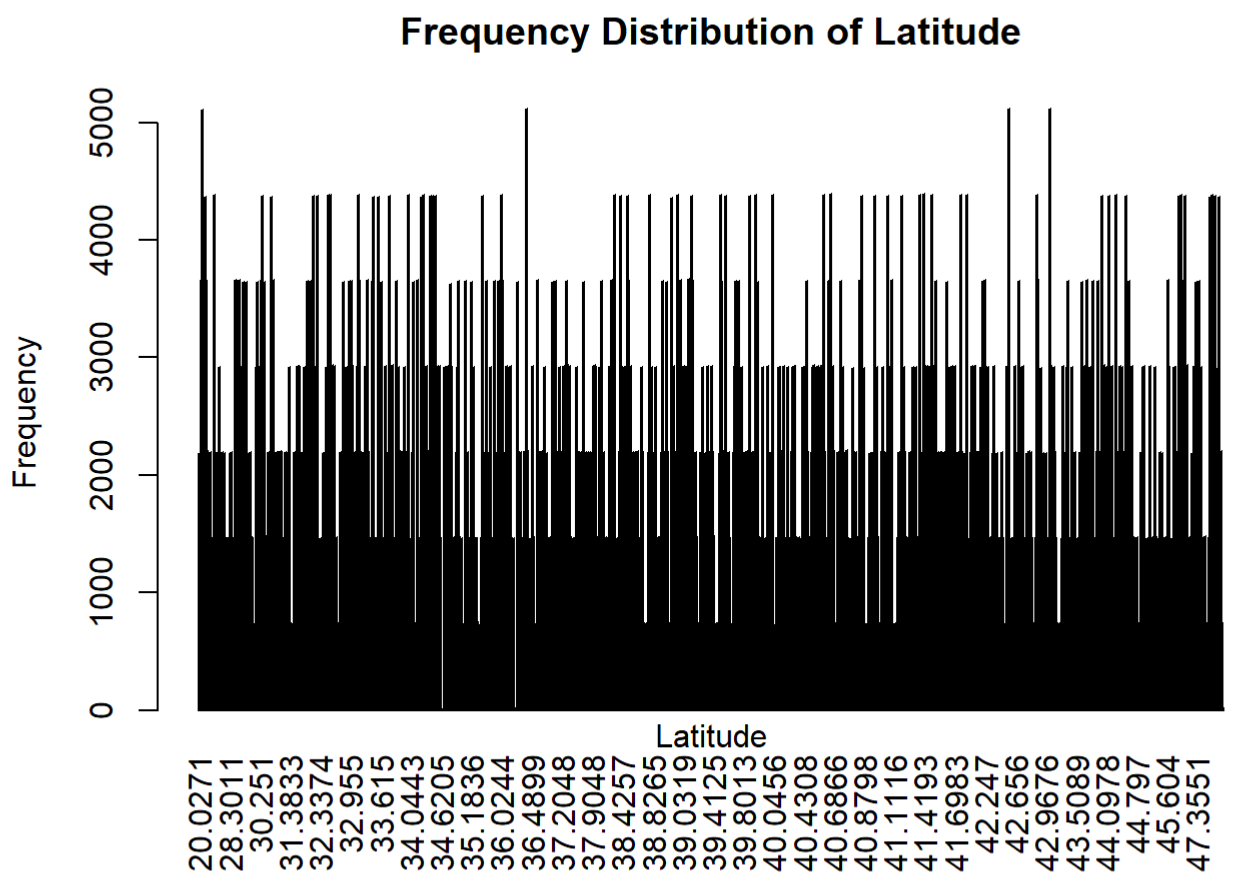
Frequency of lat

```
table_lat <- table(fraudTotal.db$lat)
head(table_lat)
```

```
## 
## 20.0271 20.0827 24.6557 26.1184 26.3304 26.3771 
##    2186    1463    3655    5108     741     732
```

Frequency Distribution of lat

```
barplot(table(fraudTotal.db$lat), las = 3, main = "Frequency Distribution of Latitude", xlab =
"", ylab = "Frequency")
mtext("Latitude", side = 1)
```



Univariate Analysis of long

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$long))
```

```
## [1] 0
```

Summary of long Column

```
summary(fraudTotal.db$long)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
## -165.67  -96.80  -87.48  -90.23  -80.16  -67.95
```

```
class(fraudTotal.db$long)
```

```
## [1] "numeric"
```

Find the Standard Deviation and Variance of long variable

```
sd(fraudTotal.db$long)
```

```
## [1] 13.74789
```

```
var(fraudTotal.db$long)
```

```
## [1] 189.0046
```

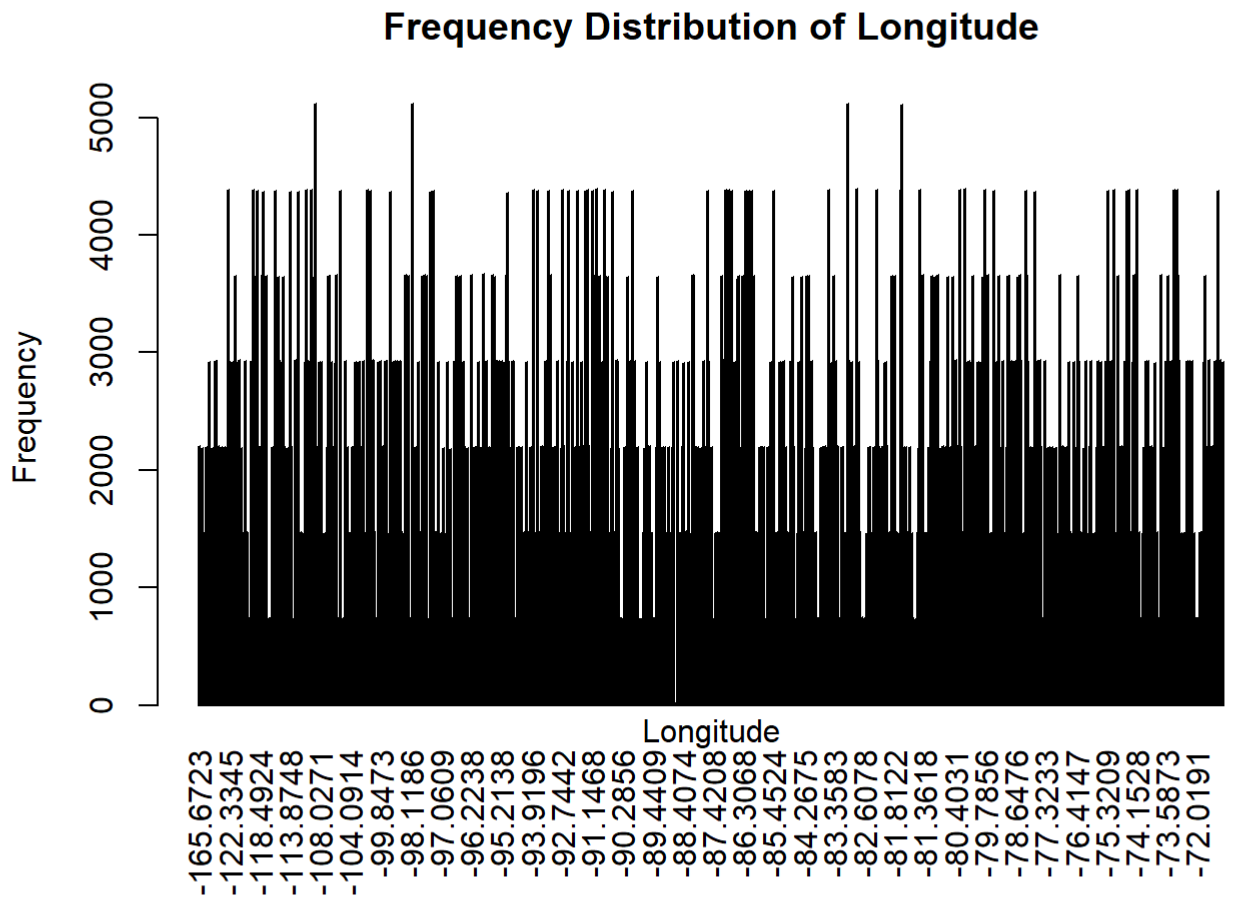
Frequency of long

```
table_long <- table(fraudTotal.db$long)
head(table_long)
```

```
## 
## -165.6723  -156.292  -155.488 -155.3697  -153.994 -133.1171 
##      2203       734      1463      2186        12        14
```

Frequency Distribution of long

```
barplot(table(fraudTotal.db$long), las = 3, main = "Frequency Distribution of Longitude", xlab =
"", ylab = "Frequency")
mtext("Longitude", side = 1)
```



Univariate Analysis of city_pop

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$city_pop))
```

```
## [1] 0
```

Summary of city_pop Column

```
summary(fraudTotal.db$city_pop)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##      23     741    2443   88644   20328 2906700
```

```
class(fraudTotal.db$city_pop)
```

```
## [1] "integer"
```

Find the Standard Deviation and Variance of city_pop variable

```
sd(fraudTotal.db$city_pop)
```

```
## [1] 301487.6
```

```
var(fraudTotal.db$city_pop)
```

```
## [1] 90894784015
```

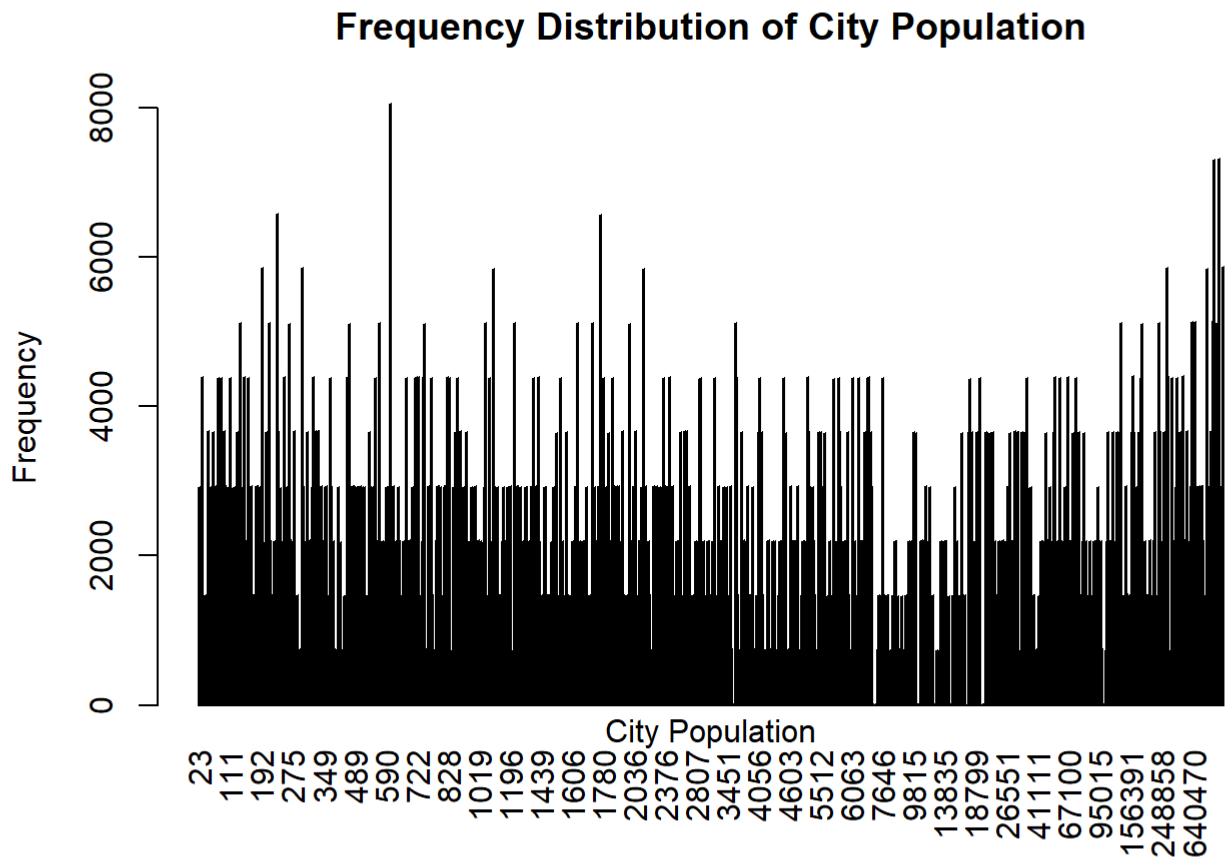
Frequency of city_pop

```
table_city_pop <- table(fraudTotal.db$city_pop)
head(table_city_pop)
```

```
## 
##   23   37   43   46   47   49 
## 2915 1469 2920 4386  734 1472
```

Frequency Distribution of city_pop

```
barplot(table(fraudTotal.db$city_pop), las = 3, main = "Frequency Distribution of City Populatio
n", xlab = "", ylab = "Frequency")
mtext("City Population", side = 1)
```



Univariate Analysis of job

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$job))
```

```
## [1] 0
```

Summary of job Column

```
summary(fraudTotal.db$job)
```

```
##                                 Film/video editor
##                                             13898
##                               Exhibition designer
##                                             13167
##                          Surveyor, land/geomatics
##                                             12436
##                                   Naval architect
##                                             12434
##                                Materials engineer
##                                             11711
##                        Designer, ceramics/pottery
##                                             11688
##                          Environmental consultant
##                                             10974
##                                 Financial adviser
##                                             10963
##                                 Systems developer
##                                             10962
##                                        IT trainer
##                                             10943
##                           Copywriter, advertising
##                                             10241
##                           Scientist, audiological
##                                             10234
##               Chartered public finance accountant
##                                             10211
##                           Chief Executive Officer
##                                             10199
##                                        Podiatrist
##                                              9525
##                                       Comptroller
##                                              9515
##                          Magazine features editor
##                                              9506
##                           Agricultural consultant
##                                              9500
##                                         Paramedic
##                                              9494
##                                               Sub
##                                              9488
##                            Audiological scientist
##                                              8801
## Historic buildings inspector/conservation officer
##                                              8787
##                                 Building surveyor
##                                              8786
##                                 Librarian, public
##                                              8773
##                                          Musician
##                                              8772
##                       Scientist, research (maths)
##                                              8768
##                                         Barrister
##                                              8767
##                     Clothing/textile technologist
##                                   Mining engineer
##                                              8765
##                                      Immunologist
##                                              8762
##                                    Water engineer
##                                              8760
##                                 Quantity surveyor
##                                              8740
##                               Mechanical engineer
##                                              8080
##                          Secondary school teacher
##                                              8062
##                                  Financial trader
##                                              8056
##                                    Prison officer
##                                              8054
##                           Land/geomatics surveyor
##                                              8054
##                            Sales professional, IT
##                                              8052
##                              Engineer, automotive
##                                              8052
##                                        Counsellor
##                                              8050
##                                Petroleum engineer
##                                              8047
##                            Psychologist, forensic
##                                              8046
##                         Claims inspector/assessor
##                                              8044
##                               Early years teacher
##                                              8042
##                                      Geoscientist
##                                              8041
##                                   Energy engineer
##                                              8041
##                               Pensions consultant
##                                              8038
##                            Psychotherapist, child
##                                              8036
##                                              Make
##                                              8036
##                                       Firefighter
##                                              8028
##                                 Chemical engineer
##                                              8021
##                                    Science writer
##                                              7334
##                                              7332
##                               Engineer, biomedical
##                                               7330
##                                  Drilling engineer
##                                               7321
##             Research scientist (physical sciences)
##                                               7319
##                       Medical sales representative
##                                               7309
##                                Librarian, academic
##                                               7307
##                                  Scientist, marine
##                                               7306
##                                Trade mark attorney
##                                               7304
##                                Electrical engineer
##                                               7301
##                              Insurance underwriter
##                                               7301
##                                     Cytogeneticist
##                                               7297
##                    Television production assistant
##                                               7297
##                            Chartered loss adjuster
##                                               7296
##                  Special educational needs teacher
##                                               7283
##                          Trading standards officer
##                                               6611
##                              Accounting technician
##                                               6595
##                            Therapist, occupational
##                                               6594
##                           Counselling psychologist
##                                               6590
##                                 Surveyor, minerals
##                                               6589
##                           Educational psychologist
##                                               6588
##                                             Dealer
##                                               6586
##                               Engineer, production
##                                               6584
##                             Race relations officer
##                                               6583
##                              Multimedia programmer
##                                               6582
##                          Radio broadcast assistant
##                                               6582
##                                  Social researcher
##                                               6580
##              Engineer, control and instrumentation
##                                               6579
##                                     Radio producer
##                                               6579
##                 Teacher, special educational needs
##                                               6578
##                             Chief Strategy Officer
##                                               6577
##                                        Fine artist
##                                               6576
##                                   Technical brewer
##                                               6576
##                                  Ceramics designer
##                                               6569
##                                    Physiotherapist
##                                               6566
##                                       Toxicologist
##                                               6555
##              Senior tax professional/tax inspector
##                                               5877
##                     Television/film/video producer
##                                               5871
##                         Further education lecturer
##                                               5865
##                              Scientist, biomedical
##                                               5862
##                                      Archaeologist
##                                               5860
##                                     Futures trader
##                                               5860
##                                  Buyer, industrial
##                                               5857
##                              Engineering geologist
##                                               5857
##                                      Lexicographer
##                                               5857
##                       Designer, industrial/product
##                                               5856
##                                  Probation officer
##                                               5856
##                        Advertising account planner
##                                               5852
##                      Development worker, community
##                                               5852
##                                            (Other)
##                                            1061906
```

```
class(fraudTotal.db$job)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$job <- as.factor(fraudTotal.db$job)
```

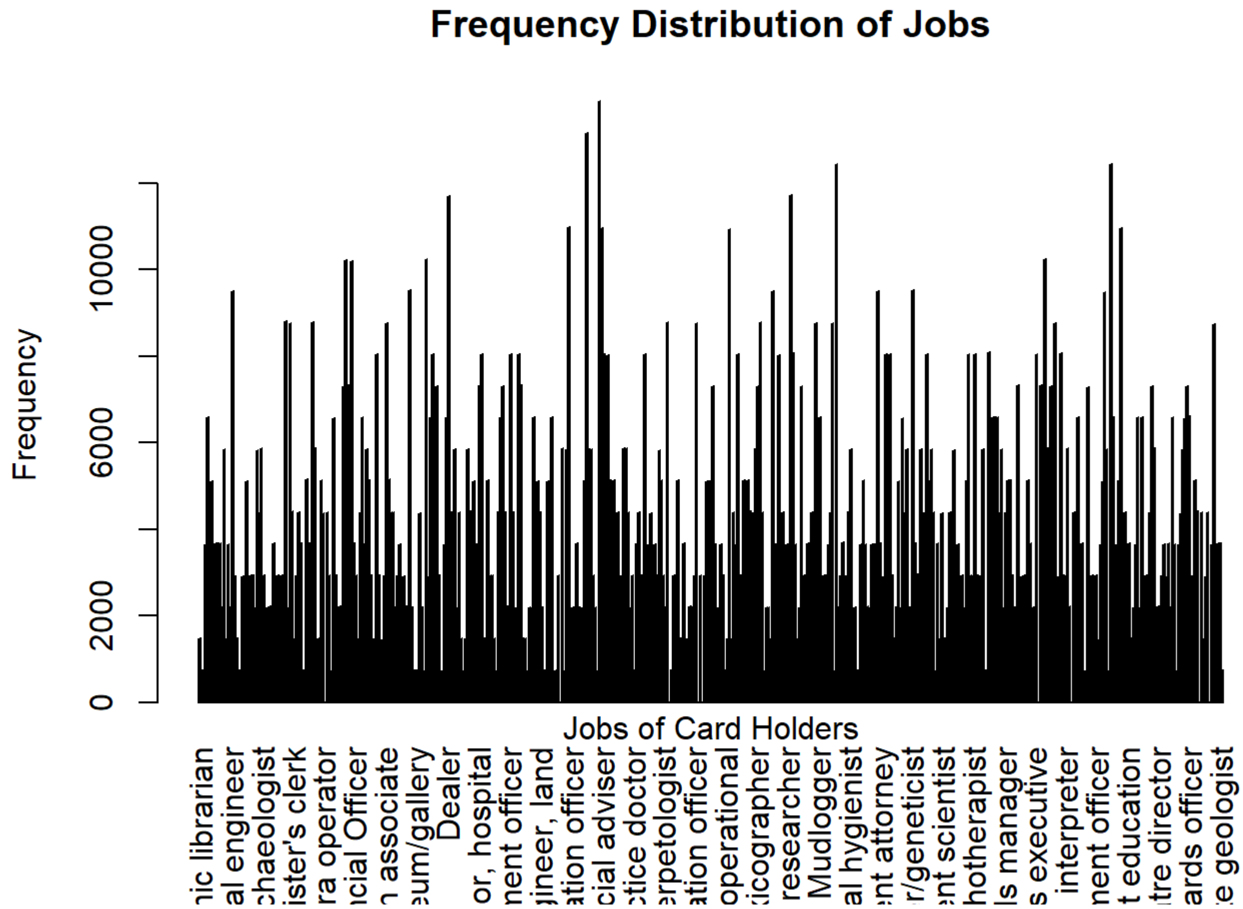
Frequency of job

```
table_job <- table(fraudTotal.db$job)
head(table_job)
```

```
## 
##                     Academic librarian                Accountant, chartered 
##                                   1467                                   11 
##      Accountant, chartered certified Accountant, chartered public finance 
##                                    751                                 3657 
##                  Accounting technician                        Acupuncturist 
##                                   6595                                 2198 
```

Frequency Distribution of job

```
barplot(table(fraudTotal.db$job), las = 3, main = "Frequency Distribution of Jobs", xlab = "", y
lab = "Frequency")
mtext("Jobs of Card Holders", side = 1)
```



Univariate Analysis of dob

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$dob))
```

```
## [1] 0
```

Converting Character to DateTime class

```
library(lubridate)

fraudTotal.db$dob <- ymd(fraudTotal.db$dob)
```

Summary of dob Column

```
summary(fraudTotal.db$dob)
```

```
##         Min.      1st Qu.       Median         Mean      3rd Qu.         Max. 
## "1924-10-30" "1962-08-13" "1975-11-30" "1973-10-15" "1987-04-23" "2005-01-29"
```

```
class(fraudTotal.db$dob)
```

```
## [1] "Date"
```

Find the Standard Deviation and Variance of dob variable

```
sd(fraudTotal.db$dob)
```

```
## [1] 6356.34
```

```
var(fraudTotal.db$dob)
```

```
## [1] 40403063
```

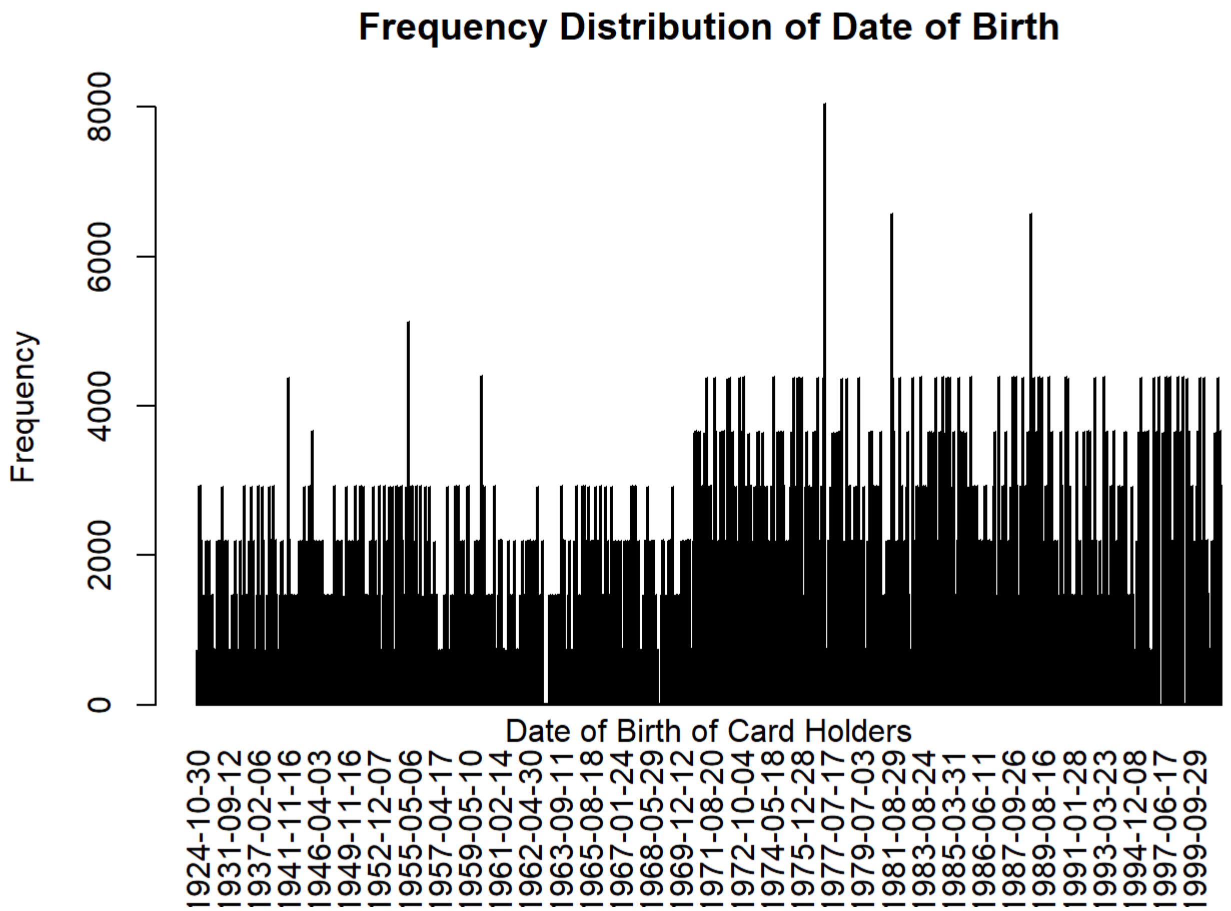
Frequency of dob

```
table_dob <- table(fraudTotal.db$dob)
head(table_dob)
```

```
## 
## 1924-10-30 1925-08-29 1926-06-26 1926-07-12 1926-08-27 1926-09-14 
##        735         11       2924       2923       2198        738
```

Frequency Distribution of dob

```
barplot(table(fraudTotal.db$dob), las = 3, main = "Frequency Distribution of Date of Birth", xla
b = "", ylab = "Frequency")
mtext("Date of Birth of Card Holders", side = 1)
```



Univariate Analysis of trans_num

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$trans_num))
```

```
## [1] 0
```

Summary of trans_num Column

```
summary(fraudTotal.db$trans_num)
```

```
## 00000ecad06b03d3a8d34b4e30b5ce3b 000014ca3f6921fe6793f88fe494f39d
##                                1                                1
## 00001ded488fddab97677128e5034d39 0000246d803d5f465cc322d8a3c3528f
##                                1                                1
## 0000258ae973a6199fca79d94947672f 0000307898b3352b5a0d66015d362794
##                                1                                1
## 0000425d184356a21be4b39933c2c0ea 000048ecb6c1d9337bcc27109b46794d
##                                1                                1
## 000051c6b92f7cd491c41b025d60a933 00005fc67bb45d98730559d40c9ca601
##                                1                                1
## 000067191c6544818ea1831c381d72c3 00006889944d759855fea412e09ecdd8
##                                1                                1
## 00006bc3a2769e9f44cbf2dde6e69ede 00007a622ab06e0ea2669fc38ba6b60b
##                                1                                1
## 000088fe170f044d2ed28c570282c7a4 00008f7ba50172eef2b057a0e06aa142
##                                1                                1
## 0000909e67e3cc52099da05d144ee403 0000948be671f10fc10b2aee0d67edee
##                                1                                1
## 0000ab27c12ae11b3317ff22750cb022 0000ad2f657e05cba8e1f3654c3317a4
##                                1                                1
## 0000b45f78355eab64e143fd8cf1d721 0000d3f43ee755ae8a702153b0fc7510
##                                1                                1
## 0000ddb86257223f2e75b951bb6b1c13 0000dfd04a508bc2bd2856186cffaf44
##                                1                                1
## 0000e82fd9660b8069fd16845d540615 000119ca11541a47cf0dd7a204c0f69a
##                                1                                1
## 000121bf3adda35bc12ffd4db959060e 000127d6a7195801ee88746dc1b0c0c2
##                                1                                1
## 00013a9a862695102316e8d566462618 00013fa083c116fed2268a7db6c6506b
##                                1                                1
## 000143780aca896d53718bdd6e2eb5a2 0001448f4aa37c0e13159e1120968ca5
##                                1                                1
## 000163efab01b5ed89ad0b3025fd2dc6 000164cb82c6ebf15ea285af35c09a4e
##                                1                                1
## 000169feeab4b72e7a95cc3203edcbfb 00016ca603ab04668d1ab1181c2fe40d
##                                1                                1
## 00017954fb9236f8b82ea2237d521f92 00017fcc7a37796b5ebf048464185744
##                                1                                1
## 00018ba4aaeb4d893423d6b2426f02db 0001911bd16f115b060cef4cdd238a5c
##                                1                                1
## 0001915769fd20ee9f76ec525525ec19 000192586db6ad7d6a51ef50971ec03d
##                                1                                1
## 00019d470fa038baf78fb7968e1dea9a 0001c39cd8006f1f0d5d3e2d96b0b3d2
##                                1                                1
## 0001d673d869467160af100fe1713eda 0001d73e875fc1db9646c7c9d12e6470
##                                1                                1
## 0001d8e944541d39e42583401464b6a4 0001e0e8d8941d4f94880d0423d674f4
##                                1                                1
## 0001f83adff5a6bd710f1a7ebd61a258 0001ffb93362b5bf185f5bbf95cd0bec
##                                1                                1
## 000200cf67b9502d76600541f03ab842 000208c59bec43c567baac1452d886e4
##                                1                                1
## 00020a7faae397ec51eb688f5d61b003 00021bf0df08095718bfa5a47a525949
##                                1                                1
## 00023214e7c5d7c16959f0c7ba07908e 00023738925895e0d02ca429693331e8
##                                1                                1
## 000245bf66ede36b681282f3d042f9e3 000246c992d4e227bc0485323dd9b4ff
##                                1                                1
## 000250d9266ac0f0949ad8be70e55eb4 000252c0b3c8107c1b9a2fc2f6b9d7c7
##                                1                                1
## 00025a7a8b21f957dd49beae5e151cee 0002711a6411d09a663371181b3c702b
##                                1                                1
## 00028fbf5d33903f9791a76eb3360ef6 00029d34a548c75ee08d5847a348bece
##                                1                                1
## 0002a4b00ac3d229c435fcb082162a07 0002aa5322e4573674783d0ac0c8ae27
##                                1                                1
## 0002bd7092d3288b42b94865e2cb9de9 0002d40e03a6bbf369b989ade1b187e3
##                                1                                1
## 0002d43eb1e12616cc80894056116860 0002d8e7cfdc154fe73665f5d5cc4db9
##                                1                                1
## 0002dae8c11316e2c03f432a59412412 0002e425f19a2e096016f0ba8244469b
##                                1                                1
## 0002f4af124d110d2eef10993744eb07 000305d19ddf67681fc32425e2e2c6ba
##                                1                                1
## 00030abda40155fa48410a650ae7abef 00030b0ba28bda80a5f587a836fb9359
##                                1                                1
## 00032e683eb5bb37425a0cae2fe6c7f9 000330b78a4c89e698afb61b5ac86416
##                                1                                1
## 000339c200c0768700aae04c433a1650 0003428c6ec591e9a312d8fc79a10880
##                                1                                1
## 0003485b55a2981084749c6c5457be09 000366dcad8abb6ae9d1a0af731324d0
##                                1                                1
## 000368b37196ae86e6a464183f3b00b6 000371f9da2ef6799d41011614317d97
##                                1                                1
## 0003811e17f1c64a8cd29beddb62b92f 000395c23ce64b297cc7634ad3d565ed
##                                1                                1
## 000396e0eb0ac07230f5dbd6e7cdb0b0 0003997b8941c298a7e6b19e6918588f
##                                1                                1
## 0003a717990bfa35265d8d6c17db3110 0003b53c257094b2a0bbde9958900215
##                                1                                1
## 0003bba77eb22ca7b2d0bf975ae110ec 0003bffa13b8686aa83a878ba68db789
##                                1                                1
## 0003c29f50b8189dd4fef7dfe2e17cb2 0003c4d86ff998ae15f12b6231ada889
##                                1                                1
## 0003cdb7d3eb2564494480e3c51ee2d1 0003d4c16719801eee99c150ebc10477
##                                1                                1
## 0003e7b6ccdca1572cb049c1a7dd1ace 0003e99641f7a899fe6ec4ac9fb7a1c5
##                                1                                1
## 0003f8eab68854eb7c0fc8c2c24fb55a                          (Other)
##                                1                          1852295
```

```
class(fraudTotal.db$trans_num)
```

```
## [1] "factor"
```

Convert Characater Class to a Factor Class

```
fraudTotal.db$trans_num <- as.factor(fraudTotal.db$trans_num)
```

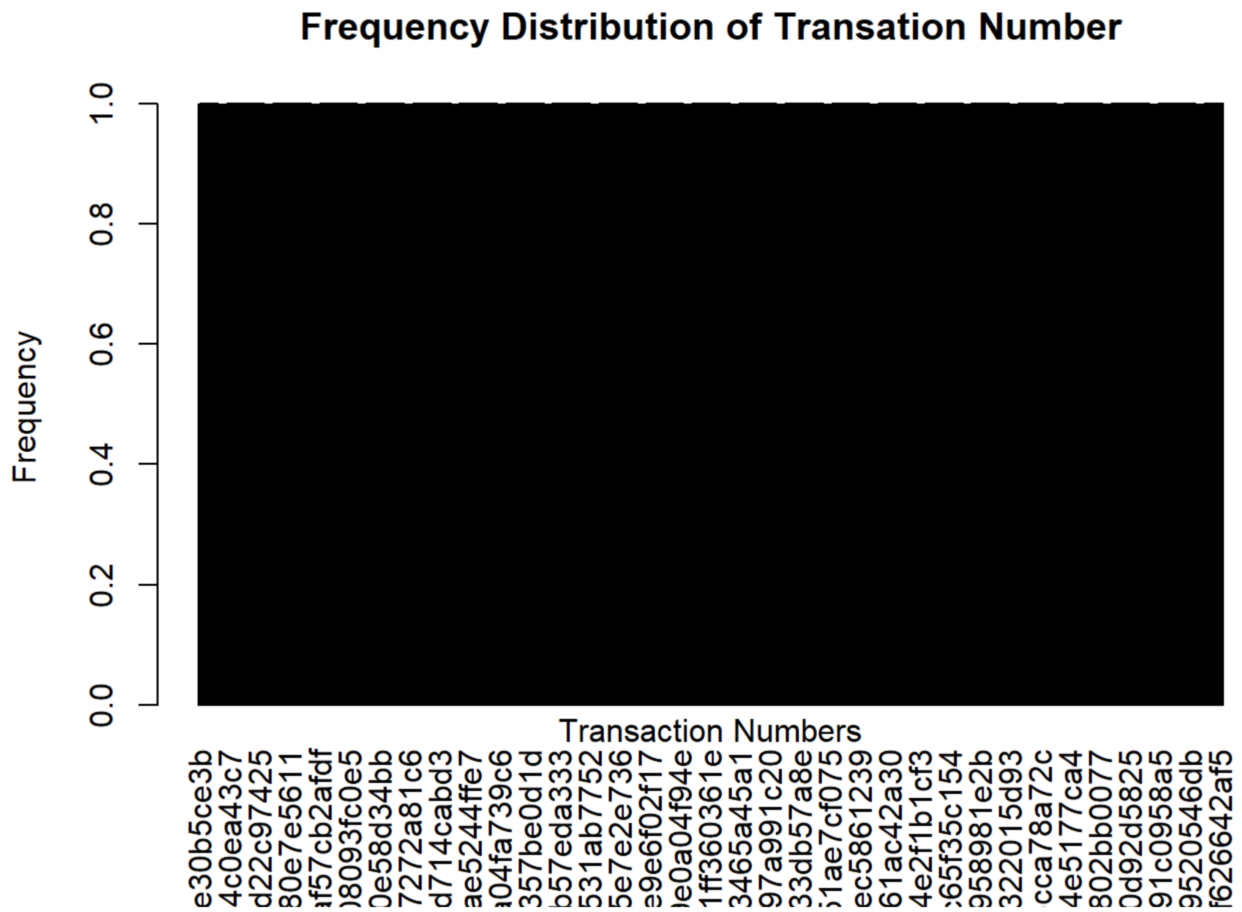
Frequency of trans_num

```
table_trans_num <- table(fraudTotal.db$trans_num)
head(table_trans_num)
```

```
## 
## 00000ecad06b03d3a8d34b4e30b5ce3b 000014ca3f6921fe6793f88fe494f39d 
##                                1                                1 
## 00001ded488fddab97677128e5034d39 0000246d803d5f465cc322d8a3c3528f 
##                                1                                1 
## 0000258ae973a6199fca79d94947672f 0000307898b3352b5a0d66015d362794 
##                                1                                1
```

Frequency Distribution of trans_num

```
barplot(table(fraudTotal.db$trans_num), las = 3, main = "Frequency Distribution of Transation Nu
mber", xlab = "", ylab = "Frequency")
mtext("Transaction Numbers", side = 1)
```



Univariate Analysis of unix_time

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$unix_time))
```

```
## [1] 0
```

Summary of unix_time Column

```
summary(fraudTotal.db$unix_time)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max. 
## 1325376018 1343016824 1357089331 1358674219 1374581485 1388534374
```

```
class(fraudTotal.db$unix_time)
```

```
## [1] "integer"
```

Find the Standard Deviation and Variance of unix_time variable

```
sd(fraudTotal.db$unix_time)
```

```
## [1] 18195081
```

```
var(fraudTotal.db$unix_time)
```

```
## [1] 331060986699918
```

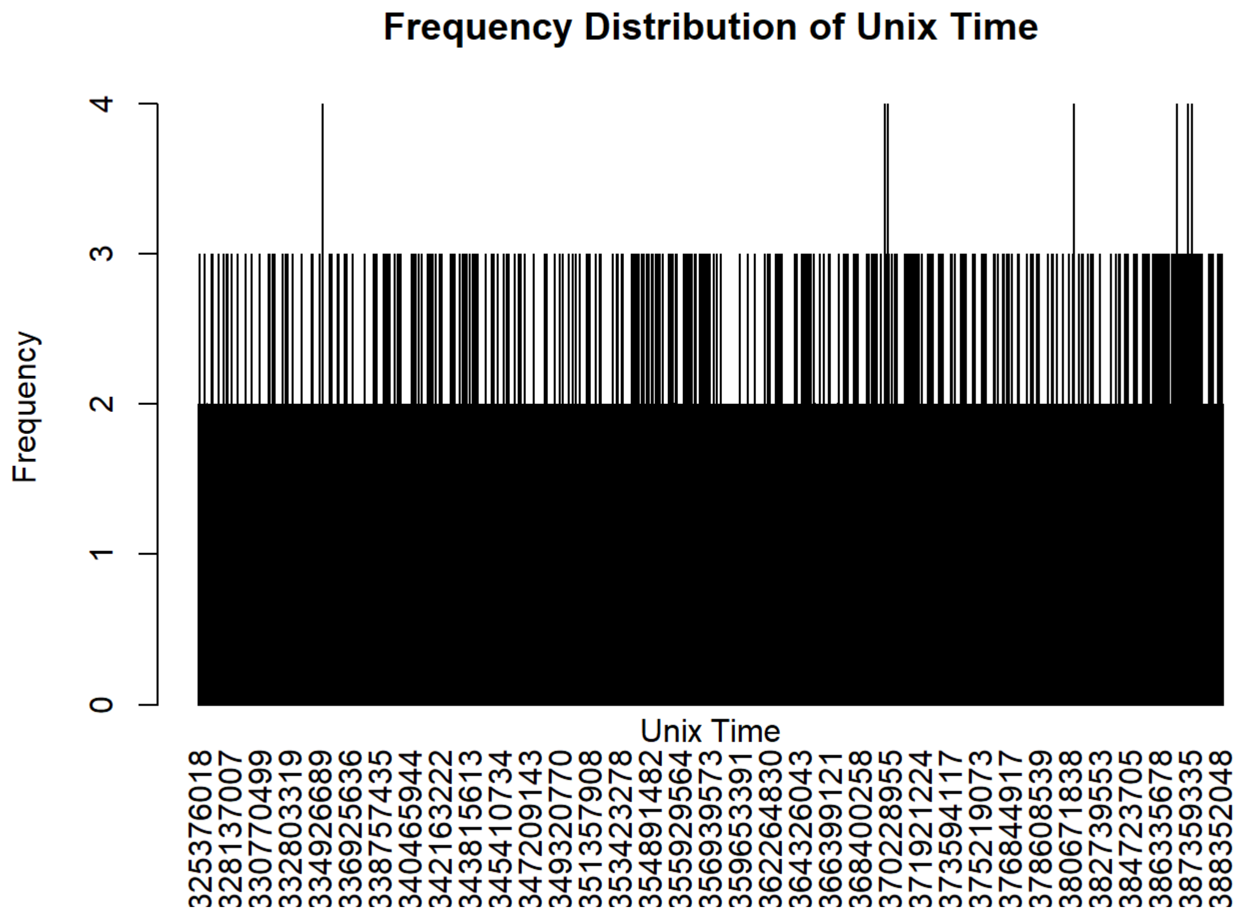
Frequency of unix_time

```
table_unix_time <- table(fraudTotal.db$unix_time)
head(table_unix_time)
```

```
## 
## 1325376018 1325376044 1325376051 1325376076 1325376186 1325376248 
##          1          1          1          1          1          1
```

Frequency Distribution of unix_time

```
barplot(table(fraudTotal.db$unix_time), las = 3, main = "Frequency Distribution of Unix Time", x
lab = "", ylab = "Frequency")
mtext("Unix Time", side = 1)
```



Univariate Analysis of merch_long

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$merch_long))
```

```
## [1] 0
```

Summary of merch_long Column

```
summary(fraudTotal.db$merch_long)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
## -166.67  -96.90  -87.44  -90.23  -80.25  -66.95
```

```
class(fraudTotal.db$merch_long)
```

```
## [1] "numeric"
```

Find the Standard Deviation and Variance of merch_long variable

```
sd(fraudTotal.db$merch_long)
```

```
## [1] 13.75969
```

```
var(fraudTotal.db$merch_long)
```

```
## [1] 189.3291
```

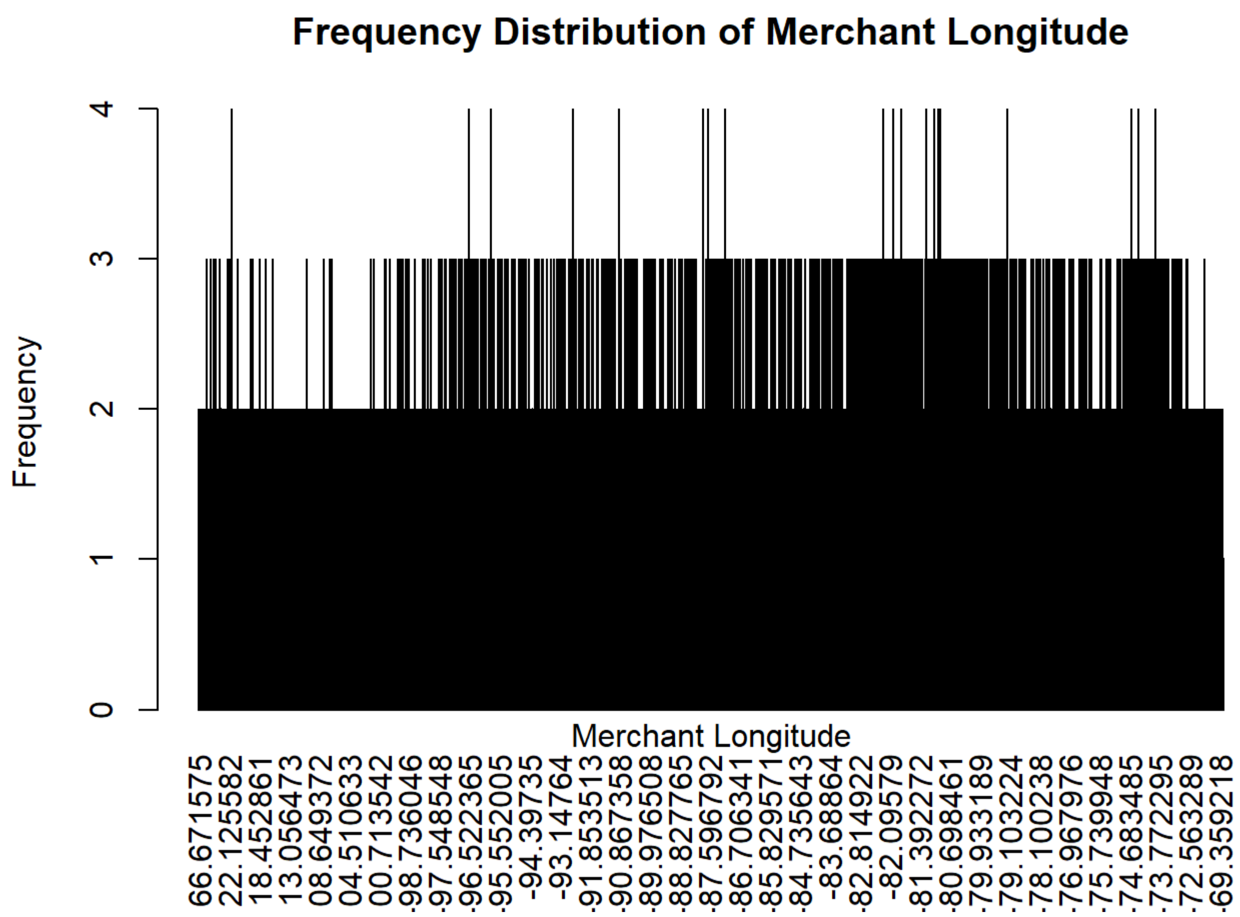
Frequency of merch_long

```
table_merch_long <- table(fraudTotal.db$merch_long)
head(table_merch_long)
```

```
## 
## -166.671575 -166.671242 -166.670685 -166.670132 -166.670006  -166.66991 
##           1           1           1           1           1           1
```

Frequency Distribution of merch_long

```
barplot(table(fraudTotal.db$merch_long), las = 3, main = "Frequency Distribution of Merchant Lon
gitude", xlab = "", ylab = "Frequency")
mtext("Merchant Longitude", side = 1)
```



Univariate Analysis of merch_lat

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$merch_lat))
```

```
## [1] 0
```

Summary of merch_lat Column

```
summary(fraudTotal.db$merch_lat)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   19.03   34.74   39.37   38.54   41.96   67.51
```

```
class(fraudTotal.db$merch_lat)
```

```
## [1] "numeric"
```

Find the Standard Deviation and Variance of merch_lat variable

```
sd(fraudTotal.db$merch_lat)
```

```
## [1] 5.105604
```

```
var(fraudTotal.db$merch_lat)
```

```
## [1] 26.06719
```

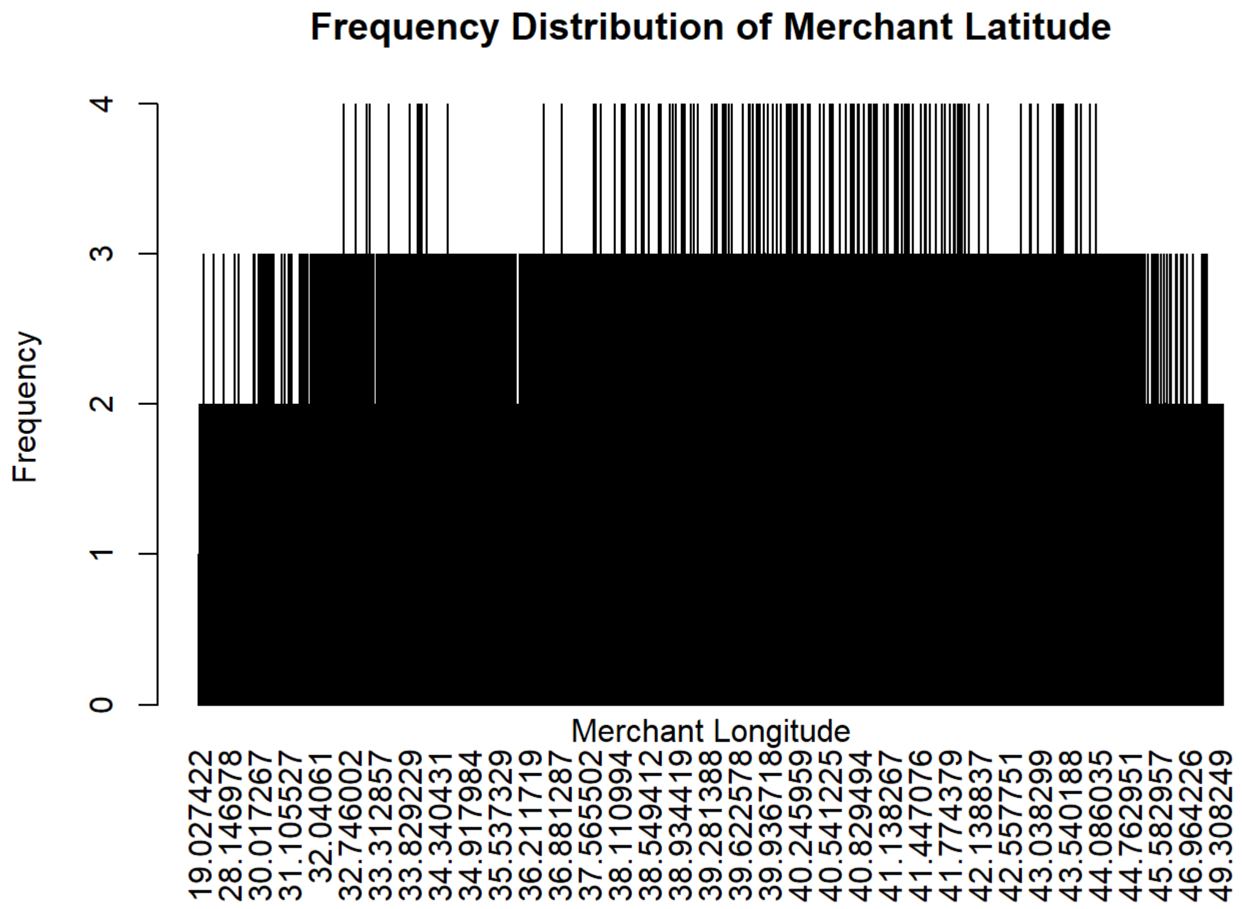
Frequency of merch_lat

```
table_merch_lat <- table(fraudTotal.db$merch_lat)
head(table_merch_lat)
```

```
## 
## 19.027422 19.027785 19.027804 19.027849 19.029798 19.031242 
##         1         1         1         1         1         1
```

Frequency Distribution of lat

```
barplot(table(fraudTotal.db$merch_lat), las = 3, main = "Frequency Distribution of Merchant Lati
tude", xlab = "", ylab = "Frequency")
mtext("Merchant Longitude", side = 1)
```



Univariate Analysis of is_fraud (Dependent Variable)

Checking to see if any NA values exist

```
sum(is.na(fraudTotal.db$is_fraud))
```

```
## [1] 0
```

Summary of is_fraud Column

```
summary(fraudTotal.db$is_fraud)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
## 0.00000 0.00000 0.00000 0.00521 0.00000 1.00000
```

```
class(fraudTotal.db$is_fraud)
```

```
## [1] "integer"
```

Find the Standard Deviation and Variance of cc_num variable

```
sd(fraudTotal.db$is_fraud)
```

```
## [1] 0.07199217
```

```
var(fraudTotal.db$is_fraud)
```

```
## [1] 0.005182873
```

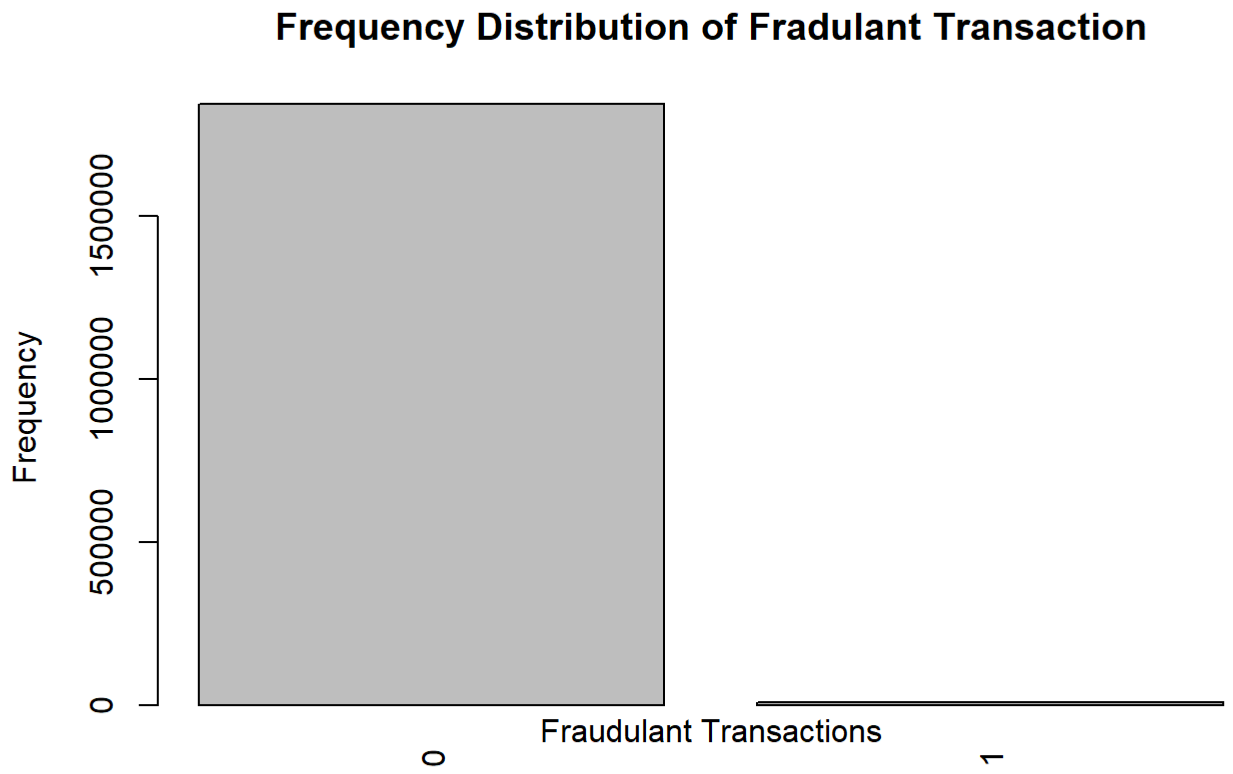
Frequency of is_fraud

```
table(fraudTotal.db$is_fraud)
```

```
## 
##       0       1 
## 1842743    9651
```

Frequency Distribution of is_fraud

```
barplot(table(fraudTotal.db$is_fraud), las = 3, main = "Frequency Distribution of Fradulant Tran
saction", xlab = "", ylab = "Frequency")
mtext("Fraudulant Transactions", side = 1)
```


Frequency Distribution of Fradulant Transaction
Frequency
0
500000
1000000
1500000
Fraudulant Transactions
0
1