

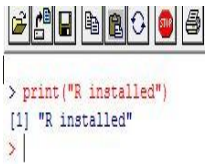
CMPE 272

Name: Arpita Dixit

Fraud Detection Using R language

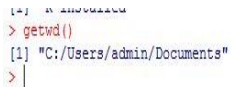
Link to Github Repo: https://github.com/ArpitaDixit/Fraud_Detection_Using-R

1. Install R using the below link
<https://cran.r-project.org/bin/windows/base/>
2. Open R and check if successfully installed.

A screenshot of the R console window. The title bar shows standard Windows icons. The console text is as follows:

```
> print("R installed")  
[1] "R installed"  
> |
```

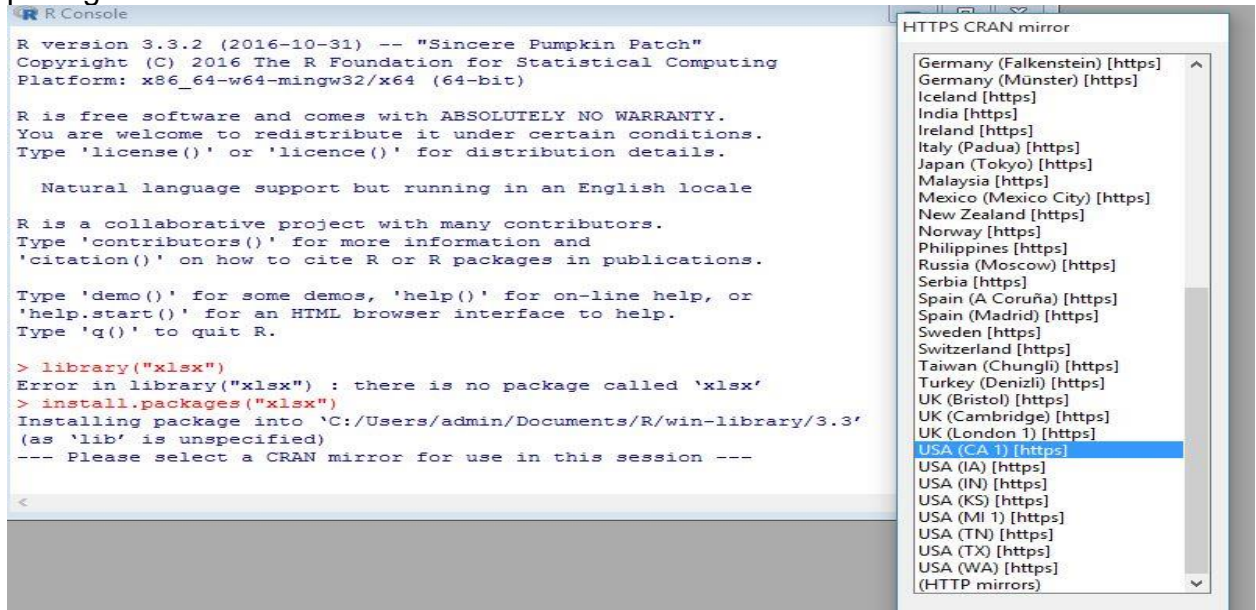
3. Download Excel file from the below link
<https://catalog.data.gov/dataset/fair-market-rents-for-the-section-8-housing-assistance-payments-program>
4. Get R's working directory using command getwd()

A screenshot of the R console window showing the output of the getwd() command.

```
[4] R session  
> getwd()  
[1] "C:/Users/admin/Documents"  
> |
```

5. Rename file as Rentdata.xlsx and save the xls file in the working directory.

6. Install xlsx Package using command `install.packages("xlsx")`. Select Appropriate CRAN package



```
6 Vs pz 104 1155 UNKN
> install.packages("xlsx")
Installing package into 'C:/Users/admin/Documents/R/win-library/3.3'
(as 'lib' is unspecified)
trying URL 'https://cran.cnr.berkeley.edu/bin/windows/contrib/3.3/xlsx_0.5.7.zip'
Content type 'application/zip' length 401439 bytes (392 KB)
downloaded 392 KB

package 'xlsx' successfully unpacked and MD5 sums checked
Warning: cannot remove prior installation of package 'xlsx'

The downloaded binary packages are in
C:/Users/admin/AppData/Local/Temp/RtmpkxX5Rx/downloaded_packages
> |
```

7. Verify if packages are installed
`any(grepl("xlsx", installed.packages()))`

```
Warning: package 'xlsx' is in use and will not be installed
> any(grepl("xlsx", installed.packages()))
[1] TRUE
> |
```

8. `library("xlsx")`
Loads required packages like rJava, methods, xlsxjars
9. Read Data from Excel
`rentdata <- read.xlsx("Rentdata.xlsx", sheetIndex = 1)`

10. Read first 5 rows of data using command head(rentdata)

```
> head(rentdata)
      fips2000 fips2010 fmr2 fmr0 fmr1 fmr3 fmr4 county State CouSub countyname Metro_code Areaname county_town_name pop2010
1 0100199999 0100199999 788 628 663 1084 1288 1 1 99999 Autauga County METRO33860M33860 Montgomery, AL MSA Autauga County 54571
2 0100399999 0100399999 762 494 643 1123 1318 3 1 99999 Baldwin County NCNTY01003N01003 Baldwin County, AL Baldwin County 182265
3 0100599999 0100599999 670 492 495 834 895 5 1 99999 Barbour County NCNTY01005N01005 Barbour County, AL Barbour County 27457
4 0100799999 0100799999 773 545 652 1015 1142 7 1 99999 Bibb County METRO13820M13820 Birmingham-Hoover, AL HUD Metro FMR Area Bibb County 22915
5 0100999999 0100999999 773 545 652 1015 1142 9 1 99999 Blount County METRO13820M13820 Birmingham-Hoover, AL HUD Metro FMR Area Blount County 57322
6 0101199999 0101199999 599 481 505 791 1061 11 1 99999 Bullock County NCNTY01011N01011 Bullock County, AL Bullock County 10914

state_alpha fmr_type metro
1 AL 40 1
2 AL 40 0
3 AL 40 0
4 AL 40 1
5 AL 40 1
6 AL 40 0
> |
```

11. Install DMwR Package using command install.packages("DMwR"). Select Appropriate CRAN package

12. library("DMwR")

Loads required packages

```
> library(DMwR)
Loading required package: lattice
Loading required package: grid
> |
```

13. Exploring the data set summary(rentdata)

```
> summary(rentdata)
      fips2000 fips2010 fmr2 fmr0 fmr1 fmr3 fmr4 county State CouSub
2300901185: 2 0100199999: 1 Min. : 405.0 Min. : 314.0 Min. : 327.0 Min. : 519 Min. : 570 Min. : 1.00 Min. : 1.00 99999 :3161
2301301465: 2 0100399999: 1 1st Qu.: 643.0 1st Qu.: 453.0 1st Qu.: 502.0 1st Qu.: 860 1st Qu.: 951 1st Qu.: 13.00 1st Qu.:21.00 02760 : 3
2301501010: 2 0100599999: 1 Median : 715.0 Median : 518.0 Median : 572.0 Median : 955 Median :1124 Median : 37.00 Median :27.00 04825 : 3
0100199999: 1 0100799999: 1 Mean : 806.6 Mean : 568.8 Mean : 639.4 Mean :1065 Mean :1213 Mean : 74.14 Mean :30.36 06260 : 3
0100399999: 1 0100999999: 1 3rd Qu.: 924.0 3rd Qu.: 646.0 3rd Qu.: 735.0 3rd Qu.:1199 3rd Qu.:1407 3rd Qu.:105.00 3rd Qu.:42.00 11800 : 3
(Other) :4760 0101199999: 1 Max. :2062.0 Max. :1291.0 Max. :1635.0 Max. :2801 Max. :3386 Max. :840.00 Max. :78.00 00380 : 2
NA's : 1 (Other) :4763 NA's :1 (Other) :1594

countyname Metro_code Areaname county_town_name pop2010
Washington County: 105 METRO14460MM11207: 114 Boston-Cambridge-Quincy, MA-NH HUD Metro FMR Area : 114 Washington County: 27 Min. : 0
Franklin County : 89 NCNTY23003N23003: 73 Aroostook County, ME : 73 Jefferson County : 25 1st Qu.: 4017
Aroostook County : 73 METRO12620N23019: 53 Penobscot County, ME (part) HUD Metro FMR Area : 53 Jackson County : 23 Median : 15034
Middlesex County : 71 METRO25540M25540: 51 Hartford-West Hartford-East Hartford, CT HUD Metro FMR Area: 51 Lincoln County : 22 Mean : 65627
Penobscot County : 67 NCNTY23029N23029: 49 Washington County, ME : 49 Franklin County : 21 3rd Qu.: 41440
Worcester County : 61 METRO44140M44140: 44 Springfield, MA HUD Metro FMR Area : 44 Madison County : 19 Max. :9818605
(Other) :4303 (Other) :4385 (Other) :4385 (Other) :4632 NA's :2

state_alpha fmr_type metro
ME : 535 Min. :40.00 Min. :0.0000
MA : 351 1st Qu.:40.00 1st Qu.:0.0000
NH : 259 Median :40.00 Median :0.0000
VT : 255 Mean :40.34 Mean :0.4013
TX : 254 3rd Qu.:40.00 3rd Qu.:1.0000
CT : 169 Max. :50.00 Max. :1.0000
(Other):2946
> |
```

14. Nlevel mismatch for fips2000 and fips2010

```
nlevels(rentdata$fips2000)
```

```
nlevels(rentdata$fips2010)
```

```
> nlevels(rentdata$fips2000)
[1] 4765
> nlevels(rentdata$fips2010)
[1] 4769
> |
```

15. Install Hmisc Package using command `install.packages("Hmisc")`. Select Appropriate CRAN package

16. `library('Hmisc')`

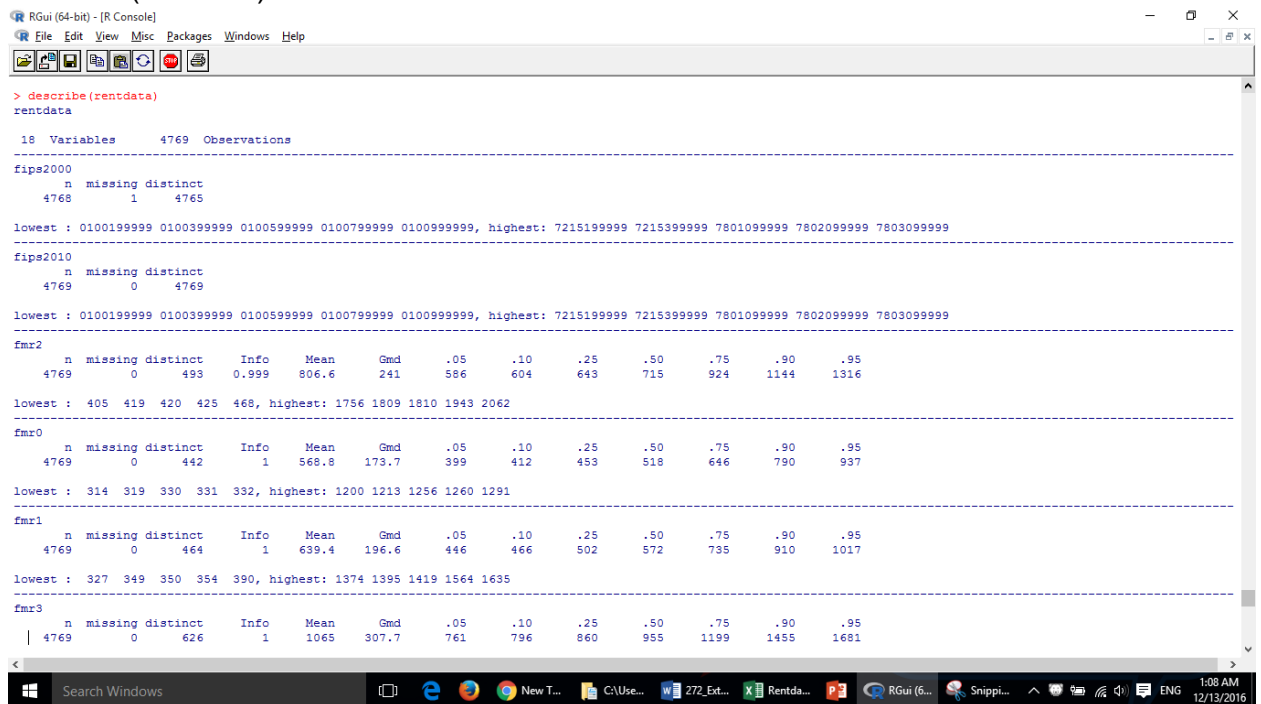
```
> library(Hmisc)
Loading required package: survival
Loading required package: Formula
Loading required package: ggplot2

Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

    format.pval, round.POSIXt, trunc.POSIXt, units
```

17. `describe(rentdata)`



```
> describe(rentdata)
rentdata

  18 Variables      4769 Observations
-----
fips2000
  n missing distinct
  4769      1      4765
lowest : 0100199999 0100399999 0100599999 0100799999 0100999999, highest: 7215199999 7215399999 7801099999 7802099999 7803099999
-----
fips2010
  n missing distinct
  4769      0      4769
lowest : 0100199999 0100399999 0100599999 0100799999 0100999999, highest: 7215199999 7215399999 7801099999 7802099999 7803099999
-----
fmr2
  n missing distinct      Info      Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
  4769      0      493      0.999      806.6      241      586      604      643      715      924      1144      1316
lowest : 405 419 420 425 468, highest: 1756 1809 1810 1943 2062
-----
fmr0
  n missing distinct      Info      Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
  4769      0      442      1      568.8      173.7      399      412      453      518      646      790      937
lowest : 314 319 330 331 332, highest: 1200 1213 1256 1260 1291
-----
fmr1
  n missing distinct      Info      Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
  4769      0      464      1      639.4      196.6      446      466      502      572      735      910      1017
lowest : 327 349 350 354 390, highest: 1374 1395 1419 1564 1635
-----
fmr3
  n missing distinct      Info      Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
  4769      0      626      1      1065      307.7      761      796      860      955      1199      1455      1681
```

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

fmr3
  n missing distinct Info Mean Gmd .05 .10 .25 .50 .75 .90 .95
4769 0 626 1 1065 307.7 761 796 860 955 1199 1455 1681

lowest : 519 527 560 599 630, highest: 2263 2420 2551 2667 2801

fmr4
  n missing distinct Info Mean Gmd .05 .10 .25 .50 .75 .90 .95
4769 0 737 1 1213 369.4 830 863 951 1124 1407 1693 1953

lowest : 570 615 660 718 725, highest: 2716 2892 3010 3061 3386

county
  n missing distinct Info Mean Gmd .05 .10 .25 .50 .75 .90 .95
4768 1 329 0.999 74.14 86.3 3.0 5.0 13.0 37.0 105.0 169.0 230.6

lowest : 1 3 5 6 7, highest: 800 810 820 830 840

State
  n missing distinct Info Mean Gmd .05 .10 .25 .50 .75 .90 .95
4769 0 54 0.997 30.36 16.96 6.0 9.8 21.0 27.0 42.0 50.0 53.0

lowest : 1 2 4 5 6, highest: 55 56 66 72 78

CouSub
  n missing distinct
4769 0 1533

lowest : 00100 00170 00260 00275 00325, highest: 87845 87910 87985 88190 99999

countyname
  n missing distinct
4769 0 1962

lowest : Añasco Municipio Abbeville County Acadia Parish Accomack County Ada County
highest: Yukon-Koyukuk Census Area Yuma County Zapata County Zavala County Ziebach County

Metro_code
  n missing distinct
4769 0 2572

lowest : METRO10180M10180 METRO10380M10380 METRO10420M10420 METRO10500M10500 METRO10580M10580
highest: NCNTY56045N56045 NCNTY66010N66010 NCNTY72923N72923 NCNTY78010N78010 NCNTY78923N78923

Areaname
  n missing distinct
4769 0 2572

lowest : Abbeville County, SC Abilene, TX MSA Acadia Parish, LA Accomack County, VA Adair County, IA
highest: Yuma County, CO Yuma, AZ MSA Zapata County, TX Zavala County, TX Ziebach County, SD

county_town_name
  n missing distinct
4769 0 3176

lowest : Añasco Municipio Abbeville County Abbot town Abington town Acadia Parish
highest: Yukon-Koyukuk Census Area Yuma County Zapata County Zavala County Ziebach County

pop2010
  n missing distinct Info Mean Gmd .05 .10 .25 .50 .75 .90 .95
4767 2 4467 1 65627 104327 424.3 1002.0 4017.0 15034.0 41440.0 122631.0 255519.2

lowest : 0 1 2 3 4, highest: 3095313 3817117 4092459 5194675 9818605

state_alpha
  n missing distinct
4769 0 54

lowest : AK AL AR AZ CA, highest: VT WA WI WV WY

fmr_type
  n missing distinct Info Mean Gmd
4769 0 2 0.098 40.34 0.6564

Value 40 50
```

```

RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

4769      0      2572

lowest : Abbeville County, SC Abilene, TX MSA      Acadia Parish, LA      Accomack County, VA      Adair County, IA
highest: Yuma County, CO      Yuma, AZ MSA      Zapata County, TX      Zavala County, TX      Ziebach County, SD
-----
county_town_name
  n missing distinct
4769      0      3176

lowest : Añasco Municipio      Abbeville County      Abbot town      Abington town      Acadia Parish
highest: Yukon-Koyukuk Census Area Yuma County      Zapata County      Zavala County      Ziebach County
-----
pop2010
  n missing distinct      Info      Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
4767      2      4467      1      65627      104327      424.3      1002.0      4017.0      15034.0      41440.0      122631.0      255519.2

lowest :      0      1      2      3      4, highest: 3095313 3817117 4092459 5194675 9818605
-----
state_alpha
  n missing distinct
4769      0      54

lowest : AK AL AR AZ CA, highest: VT WA WI WV WY
-----
fmr_type
  n missing distinct      Info      Mean      Gmd
4769      0      2      0.098      40.34      0.6564

Value      40      50
Frequency      4607      162
Proportion 0.966 0.034
-----
metro
  n missing distinct      Info      Sum      Mean      Gmd
4769      0      2      0.721      1914      0.4013      0.4806

> |

```

18. As seen in the description screen shots above there are missing values in fips2000, county, pop2010.

19. Missing pair of all possible combinations of the 3

```
length(which(is.na(rentdata$fips2000) & is.na(rentdata$county) &
is.na(rentdata$pop2010)))
```

```
length(which(is.na(rentdata$fips2000) & is.na(rentdata$county)))
```

```
length(which(is.na(rentdata$fips2000) & is.na(rentdata$pop2010)))
```

```
length(which(is.na(rentdata$county) & is.na(rentdata$pop2010)))
```

```

> length(which(is.na(rentdata$fips2000) & is.na(rentdata$county) & is.na(rentdata$pop2010)))
[1] 0
> length(which(is.na(rentdata$fips2000) & is.na(rentdata$county)))
[1] 1
> length(which(is.na(rentdata$fips2000) & is.na(rentdata$pop2010)))
[1] 0
> length(which(is.na(rentdata$county) & is.na(rentdata$pop2010)))
[1] 0
> |

```


20. Proportion on basis of metro, frm_type, county, state

```
table(rentdata$metro)/nrow(rentdata)*100
table(rentdata$frm_type)/nrow(rentdata)*100
table(rentdata$county)/nrow(rentdata)*100
```

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

> table(rentdata$metro)/nrow(rentdata)*100
 0      1 
59.8658 40.1342 
> table(rentdata$frm_type)/nrow(rentdata)*100
 40      50 
96.603061 3.396939 
> table(rentdata$metro)/nrow(rentdata)*100
 0      1 
59.8658 40.1342 
> table(rentdata$county)/nrow(rentdata)*100

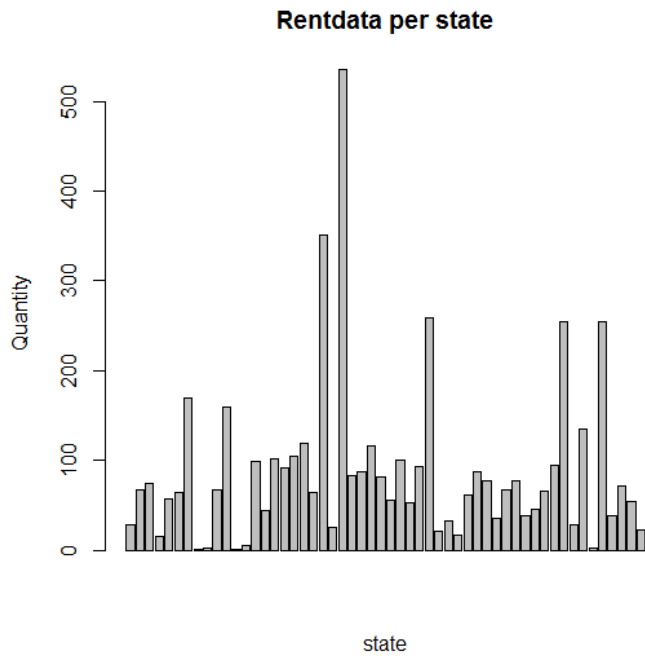
 1      3      5      6      7      9     10     11     12     13     14     15     16     17     19 
2.78884462 4.59215769 3.41790732 0.02096876 3.52275110 4.46634515 0.04193751 3.45984483 0.02096876 2.74690711 0.02096876 3.04046970 0.02096876 3.45984483 3.01950094 
20      21      23      25      27      28      29      30      31      33      35      36      37      39      41 
0.04193751 2.53721954 2.09687566 2.18075068 3.18725100 0.04193751 1.88718809 0.02096876 1.46781296 0.85971902 0.83875026 0.02096876 0.83875026 0.81778151 0.81778151 
43      45      47      49      50      51      53      54      55      56      57      59      60      61      63 
0.81778151 0.81778151 0.79681275 0.77584399 0.02096876 0.77584399 0.77584399 0.02096876 0.75487524 0.02096876 0.75487524 0.02096876 0.75487524 0.73390648 
65      67      68      69      70      71      73      75      77      78      79      81      83      85      86 
0.73390648 0.73390648 0.02096876 0.73390648 0.02096876 0.73390648 0.71293772 0.71293772 0.02096876 0.69196897 0.69196897 0.69196897 0.69196897 0.02096876 
87      89      90      91      93      95      97      99     100     101     103     105     107     109     110 
0.69196897 0.67100021 0.02096876 0.67100021 0.65003145 0.65003145 0.65003145 0.65003145 0.02096876 0.65003145 0.65003145 0.67100021 0.62906270 0.02096876 
111     113     115     117     119     121     122     123     125     127     129     130     131     133     135 
0.60809394 0.58712518 0.58712518 0.56615643 0.56615643 0.56615643 0.02096876 0.54518767 0.54518767 0.52421891 0.48228140 0.02096876 0.48228140 0.44034389 
137     139     141     143     145     147     149     150     151     153     155     157     159     161     163 
0.44034389 0.41937513 0.41937513 0.39840637 0.39840637 0.39840637 0.39840637 0.02096876 0.35646886 0.37743762 0.33550010 0.33550010 0.33550010 0.33550010 
164     165     167     169     170     171     173     175     177     179     180     181     183     185     186 
0.02096876 0.31453135 0.29356259 0.29356259 0.02096876 0.29356259 0.29356259 0.27259384 0.25162508 0.25162508 0.02096876 0.25162508 0.25162508 0.02096876 
187     188     189     191     193     195     197     198     199     201     203     205     207     209     211 
0.20968757 0.02096876 0.18871881 0.16775005 0.16775005 0.02096876 0.18871881 0.02096876 0.16775005 0.12581254 0.10484378 0.10484378 0.10484378 0.08387503 
213     215     217     219     220     221     223     225     227     229     230     231     233     235     237 
0.08387503 0.08387503 0.08387503 0.08387503 0.02096876 0.08387503 0.08387503 0.08387503 0.08387503 0.02096876 0.06290627 0.06290627 0.06290627 0.06290627 
239     240     241     243     245     247     249     251     253     255     257     259     261     263     265 
0.06290627 0.02096876 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 
267     269     270     271     273     275     277     279     281     282     283     285     287     289     290 
0.04193751 0.04193751 0.02096876 0.04193751 0.04193751 0.06290627 0.04193751 0.04193751 0.04193751 0.02096876 0.04193751 0.04193751 0.04193751 0.02096876 
0.04193751 0.04193751 0.02096876 0.04193751 0.04193751 0.06290627 0.04193751 0.04193751 0.04193751 0.02096876 0.04193751 0.04193751 0.04193751 0.04193751 0.02096876
```

```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

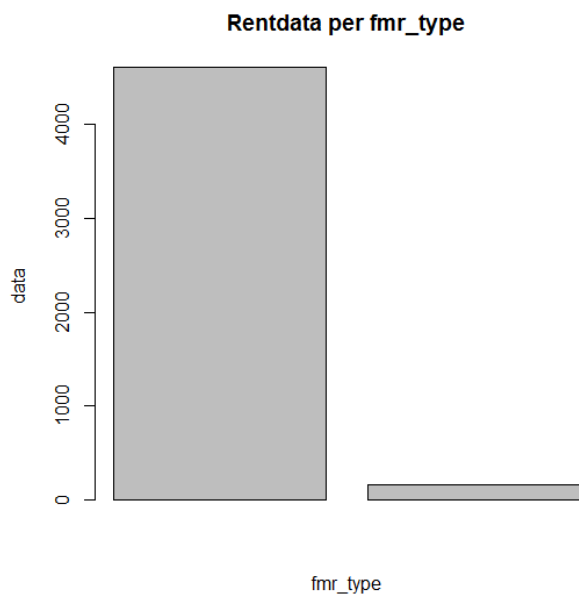
0.08387503 0.08387503 0.08387503 0.08387503 0.02096876 0.08387503 0.08387503 0.08387503 0.08387503 0.08387503 0.02096876 0.06290627 0.06290627 0.06290627 0.06290627 
239      240      241      243      245      247      249      251      253      255      257      259      261      263      265 
0.06290627 0.02096876 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.06290627 0.06290627 0.04193751 0.04193751 
267      269      270      271      273      275      277      279      281      282      283      285      287      289      290 
0.04193751 0.04193751 0.02096876 0.04193751 0.04193751 0.06290627 0.04193751 0.04193751 0.04193751 0.02096876 0.04193751 0.04193751 0.04193751 0.04193751 0.02096876 
291      293      295      297      299      301      303      305      307      309      311      313      315      317      319 
0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.04193751 0.02096876 
321      323      325      327      329      331      333      335      337      339      341      343      345      347      349 
0.04193751 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
351      353      355      357      359      361      363      365      367      371      373      375      377      379      381 
0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
381      383      385      387      389      391      393      395      397      399      401      403      405      407      409 
0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
411      413      415      417      419      421      423      425      427      429      431      433      435      437      439 
0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
441      443      445      447      449      451      453      455      457      459      461      463      465      467      469 
0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
471      473      475      477      479      481      483      485      487      489      491      493      495      497      499 
0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
501      503      505      507      510      515      520      530      540      550      560      570      580      590      595 
0.02096876 0.02096876 0.02096876 0.02096876 0.08387503 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
600      610      620      630      640      650      660      670      680      683      685      690      700      710      720 
0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
720      730      735      740      750      760      770      775      790      800      810      820      830      840      840 
0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 0.02096876 
> table(rentdata$state_alpha)/nrow(rentdata)*100
      AK      AL      AR      AZ      CA      CO      CT      DC      DE      FL      GA      GU      HI 
0.60809394 1.40490669 1.57265674 0.31453135 1.21618788 1.34200042 3.54371986 0.02096876 0.06290627 1.40490669 3.33403229 0.02096876 0.10484378 
      IA      IL      IN      KS      KY      LA      MA      MD      ME      MI      MN      MO      NE      NH      NJ      NY      NV      OH      OK      OR 
2.07590690 0.92262529 2.13881317 1.92912560 2.20171944 2.51625079 1.34200042 3.76003355 0.52421891 11.21828476 1.74040679 1.82428182 2.43237576 0.75487524 0.81778151 1.50975047 
      PA      PR      RI      SC      SD      TN      TX      UT      VA      VT      WI      WY 
1.71943804 1.17425037 2.09687566 1.11134410 1.95009436 5.43090795 0.44034389 0.69196897 0.35646886 1.30006291 1.84525058 1.61459425 0.75487524 0.81778151 1.50975047 
      WV 
1.40490669 1.63556301 0.81778151 0.96456280 1.38393793 1.99203187 5.32606416 0.60809394 2.83078213 0.06290627 5.34703292 0.81778151 1.50975047 
1.15328161 0.48228140
```

21. Variability in figure

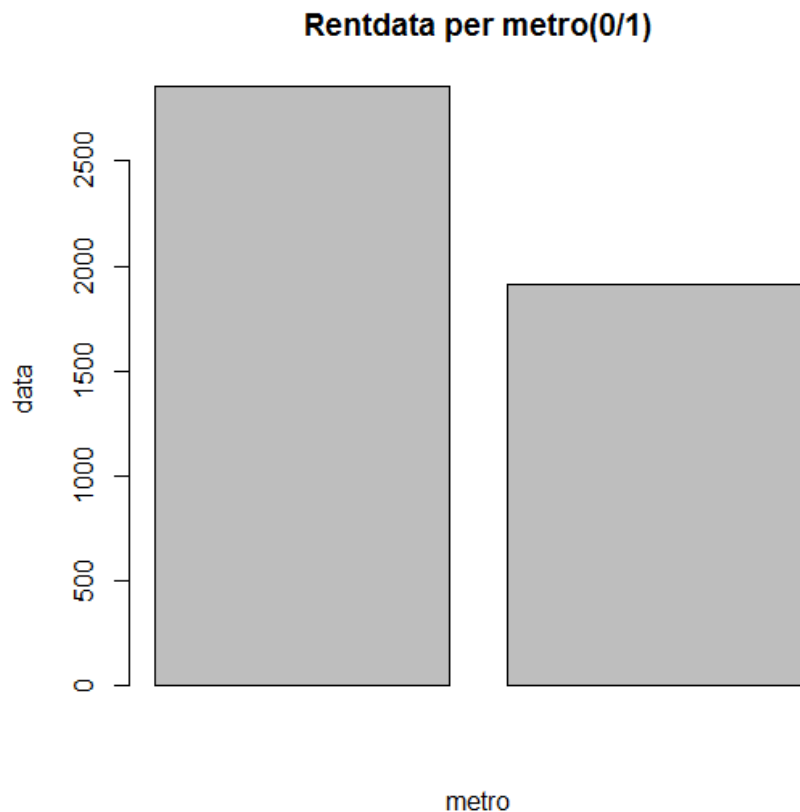
a. `totP<-table(rentdata$state)`
`barplot(totP,main="Rentdata per state",`
`names.arg="",xlab='state',ylab='Quantity')`



b. `totS<-table(rentdata$fmr_type)`
`barplot(totS,main="Rentdata per fmr_type",`
`names.arg="",xlab='fmr_type',ylab='data')`




```
c. totQ<-table(rentdata$metro)
  barplot(totQ,main="Rentdata per metro(0/1)",
  names.arg="",xlab='metro',ylab='data')
```



22. New column metro per fmr_type

```
rentdata$Uperstate<-rentdata$metro/rentdata$fmr_type
> rentdata$Uperstate<-rentdata$metro/rentdata$fmr_type
> summary(rentdata$Uperstate)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000000	0.000000	0.000000	0.009864	0.025000	0.025000

```
> |
```

23. attach(rentdata)

```
> attach(rentdata)
The following objects are masked from rentdata (pos = 3):

Areaname, county, county_town_name, countyname, CouSub, fips2000, fips2010, fmr_type, fmr0, fmr1, fmr2,
fmr3, fmr4, metro, Metro_code, pop2010, State, state_alpha

> |
```

24. Most and Least Occurring States

```
upp <- aggregate(Uperstate,list(State),median,na.rm=T)
topP <- sapply(c(T,F),function(o) upp[order(upp[,2],decreasing=o)[1:5],1])
colnames(topP) <- c('StateMost','StateLeast')
topP
```

```
> topP
      StateMost StateLeast
[1,]         6          1
[2,]        11          2
[3,]        12          4
[4,]        25          5
[5,]        34          8
> |
```

25. Convert to factor

```
tops$State <- factor(tops$State)
```

26. Top and Bottom fips2000

```
vs <- aggregate(State,list(fips2000),sum,na.rm=T)
S1 <- sapply(c(T,F),function(o) vs[order(vs$x,decreasing=o)[1:5],1])
colnames(S1) <- c('Most','Least')
S1
```

```
> S1
      Most      Least
[1,] "7801099999" "0100199999"
[2,] "7802099999" "0100399999"
[3,] "7803099999" "0100599999"
[4,] "7200199999" "0100799999"
[5,] "7200399999" "0100999999"
> |
```

27. $\text{sum}(vs[\text{order}(vs\$x, \text{decreasing}=T)[1:100], 2]) / \text{sum}(\text{State}, na.rm=T) * 100$

```
> sum(vs[order(vs$x,decreasing=T)[1:100],2])/sum(State,na.rm=T)*100
[1] 4.782263
> |
```

28. $\text{sum}(vs[\text{order}(vs\$x, \text{decreasing}=F)[1:2000], 2]) / \text{sum}(\text{State}, na.rm=T) * 100$

```
[1] 4.782263
> sum(vs[order(vs$x,decreasing=F)[1:2000],2])/sum(State,na.rm=T)*100
[1] 23.06869
> |
```

29. Data per fmr_type

```
totS <- table(fmr_type)
> totS <- table(fmr_type)
> totS
fmr_type
  40   50
4607 162
```

30. Data per metro

```
100 / 102  
> totP <- table(metro)  
> totP  
metro  
  0    1  
2855 1914  
> |
```