DATA607_Project2

Pei-Ming Chen

2023-03-05

```
library(tidyverse)
library(openintro)
library(dplyr)
library(ggplot2)
```

Dataset- Income

The database contains 32,000 records on US Household Income Statistics & Geo Locations. The field description of the database is available on Kaggle.com. Income is a vital element when determining both quality and socioeconomic features of a given geographic location. The uploaded data was derived from over +36,000 files and covers 348,893 location records.

Importing the Data

```
HI <- "https://raw.githubusercontent.com/PMCformosa/DATA607_Project-2/main/kaggle_income.csv"

Household_Income <- read.csv(file = HI, header =TRUE, sep=",")
head(Household_Income)
```

```
##        id State_Code State_Name State_ab         County           City
## 1 1011000          1    Alabama       AL  Mobile County      Chickasaw
## 2 1011010          1    Alabama       AL Barbour County     Louisville
## 3 1011020          1    Alabama       AL  Shelby County     Columbiana
## 4 1011030          1    Alabama       AL  Mobile County        Satsuma
## 5 1011040          1    Alabama       AL  Mobile County Dauphin Island
## 6 1011050          1    Alabama       AL Cullman County        Cullman
##             Place Type Primary Zip_Code Area_Code    ALand    AWater      Lat
## 1  Chickasaw city City   place    36611       251 10894952    909156 30.77145
## 2       Clio city City   place    36048       334 26070325     23254 31.70852
## 3 Columbiana city City   place    35051       205 44835274    261034 33.19145
## 4     Creola city City   place    36572       251 36878729   2374530 30.87434
## 5  Dauphin Island Town   place    36528       251 16204185 413605152 30.25091
## 6      Dodge City Town   place    35057       256  8913021     26837 34.04541
##         Lon  Mean Median Stdev     sum_w
## 1 -88.07970 38773  30506 33101 1638.2605
## 2 -85.61104 37725  19528 43789  258.0177
## 3 -86.61562 54606  31930 57348  926.0310
## 4 -88.00944 63919  52814 47707  378.1146
## 5 -88.17127 77948  67225 54270  282.3203
## 6 -86.88267 50715  42643 35886  173.3260
```

About the Dataset The dataset was taken from Kaggle.com which was originally taken from an article called “US Household Income Statistics” Datasets https://www.kaggle.com/datasets/goldenoakresearch/us-household-income-stats-geo-locations

```
glimpse(Household_Income)
```

```
## Rows: 32,526
## Columns: 19
## $ id         <int> 1011000, 1011010, 1011020, 1011030, 1011040, 1011050, 10110~
## $ State_Code <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ State_Name <chr> "Alabama", "Alabama", "Alabama", "Alabama", "Alabama", "Ala~
## $ State_ab   <chr> "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL",~
## $ County     <chr> "Mobile County", "Barbour County", "Shelby County", "Mobile~
## $ City       <chr> "Chickasaw", "Louisville", "Columbiana", "Satsuma", "Dauphi~
## $ Place      <chr> "Chickasaw city", "Clio city", "Columbiana city", "Creola c~
## $ Type       <chr> "City", "City", "City", "City", "Town", "Town", "City", "To~
## $ Primary    <chr> "place", "place", "place", "place", "place", "place", "plac~
## $ Zip_Code   <int> 36611, 36048, 35051, 36572, 36528, 35057, 36426, 36020, 356~
## $ Area_Code  <chr> "251", "334", "205", "251", "251", "256", "251", "334", "25~
## $ ALand      <dbl> 10894952, 26070325, 44835274, 36878729, 16204185, 8913021, ~
## $ AWater     <dbl> 909156, 23254, 261034, 2374530, 413605152, 26837, 91015, 17~
## $ Lat        <dbl> 30.77145, 31.70852, 33.19145, 30.87434, 30.25091, 34.04541,~
## $ Lon        <dbl> -88.07970, -85.61104, -86.61562, -88.00944, -88.17127, -86.~
## $ Mean       <int> 38773, 37725, 54606, 63919, 77948, 50715, 33737, 46319, 579~
## $ Median     <int> 30506, 19528, 31930, 52814, 67225, 42643, 23610, 40242, 395~
## $ Stdev      <int> 33101, 43789, 57348, 47707, 54270, 35886, 28256, 38941, 472~
## $ sum_w      <dbl> 1638.26051, 258.01768, 926.03100, 378.11462, 282.32033, 173~
```

Data Wrangling

```
colnames(Household_Income)
```

```
##  [1] "id"         "State_Code" "State_Name" "State_ab"   "County"
##  [6] "City"       "Place"      "Type"       "Primary"    "Zip_Code"
## [11] "Area_Code"  "ALand"      "AWater"     "Lat"        "Lon"
## [16] "Mean"       "Median"     "Stdev"      "sum_w"
```

```
colnames(Household_Income) <- c("Location_ID", "State_Code","State_Name","Abb_State_name","County", "Ci
colnames(Household_Income)
```

```
##  [1] "Location_ID"    "State_Code"     "State_Name"     "Abb_State_name"
##  [5] "County"         "City "          "Geo_location"   "Type"
##  [9] "Primary"        "Zip_Code"       "Area_Code"      "Area_Square"
## [13] "Water_area"     "Mean_H_Income"  "Sd_H_Income"    "Mean"
## [17] "Median"         "Stdev"          "sum_w"
```

```
glimpse(Household_Income)
```

```
## Rows: 32,526
## Columns: 19
## $ Location_ID    <int> 1011000, 1011010, 1011020, 1011030, 1011040, 1011050, 1~
## $ State_Code     <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ State_Name     <chr> "Alabama", "Alabama", "Alabama", "Alabama", "Alabama", ~
## $ Abb_State_name <chr> "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "~
## $ County         <chr> "Mobile County", "Barbour County", "Shelby County", "Mo~
## $ ‘City ‘        <chr> "Chickasaw", "Louisville", "Columbiana", "Satsuma", "Da~
## $ Geo_location   <chr> "Chickasaw city", "Clio city", "Columbiana city", "Creo~
## $ Type           <chr> "City", "City", "City", "City", "Town", "Town", "City",~
## $ Primary        <chr> "place", "place", "place", "place", "place", "place", "~
## $ Zip_Code       <int> 36611, 36048, 35051, 36572, 36528, 35057, 36426, 36020,~
## $ Area_Code      <chr> "251", "334", "205", "251", "251", "256", "251", "334",~
## $ Area_Square    <dbl> 10894952, 26070325, 44835274, 36878729, 16204185, 89130~
## $ Water_area     <dbl> 909156, 23254, 261034, 2374530, 413605152, 26837, 91015~
## $ Mean_H_Income  <dbl> 30.77145, 31.70852, 33.19145, 30.87434, 30.25091, 34.04~
## $ Sd_H_Income    <dbl> -88.07970, -85.61104, -86.61562, -88.00944, -88.17127, ~
## $ Mean           <int> 38773, 37725, 54606, 63919, 77948, 50715, 33737, 46319,~
## $ Median         <int> 30506, 19528, 31930, 52814, 67225, 42643, 23610, 40242,~
## $ Stdev          <int> 33101, 43789, 57348, 47707, 54270, 35886, 28256, 38941,~
## $ sum_w          <dbl> 1638.26051, 258.01768, 926.03100, 378.11462, 282.32033,~
```

```
summary(Household_Income)
```

```
##   Location_ID         State_Code     State_Name        Abb_State_name
##  Min.   :     1026   Min.   : 1.00   Length:32526       Length:32526
##  1st Qu.:  8021282   1st Qu.:13.00   Class :character   Class :character
##  Median : 29011679   Median :29.00   Mode  :character   Mode  :character
##  Mean   : 62037073   Mean   :28.62
##  3rd Qu.: 48028986   3rd Qu.:42.00
##  Max.   :480221068   Max.   :72.00
##     County             City           Geo_location          Type
##  Length:32526       Length:32526       Length:32526       Length:32526
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    Primary             Zip_Code      Area_Code          Area_Square
##  Length:32526       Min.   :  601   Length:32526       Min.   :0.000e+00
##  Class :character   1st Qu.:26362   Class :character   1st Qu.:1.907e+06
##  Mode  :character   Median :48163   Mode  :character   Median :5.023e+06
##                     Mean   :50183                      Mean   :1.166e+08
##                     3rd Qu.:76712                      3rd Qu.:3.091e+07
##                     Max.   :99950                      Max.   :9.163e+10
##    Water_area        Mean_H_Income    Sd_H_Income           Mean
##  Min.   :0.000e+00   Min.   :17.93   Min.   :-175.86   Min.   :     0
##  1st Qu.:0.000e+00   1st Qu.:34.01   1st Qu.: -97.66   1st Qu.: 46016
##  Median :2.703e+04   Median :38.93   Median : -87.14   Median : 60738
##  Mean   :6.952e+06   Mean   :37.73   Mean   : -91.30   Mean   : 66704
##  3rd Qu.:5.082e+05   3rd Qu.:41.50   3rd Qu.: -79.85   3rd Qu.: 82224
##  Max.   :2.453e+10   Max.   :71.25   Max.   : -65.50   Max.   :242857
##      Median           Stdev            sum_w
##  Min.   :     0   Min.   :     0   Min.   :    0.0
##  1st Qu.: 36046   1st Qu.: 36075   1st Qu.:  201.4
##  Median : 51874   Median : 46179   Median :  329.5
##  Mean   : 85453   Mean   : 47274   Mean   :  576.9
##  3rd Qu.: 80915   3rd Qu.: 58078   3rd Qu.:   590.2
##  Max.   :300000   Max.   :113936   Max.   :612241.9
```

```
DF_H_Income <- Household_Income %>% as_tibble()

DF_H_Income
```

```
## # A tibble: 32,526 x 19
##    Locati~1 State~2 State~3 Abb_S~4 County ‘City ‘ Geo_l~5 Type  Primary Zip_C~6
##       <int>   <int> <chr>   <chr>   <chr>  <chr>   <chr>   <chr> <chr>     <int>
##  1  1011000       1 Alabama AL      Mobil~ Chicka~ Chicka~ City  place     36611
##  2  1011010       1 Alabama AL      Barbo~ Louisv~ Clio c~ City  place     36048
##  3  1011020       1 Alabama AL      Shelb~ Columb~ Columb~ City  place     35051
##  4  1011030       1 Alabama AL      Mobil~ Satsuma Creola~ City  place     36572
##  5  1011040       1 Alabama AL      Mobil~ Dauphi~ Dauphi~ Town  place     36528
##  6  1011050       1 Alabama AL      Cullm~ Cullman Dodge ~ Town  place     35057
##  7  1011060       1 Alabama AL      Escam~ East B~ East B~ City  place     36426
##  8  1011070       1 Alabama AL      Elmor~ Coosada Elmore  Town  place     36020
##  9  1011080       1 Alabama AL      Morga~ Eva     Eva     Town  place     35621
## 10  1011090       1 Alabama AL      Talla~ Sylaca~ Fayett~ CDP   place     35151
## # ... with 32,516 more rows, 9 more variables: Area_Code <chr>,
## #   Area_Square <dbl>, Water_area <dbl>, Mean_H_Income <dbl>,
## #   Sd_H_Income <dbl>, Mean <int>, Median <int>, Stdev <int>, sum_w <dbl>, and
## #   abbreviated variable names 1: Location_ID, 2: State_Code, 3: State_Name,
## #   4: Abb_State_name, 5: Geo_location, 6: Zip_Code
## # i Use ‘print(n = ...)‘ to see more rows, and ‘colnames()‘ to see all variable names
```

Reshaping dataset

Columns of County, Geo_location, and Type are collapsed into one column named Location and nother one with values on called District_type. Also , Area_Square, Water_area, Zip_Code are collapsed into the columns called Area and Area_size. The whole data frame looks much cleaner and consolidated.

```
DF_H_Income2 <- DF_H_Income %>%  pivot_longer(
    cols = c(County,Geo_location, Type, Primary),
    names_to = "Location",
    values_to = "District_type")
```

```
DF_H_Income3 <- DF_H_Income2 %>%  pivot_longer(
    cols = c(Area_Square, Water_area, Zip_Code),
    names_to = "Area",
    values_to = "Area_size")

DF_H_Income3
```

```
## # A tibble: 390,312 x 16
##    Locati~1 State~2 State~3 Abb_S~4 ‘City ‘ Area_~5 Mean_~6 Sd_H_~7  Mean Median
##       <int>   <int> <chr>   <chr>   <chr>   <chr>     <dbl>   <dbl> <int>  <int>
##  1  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  2  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  3  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  4  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  5  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  6  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  7  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  8  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
##  9  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
## 10  1011000       1 Alabama AL      Chicka~ 251        30.8   -88.1 38773  30506
## # ... with 390,302 more rows, 6 more variables: Stdev <int>, sum_w <dbl>,
## #   Location <chr>, District_type <chr>, Area <chr>, Area_size <dbl>, and
## #   abbreviated variable names 1: Location_ID, 2: State_Code, 3: State_Name,
## #   4: Abb_State_name, 5: Area_Code, 6: Mean_H_Income, 7: Sd_H_Income
## # i Use ‘print(n = ...)‘ to see more rows, and ‘colnames()‘ to see all variable names
```

Counting possible missing Values

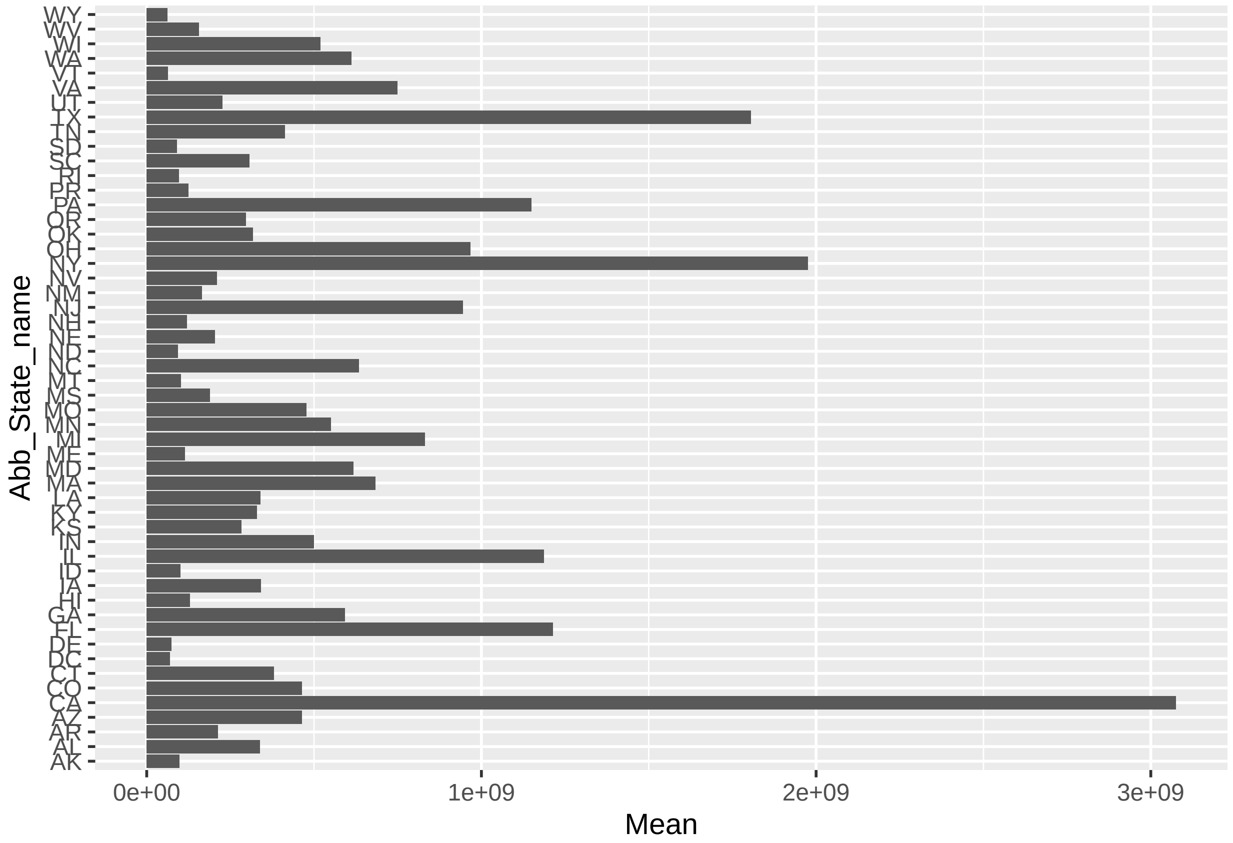
```
sum(is.na(DF_H_Income3))
```

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

Plot the graph of average income of each state

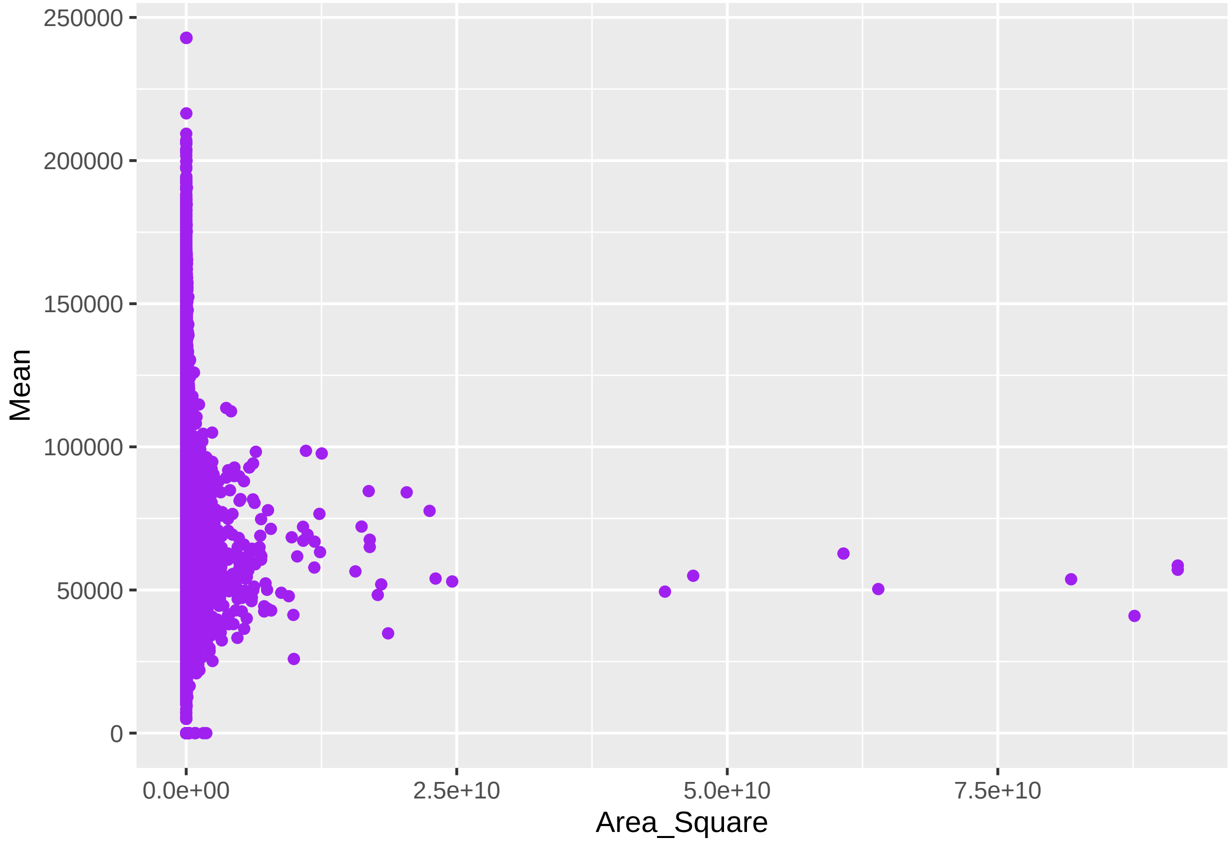
Significantl , the highest average income is observed in California state

```
ggplot(DF_H_Income3, aes(x=Mean, y=Abb_State_name)) + geom_col()
```



The relationship of size of Area and its income level is shown in the graph below.

```
Household_Income %>%
ggplot(aes(x = Area_Square, y = Mean)) +
  geom_point(color= "purple")
```



Puerto rico was chose to study here for its relatively low mean value in income . Its area-sdize and income relashionship is shown below.

```
PR_data <- DF_H_Income3[DF_H_Income3$State_Name == 'Puerto Rico' ,]
PR_data
```

```
## # A tibble: 4,560 x 16
##    Locati~1 State~2 State~3 Abb_S~4 ‘City ‘ Area_~5 Mean_~6 Sd_H_~7  Mean Median
##       <int>   <int> <chr>   <chr>   <chr>   <chr>     <dbl>   <dbl> <int>  <int>
##  1  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  2  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  3  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  4  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  5  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  6  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  7  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  8  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
##  9  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
## 10  7201587      72 Puerto~ PR      Aiboni~ 787        18.1   -66.3 22653  15565
## # ... with 4,550 more rows, 6 more variables: Stdev <int>, sum_w <dbl>,
## #   Location <chr>, District_type <chr>, Area <chr>, Area_size <dbl>, and
## #   abbreviated variable names 1: Location_ID, 2: State_Code, 3: State_Name,
## #   4: Abb_State_name, 5: Area_Code, 6: Mean_H_Income, 7: Sd_H_Income
## # i Use ‘print(n = ...)‘ to see more rows, and ‘colnames()‘ to see all variable names
```

```
summary(PR_data)
```

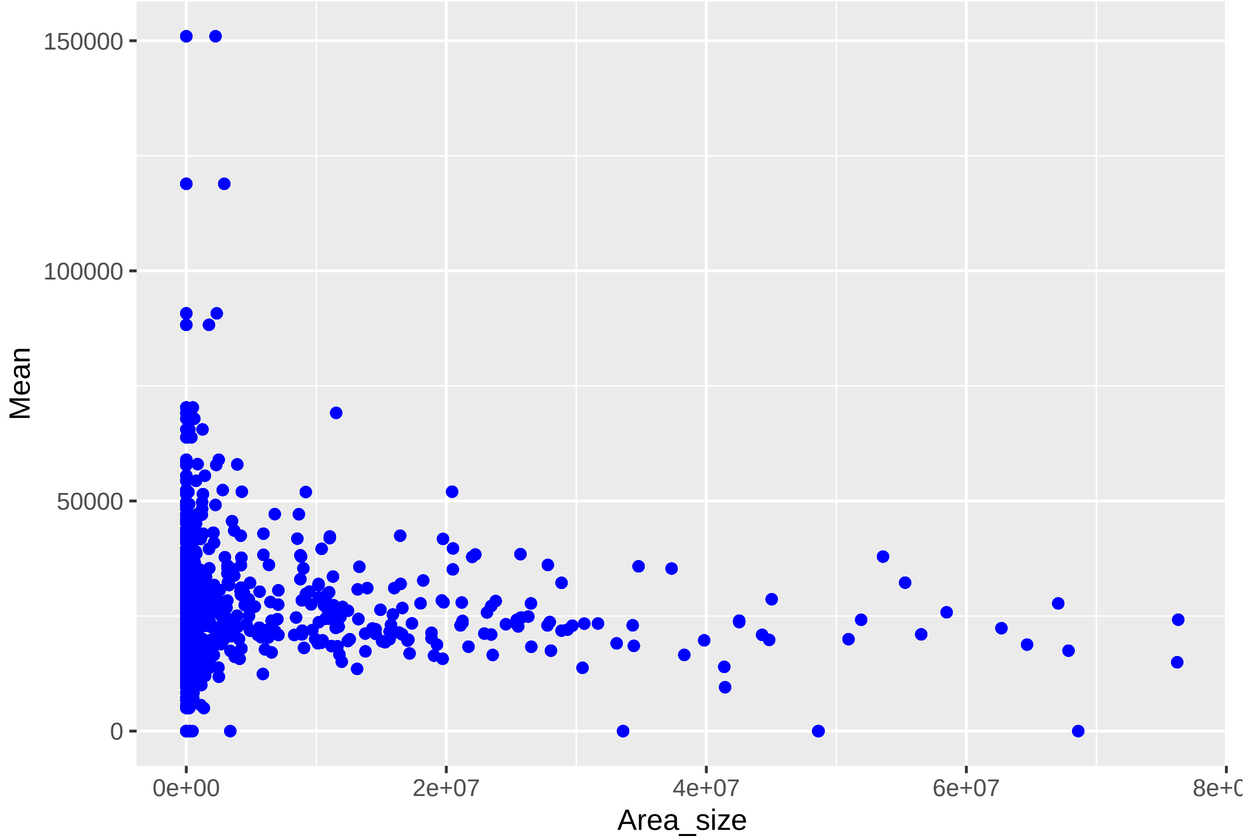
```
##   Location_ID         State_Code  State_Name        Abb_State_name    
##  Min.   :   72026   Min.   :72   Length:4560        Length:4560       
##  1st Qu.: 7202694   1st Qu.:72   Class :character   Class :character  
##  Median :72021641   Median :72   Mode  :character   Mode  :character  
##  Mean   :50357066   Mean   :72                                        
##  3rd Qu.:72022588   3rd Qu.:72                                        
##  Max.   :72023536   Max.   :72                                        
##     City             Area_Code         Mean_H_Income    Sd_H_Income    
##  Length:4560        Length:4560        Min.   :17.93   Min.   :-67.89  
##  Class :character   Class :character   1st Qu.:18.17   1st Qu.:-66.63  
##  Mode  :character   Mode  :character   Median :18.35   Median :-66.19  
##                                        Mean   :18.28   Mean   :-66.34  
##                                        3rd Qu.:18.41   3rd Qu.:-66.04  
##                                        Max.   :18.52   Max.   :-65.50  
##       Mean            Median           Stdev            sum_w        
##  Min.   :     0   Min.   :     0   Min.   :    0   Min.   :    0.0  
##  1st Qu.: 19608   1st Qu.: 13247   1st Qu.:17626   1st Qu.:  284.7  
##  Median : 24382   Median : 17896   Median :22801   Median :  452.4  
##  Mean   : 27256   Mean   : 22048   Mean   :24979   Mean   :  707.5  
##  3rd Qu.: 31604   3rd Qu.: 23083   3rd Qu.:29373   3rd Qu.:  722.8  
##  Max.   :150971   Max.   :300000   Max.   :80307   Max.   :44599.5  
##    Location          District_type          Area             Area_size       
##  Length:4560        Length:4560        Length:4560        Min.   :       0  
##  Class :character   Class :character   Class :character   1st Qu.:     719  
##  Mode  :character   Mode  :character   Mode  :character   Median :    4614  
##                                                           Mean   : 3697950  
##                                                           3rd Qu.: 1520370  
##                                                           Max.   :76299498  
```

```
PR_dataB <- PR_data[order(PR_data$Mean, decreasing = TRUE),]

PR3 <- select(PR_dataB, "Mean")
PR3
```

```
## # A tibble: 4,560 x 1
##      Mean
##     <int>
##  1 150971
##  2 150971
##  3 150971
##  4 150971
##  5 150971
##  6 150971
##  7 150971
##  8 150971
##  9 150971
## 10 150971
## # ... with 4,550 more rows
## # i Use ‘print(n = ...)‘ to see more rows
```

```
ggplot(PR_data, aes(x = Area_size, y= Mean)) + geom_point(color="blue")
```



```
CA_data <- DF_H_Income3[DF_H_Income3$State_Name == 'California' ,]
CA_data
```

```
## # A tibble: 39,360 x 16
##    Locati~1 State~2 State~3 Abb_S~4 ‘City ‘ Area_~5 Mean_~6 Sd_H_~7  Mean Median
##       <int>   <int> <chr>   <chr>   <chr>   <chr>     <dbl>   <dbl> <int>  <int>
##  1  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  2  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  3  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  4  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  5  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  6  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  7  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  8  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
##  9  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
## 10  6011848       6 Califo~ CA      Bieber  530        41.2   -121. 54602 300000
## # ... with 39,350 more rows, 6 more variables: Stdev <int>, sum_w <dbl>,
## #   Location <chr>, District_type <chr>, Area <chr>, Area_size <dbl>, and
## #   abbreviated variable names 1: Location_ID, 2: State_Code, 3: State_Name,
## #   4: Abb_State_name, 5: Area_Code, 6: Mean_H_Income, 7: Sd_H_Income
## # i Use ‘print(n = ...)‘ to see more rows, and ‘colnames()‘ to see all variable names
```

```
summary(CA_data)
```

```
##   Location_ID         State_Code  State_Name        Abb_State_name
##  Min.   :    6029   Min.   :6    Length:39360       Length:39360
##  1st Qu.: 6026676   1st Qu.:6    Class :character   Class :character
##  Median :60214874   Median :6    Mode  :character   Mode  :character
##  Mean   :41002122   Mean   :6
##  3rd Qu.:60223072   3rd Qu.:6
##  Max.   :60231269   Max.   :6
##     City             Area_Code         Mean_H_Income    Sd_H_Income
##  Length:39360       Length:39360       Min.   :32.56   Min.   :-124.2
##  Class :character   Class :character   1st Qu.:33.93   1st Qu.:-121.6
##  Mode  :character   Mode  :character   Median :34.26   Median :-118.5
##                                        Mean   :35.62   Mean   :-119.5
##                                        3rd Qu.:37.70   3rd Qu.:-117.9
##                                        Max.   :41.89   Max.   :-114.6
##       Mean           Median           Stdev            sum_w
##  Min.   :     0   Min.   :     0   Min.   :     0   Min.   :     0.0
##  1st Qu.: 53096   1st Qu.: 42017   1st Qu.: 41752   1st Qu.:   193.3
##  Median : 72332   Median : 63368   Median : 53256   Median :   328.3
##  Mean   : 78127   Mean   :100582   Mean   : 53653   Mean   :   596.1
##  3rd Qu.: 98073   3rd Qu.:105283   3rd Qu.: 66052   3rd Qu.:   597.6
##  Max.   :242857   Max.   :300000   Max.   :100375   Max.   :106035.1
##    Location          District_type          Area             Area_size
##  Length:39360       Length:39360       Length:39360       Min.   :0.000e+00
##  Class :character   Class :character   Class :character   1st Qu.:1.454e+04
##  Mode  :character   Mode  :character   Mode  :character   Median :9.411e+04
##                                                           Mean   :1.621e+07
##                                                           3rd Qu.:1.303e+06
##                                                           Max.   :1.770e+10
```

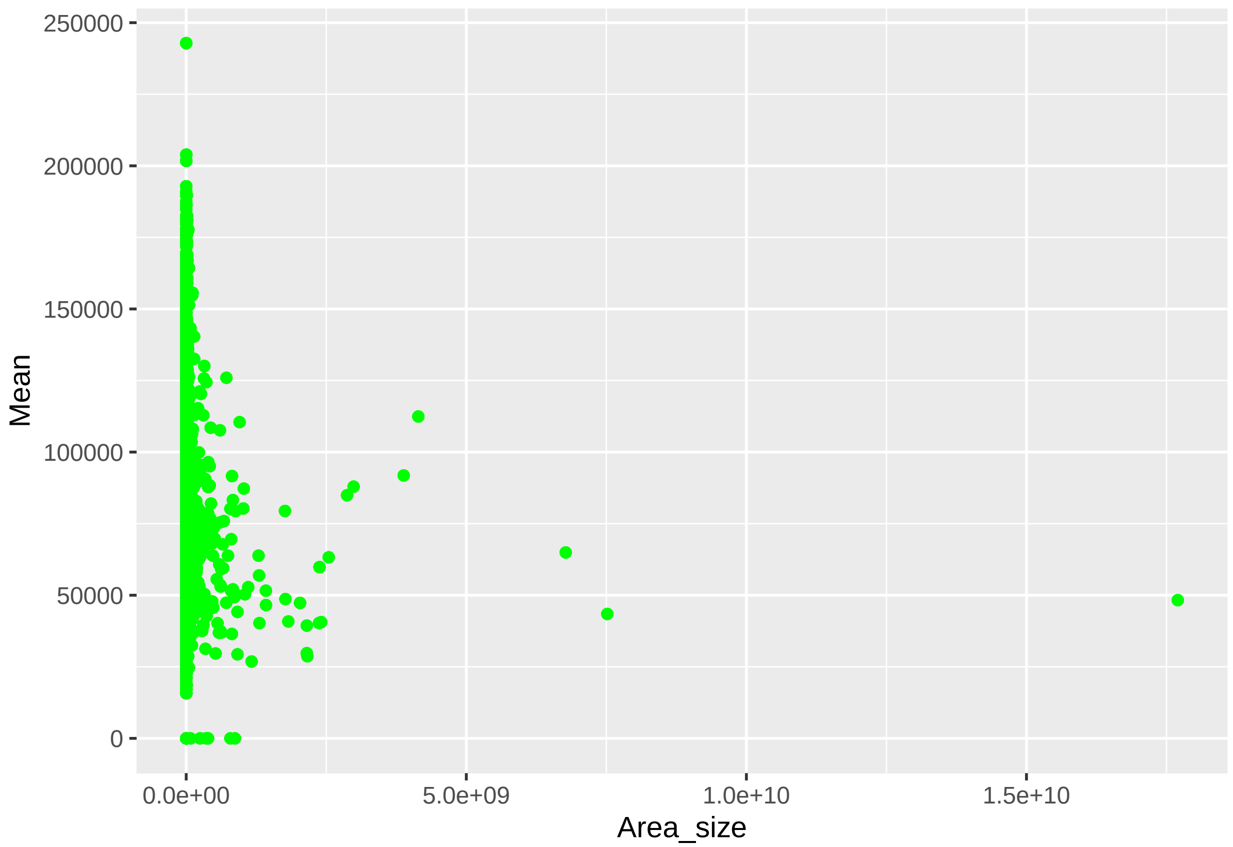
```
CA_data2 <- CA_data[order(CA_data$Mean, decreasing = TRUE),]
CA_data2
```

```
## # A tibble: 39,360 x 16
##    Locat~1 State~2 State~3 Abb_S~4 ‘City ‘ Area_~5 Mean_~6 Sd_H_~7   Mean Median
##      <int>   <int> <chr>   <chr>   <chr>   <chr>     <dbl>   <dbl>  <int>  <int>
##  1  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  2  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  3  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  4  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  5  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  6  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  7  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  8  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
##  9  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
## 10  6.02e7       6 Califo~ CA      San Di~ 619          32.7   -117. 242857 300000
## # ... with 39,350 more rows, 6 more variables: Stdev <int>, sum_w <dbl>,
## #   Location <chr>, District_type <chr>, Area <chr>, Area_size <dbl>, and
## #   abbreviated variable names 1: Location_ID, 2: State_Code, 3: State_Name,
## #   4: Abb_State_name, 5: Area_Code, 6: Mean_H_Income, 7: Sd_H_Income
## # i Use ‘print(n = ...)‘ to see more rows, and ‘colnames()‘ to see all variable names
```

```
CA3 <- select(CA_data2, "Mean")
CA3
```

```
## # A tibble: 39,360 x 1
##       Mean
##      <int>
##  1 242857
##  2 242857
##  3 242857
##  4 242857
##  5 242857
##  6 242857
##  7 242857
##  8 242857
##  9 242857
## 10 242857
## # ... with 39,350 more rows
## # i Use `print(n = ...)` to see more rows
```

```
ggplot(CA_data, aes(x = Area_size, y= Mean )) + geom_point(color="green")
```


Mean
0.0e+00
5.0e+09
1.0e+10
1.5e+10
Area_size