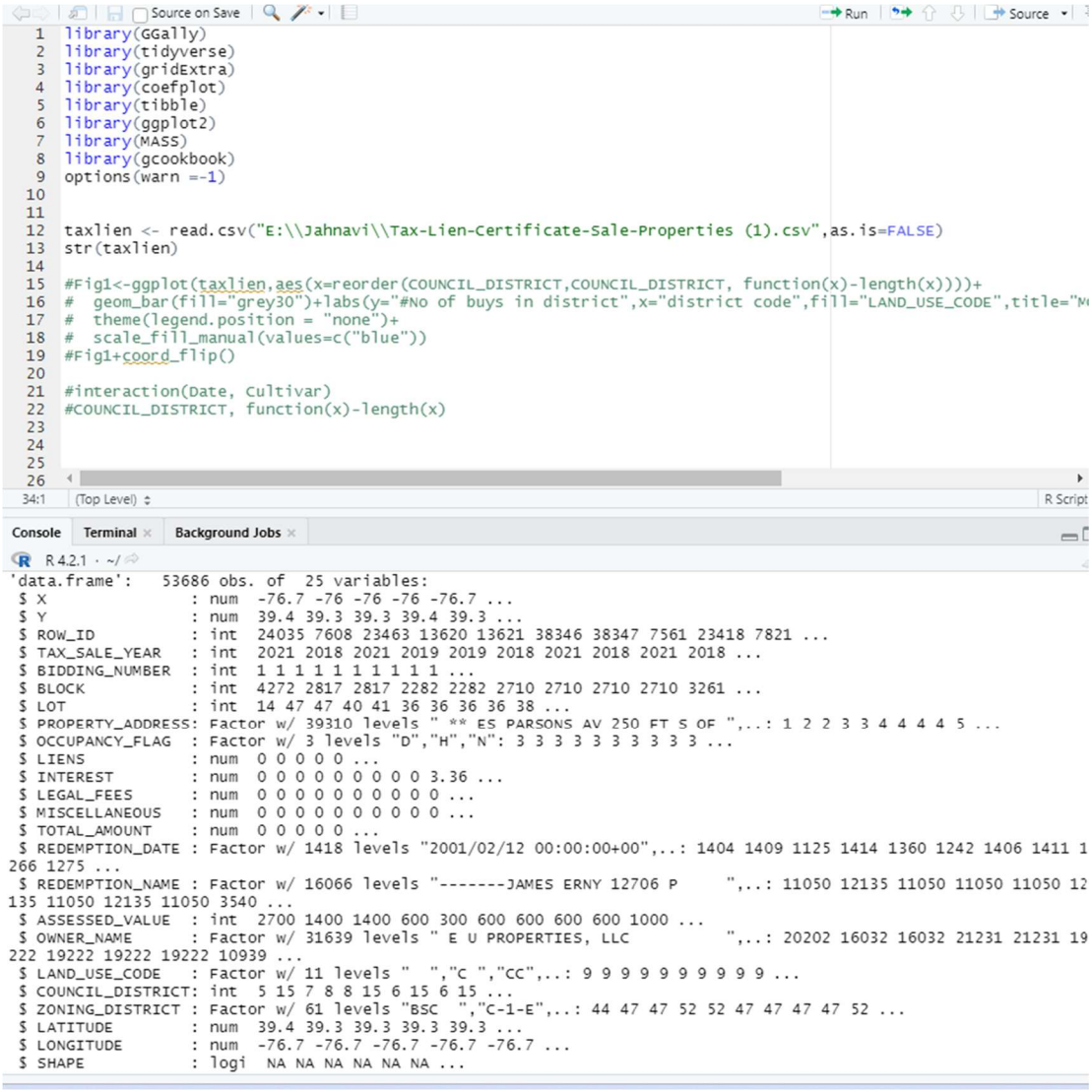
VISUALIZATION:

R SCRIPT

VISUALIZATION SHOWING THAT DATASET HAS BEEN ENTERED.

The as.is=FALSE argument specifies that the function should not convert character vectors to factors, and instead keep them as character vectors.

The str() function is then applied to the taxlien data frame to show its structure. This function displays the internal structure of an R object and is useful for examining the data types of the variables, their length, and the first few observations.



```
library(GGally)
library(tidyverse)
library(gridExtra)
library(coefplot)
library(tibble)
library(ggplot2)
library(MASS)
library(gcookbook)
options(warn =-1)


taxlien <- read.csv("E:\\Jahnavi\\Tax-Lien-Certificate-Sale-Properties (1).csv",as.is=FALSE)
str(taxlien)

#Fig1<-ggplot(taxlien,aes(x=reorder(COUNCIL_DISTRICT,COUNCIL_DISTRICT, function(x)-length(x))))+
#  geom_bar(fill="grey30")+labs(y="#No of buys in district",x="district code",fill="LAND_USE_CODE",title="M
#  theme(legend.position = "none")+
#  scale_fill_manual(values=c("blue"))
#Fig1+coord_flip()

#interaction(Date, Cultivar)
#COUNCIL_DISTRICT, function(x)-length(x)
```

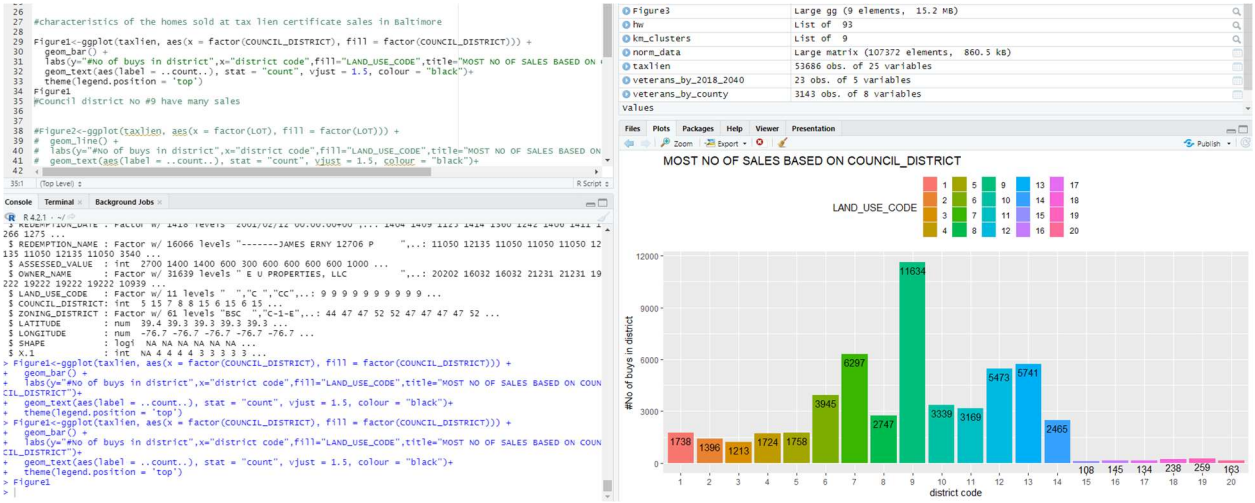
```
'data.frame':   53686 obs. of  25 variables:
 $ X                : num  -76.7 -76 -76 -76 -76.7 ...
 $ Y                : num  39.4 39.3 39.3 39.4 39.3 ...
 $ ROW_ID           : int  24035 7608 23463 13620 13621 38346 38347 7561 23418 7821 ...
 $ TAX_SALE_YEAR    : int  2021 2018 2021 2019 2019 2018 2021 2018 2021 2018 ...
 $ BIDDING_NUMBER   : int  1 1 1 1 1 1 1 1 1 1 ...
 $ BLOCK            : int  4272 2817 2817 2282 2282 2710 2710 2710 2710 3261 ...
 $ LOT              : int  14 47 47 40 41 36 36 36 36 38 ...
 $ PROPERTY_ADDRESS: Factor w/ 39310 levels " ** ES PARSONS AV 250 FT S OF ",..: 1 2 2 3 3 4 4 4 4 5 ...
 $ OCCUPANCY_FLAG   : Factor w/ 3 levels "D","H","N": 3 3 3 3 3 3 3 3 3 3 ...
 $ LIENS            : num  0 0 0 0 0 ...
 $ INTEREST         : num  0 0 0 0 0 0 0 0 0 3.36 ...
 $ LEGAL_FEES       : num  0 0 0 0 0 0 0 0 0 0 ...
 $ MISCELLANEOUS    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ TOTAL_AMOUNT     : num  0 0 0 0 0 ...
 $ REDEMPTION_DATE : Factor w/ 1418 levels "2001/02/12 00:00:00+00",..: 1404 1409 1125 1414 1360 1242 1406 1411 1
266 1275 ...
 $ REDEMPTION_NAME : Factor w/ 16066 levels "-------JAMES ERNY 12706 P     ",..: 11050 12135 11050 11050 11050 12
135 11050 12135 11050 3540 ...
 $ ASSESSED_VALUE   : int  2700 1400 1400 600 300 600 600 600 600 1000 ...
 $ OWNER_NAME       : Factor w/ 31639 levels " E U PROPERTIES, LLC          ",..: 20202 16032 16032 21231 21231 19
222 19222 19222 19222 10939 ...
 $ LAND_USE_CODE    : Factor w/ 11 levels "  ","C ","CC",..: 9 9 9 9 9 9 9 9 9 9 ...
 $ COUNCIL_DISTRICT: int  5 15 7 8 8 15 6 15 6 15 ...
 $ ZONING_DISTRICT : Factor w/ 61 levels "BSC  ","C-1-E",..: 44 47 47 52 52 47 47 47 47 52 ...
 $ LATITUDE         : num  39.4 39.3 39.3 39.3 39.3 ...
 $ LONGITUDE        : num  -76.7 -76.7 -76.7 -76.7 -76.7 ...
 $ SHAPE            : logi  NA NA NA NA NA NA ...
```

FIGURE NO 1:

CHARACTERISTICS OF THE HOMES SOLD AT TAX LIEN CERTIFICATE SALES IN BALTIMORE.

This R script is generating a bar graph using the ggplot2 package in R. The dataset used in this script is taxlien, which presumably contains data related to homes sold at tax lien certificate sales in Baltimore. The graph is intended to show the number of purchases made in each district based on the council district code.

Overall, this R script is creating a bar chart that shows the number of purchases made in each council district based on the COUNCIL_DISTRICT variable in the taxlien dataset. The chart is also annotated with text labels showing the count of observations in each bar.



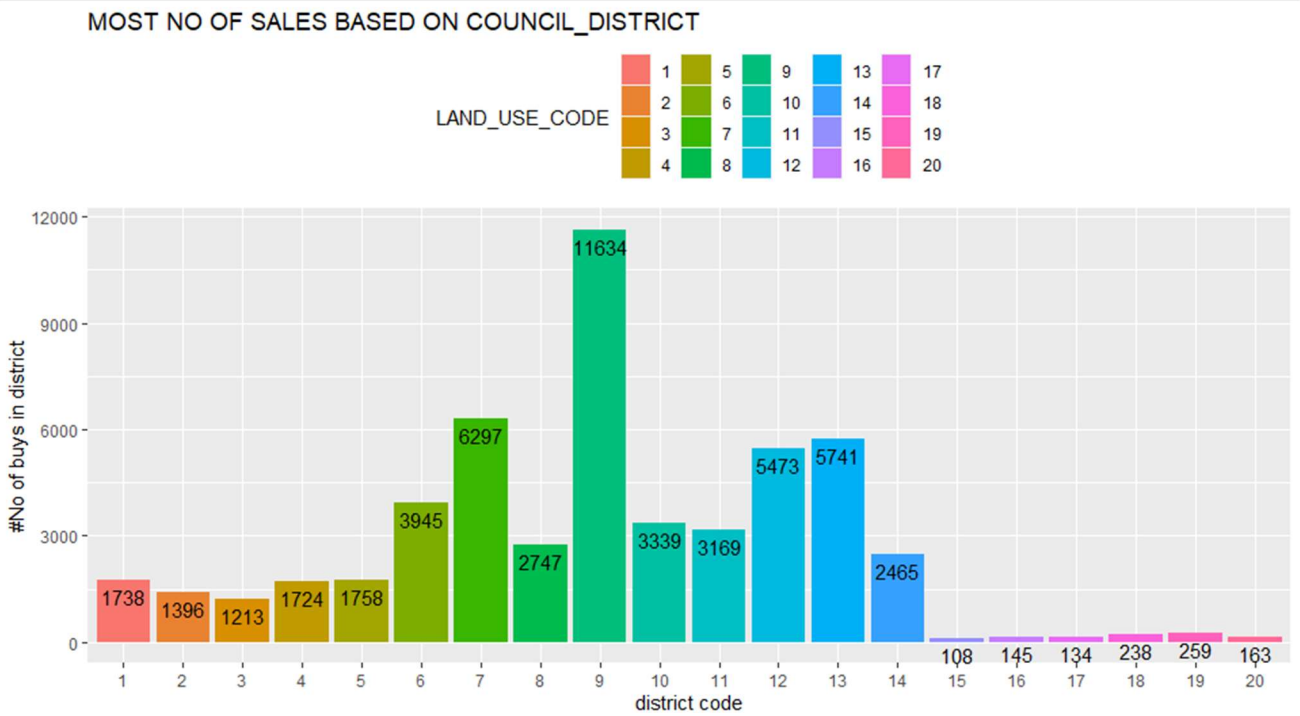
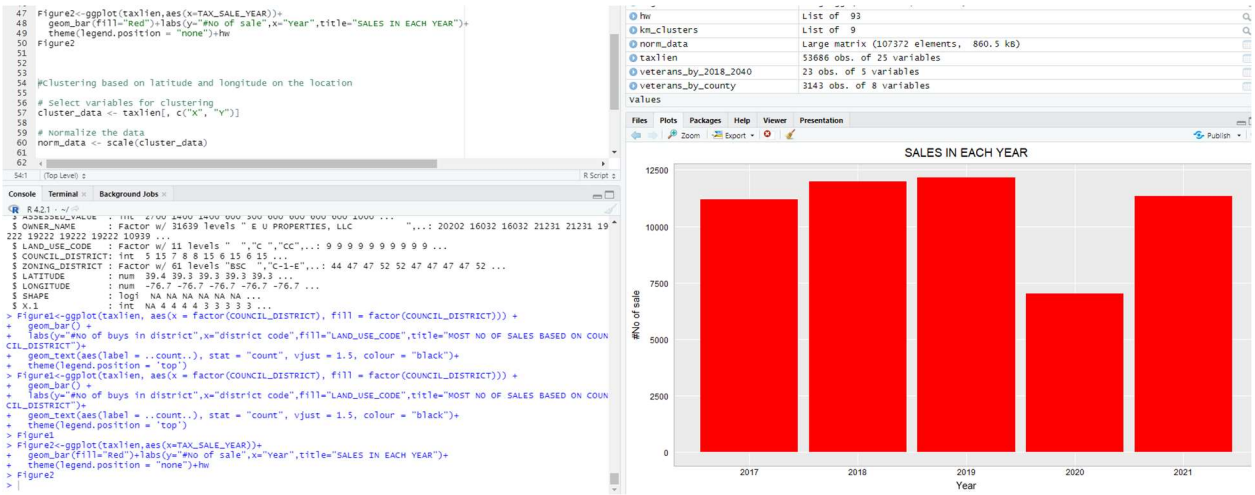
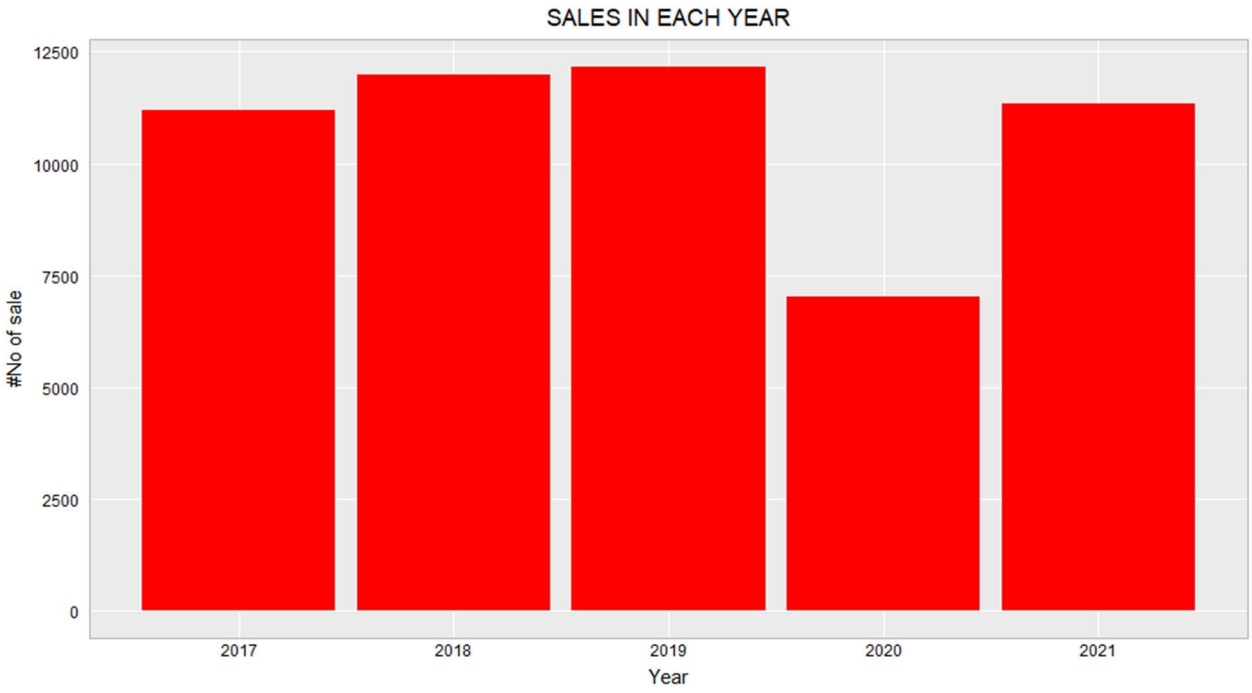


FIGURE 2:

This R code creates a bar plot using the ggplot2 package to visualize the number of sales in each year for tax lien certificate sales in Baltimore. Here is an explanation of the code:



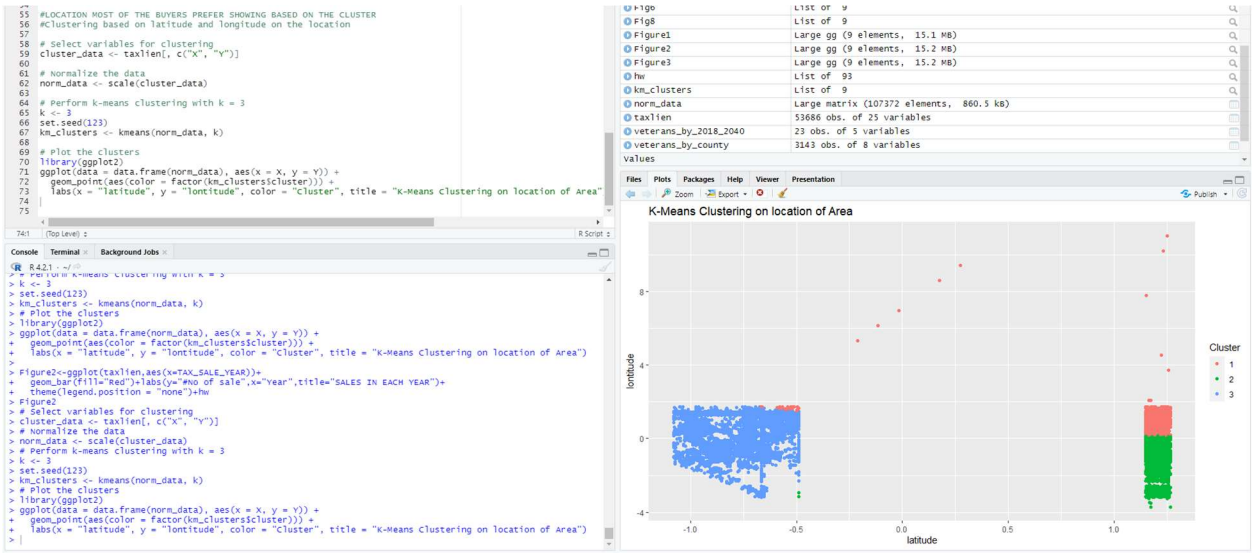


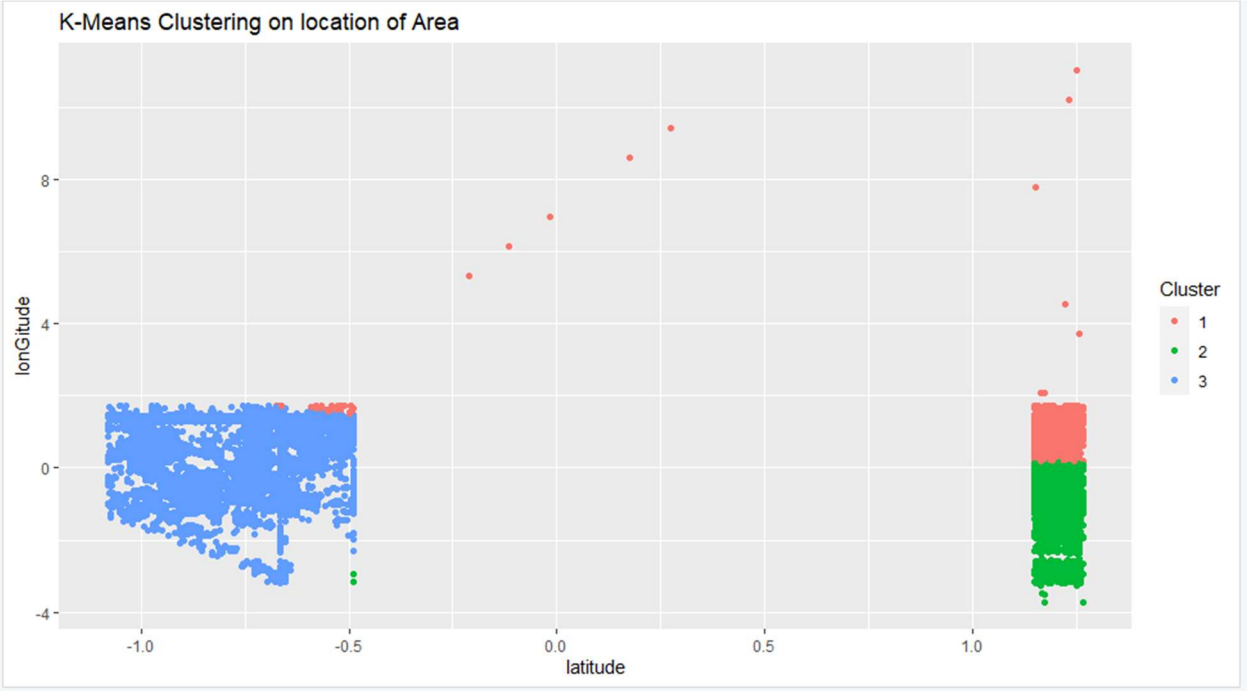
Overall, this plot shows the number of tax lien certificate sales that occurred each year in Baltimore.

FIGURE 3:

#LOCATION MOST OF THE BUYERS PREFER SHOWING BASED ON THE CLUSTER

#CLUSTERING BASED ON LATITUDE AND LONGITUDE ON THE LOCATION





This R script performs K-means clustering on the location data of homes sold at tax lien certificate sales in Baltimore, to identify clusters of areas where most of the buyers prefer to buy homes.

First, the X and Y variables (longitude and latitude) are selected from the dataset and normalized using the scale() function. Then, K-means clustering is performed on the normalized data with k=3 clusters, using the kmeans() function.

Finally, a scatter plot is created using ggplot2 library to visualize the clusters based on the location of the areas. Each point represents a home sale, and the color of the point indicates the cluster it belongs to. The plot title mentions that the clusters represent the areas where most of the buyers prefer to buy homes.