Assignment 3

Kayla Myros, Claire Tham, Khalil Kaba

11/8/2021

Introduction

Research Question

Does the housing real estate market see less activity in areas that have higher incidences of home ownership cost burden?

Data Gathering Process

#setting up the R Markdown File

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.5     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.4     ✓ stringr 1.4.0
## ✓ readr   2.0.2     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidycensus)
library(readxl)
library(knitr)
library(tigris)
```

```
## To enable
## caching of data, set `options(tigris_use_cache = TRUE)` in your R script or .Rprofile.
```

```
##
## Attaching package: 'tigris'
```

```
## The following object is masked from 'package:tidycensus':
##
##     fips_codes
```

```
library(sf)
```

```
## Linking to GEOS 3.8.1, GDAL 3.2.1, PROJ 7.2.1
```

```
library(dplyr)
```

Variable Definitions

CATEGORICAL VARIABLES

Our team created the following categorical variables using the U.S. Census Bureau, American Community Survey 1-year estimates data.

Median Household Income, 2019 Data of median household income for each Metropolitan Statistical Area categorized into "high" and "low" categories. High median income is any incomes greater than 55,000 dollars, Low median income is equal to or less than 55,000 dollars.

Mortgage Cost Burden, 2019 Mortgage cost burden data is defined as the percentage of the population within the Metropolitan Statistical Area that are homeowners with a mortgage who spend 30 percent or more of their income on their mortgage cost. The variable is defined as a category of a highly burdened area and less burdened area. Highly burdened means 30 percent or more of the mortgaged homeowner population are housing burdened. While less burdened means fewer than 30 percent of the mortgaged homeowner population are housing burdened.

CONTINUOUS VARIABLES

Our team identified the following continuous variables from the Zillow research data center.

For Sale Inventory, 2021 This data shows the number of homes on the housing market per the Metropolitan Statistical Area.It is reported monthly.

Median Sales Price, 2021 This data shows the median sales price of homes sold on the housing market within the specific Metropolitan Statistical Area. It is reported monthly.

Median List Price, 2021 This data shows the median listing price of homes on the housing market within the specific Metropolitan Statistical Area.This is what the property owner is asking for the home. This data is reported monthly.

Loading and Cleaning the Data files

```
Median_HH_Income <- read_excel("2019_medianHHincome.xlsx") %>%
  mutate(MedIncome_Quality = ifelse(Med_Income > 55000, "High", "Low"))


Mortgage_Burden <- read_excel("MortgageBurden_MSA_2019_1yr.xlsx") %>%
    mutate(MortBurd_Quality = ifelse(Pct_MortBurden > .30, "HighlyBurdened", "LessBurdened"))
    colnames(Mortgage_Burden)[2] <- c("RegionName")


ForSale_Inventory <- read_excel("MSA_For Sale Inventory.xlsx")


Med_Sale_Price <- read_excel("MSA_Med Sale Price.xlsx")
  colnames(Med_Sale_Price)[14] <- c("SaleP_Sep")


Median_List_Price <- read_excel("MSA_Median List Price.xlsx")
  Median_List_Price <- Median_List_Price[ -c(1,2,4:49) ]
  colnames(Median_List_Price)[2] <- c("listP_Sep")
```

Fixing the Census Data MSA to Match Zillow Data MSA

```
Median_HH_Income$RegionName <- gsub("Albany-Schenectady-Troy, NY", "Albany, NY", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Allentown-Bethlehem-Easton, PA-NJ", "Allentown, PA", Median_HH_Income$Region
Name)
Median_HH_Income$RegionName <- gsub("Atlanta-Sandy Springs-Alpharetta, GA", "Atlanta, GA", Median_HH_Income$Regio
nName)
Median_HH_Income$RegionName <- gsub("Austin-Round Rock-Georgetown, TX", "   Austin, TX", Median_HH_Income$RegionN
ame)
Median_HH_Income$RegionName <- gsub("Baltimore-Columbia-Towson, MD", "Baltimore, MD", Median_HH_Income$RegionName
)
Median_HH_Income$RegionName <- gsub("Birmingham-Hoover, AL", "Birmingham, AL", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Boston-Cambridge-Newton, MA-NH", "Boston, MA", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Buffalo-Cheektowaga, NY", "Buffalo, NY", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Charleston-North Charleston, SC", "Charleston, SC", Median_HH_Income$RegionN
ame)
Median_HH_Income$RegionName <- gsub("Charlotte-Concord-Gastonia, NC-SC", "Charlotte, NC", Median_HH_Income$RegionN
ame)
Median_HH_Income$RegionName <- gsub("Chicago-Naperville-Elgin, IL-IN-WI", "Chicago, IL", Median_HH_Income$RegionN
ame)
Median_HH_Income$RegionName <- gsub("Cincinnati, OH-KY-IN", "Cincinnati, OH", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Cleveland-Elyria, OH", "Cleveland, OH", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Dallas-Fort Worth-Arlington, TX", "Dallas-Fort Worth, TX", Median_HH_Income$
RegionName)
Median_HH_Income$RegionName <- gsub("Dayton-Kettering, OH", "Dayton, OH", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Deltona-Daytona Beach-Ormond Beach, FL", "Daytona Beach, FL", Median_HH_Inco
me$RegionName)

Median_HH_Income$RegionName <- gsub("Denver-Aurora-Lakewood, CO", "Denver, CO", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Des Moines-West Des Moines, IA", "Des Moines, IA", Median_HH_Income$RegionNa
me)

Median_HH_Income$RegionName <- gsub("Detroit-Warren-Dearborn, MI", "Detroit, MI", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Durham-Chapel Hill, NC", "Durham, NC", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Grand Rapids-Kentwood, MI", "Grand Rapids, MI", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Greensboro-High Point, NC", "Greensboro, NC", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Greenville-Anderson, SC", "Greenville, SC", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Hartford-East Hartford-Middletown, CT", "Hartford, CT", Median_HH_Income$Reg
ionName)

Median_HH_Income$RegionName <- gsub("Houston-The Woodlands-Sugar Land, TX", "Houston, TX", Median_HH_Income$Regio
nName)

Median_HH_Income$RegionName <- gsub("Indianapolis-Carmel-Anderson, IN", "Indianapolis, IN", Median_HH_Income$Regi
onName)

Median_HH_Income$RegionName <- gsub("Las Vegas-Henderson-Paradise, NV", "Las Vegas, NV", Median_HH_Income$RegionN
ame)

Median_HH_Income$RegionName <- gsub("Little Rock-North Little Rock-Conway, AR", "Little Rock, AR", Median_HH_Inco
me$RegionName)

Median_HH_Income$RegionName <- gsub("Louisville/Jefferson County, KY-IN", "Louisville-Jefferson County, KY", Medi
an_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Memphis, TN-MS-AR", "Memphis, TN", Median_HH_Income$RegionName)

Median_HH_Income$RegionName <- gsub("Memphis, TN-MS-AR", "Memphis, TN", Median_HH_Income$RegionName)

Median_HH_Income$RegionName <- gsub("Miami-Fort Lauderdale-Pompano Beach, FL", "Miami-Fort Lauderdale, FL", Media
n_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Milwaukee-Waukesha, WI", "Milwaukee, WI", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Minneapolis-St. Paul-Bloomington, MN-WI", "Minneapolis-St Paul, MN", Median_
HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Nashville-Davidson--Murfreesboro--Franklin, TN", "Nashville, TN", Median_HH_
Income$RegionName)

Median_HH_Income$RegionName <- gsub("New Haven-Milford, CT", "New Haven, CT", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("New Orleans-Metairie, LA", "New Orleans, LA", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("New York-Newark-Jersey City, NY-NJ-PA", "New York, NY", Median_HH_Income$Reg
ionName)
Median_HH_Income$RegionName <- gsub("Ogden-Clearfield, UT", "Ogden, UT", Median_HH_Income$RegionName)

Median_HH_Income$RegionName <- gsub("Omaha-Council Bluffs, NE-IA", "Omaha, NE", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Orlando-Kissimmee-Sanford, FL", "Orlando, FL", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Philadelphia-Camden-Wilmington, PA-NJ-DE-MD", "Philadelphia, PA", Median_HH_
Income$RegionName)

Median_HH_Income$RegionName <- gsub("Phoenix-Mesa-Chandler, AZ", "Phoenix, AZ", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Portland-South Portland, ME", "Portland, ME", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Providence-Warwick, RI-MA", "Providence, RI", Median_HH_Income$RegionName)

Median_HH_Income$RegionName <- gsub("Raleigh-Cary, NC", "Raleigh, NC", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Riverside-San Bernardino-Ontario, CA", "Riverside, CA", Median_HH_Income$Reg
ionName)
Median_HH_Income$RegionName <- gsub("Sacramento-Roseville-Folsom, CA", "Sacramento, CA", Median_HH_Income$RegionN
ame)

Median_HH_Income$RegionName <- gsub("San Diego-Chula Vista-Carlsbad, CA", "San Diego, CA", Median_HH_Income$Regio
nName)
Median_HH_Income$RegionName <- gsub("San Francisco-Oakland-Berkeley, CA", "San Francisco, CA", Median_HH_Income$R
egionName)
Median_HH_Income$RegionName <- gsub("San Jose-Sunnyvale-Santa Clara, CA", "San Jose, CA", Median_HH_Income$Region
Name)

Median_HH_Income$RegionName <- gsub("Seattle-Tacoma-Bellevue, WA", "Seattle, WA", Median_HH_Income$RegionName)
Median_HH_Income$RegionName <- gsub("Tampa-St. Petersburg-Clearwater, FL", "Tampa, FL", Median_HH_Income$RegionNa
me)
Median_HH_Income$RegionName <- gsub("Virginia Beach-Norfolk-Newport News, VA-NC", "Virginia Beach, VA", Median_HH
_Income$RegionName)


Median_HH_Income$RegionName <- gsub("Washington-Arlington-Alexandria, DC-VA-MD-WV", "Washington, DC", Median_HH_I
ncome$RegionName)
Median_HH_Income$RegionName <- gsub("Worcester, MA-CT", "Worcester, MA", Median_HH_Income$RegionName)
```

```
Mortgage_Burden$RegionName <- gsub("Albany-Schenectady-Troy, NY", "Albany, NY", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Allentown-Bethlehem-Easton, PA-NJ", "Allentown, PA", Mortgage_Burden$RegionNa
me)
Mortgage_Burden$RegionName <- gsub("Atlanta-Sandy Springs-Alpharetta, GA", "Atlanta, GA", Mortgage_Burden$RegionN
ame)
Mortgage_Burden$RegionName <- gsub("Austin-Round Rock-Georgetown, TX", "   Austin, TX", Mortgage_Burden$RegionNa
me)
Mortgage_Burden$RegionName <- gsub("Baltimore-Columbia-Towson, MD", "Baltimore, MD", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Birmingham-Hoover, AL", "Birmingham, AL", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Boston-Cambridge-Newton, MA-NH", "Boston, MA", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Buffalo-Cheektowaga, NY", "Buffalo, NY", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Charleston-North Charleston, SC", "Charleston, SC", Mortgage_Burden$RegionNam
e)
Mortgage_Burden$RegionName <- gsub("Charlotte-Concord-Gastonia, NC-SC", "Charlotte, NC", Mortgage_Burden$RegionNa
me)
Mortgage_Burden$RegionName <- gsub("Chicago-Naperville-Elgin, IL-IN-WI", "Chicago, IL", Mortgage_Burden$RegionNam
e)
Mortgage_Burden$RegionName <- gsub("Cincinnati, OH-KY-IN", "Cincinnati, OH", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Cleveland-Elyria, OH", "Cleveland, OH", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Dallas-Fort Worth-Arlington, TX", "Dallas-Fort Worth, TX", Mortgage_Burden$Re
gionName)
Mortgage_Burden$RegionName <- gsub("Dayton-Kettering, OH", "Dayton, OH", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Deltona-Daytona Beach-Ormond Beach, FL", "Daytona Beach, FL", Mortgage_Burden
$RegionName)

Mortgage_Burden$RegionName <- gsub("Denver-Aurora-Lakewood, CO", "Denver, CO", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Des Moines-West Des Moines, IA", "Des Moines, IA", Mortgage_Burden$RegionName
)

Mortgage_Burden$RegionName <- gsub("Detroit-Warren-Dearborn, MI", "Detroit, MI", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Durham-Chapel Hill, NC", "Durham, NC", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Grand Rapids-Kentwood, MI", "Grand Rapids, MI", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Greensboro-High Point, NC", "Greensboro, NC", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Greenville-Anderson, SC", "Greenville, SC", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Hartford-East Hartford-Middletown, CT", "Hartford, CT", Mortgage_Burden$Regio
nName)

Mortgage_Burden$RegionName <- gsub("Houston-The Woodlands-Sugar Land, TX", "Houston, TX", Mortgage_Burden$RegionN
ame)

Mortgage_Burden$RegionName <- gsub("Indianapolis-Carmel-Anderson, IN", "Indianapolis, IN", Mortgage_Burden$Region
Name)

Mortgage_Burden$RegionName <- gsub("Las Vegas-Henderson-Paradise, NV", "Las Vegas, NV", Mortgage_Burden$RegionNam
e)

Mortgage_Burden$RegionName <- gsub("Little Rock-North Little Rock-Conway, AR", "Little Rock, AR", Mortgage_Burden
$RegionName)
Mortgage_Burden$RegionName <- gsub("Louisville/Jefferson County, KY-IN", "Louisville-Jefferson County, KY", Mortg
age_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Memphis, TN-MS-AR", "Memphis, TN", Mortgage_Burden$RegionName)

Mortgage_Burden$RegionName <- gsub("Memphis, TN-MS-AR", "Memphis, TN", Mortgage_Burden$RegionName)

Mortgage_Burden$RegionName <- gsub("Miami-Fort Lauderdale-Pompano Beach, FL", "Miami-Fort Lauderdale, FL", Mortga
ge_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Milwaukee-Waukesha, WI", "Milwaukee, WI", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Minneapolis-St. Paul-Bloomington, MN-WI", "Minneapolis-St Paul, MN", Mortgage
_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Nashville-Davidson--Murfreesboro--Franklin, TN", "Nashville, TN", Mortgage_Bu
rden$RegionName)

Mortgage_Burden$RegionName <- gsub("New Haven-Milford, CT", "New Haven, CT", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("New Orleans-Metairie, LA", "New Orleans, LA", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("New York-Newark-Jersey City, NY-NJ-PA", "New York, NY", Mortgage_Burden$Regio
nName)
Mortgage_Burden$RegionName <- gsub("Ogden-Clearfield, UT", "Ogden, UT", Mortgage_Burden$RegionName)

Mortgage_Burden$RegionName <- gsub("Omaha-Council Bluffs, NE-IA", "Omaha, NE", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Orlando-Kissimmee-Sanford, FL", "Orlando, FL", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Philadelphia-Camden-Wilmington, PA-NJ-DE-MD", "Philadelphia, PA", Mortgage_Bu
rden$RegionName)

Mortgage_Burden$RegionName <- gsub("Phoenix-Mesa-Chandler, AZ", "Phoenix, AZ", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Portland-South Portland, ME", "Portland, ME", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Providence-Warwick, RI-MA", "Providence, RI", Mortgage_Burden$RegionName)

Mortgage_Burden$RegionName <- gsub("Raleigh-Cary, NC", "Raleigh, NC", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Riverside-San Bernardino-Ontario, CA", "Riverside, CA", Mortgage_Burden$Regio
nName)
Mortgage_Burden$RegionName <- gsub("Sacramento-Roseville-Folsom, CA", "Sacramento, CA", Mortgage_Burden$RegionNam
e)

Mortgage_Burden$RegionName <- gsub("San Diego-Chula Vista-Carlsbad, CA", "San Diego, CA", Mortgage_Burden$RegionN
ame)
Mortgage_Burden$RegionName <- gsub("San Francisco-Oakland-Berkeley, CA", "San Francisco, CA", Mortgage_Burden$Reg
ionName)
Mortgage_Burden$RegionName <- gsub("San Jose-Sunnyvale-Santa Clara, CA", "San Jose, CA", Mortgage_Burden$RegionNa
me)

Mortgage_Burden$RegionName <- gsub("Seattle-Tacoma-Bellevue, WA", "Seattle, WA", Mortgage_Burden$RegionName)
Mortgage_Burden$RegionName <- gsub("Tampa-St. Petersburg-Clearwater, FL", "Tampa, FL", Mortgage_Burden$RegionName
)
Mortgage_Burden$RegionName <- gsub("Virginia Beach-Norfolk-Newport News, VA-NC", "Virginia Beach, VA", Mortgage_B
urden$RegionName)


Mortgage_Burden$RegionName <- gsub("Washington-Arlington-Alexandria, DC-VA-MD-WV", "Washington, DC", Mortgage_Bur
den$RegionName)
Mortgage_Burden$RegionName <- gsub("Worcester, MA-CT", "Worcester, MA", Mortgage_Burden$RegionName)
```

Joining the datasets & filtering to variables needed

```
data <-  left_join(Median_HH_Income, Mortgage_Burden) %>%
  left_join(ForSale_Inventory) %>%
  left_join(Median_List_Price) %>%
  left_join(Med_Sale_Price)
```

```
## Joining, by = c("geoID", "RegionName")
```

```
## Joining, by = "RegionName"
## Joining, by = "RegionName"
```

```
## Joining, by = c("RegionName", "RegionID", "SizeRank", "RegionType", "StateName")
```

```
data <- data[ -c(3,4,6:12,15:26, 29:36) ]

#kable(head(data))
```

Filter out the NA Data

```
data <- data %>% drop_na(SaleP_Sep)
```

Observations (UPDATED)

There are 75 rows of data, this represents 75 Metropolitan Statistical Areas (MSA) and there are 8 variables in the columns, 4 of which are our variables for the regression analyses. The additional columns are the GEOID, and raw data from which we calculated our Categorical Variables.

ASSIGNMENT 3

Our dependent variable is the For Sale Housing Inventory of each MSA. Our independent variables are median household income, mortgage burden, median home sale price, and median home list price.

Continious Variables: Correlation + Regression

This is a correlation test between the dependent variable (For Sale Inventory) and an independent variable (Median Sale Price in September 2021). The correlation is 0.25. This means that there is a positive correlation between median sale price and the size of the for sale inventory.

```
cor.test(~ Inv_Sep + SaleP_Sep, data = data)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Inv_Sep and SaleP_Sep
## t = 2.2049, df = 71, p-value = 0.03071
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02450486 0.45661860
## sample estimates:
##       cor
## 0.253145
```

```
# Correlation is 0.253145
```

This is a regression test between the dependent variable (For Sale Inventory) and an independent variable (Median List Price in September 2021). The p-value is 0.031. There is a small, positive correlation between median list price and the size of the for sale inventory.

```
ListPrice_model <- lm(Inv_Sep ~ listP_Sep, data = data)

summary(ListPrice_model)
```

```
##
## Call:
## lm(formula = Inv_Sep ~ listP_Sep, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -17167  -4846  -3412   1161  51719
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.876e+03  2.744e+03   1.048   0.2980
## listP_Sep   1.386e-02  6.288e-03   2.205   0.0307 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10090 on 71 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.06408,    Adjusted R-squared:  0.0509
## F-statistic: 4.861 on 1 and 71 DF,  p-value: 0.03071
```

This is a correlation test between the dependent variable (For Sale Inventory) and an independent variable (Median List Price in September 2021). The correlation is 0.25. This means that there is a positive correlation between median list price and the size of the for sale inventory.

```
cor.test(~ Inv_Sep + listP_Sep, data = data)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Inv_Sep and listP_Sep
## t = 2.2049, df = 71, p-value = 0.03071
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02450486 0.45661860
## sample estimates:
##       cor
## 0.253145
```

```
# Correlation is 0.253145
```

This is a regression test between the dependent variable (For Sale Inventory) and an independent variable (Median Sale Price in September 2021). The p-value is 0.031. There is a small, positive correlation between median sale price and size of the for sale inventory.

```
SalePrice_model <- lm(Inv_Sep ~ SaleP_Sep, data = data)

summary(SalePrice_model)
```

```
##
## Call:
## lm(formula = Inv_Sep ~ SaleP_Sep, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -17167  -4846  -3412   1161  51719
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.876e+03  2.744e+03   1.048   0.2980
## SaleP_Sep   1.386e-02  6.288e-03   2.205   0.0307 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10090 on 71 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.06408,    Adjusted R-squared:  0.0509
## F-statistic: 4.861 on 1 and 71 DF,  p-value: 0.03071
```

Categorical Variables: T-Test and Regression

This is a t-test between dependent variable (For Sale Inventory) and independent variable (Mortgage Burden). The mortgage burden quality variable is considered high if the 30 percent or higher of an MSA's homeowner population is considered mortgage burdened.

```
t.test(Inv_Sep ~ MortBurd_Quality, data)
```

```
##
##  Welch Two Sample t-test
##
## data:  Inv_Sep by MortBurd_Quality
## t = 2.6747, df = 70.904, p-value = 0.00928
## alternative hypothesis: true difference in means between group HighlyBurdened and group LessBurdened is not eq
ual to 0
## 95 percent confidence interval:
##  1366.680 9373.395
## sample estimates:
## mean in group HighlyBurdened   mean in group LessBurdened
##                     10102.204                     4732.167
```

The results of the t-test show that the p-value is less than 0.05, (it is 0.00928). Therefore, there is a difference of homes for sale in MSAs with a high mortgage burden compared to MSAs with a low mortgage burden.

This is a t-test between dependent variable (For Sale Inventory) and independent variable (Median Household Income Quality). The median household income is considered high if the MSA's median income is above $55,000 annually, and it's considered low if it is below that.

```
MedInc_model <- lm(Inv_Sep ~ MedIncome_Quality, data)

summary(MedInc_model)
```

```
##
## Call:
## lm(formula = Inv_Sep ~ MedIncome_Quality, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7618  -6074  -2870    638  54434
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)              8979       1260   7.128 6.83e-10 ***
## MedIncome_QualityLow    -6702       4068  -1.648    0.104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10230 on 71 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.03683,    Adjusted R-squared:  0.02326
## F-statistic: 2.715 on 1 and 71 DF,  p-value: 0.1038
```

The p-value is not significant, as it is above 0.05, the p-value equals 0.104.

#Regression 2

This is a correlation test between the dependent variable (For Sale Inventory) and an independent variable (Mortgage Burden). The correlation is 0.1955168. This means that there is a positive correlation between median sale price and the percentage of Mortgage Burden.

```
cor.test(~ Inv_Sep + Pct_MortBurden, data = data)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Inv_Sep and Pct_MortBurden
## t = 2.1659, df = 71, p-value = 0.03368
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02003645 0.45307288
## sample estimates:
##       cor
## 0.2489561
```

```
# Correlation is 0.1955168
```

This is a regression test between the dependent variable (For Sale Inventory) and an independent variable (Motgage Burden). The p-value is 0.001. There is a positive but limited correlation between median list price and mortgage burden.

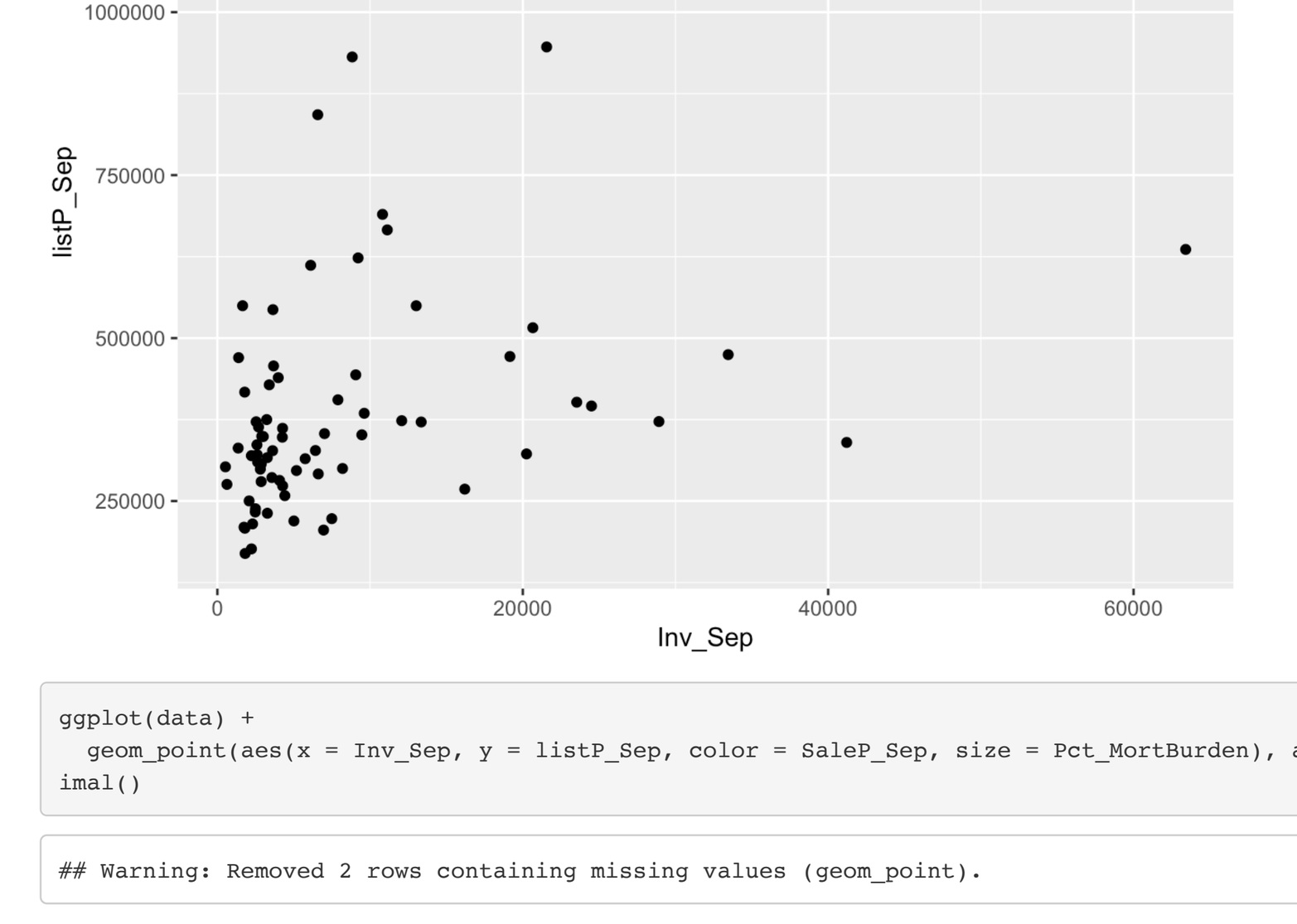
```
MortBurd_mode <- lm(Inv_Sep ~ Pct_MortBurden, data = data)

summary(MortBurd_mode)
```

```
##
## Call:
## lm(formula = Inv_Sep ~ Pct_MortBurden, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12174  -4703  -2937    269  52556
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2260       3044   0.743   0.4602
## Pct_MortBurden    15202       7018   2.166   0.0337 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10100 on 71 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.06198,    Adjusted R-squared:  0.04877
## F-statistic: 4.691 on 1 and 71 DF,  p-value: 0.03368
```

```
ggplot(data) +
  geom_point(aes(x = Inv_Sep, y = listP_Sep))
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```




```
ggplot(data) +
  geom_point(aes(x = Inv_Sep, y = listP_Sep, color = SaleP_Sep, size = Pct_MortBurden), alpha = 0.50) + theme_min
imal()
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```