Real Estate Prices and Inflation

Michael Grodecki

December 8th, 2022

INFO 5652 CU Boulder

1. Introduction

Throughout 2021 and 2022 the inflation rate averaged at 7.7%, which is a figure that is several times higher than in previous years, where the yearly inflation rate averaged at around 2%-3%. The effects of inflation have had quite a noticeable impact on the prices of goods and services, affecting consumers in a largely negative way. In the years following the COVID-19 pandemic, home prices have skyrocketed to record highs, diminishing the prospects of home ownership for many Americans. In 1985, the average single-family home cost was $82,800, compared to the average cost of $454,900 in the third quarter of 2022.[1,2] When looking at data from the first quarter of 2020, right at the start of the pandemic, the average home cost was at $329,000.[2] This 38% increase in average home prices over the last two years is much higher than the average rate of inflation. Looking back at older statistics, housing prices have been rising at a rate that is significantly higher than the rate of inflation. If housing prices more or less followed the rate of inflation since 1970, the average home price today would be at around $178,000.[1]

The increase in housing prices have been especially notable during the past two years, and the goal of this project is to determine what effects inflation has had on average housing prices, using observational data about housing prices and economics collected over a long period of time. By comparing national average housing prices to inflation data, it will be possible to analyze the correlation between these variables.

There are several past studies that have analyzed the correlation between inflation and housing prices. One of the most cited studies in the literature analyzes this question by comparing housing prices to CPI data.[3] In this study, which was published in 2002, regression analysis of the correlation between housing prices and the non-housing prices of goods and services found that there is significant correlation between the two variables, generating a Fisher coefficient of 1.2-1.54 when comparing monthly data from the years 1973-2000.[3] A study from 1981 also analyzes the relationship between housing prices and the consumer inflation rate, using the Consumer Price Index, and found that housing prices increased at a rate that was 33% higher than the CPI between 1963-1978.[4] A related study from 1982 analyzes this same correlation, suggesting that the CPI has significantly affected housing prices, when analyzing data from the past decade. [6] A recent study from 2022 suggests that changes in mortgage rates have had a significant impact on housing prices since the start of the pandemic, along with other factors such as the government monetary stimulus and the rising price in construction materials.[5]

This project will use data about the CPI, inflation rate calculated from the CPI, 30-year mortgage rates and the US housing price index (HPI) to explore the relations between these variables and how they have changed over a long period of time. In

addition to analyzing national HPI data, regional HPI's from the Mountain Census Region, Colorado, Denver, and Boulder will be analyzed and compared to each other. The hypothesis is that Boulder and Denver housing prices will be significantly higher than the state and national averages.

2. Data

Economic data and real estate data from various sources are used in the analysis. The datasets used are all numerical datasets and consist of observational data. Data containing the CPI (Consumer Price Index), 30 year fixed mortgage rates, and the HPI (Housing Price Index) across various geographical locations are used in the analysis.

The CPI is a measure of change in the prices of various goods measured over time, showing the change in average prices for consumer goods over a period of time. The CPI dataset used in the analysis is sourced from usinflationcalculator.com. This data is derived from statistics provided by the U.S. Department of Labor Bureau of Labor Statistics and shows the monthly change in CPI from 1913-2022. The CPI is a useful measure of inflation, and inflation rates are often calculated using CPI data.

A 30-year fixed mortgage rate is a fixed-rate loan that will be paid off in increments over a 30 year period, meaning that the interest rate will stay constant over the 30 year period regardless of inflation. These loans are given by a lender typically to finance the purchase of a property. Mortgage data used in the analysis is derived from Freddie Mac (The Federal Home Loan Mortgage Corporation), a company that frequently reports on economic and real estate data to the federal government. The data is numerical and shows the monthly changes in mortgage rates over a monthly time interval from 1971-2022.

The HPI (Housing Price Index) is an index that measures changes in single-family home values in the United States over a periodic interval. The HPI data used in the analysis is taken from the FHFA (Federal Housing Finance Agency) HPI dataset, which is a public nationwide dataset of changes in single-family housing prices using data gathered from all 50 states and over 400 cities. The data spans from a time period of 1991-2022 for national data and 1978-2022 for regional data that is used in this project's analysis. This dataset was especially important in the analysis due to the insight it gave on national, state, census division, and city-specific data about changes in housing prices.

An initial look at the data suggested that there may be some relationship between the CPI and average housing prices. Both of these variables have been rising significantly since the 1970s with noticeable downturns during financial crises such as the 2008 recession. Mortgage rates, on the other hand, have been fluctuating significantly over time and looking at the data does not offer any insight concerning the relationship between inflation rates and mortgage rates. The correlation between mortgage rates and

average housing prices was also not evident from an initial look at the data.

An initial look at the data suggested that there were fluctuations in the statistics in 2008 and 2020, which were both periods of economic recession. Some additional research had to be conducted about the economic and financial conditions over the past fifty years in order to provide possible explanations for the evident fluctuations in the datasets, which were present in all the datasets over a series of key dates. Comparing the datasets side-by-side using graphs and scatter plots was done in order to further help illustrate the datasets and help determine which statistical methods, such as linear regression, might be useful in analyzing the correlation between variables.

By deriving the rate of inflation from the CPI data and analyzing the correlation between mortgage rates and average house prices, the analysis gains insight into the effects of inflation on housing prices nationwide.

3. Analysis

The primary goal of the analysis is to determine the correlation between housing prices and the rate of inflation. Linear regression analysis and k-means clustering were the primary methods used to determine the amount of correlation between different variables.

3.1 Exploratory Visualizations

All four of the datasets were imported as csv files into R and were properly formatted and free of any obvious errors or deviations. In terms of data cleaning, the different time spans of the datasets had to be taken into account, as well as the time increments, since the regional HPI datasets were reported quarterly while all the other datasets were reported on a monthly basis. The months were labeled numerically as a separate variable, as well as the quarters in the regional HPI datasets. In order to account for this discrepancy, the month and the quarter variables and the year variables were updated as decimal numbers. For example, data listed for year: 1991 month: 6 was presented as 1991.5 in the year variable. Updating the year variable had to be done in order to make the datasets amenable to further analysis.

The national HPI data was downloaded as a single dataset containing thousands of entries for all fifty states and over 400 cities across the country. The HPI data for the Mountain Census Region (Montana, Idaho, Wyoming, Utah, Colorado, Nevada, Arizona, New Mexico), Colorado, Denver, and Boulder were all subsetted as new datasets from the national HPI dataset.
The first analysis that was performed was a simple graphical representation of each individual dataset, in order to help visually determine what correlations were present between the different datasets. A direct comparison between the variables was possible due to the fact that all the datasets are measuring the change of a variable over time. Plots were created for all six datasets.

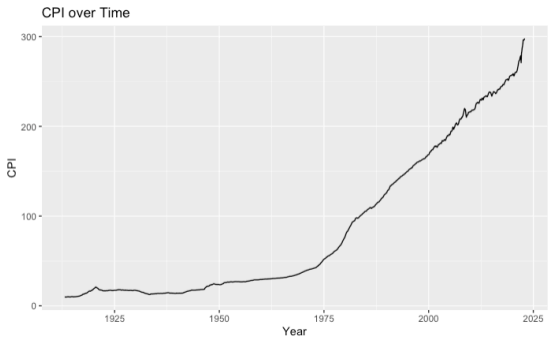


Figure 1: Graphical representation of CPI from 1913-2022.

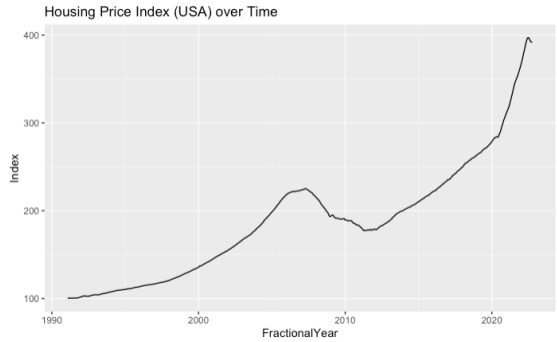


Figure 2: Graphical representation of national HPI data from 1991-2022.

Figure 3: Graphical representation of 30-year fixed mortgage rates from 1971-2022.

The most obvious association drawn from the initial look at the data was that trends in HPI were similar across regions and were comparable to the national data, with only small variations in overall trends. (Additional graphs located in appendix). Both the CPI scores and the HPI scores are rising over time, but there is no obvious association between them to suggest that they are directly correlated. Mortgage rates seem to have no correlation between either of these variables. Two scatter plots were created to further illustrate the correlation between national and regional HPI's over time. An obvious association between different regional HPI's can be seen when graphed against the CPI.

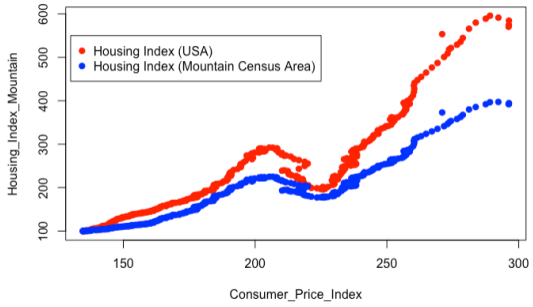


Figure 4: Scatter plot showing correlation between the national HPI and the Mountain Census Area HPI, graphed against the CPI. Points for overlapping time periods in the three datasets.

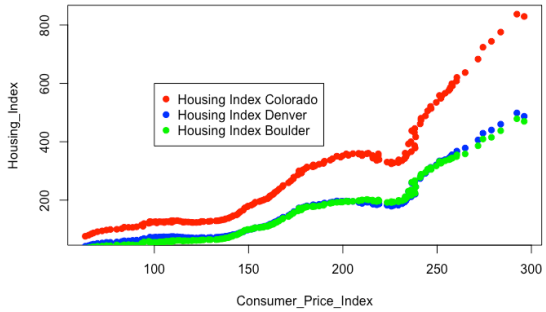


Figure 5: Scatter plot showing correlation between the regional HPI's (Colorado, Denver, Boulder) graphed against the CPI. Points for overlapping time periods in the three datasets.

The data was reorganized into new dataframes, with new variables created to show inflation rates. Consumer price inflation rates and housing price inflation rates were determined by calculating the percentage changes in prices on the date intervals. Five new data frames were organized (df, df1, df2, df3, df4). These data frames were organized to ease the

comparison between different variables. For example, df2 contains the HPI's for Colorado, Denver, and Boulder, the housing inflation rates for these three areas, mortgage rates, and CPI and consumer inflation rates. To determine which variables may be candidates for further analysis, a correlation plot was made. The variables with the highest correlation score were noted.

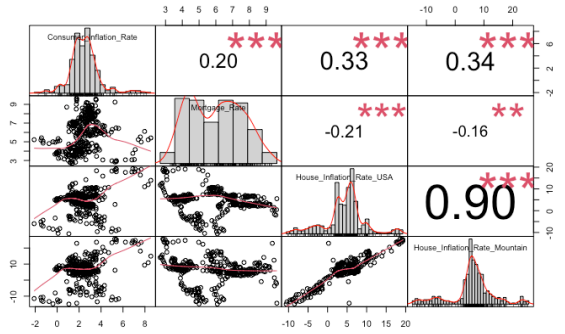


Figure 6: Example of a correlation plot. This correlation plot shows data for consumer price inflation rates, mortgage rates, national housing price inflation rates, and Mountain Census Area housing price inflation rates. Variables with high correlation have the highest correlation score. Scatter plots showing the correlation were also generated.

All of the housing price inflation rate datasets when compared against different regions in Colorado showed high correlation scores. When compared individually to the national housing price inflation rate, the regional housing price inflation rate scores showed a much lower correlation, averaging with a score around 0.5. There was correlation between the consumer price inflation rate and the housing price inflation rate scores, with the correlation score between 0.3-0.4. Regional and national housing inflation rates had a correlation score between 0.25-0.35 when compared to consumer inflation rates. Surprisingly, housing inflation rates had a very small negative correlation when compared to mortgage rates, which went against the initial hypothesis that these variables may have a statistically significant correlation.

3.2 Linear Regression

The results of the exploratory visualizations showed that certain variables were good candidates for further analysis. Linear regression analysis was performed to more accurately determine the relationship between variables by finding a line of best fit between data points. The goal of the linear regression analysis was to end up with a line that best fits the data. The total prediction error is minimized, and is depicted on the graph as the shortest distance between each data point and the regression line, when two variables are plotted against each other.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Equation 1. Linear regression equation. Y is all the observed values for the dependent variable, β_0 is the y-intercept (bias), β_1 is the slope (coefficient), X is all observed values of the independent variable, and ε is the error.

The Pearson correlation coefficient is displayed for each linear regression analysis that was made. The Pearson correlation coefficient describes the association between variables and is based on the method of covariance.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Equation 2. Pearson correlation coefficient equation. r is the Pearson correlation coefficient, x_i is the x variable samples, x bar is

the mean of values in the x variable, y_i is the y variable sample and y bar is the mean of values in the y variable.

Additionally, a number of statistics are calculated including the adjusted R-squared statistic which shows how well the regression model explains observed data.

With the linear regression analysis a more accurate relationship between the different variables could be determined.

3.3 K-means Clustering

K-means is a form of cluster analysis that involves partitioning a group of observations into *k* amounts of clusters. Each observation belongs to the cluster with the nearest centroid. The k-means algorithm starts out with an initial group of randomly selected centroids and uses an iterative process to optimize the centroids. An initial amount of centroids is selected, and the optimal amount of centroids for the dataset is determined by generating graphs with different numbers of centroids. The optimal amount of clusters is determined, and these clusters can be selected for further linear regression analysis to determine the correlation between variables. The goal of k-means clustering in the analysis is to further refine the similarity between variables, and to take clusters of similar points and determine how correlated they are.

4. Results

The primary analyses used to determine the relationships between different statistical variables were linear regression and k-means clustering. Several meaningful associations were discovered.

4.1 Linear Regression

Linear regression was used to investigate associations between several different variables.

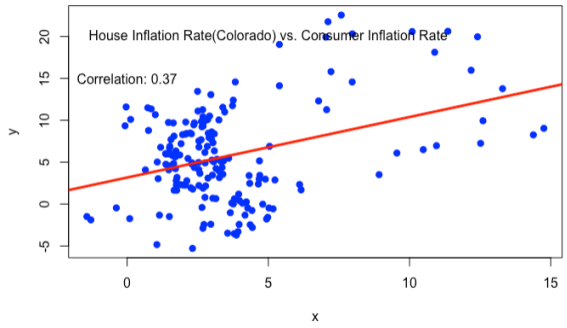


Figure 7: Linear regression line of housing inflation rate in Colorado vs the consumer inflation rate. Pearson correlation coefficient of 0.37. Adjusted R-squared: 0.1329.

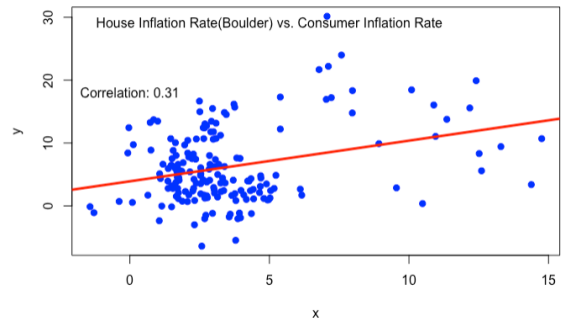


Figure 8: Linear regression line of housing inflation rate in Boulder vs the consumer inflation rate. Pearson correlation coefficient of 0.31. Adjusted R-squared: 0.09375.

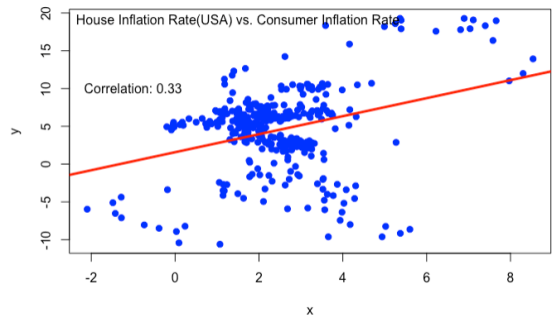


Figure 9: Linear regression line of the national housing inflation rate vs the consumer inflation rate. Pearson correlation coefficient of 0.33. Adjusted R-squared: 0.1042.

This linear regression analysis sheds light on the relationship between regional housing inflation rates and the national consumer inflation rate. House inflation rates and consumer inflation rates are mildly correlated, based on the value of the Pearson correlation coefficient, implying that the rising cost of consumer goods is associated with the rising cost of single family homes. The correlation coefficient of Colorado data and Boulder data is similar to the coefficient of the national data, with the Colorado coefficient slightly higher and the Boulder coefficient slightly lower than the national average inflation coefficient. In all cases the adjusted R-squared is very low implying that this mild level of correlation is not well explained by the linear model.

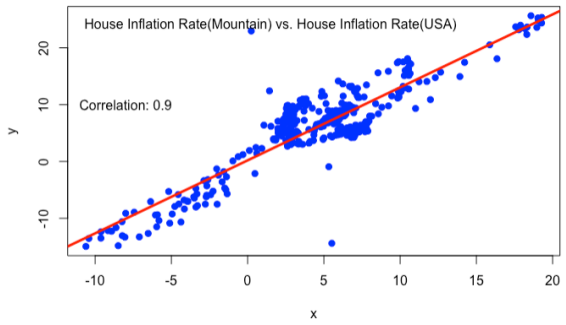


Figure 10: Linear regression line of housing inflation rate (Mountain Census Region) plotted against the national housing inflation rate. Correlation coefficient of 0.9. Adjusted R-squared: 0.8099.

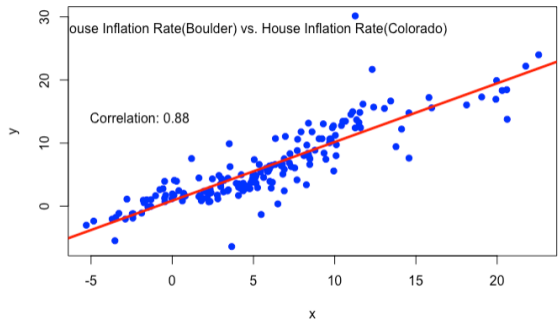


Figure 11: Linear regression line of Boulder housing inflation rate plotted against the Colorado housing inflation rate. Correlation coefficient of 0.88. Adjusted R-squared: 0.7732.

The graph comparison chart showed that regional housing inflation rates were strongly correlated with national house inflation rates, and this was shown in the results of the linear regression analysis. Housing inflation rates in the Mountain Census Region are strongly correlated with national housing inflation rates, with a coefficient of 0.9. Boulder housing inflation rates are also strongly correlated with national house inflation rates with a correlation coefficient of 0.88. In both cases the adjusted R-squared is close to 0.8. This goes against the initial prediction that Boulder housing inflation rates might have been slightly higher than the state averages and national averages, due to the Boulder area being a desirable location. In fact, all three housing inflation rates are strongly correlated.

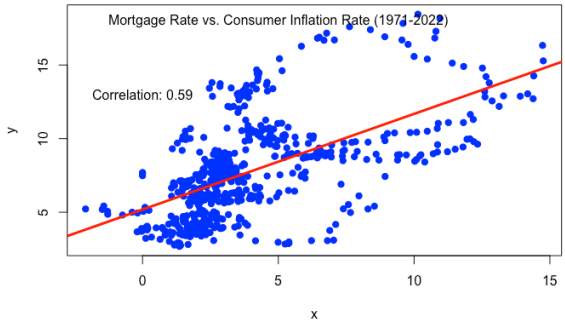


Figure 12: Linear regression line of mortgage rate plotted against the consumer inflation rate. Correlation coefficient of 0.59. Adjusted R-squared: 0.3507.

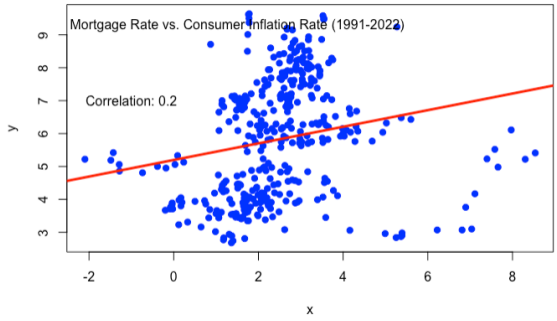


Figure 13: Linear regression line of mortgage rate plotted against the consumer inflation rate. Correlation coefficient of 0.2. Adjusted R-squared: 0.03833.

Another interesting association that was discovered using linear regression was the relation between mortgage rates and the consumer inflation rate. When using mortgage rate data stretching back to 1978, these two variables are moderately correlated, with a correlation coefficient of 0.59 and the adjusted R-squared of 0.35. When performing the same regression analysis, but with data stretching back to 1991, the correlation between mortgage rates and the consumer inflation rate is much lower, with a correlation coefficient of 0.2. The adjusted R-squared of 0.04 signifies essentially no linear relationship between these two rates for this time period. This implies that between 1978-1991, mortgage rates were associated with the consumer inflation rate, since the data points from this time period have such heavy weight on the line of best fit. This implies that since the 1990s, this association has significantly weakened. These findings led to some further investigation using cluster analysis, to determine exactly which data points were correlated.

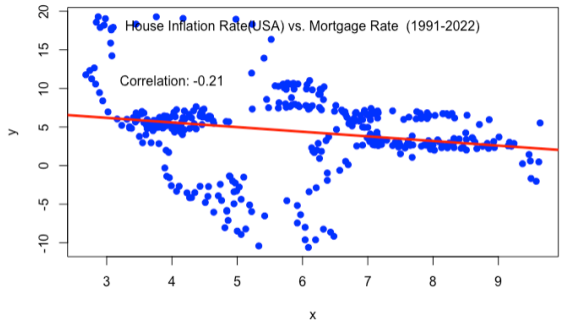


Figure 14: National housing inflation rate plotted against the mortgage rate. Correlation coefficient of -0.21. Adjusted R-squared: 0.03955.

This analysis shows that national housing inflation rates had no correlation with mortgage rates. This finding is interesting, considering that consumer inflation rates had a much higher correlation with mortgage rates, when comparing data from the same timeframe.

4.2 K-means Clustering

K-means clustering used varying numbers of centroids in each analysis that was conducted. The chart with the optimal number of centroids is shown in this section, with the other amounts of centroids shown in the appendix.

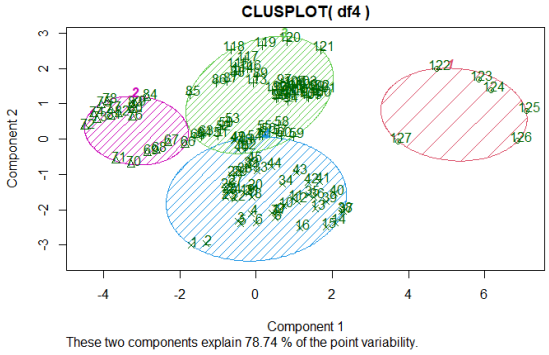


Figure 15: Clustering plot using data from df4, dataframe that contains all inflation rate indexes together. Consumer inflation rate

and mortgage rates included. Data from 1991-2022. Four clusters found to be optimal.

A k-means analysis was conducted on df4, the data frame that contains all inflation data combined, including housing inflation, mortgage rates and consumer inflation. Four distinct clusters were created in this chart containing data from 1991-2000. Further inspection of the clusters formed found that each cluster corresponds to a specific date range. Regression analysis was performed on the individual clusters to determine the exact correlation between mortgage rates and consumer inflation rates in the specific date range.

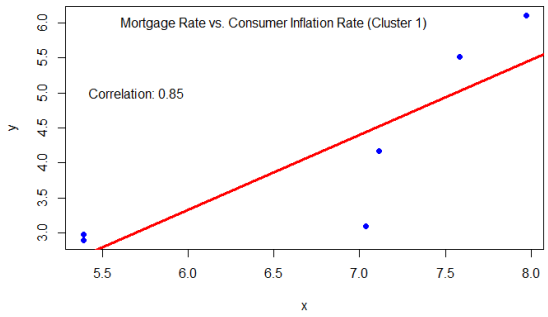


Figure 16: Linear regression graph of cluster 1 corresponding to the date range between 2021 and 2022. Correlation coefficient is 0.85. Adjusted R-squared: 0.6468.

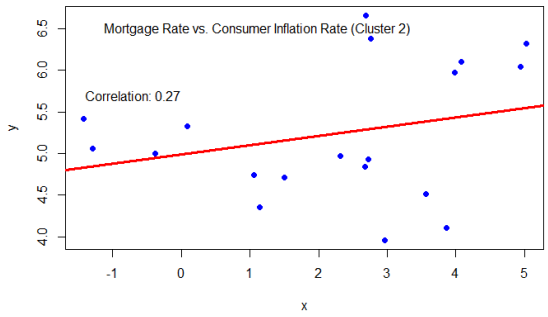


Figure 17: Linear regression graph of cluster 2 corresponding to the date range between 2007 and 2012. Correlation coefficient is 0.27. Adjusted R-squared: 0.01752.

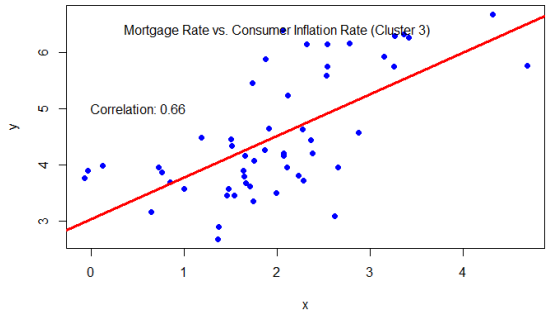


Figure 18: Linear regression graph of cluster 3 corresponding to the date range between 2003-2007 and 2012-2021. Correlation coefficient is 0.66. Adjusted R-squared: 0.4182.

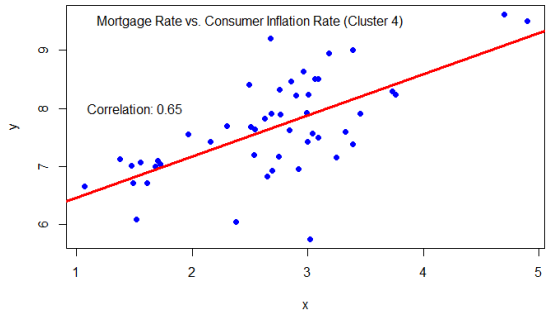


Figure 19: Linear regression graph of cluster 4 corresponding to the date range between 1991 and 2003. Correlation coefficient is 0.65. Adjusted R-squared: 0.4151.

The correlation between mortgage rates and consumer inflation rates varies significantly between these date ranges. The correlation was moderate between 1991 and 2022, with a correlation coefficient of about 0.7 and the adjusted R-squared of about 0.4. There was no correlation between mortgage rates and consumer inflation rates in the period around the housing crisis between 2007 and 2012, most likely owing to an unprecedented amount of liquidity provided to the economy as part of the banking system bail out. The most recent data between 2021 and 2022 is an outlier in the dataset, with a significantly stronger correlation between the two variables but the mortgage rates much lower

than the consumer inflation rates. This implies that the mortgage rates are likely to increase if the historical trends were to apply. These findings led to the generation of a second k-means clustering analysis, but with data that reaches back further in time to 1971 comprising just the mortgage rates and the consumer inflation rates.

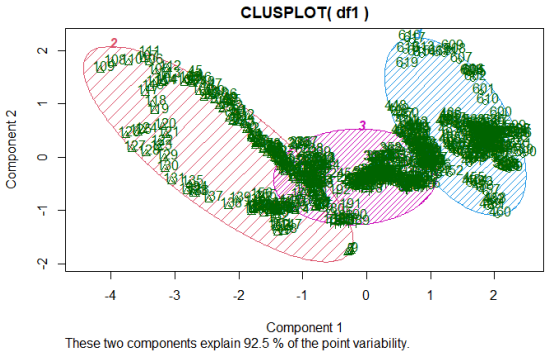


Figure 20: Clustering analysis for the 1971-2022 data with only the consumer inflation rate and mortgage rates (df1, monthly). Three clusters found to be optimal.

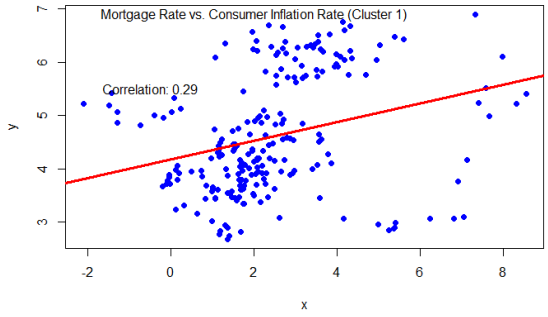


Figure 21: Linear regression graph of cluster 1 corresponding to the date range between 2004 and 2022. Correlation coefficient is 0.29. Adjusted R-squared: 0.07853.

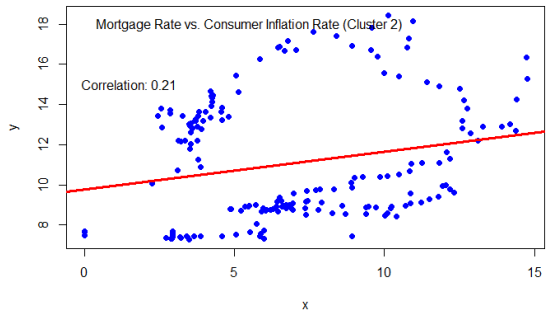


Figure 22: Linear regression graph of cluster 1 corresponding to the date range between 1971 and 1986. Correlation coefficient is 0.21. Adjusted R-squared: 0.04064,

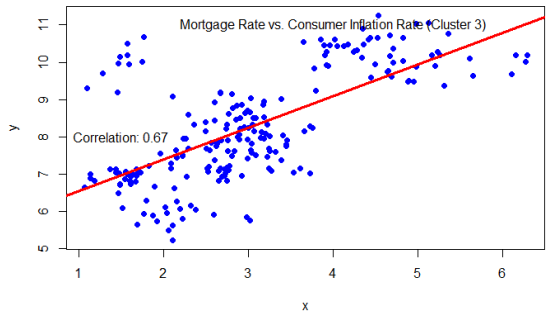


Figure 23: Linear regression graph of cluster 1 corresponding to the date range between 1986 and 2004. Correlation coefficient is 0.67. Adjusted R-squared: 0.4506.

Using a larger dataset that goes further back in time, the general trend discovered is that the association between mortgage rates varies with time. The two variables are moderately correlated in the time period between 1986 and 2004. The association between the two variables is particularly weak between 1971 and 1986 – the period of high inflation. It has also been generally weak since 2004, the period of time associated with low consumer inflation rates. It is interesting to observe that the data points from 2020-2022 are outliers, and are located on the lower right part of cluster 1,

likely skewing the regression results for cluster 1 representing the time period between 2004 and 2022.

The results of the analyses conducted show a couple of interesting findings. Changes in housing inflation rates are comparable between different locations in Colorado, and the Colorado rates are comparable to the national average rates. This finding was not initially expected, but the trends in the data show that there is a correlation between these housing inflation rates. Surprisingly, mortgage rates have little impact on the housing inflation rate. Mortgage rates are significantly impacted by consumer inflation rates, and this association has changed notably across different time periods.

5. Conclusion

Overall this report gained insight into the impact of various economic factors on housing prices using observational numeric data. The analysis was largely exploratory in nature and led to the discovery of some interesting associations between variables that were not initially predicted.
The inflation rate in Mountain Census Area housing prices, Colorado prices, Denver prices, and Boulder prices were all found to be closely correlated to the national average inflation rate in housing prices, which went against the initial prediction that Boulder and Denver price inflation rates would be significantly higher.

A recent study suggested a correlation between mortgage rates and housing prices resulting from government policies in the last two years since the COVID-19 pandemic. [5] In this project, an analysis of mortgage rates and consumer inflation rates found found that the correlation between mortgage rates and consumer inflation rates varies significantly across The correlation was moderate between 1991 and 2022. However, the most recent data between 2021 and 2022 is an outlier in the dataset, with a significantly stronger correlation between the two variables but the mortgage rates much lower than the consumer inflation rates.

This linear regression analysis shed light on the relationship between regional housing inflation rates and the national consumer inflation rate. House inflation rates and consumer inflation rates are mildly correlated, based on the value of the Pearson correlation coefficient, implying that the rising cost of consumer goods is associated with the rising cost of single family homes. The correlation coefficient of Colorado data and Boulder data is similar to the coefficient of the national data, with the Colorado coefficient slightly higher and the Boulder coefficient slightly lower than the national average inflation coefficient. In all cases the adjusted R-squared was very low implying that this mild level of correlation was not well explained by the linear model. These results confirm the studies earlier studies conducted in past decades that these two variables had a statistically significant correlation.[4,6]

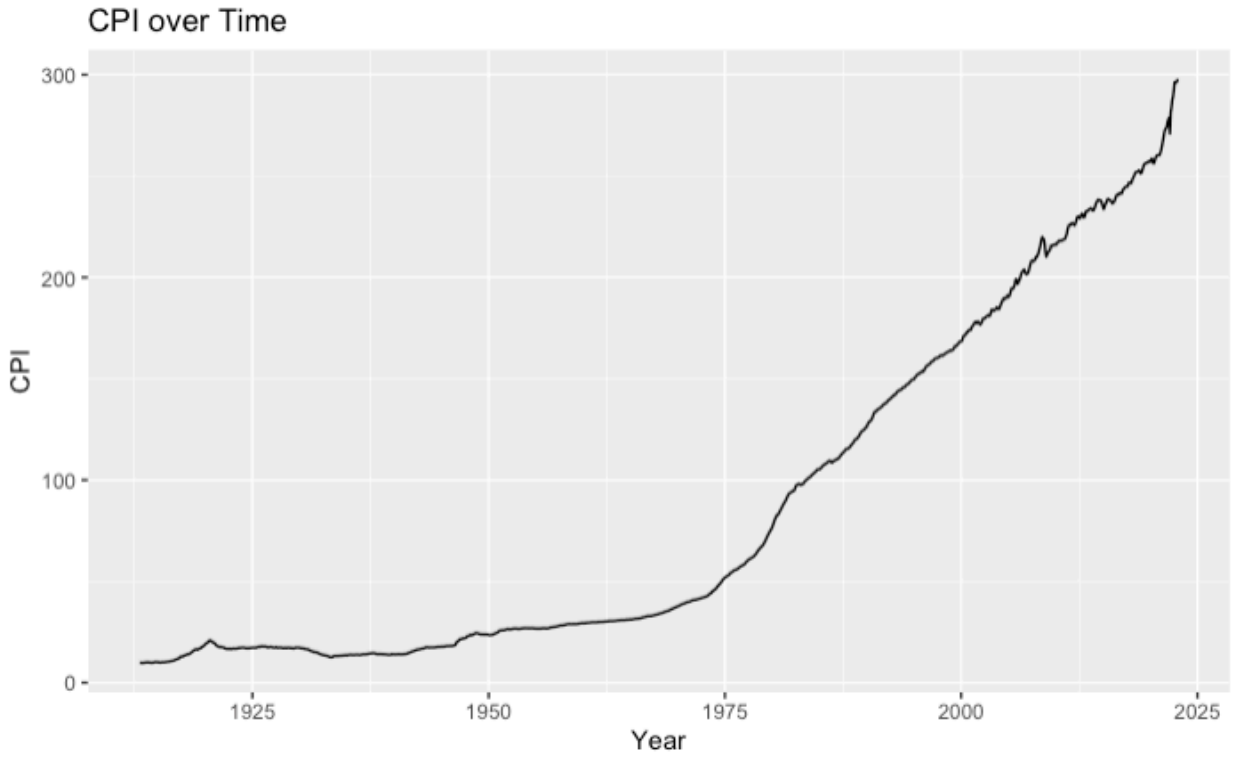
A suggestion for further research would be to conduct further research on historical factors that significantly impacted economic and housing factors in decades past, in order

to find probable causes for certain trends that were discovered in the datasets. The impact of government policies, fiscal policies, consumer trends, trends in real estate development and the economic conditions present at certain periods in time all likely had an impact on the datasets. Further research into these topics would help find possible explanations for some of the trends that were discovered in this analysis.

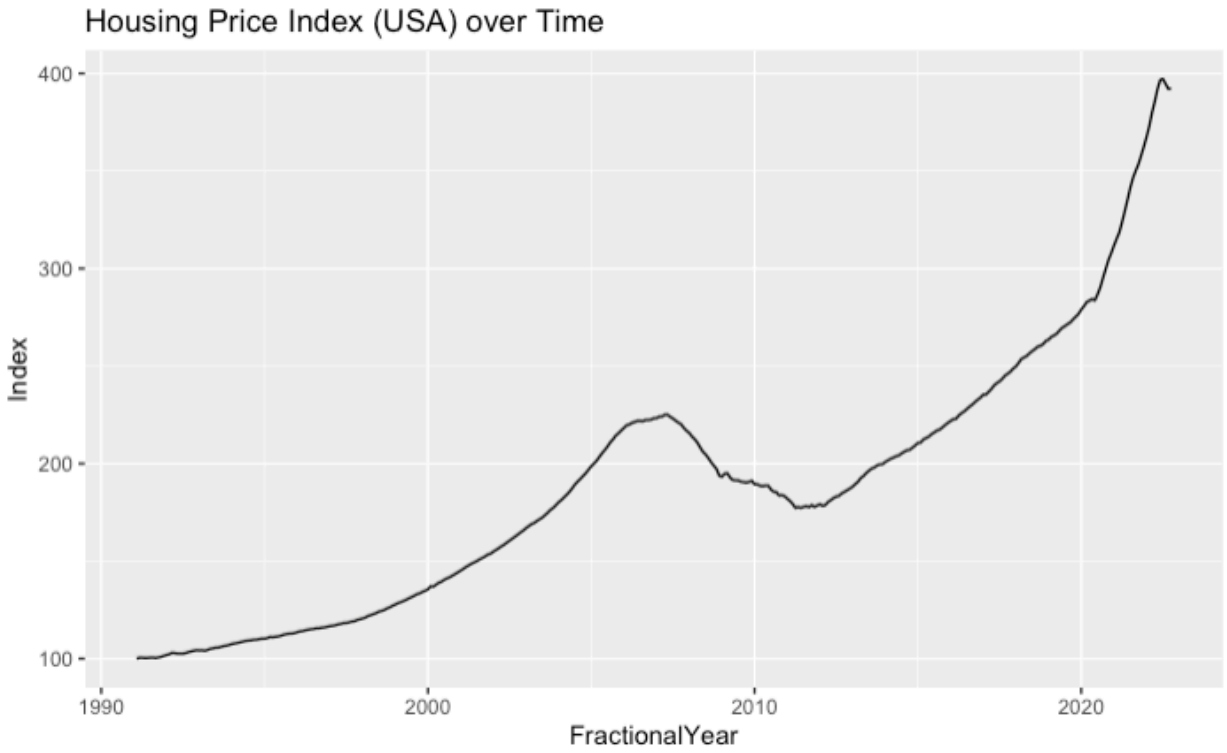
6. Appendix

The appendix shows all of the plots that were generated in the analysis, in greater resolution.

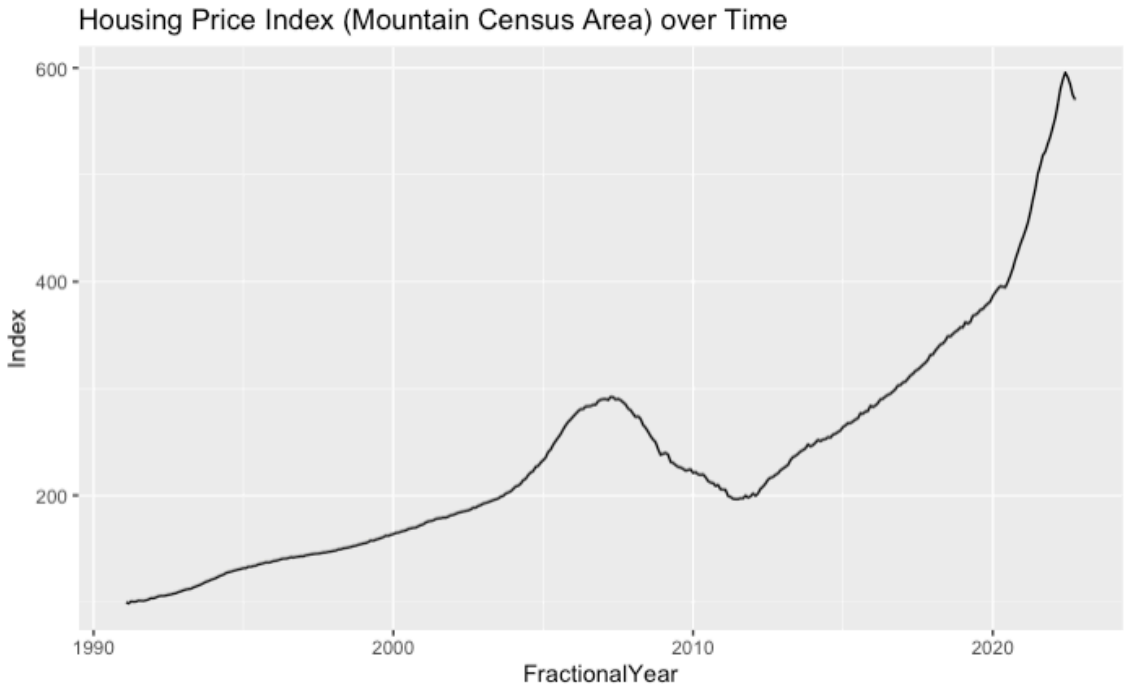
6.1 Plots



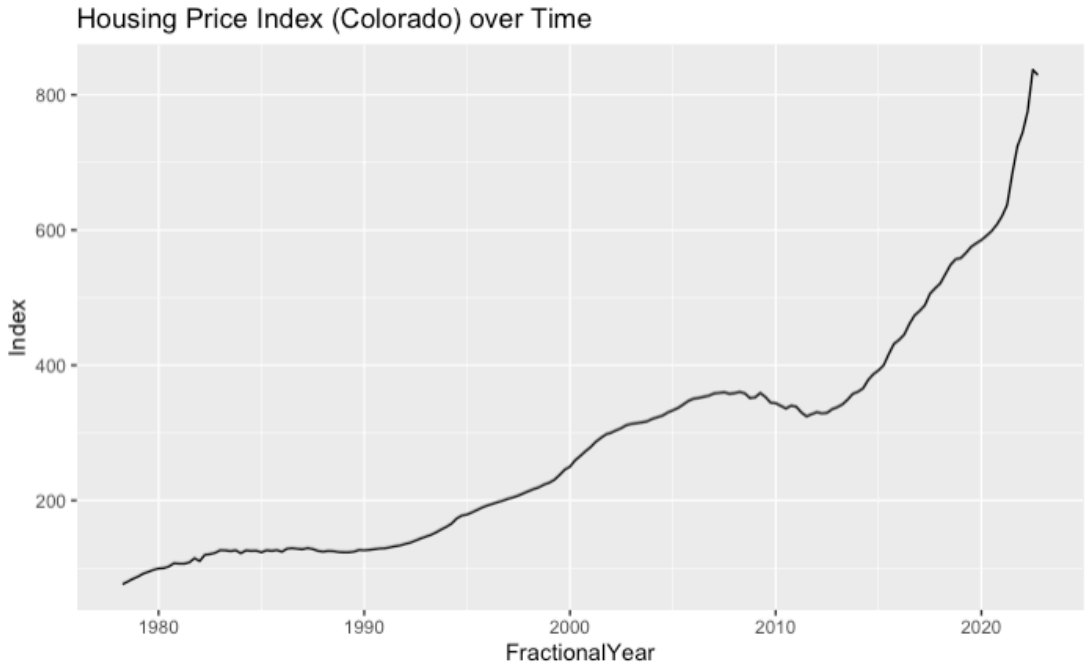
Linear graph representation of CPI dataset from 1913-2022.



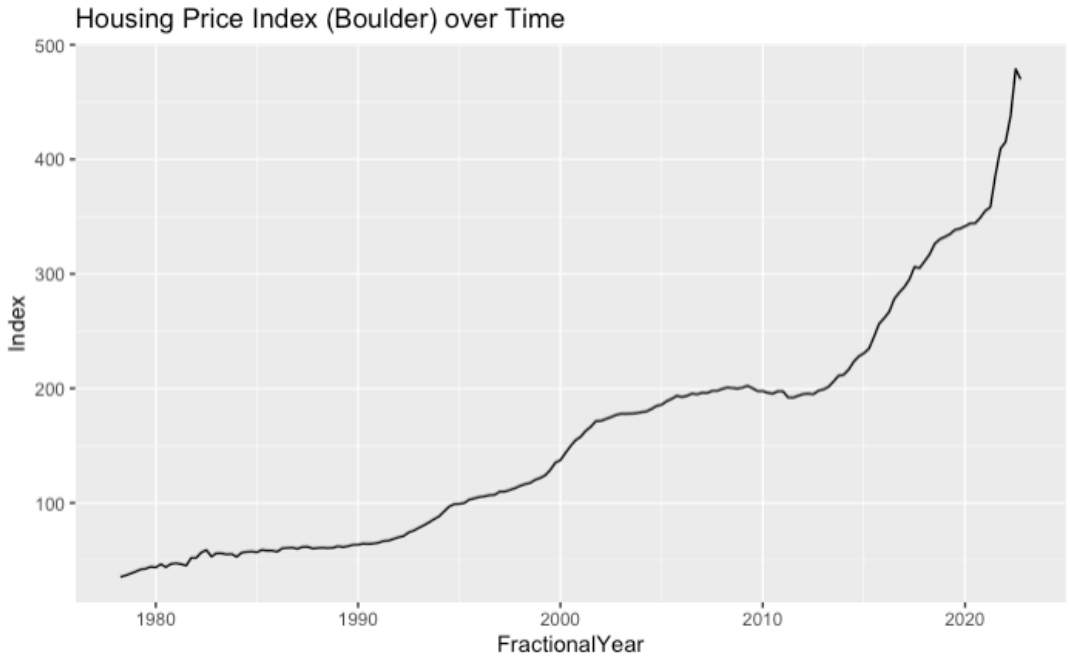
Linear graph representation of national HPI dataset from 1991-2022.



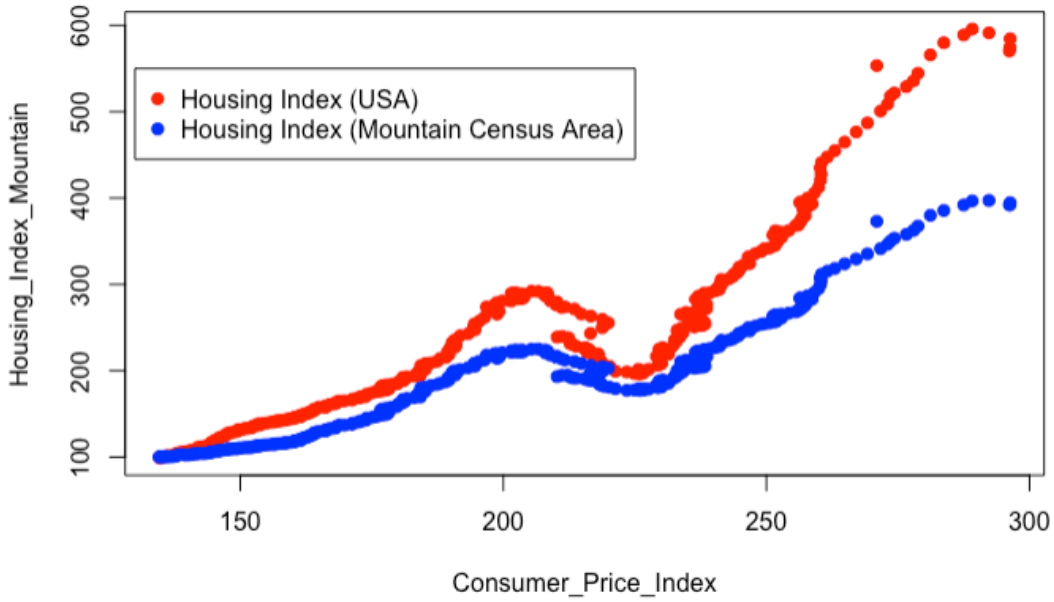
Linear graph representation of Mountain Census Area HPI dataset from 1991-2022.



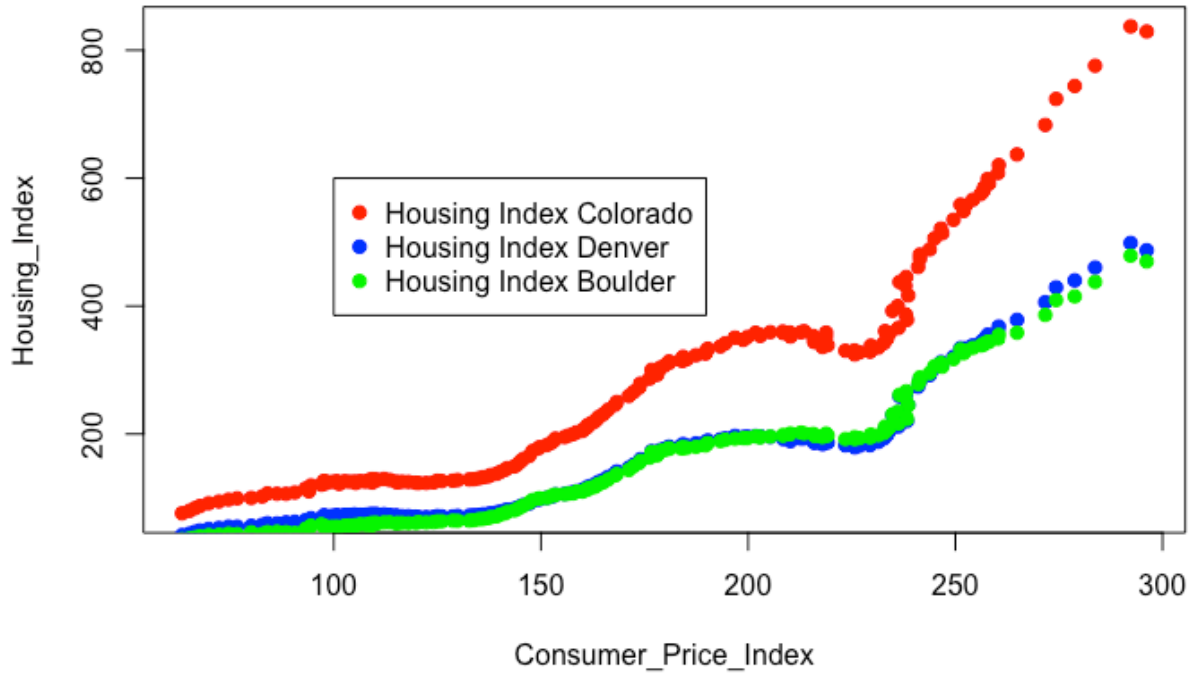
Linear graph representation of Colorado HPI dataset from 1978-2022.



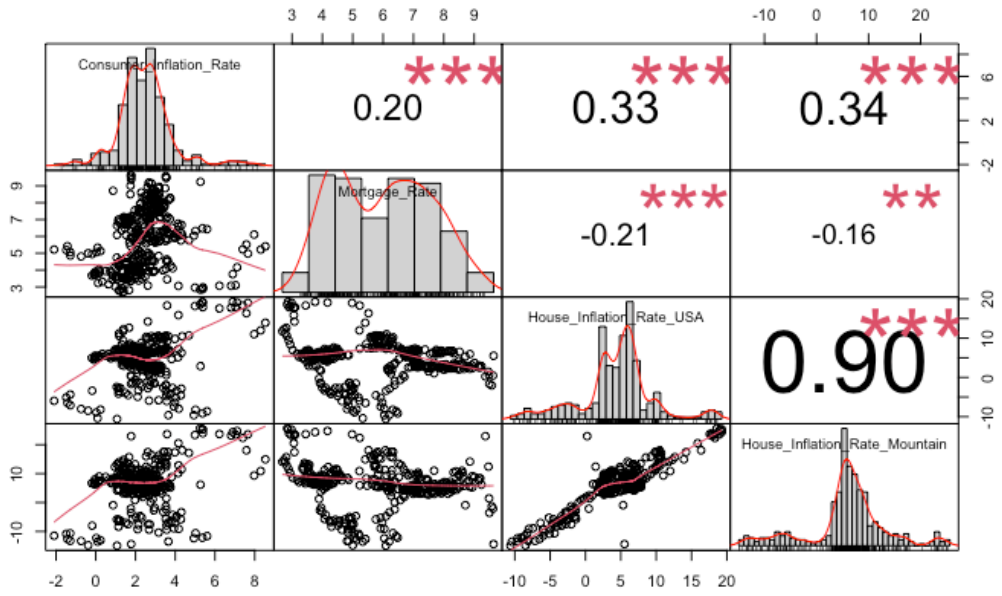
Linear graph representation of Boulder HPI dataset from 1978-2022.



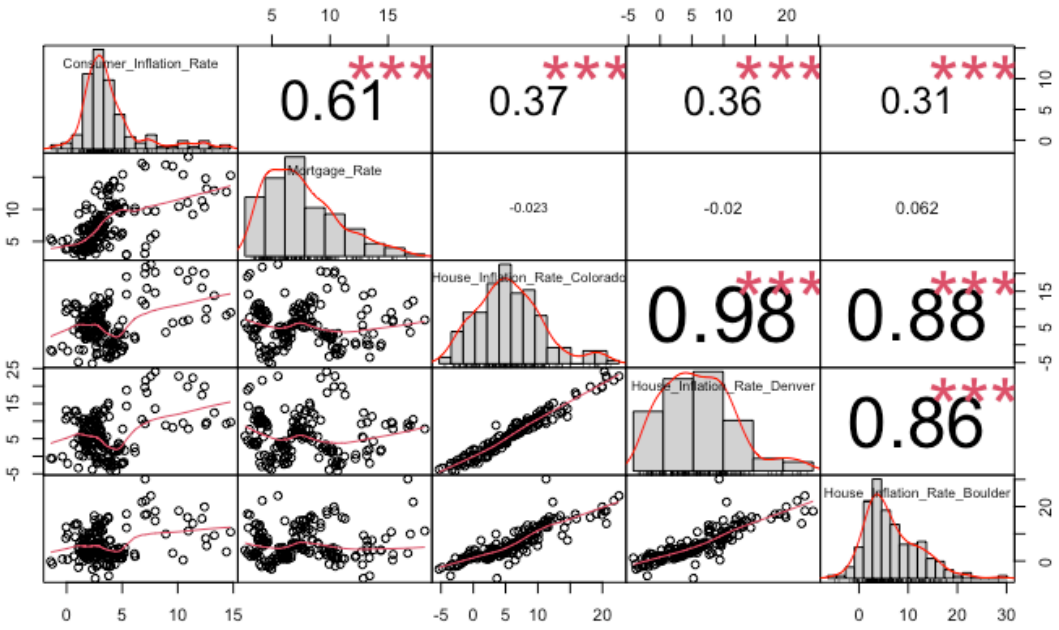
Scatter plot comparing the Mountain Area housing index and the CPI.



Scatter plot comparing the consumer inflation rate, mortgage rate, national house inflation rate, and the mountain house inflation rate.

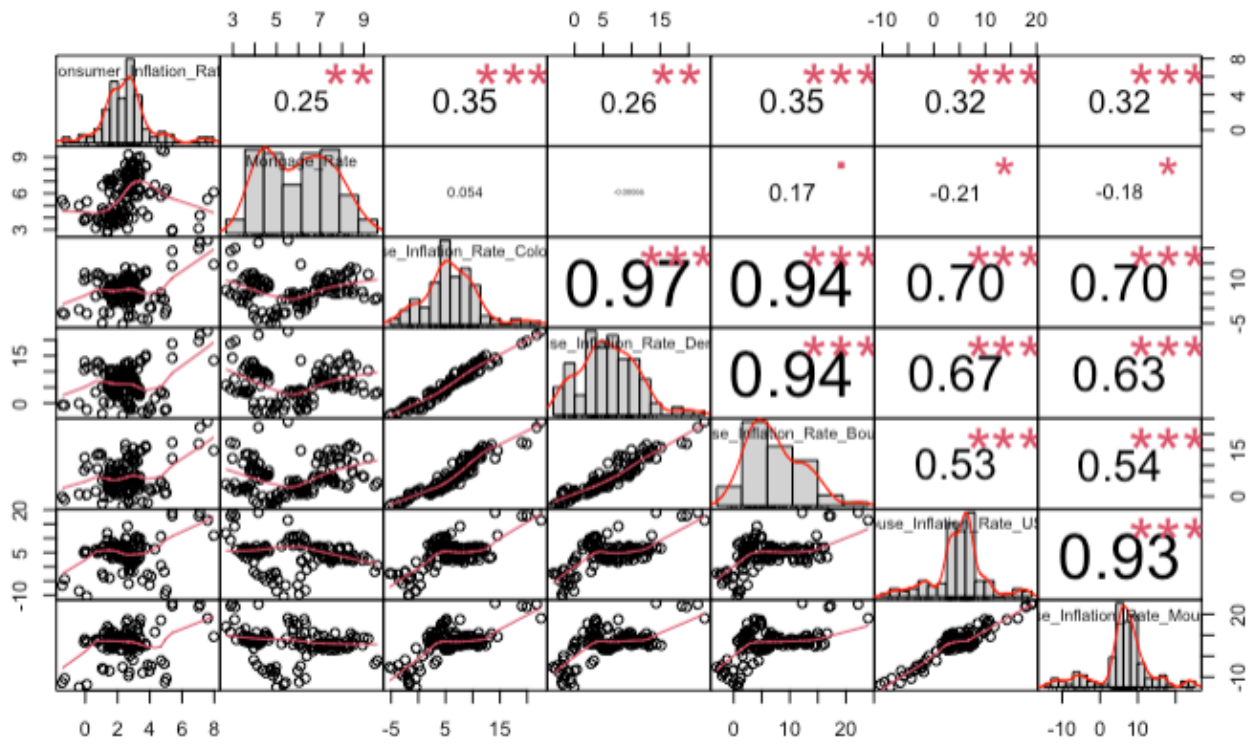


('Consumer_Inflation_Rate','Mortgage_Rate','House_Inflation_Rate_Colorado','House_Inflation_Rate_Denver',
'House_Inflation_Rate_Boulder')



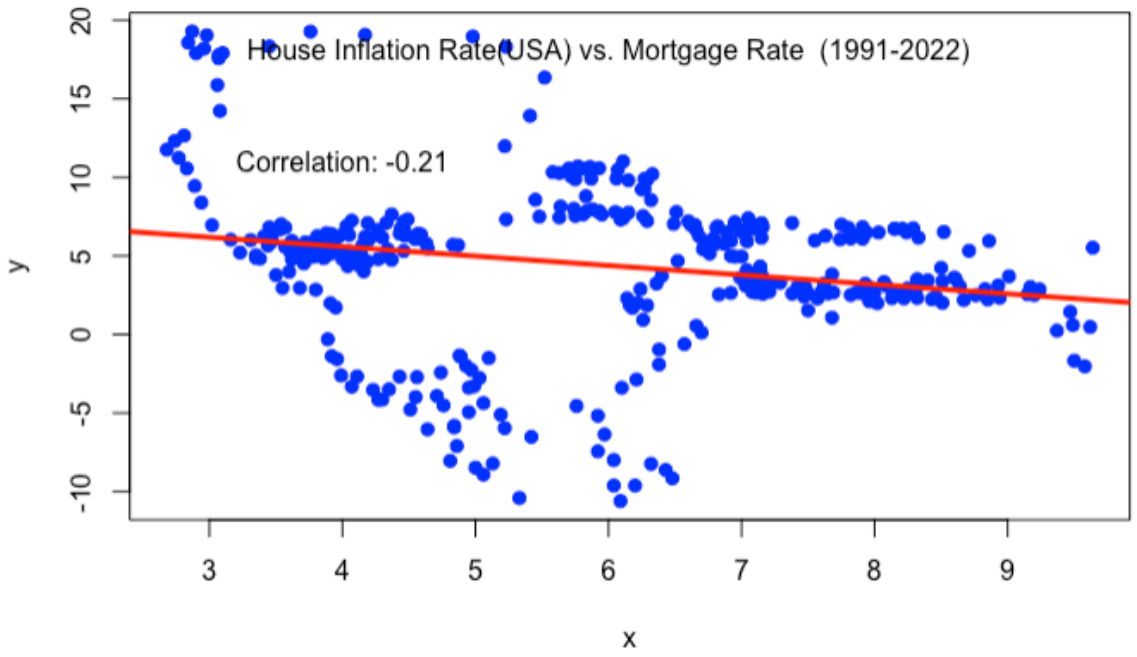
('Consumer_Inflation_Rate','Mortgage_Rate','House_Inflation_Rate_Colorado','House_Inflation_Rate_Denver',

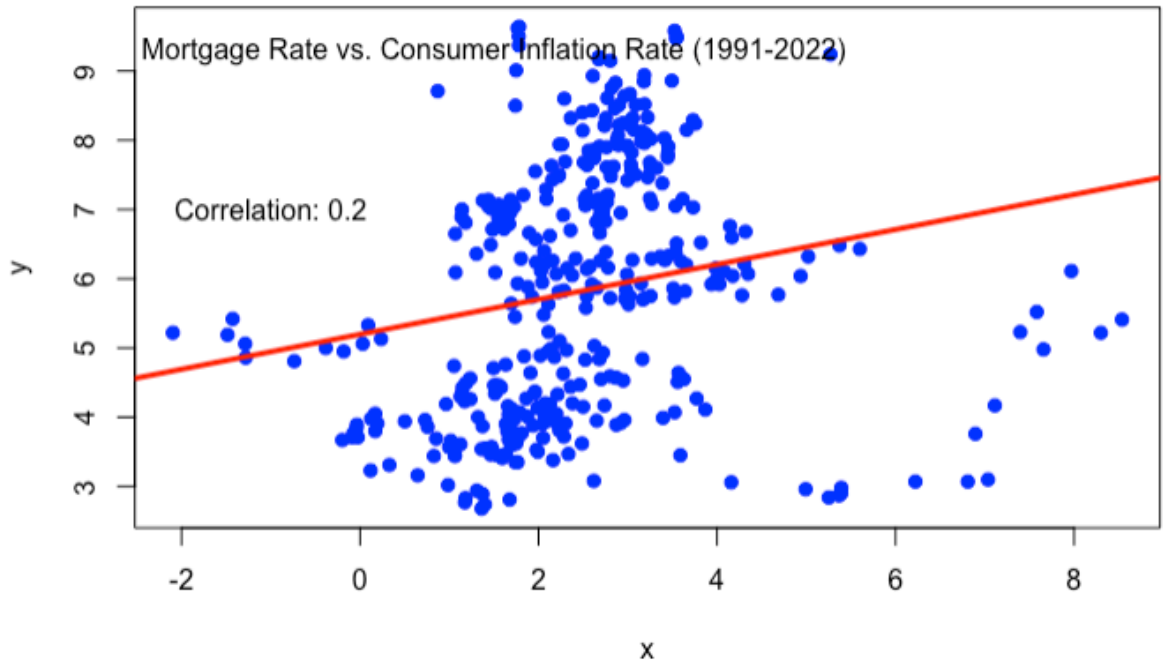
'House_Inflation_Rate_Boulder','House_Inflation_Rate_USA','House_Inflation_Rate_Mountain')

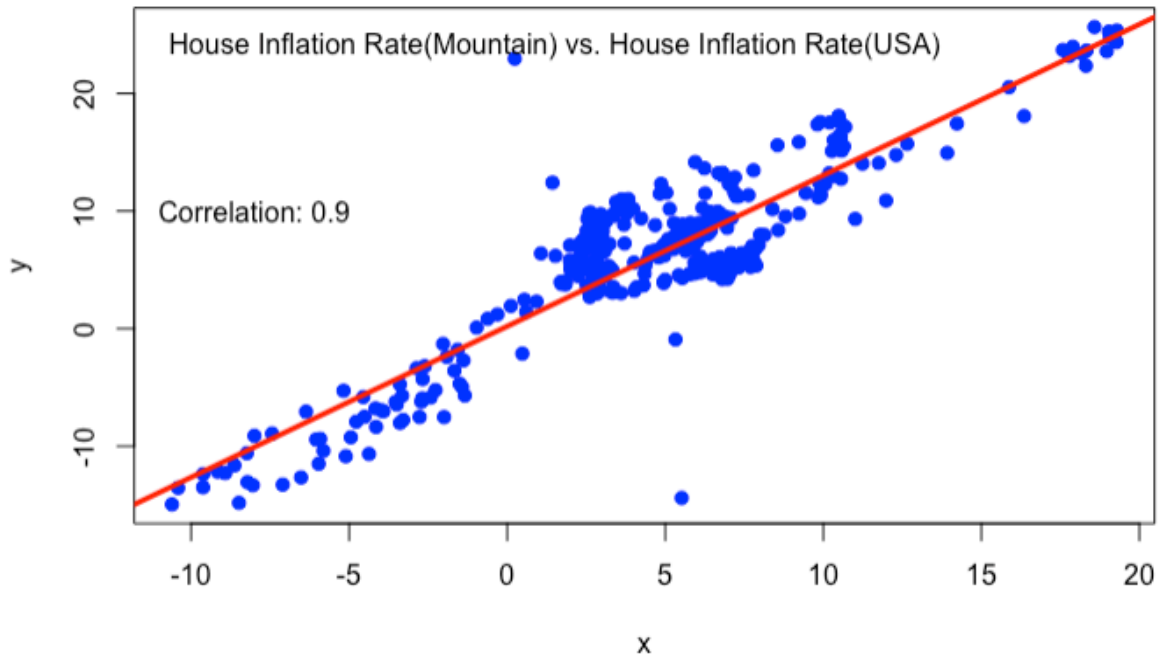


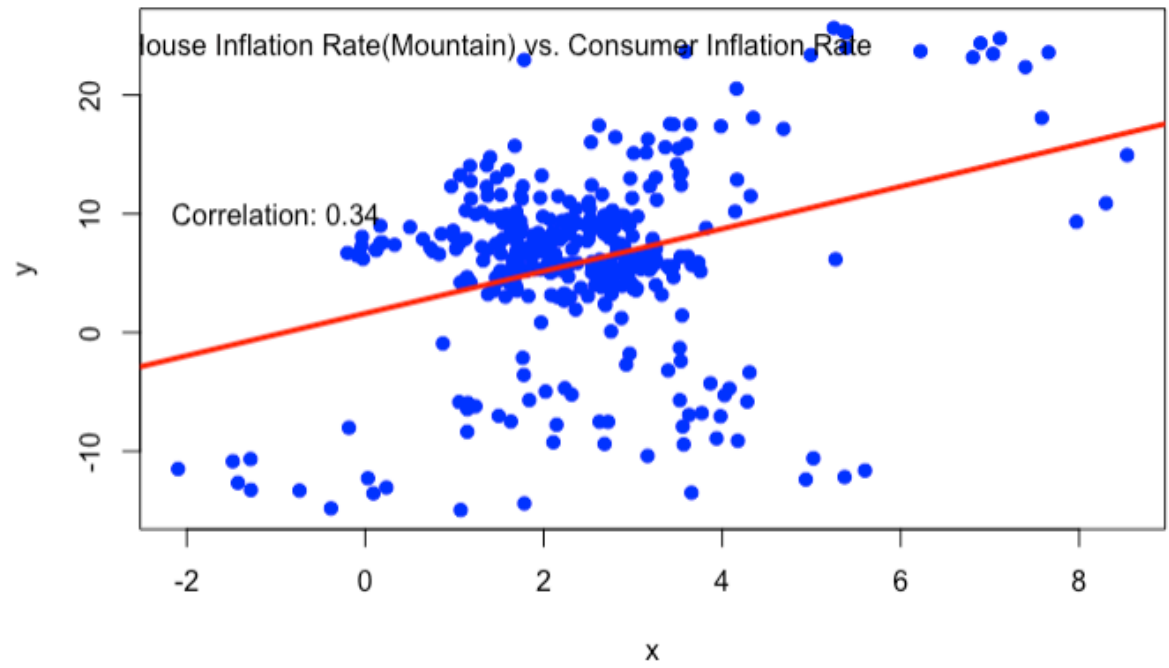
Correlations for monthly data 1991-2022

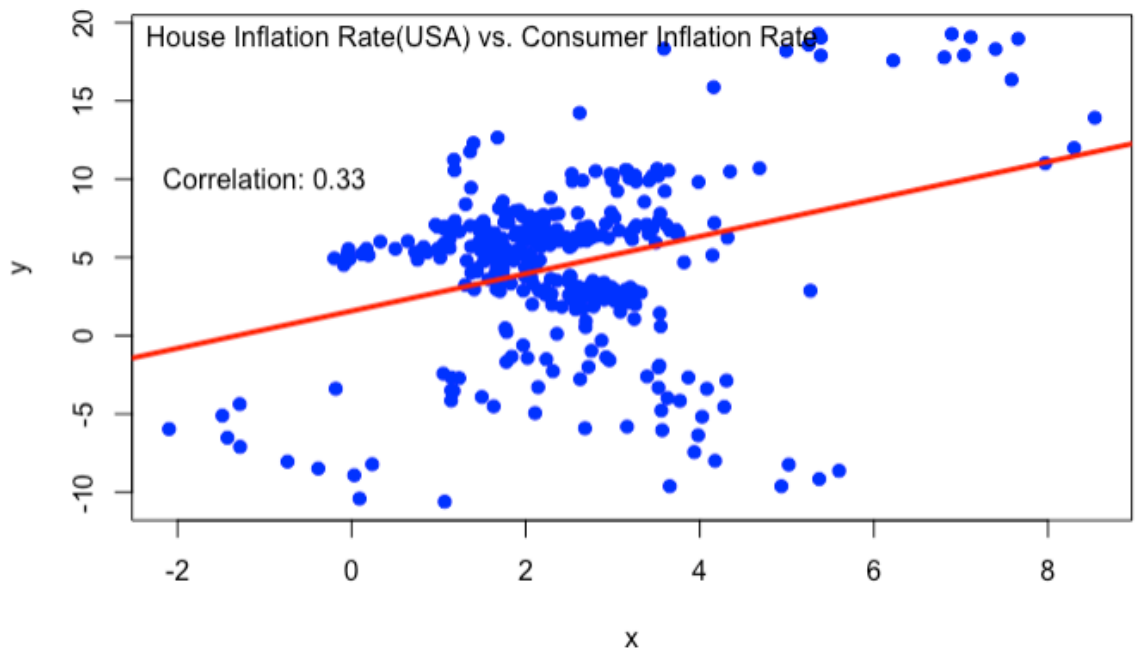
(Pearson correlation coefficient)


House Inflation Rate(USA) vs. Mortgage Rate (1991-2022)
Correlation: -0.21
y
x
20
15
10
5
0
-5
-10
3
4
5
6
7
8
9

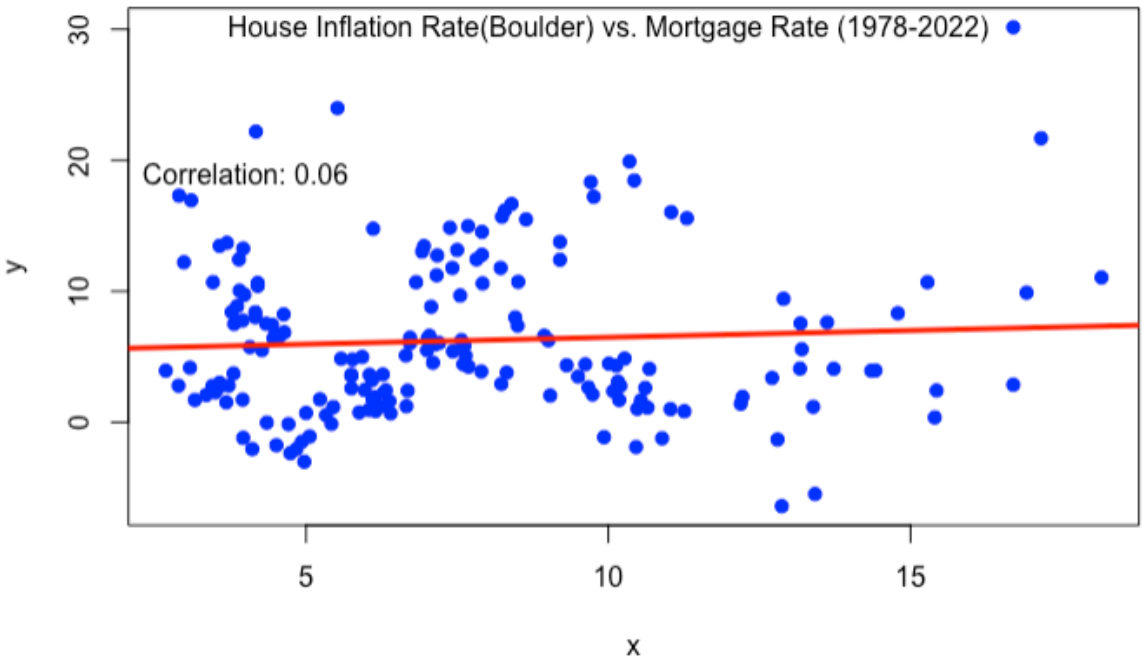

Mortgage Rate vs. Consumer Inflation Rate (1991-2022)
Correlation: 0.2
y
x
9
8
7
6
5
4
3
-2
0
2
4
6
8

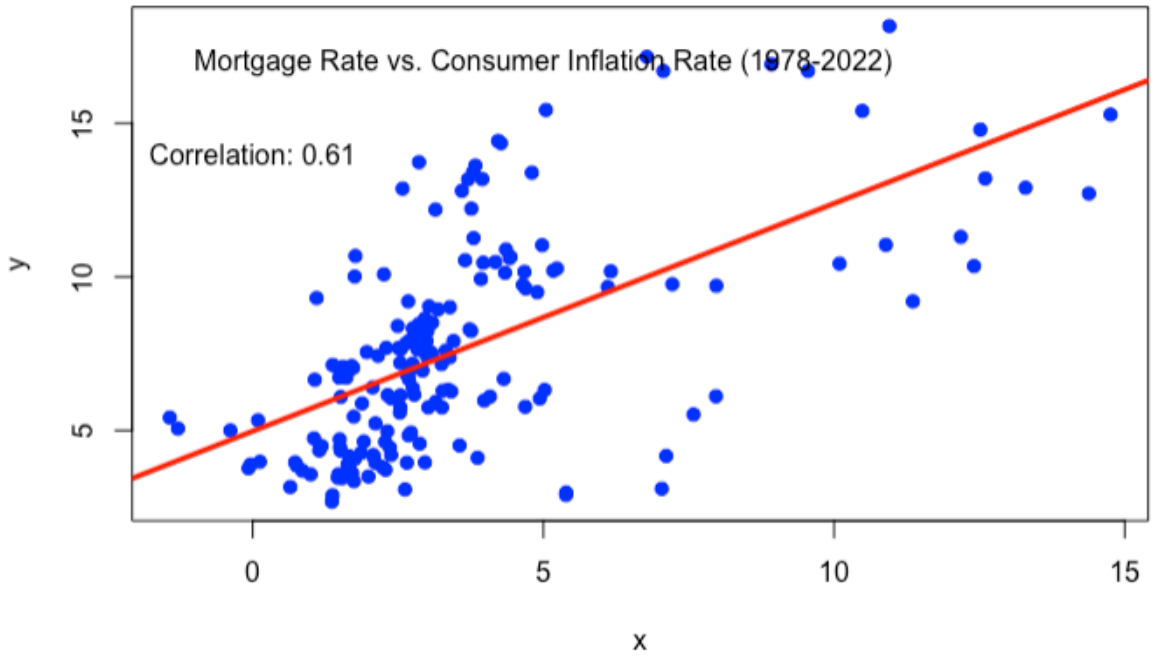

House Inflation Rate(Mountain) vs. House Inflation Rate(USA)
Correlation: 0.9
y
x
20
10
0
-10
-10
-5
0
5
10
15
20

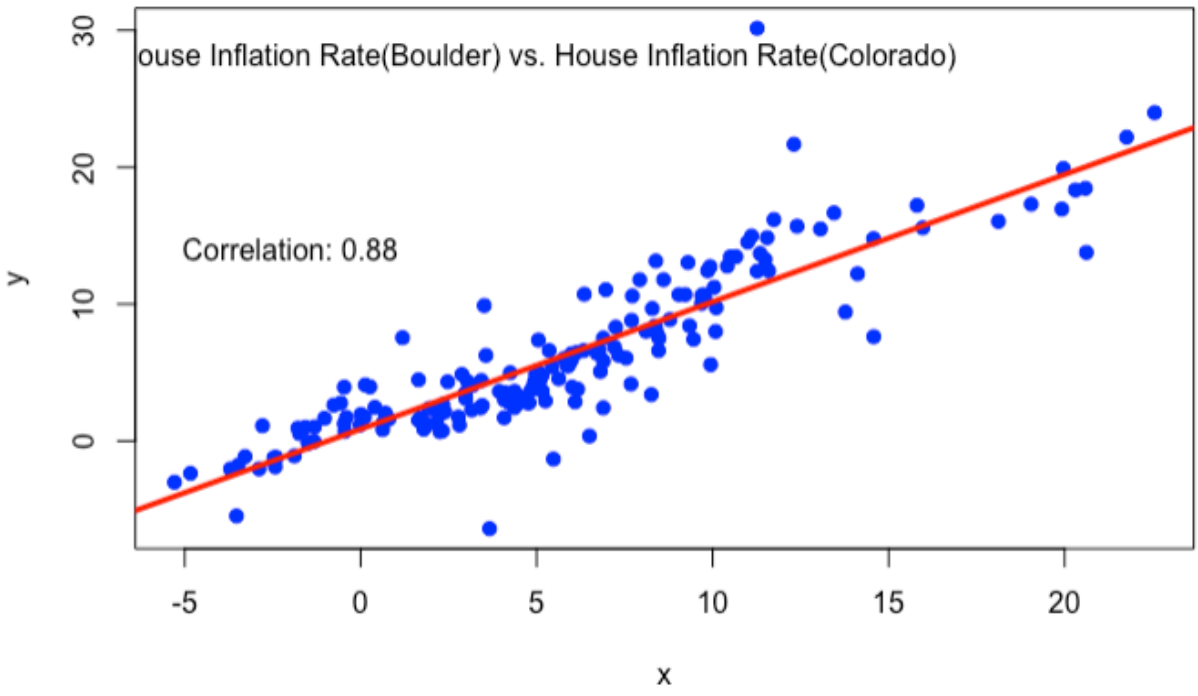

louse Inflation Rate(Mountain) vs. Consumer Inflation Rate
Correlation: 0.34
y
x
20
10
0
-10
-2
0
2
4
6
8

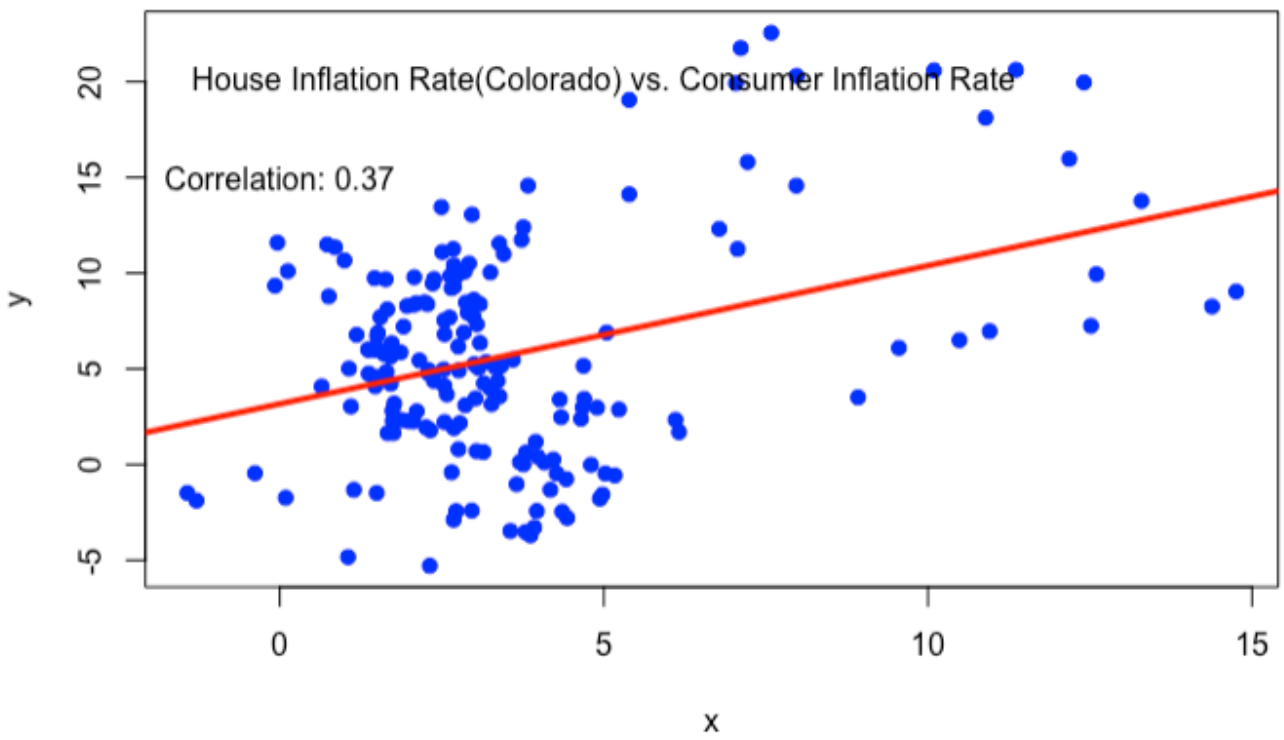


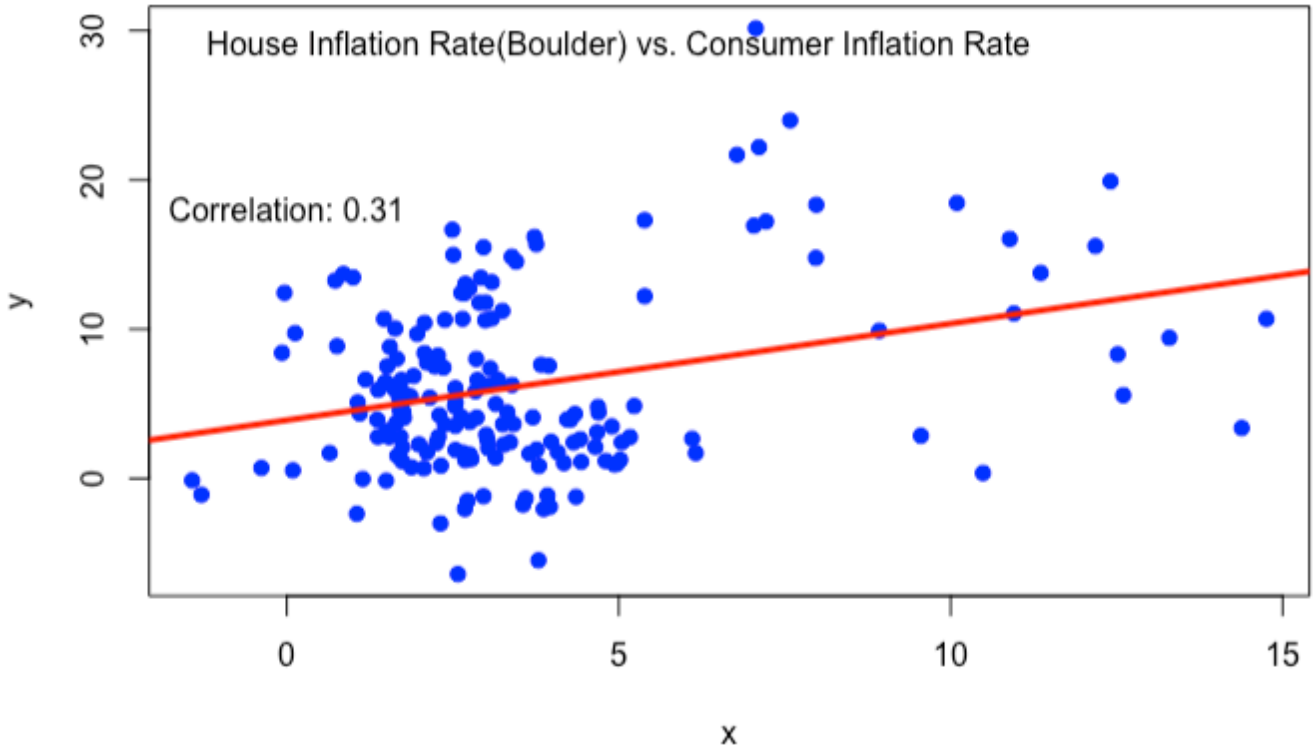
Correlations for quarterly data 1978-2022



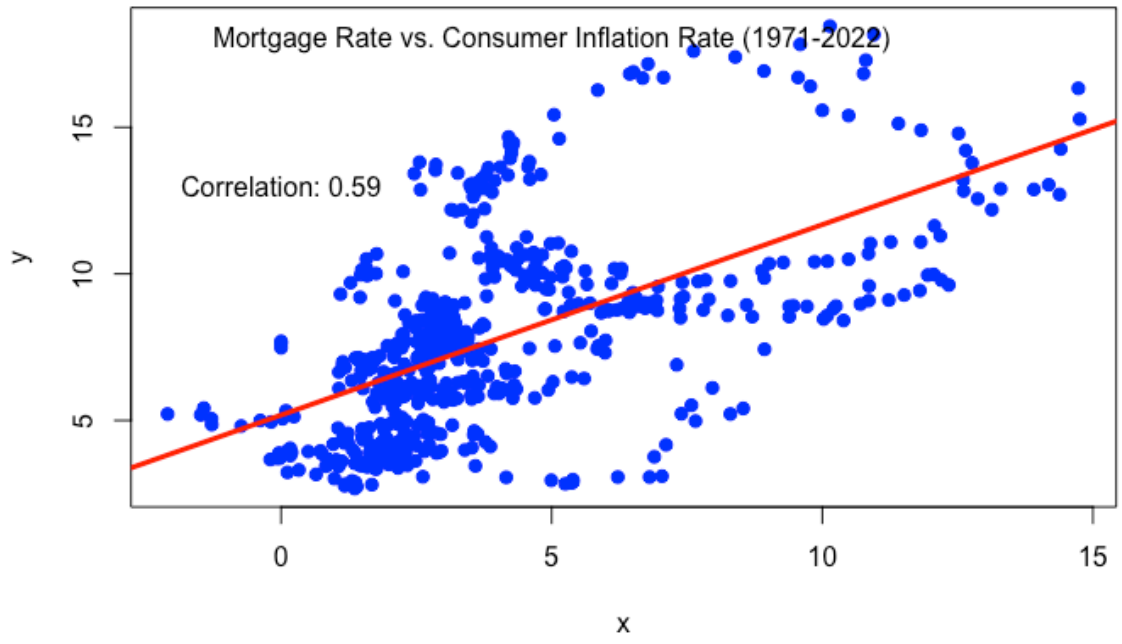

Mortgage Rate vs. Consumer Inflation Rate (1978-2022)
Correlation: 0.61
y
x
0
5
10
15


ouse Inflation Rate(Boulder) vs. House Inflation Rate(Colorado)
Correlation: 0.88
y
x
-5
0
5
10
15
20
30


House Inflation Rate(Colorado) vs. Consumer Inflation Rate
Correlation: 0.37
y
x
-5
0
5
10
15
20
0
5
10
15


House Inflation Rate(Boulder) vs. Consumer Inflation Rate
Correlation: 0.31
y
x
0
10
20
30
0
5
10
15

#Analyze individual correlations for 1971-2022 data (df1, monthly)

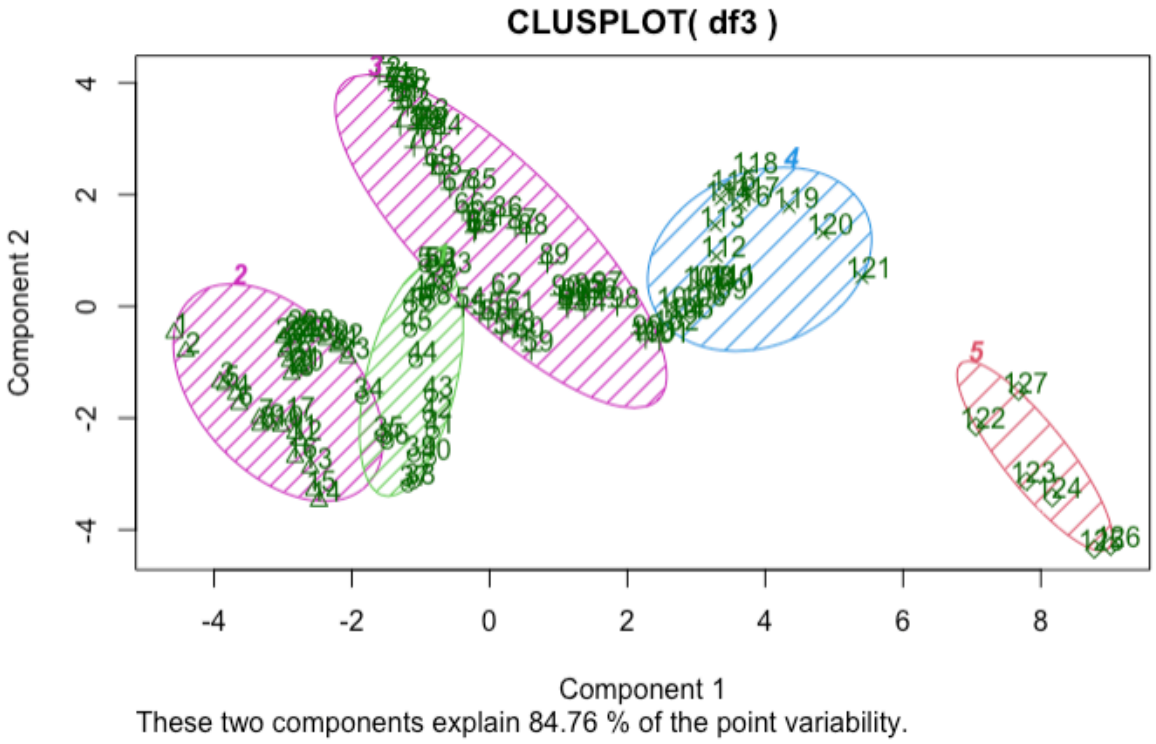


#clustering analysis for the 1991-2022 data with all indexes (df3, quarterly)

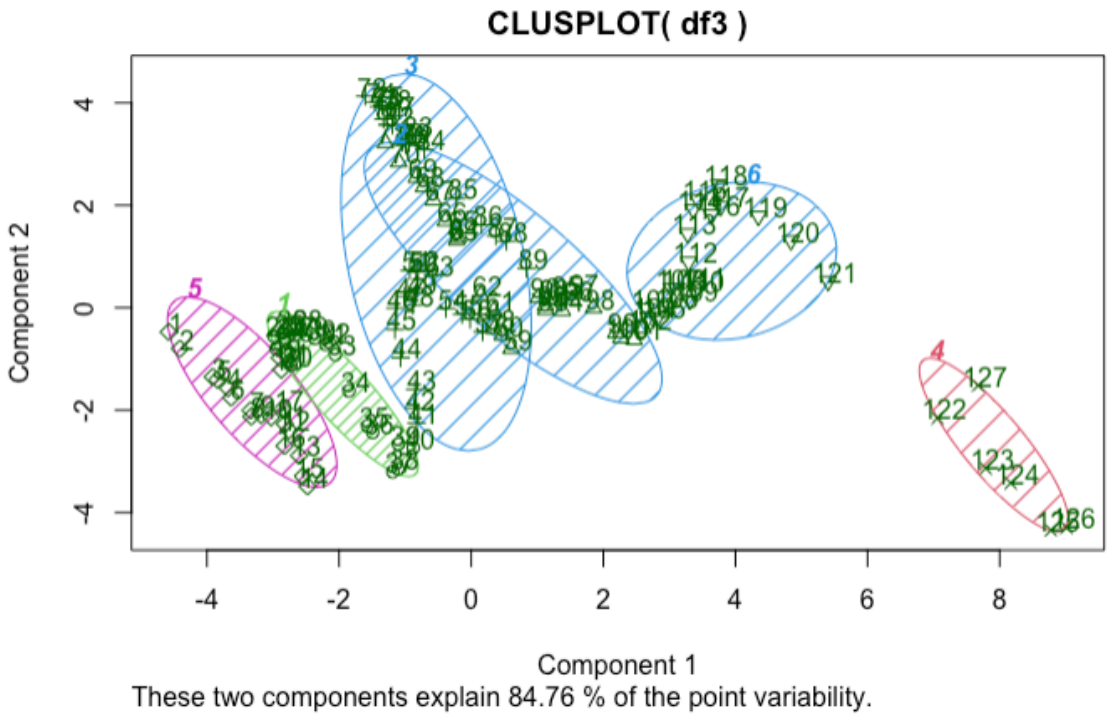
##set up parameters and run k-means

k = 4
(kmeansModel4 <- kmeans(df3, k, algorithm = c("Hartigan-Wong", "Lloyd", "Forgy", "MacQueen")))
(kmeansModel4[1])
(ideal amount of clusters chosen)

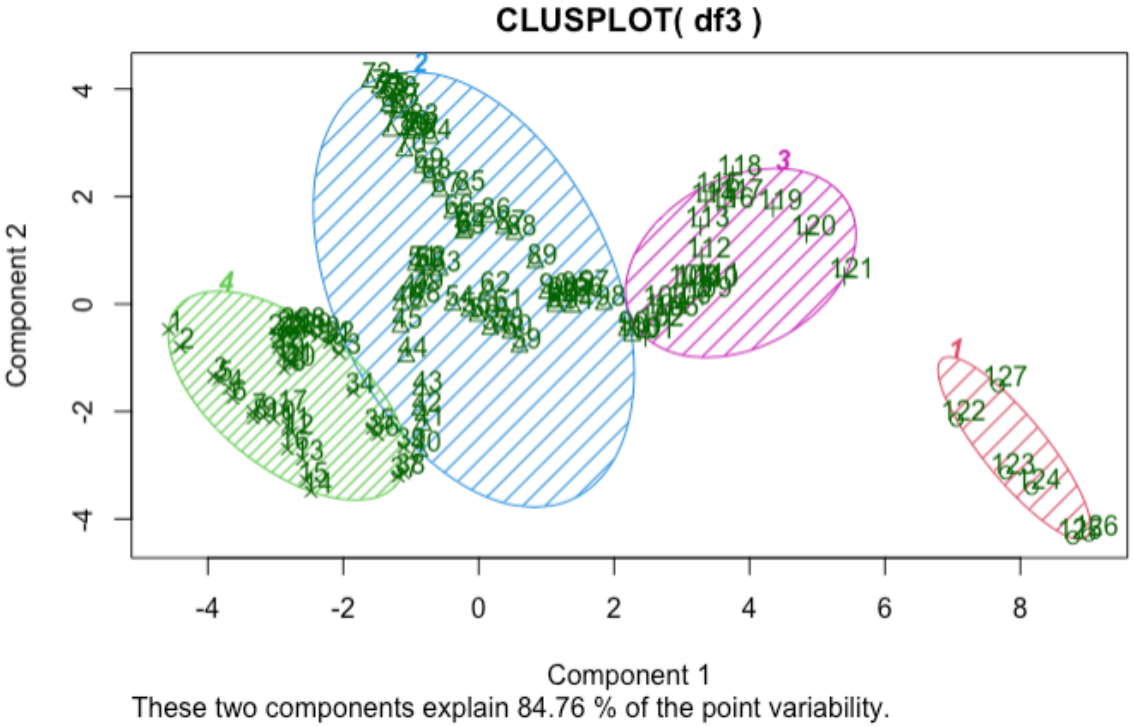
Df3 k = 5


CLUSPLOT(df3)
Component 2
Component 1
These two components explain 84.76 % of the point variability.

Df3 k = 6

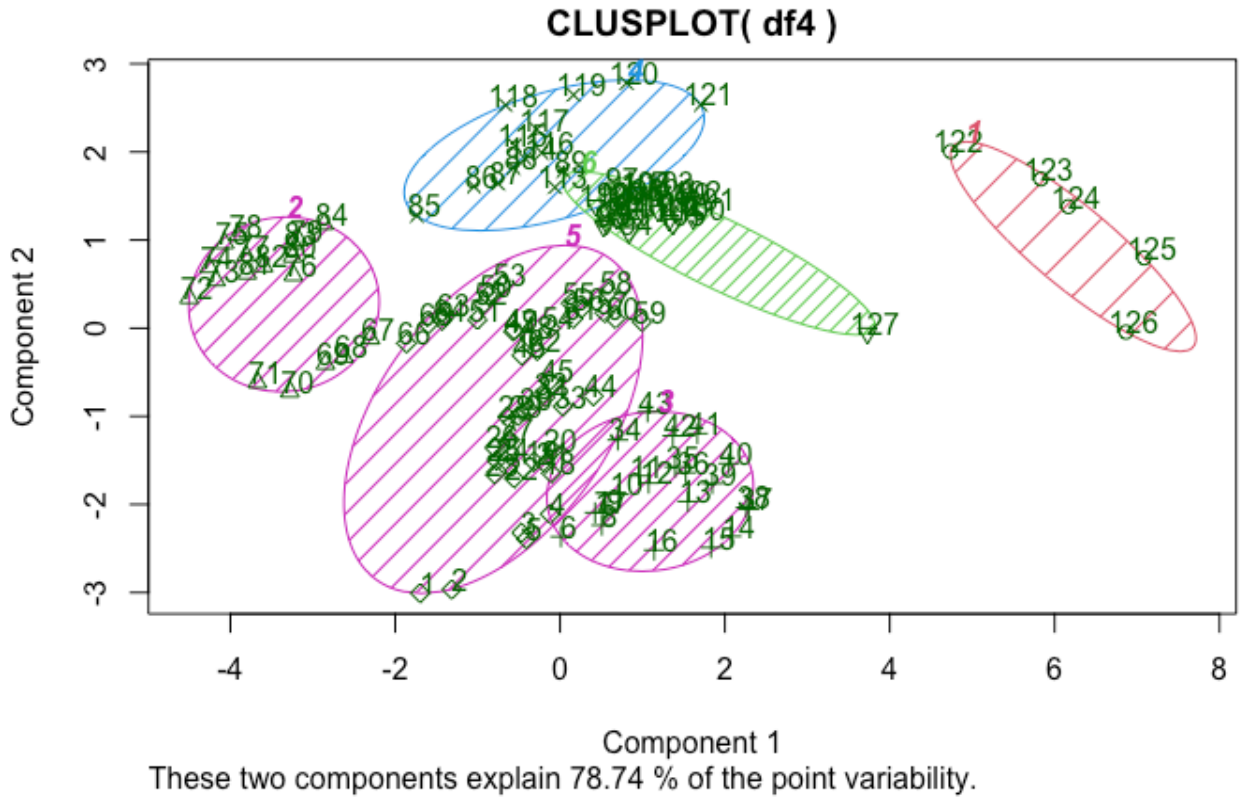

CLUSPLOT(df3)
Component 2
Component 1
These two components explain 84.76 % of the point variability.

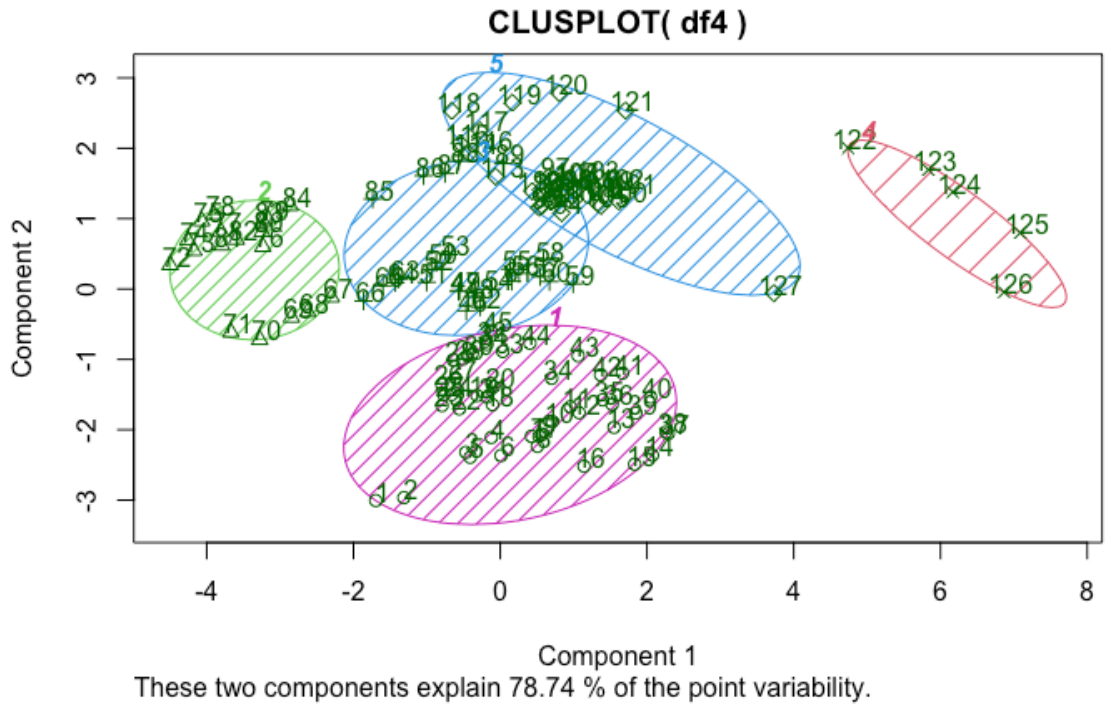
Df3 k = 4

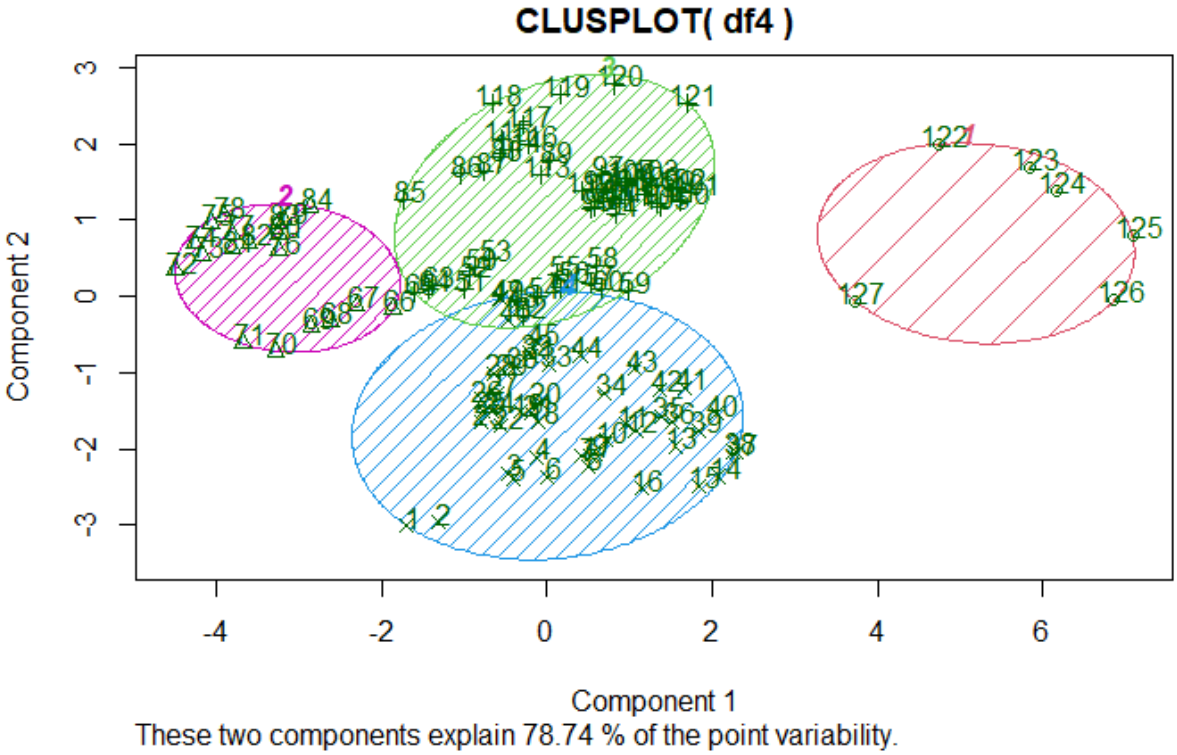

CLUSPLOT(df3)
Component 2
Component 1
These two components explain 84.76 % of the point variability.

#clustering analysis for the 1991-2022 data with only the inflation indexes (df4, quarterly)

K = 4,5,6

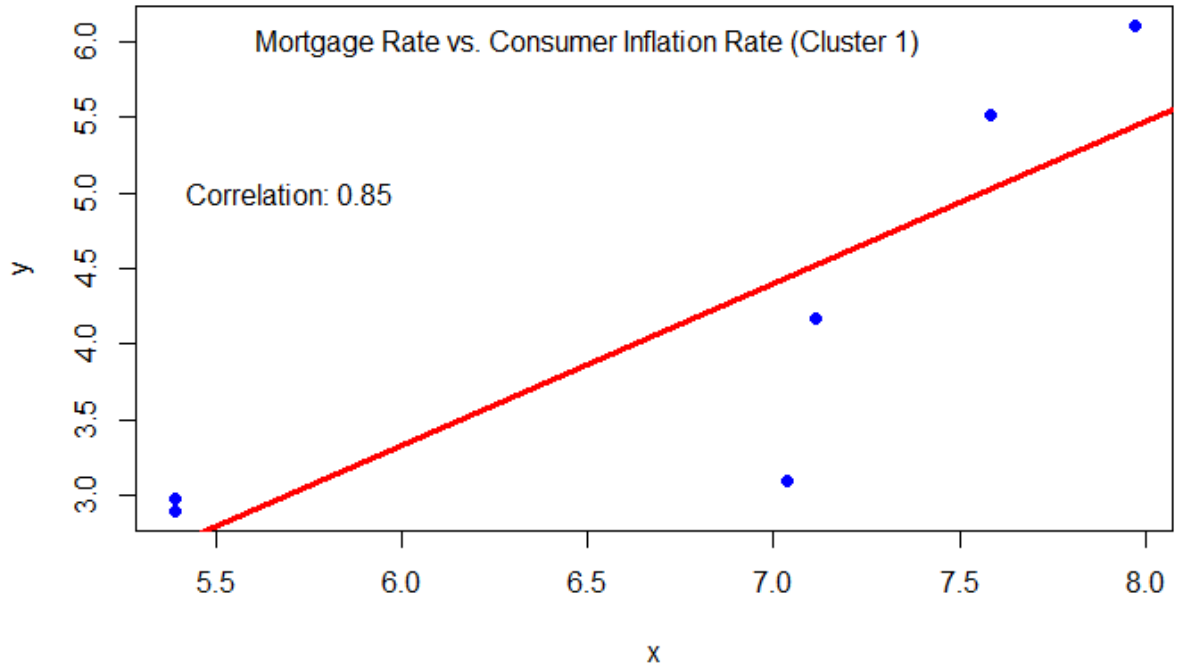




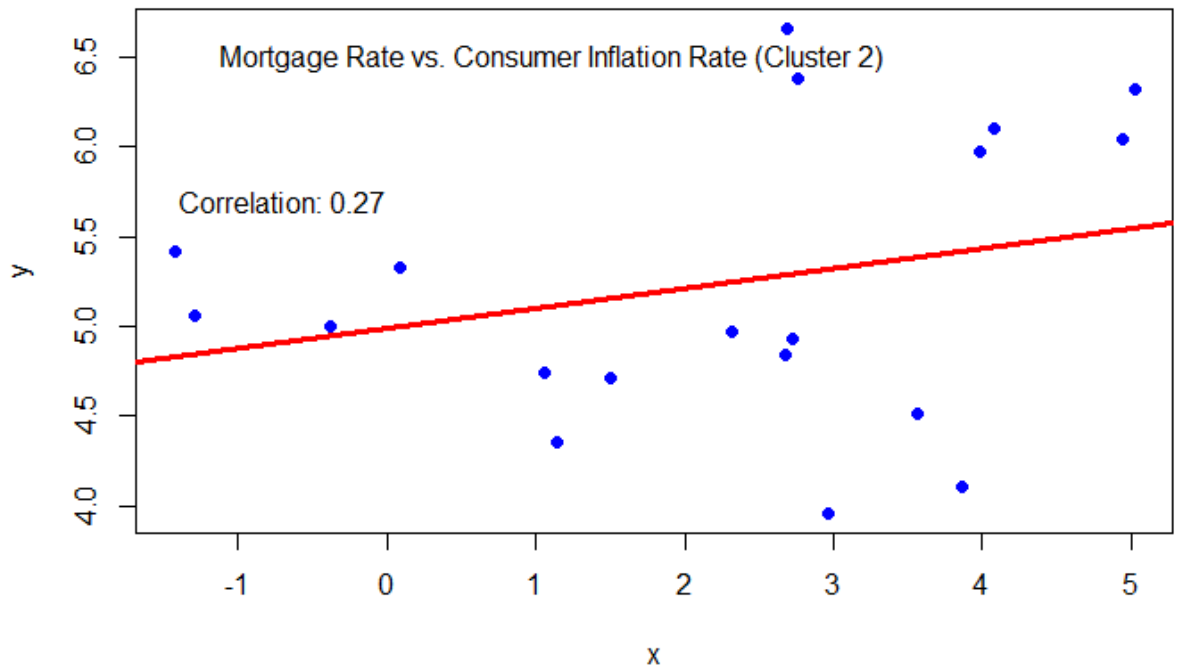


#Analyze individual correlations for each cluster for the 1991-2022 data

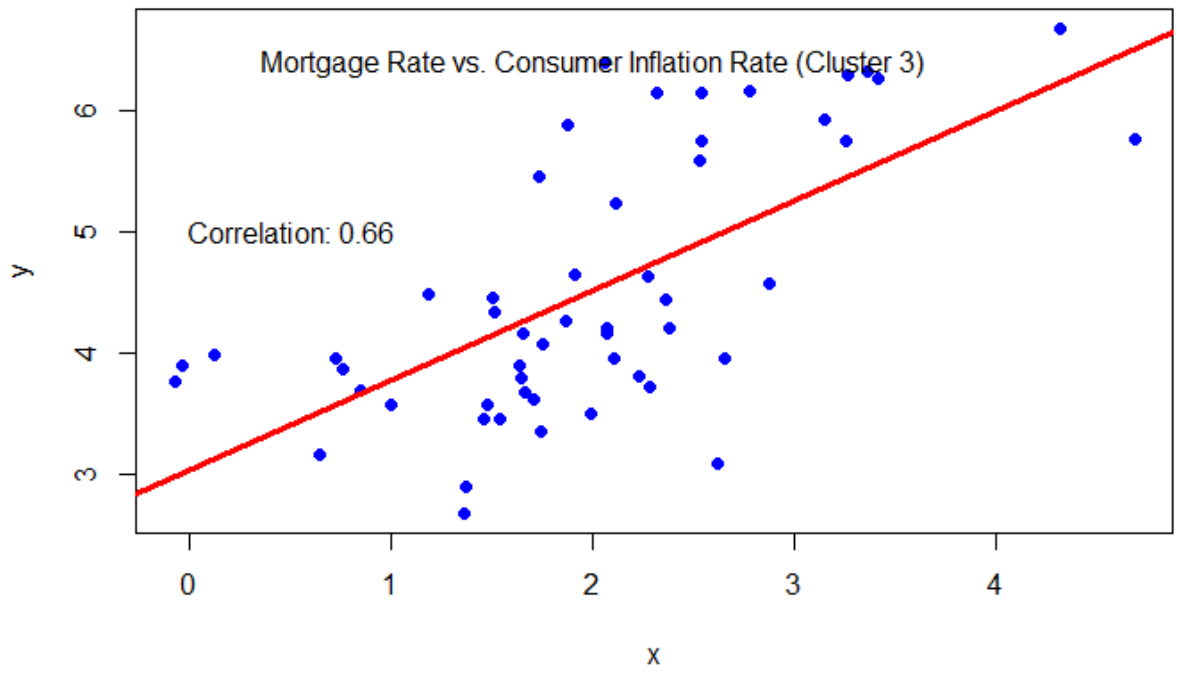
#scatterplot of Mortgage_Rate vs. Consumer_Inflation_Rate for Cluster 1



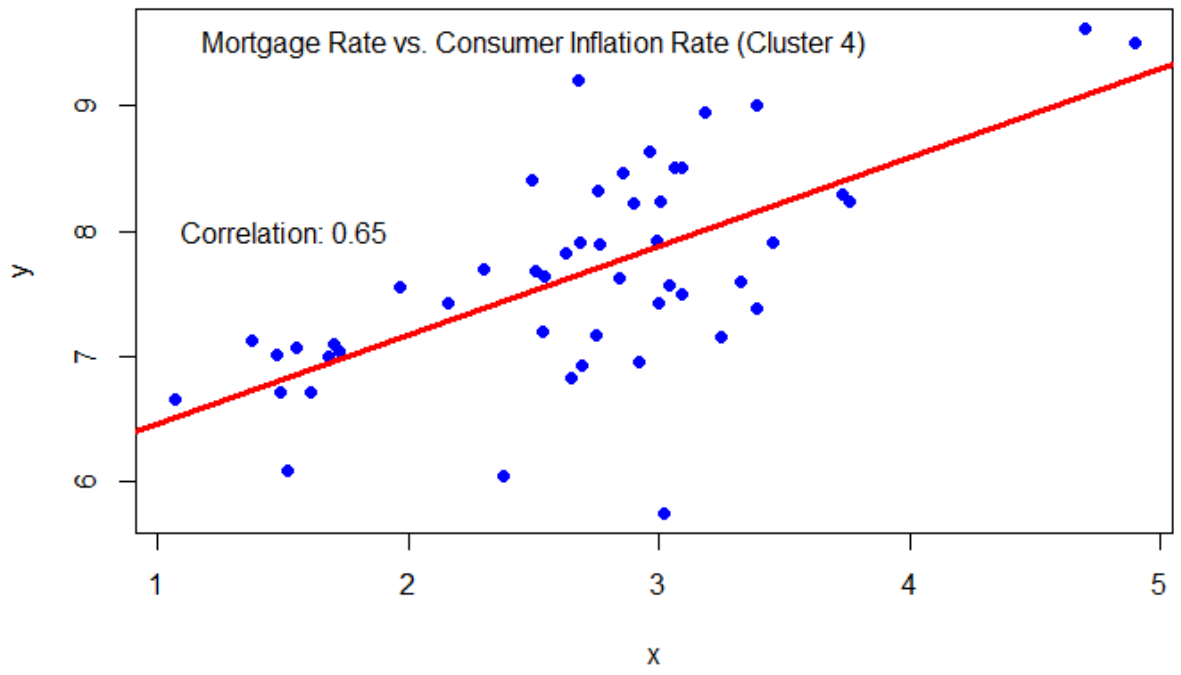
#scatterplot of Mortgage_Rate vs. Consumer_Inflation_Rate for Cluster 2



#scatterplot of Mortgage_Rate vs. Consumer_Inflation_Rate for Cluster3

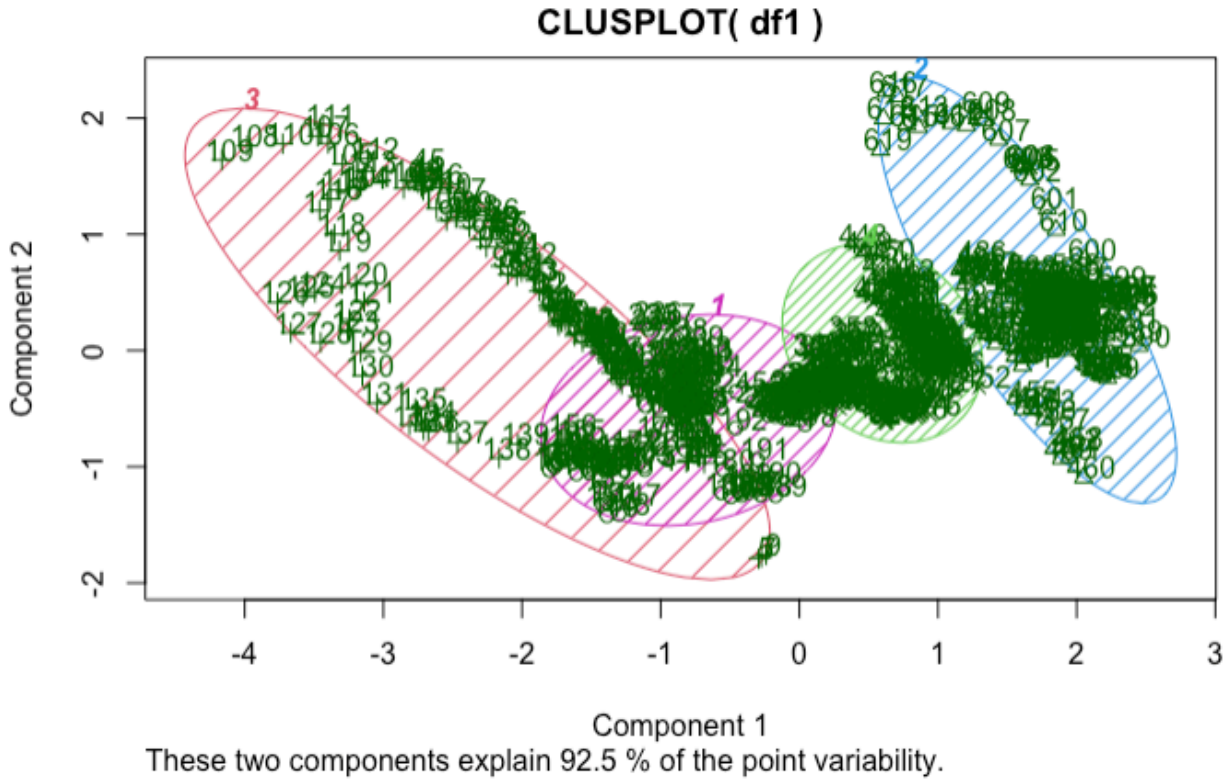


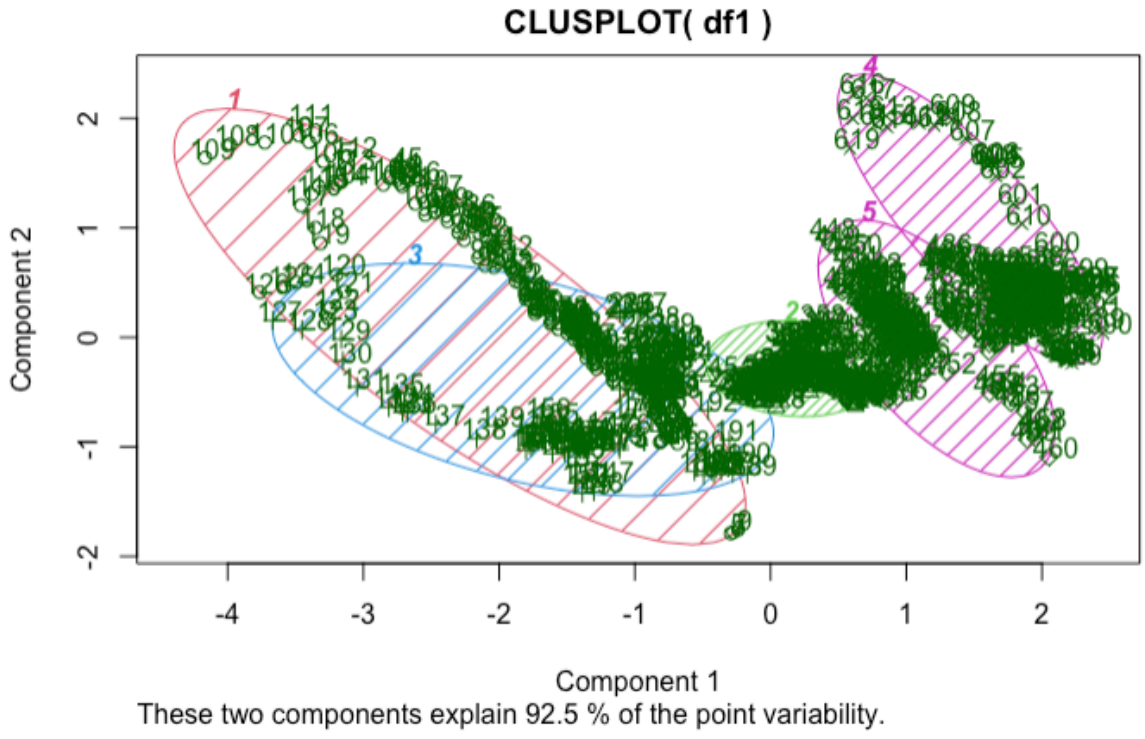
#scatterplot of Mortgage_Rate vs. Consumer_Inflation_Rate for Cluster 4

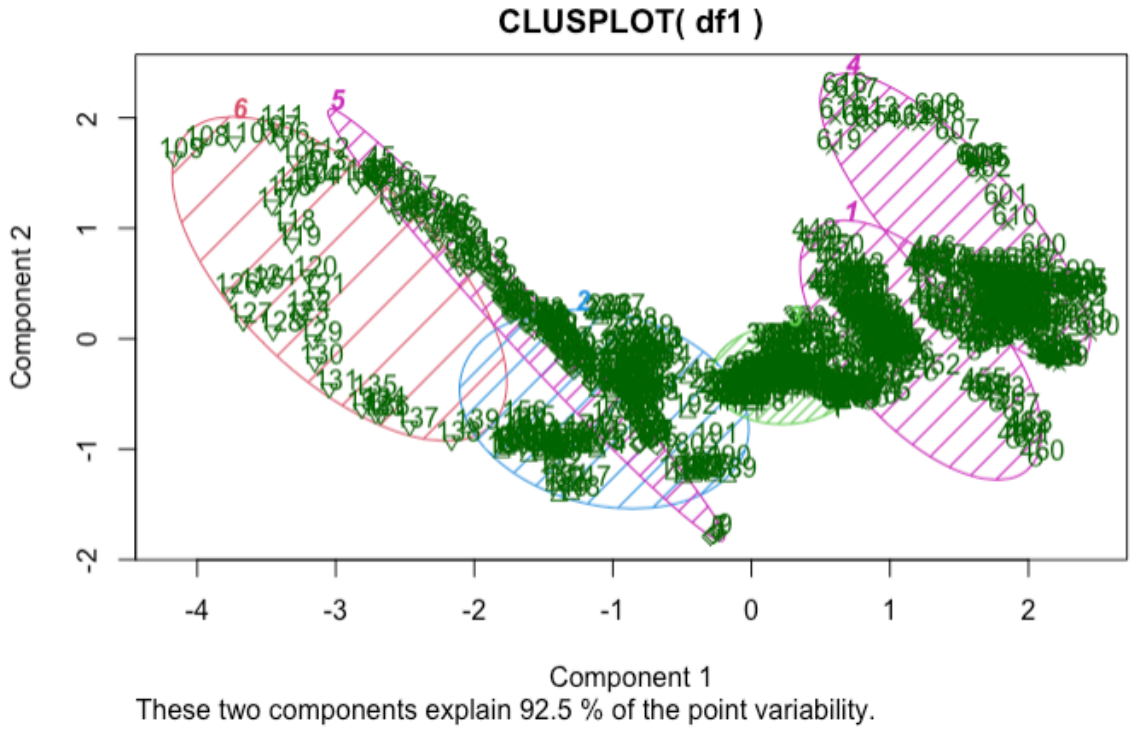


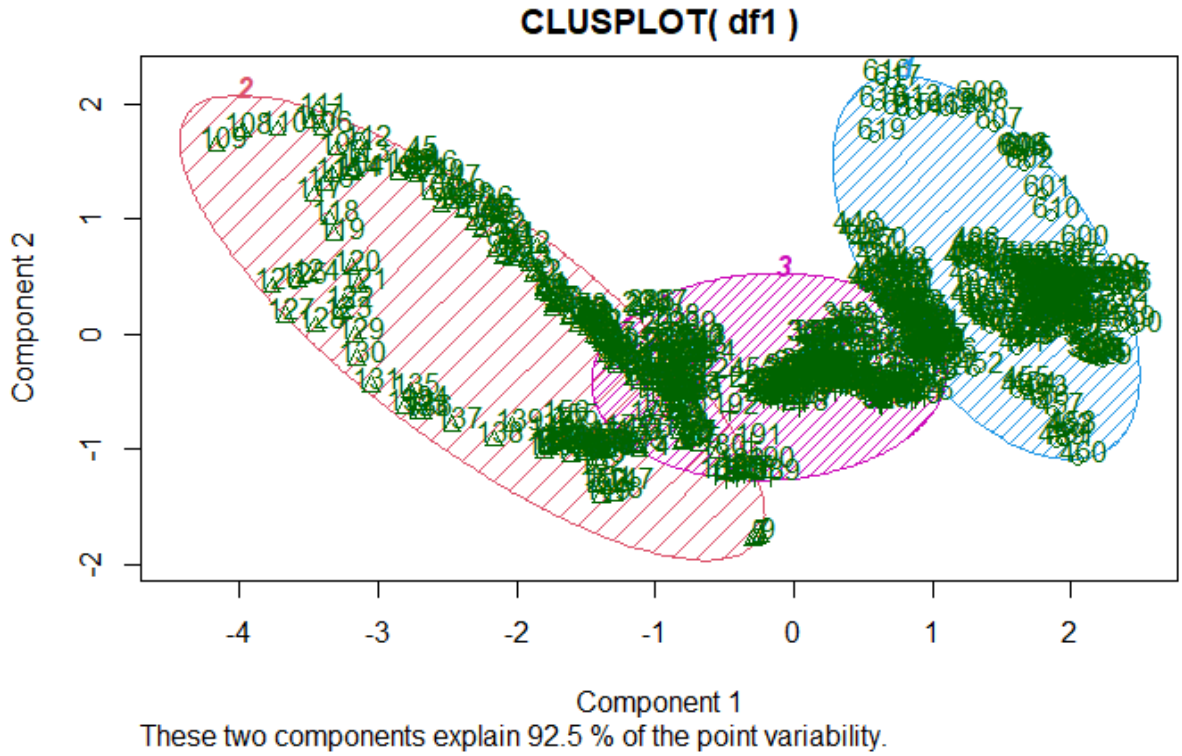
#clustering analysis for the 1971-2022 data with only the consumer inflation index and mortgage rates (df1, monthly)

k=4,5,6,3


CLUSPLOT(df1)
Component 2
Component 1
These two components explain 92.5 % of the point variability.

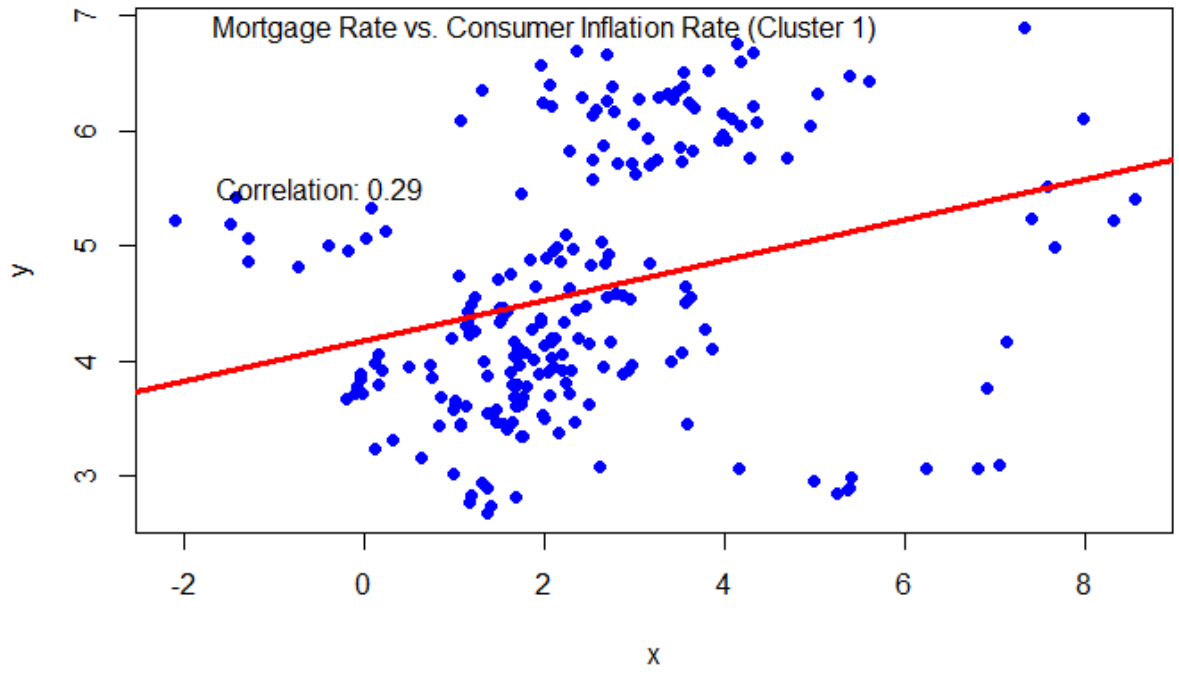

CLUSPLOT(df1)
Component 2
Component 1
These two components explain 92.5 % of the point variability.



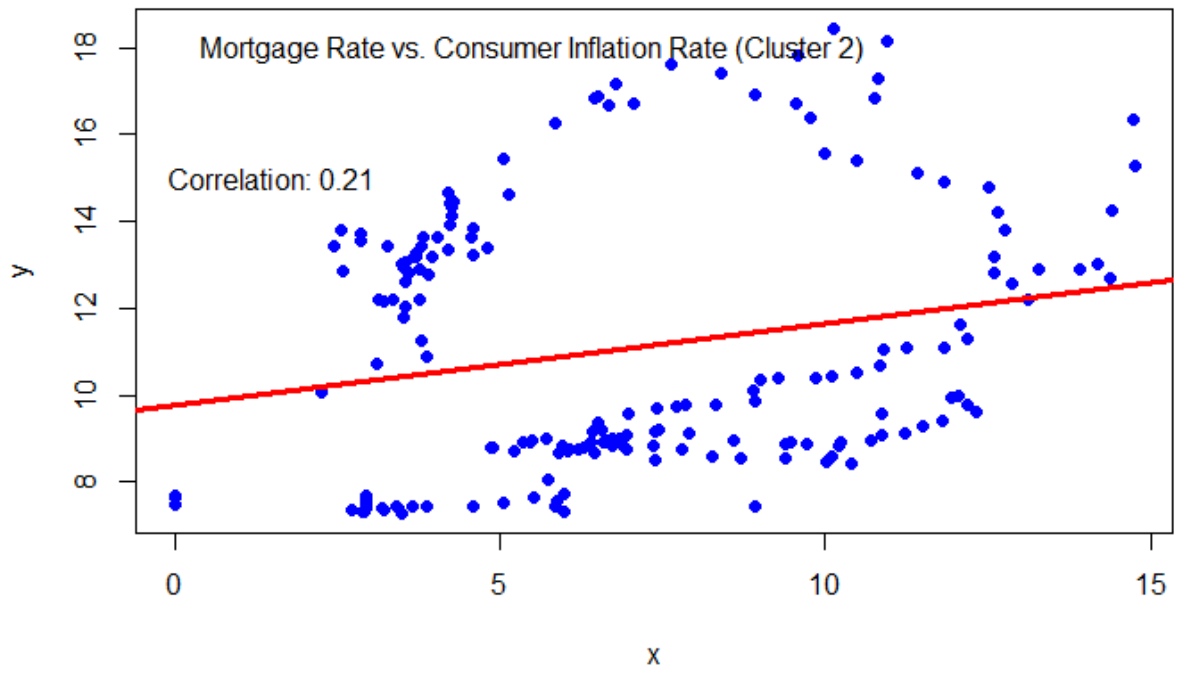


#Analyze individual correlations for each cluster for the 1971-2022 data

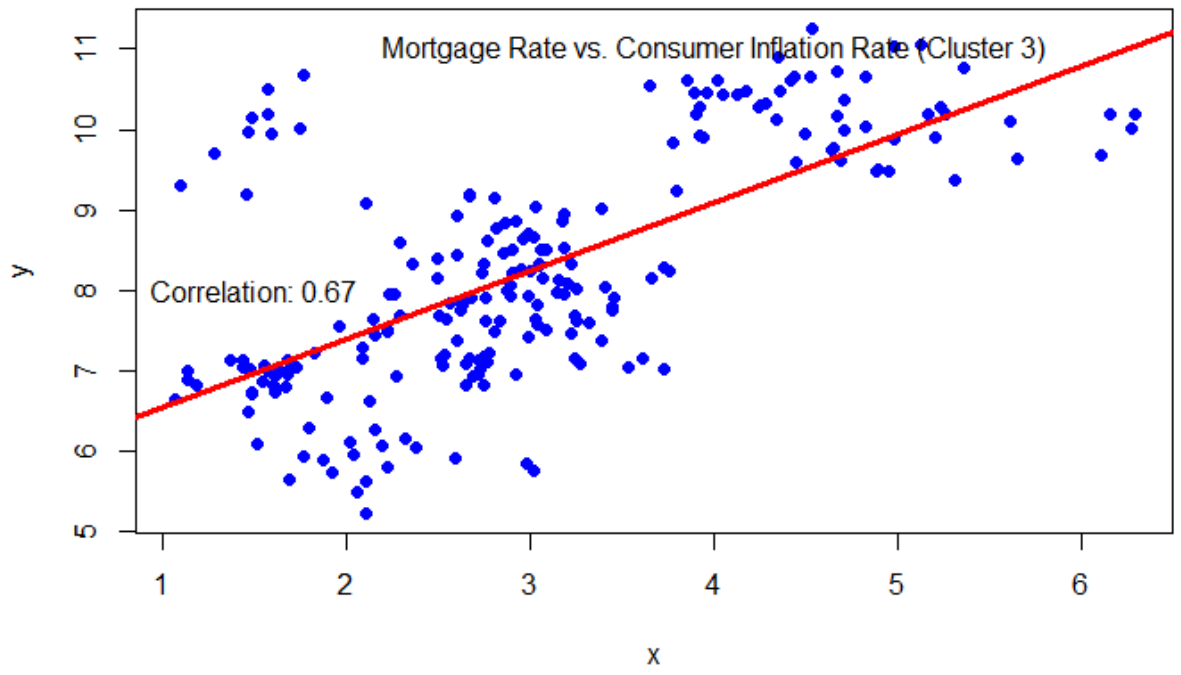
#scatterplot of Mortgage_Rate vs. Consumer_Inflation_Rate for Cluster 1



#scatterplot of Mortgage_Rate vs. Consumer_Inflation_Rate for Cluster 2



#scatterplot of Mortgage_Rate vs. Consumer_Inflation_Rate for Cluster 3



6.2 Code

All code for the analysis is located in Final Project.rmd. All csv data is located in the zip file. File path names for the csv files in the code will need to be updated to path names stored on the local device in order for the data to be read in and for the analysis to run.

7. References

[1] Candiloro, Taelor. “Why Millennials Can't Afford Homes: Housing Prices vs. Inflation.” *Anytime Estimate*, 12 Sept. 2022.

[2] “Median Sales Price of Houses Sold for the United States (MSPUS)” https://fred.stlouisfed.org/series/MSPUS.

[3] Anari, A and Kolari J. “House Prices and Inflation” Real Est Econ” , vol. 30, no. 1, pp. 67-84, 2002.

[4] Spellman L.J. “Inflation and Housing Prices” Real Est Econ” , vol. 9, no. 3, pp. 205-222, 1981.

[5] Diamond, William and Landvoigt, Tim and Sanchez, German, Printing Away the Mortgages: Fiscal Inflation and the Post-Covid Housing Boom (November 20, 2022). Available at SSRN: https://ssrn.com/abstract=4283119 or http://dx.doi.org/10.2139/ssrn.4283119

[6] Dougherty, A., Van Order, R. “Inflation, Housing Costs, and the Consumer Price Index. The American Economic Review”, [s. l.], v. 72, n. 1, p. 154–164, 1982.