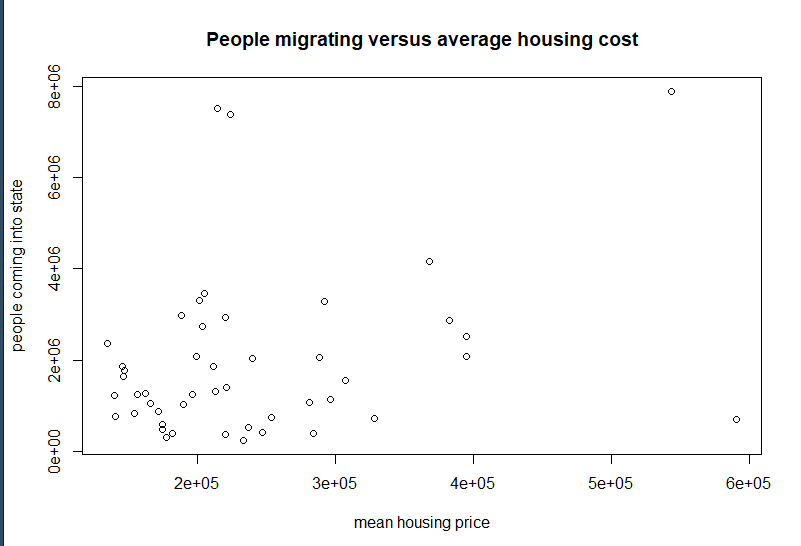
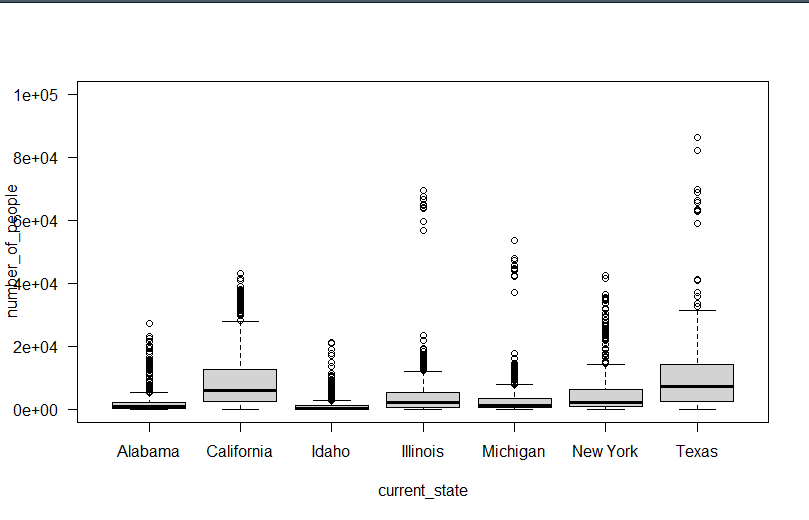
Goal: Using Rstudio to make visualization and perform research on the shared datasets. After combing through the datasets, some questions were raised about what’s happening with the housing market in the United States.

1. Is there a correlation between the mean housing price in the united states and the influx of people moving into those states?

people\_migrating\_versus\_mean\_housing\_price\_over\_all\_states\_all\_years.PNG

We can see that through the graph that people migrating into states is not solely from the housing price, or else the trend would be a reverse relationship. Upon viewing this graph, it is clear that mean housing price is not the only reason for people migration, if at all.

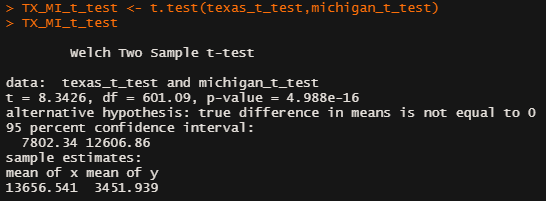
1. Show that California and Texas have significantly higher rates of movers than other states by comparing random states to each other. Use t-test or box-and-whisker graph to quantitatively prove this claim.



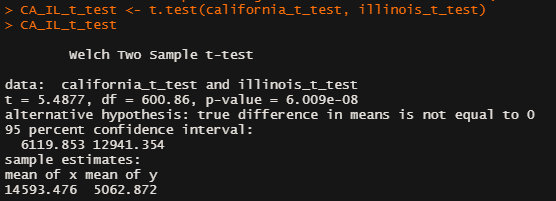
box\_whisker\_selected\_current\_state\_versus\_people\_moving\_ZOOMED.PNG

This box-and-scatter plot shows a random selection of states including Texas & California. The box and whisker for both California and Texas are significantly larger than the other states’, indicating that our states of interest have a statistically significant difference from others.

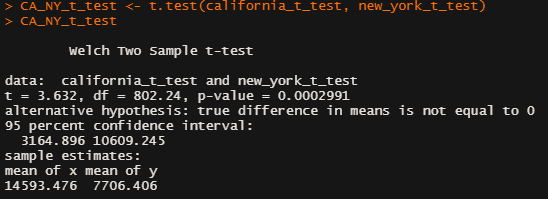
Note: California had significant outliers in the graph that were omitted from the diagram to allow proper display of the graph.



TX\_MI\_t\_test\_number\_of\_people.PNG



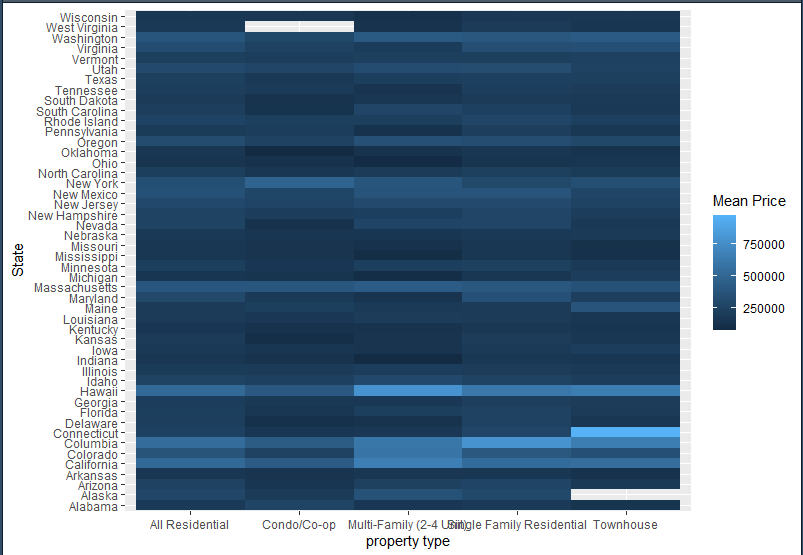
CA\_IL\_t\_test\_number\_of\_people\_migrating.PNG



CA\_NY\_t\_test\_number\_of\_people\_migrating.PNG

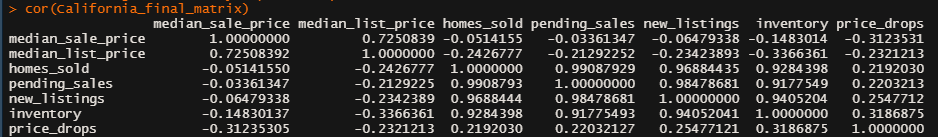
To obtain quantitative values to compare the number of people migrating into a state, a T-test is used to show if there is any statistical significance. If the t-test is greater than 2.0, then there is a noticeable difference between the two states being compared. This can be observed in all three correlation matrices, which supports that California & Texas have a much higher population moving in. Furthermore, the p-values for all comparisons are less than the general 5% rule, indicating that there are one or more variables at play that are causing this massive difference in people moving into Texas & California as opposed to other states.

1. Do property types affect the mean housing prices in the United States?

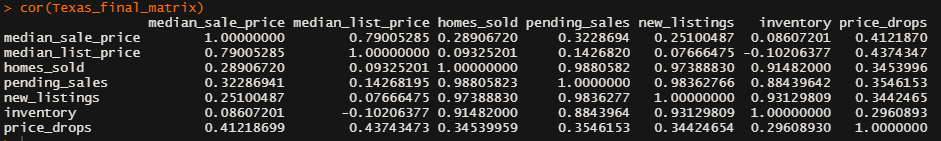


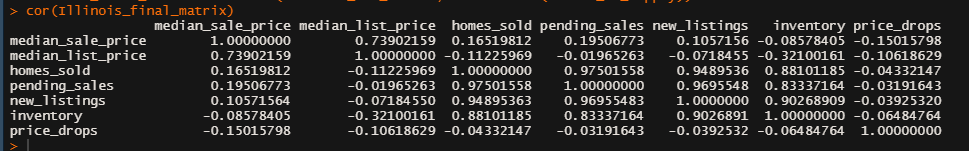
HEAT\_map\_state\_property\_type\_median\_price.PNG

The heat map presented shows the states versus property type, with the color-filled boxes representing the mean price throughout the years. One observation is that the “Single Family Residential” property type tends to have a higher mean price than its counterparts. However, not all states reflect this trend so we can say that the mean price is not solely affected by the type of property, though it could give some influence.



California\_cor\_matrix.PNG

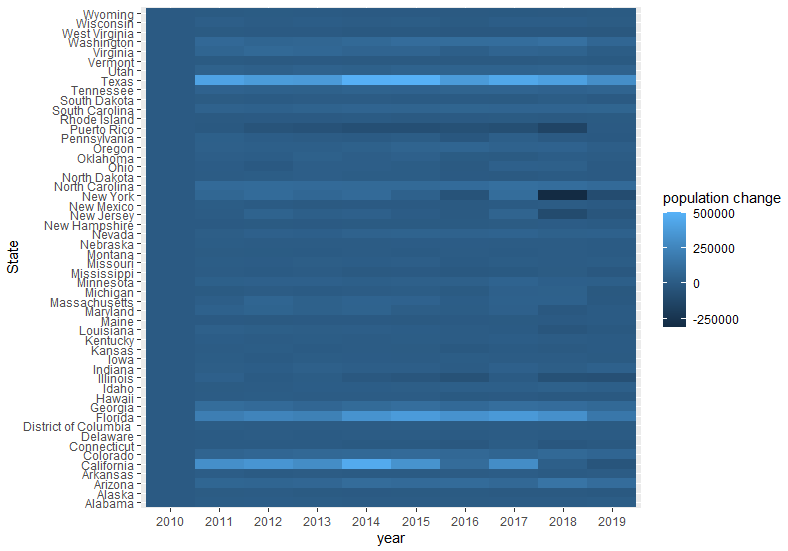
Texas\_cor\_matrix.PNG

Illinois\_cor\_matrix.PNG

We can see that besides “median\_list\_price”, there are no correlations between “median\_sale\_price” among the other variables. It’s important to note that the median sale price in this dataset is not affected by the “homes\_sold”. This data suggest that if the median sale price goes up as people are moving into states & buying the available homes, the correlation would be an inverse (< -0.7) relationship of homes sold. Since this is not the case, this further indicates that there are other variables at play in the rise of mean housing prices.

5. Is there a clear trend throughout time in the United States that shows which states had the most migration?

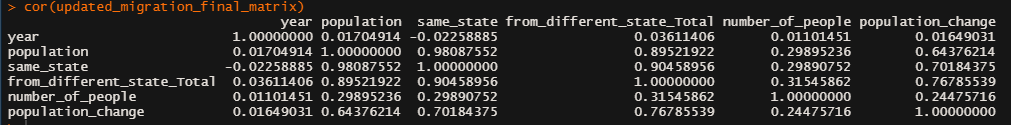
As expected, California and Texas are showing a lighter color, indicating a higher than normal rate of migration influx into those two states throughout the years. Though it appears Florida is also a state that received large amount of people moving in as well, majority of states have a lower rate of population migrating in throughout the years with small variance. Keep it mind that it is possible for California, Florida, and Texas can skew the color representation due to scaling of the population change.



Heat\_Map\_year\_state\_pop\_change.png

##### 6. Are there any variables in the other dataset that could provide insight in the United States?

The population is expected to have a positive correlation with the populations in the same state and different states as it is expected for the population to grow over time. If there are more people present, then we can expect these people to exist in one state or another. Lastly, it appears that none of the variables are dependent on time.



updated\_migration\_correlation.png