Austin Housing & Income Analysis

Springboard Capstone Report
By Brian Mc

Problem

Are incomes keeping up with rising home and rent prices prices in the Austin Metro Area of Texas? The Austin Board of Realtors says the median home price was up 14 percent in April 2015, compared to the same month last year (2014), to $274,000*. Furthermore, in October 2015, the Austin Metro Area population surpassed 2 million people. Only five years earlier in 2010, the population was ~1.7 million, this is an increase of 23.53%. With such an increase in demand, is the supply remaining affordable? Are rent and housing values rising? Are incomes rising as well? Using statistical analysis methods available in RStudio recent income, home values, and rent prices of the Austin Metro Area (by zip code and city) will be analyzed to identify changing trends.

Finding the Data

To analyze this questions I pulled data from the Zillow data center (http://www.zillow.com/research/data/) and the Amercian Community Survey. These data sets contain a plethora of data that can be analyzed to address this issue. For this project, I focused on the best available data (most current) and most detailed that could be reasonably analyzed. Thus I downloaded data from 2011-2015 and focused on data sets that could be broken down by metro area, city, and zip code. I downloaded the zillow time series median home values and the all homes rent values for 2011-2015, and the American Community Survey (ACS) 5-year household income for 2011-2014. When a comparative analysis was made, the time sets for the home and rent values were reduced to just 2011-2014 to match the ACS household income values. The 2015 ACS household income values were not available at the time of this report. These data sets were downloaded as csv files for the entire country. They were filtered to include on zip codes in the Austin Metro Area for the state of Texas. These results were then imported into RStudio for analysis.

Analysis

Using Rstudio, begin the analysis by importing the 3 csv files (ACS household income, Zillow rental and home value datasets), and load dplyr, ggplot2, data.table, tidyr, and scales using the library command.

Data Wrangling

After importing both zillow data sets, the column headers were renamed to provided more meaningful titles. A description of the type of data was added to the beginning of each year name. For example, the rent 2011 values header was changed from "X2011" to "Rent2011", and the home values were changed from "X2011" to "Own2011". For the ACS data, the titles were changed from Geoid1, Geoid2 to, Geoid1 and Zipcode (Geoid1 is a unique numeric number used by the ACS and Census, so it remained unchanged, and was not used in this analysis). The ACS data was only described by zip codes, it did not come with any other means of organization, so to help sort the data, the city of each zip code was added to the dataset. This was done by joining (full_join, by zip code) one of the zillow data sets (which contained zip codes, cities) to the ACS data. The unneeded columns from the zillow data were removed from the ACS dataset.

There were too many zip codes to graphically display and analyze, the datasets were summarized (command: summarise) by cities to find the average (mean) values (rent, home values, and household income) for the included zip codes.

All three data sets were joined together to create a master data set. The Price to rent ratio were found for each zip code and re analyzed for each city as well (after the join). Zillow defines Price to rent as a ratio as the estimated home value is divided by 12 times its estimated monthly rent price. This is useful to compare home and rent values. After wrangling all of the data, ggplot2 can now be used to help graphically view the data.

Ggplot (all graphics are in slidedeck)

To graph the data, each of the value columns (Rent2011, Own2012, etc) had the be organized into one single column to be graphed against a time series. Using the *gather* commanded each set of Rent/Own/Income values were combined into one column. Once gathered, the minimum and maximum values to be graphed were found. This allowed the graphs y min and max to maximize the visible range of the data. The resulting graphs compared the rent, income, and home values versus time (2011-2014/5) to show a steady increase in all values. After viewing the graphs it can be seen that while all datasets did increase, the home and rent values increased substantially more than the income.

Statistical Analysis

Using linear model in rstudio, we can compare the rent values and home values to the household income (*lm* function). Use the summary and plot commands in the r script to view the results. After viewing the results we can see that there is a correlation between the two datasets.

Further Analysis & Recommendations

While it was found that yes, home values and rent are tied to household income increases, they are not proportionately keeping up. Since this study focused at the city level, and more detailed zip code study (and graphics, mapping of each zip code) could be made. Other values could contribute as well. Are more homes being sold? Are the incoming owners having higher incomes than the outgoing owners? While this study helped begin to answer housing and income relations, it is not the end. As more data becomes available, a more recent analysis can be made.

Sources:

* http://kut.org/post/austin-home-prices-hit-another-record-april-2015

**http://www.bizjournals.com/austin/news/2016/05/19/atxsubcounty-population-growth

.