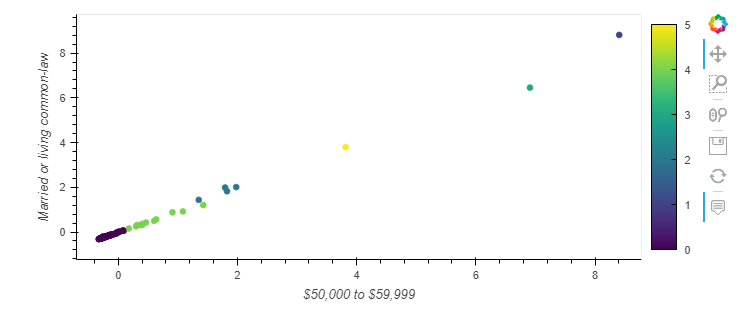
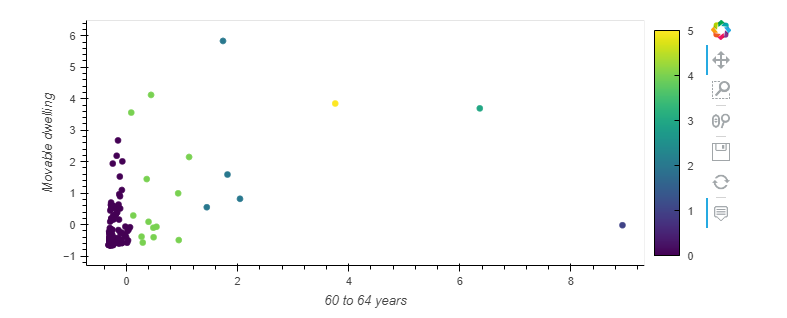
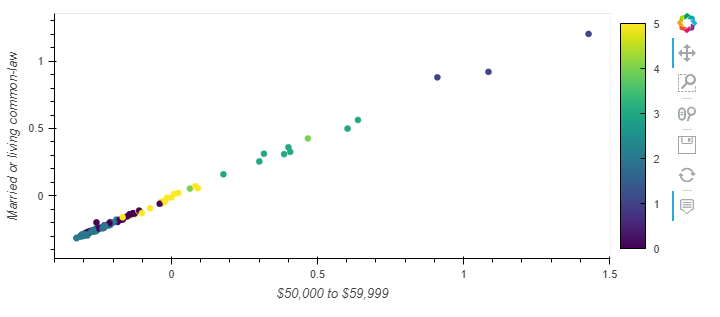
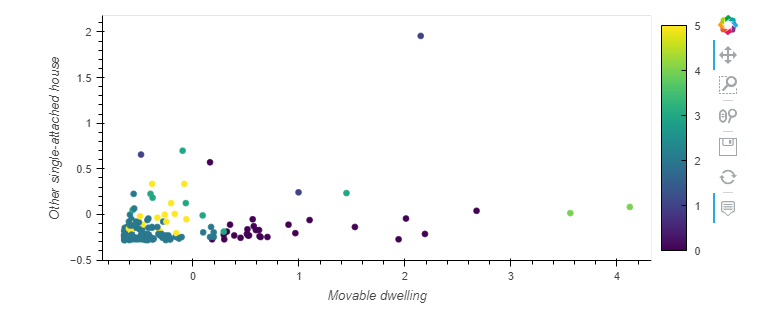
***Cluster Analysis***

**Initial Cluster Analysis:** Initial analysis proved difficult with the scaled data, because a large amount of the different features when compared together, resulted in incredibly similar graphs that were linear, and unhelpful. There were incredibly strong outliers at the CMA level, such as Toronto, Montreal and Vancouver, which seemed to heavily skew our data. Using some initial PCA, we were able to determine the two most potent features, Ages 60 to 64 years and Movable Dwellings. Comparing a graph of them both to one using other regular features can show very blatant changes and just how much they both affect the data even with the outliers.  
  


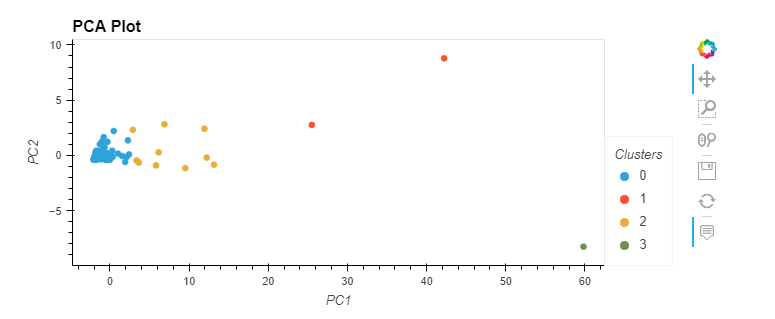


As such it was decided that we needed to remove the outliers, in order to get stronger, more focused cluster data.

**Focused Cluster Analysis:** With the removal of major outliers, we can still see a great deal of similarities between graphs using different features, except once again determining the important features through PCA. This time however, the strength has shifted where Movable Dwellings are still incredibly strong, but Other single-attached house is the next strongest. Meaning there has been a noticeable change in what defines the CMA.  
  




As you can see there are still some much smaller outliers in comparison to the first set of clusters, but you can still see various clusters forming in the larger, much more common base. But, even without the strongest features, it is still much easier to see common cities or towns that exist in a very similar way when compared with most features. This makes it easier for marketing to determine where to focus specific products. Including areas like: Thunder Bay and Brantford, Regina and Sherbrooke (this can be seen on the full HTML charts, as these are merely images).  
  
**FULL PCA:** Initial PCA determined that while there was very little variance because of the outliers, it was able to determine that initially Ages 60 to 64 held an accuracy of 94%, meaning the feature controlled a VAST majority of change and variance in the data alone, with Movable Dwelling being second, at a much MUCH smaller rate. This was determined to be a factor mainly caused by the outliers, because once they were removed we did another PCA test which determined that Movable Dwelling was the strongest feature, alone holding an accuracy of 79%, and it appeared again as the second strongest, for 16%. As such, we can determine that for the outliers market; Toronto, Vancouver and Montreal, that retirees (those aged 60-64 preparing to retire) have a VERY strong control of their market. While for the rest, Mobile Dwelling ownership can account for a strong control of their markets, as far as customers to look for.  
  
It also allows us to see the similarities in towns to look for, just like the individual cluster analysis equating one town to another, for the sake of looking for future profits.

Original Cluster:  
  
  


New Custer:

