**Should I Stay or Should I go?**

Ben Tankus

May, 13th 2020

**1. Introduction**

**1.1 Background**

We are considering a move to Loma Linda California for my girlfriend’s graduate school. We really like Portland Oregon, which is where we currently live, and want to see how different Loma Linda will be. I am also worried about finding a job and housing there.

**1.2 Problem**

We really like Portland, Oregon and we want to make sure Loma Linda California is fairly similar before we move.

**1.3 Interest**

Moving to a different city is a big risk but identifying the best possible location for us to move before we arrive using data will make the move a bit less stressful.

**2. Data acquisition and cleaning**

**2.1 Data sources**

I used the Quandl API system to get housing prices and for-sale house count. This provided great historical information on house price and went all the way back to the 1970s! Unfortunately, the data was only published until 2017, so the most recent trends are behind a paywall. The for-sale count however had data as recent as last month!

For job data most sources were behind paywalls, so I had to scrape Google job listings. This created some problems with cleaning that I will discuss in the next section.

**2.2 Data cleaning**

The housing data was mostly clean from the start. Each column was published along with the last day of the associated month which made it handy to join between columns. The price data went from 1975 (when houses were only $20,000!!!) to 2017, and the housing counts went from 2009 to 2020. The only real cleaning was done when plotting to remove the nulls where there was price but no count data.

Job data was a bit more difficult to wrangle. Because I got the data from scrapping it was inherently dirtier. In the results there were links that did not go directly to jobs. I tried to filter the dataframe for results that only included the job title, but that didn’t work. I also tried compiling a list of job sites that publish these jobs, but that would filter out some results. I ended up not filtering anything out under the assumption that both PDX and LL data would be equally affected.

**3. Conclusions**

### 3.1 Housing Cost

It looks like Loma Linda had a huge cost spike right before the 2008 recession and is only now recovering. Portland on the other hand, really didn't have as much of a bubble burst and average housing values are now almost a full $100k higher.

### 3.2 Housing Inventory

Loma Linda and Portland actually started fairly close in 2009 and held a fairly consistent gap until 2013 where Portland began listing more houses more frequently. Portland took another huge jump around 2015 where people really began building houses faster.

### 3.3 Overall Housing Conclusion

The prices track fairly close between Loma Linda and Portland. As Loma Linda is a smaller city it is likely more sensitive to large economic impacts, as seen by the recession. Once housing data is published for the Coronavirus epidemic I would expect to see Loma Linda housing prices fall, and houses available climb, both more substantially than Portland.

# 3.4 Job Plot Interpretation:

These data are very limited, so no substantial conclusions can be gained from them. Empirically, it looks as though Portland has about 2x as many Data Analyst jobs as Loma Linda, which is less than I expected, as the population is close to 8x the size. Maybe Loma Linda has a better job market? This will likely need to be further analyzed in the future.

**4. Final Word:**

While moving to a new town will always be risky, this analysis makes me a bit more relaxed. The housing value seems similar to Portland, and while there are less jobs, they do seem to be available. I think I’ll go and see if my predictions are accurate!