**Data Limitations**

**Project: Where are the people, where are the jobs?**

**1.7.19**

Data Sources

The central data source for this analysis is the University of Southern California’s Neighborhood Data for Social Change website. Their public data sets under the “Transportation,” “Health,” and “Environment” categories were the foundation for this analysis. These data sets were cleaned versions of American Community Survey (ACS) data, collected by the United States Census Bureau.

Quality Issues

Thanks to the data cleaning and preparation done by USC’s Neighborhood Data for Social Change, the quality of the data sets we used was very good. The main quality issue was the occasional lack of reported information — particular variables in a given census tract for which there was no ACS data. However, these null values represented very small portions of the datasets, and did not inhibit our ability to answer our research questions.

Missing Years

Our biggest data challenge was that in their ACS data, the Census Bureau does not collect the same information each and every year. As a result, some of our data had different spans, and some of our data has gaps in collection years. Below are the years the datasets cover for all of the USC datasets we used.

* Commute Times: 2010 - 2016, yearly
* Public Transit Use: 2010 - 2016, yearly
* Vehicle Ownership: 2010 - 2016, yearly
* Air Quality: 2014, 2017
* Asthma: 2014, 2017

Note that for many of the data sets, most recent years (2017 & 2018) were not included. This presented an analysis challenge because we were lacking the ability to assess very recent changes and trends regarding commuting.

Lack of Holistic Information regarding the Job Industry

The job market can vary greatly due to factors not included in the dataset such as economic data and the growth (or lack of) in specific industries over time. As such, it is possible that our data can over-represent an industry that is doing well and under-represent a struggling industry in a period of time, leading to a skewed view of the job market in LA.

Lack of Holistic Data regarding the Cause of Short and Long Commute Times

Our data uncovered trends in the most ‘extreme parts’ of LA. For example, we were able to deduce that areas further from the city center typically had longer commute times. However, we cannot say the same for the more central parts of LA as we were unable to cover why they had longer or shorter commute times. This could be attributed to a lack of data. Other pieces of data such as traffic or the concentration of workplaces could uncover why certain parts of central LA had longer commute times than others.