# Final Project Information

**Project 1**

Project Objective**:**The [American Community Survey  (Links to an external site.)](https://www.census.gov/programs-surveys/acs/)is a rolling five year survey that is used to acquire information on education, employment, income, and housing and transportation costs among others. These data are not full census data and are used to direct federal funding across several important programs. There is some question as to whether these data are sufficiently representative. The purpose of this project is to compare the Los Angeles rental data (required by the City when a landlord rents a unit in a multi-family unit) to that reported in the ACS. In addition, if the data sources are markedly different, are there any factors in which data vary along?

Data and Information Resources:

* [Background JPER paperreview the documentiew in a new window](https://canvas.ucdavis.edu/courses/80796/files/831216/download?wrap=1)
* [Data Dictionaryreview the documentiew in a new window](https://canvas.ucdavis.edu/courses/80796/files/831217/download?wrap=1)
* [LA Rental Unit Datareview the documentiew in a new window](https://canvas.ucdavis.edu/courses/80796/files/831218/download?wrap=1)
* [ROI Datareview the documentiew in a new window](https://canvas.ucdavis.edu/courses/80796/files/831219/download?wrap=1)

The two datasets we will be using are:

Palm-ROI\_downloadable\_data.xls

Palm-LosAngelesRentalCensus.csv

Note that Palm-ROI\_downloadable\_data.xls has multiple tabs so we will have to determine which is best to match the ROI data. Select the tab “ROI\_12.15.14,” save this as a CSV and rename as “Palm-ROI\_12.15.14.csv” Remove the first row of meta column header. Then upload to R.

STRATEGY

# Re-format and upload the datasets.

# DATA CLEANING

# 1) Begin with data cleaning. Remove duplicates.

# 2) Determine useful indicators to address the research questions.

## 3) Define parameters for those indicators. Identify inconsistencies in definitions such as region. Define 'reasonable' ranges for upper and lower boundaries to remove externalities. See Table 2 (within the middle 99.6% of each variable's distribution)

# 3) Create a filtered data set based on those 'reasonable' ranges.

# MERGE THE DATA

# Merge with the Census data

# DATA ANALYSIS

# 1) Analyze the data by region/district to assess housing characteristics and detect patterns.

# 2) Analyze the data in relation to the census data: e.g. compare to national averages.

# 3) Map the geolocated data to analyze spatial patterns within and between districts.

# PRESENT FINDINGS

# ggplot, graphing

# Make conclusions about the original challenge/research question.

# RMarkdown

#Ways to match up the data:

ROI\_12.15.14: County name

City Revenue: County name

#questions about the data

1) What is the difference between ROI\_12.15.14 and ROI\_old. Should we bother to look at the old one?

**Los Angeles Rental Census Data Dictionary**

**V1-Jan 11 2017**

1. GEOID
   1. Census tract identification number
2. ACS\_MFRENTALS
   1. The number of rentals in multi-family units in a given tract, from the 2015 5 year wave ACS Data Table B20532
3. ACS\_Owners
   1. The Number of owner occupied units of all types, from the 2015 5 year wave ACS Data Table B20532
4. ACS\_SF\_Rents
   1. The number of rentals in single family and single-family attached units in a given tract, from the 2015 5 year wave ACS Data Table B20532
5. LUPAMS
   1. Data from the LUPAMS office on the number of multi-family rentals registered on site (as required by the City’s proactive rental inspection program).
6. HCIDLA
   1. Data from the HCIDLA office on the number of multi-family rentals registered on site.
7. LUPAMS\_RC
   1. Data from the LUPAMS office on the number of *rent-controlled* multi-family rentals registered on site (as required by the City’s proactive rental inspection program).
8. HCIDLA\_RC
   1. Data from the HCIDLA office on the number of *rent-controlled* multi-family rentals registered on site.

**Class Presentation Notes 2017-2-22**

Matt Palm, lead researcher with CAL HFA Housing Finance Agency.

<https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>

300 million people. Goal used to be 5% of population for Census data. Since expensive, transitioned to new mode. Starting 2000s, smaller samples more often, pooled into estimates that are statistically robust.

<http://www.urbandisplacement.org/map/la>

Measures gentrification for LA and the Bay Area. Data gathered on a rolling basis between 2009 and 2013; those estimates may be off because we were coming out of a recession. Mapping tools like this are being used by policymakers for housing, etc. But when data is on such a time lag, this is an issue; gentrification is already happening. In 2017, we are using data from 2009-2013 to measure changes in our neighborhoods and this is problematic. We want to understand how using richer data could change the narrative. In 2002, city pass a pro-rental inspection program in SF: all rental housing units in the city are inspected to make sure public health needs are met, this is especially important in immigrant communities—this is a census of registered housing stock. Our task is comparing that to what the census is telling us. Why this matters: rental housing is important for low-income families, temporary workers, and immigrant families.

Beurocrats make people fill out a lot of forms; that data does not always get used.

This project is an attempt to augment the American community survey with better data so that you can address problems in real time. We want to see if existing data can function as a leading indicator, so that we can take measures to invest in affordable housing, etc.

LA, Santa Cruz, Sacramento—cities with this data from landlords.

<http://interact.regionalchange.ucdavis.edu/roi/>

Regional Opportunity Index: Rental exposures to air quality issues, etc. Compare this to the census data.

The spatialized nature of social change is working against us in terms of how the census actually operates. Tip: A lot of the data you might think you need to gather might already exist, especially equity issues (US).

RESEARCH QUESTION:

[Mattdpalm@gmail.com](mailto:Mattdpalm@gmail.com) (cc Deb on emails to him). Research questions:

* **Educational data**
* Jobs access
* **Public health**, in particular, air quality exposure. Infant mortality rate.

Does looking at census data collected by the city give a fundamentally different story than data collected by the ACS?

Users: city of LA, append to the regional opportunity index, state of California—where to target affordable housing (the problem with using ROI is that it is bringing back red lining—reproducing the inequality), UC Berkeley