Data Description: IPI5

Ryan Pacheco, Ashley Gilbert, Ben Whitehead

Our project is to use data from the Census, American Community Survey data, and other sources to determine whether a city is growing in a sustainable manner.

1. What data source(s) are you using?

- American Community Survey Data
- Census Data
- Envirofacts Data Service API
- DARTE Annual On-road CO2 Emissions

2. How large are these data sources? List tables, record counts, and (relevant) columns or attributes.

We are planning on using the following variables from the census data:

NAME:	City Name
P002001:	Total Population
P002002:	Total Urban Population
P002005:	Total Rural Population
H001001:	Housing Unit Count
P013001:	Median Age (both sexes)
H003001:	Occupancy Status For Housing Units
P027001:	Presence of Non-Relatives
H005001:	Total
H005002:	For rent
H005003:	Rented, not occupied
H005004:	For sale only
H005005:	Sold, not occupied
H005006:	For seasonal, recreational, or occasional use
H005007:	For migrant workers
H005008:	Other vacant

We are planning on using the following from the ACS data:

B23024_001E:	total of POVERTY STATUS IN THE PAST 12 MONTHS BY DISABILITY STATUS BY EMPLOYMENT STATUS FOR THE POPULATION 20 TO 64 YEARS
B25091_001E:	MORTGAGE STATUS BY SELECTED MONTHLY OWNER COSTS AS A PERCENTAGE OF HOUSEHOLD INCOME IN THE PAST 12 MONTHS
B01003_001E:	Total Population
B27001_001E:	HEALTH INSURANCE COVERAGE STATUS BY SEX BY AGE
B00002_001E:	Unweighted sample housing units

We are still looking into the Envirofacts Data Service API. The documentation is very messy, and we just learned about this resource procured by the EPA. We're hoping to find something along the lines of pollutants that we can correlate with location/FIPS data.

Similar to the Envirofacts Data Service, the DARTE CO2 Emissions Data should allow us to tie some kind of emissions rating to a particular location. Helping us to identify sustainability trends for specific cities. We are still digging into this data set, because we just discovered it as a part of a Boise State Public Radio story recently written.

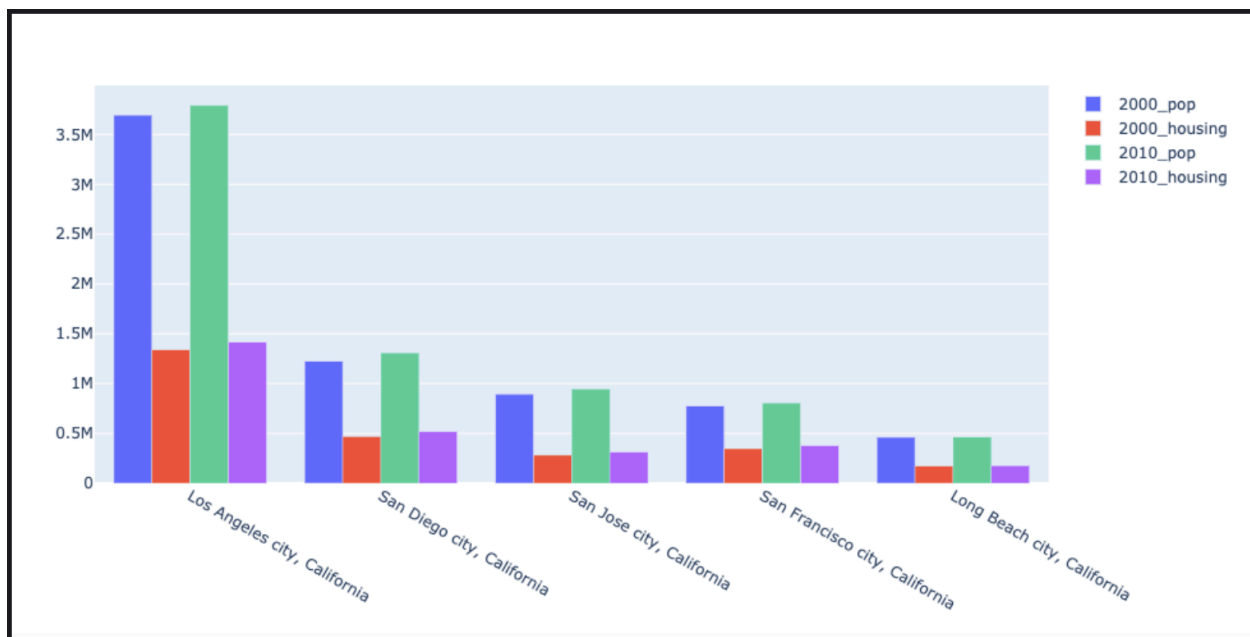# 3. What entities do the sources describe? How many of each entity do you have?

- ACS gives us access to the demographical data associated with a particular region.
- Census gives us population, and vocational data
- Envirofacts Data Service API gives us access to greenhouse gas emissions data per location.
- the DARTE data also gives us locational CO2 emissions data.

4. If relevant, how are records or key attributes distributed? If you have a 'time' or a 'count' attribute, showing that distribution is useful; this may be a distribution of aggregates (e.g. the distribution of average per-day counts at different locations).

Most of our data is a point in time count. For the census, it is point in time counts every ten years.

For the Envirofacts data, it looks like the data is collected per facility, so there may be a significant task of mapping facilities to FIPS regions and times we have for the other data sources. The data seems to be reported yearly.

Darte data is a count for a location in time.

Relative and Actual population growth: