# Artificial Data:

## Inspiration:

At Census Beauro website, the data about [Population & Housing](http://www.census.gov/prod/www/decennial.html) is giving in pdf format from 1790-2010 that requires OCR to extract data, a very good resource found that can extract the desired data over the time till 1920 [NHGIS Data Finder](https://data2.nhgis.org/main). But over the time period new variables are introduced in the census, so, it would be difficult to make data points on time series over multiple years i.e. we can backtrack only two variables “Total Population & Person by Sex” from 1920 to 2010. So, question arises how to arrange the data that can at least include all variable from the latest two census held in 2000 & 2010.

The Naïve Answer came in mind from the [Census 2010 History Overview](https://www.census.gov/history/www/through_the_decades/overview/2010_overview_1.html), which announced the shortest census of history only considering 10 question in questionnaire related to name, gender, race etc and rest of data collection from the ACS; “ARTIFICIAL DATA”. So, we can make data points over the years from the various surveys e.g. ACS, Economic Census, Business Dynamic Statistics etc. It would be clearer by considering the following figure.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|  |  |  |  |  |  |  |  |  |  |

Where Rectangle show the Census, ACS, Economic Census, Business Dynamic Statistics; Blue as Census, Yellow as ACS, Green as Economic Census and Orange as Business Dynamic Statistics. And triangle represents the variables under focus.

## Data Variables:

[United States Census 2010](https://www.census.gov/2010census/data/) provides data products of census 2010, which contains Summary files, Demographic Profiles, Households and Families, The Asian Population, Hosing characteristics etc. Some of data tables (summaries) of Census 2010 are extracted from the website and shown below the table No.1 with their reference files.

Considering three variables, I have searched in American fact finder and found the topic ids, shown below:

1. Asian Population

Key word: [B02011](http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t)

1. House hold by type:

Key word: [S1101](http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t)

1. Housing Units:

Key word: [B25001](http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t)

|  |
| --- |
| Asian Population: <https://www.census.gov/prod/cen2010/briefs/c2010br-11.pdf> |
| House hold by type : <https://www.census.gov/prod/cen2010/briefs/c2010br-14.pdf> |
| Housing Units: <https://www.census.gov/prod/cen2010/briefs/c2010br-07.pdf> s |

Table No.1. Census 2010 Data Tables of variables.