Part 2: Full Specification

App Name: HUD Fair Market Rents Data Analyzer

Runtime: Python 2.7

Dependencies

The following external modules are expected to be dependencies of this application:

- Tkinter, etc. (tkFileDialog, ttk)
- Matplotlib, mpl_toolkits.basemap
- Numpy, scipy
- json
- urllib
- csv

Concept of Operation

The following ordered list describes the expected user experience and high level use cases.

1. Assumption: The program folder structure (including program code and initial data set) will be available on the user's machine. The user might download a ZIP archive from an internet website, or access the folder structure from a CD or flash drive.
2. The main python script will be launched in python v2.7.
3. The main user interface form will display.
4. The user will invoke the function to load the data set files.
    a. This could be initiated automatically at app launch (TBD).
5. After data is loaded, the user may select various properties (Year, County, etc) and launch the data analysis and/or visualizations.
    a. Map Visualization for a given year and # of bedrooms.
        i. This function could be used to gain an understanding of relative rent differences between various locations across the United States.
    b. Multi-year Analysis for a given County/State.
        i. This function could be used to understand how rent values are changing from year to year for a given County and estimate future rent values.

Source Dataset

The Housing and Urban Development (HUD) data for Fair Market Rents (FMR) are published annually, and periodically revised. The FMR data points are required to be published by HUD as part of Section 8 of the United States Housing Act of 1937 and estimate gross rent for the given location including shelter rent plus the cost of all tenant-paid utilities except telephone, cable or satellite television service, and internet service (Office of Policy Development & Research).

Microsoft Excel spreadsheets are available on the HUD Data Sets portal at the following link:

http://www.huduser.org/portal/datasets/fmr.html

Each year is published in a separate Excel spreadsheet. The data sets for an 11 year period from 2005 – 2015 (effective Oct 1, 2014, as revised) were downloaded, saved and converted to comma separated value (CSV) format. These CSV files will be included with the program package for distribution.

Original Data Format

Unfortunately, the HUD data sets do not follow an identical format from year to year. Generally, they include a basic set of common data and this common data will be used during the project. The data columns appear is different orders and with different header names from year to year. As such, a data mapping feature is expected to be coded as part of the data loading function to ensure common data points are matched across multiple files. This functionality will be transparent to the user unless an unknown year is imported, in which case an error will be reported.

The data points which will be loaded and mapped are described in the following table.

Data Point	Description
state_alpha	The 2 character abbreviation for the state name.
countyname	The plain text name of the county in which this area or town exists.
Areaname	A common name for the geographic area for which the rent value data is relevant.
fmr0	Gross rent in USD for a Studio (zero separate bedrooms)
fmr1	Gross rent in USD for a 1 bedroom
fmr2	Gross rent in USD for 2 bedrooms
fmr3	Gross rent in USD for a 3 bedrooms
fmr4	Gross rent in USD for a 4 bedrooms

Geographical Location

The raw data does not include latitude/longitude information. As such, the county name and state abbreviation will be used with the Google Geocoding API to augment the data set with latitude and longitude values. This is being done with the intention of using the matplotlib basemap as a vehicle for visualizing the rent values throughout the United States as described in the Data Visualization section of this specification.

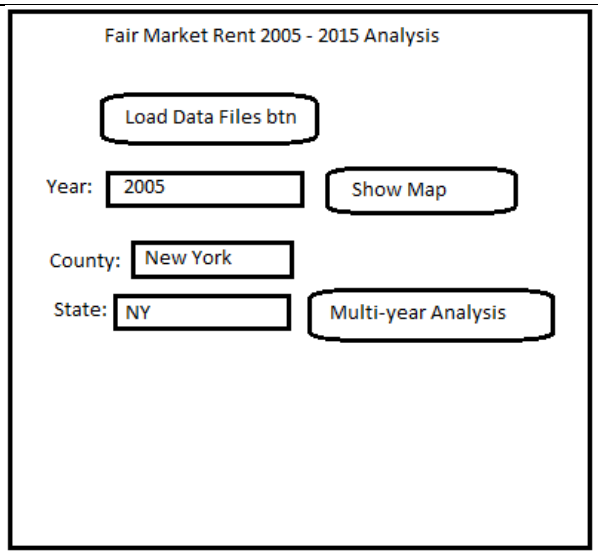
Data Analysis

The key analysis which will be performed is linear regression on a particular location's rent prices. After selecting a state and county, the rent values for each of the 0 – 4 bedrooms across the eleven year period pass through scipy's `curve_fit` function for linear regression. The intercept and slope values will be output to the user in addition to a visualization of the location's data and linear model. As time permits, additional functional models may be incorporated.

User Interface

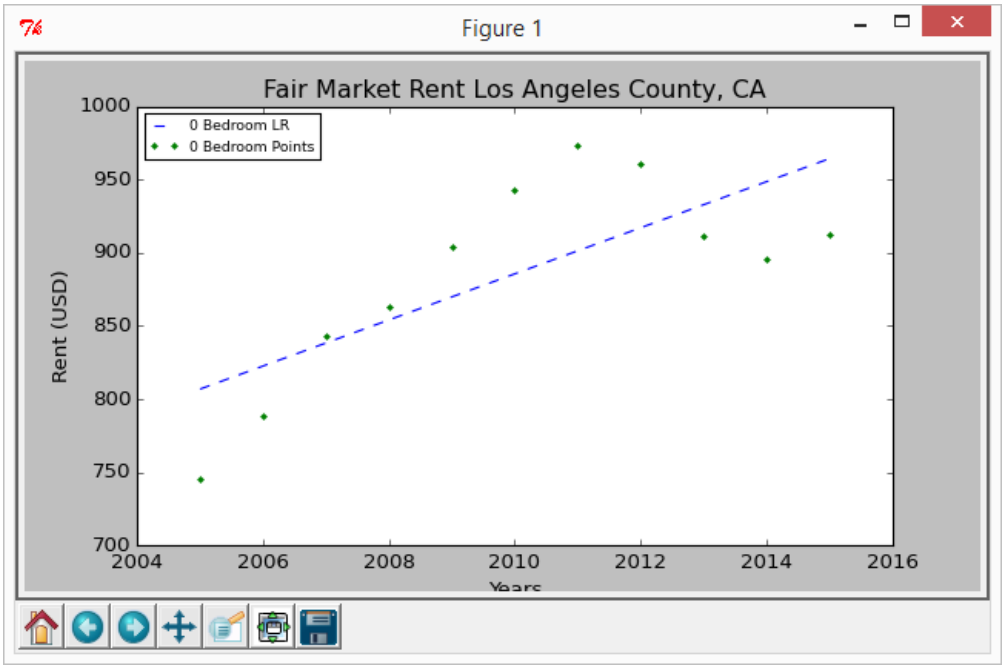
A basic user interface will be provided via Tkinter. A wire frame diagram of the user interface and description of the controls follow.

- Load Data Files (button): Loads the data files from the expected subfolder (.\Data\).
- Text field for entering year used during map visualization.
- Show Map (button): configures and launches the map visualization.
- Text fields for entering County & State for multi-year analysis
- Multi-year analysis (button): launches the linear regression and charting visualization for the specified county/state.



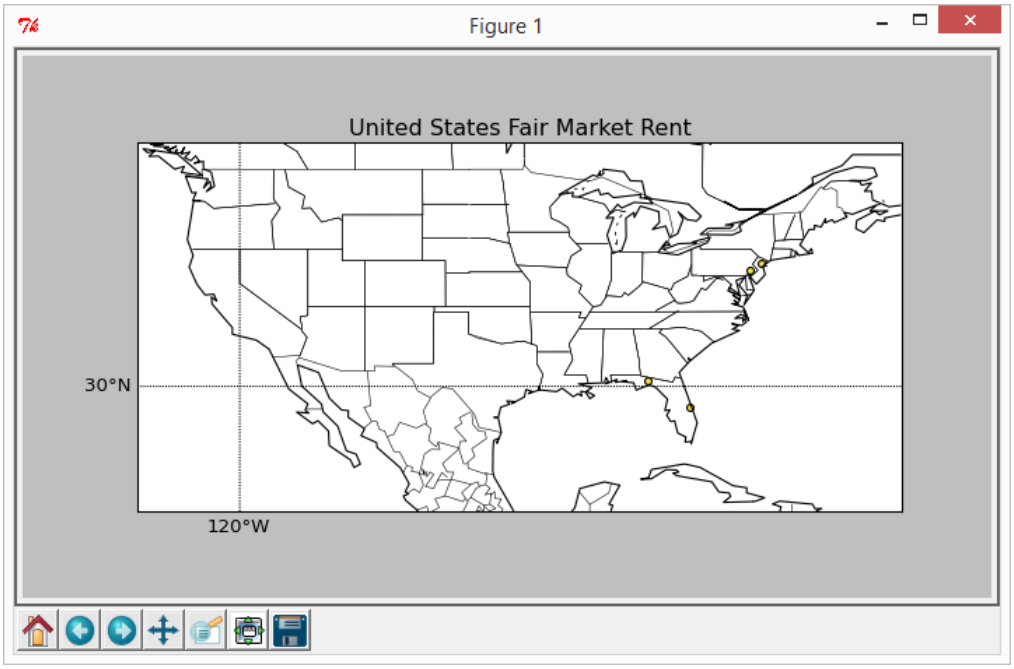
Data Visualization

The Linear Regression for specific counties will result in a plot being displayed similar to the following screenshot.



Using the latitude/longitude information from the Google Geocoding API, the Fair Market Rent values will be overlaid on a map of the United States. The following figure provides an illustration of the

concept using four data points (Union County, NJ; Delaware County, PA, Leon County, FL; Brevard County, FL)



References

Office of Policy Development & Research. *Fair Market Rents*. U.S. Department of Housing & Urban Development, July 2007. Web. 9 Nov. 2014. <http://www.huduser.org/portal/datasets/fmr/fmrover_071707R2.doc>.