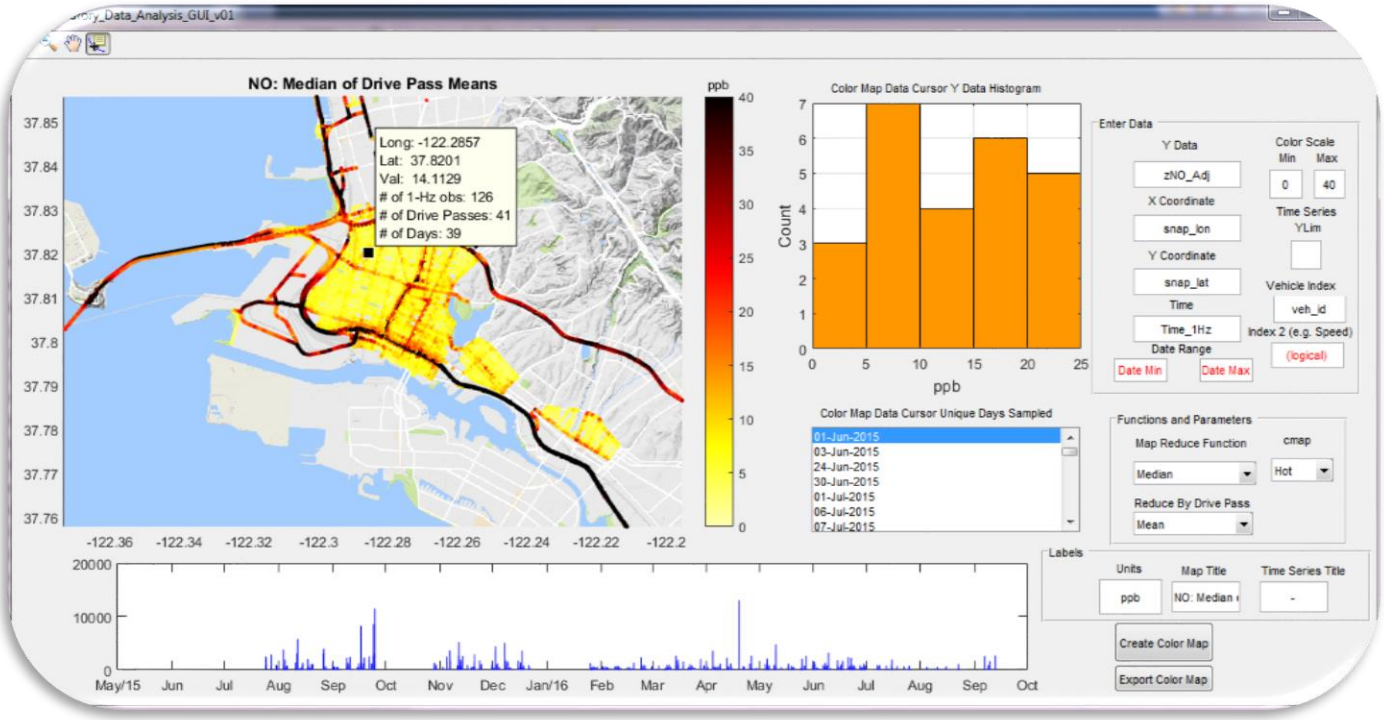
Exploratory Data Analysis GUI Tutorial



Version 1, Aug. 25th, 2017

Please send bugs and future requests to: Kyle Messier – kpmess@gmail.com

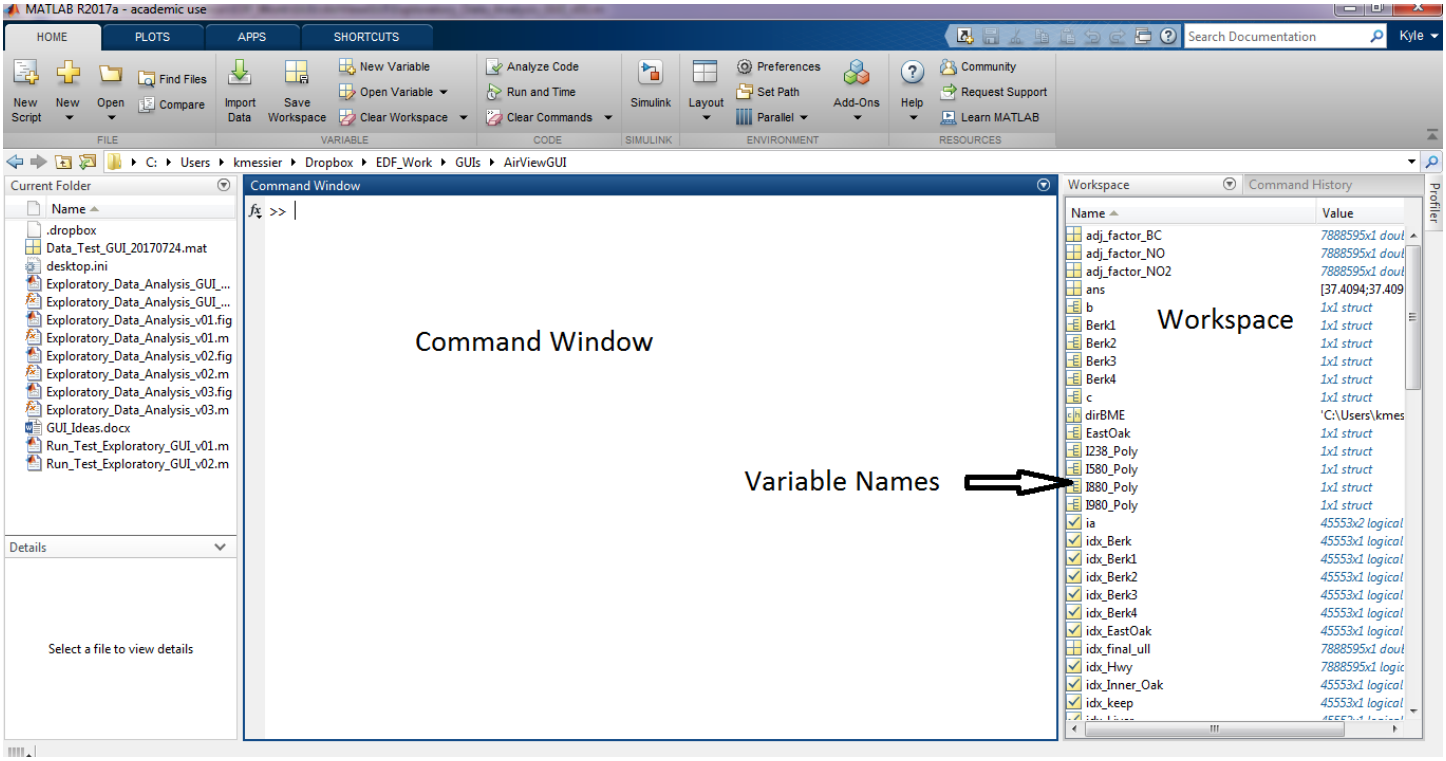
i. Before Getting Started

1) Matlab GUI's consist of two files: .m and a .fig file with the same base name. The name of this GUI's files are:

Exploratory_Data_Analysis_GUI_v01.m

Exploratory_Data_Analysis_GUI_v01.fig

2) This GUI requires beginner level experience in Matlab. A basic understanding of the workspace, command window, and Matlab variables will make this a lot smoother.



3) ***There are a number of files that are requirements for this to run correctly.*** I've outlined the file names and where to get them below:

- plot_google_map.m: Available on the Matlab File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/27627-zoharby-plot-google-map
- dynamicDateTicks.m: Available on the Matlab File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/27075-intelligent-dynamic-date-ticks
- export_fig.m: Available on the Matlab File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/23629-export-fig
- diverging_map.m: Available online at the following website: http://www.kennethmoreland.com/color-maps/
- create_Drive_Pass_Index.m: A function I created. Available from directly or through my GitHub. https://github.com/kpmessier/SV-Library/tree/master/SV_Library2.0/Basic_Functions

4) Once you have all of those files downloaded, you need to make sure they are accessible to Matlab. You do this by adding them to your search path in the command window:

>> addpath('C:/Path/to/your/files/)

I. Starting the GUI

1) A few options:

A. Use the navigation tool at the top of the command window to put your working directory to where the GUI and data are.

B. Use addpath to add the file location of your GUI and example data to the search path
C. A combination of those for the GUI and example data

2) The GUI requires that the data be read into your Matlab workspace. Examples of how to read in data can be found online:

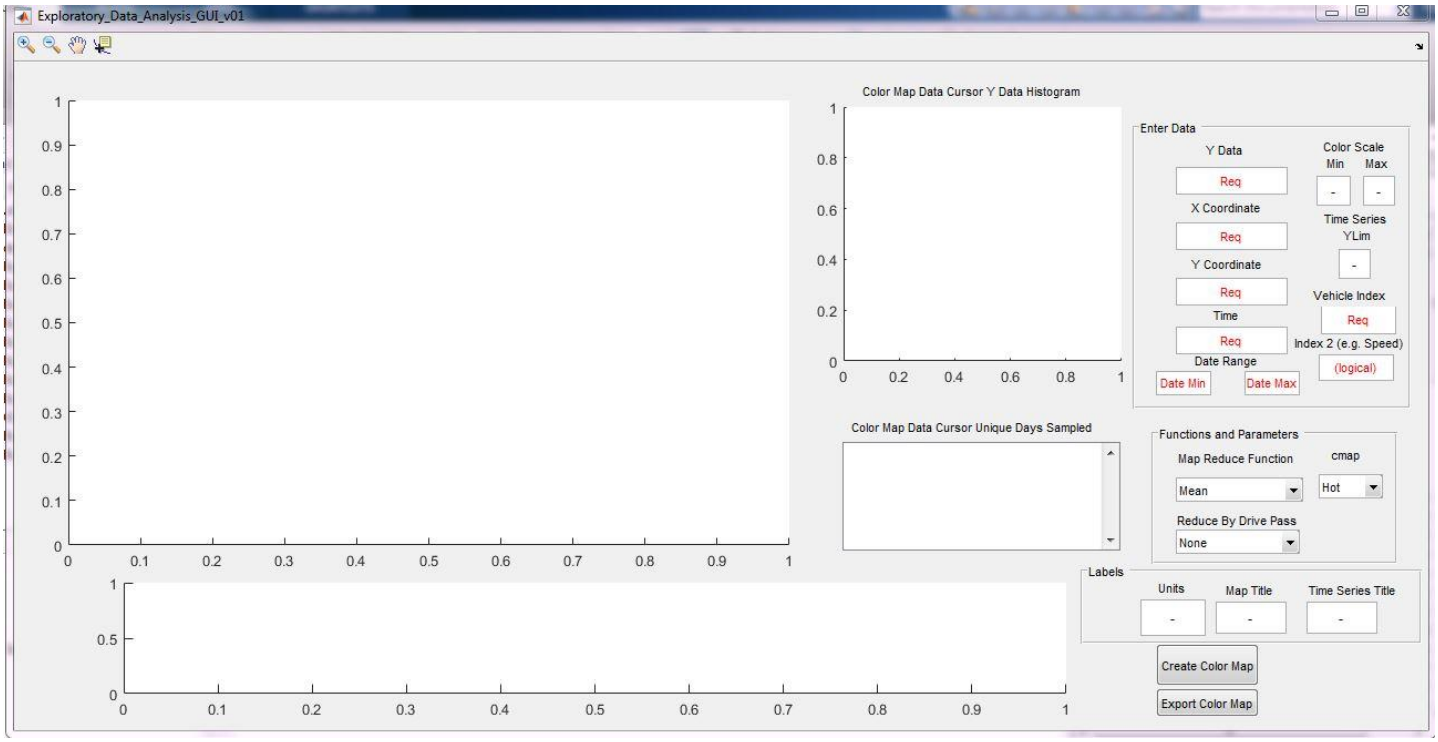
For this example, load in the example .mat file via the command window:

>> load Data_Test_GUI_20170724

2) Call the GUI in the command window:

>> Exploratory_Data_Analysis_GUI_v01();

3) This opens up the GUI:



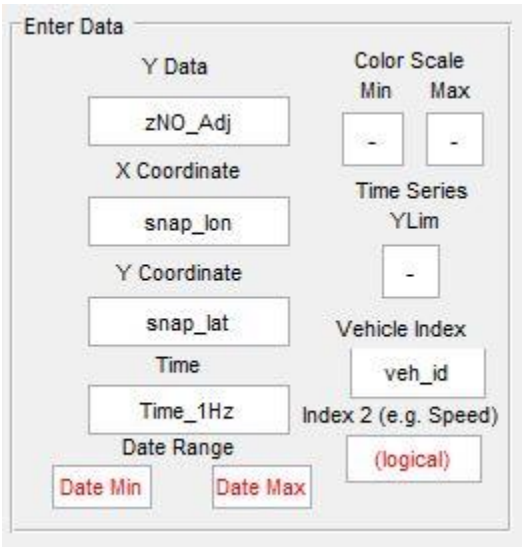
II. Input Data

1) Data is input by typing it into the windows on the right.

2) There are 5 required input variables, which are denoted with Req (i.e. Required)

- Y Data – this is the pollutant data you want to map
- X coordinate – e.g. longitude
- Y coordinate – e.g. latitude
- Time
- Vehicle Index – this is required to properly calculate the drive pass means. If there is only one car, then you can create a vector of ones >> veh_id = ones(length(YData),1);

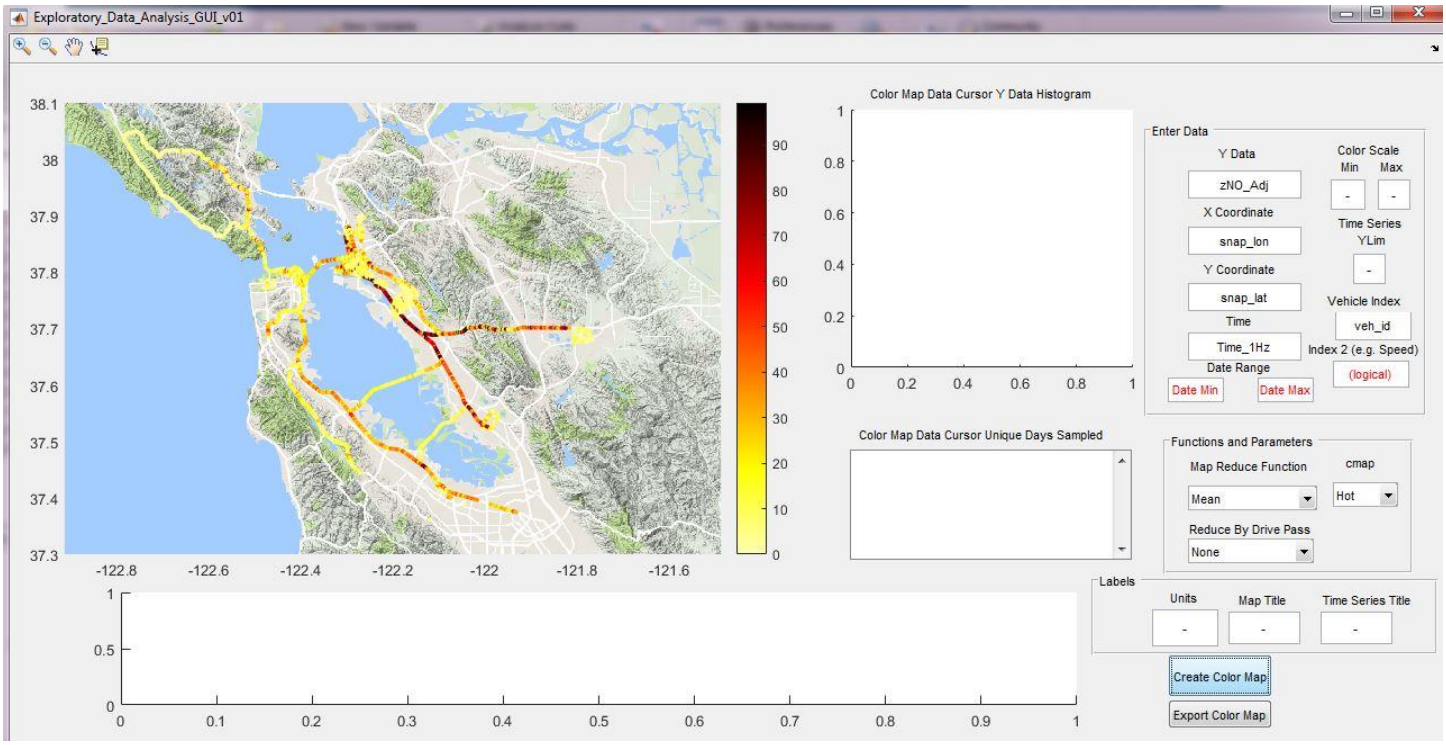
3) Type the following field names into the GUI



4) Once you have the required input, you can create the color map.

Press the “Create Color Map” Button at the bottom right of the GUI

Notes about the “Create Color Map” Button: Anytime you change input including the YData, coordinates, time, data range, indices, map reduce functions, color scale, and labels, then *you have to press the button again* for the changes to be reflected on the map. Also, note that the mouse changes to a “thinking” while it is creating the map, so don’t try to do other things on the GUI and screw everything up. Using a drive pass reduce function (e.g. Mean of drive passes) takes a little longer, so you’ll have to be patient.



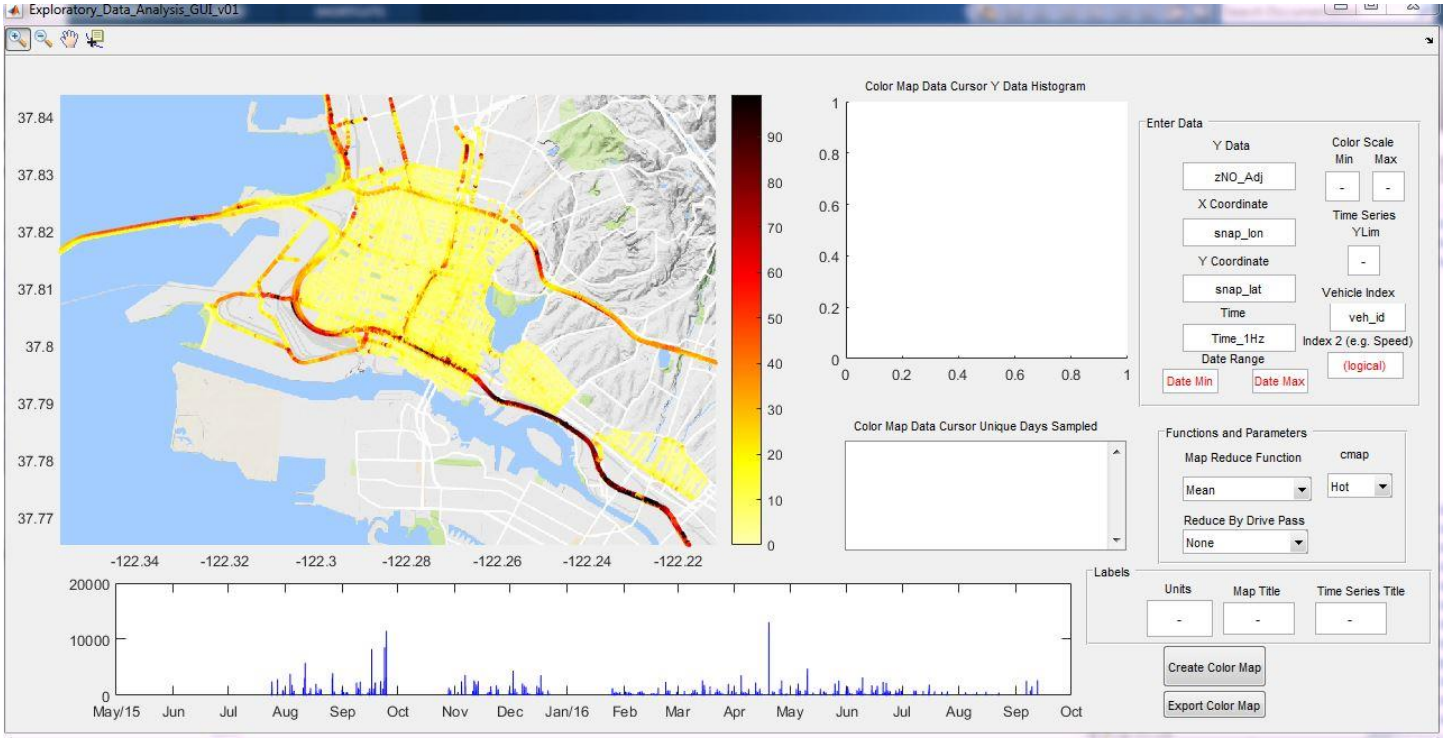
5) The color map plots all of the locations with the default reducing function (mean) and with the default color scale (0 to 95th percentile of YData).

II. Exploring the data

1) You can use toolbar at the top to zoom in and the data cursor to explore individual data points.

2) ***Zoom In:*** The zoom in tool adds the entire time series of data within the window to the plot below.

Try using the zoom tool to zoom to an area (e.g. West Oakland)

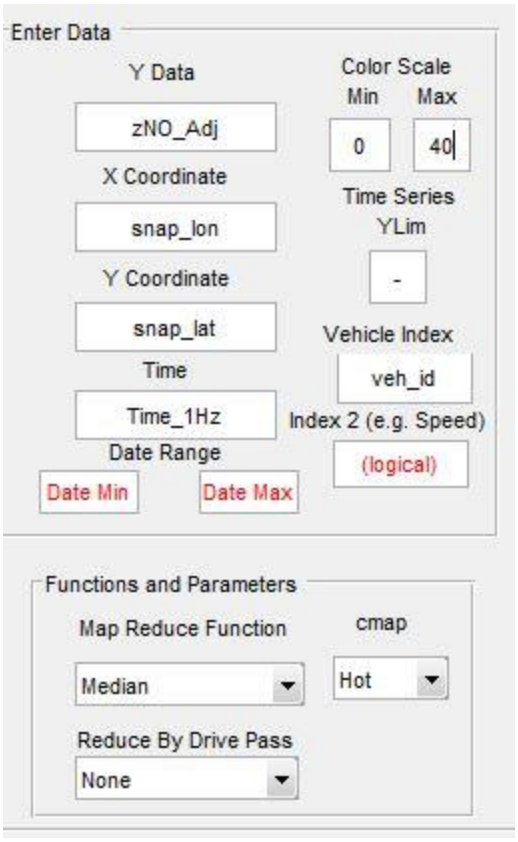


3) Change the reducing functions.

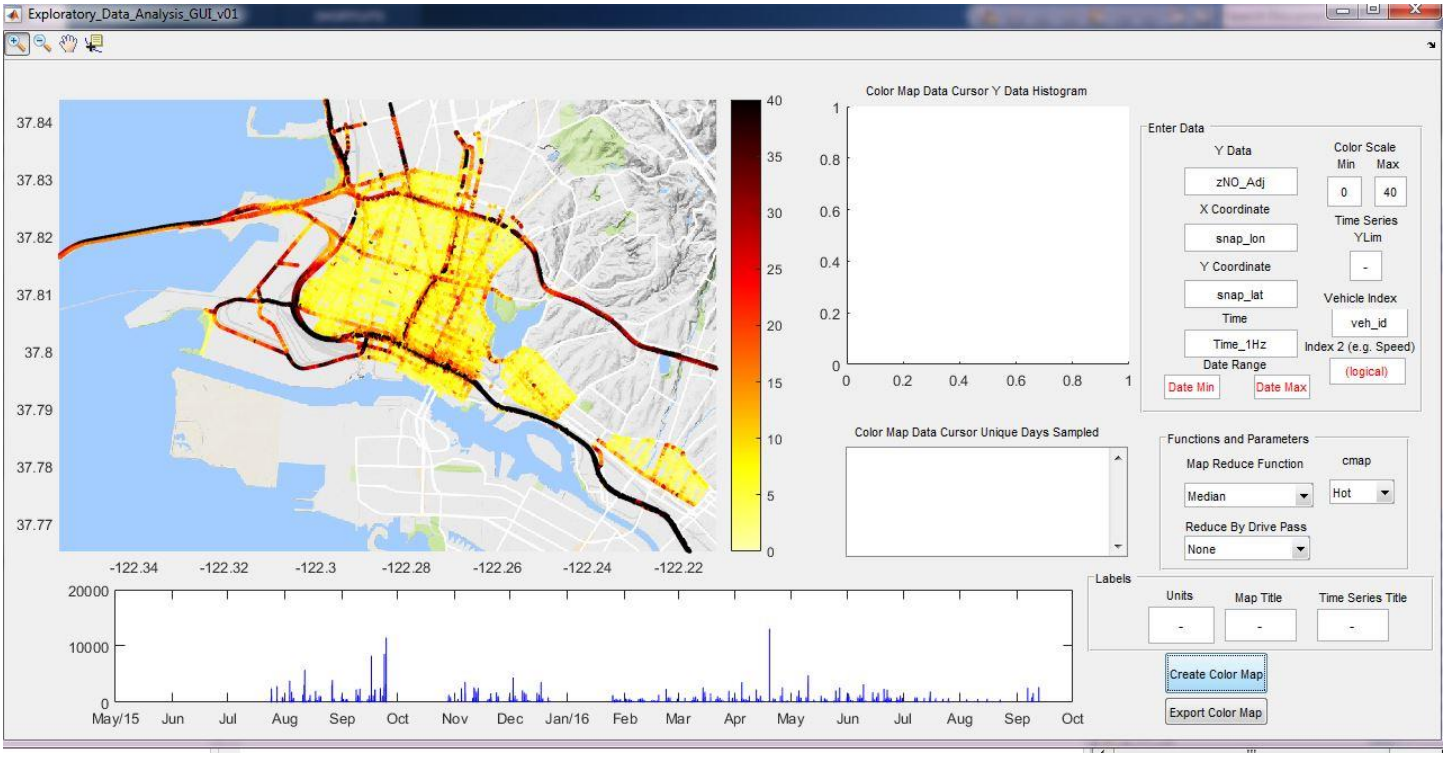
By reducing function, we mean the function that is calculated at each individual snapped point. So the mean results in a map showing the mean of concentration as each point.

You can see all of the options from the pull down menu. Select ‘median’.

4) Let’s also change the color scale. In the top right, set the min to 0 and max to 40.

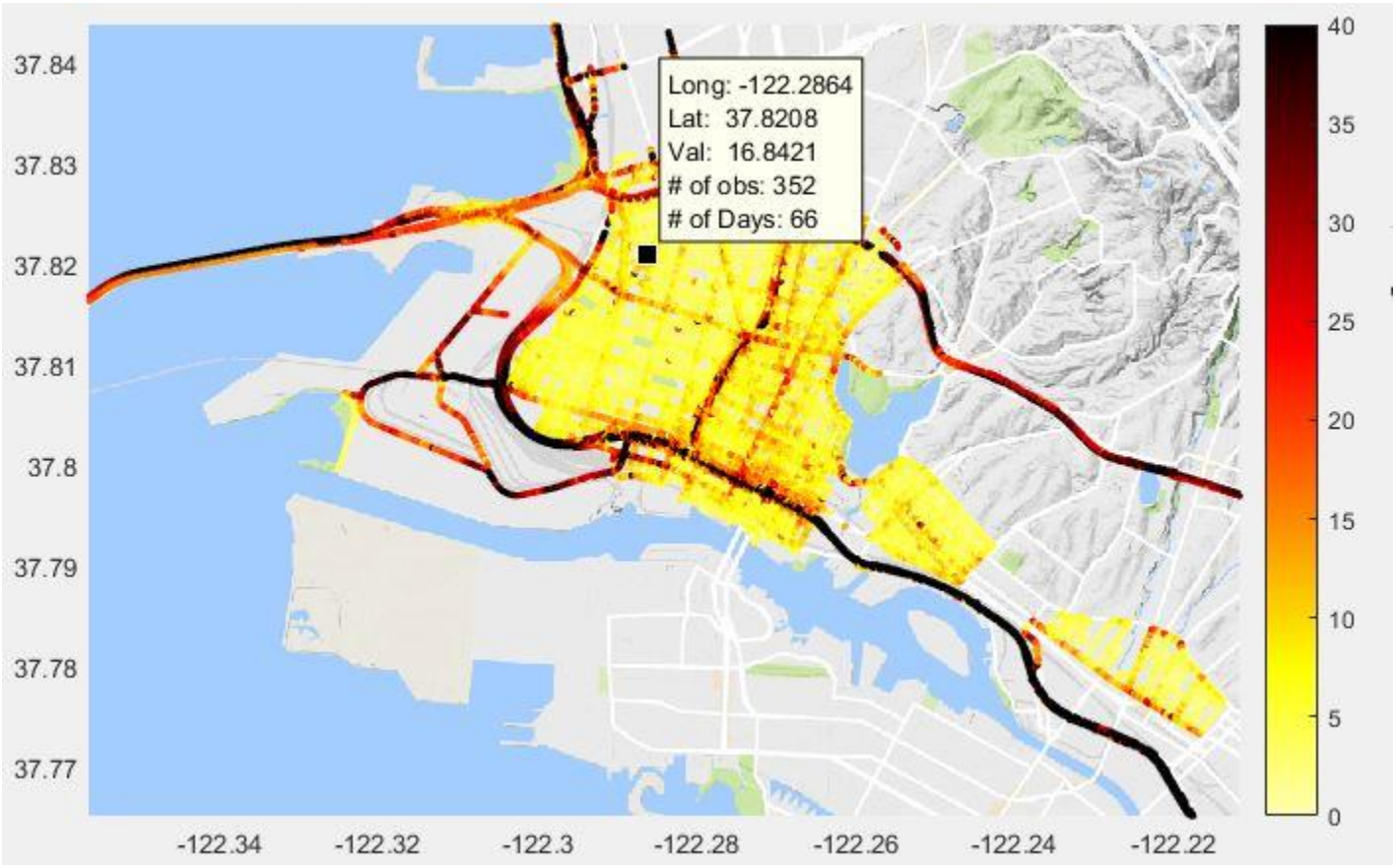


5) Then press ‘Create Color Map’ button.

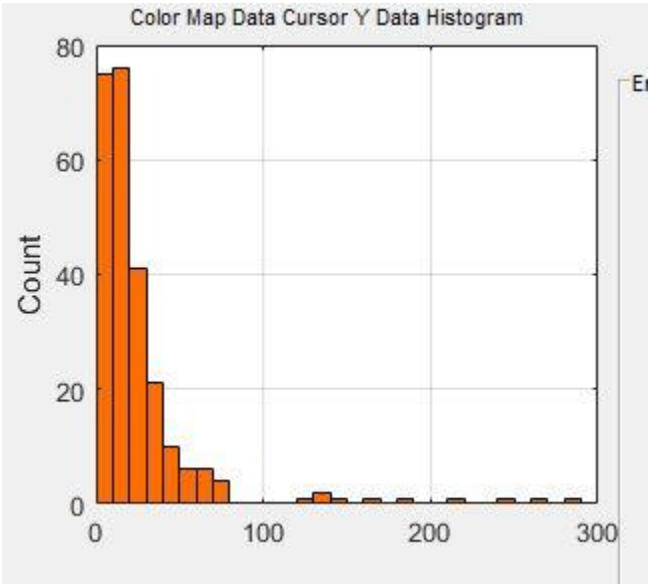


6) You can use the data cursor button  at the top of the GUI to select individual points.

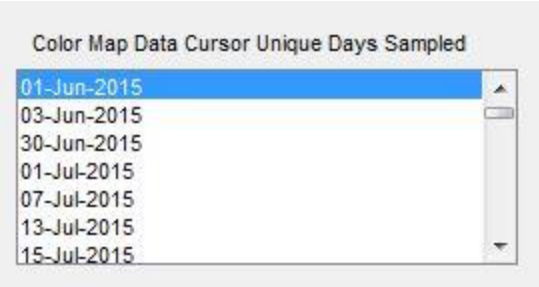
Use the tool to select a point.



The tool shows a pop-up window called a ‘data-tip’ that shows the coordinate longitude, latitude, the value being mapped (i.e. mean, median, etc.), the number of 1-Hz observations, and the number of unique drive days.



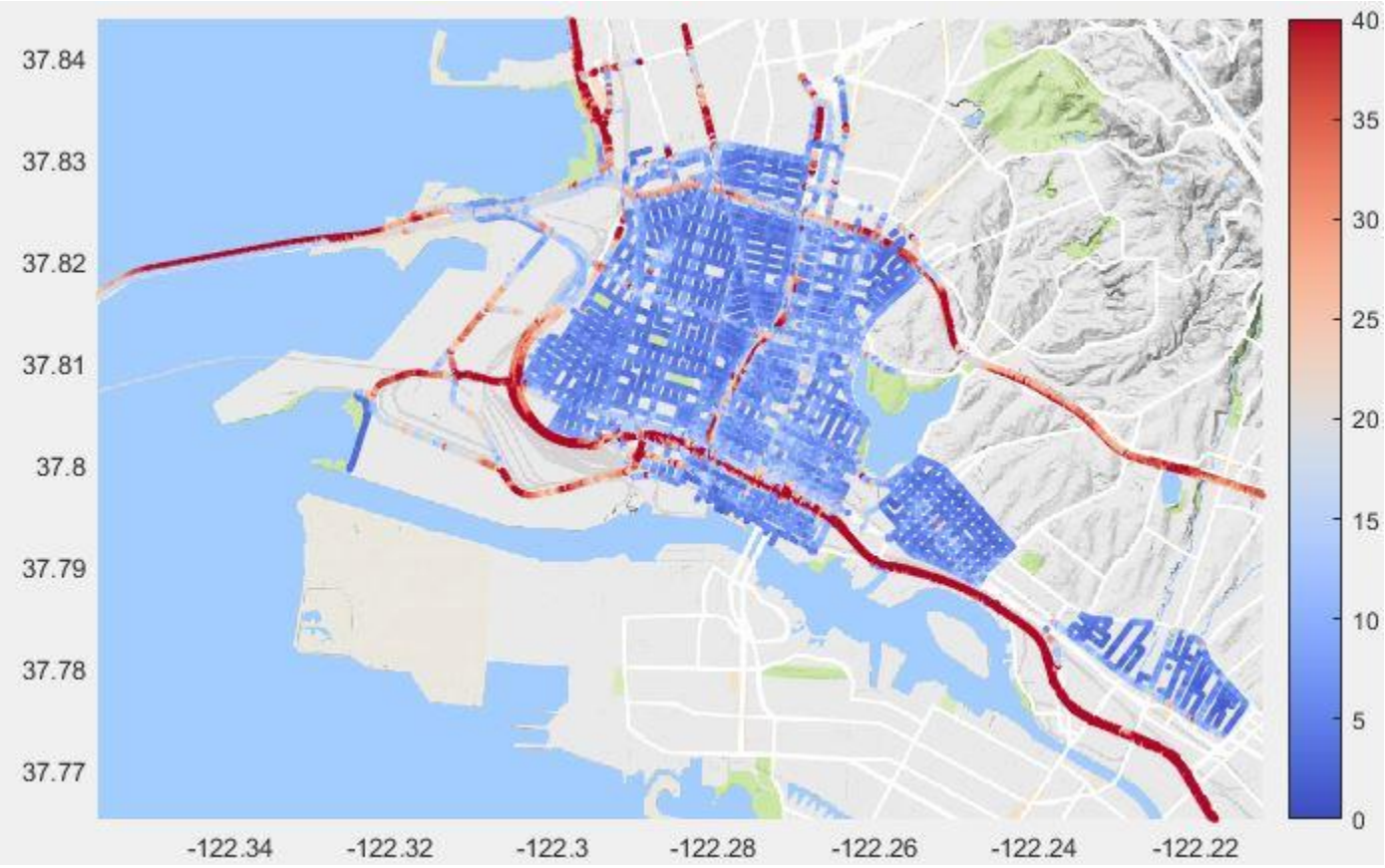
It also shows the 1-Hz (or drive pass average) data for the selected point. The color of the histogram corresponds to the value represented by the color scale.



The unique days that have been sampled at that location are shown in this window. You can scroll up and down to see them all.

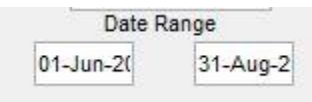
The time series of the individual is also seen in red superimposed over the time series of the map window. It may be hard to see for a lot of points.

7) Adjust the color scheme or cmap. There is a pull down menu with a few options for colormaps. The map updates automatically.



8) Adjust the time range

To adjust the time range you can input the minimum date and maximum dates in their respective boxes. Any date format acceptable to Matlab is okay, but it is easiest to use MM/DD/YYYY format.



The date shown changes automatically to a string format (e.g. 01-Jun-2016)

Press the ‘Create Color Map’ button to update the map. The time series will update with the next zoom.

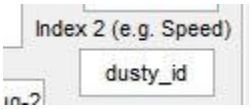
9) Add another index to the data.

There is another optional input area where you can put in a logical index. The index should be the same size as the YData and be 1 (true) for data you want to map and 0 (false) otherwise.

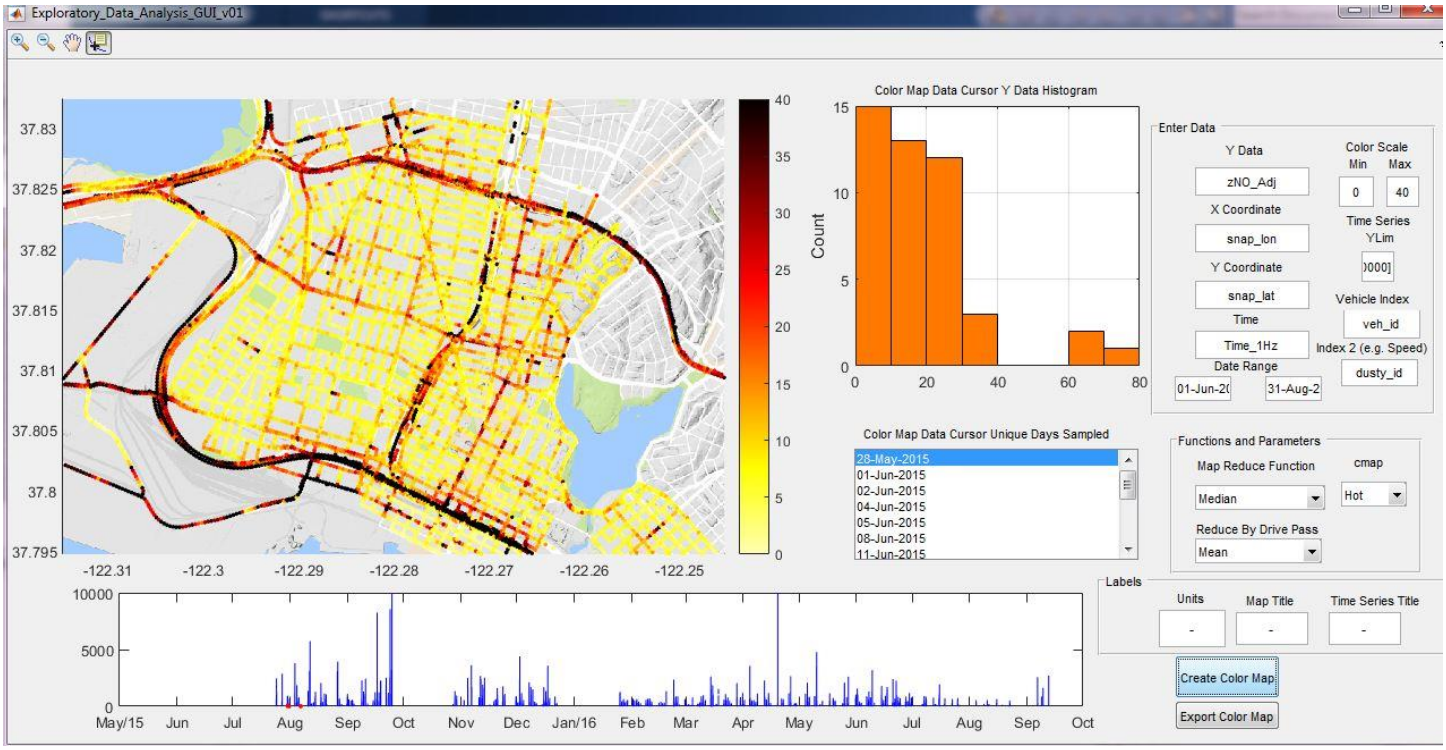
You can create logical indices such as limiting to only 1 car or when the speed is greater than 0 kmh.

In the command window, create an index for only the car “Dusty”

>> dusty_idx = veh_id == 1;



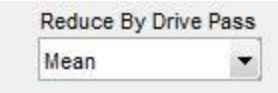
Click the “Create Color Map” button.



III. Reduce by Drive Pass

The default reducing function at each point is applied to all of the 1-Hz data that occur at that point. An alternative is to first calculate the mean or median for each drive pass, that is all of the 1-Hz data that are sampled as the vehicle is driven past this particular point. In practice, this may range from one or two 1-Hz samples, to dozens or more than a hundred in the more extreme cases such as long wait intersections.

To apply a drive pass reducing function use the drop down menu: the options are mean or median. Then the map reduce function will be applied to these mean or medians.

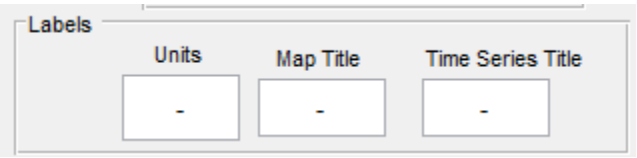


The data-tips now display the number of unique drive passes and the histogram plots the drive pass mean or medians. Note how the [# of 1-Hz samples ≥ # of Drive Passes ≥ Unique Drive Days]

IV: Errors

When entering input into the GUI, if you enter in a variable that does not exist then it will tell you with a pop-up display box. Simply re-enter the correct variable name.

V: Labels



You can enter in the units label, map title, and time series title into these boxes. They will appear once you re-create the color map or time series.

VI. Exporting the Color Map

You can export the color map to a picture file that can be brought into other programs by clicking the 'Export Color Map' button in the bottom right corner.

When you click it, there are a couple dialog boxes asking you to input the filename and format.