

# Creating a Tableau Data Visualization on Cincinnati Crime

## By Jeffrey A. Shaffer

### Step 1 – Gather and Compile the Data:

This data was compiled using weekly files provided by the Cincinnati Police. Each file includes 26 fields, including these important fields; incident number, date & time reported, date & time from/to, offense, address and neighborhood.

The data visualization shows all crimes reported in the Cincinnati area for 2013. Since the data is reported weekly it was necessary to compile 25 weekly spreadsheets into a single spreadsheet. It's certainly possible to combine these manually, but there are a number of tools available that will do this in a more automated fashion. One useful tool is an Excel Macro written by Chris Kent that makes combining multiple Excel files really easy (<http://tinyurl.com/MergeExcelFiles>). Creating and using a quick keyboard shortcut to the macro it took no time at all to combine these files. Also, it would be fairly easy to customize this code if there were a need to combine hundreds or even thousands of Excel files.

### Step 2 – Geocode the Addresses

The next step is to geocode the addresses. This is the process of converting an address into latitude and longitude. In this case there are 16,612 addresses that need to be geocoded. There are a number of free batch geocoding tools available, each with different daily limits. Here are three recommended tools for geocoding.

<http://www.findlatitudeandlongitude.com/> - this is an excellent tool that offers many options, including geocoding a single address, locating a point on a map, reverse geocoding (i.e. converting a latitude and longitude to an address) and batch geocoding (click on **Batch Geocode** from the Menu options box near the bottom of the home page) .

<http://geoservices.tamu.edu/Services/Geocode/> - another excellent tool that allows for 2,500 free searches at a time and allows registered partners to search more. This tool also allows the user to upload a file, monitor progress and then download the finished results. They also offer cheap prices for very large geocoding jobs.

<http://www.juiceanalytics.com/tags/geocoding/> - Juice Analytics has created a terrific tool for geocoding that works directly from MS Excel using a Yahoo API. It is necessary to have a Yahoo account (which is easy to do and free) and then obtain a Yahoo Search ID. There is also a nice feature allowing the user to download the KML file.

In order to geocode and address these tools require the full address, city and state. In the Cincinnati Crime dataset there is only address so it was necessary to append this to the address. This is simple to do right from excel using the **concatenate()** function or simply combining the strings using the **&** symbol.

Example:

	A	B
1	<b>Address</b>	<b>Combine Address with City and State</b>
2	123 Main Street	123 Main Street, Cincinnati, OH
3	123 Vine Street	=A2&" , Cincinnati, OH"
4	<b>OR</b>	
5		
6	123 Main Street	123 Main Street, Cincinnati, OH
7	123 Vine Street	=CONCATENATE(A6, ", Cincinnati, OH")

After appending a new column for address with city and state it is best to copy the entire column and replace with values so that there are no formulas in the spreadsheet. It's always best to import true values and not formulas into Tableau.

Note: zip code is also necessary for geocoding with the Texas A&M tool, however, the zip code does not need to be accurate to get the results. For example, choosing a central zip code for Cincinnati would allow the tool to properly geocode.

There are other fields that are appended to the Cincinnati Crime dataset for the visualization, but these are the essential fields that are necessary in order to create the base data visualization in Tableau. So the next step is loading the Excel file into Tableau.

### Step 3 – Load the data into Tableau and prepare it for analysis

Loading an Excel file into Tableau is very simple. First open Tableau 8 and then select **Connect to Data** and choose **Microsoft Excel**. After choosing the Excel file click **import all data** so that Tableau creates a full data extract. Tableau will then load all of the data.

The next step is to verify that the fields imported correctly into Tableau. Typically it's necessary to review the dimensions and measures to make sure that Tableau organized them correctly, however, in this case the only measures that are in the dataset are latitude and longitude which Tableau handled correctly on import. It's easy to check this. From the Measures pane, drag **Longitude** to the Columns shelf and **Latitude** to the Rows shelf.

**It's always important to question the data. Never assume that the data is correct or imports correctly.**

In this dataset there were a few addresses that did not geocode correctly and this is immediately apparent when mapping the points. There are 2 addresses in the dataset that did not geocode correctly using one of the geocoding tools and they ended up with a latitude and longitude in South America instead of Cincinnati, Ohio where they should be.

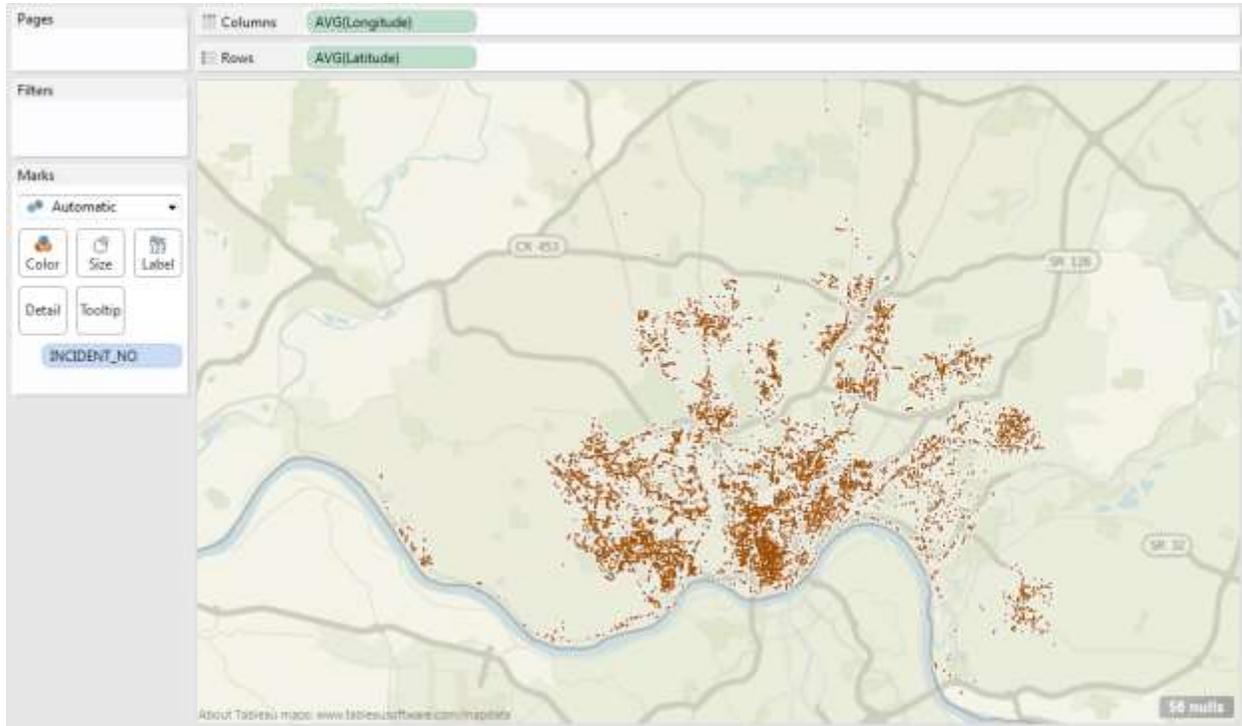
Examples:

2700 CENTRAL PY and 2020 CENTRAL PY

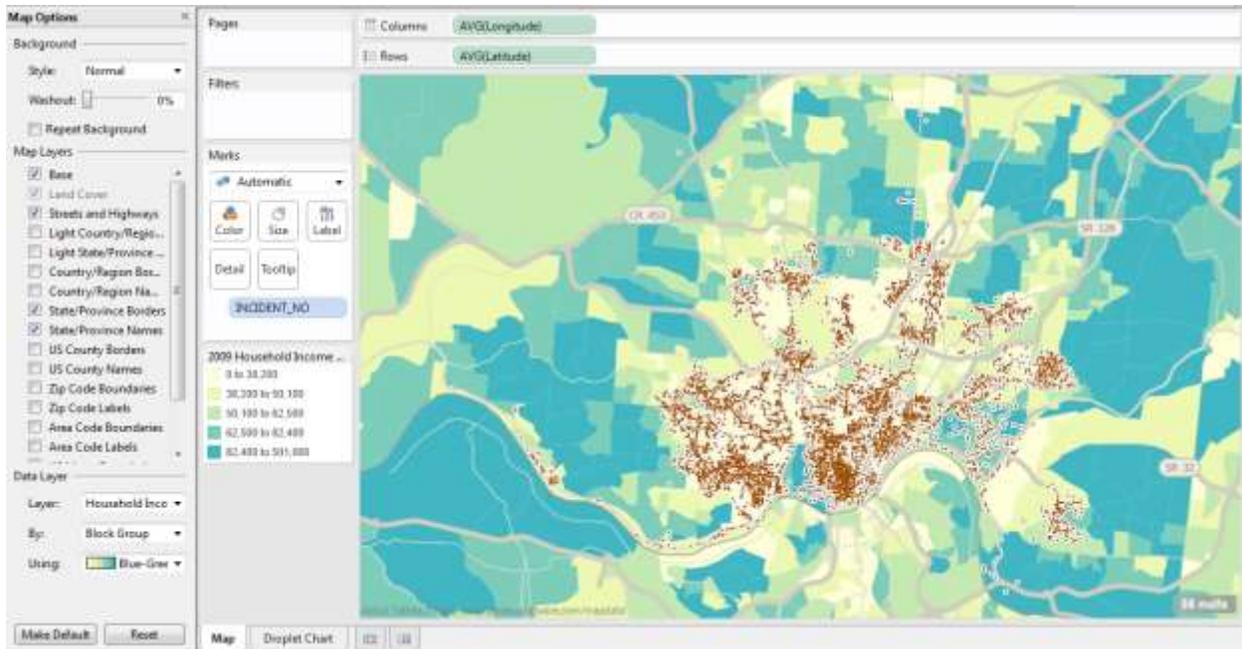
Using PARKWAY or PKWY worked in the original tool, but the other tools accepted the PY in the address for Parkway.

One option is to drop these records from the dataset or filter them out in Tableau, but these are simple to correct and update. After correcting these 2 addresses in the original Excel file it is necessary to refresh the dataset in Tableau (Data -> Refresh All Extracts). Once this is done then all of the points correctly show in the Cincinnati area (there were also 56 NULL values which will be excluded later).

After changing the size of points to make them smaller, change the color of the points to orange (instead of the default blue) and adjust some of the features of the base map, the map should look something like this.

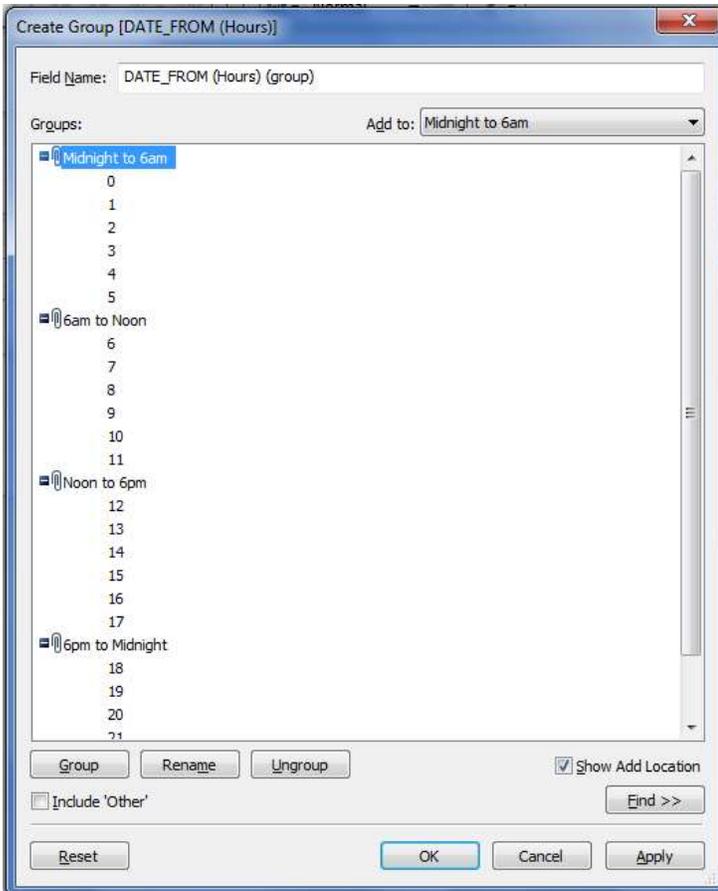


By adding a data layer that is built into Tableau we can add another dimension to the analysis. In this case add **Household Income** from the Data Layer dropdown menu and select by **Block Group**. The default color scheme will work well in this case. Now the map will look something like this.



The next to do is to create some custom groupings. The crimes in this data set occur at all hours of the evenings. For analysis purposes it will be necessary to group them into some sort of subgrouping and when examining crime data it's fairly common to use an increment between 4 to 8 hour (see *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis* by Collen McCue, page 94 on Time Groupings for more information).

Tableau has a custom calculation that can be used for custom time groupings. First right-click on the date field, in this case "Date\_From", and select **Create Custom Date**. This will bring up a dialog box to choose options. The goal in this case is to create a custom field on hour, so the options are **Hours** under the detail dropdown menu and **data part**. The next step is to group them by right-clicking the newly created field, in this case "Date\_From (Hours)", and selecting **Create Group**. Select each group of hours that need to be grouped and group them together. The final result will look something like this.



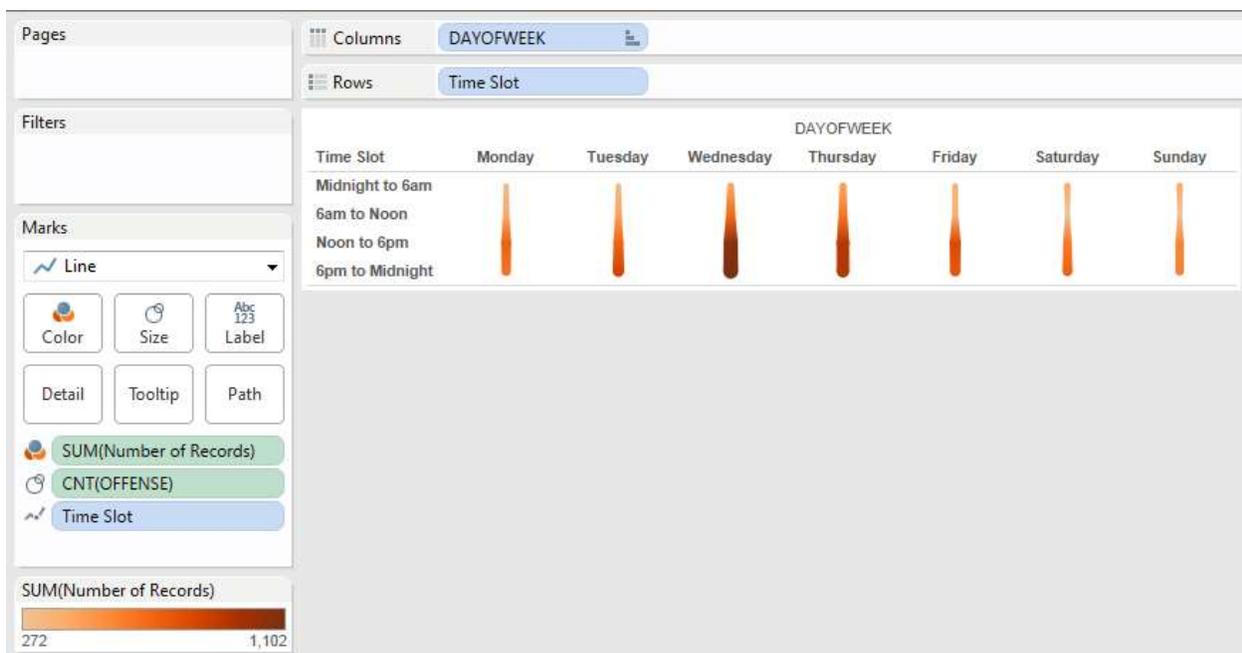
After creating the new grouping rename it to something appropriate and easy to understand, for example “Time Slots”. This will be a field that is crucial for the visualizations that follow.

#### Step 4 – Create some visualizations

Before creating any of the visualizations, put the data in a simple crosstab format. From the Dimensions pane, drag **DAYOFWEEK** to the Column Shelf, **Time Slot** to the Row Shelf and from the Measures pane, drag **Number of Records** to the Text shelf (or to the body of the cross tab). This will create the following crosstab with the total number of crimes by day of week and by hour in a layout that is similar to a calendar format (in this case moving Sunday to the end of the week for the weekend).

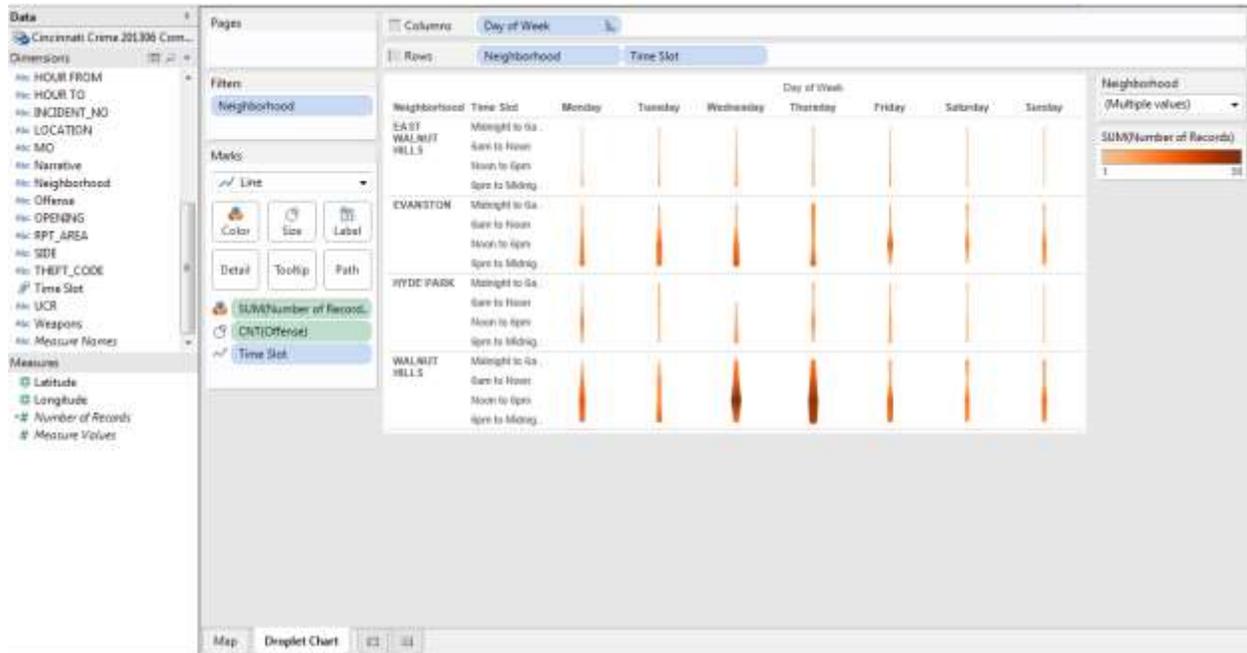
		DAYOFWEEK						
		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Time Slot	Midnight to 6am	310	320	435	448	359	411	405
6am to Noon	420	445	655	611	381	278	272	
Noon to 6pm	704	693	1,051	919	803	608	571	
6pm to Midnight	643	822	1,102	920	743	698	585	

Creating the “Droplet Chart” is a little tricky at first, but it’s easy once you get the hang of it. First, chose “Line” in the dropdown options on the Marks card. Next, from the Dimensions pane, drag **Time Slot** to the Path option on the marks card. From the Marks card, drag the **Sum(Number of Records)** up to Color. Edit the color and choose the orange palette. Next, from the Dimensions pane bring **Offense** to the Size shelf on the Marks card. The goal here is to count the offense and since **Offense** is a dimension click on **Offense** in the Marks card and select **Measure** and **Count**. After following these steps the droplet chart should look something like this.



The next step is to add **Neighborhood** to the Filters and then click on **Neighborhood** on the Filters card and select **Show Quick Filter**. On the quick filter select **Multiple Values (Dropdown)**. For this analysis the selected neighborhoods are “Hyde Park”, “East Walnut Hills”, “Walnut Hills” and “Evanston”. Once these are selected in the dropdown box then add **Neighborhood** to the Rows shelf in front of **Time Slot**.

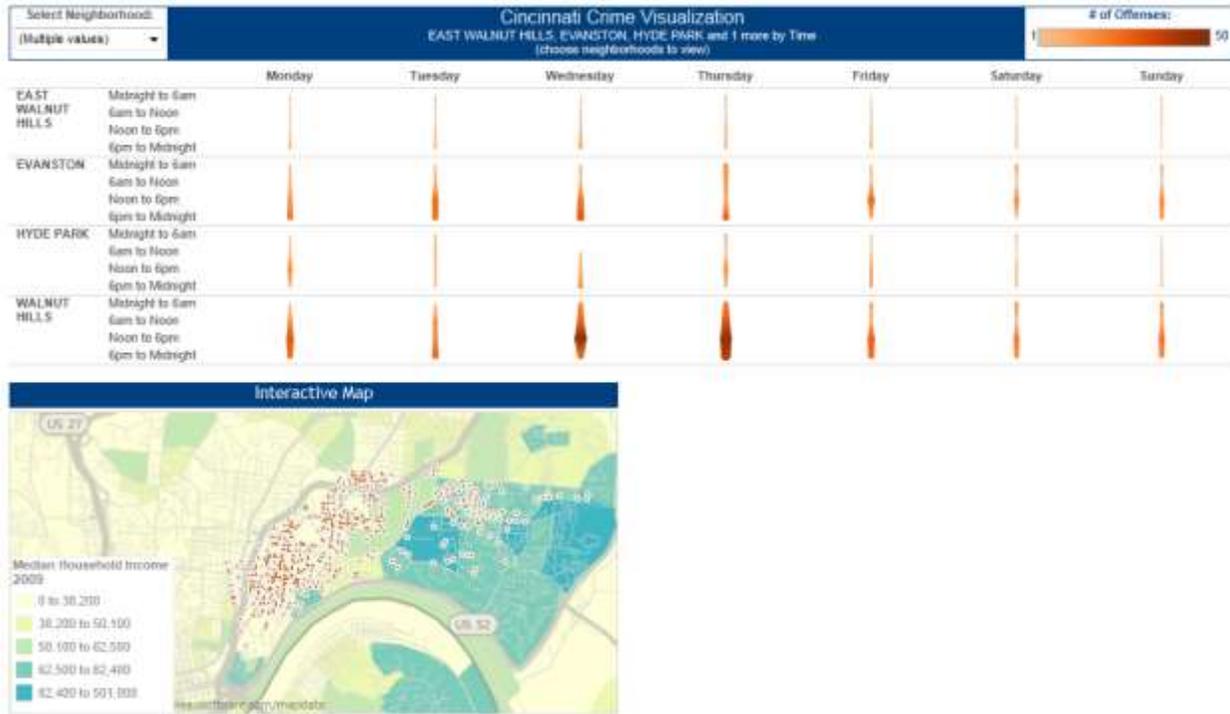
After renaming a few things and moving the legend around, the visual will look something like this:



The next step is to create a Dashboard. After creating a Dashboard page, drag the Droplet Chart on the dashboard. Format the title by double clicking on it. The title in this visualization uses a dynamic title based on the Neighborhood selected in the quick filter. The title is formatted in the following manner.

Cincinnati Crime Visualization  
<Neighborhood> by Time  
(choose neighborhoods to view)

By using floating tiles the dropdown quick filter and the legend can be placed next to the title. Also, by using a floating tile on the map legend it can be placed directly on the map for efficient use of dashboard space. The dashboard now looks like this (note – there is a Blank placeholder here for the next part).



The next step is to embed a Google Map which will be linked via a URL Action. The first step is to drag a Web Page onto the blank area of the Dashboard (in this case the bottom right hand corner). A dialog box will appear asking to set the URL. You will want to use the URL that you want as the default. In this case I want to Google map to mimic the Tableau map in location and size. The following URL will embed that map.

<http://maps.google.com/maps?q=Cincinnati,+OH&hl=en&ll=39.1281,-84.4766&z=12&iwloc=near&output=embed>

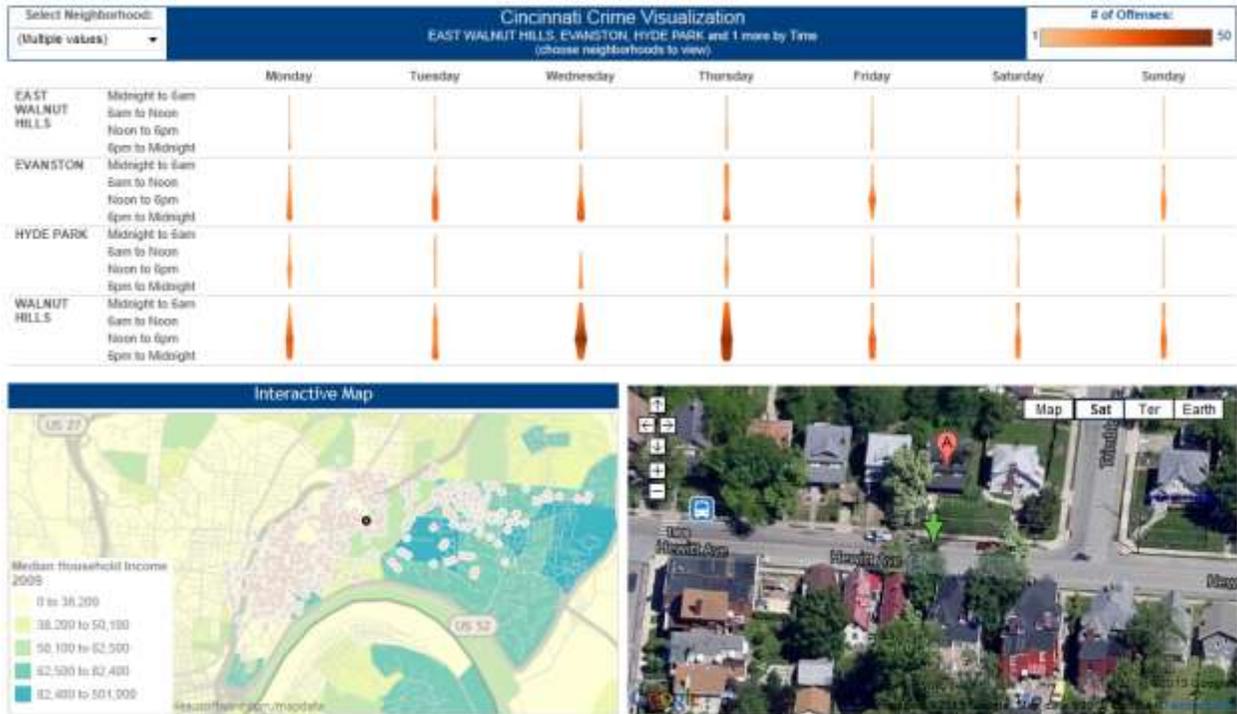
The term &iwloc=near is used to remove the bubble pop up in the Google window and z=12 is the zoom level which is set to match the Tableau map. The parameter &ll is the latitude and the longitude which is set at the center of the Tableau map, again, this is so the default location and size will look like the Tableau map.

On the Interactive Map worksheet, from the Measures pane, drag Latitude and Longitude to the Details shelf. They will appear as **AVG(Latitude)** and **AVG(Longitude)**. Right-click each one of them and select Dimension to convert them to dimensions. After this is complete they should show up in the Marks pane as **Latitude** and **Longitude** and these will be used as a parameter in the Google map.

The next step is to create a URL Action. Click **Dashboards** on the top menu and select **Actions**. Click **Add Actions** and choose **URL**. Select the Droplet Chart and the Map from the list of Tableau Sheets. Click the **Select** button on the right hand side. Now enter the following website address into line marked URL:

<http://maps.google.com/maps?q=@<Latitude>,<Longitude>&ll=<Latitude>,<Longitude>&spn=.0005,.0005&t=h&hl=en&output=embed>

This link will embed an overhead view from Google maps for the point selected on the Tableau map. The Dashboard should look like this and when a point is selected on the Tableau map the Google map should immediately update to an overhead view of the selected address. The visualization should look like this:



The next step is to make the map selection also filter the droplet chart. This is very easy in Tableau. Simply selecting the dropdown from the top right corner window of the Interactive Map window and selecting **Use as Filter** and the droplet chart will immediately filter. However, when doing this the calendar days will be messed up, only showing the one individual point in the first column, regardless of the day of the week and this also affects the other formatting. For example:



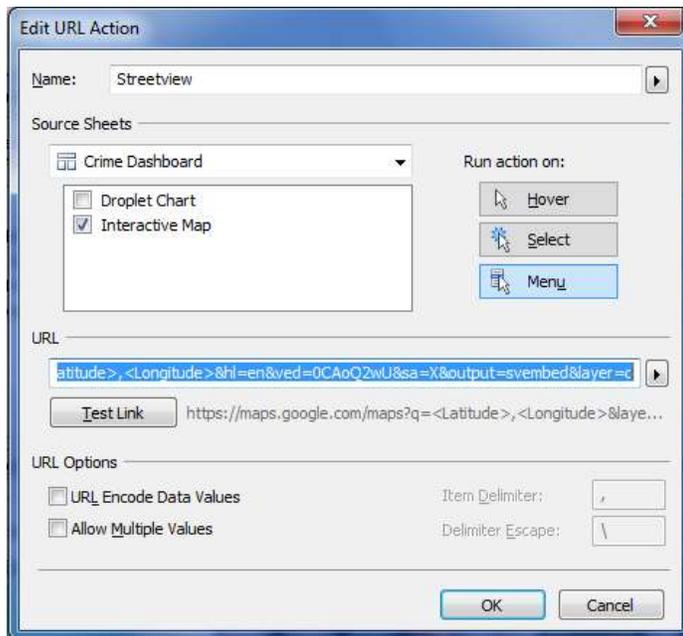
Don't despair, Tableau has unlimited undo available, so just click the undo button twice, once to remove the filter on the point and once more to remove the filter. Instead of filtering we will use the Highlighting feature and this will solve our problem and provide more features. To do this select the Interactive Map sheet and from the Dimensions pane, drag the **Day of the Week**, **Time Slot** and **Neighborhood** to the Details shelf. Return to the Dashboard. Select **Dashboards** on the top menu and select **Actions**. Click **Add Actions** and this time choose **Highlight**. Select **Selected Fields** under Target Highlighting and check the three dimensions that were just added to the Details shelf on the Interactive Map and click OK.



Embedding Streetview will add yet another feature to the visualization. Currently Tableau does not have the capability to dynamically embed two different URL's on the same Dashboard. In this case there isn't much room left anyway, so we will add this feature to the tooltip of the Interactive Map.

Click **Dashboards** on the top menu and select **Actions**. Click **Add Actions** and choose **URL**. Enter "Streetview" as the Name. Select the Interactive Map from the list of Tableau Sheets. Click the **Menu** button on the right hand side. Now enter the following website address into line marked URL:

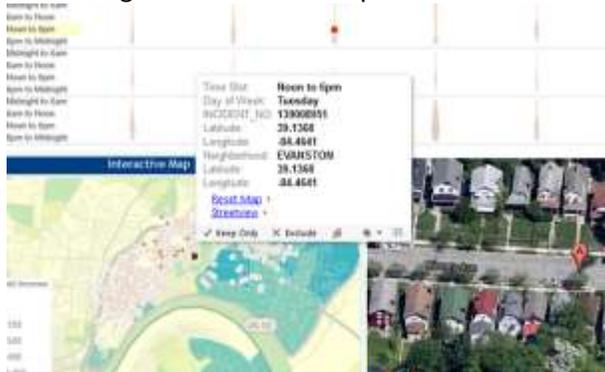
<https://maps.google.com/maps?q=<Latitude>,<Longitude>&layer=c&z=17&sl=<Latitude>,<Longitude>&cbp=13,276.3,0,0,0&cbll=<Latitude>,<Longitude>&hl=en&ved=0CAoQ2wU&sa=X&output=svembed&layer=c>



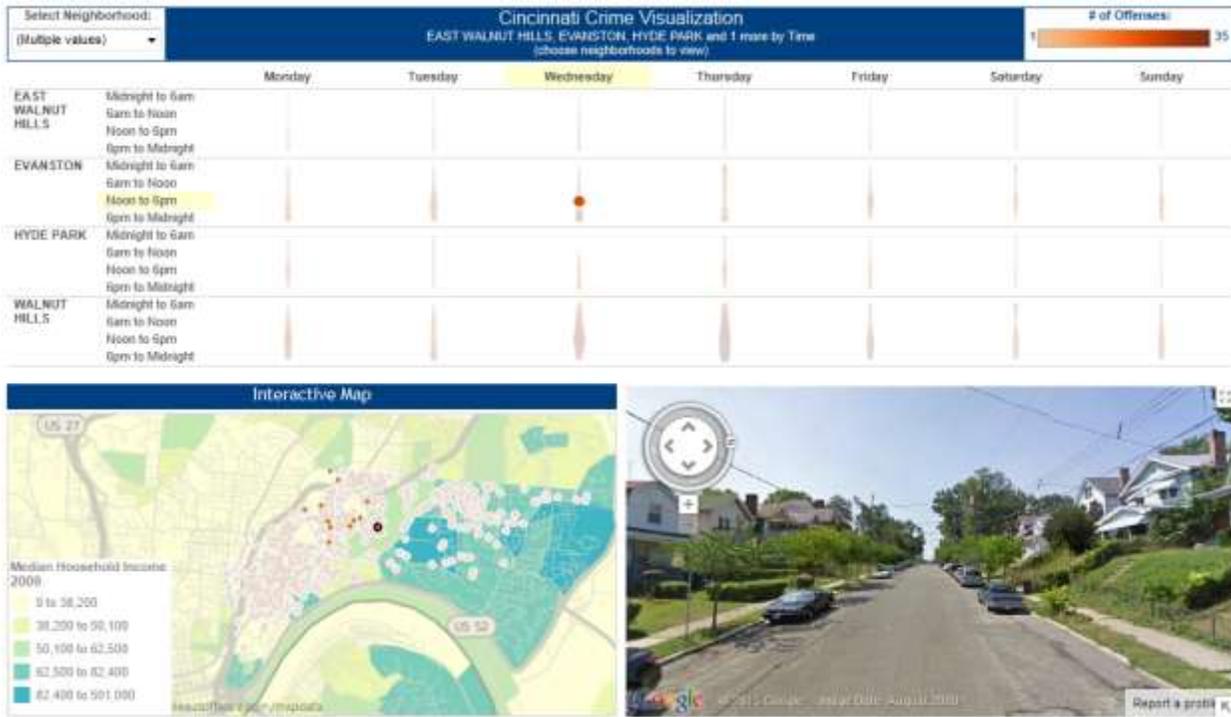
Click OK. Click **Add Actions** again and choose **URL**. Enter "Reset Map" as the Name. Select the Interactive Map from the list of Tableau Sheets. Click the **Menu** button on the right hand side. Now enter the following website address into line marked URL:

<http://maps.google.com/maps?q=Cincinnati,+OH&hl=en&ll=39.1281,-84.4766&z=12&iwloc=near&output=embed>

Click OK and OK again. Now the Tooltip on the Interactive Map should look like this:



Clicking “Reset Map” will reset the Google Map back to the original default and clicking “Streetview” will now change the embedded Google map to a Streetview version that gives the user complete control to navigate around the address.



There is a parameter `&iwloc=A` that was removed from the Streetview URL which embeds a popup bubble on the map with an address. `cbp=` is used in the URL for the Street View window and accepts 5 parameters.

Example: `cbp=11,180,0,0,5`

1. Street View/map arrangement
  - a. 11=upper half Street View and lower half map
  - b. 12=mostly Street View with corner map
2. Rotation angle/bearing (in degrees: 0-360)
3. Tilt angle, -90 (straight up) to 90 (straight down)
4. Zoom level (0-2)
5. Pitch (in degrees) -90 (straight up) to 90 (straight down), default 5

After some formatting adjustments to the tooltips, the dashboard looks like this.

