Data exploration and visualisation

—— Criminal record in Western-Australia

Jianhua Tao

Table of Contents

Problem description and motivation 3

Data collection and wrangling 3

Data exploration (by tableau) 5

Interactive narrative visualization 6

The package used: 11

Shiny 11

Googlevis 11

Leaflet 11

Plotly 11

Readxl 11

Dplyr 11

Description of implementation 11

Exploration (by R) and Instruction: 11

Conclusion 14

References: 15

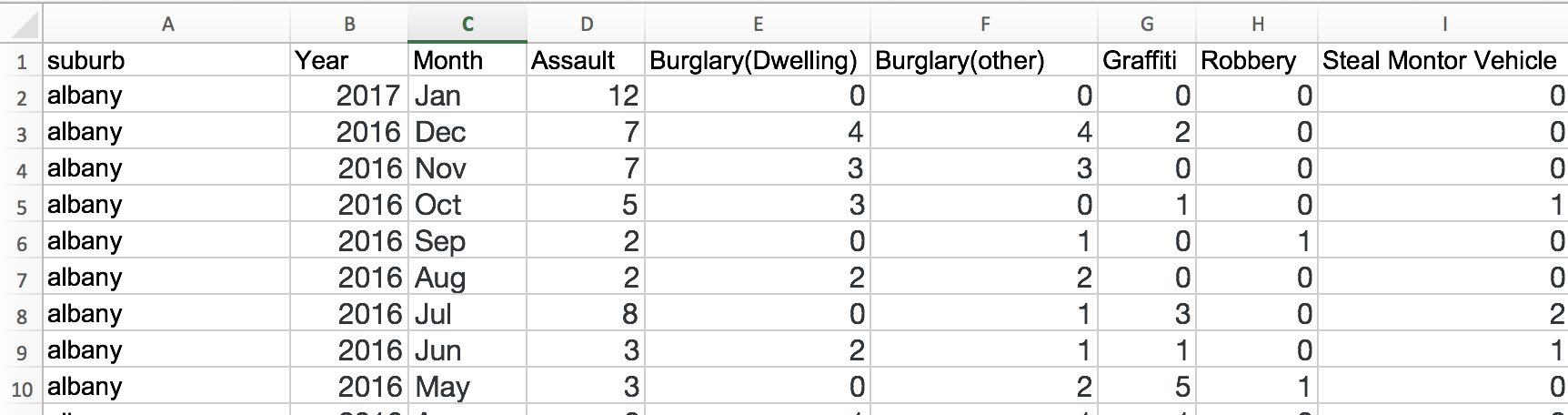
## Problem description and motivation

Criminal is always the issue to prevent residents from living with peace and harmony. Although it is inevitable to eliminate this phenomenon from this planet, controlling and suppressing of these issues is vitally important to maximize the profit for individuals. There are many ways to satisfy such purpose which can various on pre-action (governance, surveillance, education, etc.) and post-actions (penalty, trial, etc.).

Therefore, the motivation of this report is to provide the image on how criminal occur on western Australia in different countryside over last 16 years, as well as the relationship between income, population and the criminal records. So that the analysis and some prediction can be made (based on the given information) to guide the governor to take corresponding actions. Meanwhile, the contribution of this information not only can benefits government to make decision of what amount of police should be distributed in these communities, but also helpful for traveler who want to know more about criminal circumstance in western Australia.

## Data collection and wrangling

To achieve the goal described above, data is pivotal and has been selected from website named Western Australia Police [1]. One of the challenge for picking the raw data is that there is no exist file for the tremendous criminal records, so all the data are copy-paste manually to the Excel file which is time-consuming in terms of the huge amount of records. Fortunately, all the data has been well-formatted in the website so it just required a minor modify (such as font-type and font-size, etc.) when store data in the new Excel file at the vary initial stage. Feature 1 is the data format in this step.



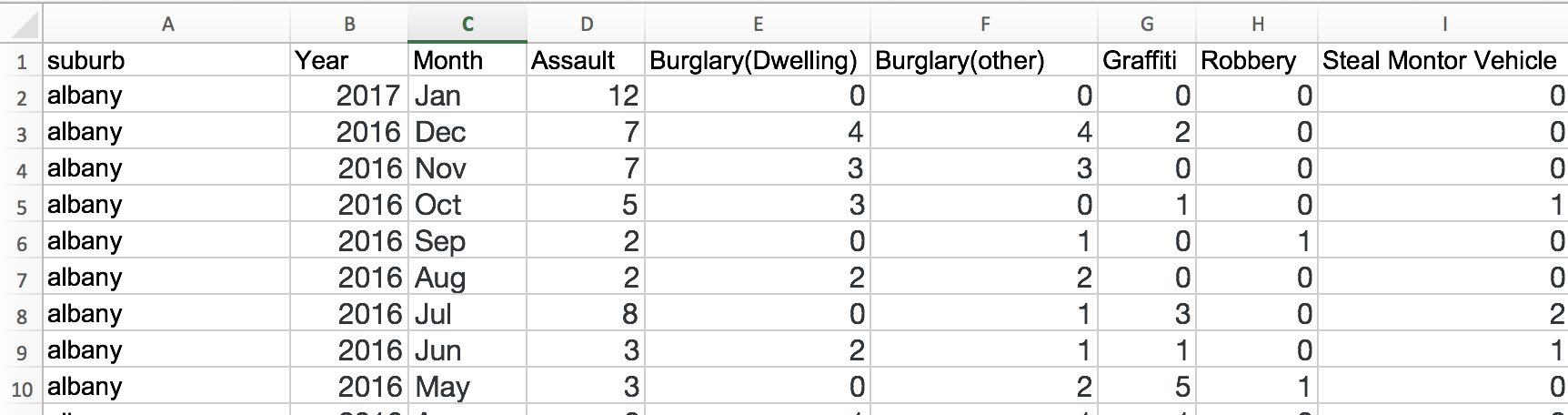
Feature 1

*Notes: All the data stored in file “Assignment data source2.xlsx”, which contains 22907 pieces of record.*

After the first stage, data has been imported into Tableau when the data is ready for utilization. And thanks to the help of Tableau, the basic visualization of the data was provided which not only render several interesting aspects of the data, but imply the issues of the data as well. For example, there is no population and income within this data sets, unnecessary to process data in period of month because the time in year is enough to illustrate the crime scenario, etc. These issues become the engine of requiring more process and exploration of data. Australia Bureau of Statistics [2] then was found under this situation, which is the website to supply the population and income of these countryside. Besides, data wrangling has also been executed during that time, which include:

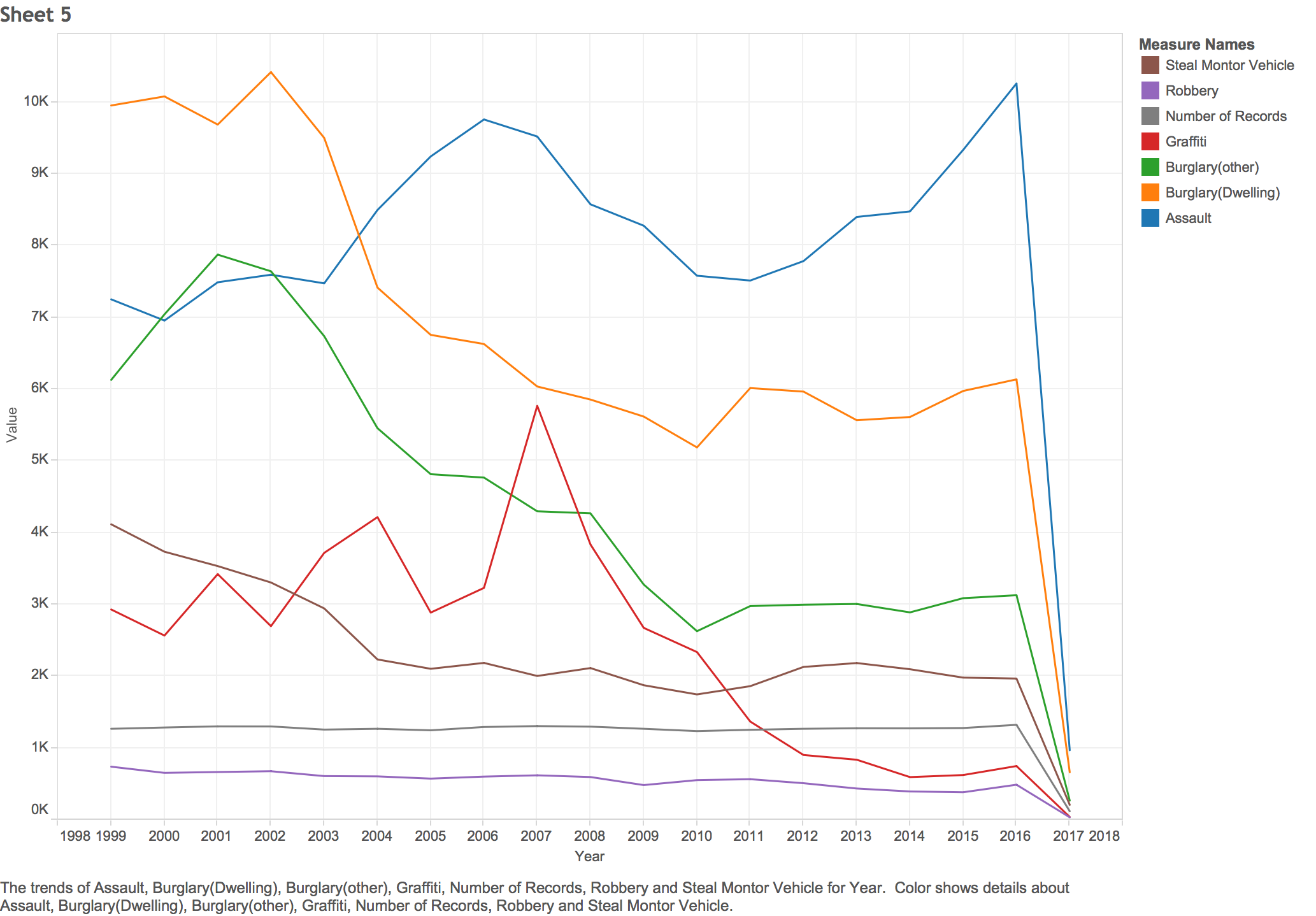
* aggregating all the criminal records based on the countryside within one year by utilizing function called “aggregate” in Rstudio;
* using method “geocode” in Rstudio to found out the latitude and longitude of all countryside in western Australia (preparation for drawing map);
* aggregating the records only based on period of year and then reckoning the composition of different criminal type on different years (data stored in file” total.xlsx” and” pie.xlsx” separately for draw line chart and pie chart).

Moreover, more data regarding population and income were retrieved and selected from Australia Bureau of Statistics, which is in specific year 2006 and 2011. Therefore, after aggregating the data by year, the data of population and income has been added to the file called “dataByYear.xlsx”. Here is the data format in this file:



Feature 2

On the other hand, the record of crime data in 2017 has been removed because the collection of that record has been done only until March, which is not in the same level as the data in other years. See feature3 below:

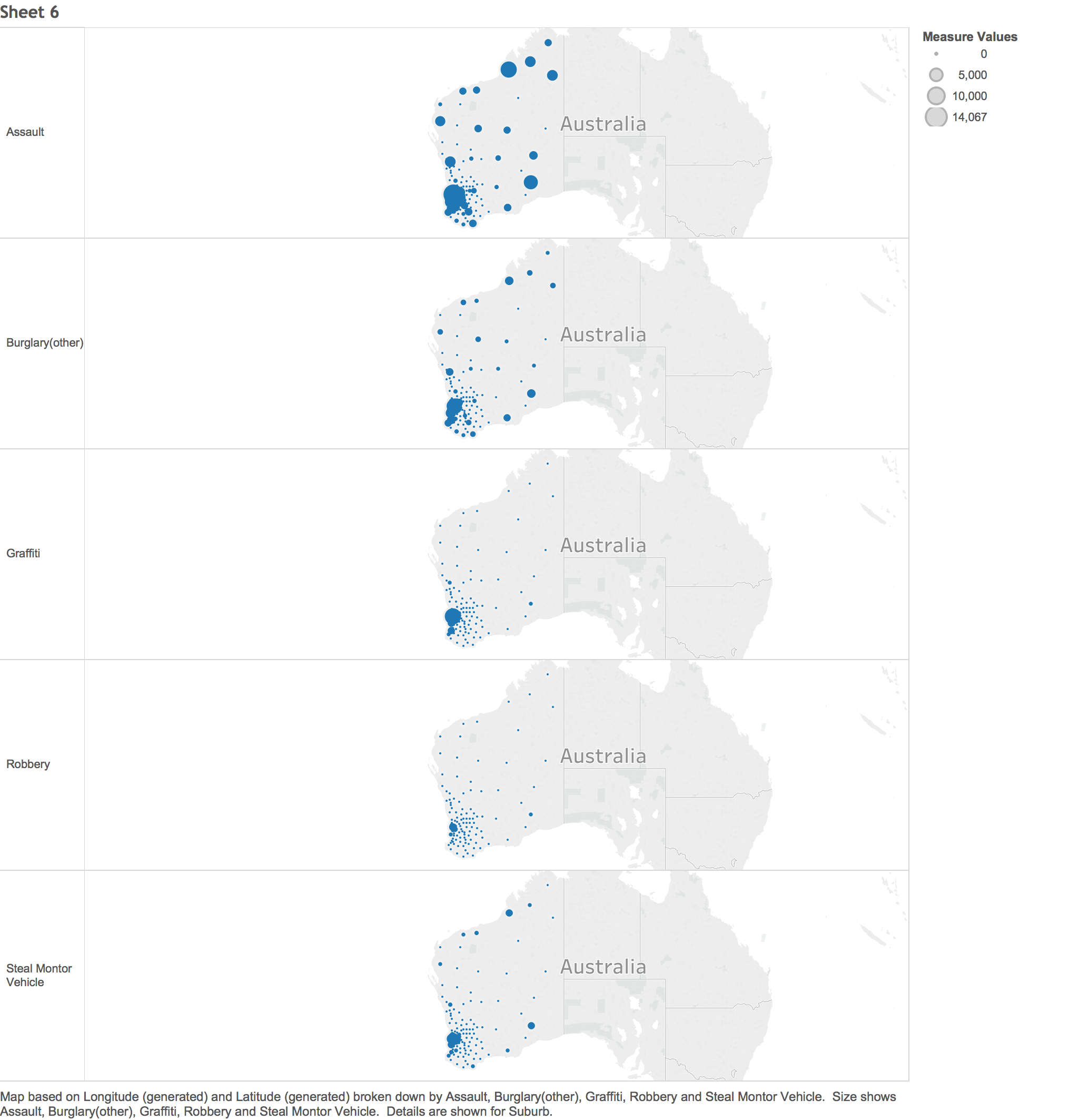


Feature 3

*Note: Feature 3 also show the problem of insufficient data, such as the absent of population and income.*

## Data exploration (by tableau)

To find out the interesting aspects of the data, Tableau has been used which provide very general ideal of how the crime distributed in Western Australia with the corresponding amount. As we see the feature 4 below, the crime occurs the most is obviously the assault, along with the lowest crime Robbery. Furthermore, the area with the most frequent crime always placed on the south-western Australia, which need more exploration to find out the causes.

****

Feature 4

More content will be illustrated later by using R.

### 

### The package used(R):

Shiny This package helps building interactive web applications. Automatic "reactive"

binding between inputs and outputs and extensive prebuilt widgets make it

possible to build beautiful, responsive, and powerful applications with minimal

effort.

**Server-side:** In charge of providing the function, such as controlling the

popup that show in the map, controlling how data will be show in the user

interface.

**Client-side:** Implement the layout and graph output.

**Global-file:** load the package and data for global utilization.

GooglevisThe package in charge of the charts showing in the screen. Such as the LineChart,

BubbleChar and PieChart.

LeafletThe package provides the function for plotting the map.

PlotlyUtilization for drawing BarChart.

ReadxlReading data from the Excel file.

DplyrProviding function called “filter’, which can select sub-dataset by specific

conditions so it is possible that user can interact with the web application.

### Description of implementation

After filtering all the information by processing 5-design sheet methodology, the final design has been finalized, which including the implementation of different charts and map. The reason for taking this design as final decision is that the combination of different charts can tell more story than show them individually. Moreover, the interactive is also vital because user may want to interact with the charts so they will engage more on the visualization rather than distracted by other stuff.

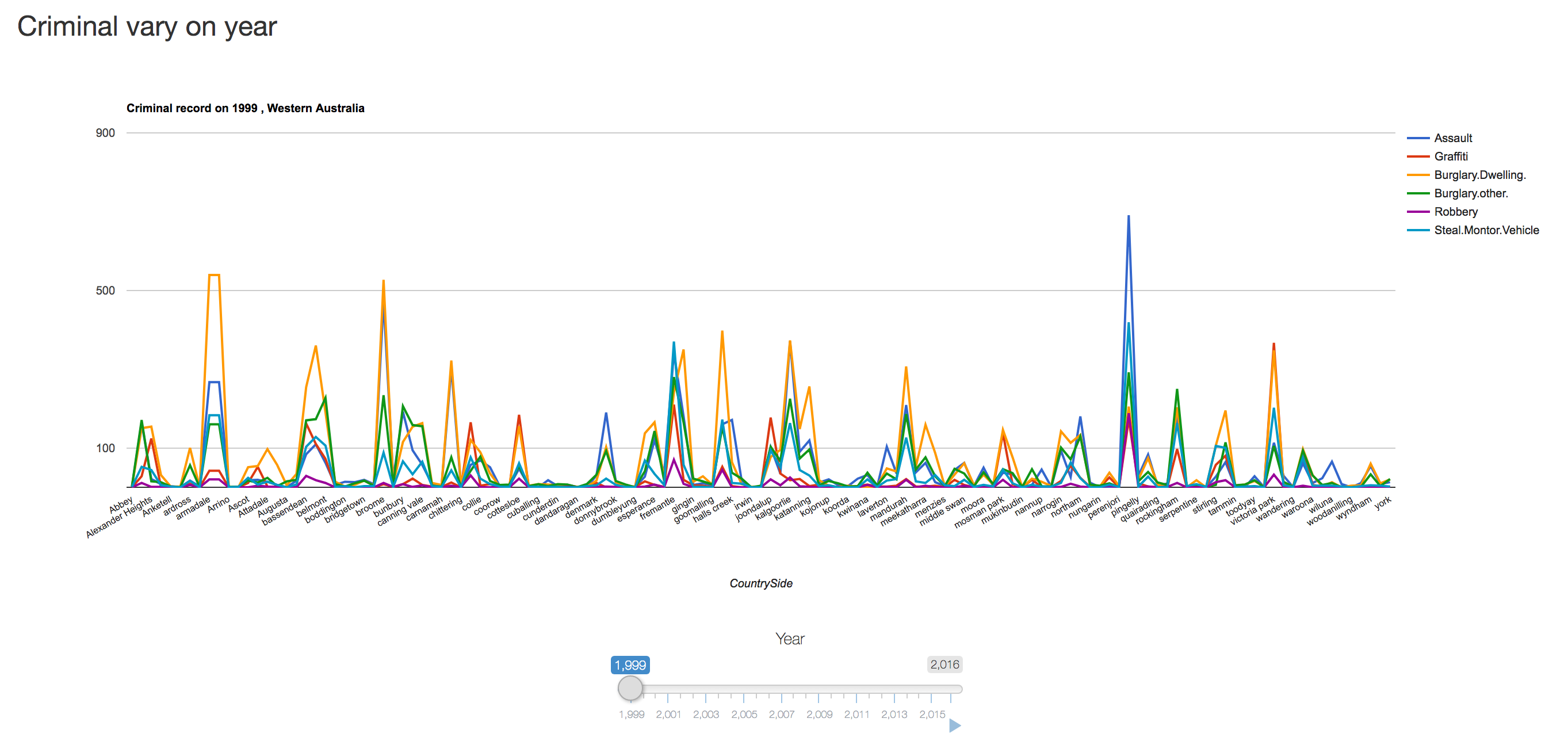
By using the package mentioned above, the output in the server side has been deployed. Almost all the graph in the screen have support from the server side such as the LineChart, BubbleChart, and BarChart. To be more specific, the graph was prepared in the server side by some relevant code and then was plotted in the user interface. When dynamic view needed, function observe and input has been called to meet these requirements.

Thanks to the leaflet, which makes drawing map in shiny possible. The same mechanism as what described above, which is about rendering leaflet in the server side and plotting in user interface; Furthermore, the combination of **CSS** and **shiny** is vitally important to decorate the map, which also provide some practical function such as the draggable panel with some relevant content.

## Exploration (by R) and Instruction:

Overall viewpoint of the dataset is about the crime that occur in different place during different years from 1999 to 2016. Because of the visualization, it is more direct to see what kind of crime happen the most during which year in which suburb (Feature 4 is the crime record visualization in 1999). After viewing all the crimes that happen between 1999 and 2016, we can find that the crime happens the most are Assault and Graffiti, which concentrate in Perth, Broome, and Derby, etc. The amount of Robbery and Burglary stay relevant low level during this 17 years. One other interesting thing is Steal of Motor occur the most in 2002 and then decreased.

Instruction for this page: selecting the year by draw the circle in the year bar on the bottom of the page. Or click the “play” button which will trigger the graph that run by itself.



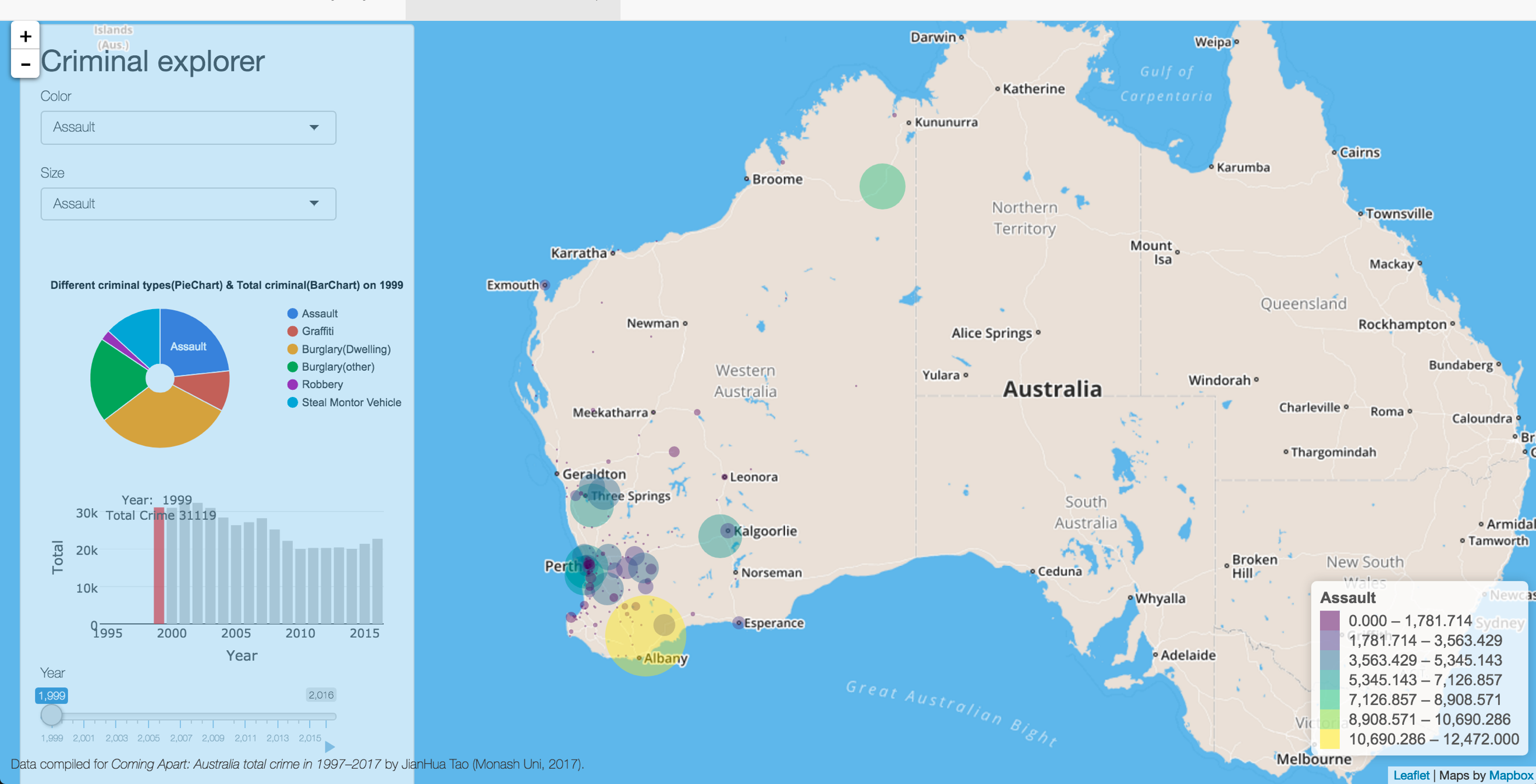
Feature 4

When comes to the geographic view, it is obvious that most of the criminal happen in south western of Western Australia, around Perth to be more specific. And the situation is similar in terms of all the different criminal type. Besides, one significant point on this stage is that the total amount of crime reach the peak in 2001 and start decrease gradually after that year, eventually become stable since 2009.

Instruction for this page:

The color and size can be determined by different criminal type. But to generate the best view of this map, the criminal type is highly recommended to be selected consistently for both size and color, otherwise the size and the color may result in a mess if different criminal types picked.

One more complementary information this page can provide is the total criminal that show in different years, as well as the percentage of different type. So basically, by selecting the year offered in the panel, user can see different aspect of the chart in that page.

****

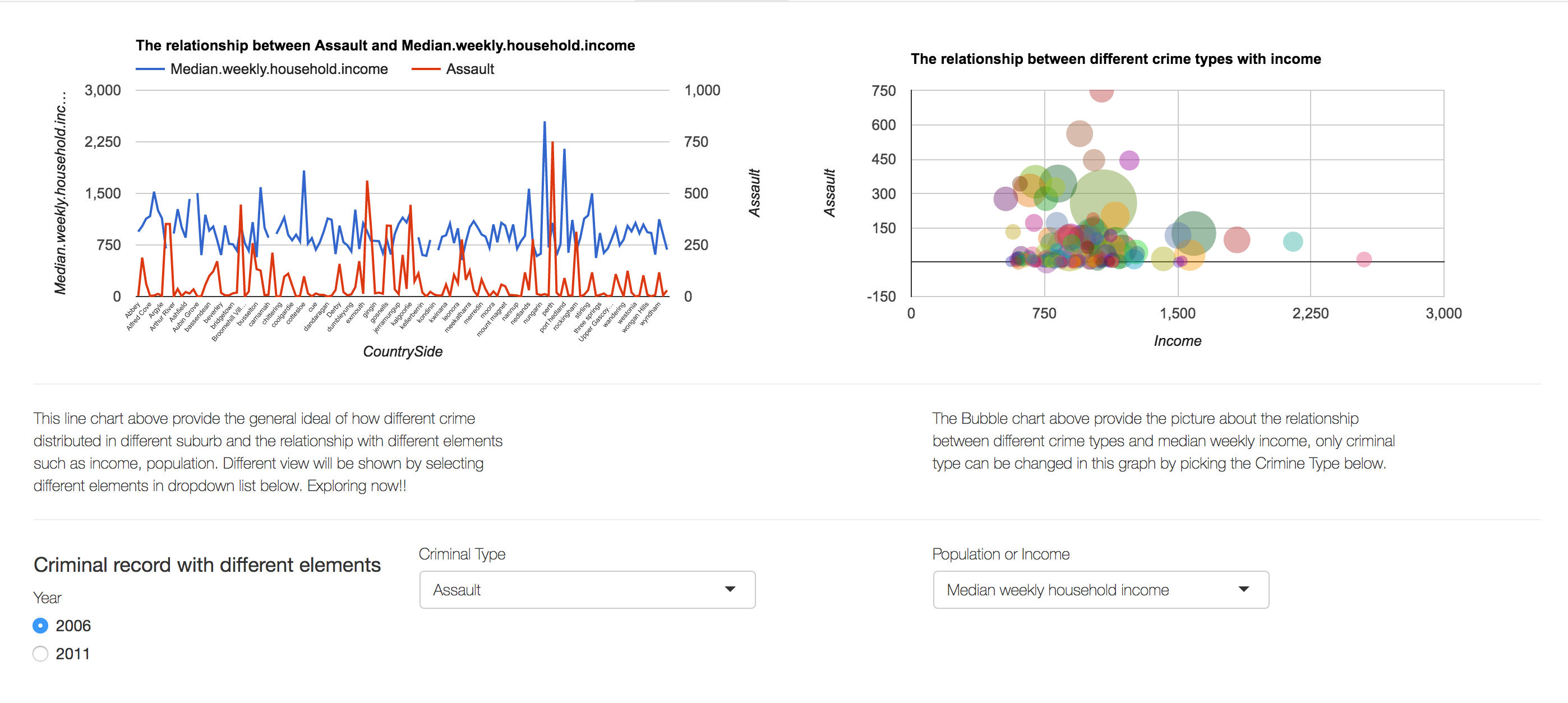
Feature 5

The information provided so far, which can contribute to police or relevant department make sense of how to act on these criminals, for example:

* For the suburb with high frequent occurrence about Steal Motor or Robbery, more policeman should be allocated to that place for patrol, and vice versa. So that human power distribution can be more efficient.
* For suburb that always bothered by Assault, more skillful police who expertise on this type of crime should be distributed on this area, to improve the situation here.
* For countryside that suffer Stealing and Graffiti, more camera surveillance should be placed in this place with sign to frighten criminal.

On the other hand, such information also benefits the rating of this place if the government want judge how is that specific suburb performs.

Nevertheless, how phenomenon of criminal distribution formed has become another topic of this data exploration and visualization. Thus, Median Resident Weekly Income and Population has been selected to generate the basic viewpoint of this issues. See Feature 6 below:



Feature 6

Basically, the two graph describe the similar issue but focus on different parts.

The graph in the left show the relationship between the different criminal types and the elements such as income and population. It is more about the relation. The graph in right yet provide the trend of how things change in different year. Since the graph designed for interactive, so it is possible to manipulate with these elements by select corresponding menu.

Instruction for this page: Year can be selected by given choice; Criminal type can change the shape of both charts but the selection of Population or Income will only result in changing on the LineChart, because the horizontal axis in the bubble is already fixed.

## Conclusion

This data visualization based on criminal records can play a significant role in contemporary world, since people are concern more about their well-being and happiness nowadays. By visualizing the criminal record over last 16 years, people can make a general ideal of how crime occur on where, which help individuals know more about their country especially for people who live in Western-Australia. More realistically, the given visualization can also benefit some relevant departments who can act on guarding residents by taking proof actions, such as placing camera on the place that criminal occurring frequently, dispatching policeman for patrol, etc. Nevertheless, it is undeniable that the limitations of this work also exist. Because of the data missing somewhere on the source website, which makes the graph not consistent and completed.

## References:

[1] Crime Statistics Portal

*<https://www.police.wa.gov.au/Crime/Crime-Statistics-Portal/Statistics>*

[2] Australian Bureau of Statistic

*http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Census?OpenDocument&ref=topBar*

[3] Kim Marriot, *Data exploration and visualization*, Retrieved from ‘http://moodle.vle.monash.edu/pluginfile.php/5383405/mod\_resource/content/1/FDS.pdf’