FES Industry of Occupation Between 2012 and 2017

Ye Cai
The University of Sydney
Sydney, Australia
ycai6198@uni.sydney.edu.au

I. The Nature of the Data

A. Data Collection

The city of Sydney undertakes the Floor Space and Employment Survey (FSE) every five years to collect the data on all industries for every buildings in the area. This project started in 2012 and has finished three times of collection, each survey took around 10 months to complete. During the data collection phase, surveyors visited each business in the city of Sydney to record specific data needed in the surveys.

B. Data Attributes

In this report, only two years of data are selected to analyze the changes in the city of Sydney from 2012 to 2017, including six attributes: Longitude, Latitude, Classification Number, Classification Name, Year, and Address.

The first two elements are Longitude and Latitude that represent the precise coordinates of each company. Classification number denotes the code of each type of company. There are dozens of classifications in the data and each type has a unique code to identify. Meanwhile, each code corresponds with the name of different sorts which was stored in the Classification name. The fifth attribute is Year that represents the time data was collected and this data only has two different years, 2012 and 2017. The last one is the Address which records the street name and postcode of companies.

II. The Customers of the Data

Stakeholders like competitors and job seekers especially entrepreneurs who want to start a business or establish an industry in Sydney have the greatest demand for this data. These data could supply more information and support for them through several aspects.

First of all, increasing employment opportunities. For job seekers, they could achieve more information about the needs of employees in different industries by looking at the distribution and number of companies. After comparing the changes between 2012 and 2017, they could also know the fluctuations in the demand for employees timely.

Secondly, demonstrating the trend of business development. The information about the changes in different kinds of companies in different areas between 2012 and 2017 could give huge support for competitors to better understand the trend of the current market and make policy designing, risk prediction, and strategic realignment to increase their competitiveness.

Finally, conducive to decision-making. Entrepreneurs who just start their business normally have less experience and lack information about the situation of the market. Through analyzing the distribution and the changes of different types of companies in Sydney, they could grasp the situations and environments of different industries more easily and make a better choice to reduce the risk and increase the possibility of success.

III. Data Types

Six Attributes	
Data Type	***Attribute Name***
Nominal data	Classification Number
	Classification Name
	Address
Interval data	Longitude
	Latitude
	Year

There are six attributes in the data that convey different information to the stakeholders and can be divided into two types of data: nominal data and interval data.

A. Nominal Data

Nominal data includes classification number, classification name, and address. Nominal data is a type of data used to label variables without any quantitative values like ranking and ordering. It just likes a label, but it can be qualitative and quantitative.

Both Classification name and Address are qualitative that display the name of the companies and locations and none of them have any numerical significance. They could be used as data classification and display as different pure colors because the color hue is best to show the qualitative difference on the chart. For example, as shown in Figure 1, the data are divided by Classification name in a pie chart can better show the distinction between numbers of five kinds of companies. The audience could clearly grasp the cafe and restaurant are far more than other types, accounting for almost half. Similarly, using colors to classify the Address in data could provide better visualization for audiences to observe the distinction between

regions and the circumstances of different companies in each area.

Fig. 1. Example of a pie chart

Despite Classification number is quantitative, but the order is arbitrary and has no calculations and relationships between numbers. Therefore, classification number is also nominal data and can be grouped together into categories and use statistical methods to analyze. In addition, Classification number can replace Classification name as they are corresponding which could improve the visualization of charts when the name is too long. For instance, the bar chart (shown in Figure 2) chooses Classification number as a column instead of using Classification name. Using Classification number in the bar could bring more convenience for customers to view the data as the name cannot be completely shown in the axis.

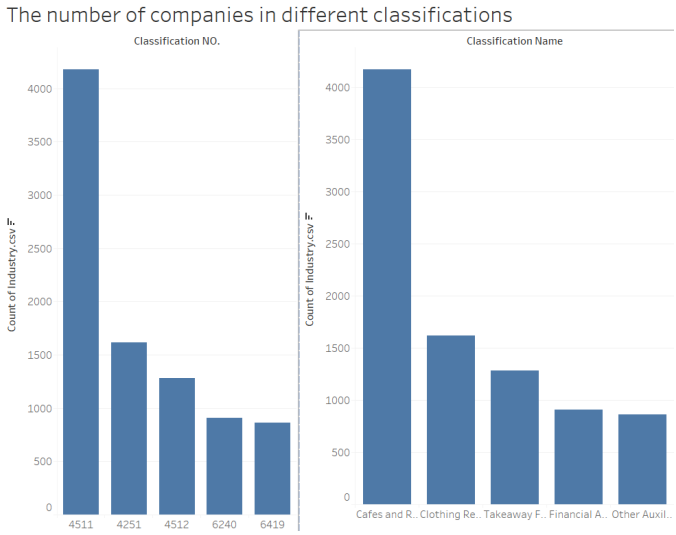

Fig. 2. Example of a bar chart

B. *Interval Data*

Interval data is a type of data that is measured by the scale and it cannot be multiplied or divided. Longitude, Latitude, and Year belong to interval data as all of them are quantitative and the difference between two values is meaningful. Moreover, they do not have an absolute zero and can be added or subtracted.

Longitude and latitude are the precise coordinates at the geographic coordinate system. Building a map is the best way to use these data to display the distribution of companies in Sydney which shown in Figure 3. The map directly shows the density of different areas could give more support for audiences to deduce and analyze.

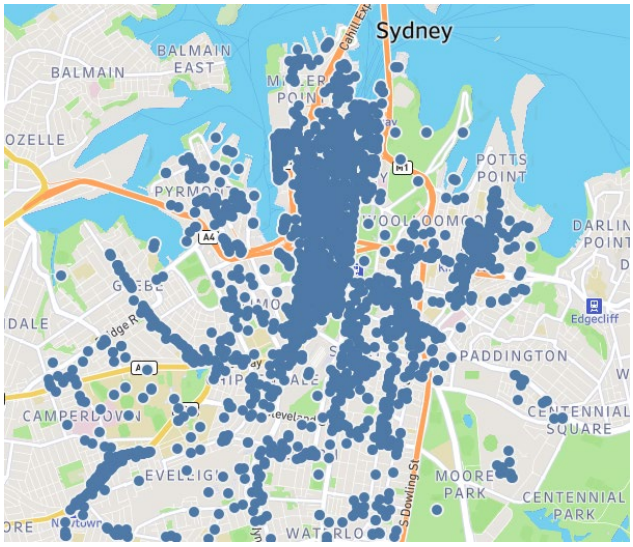

Fig. 3. Example of a map

Year is another interval data that distinguishes the data from 2012 and 2017 and normally used when showing the changes between two years. It can be in several charts like a bar chart, line chart, and circle view. All these charts could convey the changes between years to customers.

IV. Typical Mistakes

As the longitude and latitude in the data, the map is the most possible to be considered when creating a visualization chart but only one dimension can be used in the map. If people want to show more information, creating filled maps with pie charts is a normal choice to demonstrate both the distribution of the companies and the percentage of different types. However, there will be several problems that impact audience judgment and create confusion.

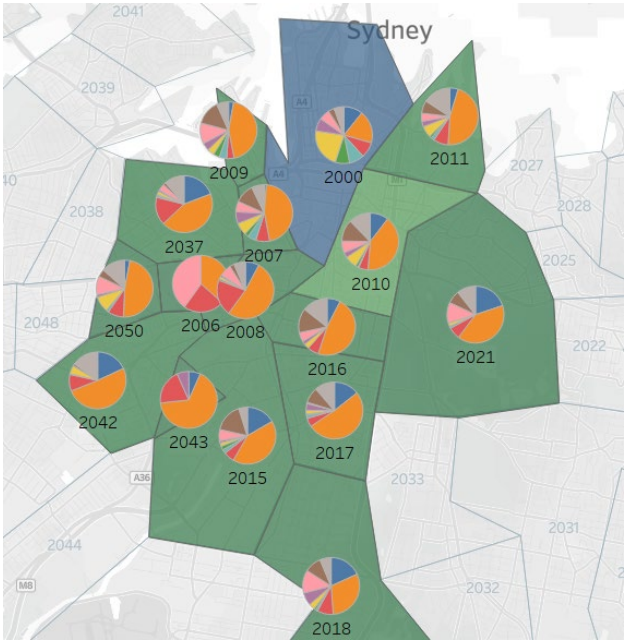

Fig. 4. Example of set-size effect

A. *The Number of objects*

When customers view the data on charts, working memory plays an important role in several behaviors like deducing and decision-making. Alvarez(2004) states that working memory includes two parts: a high-capacity sensory memory and a relatively limited-capacity short-term memory. However, the most regular mistake in visualization is about short-term memory which is the number of objects.

It is correct to show more specific details for customers which could supply more help to make decisions but it would have a negative impact if there are too many objects in one chart. The number of objects cannot more than four as it would impact the total amount of visual information which is calculated by multiply the number of objects and the visual information per object(Alvarez, 2004). If the number of objects exceeds four, customers would be unable to remember all the objects to analyze and would influence the final decision.

B. *Set-size Effects*

The Size-set effect is another common mistake in visualization. Palmer(1990) argued that the correct recognition will be impacted by the set size. It is important that choose the suitable size of each element in the chart. For example, in Figure 4, the pie charts on the map are too small and cannot be enlarged by zooming the map that increases the difficulty for audiences to recognize each part of the angles. Meanwhile, some parts of pie charts overlap it would also create confusion for the audience to analyze the picture. Therefore, set-size is another factor that needs to be noted and avoid during visualization.

V. Data Visualization

A dashboard that includes two diagrams was used to represent these data: symbol map and bar chart (Figure 5).

The first is the symbol map that shows the distribution of companies in the city of Sydney in 2017. Year is a filter in the map to remove data in 2012 and just conserve the data in 2017. After filtering data, Longitude and Latitude act as rows and columns respectively to establish the map. This is the best way to utilize coordinates to illustrate each position of companies. Both Address and Classification Number are nominal data so they are suitable to classify data. Address is firstly used to divide the map into 16 areas which allows audiences to analyze the data by region. Then, a bar chart combines Classification Number and a new calculation field which count the number of industries to show the number of different types of industries in each region. To avoid too many objects or information in one chart, only the top three are displayed in the bar chart. The total numbers of companies in each area were also calculated and demonstrated by color intensity on the map. The map mainly conveys information about the distribution of companies in the city of Sydney in 2017.

The bar chart is the second chart displays the changes in different types of industries between 2012 and 2017. The Year and Classification Number make up the x-axis to show the top 5 types divided into two years. There are dozens of types in the data and it is impossible to show all of them in one chart and it is unnecessary. The five most popular kinds are enough to provide a guide for customers to analyze. The number of industries acts as rows to show the number. This chart provides

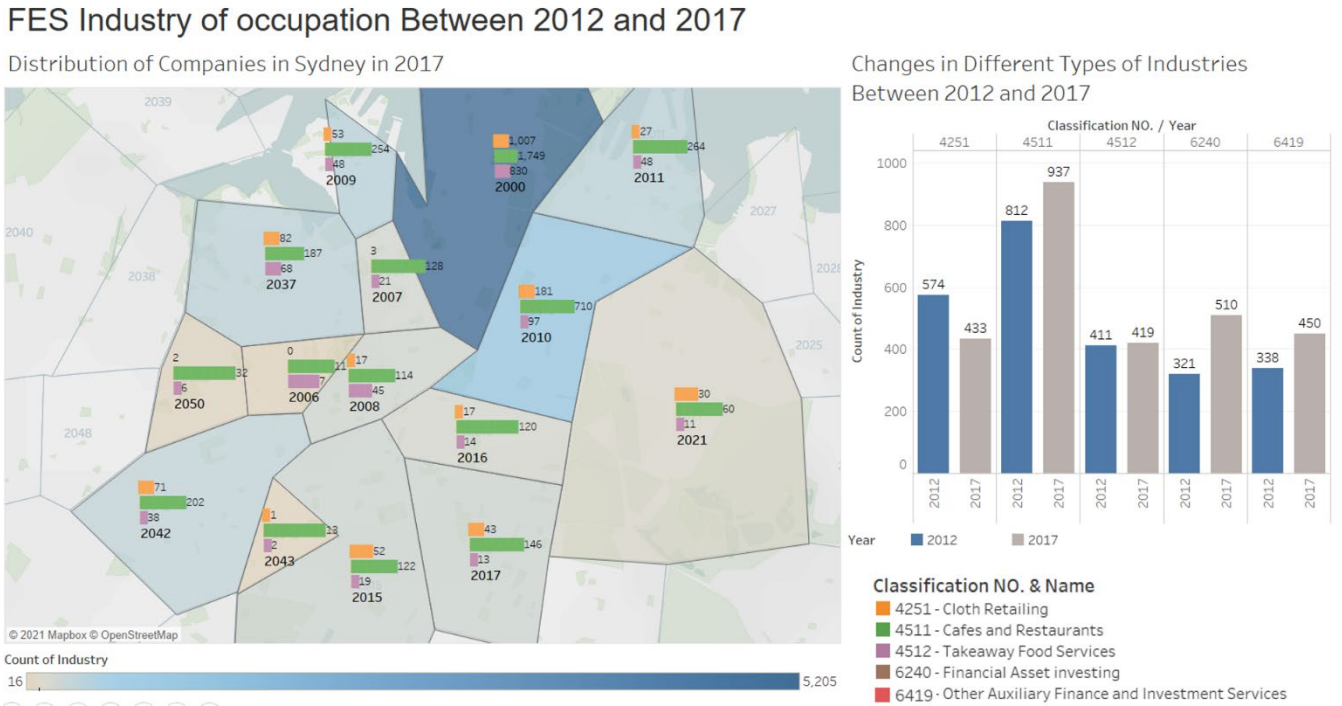

Fig. 5. Data Visualization

help for stakeholders to discover the situation of business development after analyzing the quantity fluctuation of companies.

Both two charts in the dashboard use Classification Number instead of Classification Name as the length of names are unequal and some of them are too long which would affect the display and inconvenience for customers to view. However, there is a list that includes colors, Classification numbers, and names show at the bottom right of the dashboard and customers can find the corresponding name from the list at any time.

VI. Symbolic Representations

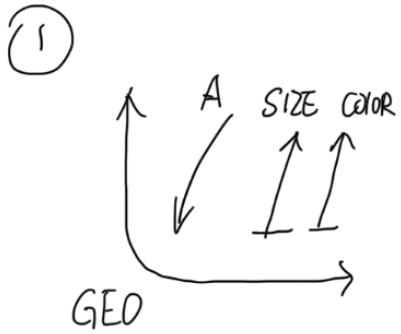

Fig. 6. Symbolic Representation 1

The first symbolic representation corresponds to the map(shown in Figure 6). The visualization chart uses the geographic map and all elements follow the geographic order so the group of imposition is a map. After comparing the principal types of construction, we can finish the first step: choose the right imposition which can be seen in Figure 6. Secondly, the conventional sign should be drawn. On the map, color intensity uses to display the distinct of numbers in different areas so both area and color should add as conventional signs. In addition, bar charts also apply in the chart to show the number of companies in different types. Therefore, the third step is to add a retinal variable: size into the graphic representation.

The second one is consistent with the bar chart, given in Figure 7. The group of imposition is a diagram and the type of imposition is orthogonal as the chart is a bar. In the bar chart, two colors represent two years respectively so the color is one of the conventional signs. Moreover, the values in the row are quantitative as they are the count of industries. In the column, there are two attributes: year and classification number. The data was divided by classification first and then partition into two years. Consequently, the year should be denoted by a dotted line and it is orderable. Classification under the year use a solid line and multiply 5 as there have 5 different classifications.

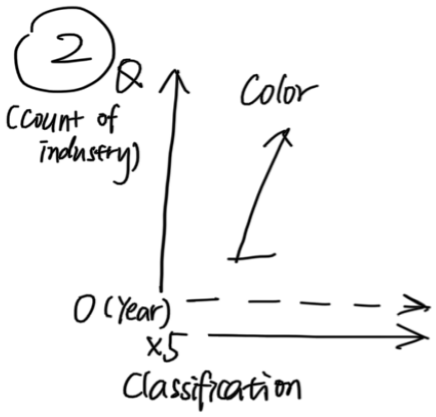

Fig. 7. Symbolic Representation 1

VII. Equivalent Symbolic Representations

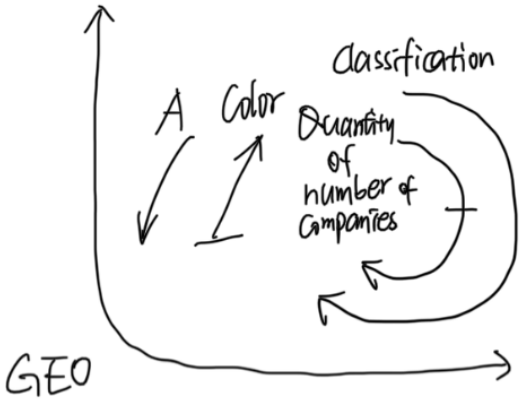

Fig. 8. Symbolic Representation 3

Comparing with the first symbolic representation, all the things stay the same except the retinal variable changes from size to the circular utilization of the plane. The previous representation uses size to display the number of different companies and uses color to divide. The circular utilization can also demonstrate the same information as pies can also show the numbers of companies and divided by colors. Meanwhile, the number of companies in the pie charts is quality so the final symbolic representation is shown in Figure 8.

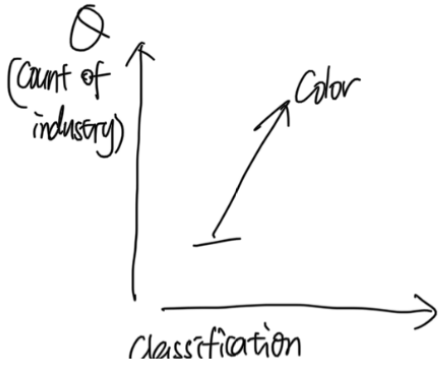

Fig. 9. Symbolic Representation 4

The second equivalent but different symbolic representations also use the bar and the y-axis stays the same. The year on the column was deleted and it would be represented by color so the color should add as a conventional sign. Finally, the dotted line was be removed and just conserve a solid line without multiply 5 as there has only one attribute on the x-axis and the number of images or the figure is 1. (Figure 9).

VIII. Alternative Visualization

The first is the map with pie charts which display the intensive of different companies in Sydney(Figure 10). Similar to the previous map, the map use filter to delete the data in 2012 first and then use Longitude and Latitude to create a map. The color intensity shows the numbers of industries in areas. In this chart, pie charts were used in each region to show the proportion of different kinds of companies, and classification numbers are used to partition. A new calculation field was used to count the industries and displayed them as pies and use a filter to output the top 3.

The second one is a bar chart that uses the classification number as the column and the count of industries as the row. The number of industries is calculated by a new field calculation use count function. A filter was also created to export the five most popular types. Different from the first bar, the year moves from the x-axis to the mark card and distinct by colors. Furthermore, the list of color, classification name, and number still appears at the bottom right of the dashboard for customers to find corresponding names. Through, this chart audience could directly observe the changes in two years and the difference between the total numbers of different companies.

References

[1] Alvarez, G. A., & Cavanaugh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106-111. Retrieved from http://ezproxy.library.usyd.edu.au/login?url=https://www.proquest.com/scholarly-journals/capacity-visual-short-term-memory-is-set-both/docview/195578639/se-2?accountid=14757.

[2] Palmer, John. (1990). Attentional Limits on the Perception and Memory of Visual Information. *Journal of Experimental Psychology: Human Perception & Performance, 16*(2), 332-350. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=ovfta&NEWS=N&AN=00004788-199005000-00008.

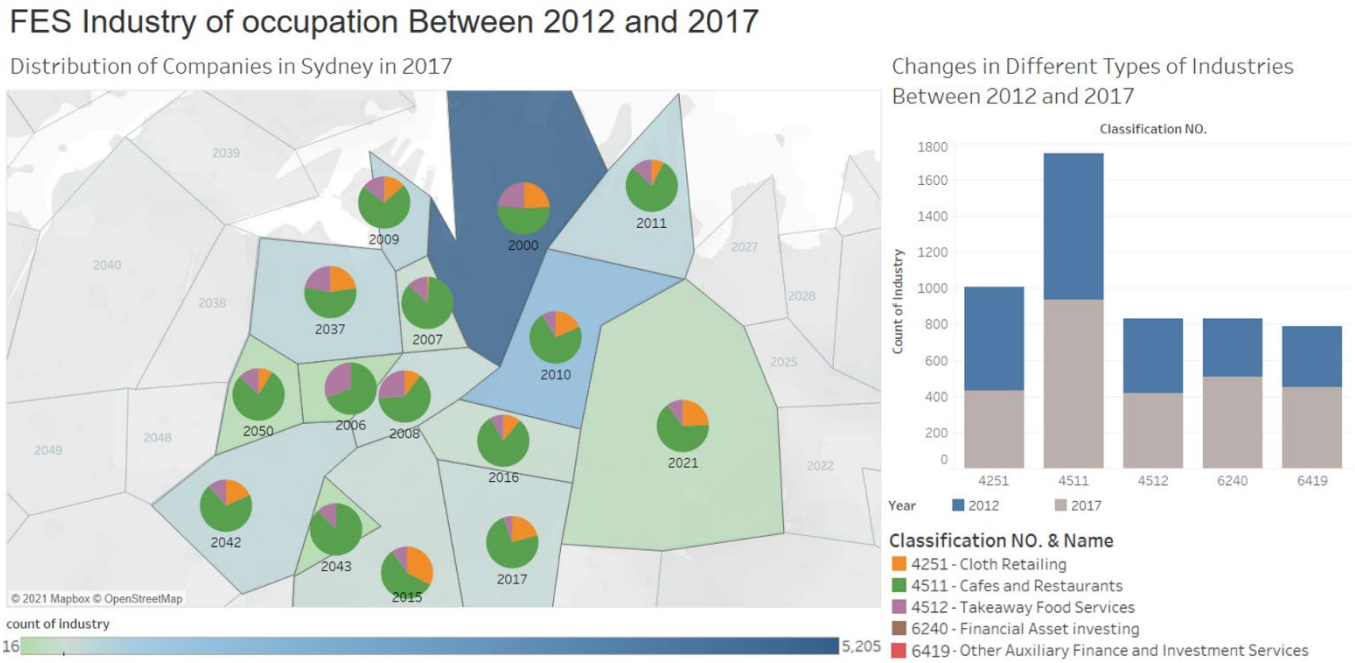

Fig. 10. Alternative visualization