Data Visualisation - Semester Project

Jephy Rapheal
No: 15232756

Data Selection

Dataset which I have chosen for visualisation is from data.gov.ie (Ireland's Open Data Portal). From this portal I have selected 'Vehicle Registration Tax' data published by the Revenue Commissioners [1].
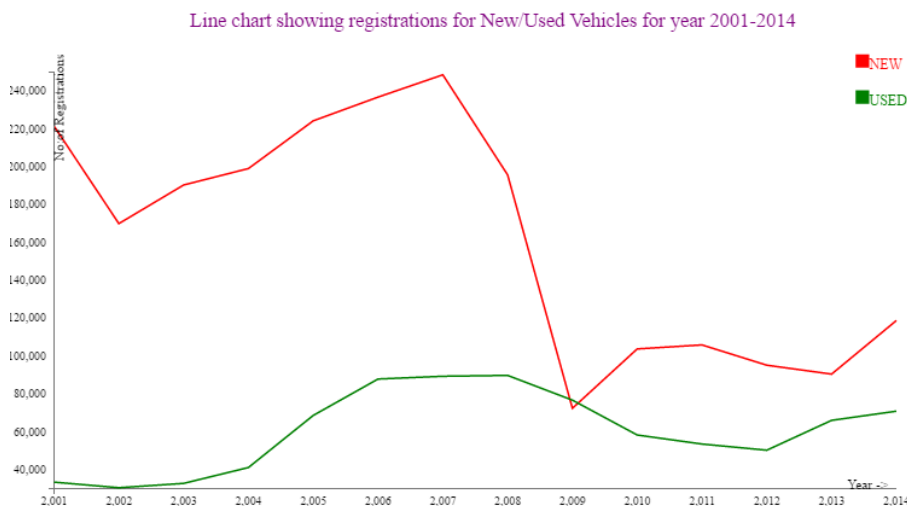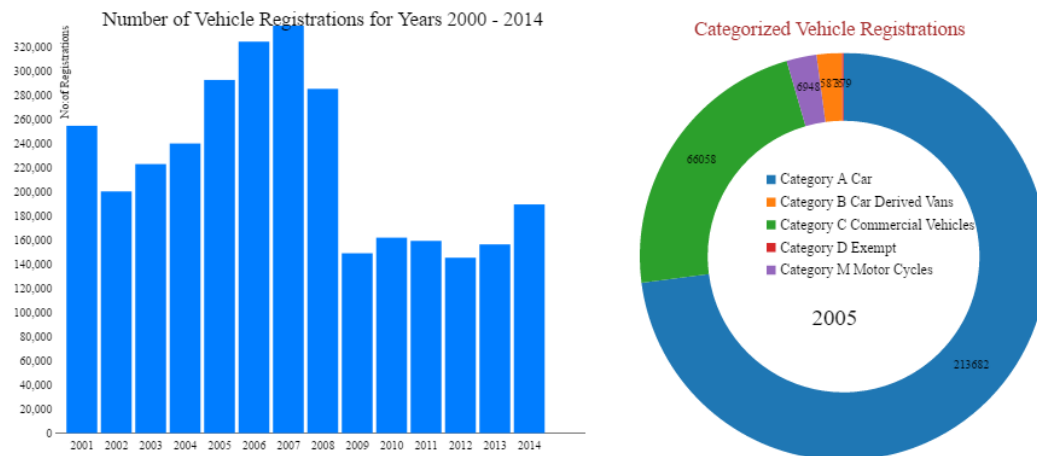Before making the choice of dataset, I had also considered datasets like Road Accidents details in 2014 for Dublin, again provided by data.gov.ie and also Flooded Roads in Galway for February 2014. These datasets were good, but I could not find any patterns or interesting facts. Moreover accidents and weather conditions were mostly unpredictable.

Vehicle registration sounded interesting as I did a quick analysis of the data, as increase or decrease in the number of registration could vary depending on the economic changes within a country. So it would be of much value to analyse and visually present the trends within the number of vehicles are getting registered.

The data file is in csv format and contains details like number of vehicles registered for each year, category of vehicle, vehicle condition (New/Used) and vehicle_receipts_€. These details are provided from the year 2001 to 2014.

Visualization

The design of visualisation which I have attempted to implement is below. It includes a bar chart, a Donut chart (using the pie Layout) and a line graph. The visualisation tries to give user insight on the variations in the number of vehicles getting registered in Ireland from the year 2001 to 2014. This is accomplished using a bar chart and line chart which shows the overall picture of registrations across all years. If the user is interested in analysis of any particular year, it can be viewed on a donut chart which would show a detailed category-wise division of vehicles registered.

Number of Vehicle Registrations for Years 2000 - 2014

Categorized Vehicle Registrations

- Category A Car
- Category B Car Derived Vans
- Category C Commercial Vehicles
- Category D Exempt
- Category M Motor Cycles

2005

Line chart showing registrations for New/Used Vehicles for year 2001-2014

■ NEW
■ USED

Bar Chart

The first visualisation component is a bar chart which present the entire statistics of the number of vehicle registered per year from 2001 till 2014. Years are presented along X-axis and Number of registrations along Y-axis. The bar chart is constructed from a TSV file 'data.tsv'

I have incorporated visual effects by highlighting bars on mouse interaction which would help users to navigate and to focus on the year of interest. Upon mouse over a tooltip text would present the actual number of registration for that particular year (For eg., figure below shows toottiptext for the year selected as 'Vehicle Registrations for year 2006 is 324712').

Moreover on click of a particular bar(Year) we get to see the category-wise breakdown of numbers of vehicles registered in a pie chart which is explained in the next part.

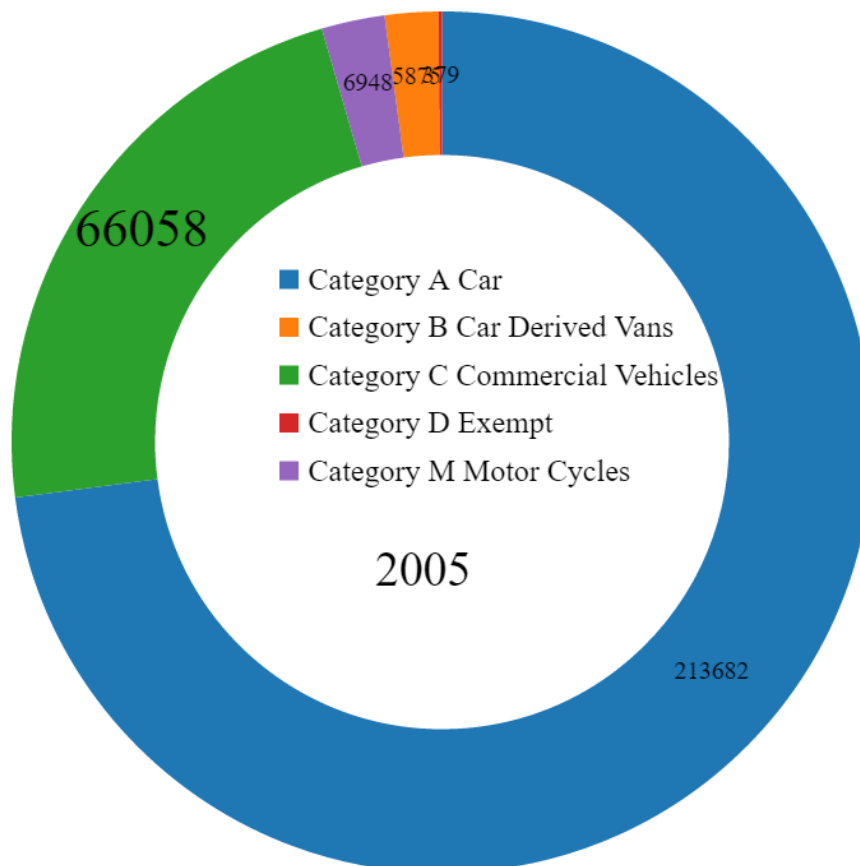## Number of Vehicle Registrations for Years 2000 - 2014
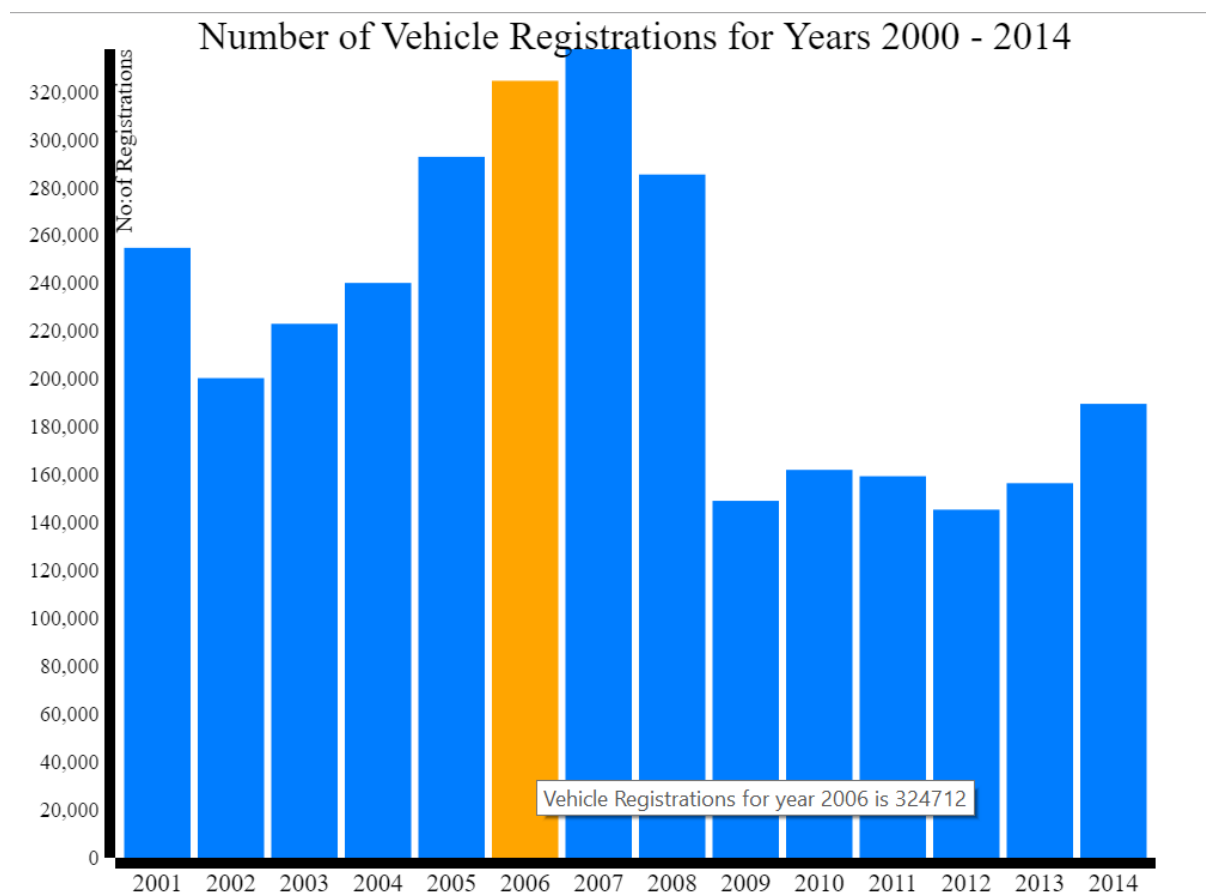
Pie Chart

The second chart which is a Pie Chart shows the category-wise break down of each vehicle (Category A Car, Category M Motor Cycles etc.,) for each year. Initial Pie chart shows the category-wise division for the entire year(from 2001 to 2014). However upon click of a particular Year in the bar chart we get to category-wise break-down of that particular year. This is to get an in-depth view of whether there was increase or decrease in the registrations for a particular category of vehicles. Value of Pie chart is calculated dynamically on click of a bar.

The Chart also has a legend to help user to identify which colour indicated which category of vehicle. It also shows the year for which the break-down of categories is being presented (for eg., figure below shows the category-wise breakdown for the year 2005) .

Moreover on mouse over of the figures, it gets highlighted so as to give a better and large view of the numbers.

Categorized Vehicle Registrations

Category A Car
Category B Car Derived Vans
Category C Commercial Vehicles
Category D Exempt
Category M Motor Cycles

2005

6948 5879
66058
213682

Line Graph

The third visualisation component is a line chart which present the entire statistics of the number of vehicle registered categorised as New and used vehicles, per year from 2001 till 2014. Years are presented along X-axis and Number of registrations along Y-axis. The line chart is constructed from json data. Number of registration for New and Used vehicles are plotted along two different lines in two different colours (Red and green) respectively.

Jephy Rapheal
No: 15232756

Line chart showing registrations for New/Used Vehicles for year 2001-2014
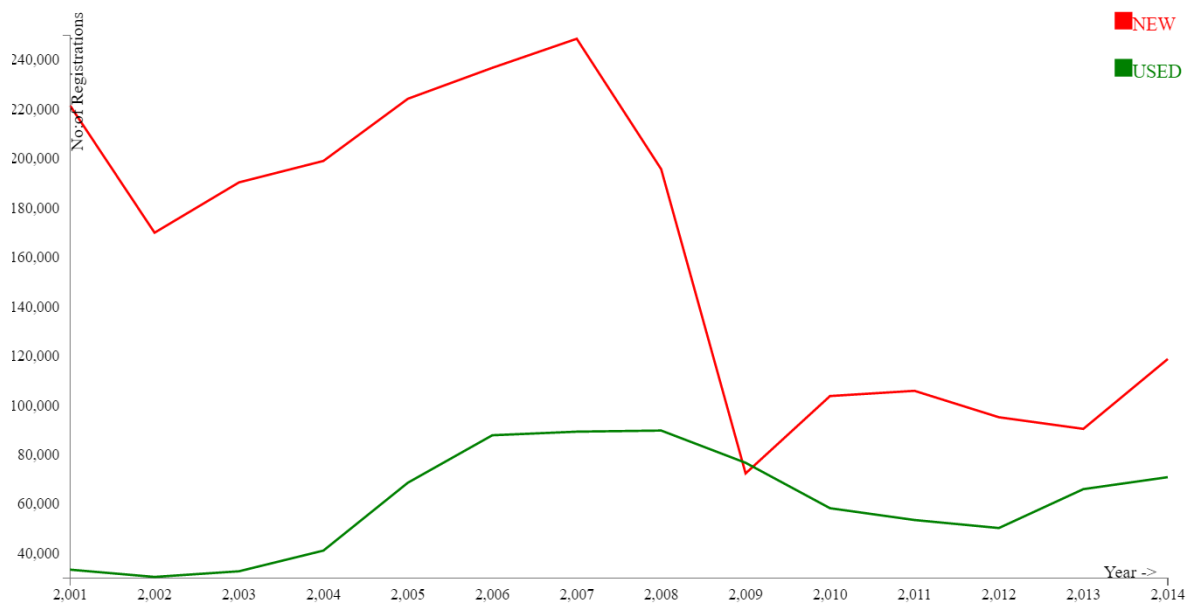
Cleansing / Parsing

To bring it to a shape to fit the chart for visualisation I have followed the below cleansing steps

For Bar Chart

As this chart summarises the entire data of number of registrations across the years, we need to summarise the number of vehicle registrations with total vehicle registration amount for each year, irrespective of the category of vehicle and vehicle condition (New or Used).

Hence I have considered only columns: **year,** number of **vehicle_registrations, vehicle_registrations_amt.** I have excluded the columns like vehicle_category and vehicle_condition. I have grouped rows based on year and took sum of vehicle_registrations and **vehicle_registrations_amt** column values.

For Donut Chart

Donut chart shows the division categories of vehicles for each year. Here I have considered only columns corresponding to **year, vehicle_category,** number of **vehicle_registrations.** I have not considered vehicle_registrations_amt and vehicle_condition here. To fit the visualisation I have calculated category-wise number of registration for each year. Upon selection of each year in the bar chart, data is dynamically bind and chart is refreshed. Initial chart shows the total number of vehicles registered for all the years and hence we all have added total number of vehicles registered under each category.

Data Visualisation - Semester Project

Jephy Rapheal
No: 15232756

For Line Chart

I have included details as json data stored in two different variables for New and Used Vehicles within the code. Key/Value pair in json data is Registrations/Year.

Reflections

As we analyse the bar chart,

- There was a gradual increase in the number of vehicles being registered from the year 2002 to 2007.
- It reaches the peak in 2007.
- There is a sudden decline, which reaches lowest point in 2009.
- Data dips from 320,000 in 2007 to around 150,000 in 2009.

Inference

This might have been because of the recession in the Banking, IT and other sectors across globe. Numbers shows a small gradual increase from 2009 till 2014 which indicates improvements in global economy.

If we analyse the line chart, which shows the number of registrations for New Vs Used Vehicles from 2001 to 2014 shows that:

- New Vehicles contributed more towards the dip in 2008/2009(recession period)
- Variations in New Vehicles registration is more compared to the Used Vehicles.
- Numbers in Old vehicles registration were not much affected by recession.

Challenges faced

There were several moments during my time while working with the design the visualisation, where I just wanted to give up. I initially use considered the entire dataset together without grouping and using the summarised data and got stuck up without coming up with the overall feel just was not coming along.

It took a long time and quite a lot of visualisations were considered and discarded as it was difficult to implement with the data which I have chosen. This was mainly because of limited experience in working with interactive visualisation. Finally I am quite satisfied with the final result with bars and pie chart.

Acknowledgements

I had referenced many resources and sample visualisations available in internet and had practised example of visualisations provided in tutorials provided with module. My project study was majorly inspired by a bar chart constructed from a TSV file storing the frequency of letters in the English language [2]. However this was a static chart without any user intervention and interactions on the data being presented.

I have tried to improve this by incorporating user interactions by updating a Pie Chart which is constructed dynamically based on the year selected in the bar chart. Pie Chart would visualise a breakdown of Category-wise division of number of vehicles registered of vehicles. I have also been inspired by a Simple D3JS Dashboard [2] which shows a dashboard with Piechart as main element and bar chart which is presented based on the section clicked in piechart. The chart shows breakdown of revenue share by a group of people and the amount they spend on various fruits.

References:

[1] - https://data.gov.ie/dataset/vehicle-registration-tax

[2] - http://bl.ocks.org/mbostock/3885304

[3] - http://bl.ocks.org/diethardsteiner/3287802

[4] - http://code.tutsplus.com/tutorials/building-a-multi-line-chart-using-d3js--cms-22935