CA682 Data management and visualisation

Name	Lei Cao
Student Number	16212709
Programme	MSc computing in software Engineering
Module Code	CA682
Assignment Title	Data Visualisation
Submission date	09/12/2016
Module coordinator	Suzanne Little

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found recommended in the assignment guidelines.

Name:_____Lei Cao_____ Date: _09/12/2016_____

Introduction

Data visualization is very important to our daily life, it tells a story and it is a way for human being to communicate with information. Good data visualizations are explanatory and exploratory.To achieve that goals . In this assignment, I created 4 graphs to show the number of new and used vehicles and the sales dollars respectively sold by month from 2002 to 2015 with D3.js, Javascript. Each graph has strong connection, the data shown will be changed as the first table graph data is changed.

d3.sj mainly is use to create lnteractive and annimated linear graph and import data, and I also use javascript to restructured my dataset from single dimensional array to two deimesional array, in addition, I used d3 api to import my csv format data to my application, and all of the work I have done by sublime text 3.

To make a comparison of total number of new and used vehicles respectively for each year, I have done a interactive linear graph which automatically shows the number of new and used vehicles for different month in one year when click the different year row in first table graph.

Dataset

The dataset was from MVA Vehicle Sales Counts by Month for CY 2002 - 2015. This dataset is intended for public access and use. At beginning, I got some inspiration from http://thestoryoftheseason.com/ which author Anna Powell Smith created a excellent interactive annimation linear graph that shows football match points and position of each team in UK, but when I've download dataset, I reconized that it is not complete, so I sent her email and asked for the rest of data, she told me that actually she calcauted that data from raw data, so I changed my mind to do this graph, but her graph gives me very good start point to think of what I am going to do and make a big picture for that. I want to find some similar structure dataset and went to a lot of dataset website and search for my ideal dataset. After 3 days, Finally I got it at DATA.GOV. I download it as csv format and later transfer them to JSON format in javascript. This dataset contains data shown like below

	A	B	C	D	E	F	G
1	Year	Month	New	Used	TotalSalesNew	TotalSalesUsed	
2	2002	JAN	31106	49927	$755015820	$386481929	
3	2002	FEB	27520	50982	$664454223	$361353242	

Process

1.Import dataset

Firstly, import dataset to my application by dataset API

```
d3.csv("MVA_Vehicle_Sales_Counts_by_Month_for_CY_2002_-_2015.csv", parser1,
```

2. Convert dataset

convert dataset to from csv format to json format in javascript after importing data by

```
function csvJSON(csv){

  var lines=csv.split("\n");

  var result = [];

  var headers=lines[0].split(",");

  for(var i=1;i<lines.length;i++){

      var obj = {};
      var currentline=lines[i].split(",");

      for(var j=0;j<headers.length;j++){
          obj[headers[j]] = currentline[j];
      }

      result.push(obj);
```

First I take the whole CSV file and split it into an array of lines. Then, I take the first line, which should be the headers, and split that by a comma into an array. Then, I loop through all the lines (skipping the first one) and inside, I loop through the headers and assign the values from each line to the proper object parameters. At this point, you probably want to just return the JavaScript object, but you can also JSON.stringify the result and return the JSON object.

3. Processing data

calculate data for first table data, in my dataset , each year has 12 records because of 12 months, so I want to get sum of the number of new and used cars so that I can only show 2002 to 2015 14 rows data.

```
data.forEach(function (elem){
    if(elem.Year == tempYear || !tempYear){
        tempYear = elem.Year;
        row.year = elem.Year
        row.newCarsNumber += parseNumber(elem.New);
        row.userCarsSold += parseNumber(elem.Used);
        row.totalSalesNewValue += parseNumber(elem.TotalSalesNew);
        row.totalSalesUsedValue += parseNumber(elem.TotalSalesUsed)
        row.monthsData.push({
            month: elem.Month,
            newCarsNumber: parseNumber(elem.New),
            userCarsSold: parseNumber(elem.Used),
            totalSalesNewValue: parseNumber(elem.TotalSalesNew),
            totalSalesUsedValue: parseNumber(elem.TotalSalesUsed)
        })
```

so I read in each row data to temp object by a loop, testing if the year is same, then add the the number of new and used vehicles to previous record.

4. Making graph using SVG

Take line graph as example.

Making date format

```
var parseDate = d3.time.format("%d-%b-%y").parse;
```

Add X and Y axis (two Y axis here)

```
  var x = d3.time.scale()
      .range([0, width- margin.right - margin.left])
      .domain(d3.extent(localData, function(d) { return d.date; }));

  var y = d3.scale.linear()
      .range([height- margin.top - margin.bottom, 0])
      .domain([min, max]);
```

Draw lines (two Lines)

```
  // Scale the range of the data

  var path = svg.append("path")        // Add the valueline path.
      .attr("d", valueline(localData))
```

5. Graph show

Year table graph shows the data which makes a sum of the number of new and used vehicles and sales for each year respectively.

Second graph is interactive linear graph, which shows the counts of new and used vehicles for each month of a specific year, the data can be refreshed when click and choose the year row in first table

Month table shows data of each month and also can be changed via first year table

Graphs graphs give information about percentage of new and used vehicles

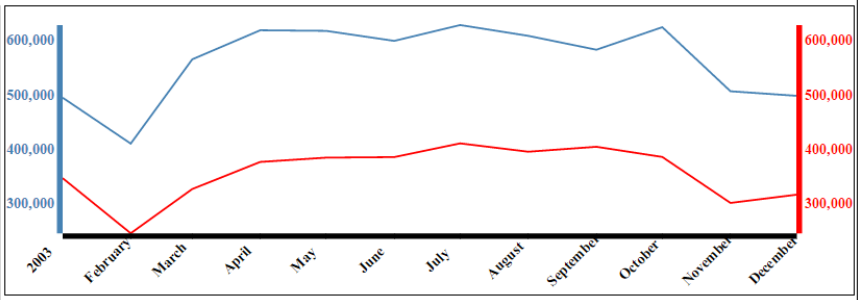
MVA vehicles counts and sales by month from 2002 to 2015

Sold cars in years

New cars | Secound hand cars

year	sold	vale $	sold	value $
2002	4021640	95725548760	6560890	49402097720
2003	4282520	106231483390	6753580	51660714970
2004	4285080	109729560240	6996770	53855489470
2005	4218340	110067339220	7038350	56398035150
2006	3992820	104116572060	6969680	58656873180
2007	3781840	99979174830	6785490	57942138690
2008	3091590	79754019870	6178850	48961678400
2009	2489280	66862179140	6088890	45100405270
2010	2680220	74609155220	6260450	50548023030
2011	2876690	83317324020	6257280	53450740830
2012	3167620	92849943390	6276780	55821270710
2013	3352090	101134900160	6447550	59761300470
2014	3412190	105772330710	6564660	63451445130

New Cars $ | Total income of sold cars in months | Secound hand cars $



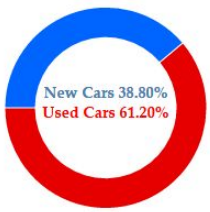
New Cars | Secound hand cars

month	sold	vale $	sold	value $
JAN	347070	8607307840	494810	3924388330
FEB	245850	6157499480	410480	3232970770
MAR	327050	7917280690	565130	4138995740
APR	376820	9097782200	618560	4581279730
MAY	384720	9173396080	617420	4697346470
JUN	385730	9418983870	599040	4551178900
JUL	410880	9900417540	628110	4893743620
AUG	395560	9853753780	608290	4603167260
SEP	404640	10124049870	582790	4461195250
OCT	385970	9803656780	624170	4705209060
NOV	301550	7796749570	506550	3851646630
DEC	316680	8380605690	498230	4019593210

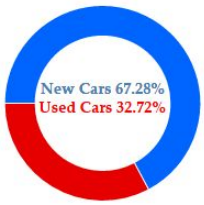
Number of sold cars

Used cars

New cars



Total income of sold cars



I used sublime text 3 to do my code, which is easy to save and small installer package,

Result

Kirk's principles of good data visualization ever suggest that good data visualizations are trustworthy , accessible, Elegant. In term of trustworthy, using appropriate color or fonts. As you can see the interactive linear chart , I used two different color blue and red to represents counts and sales respectively.

good data visualization can make communication with users, in my case, no matter which year users want to see, just click the row with a certain year and the second table will shows details of data for each month , and the linear graph will show data of this year. it will be easy for users to see the trend and changes within one year, instead of one graph shows all data and it will be messy and difficult to distinguish the data for different year.

Additionally, I used donut graph to show percentage of new cars and used cars, it is very useful and understandable,

but it does not achieve my initial goals, for the year table, I think I need to put more effort and change it to a bar chart, it can solve a specific problem such as which year has highest sales or counts of new and used vehicles. Additionally, interactive mouseover to view details is another features which i want to achieve, I have done a simple interactive mouseover to view details linear graph, but because of time limitation, I give up to further update my graphs.