Drive To You

CSI703 Final Project Process Book

”Trenton” Chen Jin[1] , Qianyi Huang[2] , Xiaoqian Cai[3]

[1] cjin5@masonlive.gmu.edu

[2] qhuang5@masonlive.gmu.edu

[3] xcai2@masonlive.gmu.edu

May 11, 2019

1. Overview and Motivation:

Life without cars in most areas of this Country is impossible. It doesn't matter whether you're a recent grad with your first real job or someone who has driving to work for decades, the car-buying process is never filled smooth. One of the reasons is that car is complicated and there is no simple way to compare one car with another by a one-dimensional vector comparison. The vectors need to be brought into the comparison equation includes speed, weight, dimensions, horsepower, torque, wheelbase, convenience and so on. Currently, all website with car comparison available would provide very similar table comparison by listing key factors on two columns. The weakness of this traditional method is that the information is easily biased by the large numbers and readers cannot quickly understand information and picture overall graphs of two cars. In this project, we dedicate to utilize visualization tools to make a user-friendly interface for interactive car comparisons.

2. Related Work

Cars are related to everyone's daily lives, either directly or indirectly. As a car enthusiast, I like watching car related TV shows, like Top Gear, Grand Tour, and listen to Everyday car driver on podcast. In some of the car shows, they introduce some head-to-head car comparison. While, if I want to search on the Internet and compare two cars myself, it is always being a difficult target. There are tons of websites for car purchasing guidance, such as U.S. News cars (https://cars.usnews.com/cars-trucks), Edmunds (https://www.edmunds.com), and so on, but they all provide a similar table for car comparisons like follows. The table always throws hundred rows of numbers, which is not a very intuitive and user-friendly way to read the differences. It is not an efficient way for general users to understand and compare the information. Therefore, we think having a good visualization of those data is crucial.

Add to Comparison	 2019 Audi A4 #1 in Luxury Small Cars	 2019 Honda Accord #4 in Midsize Cars
	Premium 40 TFSI	LX 1.5T CVT
	20703 Find Best Price	20703 Find Best Price
Prices		
MSRP	$37,400	$23,720
Avg. Paid	N/A	N/A
U.S. News Scores		
Overall	8.9 /10	8.7 /10
Critics' Rating	9.3 /10	9.2 /10
Performance	8.6 /10	8.4 /10
Interior	8.8 /10	8.4 /10
Safety	9.5 /10	9.9 /10
Reliability	3.5 /5	N/A

3. Questions

There are some primary questions we want to answer for this project.

- Which aspects of car are going to presented in our website?
- Who are our main audiences?
- Why our visualization is better than the traditional car comparison website?
- What methods and skills to make our visualization most user-friendly?
- How to classify data and create general visualization templates for customized styling?

Each car is assembled by thousands to millions of parts and that causes every single car born with different and unique character, for instance dimensions, horse powers, curb weight etc. The first issue needed to be addressed is that which aspects would be presented in our project. The aspects must follow two rules: 1. it is an important car feature, 2. the graph visualization of this character is more efficient than words descriptions. Answering who are our main audiences will narrow our result to be a user-oriented visualization tool for car comparisons. To answer the rest of the questions, we need to investigate on some current car comparison websites and think out the best resolution for our visualization design. We will learn arts design, data collection, user oriented, d3, data visualization, and more through this project. Our big goal is that the final result serves more than just a course project but could benefits many car buyers or car enthusiasts.

4. Data

We collect data on reliable and authorized websites: https://cars.usnews.com/cars-trucks (for basic dimension and engine information), https://www.zeroto60times.com (for acceleration statistics), https://www.goodcarbadcar.com (for sale information), https://www.carcomplaints.com (for safety statistics).

1. https://cars.usnews.com/cars-trucks
2. http://usedfirst.com/
3. https://www.carcomplaints.com
4. https://www.goodcarbadcar.com
5. https://www.zeroto60times.com

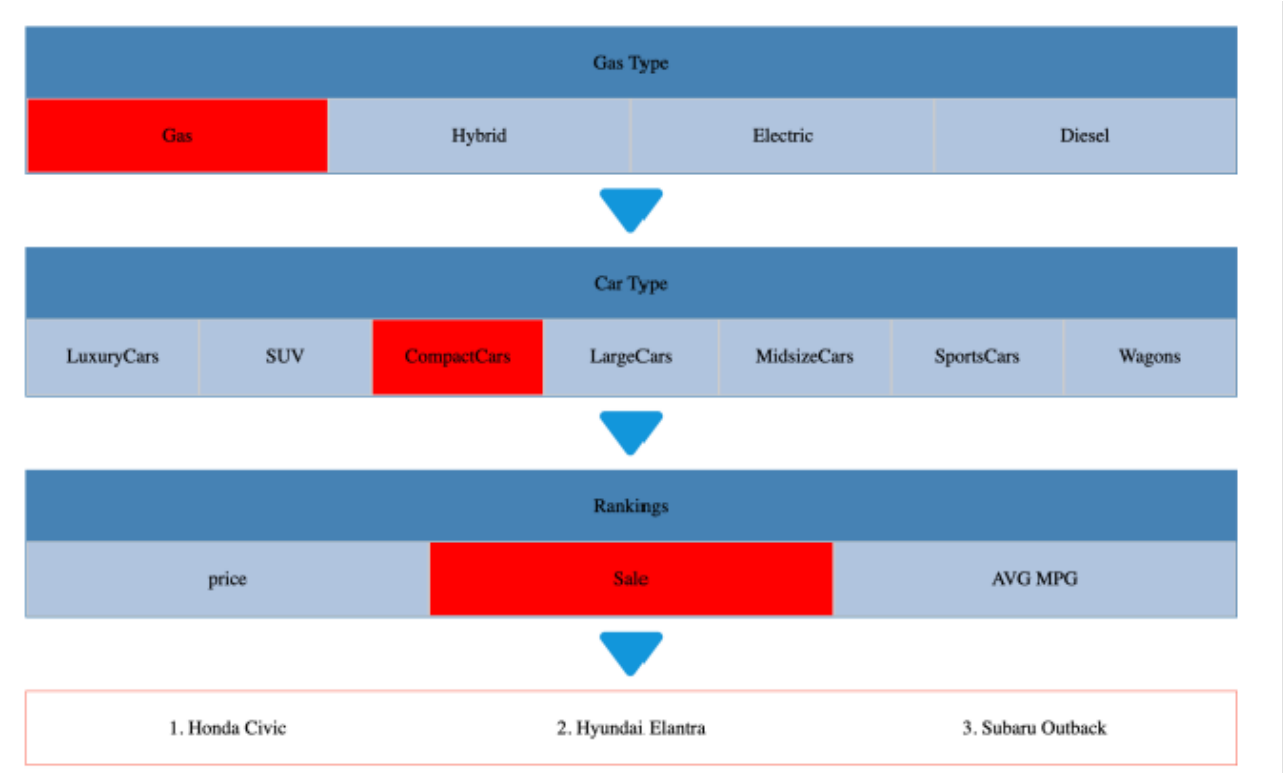
5. Exploratory Data Analysis

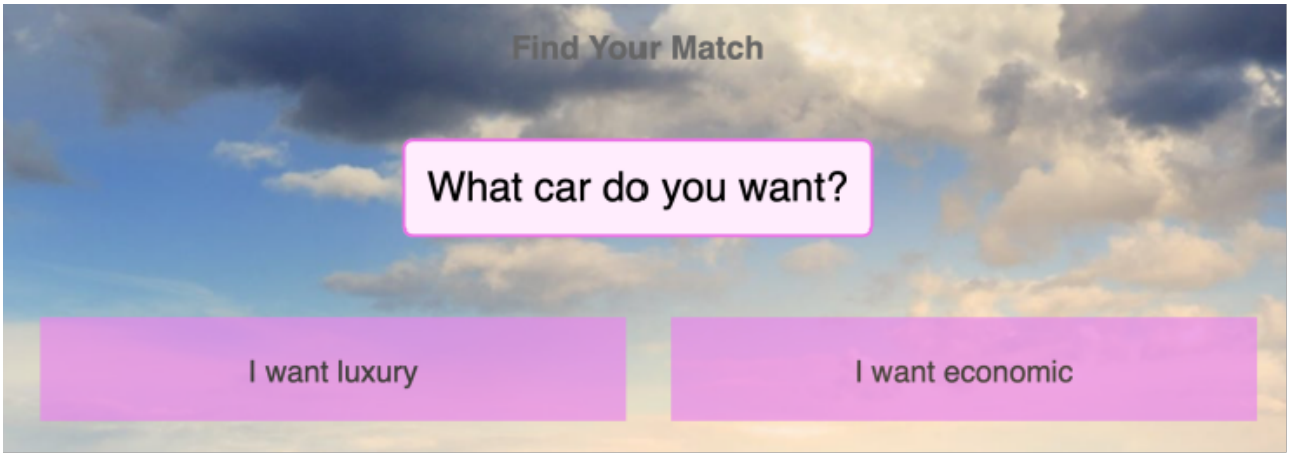
We will collect data mainly from the website https://cars.usnews.com/cars-trucks. At first we fused Python to establish a web crawler algorithm, and with this code we could obtain all the basic statistics about every given make and model from https://cars.usnews.com/cars- trucks/ browse. However, when it came to year or specification, the website directed to different pages (For example: https://cars.usnews.com/cars-trucks/acura/ilx/2019/specs/ilx-sedan-403951 and https://cars.usnews.com/cars-trucks/acura/ilx/2019/specs/ilx-sedan-wtechnology-pkg-403954). Thus, our code no longer supported the crawler. Since the en- gine or even dimension difference between specifications could not be ignored, we turned to a soft- ware called Houyi to do the web crawler for us. The output of web crawler is .xls file, so that we can clean the data in Excel. From this website (https://cars.usnews.com/cars-trucks), we obtain two datasets: 1. Dimensions, engine, fuel statistics listed by make, model, spec, and year; 2. Safety, classification information by make and model. Finally we collect the third dataset about sale statistics in the U.S. in 2018 from http://www.goodcarbadcar.net/2019/01/december-2018-ytd-us-passenger-car-sales-rankings-best-selling-cars-in-america/. This data is also listed by make and model.

6. Design Evolution

For the find-your-match part, we deigned a filter machine to list the target cars which meet the requirement of the users, and this part comes last. However, finally we decided to move it at the beginning of the webpage, since this part is designed to provide a recommendation for people who are not familiar with cars or those who have no specific target cars. And we also changed the form

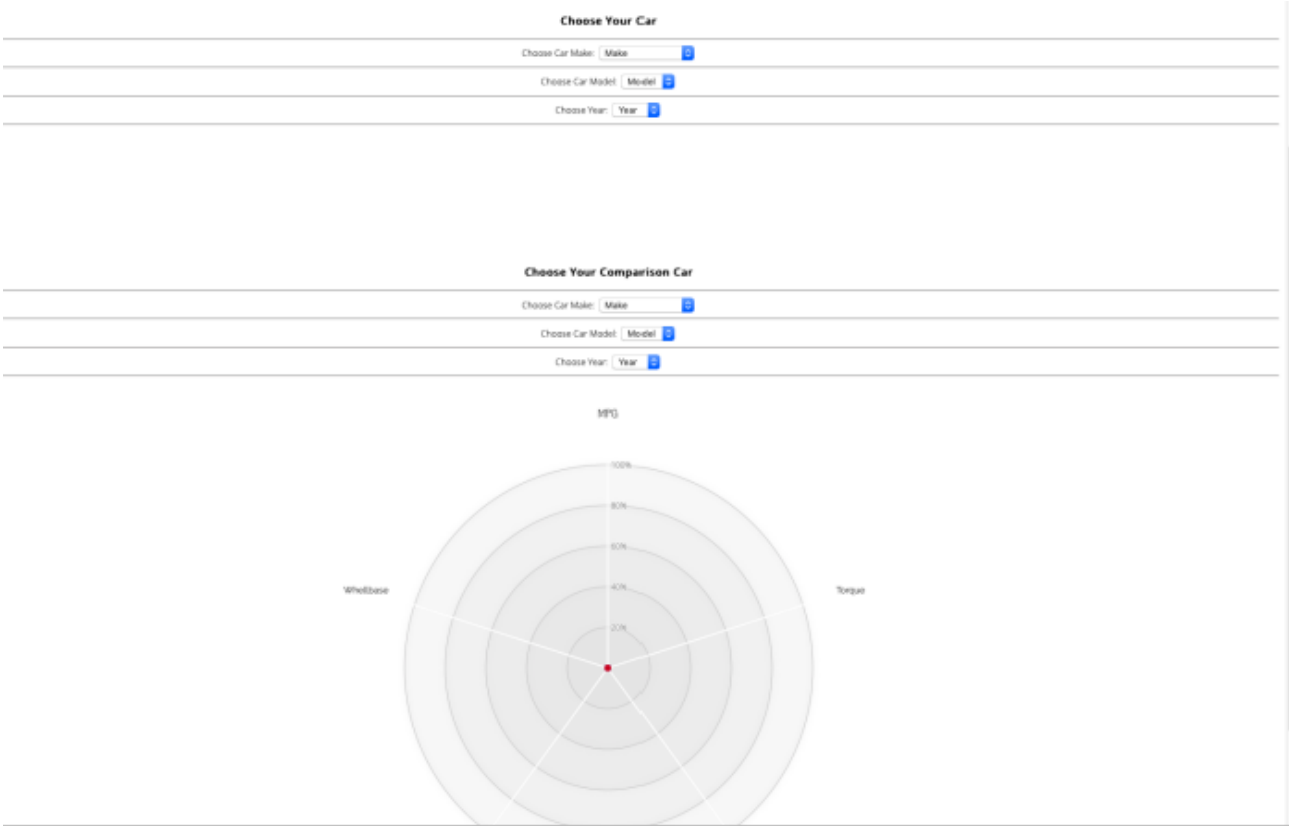
of this part. We are inspired by Toyota's official website, and we think we should do the filter for the users according to our knowledge about cars. In this way, the users can find the cars which really meet their demands.

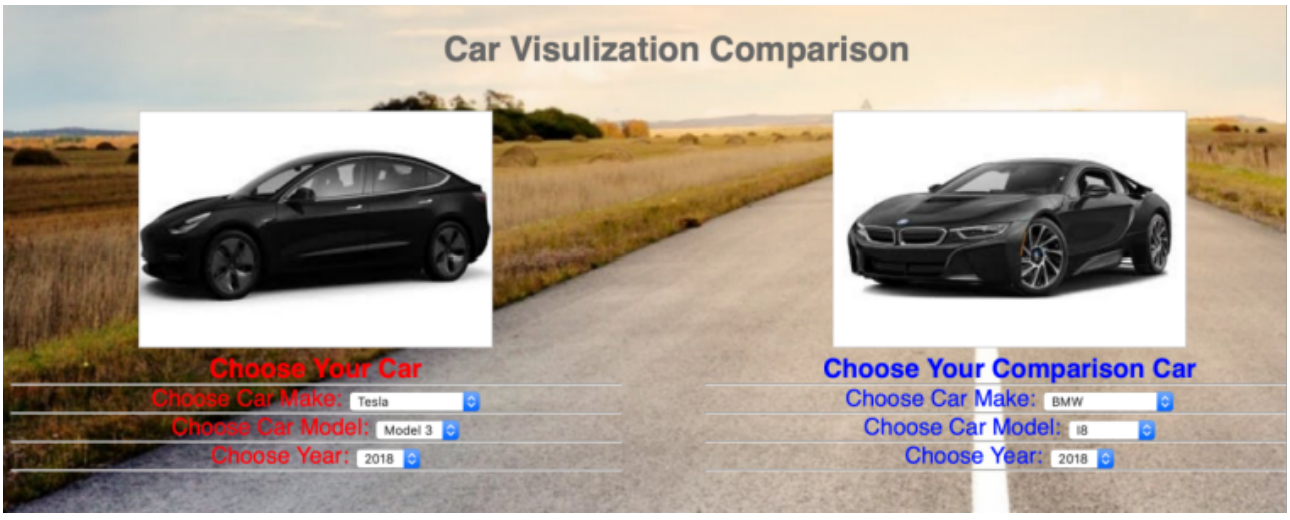




For the drop-down list part, the two cars are listed vertically in the first draft but there is no space

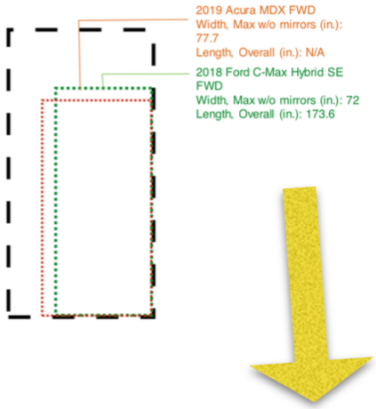
for the drop-down list. Thus, we change the way they are listed and the two cars come horizontally.

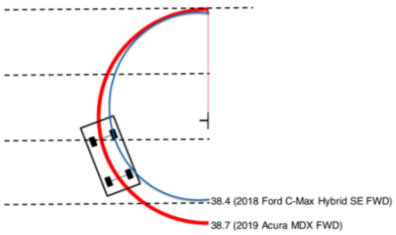




For the dimension and the turning diameter comparison, we decided to delete this part because the difference of length/width/height between two cars is not that obvious, and it's hard to visualized. We think this part is not necessary since we can also compare the wheelbase in the radar chart.

For the comparison of comfortables, we originally designed this part to show the difference of features like fog lamp and keyless start. We still insist that this part of comparison is useful to most users. However, since we combine the trims under the same model/make/year, and the features between different trims vary, it is not workable to add this part any more. Taking the size of the database into consideration, we delete this part.





Hover

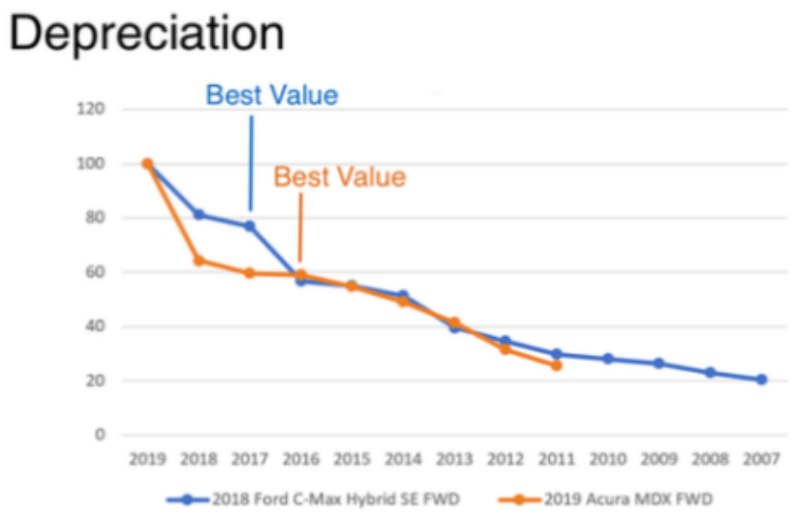
Animation

For the 0-60s acceleration, we designed animations in the first draft. However, we also delete this part because it is not easy to observe the 0.1s difference by eyes. It is a fancy part, but it is sad that we have to delete it.

For the depreciation part, we designed the year of best value in the first draft. We collected the data from the goodcarbadcar website. According to our knowledge of cars, we find the data provided by the website not convincing. And we also failed to find an object method to calculate the year of best value. Therefore, we delete the data of best value year. Moreover, to explain the concept of depreciation to new users, we also add some illustrations.

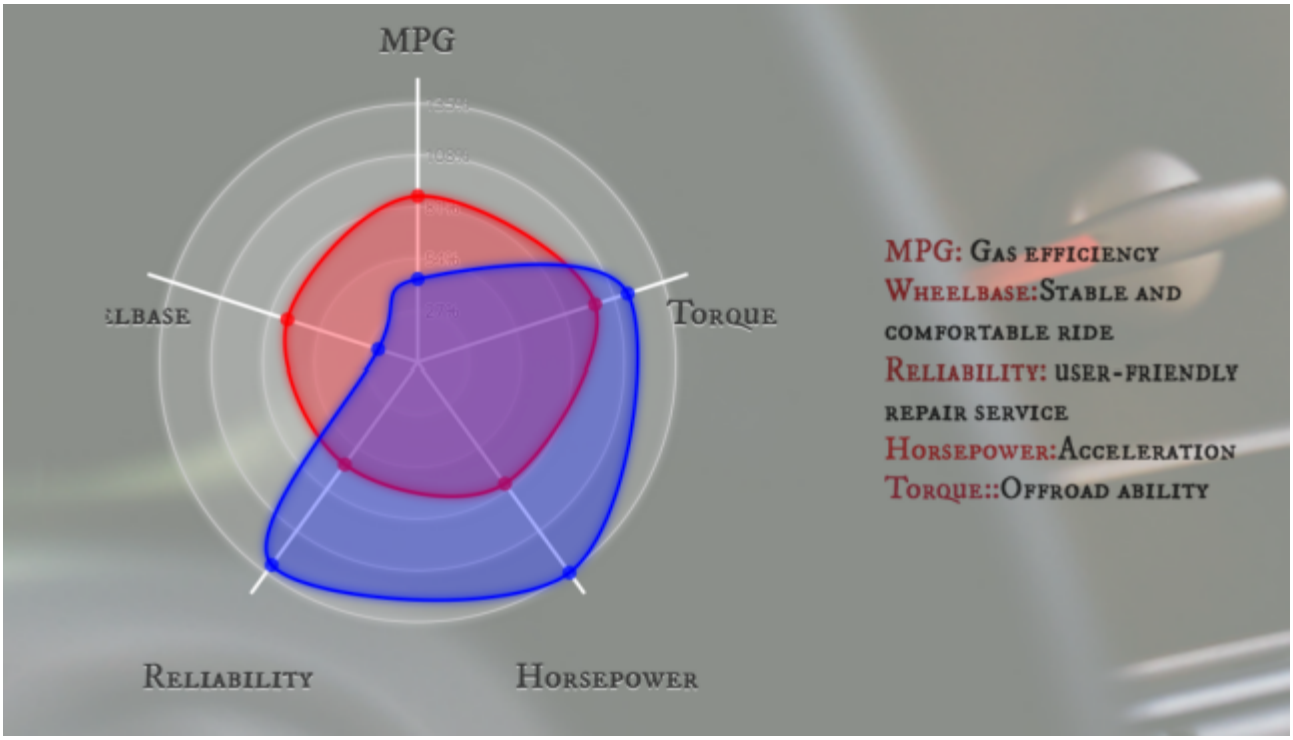
For the sales part, we wanted to put on a map of the US, and list the most popular car of each state. However, we find that most states share the same car like Ford F-150. And we do not think this data can provide useful information when users are looking for a car to buy. Hence, we change the form

to present the sales data. We use the scatter plot to describe how well the car is sold. Also, we add



some illustrations to help users to understand the graph.

For the radar chart part, we did not make big changes because we think radar chart can provide users with a straight overview of cars. Though the area of the chart does not direct to a conclusion of a good car or a bad car precisely, cars with bigger area do tend to be a good car. This part is especially friendly to people who are not familiar with cars. Moreover, expert people can also easily to find the specific part like wheelbase/torque/horsepower/MPG that they are mostly interested in.



7. Implementation

Radar Graph

Radar Graph

To bring dynamic, interactive data visualizations in web browsers, we combined d3.js, HTML, and CSS standards together to implement our project, including data manipulation and function building in JavaScript, designing layout in html, and polishing the style in CSS.

Out Radar graph is implemented to give the polygon area comparisons of car dimensions between targeted cars and comparison car. First, we need to build five axis to present five car dimensions in five circles with same center point, and each dimension stands for one attribute of cars' performance (refer to table 1).

Table 1: Car Dimensions & Car Performance

Dimensions	Performance
MPG	Gas efficiency
Horsepower	Acceleration
Reliability	User-friendly maintain and repair service
Torque	Off-road Performance
Wheelbase	Stable and comfortable ride

Second, we map our original dimension variables to percentage scale data to make sure each axis can share the same tick marks and scale. Specifically, We calculate the range of each dimension variable, and then use the quantile of each car's dimension record as the percentage scale. With five percentage scaled data, we can draw the polygon.

Next, we add highlight and hover functions to present the polygons of different cars, as well as adding labels, and fulfill different colors for different car selections. So far, we wrap up all the code together to build the radar building function (refer to radarbuild.js).

Finally, we need to retrieve data correctly to use our radar building function for radar graph. In our case, we create a new dataset by filtering the car dimension datasets based on user's car selection. Before the user make the car selection, we provide a dataset with only zero values for all five dimension variables, and build it as the default mode for our radar graph.

8. Evaluation

1. We only collect data at the level of model, not the trim. As a visualization design of what data we have, it is not a problem. However, as an improvement of current cars website, which is mentioned in our motivation, it is still not big enough.
2. For electric cars, we still need to adjust the value of MPG, or we should add some illustrations to explain that the concept of MPG does not apply to such cars. Or we can also do some calculation to measure the cost of electricity to compare with the cost of gasoline.
3. Actually, we try different cars selection including both economic and luxury cars in the radar chart. In general, the area of luxury car's polygon is better than the area of economic car's polygon, which is not contradict to our common sense that consumers pay more to enjoy better ride performance. Also, all the function of our radar graph runs smoothly, such as radar graph changes on the selection of cars, the area of polygon is highlighted once the mouse is hovering on the polygon. Thus, our evaluation of radar graph is informative and functional.
4. We collect the data of reliability on JDPower website. However, some models do not have reliability score under year 2018. To fill the missing data, we collect the score of those models under year 2017, 2016 or even 2015. Since the score of reliability does not keep the same in these years, this data is not that reliable.
5. We can add more guidance to our website since every part is not separated from other parts. We can lead people to explore the cars and the detailed statistics of the cars.
6. Overall, we finish most of our designs. The deleted parts are not reasonable or necessary in our visualization, and we also add some parts that can benefit our webpage.