

NYPD Motor Collision Data Visualization

Elly Rath

Manish Shah

Data Visualization | W209 | 04/19/2018

# Post Demo changes:

Based on the feedback we received, we made the following changes.

1. Changed the animation on the d3 map interaction.
2. Changed the colors on the tableau charts from categorical to sequential.

# Project URL:

* http://people.ischool.berkeley.edu/~ellyrath/

## Participation:

|  |  |  |
| --- | --- | --- |
| Task | Manish | Elly |
| Pre-Processing | 50% | 50% |
| User Interviews | 50% | 50% |
| Web page design | 50% | 50% |
| Coding | 50% | 50% |
| Performance Testing | 50% | 50% |
| Final Presentation | 50% | 50% |
| Final Paper | 50% | 50% |

## 

## Project Description:

New York City is divided into 5 boroughs (districts). New York city is one of the busiest city in US and has very heavy traffic. The winter weather also creates treacherous traffic conditions. New York Police department publishes the traffic accident report on the New York city website (<https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95>) as part of public safety initiative. The data is listed by date, time, intersection, geo coordinates and number of injuries and death of pedestrians, cyclists and motorists. The data set contains 1.23 M incidents from 2012 to 2018.

The purpose of this project is to provide a visualization of the traffic data to understand the intersections and areas where there have been repeated accidents.

We also integrated the weather data from NOAA (<https://www.ncdc.noaa.gov/data-access>) to understand the impact of weather on traffic incidents. This site provides hourly data for the weather and includes precipitation and temperature based on the sensor locations. There are 109163 records in the weather dataset.

## Preprocessing:

We found that for many of the rows in the NYPD data didn’t have latitudes and longitudes. We excluded those rows for the map charting but included them in the statistical charts.

The data is in CSV format and for some of the rows, the commas were misaligned. We looked for relevant information and if present selected the data else discarded it. This changes reduced our data size to be around 1M incidents.

For mapping the weather data to the NYPD data, for each incident, we looked at the nearest sensor location and picked up data from that location. If the sensor location had missing data, we picked up the next sensor location. For New York City, the weather data had two sensors.

We found that visualizing all the 1M data on the website was slow and provided little insights into the data. To improve this we did the following.

1. Restricted the map charts to top 50 fatal incidents by year. The interaction can overlay the data to give a picture of most accident prone areas.
2. Extracted relevant information from the dataset and transmitted only those data that were useful for the visualization. We extracted the data and created a geojson file with the coordinates to provide statistical information via tooltip for that location.

## Visualization:

We selected Tableau and D3 for our visualization and interaction. We implemented the statistical charts and maps in Tableau. We implemented the top 50 fatal accident in D3 in both the high level view and street map in D3.

For the high level map, we converted the shape file and converted it to geojson for rendering. We then over layed the data extracted from the NYPD dataset using geojson. The coordinates was used to place the spots and the meta data was used for displaying statistical data in the tool tip. The meta data had data arranged by year and we used d3.filter to extract data by year.

To display the borough data, we used d3.nest to extract data by borough and year.

For the street view, we used the leaflet library and over layed the data from the NYPD dataset to display the accident spots.

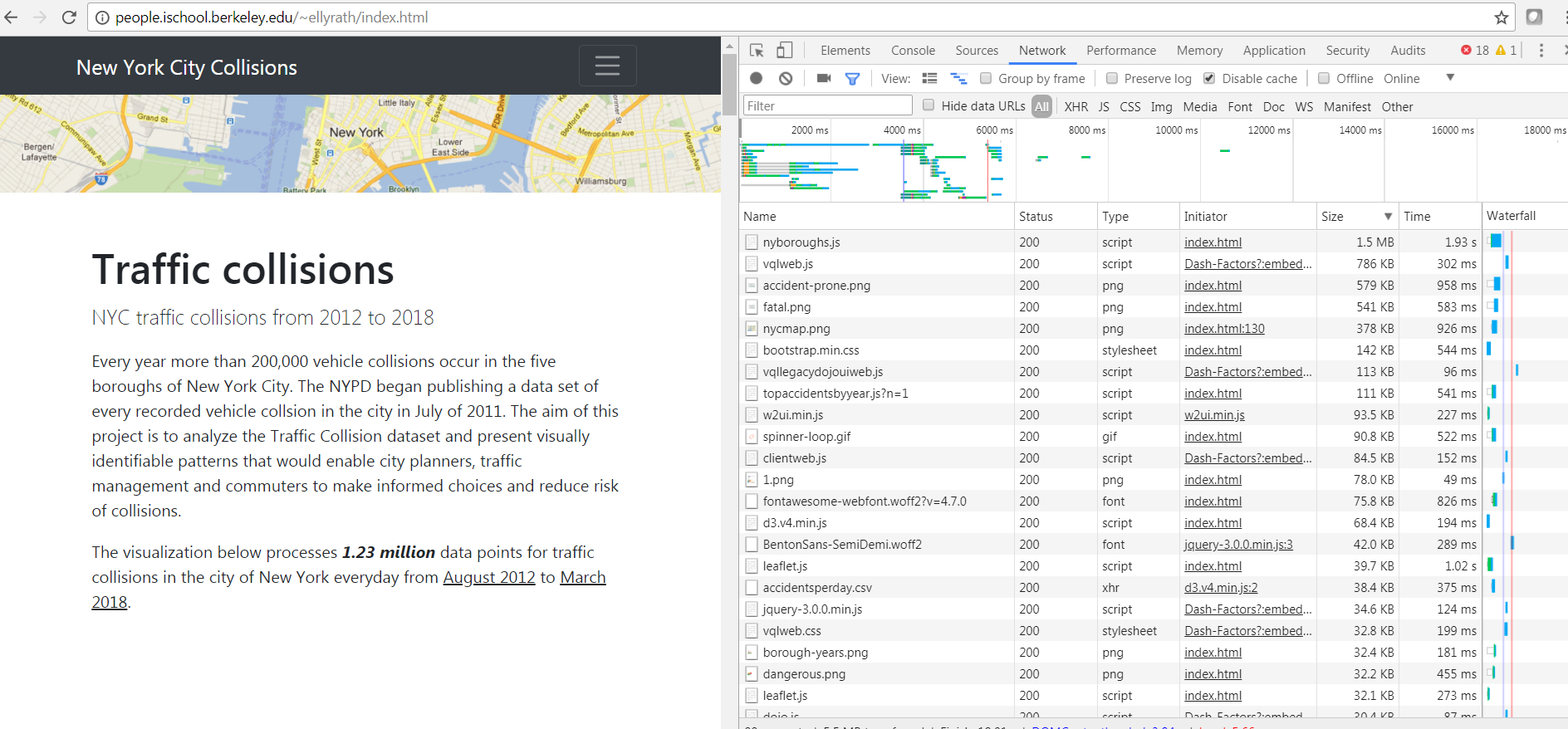
For the time series chart, we extracted the needed data (data, total accidents) from the NYPD data set and created a csv to render the data.

Use cases:

The visualization is intended for tourists planning to visit New York city or traveling from one borough to another, new drivers learning to drive and daily commuters, pedestrians and cyclists. The people can identify intersections where there have been multiple fatal accidents and be alert while traveling in those intersections. This visualization can also be used by city planners to put advisory in accident prone areas.

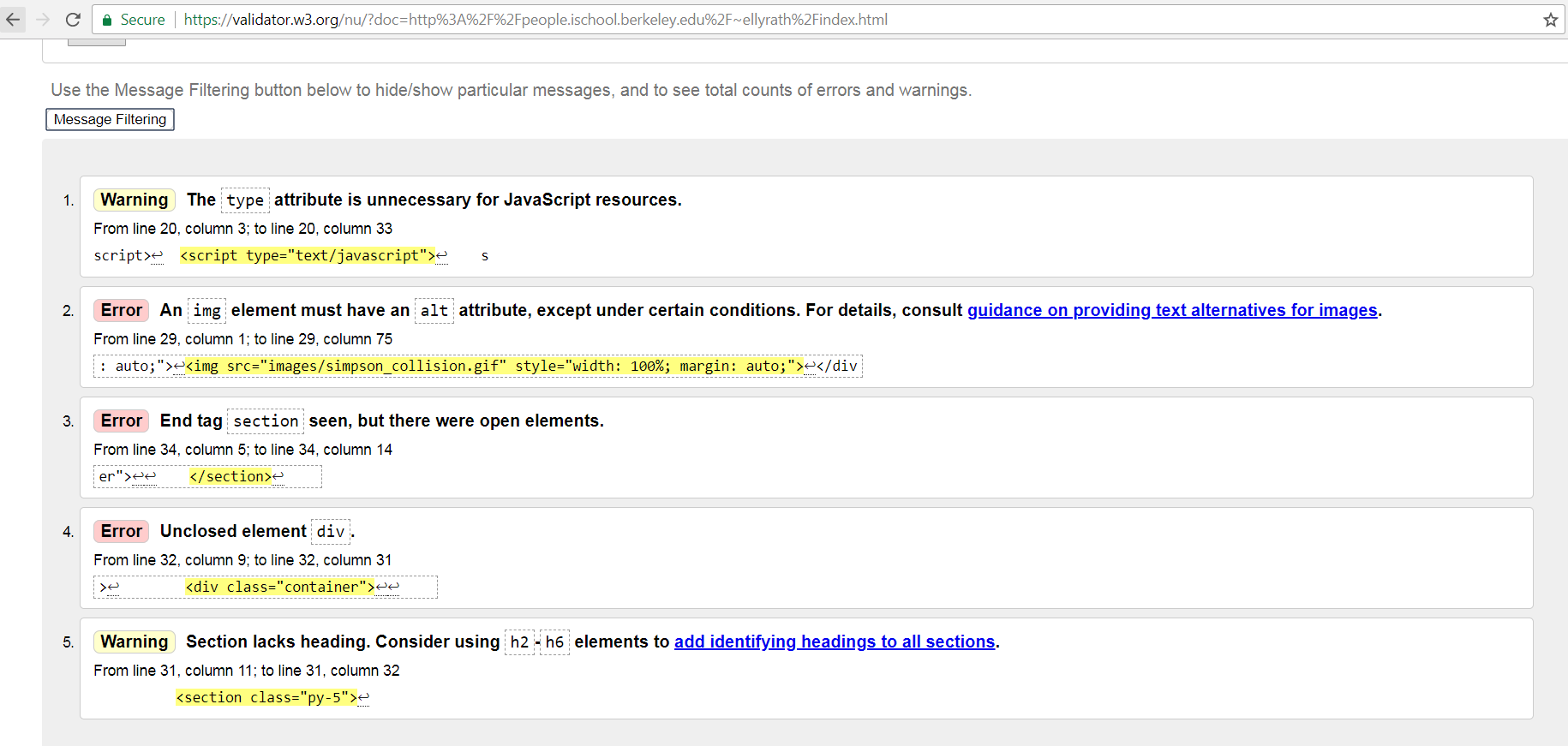
Performance:

It takes 6000ms to load the page. The largest data size transfer is 1.5MB.

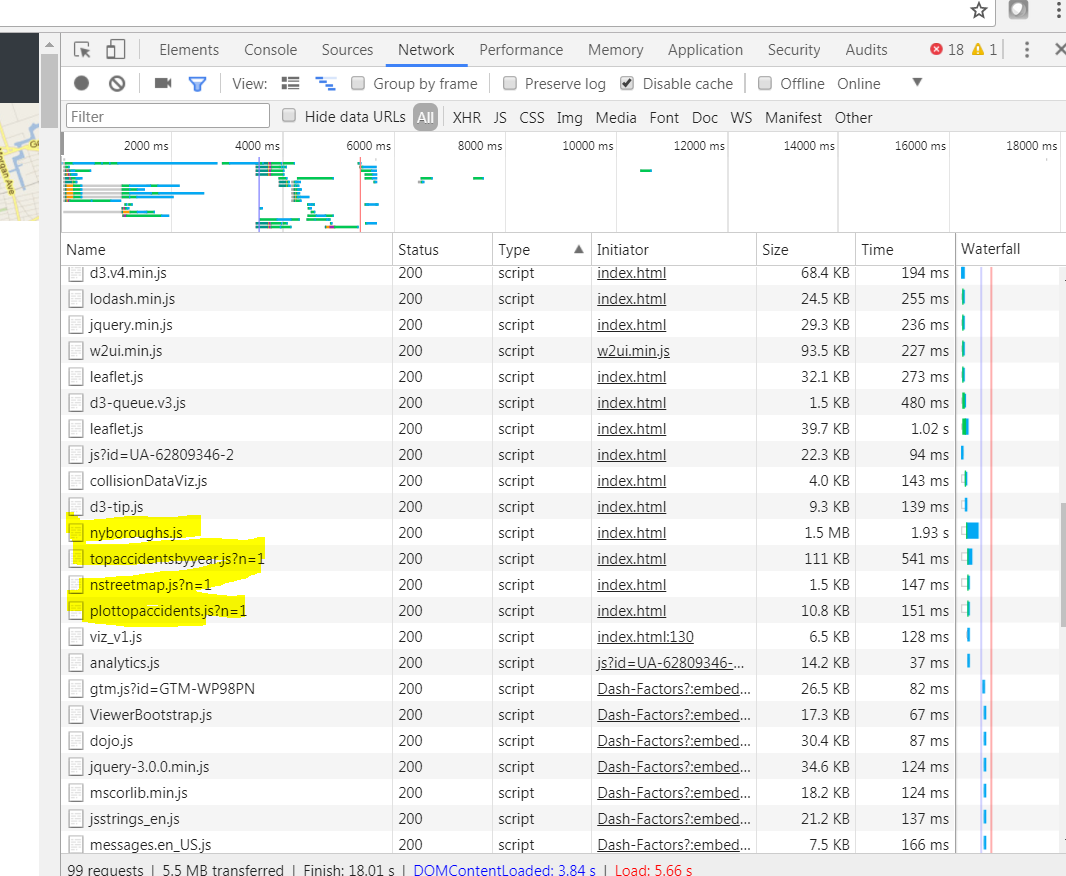


Code Validation:

We used <http://validator.w3.org> website to validate our code for proper use of html, js etc. The site identified minor errors which we corrected.

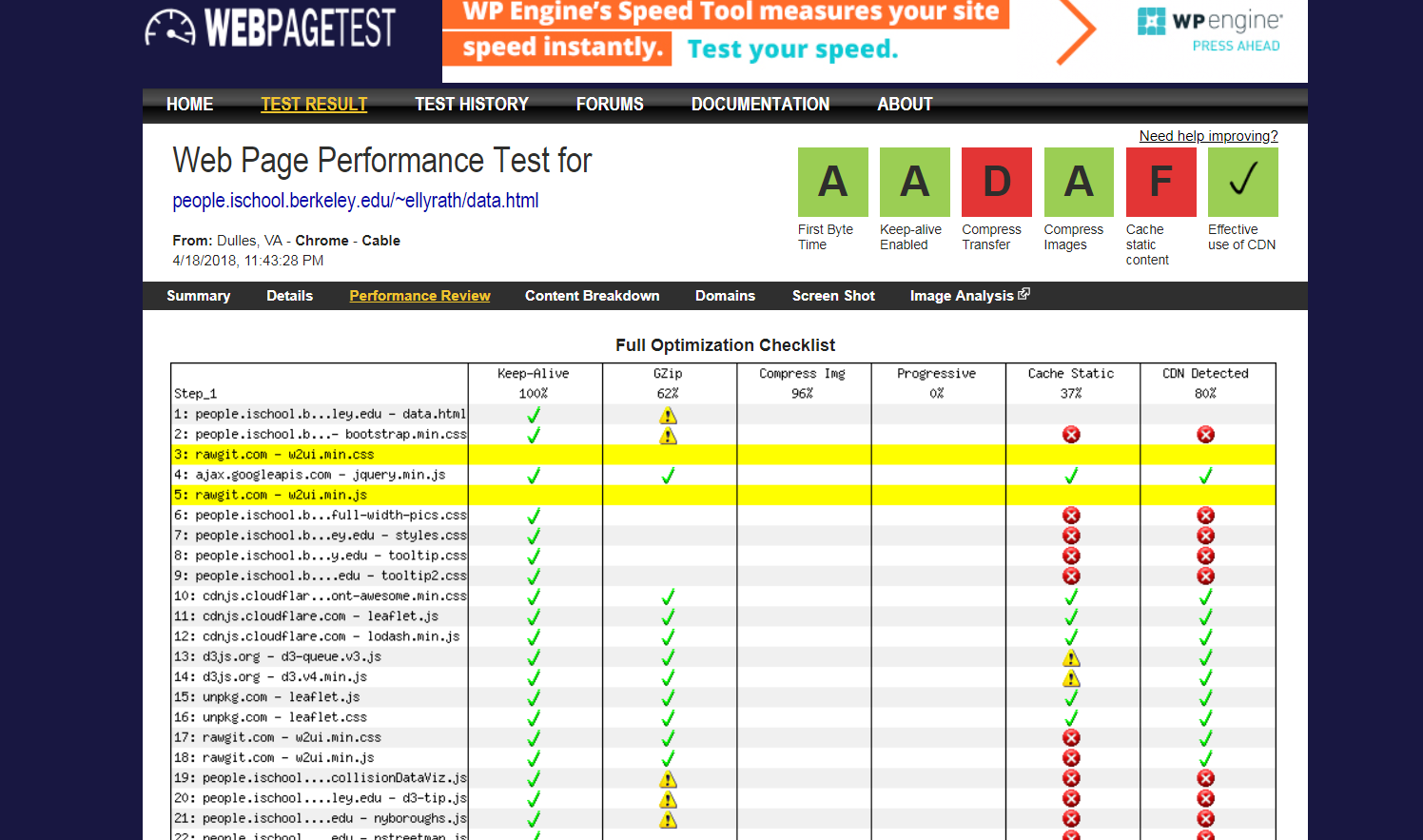


Our script loading is fast as we transfer only the extracted data meaningful for the visualization.



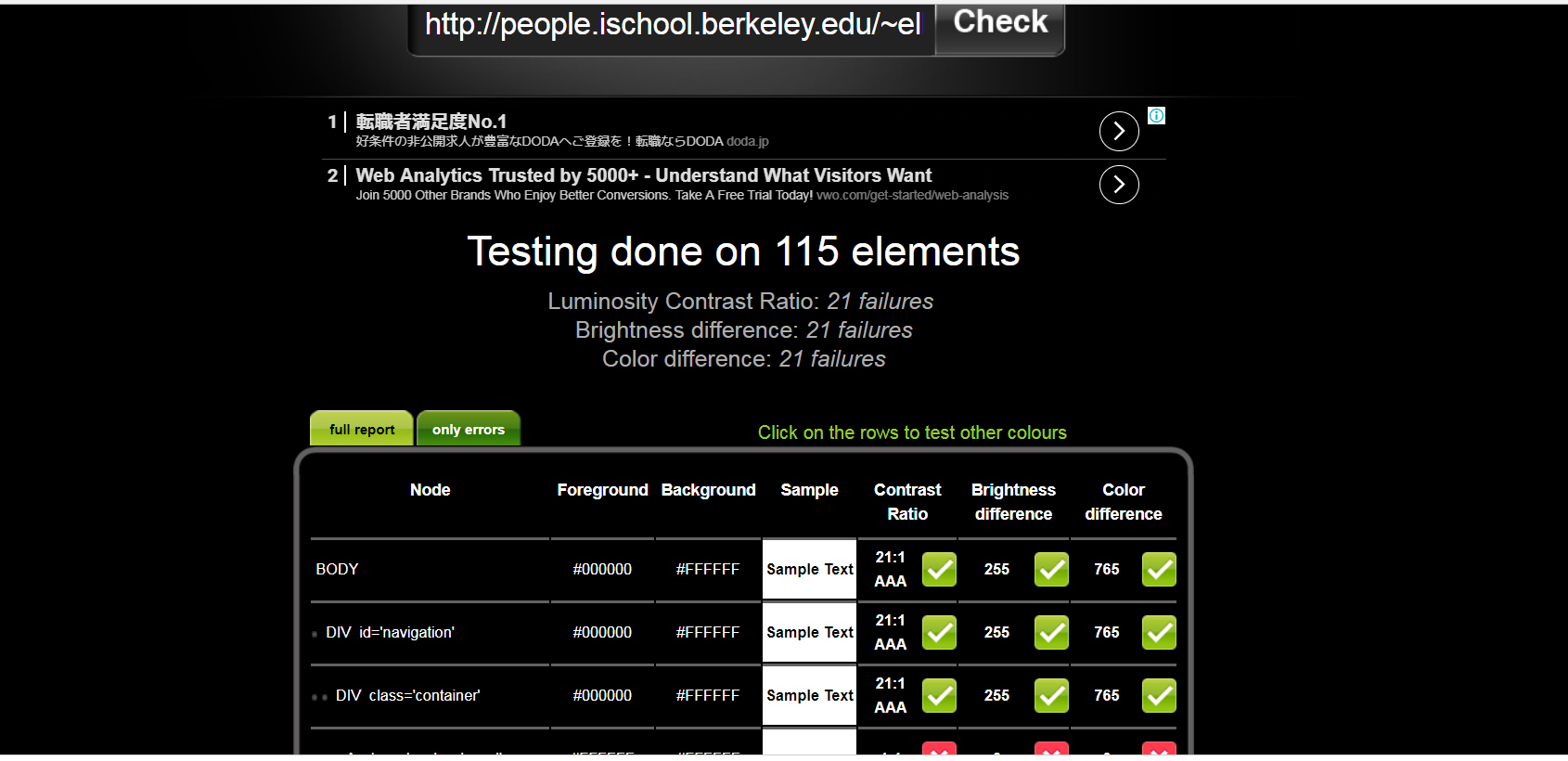
Webpage test:

We tested our web page statistics using the webpagetest website. The site gave us grading on different factors. We got an F on the static content caching. We can improve it by defining the rules in the access file of the web server.



Color test:

We did color test using the website <http://checkmycolors.com>. It complained about colors of the images that we were using but the other colors were fine.



## User Feedback Status:

## Must haves:

|  |  |  |
| --- | --- | --- |
| Feedback | Comment | Status |
| Impact of weather on accidents | Data is ready, just need to add a graph | Completed |
| Help Information | Added Usage button | Completed |
| Performance was slow | Selected top 50 incidents | Completed |
| Two users wanted historical information but one wanted current data | Already had the D3 time series. Created tableau viz where the user can chose to show all years or filter based on a specific year | Completed |
| “Vehicle 1” label on Y-axis | Removed label | Completed |
| Display season vs month information | There is a new graph that shows seasonal impacts | Completed |
| Spell out the drop down list | Spelled out the options of the drop down list | Completed |

## Should have:

|  |  |  |
| --- | --- | --- |
| Feedback | Comment | Status |
| Titles not intuitive | Fixed titles | Completed |
| Statistical information | Added data for each borough and incident | Completed |
| Introduction section before graphs | Added brief introductions | Completed |
| Filter on boroughs for dangerous intersections | Added Borough outline Maps as filters | Completed |
| The top accident prone areas in Tableau has redundant information we should remove that visualization | Removed | Completed |

## Could have:

|  |  |  |
| --- | --- | --- |
| Feedback | Comment | Status |
| Filtering based on zip codes |  | future |
| Color coding of accident spots by year | working on it | bin progress |
| Pareto of root causes with years | time limiting | future |

## Won’t have

|  |  |  |
| --- | --- | --- |
| Feedback | Comment | Status |
| Mobile application | time limiting | future |
| Integration with Navigation apps | time limiting | future |
| Filter based on address | time limiting | future |
| Contributing factor based filtering | time limiting | future |
| Adding distance measure to the maps | time limiting | future |