Web Browser Visualization
PROCESS BOOK

Brandon Tineo

INTRODUCTION

▶ This project analyzes web browser usage across the world through a variety of carefully selected visualizations. Within my project, I provide web browser data on each country for the past 4 years.

▶ The user can select what information they want the map to display in terms of date and coloring. They can also choose to see specific information about the breakdown of web browser usage in a country with either a pie chart or a bar chart

▶ The user can also choose to plot the entire history of browser usage for a particular country.

▶ Finally, the user can choose to see the map change dynamically annually, bi-annually, or monthly.

INITIAL STEPS

Background & motivation, initial sketches, data research

BACKGROUND & MOTIVATION (1/2)

▶ Over the past several years, internet usage has gone up quite a bit and along with it the variety of platforms that internet users can utilize.

▶ In working with some of my clients' websites, I became even more aware of the cross browser compatibility issues that certain libraries such as JavsScript's d3 and web-toolkit libraries, and certain animations such as CSS3 face.

▶ Running into these problems made me recall that I grew up using Internet Explorer, not because it was a particularly good browser, but rather because it was the default browser installed in my computer. Then when I started using Apple products I quickly transitioned to Safari. It wasn't until I started getting involved with computer science that I actually paid attention to other web browsers aside the one I had installed by default.

BACKGROUND & MOTIVATION (2/2)

▶ Given that internet usage varies quite a bit around the world, I began to wonder what web browsers different parts of the world used, and if there were any noticeable trends.

▶ I did some research online and found some cool information and visualizations that addressed some of my curiosity, but nothing that really caught my attention.

▶ Given that I had been looking to spend my last couple of days of the semester working on some independent coding, I decided it would be a cool project to gather some data and address these questions and more.

INITIAL SKETCHES

▶ When I first started thinking about actually visualizing this project, I wasn't sure what kind of data I would find/by using and so my sketches were very rough.

INITIAL SKETCHES

▶ Nevertheless, I knew I wanted to have some sort of a way to dynamically see the changes in the map as if time were passing.

FINDING DATA

▶ Finding and collecting data for this project was not too challenging.

▶ Whereas I first thought about looking at individual web browsers' sites/stats and trying to find data through them, I began to think that this is probably information that has already been aggregated before.

▶ In my research, I came across a website called Stat Counter which provides global statistics on a variety of topics. I was lucky to find that Stat Counter provides monthly CSV files with information on Web Browser usage in all countries in the world.

▶ I decided to download three CSV files from three different years and months to look at them and see what the data looked like.

▶ The CSV files were organized nicely, but for the purposes of my project would need to be wrangled.

NEXT STEPS

Building basic world map, data aggregation and wrangling

BUILDING A BASIC WORLD MAP

▶ As with most of my projects, I always like starting with one of the key visualizations and then going on to work on my data.

▶ The key visualization in this project was to be the map of the world.

▶ To build the world map I used the d3 library.

▶ From previous projects I had also scraped and collected data on the top 100 cities in the world and so I decided to plot these in case I could find some data on these cities.

DATA CLEANING OVERVIEW

▸ Actual values associated with each web browser were converted fro strings to floats and then passed through a percentages function to ensure we had valid data that added up to 100%.

▸ Countries that had no data were given a value of 0, and category of None so be colored accordingly in the map.

DATA AGGREGATION & WRANGLING

▶ Challenges:

  ▶ The way in which StatCounter labeled their countries did not always match the way in which the JSON filed I used to create my map labeled the countries (i.e. United States vs. United States of America).

  ▶ Different minor web browsers appear each year, so the categories in one date won't match the categories in another date.

  ▶ Some countries have no data

  ▶ The CSV file doesn't include an aggregation of the entire World's usage of web browsers for each date

  ▶ Objects cannot be sorted by values, but I needed a way to identify the Web Browsers in ascending order

DATA AGGREGATION & WRANGLING

▶ Solutions:

   ▶ To match the country names I had to write some script to parse through the country names in the CSV file match individual words within the country name (excludingwords such as "the" "of" "Republic" "Democrat" which are included in various country names). Thus my script would match United States and United States of America, in similarly match "Bolivia, Republic of" with "Bolivia".

   ▶ To resolve the different categories issue, I decided to create two categories that would serve as aggregators: "Mobile" & "Other".

      ▶ Not being able to display what categories are within the "Other" categories was something I debated about for a while.

      ▶ I also combined IEMobile with IE, Firefox with Mozilla, Ipod/Itouch with Safari.

   ▶ To gather "World Data" I iterated through all of the data and aggregated values, and then scaled them down using a percentages function I wrote.

DATA AGGREGATION & WRANGLING

▶ Solutions:

  ▶ To resolve the issue of not being able to sort objects by value, I decided to import the Underscore.js library and convert the key value pairs into a lists of lists. I then sorted this lists of lists, identified the 1st, 2nd, and 3rd most used web browser, added it to the JavsScript object I was creating and then converted the lists of lists back into a JavaScript object to be combined with the rankings object.

▶ Overview:

  ▶ All of the data for countries provided by the CSV file was matched with the data provided by the JSON file used to create the map. Categories were then calculated. Values we're then scaled to add up to 1 (percentages). The object was sorted to include the top three ranked Web Browsers.

NEXT STEPS

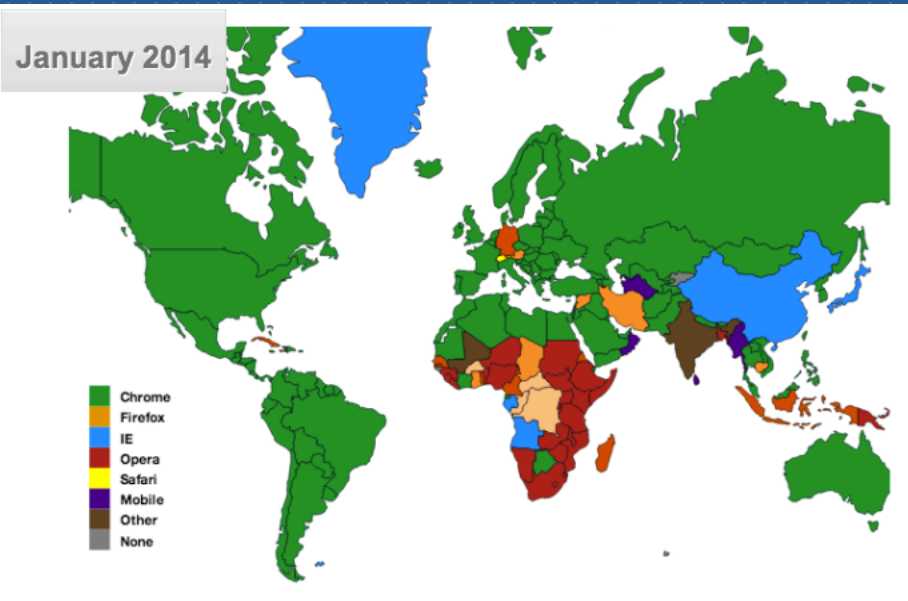Main Visualization, Side Visualizations, User Interaction

CHLOROPLETH MAP STRUGGLES

▶ One of the challenges that I faced while conceptualizing the chloropleth map toward the beginning of the project was how I was going to keep track of all the different colors and the gradient of colors these countries would reflect. Furthermore, when it came time to building the legend I was unsure of how to reflect that all shades of blue for example were IE.

▶ I started by coloring the map in a basic way.

▶ Then, to solve this problem I ended up using the colorBrewster.js library and build a scale for of each the categories in the respectively chosen color. Then I made an object where the keys were the categories and the values were one color which I selected from the gradient of colors each category was represented by
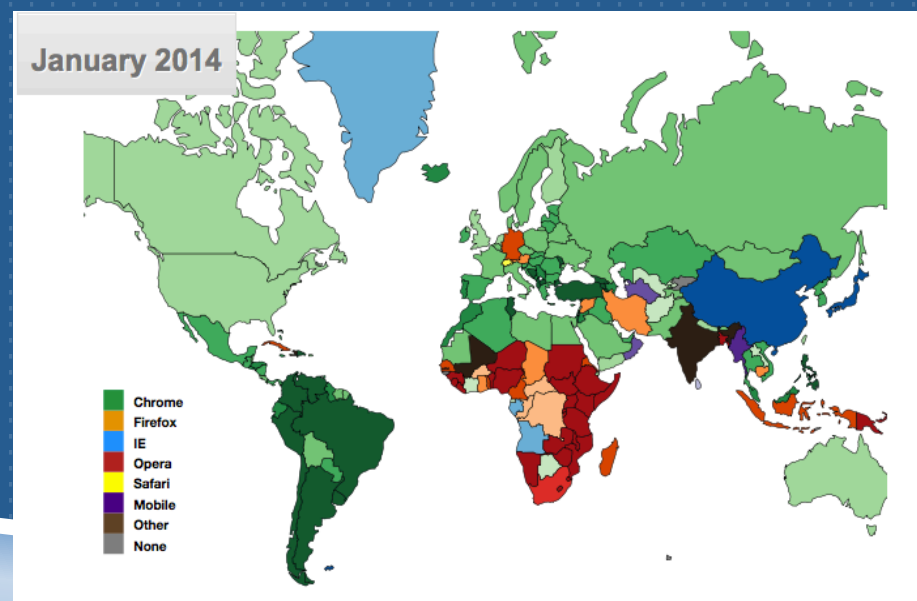
CHLOROPLETH MAP

▶ After tweaking the colors here and there, and making sure the scales were easy distinguish, I had my map ready:

▶ At this point I could definitely see some trends but I held off from analyzing the visualization until I could build more material to analyze with

▶ INITIAL COLORING

▶ FINAL COLORING

SIDE VISUALIZATIONS

▶ Pie Chart

  ▶ In thinking about the side visualizations, the first visualization that came to mind was a Pie Chart. A pier chart allows users to see the way in which a lot of bits and pieces effectively add up to one complete thing. This was thus an effective way to really show some Web Browsers compare to other web browsers.

  ▶ In building the pie chart, one of the challenges I faced was formatting the labels. I had experience calculating the hypotenues of each slide to position the labels at the edge of the slice, but I had not dealt with data where there are always really small slices. To solve this issue I removed the data where the percentage value was 0, and then wrote a function that kept track of the X and Y coordinate of each label and if any of the labels overlapped with the previous label, I moved them up by some value.
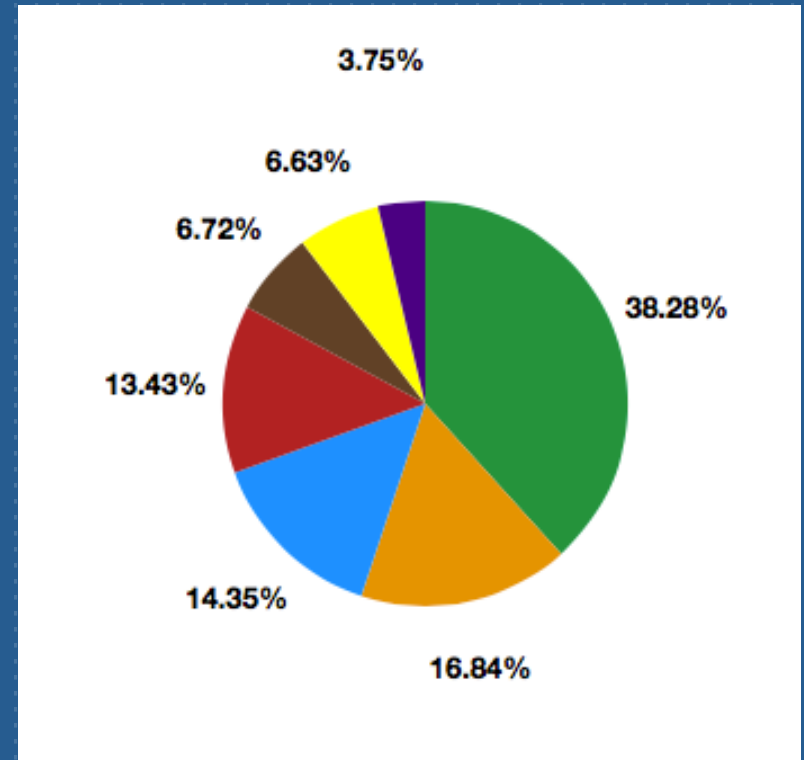
SIDE VISUALIZATIONS

▸ Double Bar Chart

 ▸ While the Pie chart gave users the ability to compare different web browser usage within one country, it didn't allow them to compare the country to the entire world, or to other countries for that matter which is something I wanted to go back to include.

 ▸ In thinking about a visualization that could do this, I thought a Bar Chart would be effective. I have experience building bar charts and believe they accurately represent the differences between two separate data sets.

 ▸ I face relatively few challenges in making the bar chart.
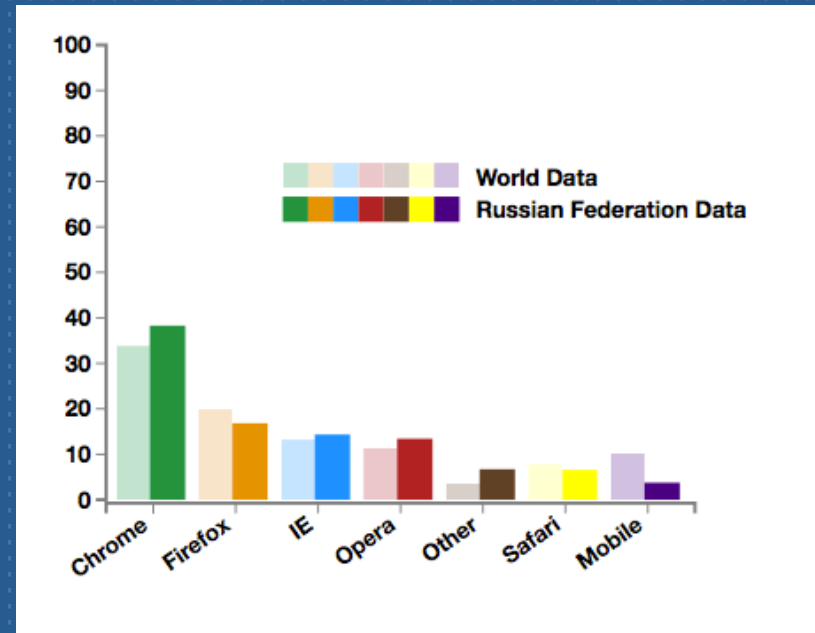
USER INTERACTION

▸ I wanted to give the user the ability to change what information they were looking at dynamically.

▸ To do this I decided to include buttons so they could alternate between the pie chart and the bar chart. I think this was a wise choice since seeing both at once is (1) overwhelming and (2) somewhat repetitive.

▸ I also decided to add a drop down menu that included the name of all the countries so users could dynamically change the information displayed by the bar chart / pie chart.

▸ I made it so the bar chart and pie chart would update dynamically on mouseover. This decision was troubling since mouseover changes are sometimes too quick and can cause users to switch the information dispalyed accidently. Thus I later considered changing it to be an on click changing.

▸ Another drop down menu was then added so users could switch how they wanted to color the map (by 1st, 2nd, or 3rd most used Web Browser)/

▸ Finally, on click of an country, I added a zoom so users could get a closer look of the country they're analyzing.

USER INTERACTION
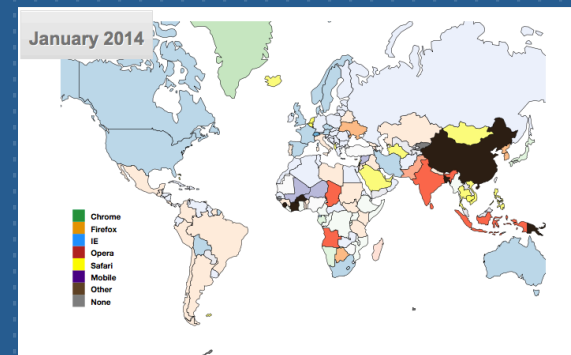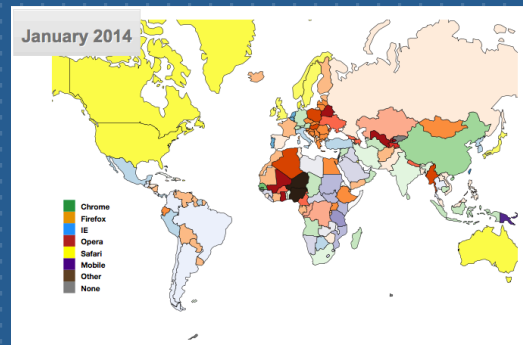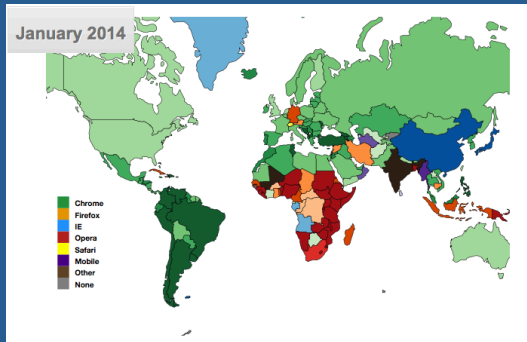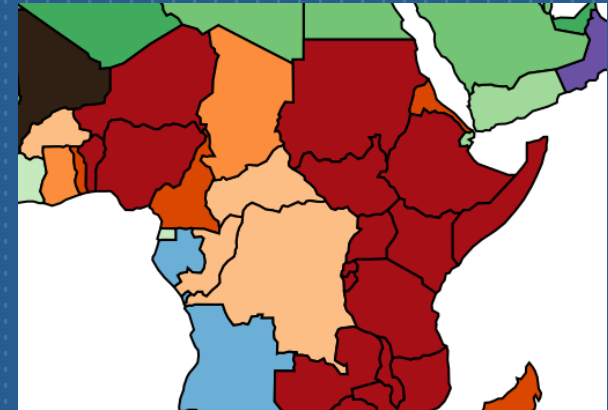
NEXT STEPS

Gathering more data, overview visualization/features, iframes, etc

MORE DATA

▶ Up until this point I had been working with one CSV file: January 2014. I had made this decision because I thought it would be easier to get down all of the key parts of the website with just own data set before scaling up.

▶ Once I had these visualizations down, I decided to read into my website CSV files with information for every month from January 2008 to May 2014.

▶ One challenge I had in this area, which I wish I could have found a better solution was reading in all of the files before initializing. I think there is perhaps a better way to do this, where the program reads in the files dynamically. Didn't look further into it though since it didn't seem to impact the speed of my website.

MORE DATA

▸ With much more data now read in, I was ready to give the user the ability to see more data.

▸ I added a drop down menu to give the user the ability to select which month from which year they wanted to see.

▸ I then decided to add button that would simulate the passage of time while dynamically changing the bar chart or pie chart and the coloring of the map. This was a really cool feature that I was excited to add.

| History By Month (Fast)	History By Month (Normal)	History By 6 Months	History By Year
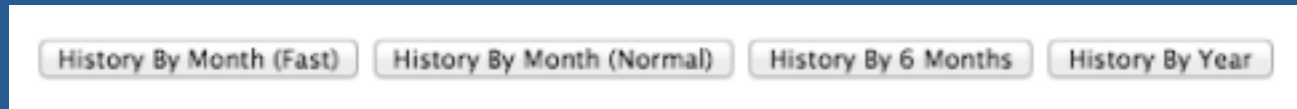
▸ Thus when users clicked this button the entire webpage would begin to change dynamically, giving a month by month breakdown of the changes in web browser usage.

MORE DATA -> OVERVIEW VISUALIZATION

▶ In adding the buttons to simulate the passage of time, I realized that something that was missing from this project as a visualization that could provide a complete overview for any given country all at once.

▶ Thus I decided to implement a line graph which plotted all of the web browser categories over the period of 4 and a half years.

▶ One of the major challenges in this area, which I still haven't figured out how to overcome is the run time. When users create the line graph, it takes about 5-7 seconds because the program is aggregating/ merging/ cleaning and compiling over 60 CSV files to create the line graph. If this project were not an independent project, I would perhaps choose to write a separate script that does all of the merging and compiling for me and reads in just one huge CSV file to make the line graph.
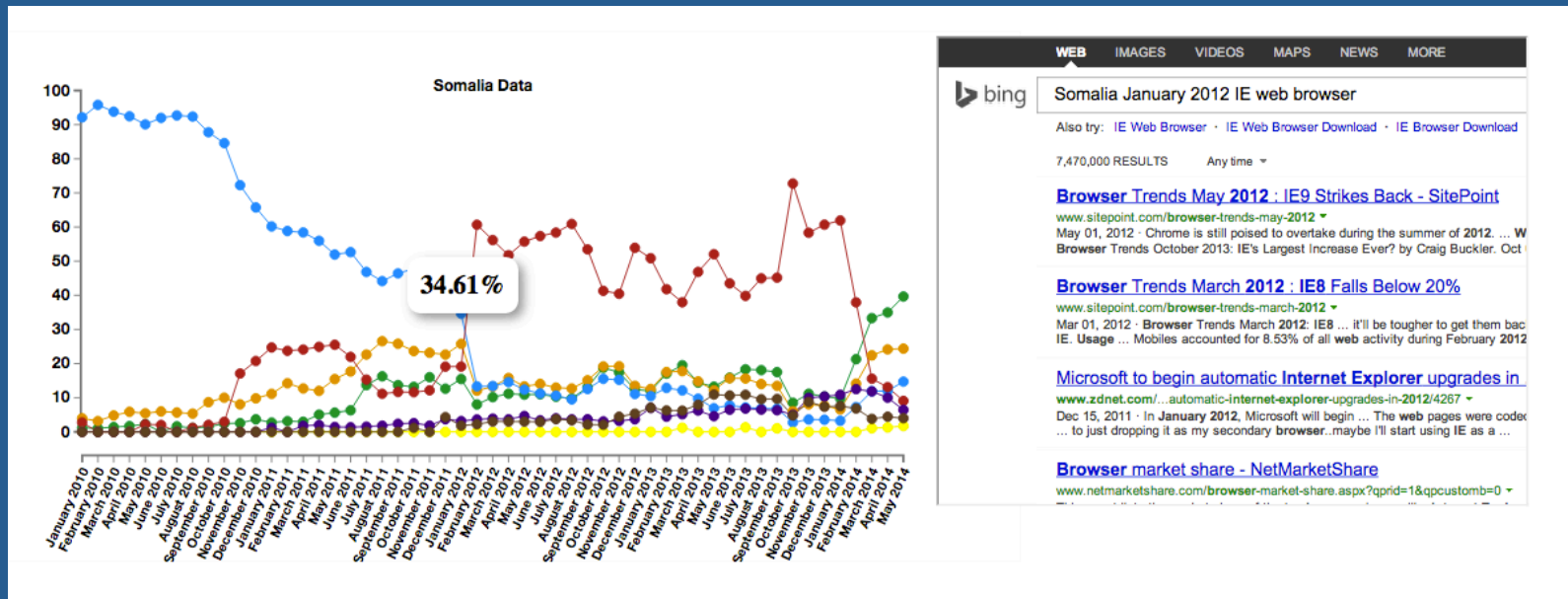
OVERVIEW VISUALIZATION

▶ After I got a very basic implementation of the line graph down, I decided to go back and make some changes to it. I added a circle to every point to make it easy to see the value

▶ I added a clear tooltip to show the exact percentage

▶ I shifted the x-axis labels to make them visible
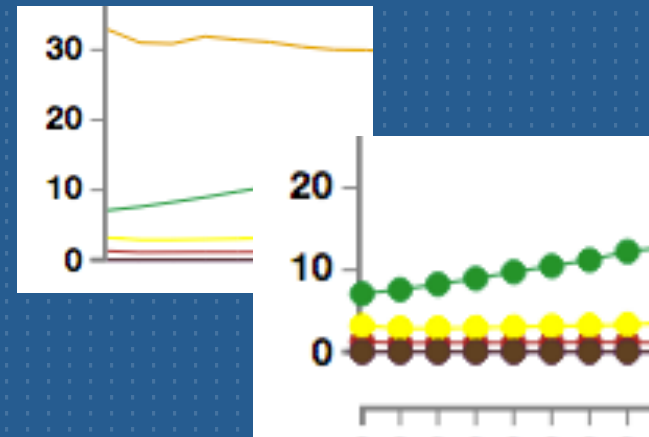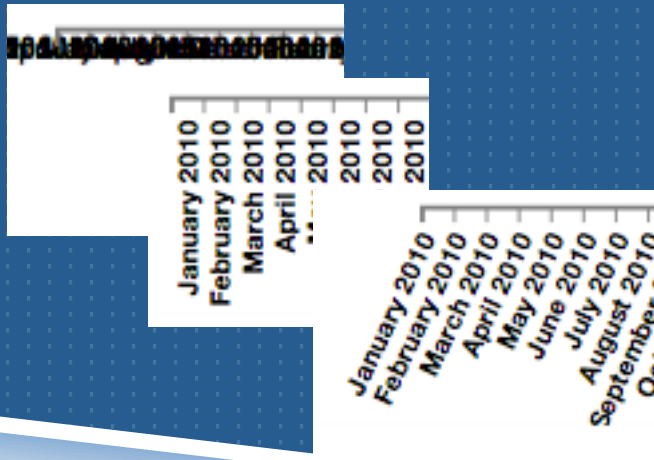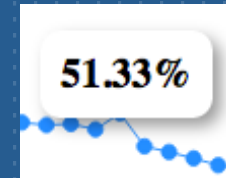
OVERVIEW VISUALIZATION -> IFRAMES

▸ As a cool additional feature I decided to add a sort of google search feature to my project whereby the user automatically submitted a search query when they click on any point on the line visualization.

STYILISTIC CHANGES

Layout format, uniform style, website appeal

STYLE CHANGES

▸ At this point most of my website was done (the key visualization, the side visualizations, and the overview visualizations as well as some cool features) and I had started to add a little bit of style to the different parts of the website but the website definitely need to work.

▸ I started by adding margins to the website which proved to be a little challenging since I didn't do it to begin with.

▸ I then moved on to format the buttons and the drop down menus in a similar way. I wanted to give them a sort of 3D between gray and white look.

▸ I wanted the bottom and the top of the website to match and be appealing but not distracting to the user. I thus decided to make the header and the footer a banner of the main web browsers being used. I changed the opacity to make these less distracting. On click, the banners refresh the page for convenience.

STYLE CHANGES
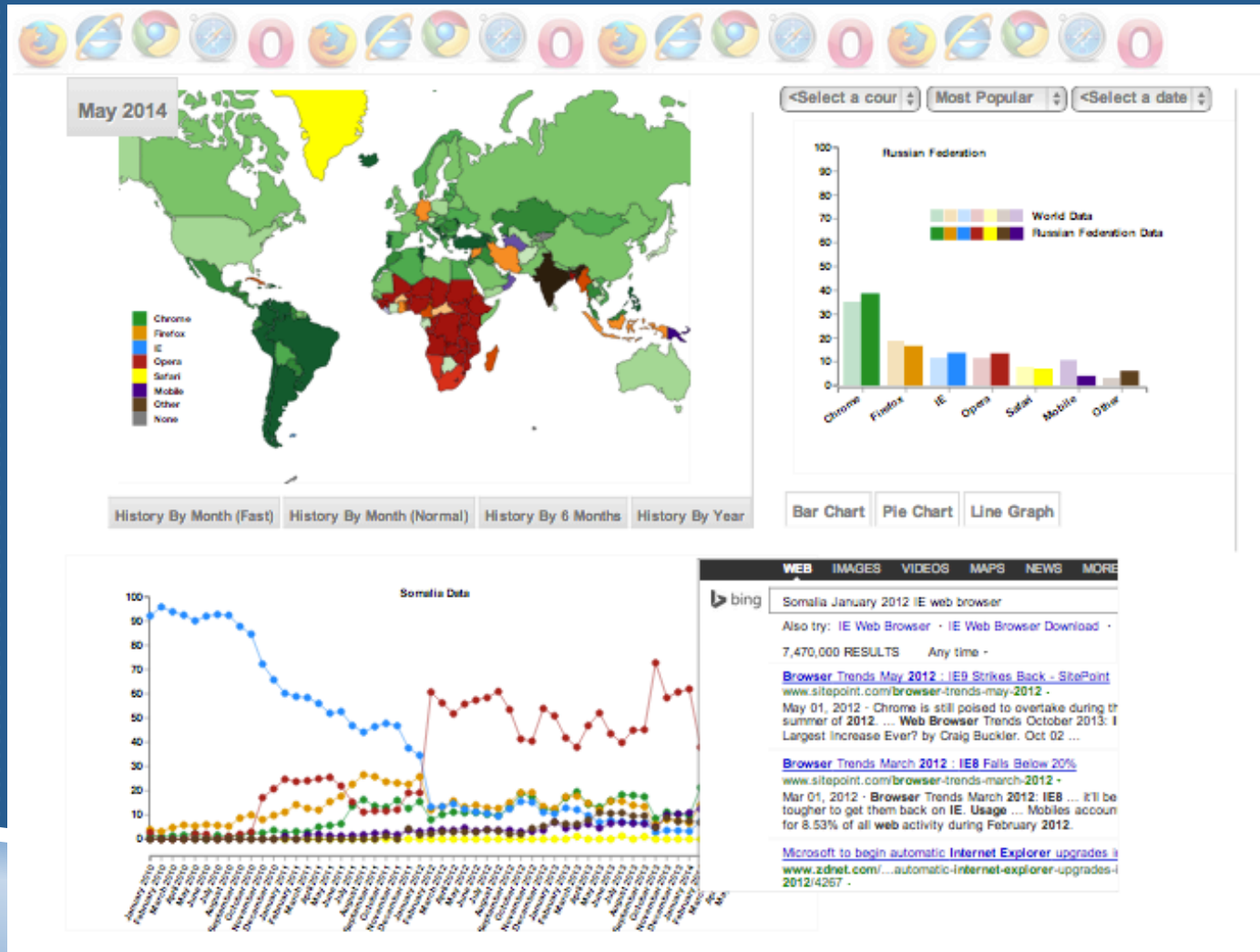
CONCLUSION

Learning from data visualizations, future work, project experience

LEARNING FROM VISUALIZATIONS

▶ The visualizations provide some key insight into the development and progression in popularity in all of the different web browsers. Primarily the visualizations all show the general decline in Internet Explorer usage as Chrome gained quite a bit of momentum starting 2013.

▶ Something particularly interesting is the popularity of Opera in Africa. It seems to follow the general pattern whereby web browser usage is often concentrated regionally.

▶ As can be seen by clicking on many of the random spike of popularity for browsers in the line graph, and reading through the search results, several spikes in popularity, especially in urban areas, seems to align with the time when new versions or updates of certain web browsers are released.

▶ Web browser usage across the world might very quite a bit, but a lot of the countries seem to follow the same pattern of decline / increasing popularity.

LEARNING FROM VISUALIZATIONS

▶ As can be seen by clicking on the options to see the 2nd and 3rd most popular web browsers, all of the coloring on the map is extremely light, whereas the coloring for the 1st most popular web browser causes pretty dark coloring for most countries.

▶ This trend shows that most countries tend to find one web browser very popular, and after that all other web browsers are relatively / equally unpopular. This fact gives insight into the notion that web browsers often monopolize within countries instead of competing with other web browsers.

FUTURE WORK

▶ Future work on web browser visualizations could find a way to automatically import/read in all available CSV files from Stat Counter every time the web browser is started. That way these visualizations would ALWAYS be up to date.

▶ More work on this project could also find a way to reveal more information about the "Other" category which currently lumps together quite a bit of information.

▶ Perhaps a good idea would be to add an option where we just see information on the key 5 web browsers: Chrome, Firefox, IE, Opera and Safari. Leaving out Other and Mobile might make the percentage values shown be in accurate so that should also be handled.

PROJECT EXPERIENCE

▶ This is one of the first extended data visualizations projects that I start, work on, and finish on my own purely for the purpose of fun, learning and exploration.

▶ It was great to come across challenges on my own and have to figure out their solutions on my own as well.

▶ I enjoyed being able to see the progress of my visualization little by little and will probably continue to tweak the website layout/appeal a little for the next couple days until it looks as professional and nice as I want it to.

▶ Data Visualization is the key to providing insight into big data and thus I look forward to finding new technique to leverage this insight.