Data Visualization

*Consumer Complaints of Bank Services*

Student report by: Tzach Ben Hamo – 316319979 Elad Talic – 315539551

Lecturer: Dr Peter Bak

University of Haifa

8/31/2016

Introduction

Our project deals with the analysis of consumer complaints of bank services using the tools of *Data Visualization*.

In our project we dealt with large data of consumer complaints and focused specifically on creating *Data Visualization* tools for exploring our data in order to obtain meaningful insights.

Data

data description:

Our goal was to find large data that we would like to study.

We've searched over the internet and after a while found the site [www.kaggle.com](http://www.kaggle.com), which contains interesting data sets. The data set of "US Consumer Finance Complaints" seemed the most interesting to explore. This data set contains rich information of more than 500,000 complaints of finance services consumers in the US.

The data is interesting since:

1. We can compare complaints on various companies.
2. We can give consumers valuable information that will help them choose the best bank for them.
3. We can give the companies information about their services and how to improve them.
4. We can give bank companies information about their competitors.

the data set includes:

* date received (dd/mm/yyyy)
* product (class of services ~ 10)
* sub product (description)
* issue (issue by categories)
* sub issue (description of consumer's issue)
* consumer complaint narrative (description)
* company's public response (description)
* company (name)
* state (initials)
* zip code (number)
* tags (tags corresponding to the consumer)
* consumer consent provided (description if yes or no)
* submitted via (Email/Fax/Referral/Web/Postal Mail/Phone)
* date sent to company(dd/mm/yyyy)
* company's response to consumer (description)
* timely response (yes/no)
* consumer disputed? (yes/no)
* complaint id (number)

we've used the following attributes:

* date received (dd/mm/yyyy)
* product (class of services ~ 10)
* company (name)
* submitted via (Fax/Web/Postal Mail/Phone)
* consumer disputed? (yes/no)

User Task

Our user task was to explore the Data Set and get valuable insights on the data through the making of *Data Visualizations*.

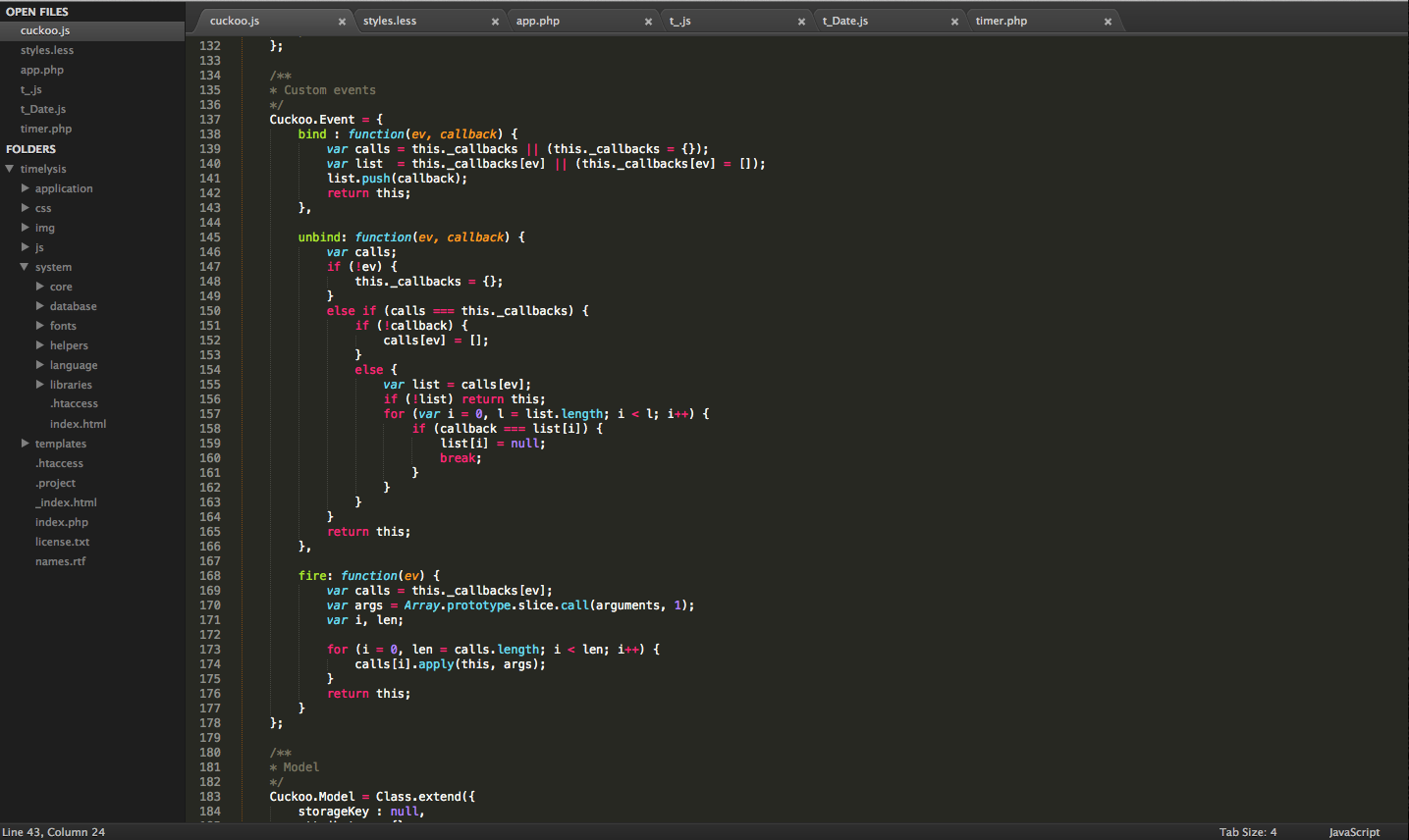
In the making of the *Data Visualizations* we mainly focused on creating dynamic visualizations that allow a high degree of freedom in order to allow the user to ask the questions he seeks to get information about.

We found special interest in answering the questions:

1. Is the way you submit your complaint influences your disputation of the service?
2. Are there points in time when the services changed their quality?
3. How large banks handle complaints and what are the differences between them?

Workspace and Software

Our *Workspace* was “sublime” editor:

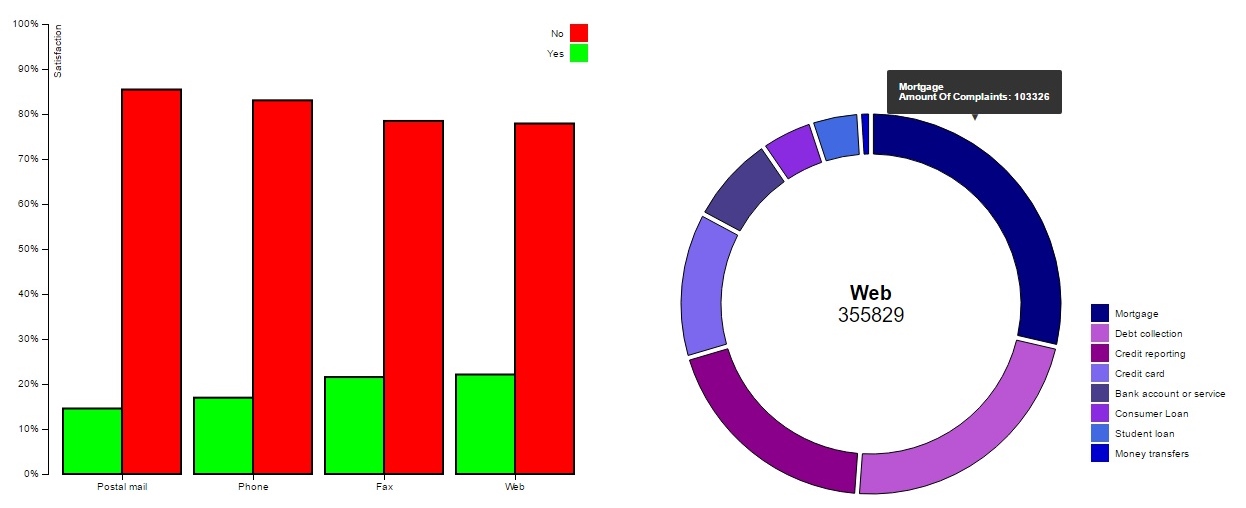


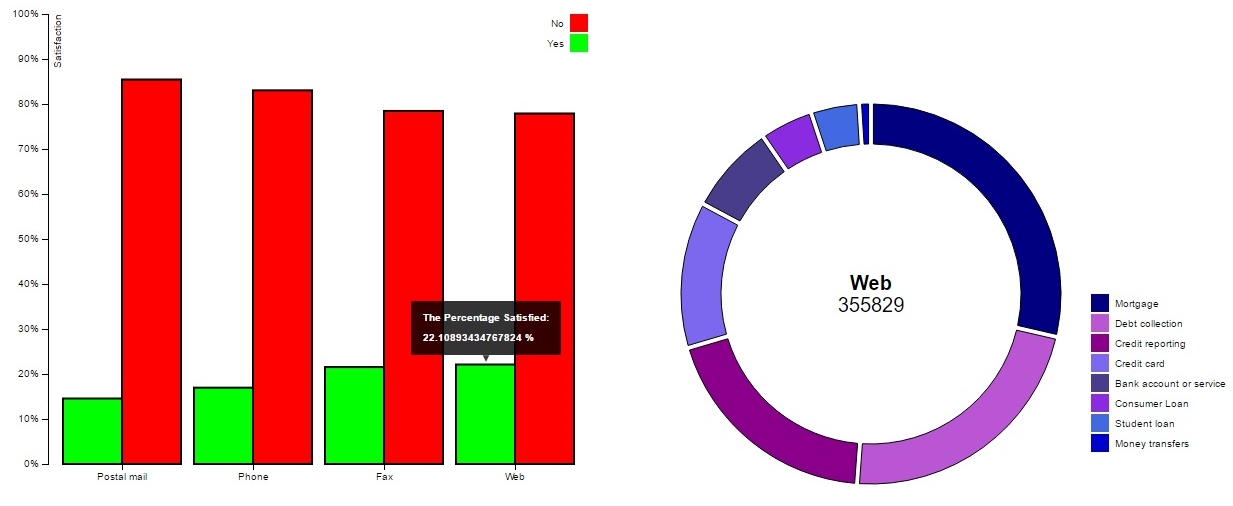
Software development tools:

* HTML
* CSS
* JavaScript
* JSON
* D3

Visualization in D3:

D3 (stands for Data Driven Documents) is a JavaScript library for producing dynamic, interactive *Data Visualizations*. D3 makes use of *SVG, HTML, CSS* in order to create *Data Visualizations*.

Submission Method Visualization



Description of Visualization:

The visualization is made of a grouped bar chart and an arc chart which interact with each other. The visualization starts with an amination which improves the user experience. The grouped bar chart shows how the submission method influences the percentage of the satisfied consumers. The tool tip provides the user the exact information he needs.

When the user wants to further investigate connections between submission method and different products he can just run the cursor on the submission method in the bar chart and look at the padded arc chart.

Visual Mapping:

* product – Color and size in the sorted arc chart.
* submitted via – X-axis.
* consumer disputed? - Y-axis and color.

Value of Visualization:

V = T + I + E + C

Time – The visualization saves time on shallow questions. As the questions get more complex the visualization is less useful in saving time.

Insights – The visualization gives the possibility to get insights about questions concerning correlation between submission way and complaint distribution by product.

Essence – The grouped bar chart definitely convey essence about how submission method affects disputation.

Confidence – The visualization increases the confidence slightly when showing the amount of complaints in the padded arc chart.

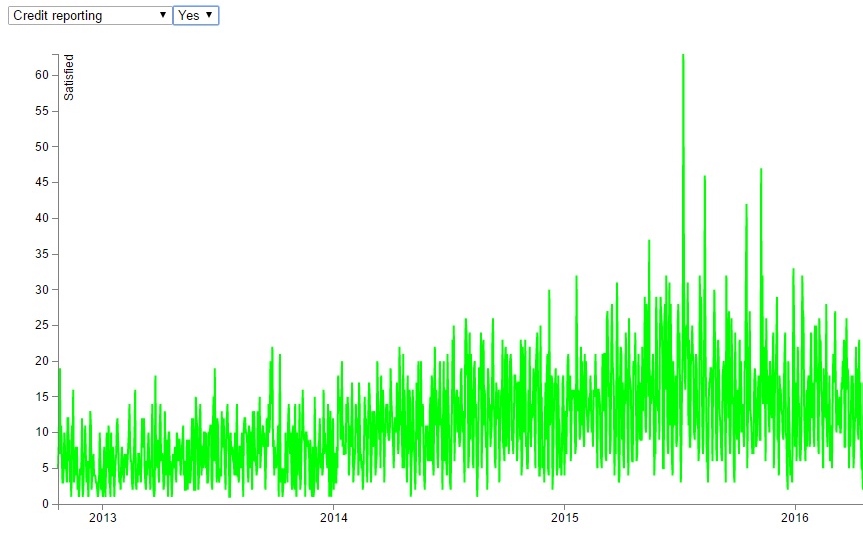
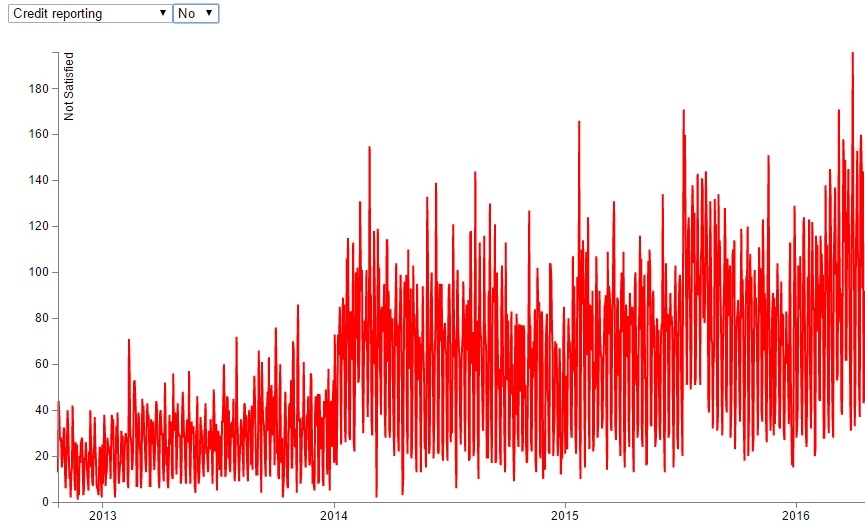
Pros and Cons of Visualization:

Pros:

* The visualization is easy to learn and use.
* The tooltip makes it easy to get information from the bars and arcs.
* Sorting the bars in the chart makes it easier to compare between different submission methods.

Cons:

* The visualization is simple and cannot answer a large variety of questions, especially it cannot answer complex questions.
* Grouping the bar chart is not informative.
* Arc chart doesn’t handle well small variables.

Disputed by Product Visualization

Description of Visualization:

The visualization is a line chart that gives the option to look at the disputation for each product separately. The visualization starts with an animation and has dynamic update which also improves the user experience. The main reason we chose this visualization is in order to give the option to analyze the services for each product by time.

The major problem in our visualization is the overplotting. Clustering per month may give better result in giving more consistent, easy to read visualization but lacks the accuracy in telling when changes actually happened.

Visual Mapping:

* product – Picked in order to visual map
* date received – X-axis
* consumer disputed? Picked in order to map. Mapped into Y-axis with color.

Value of Visualization:

V = T + I + E + C

Time – The visualization doesn't save a lot of time since it represents complex information that takes time to analyze.

Insights – The user can look for changes in services over time and find points in time that change occurred. When these points are known one can conject what made the service better/worse.

Essence – When big changes occur the visualization is definitely conveying. When only small changes occur it is definitely not conveying.

Confidence – Insights that have been deducted from the line chart are believable. Highly reported products give more consistent and confident results.

Pros and Cons of Visualization:

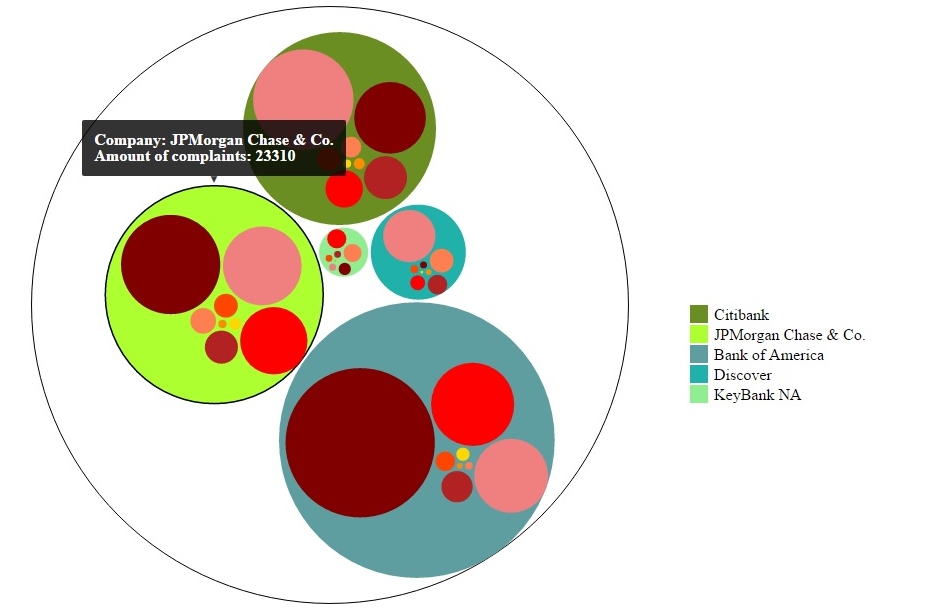
Pros:

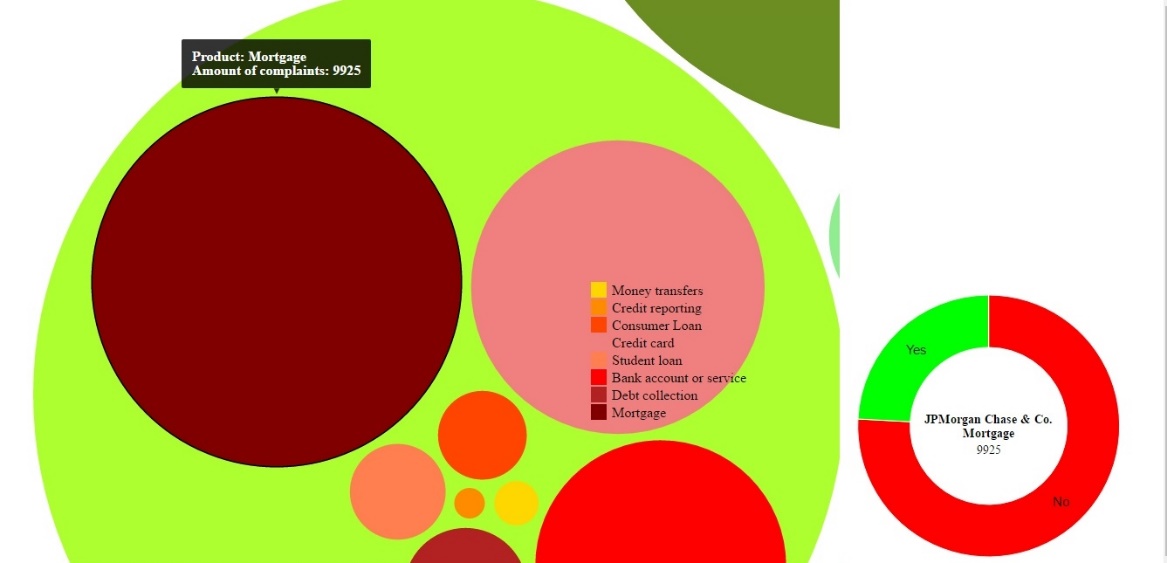
* The line chart is simple and easy to use.
* The line chart gathers large amount of information.
* Insights that regards to big differences between disputed and not disputed are highly believable
* Sharp changes in disputation can be identified quickly.

Cons:

* Line charts are fairly simple
* There is an overplotting in the visualization.
* Insights are not consistent when the product has small amount of complaints and when the differences between charts are small.

Largest Banks Comparison Visualization





Description of Visualization:

The goal of this visualization is to compare complaints on the 5 largest banks in the data. We chose a circle packing visualization since it gives the user intuition about how many complaints each bank gets, how the complaints are distributed between products and how the bank handles the complaints.

We've added an interactive arc chart for bank products for the purpose that the user can learn how different banks deal with complaints on different products.

Visual Mapping:

* product – Mapped into the packed circle, size of specific packed circle, color in packed circle.
* company – Mapped into the large circles, size of circle, color of circle.
* consumer disputed? Mapped into size in the arc chart and color.

Value of Visualization:

V = T + I + E + C

Time – The visualization saves time in providing the user specific information and to make him "feel" how different stats scale.

Insights – The user can get insights about how those banks differ, how each bank handles complaints on different products and more…

Essence – The visualization is scaled good and colorful and gives the user memorable images to think about. Moreover, the information that it gives is precise and that should convey the user.

Confidence – One can take away and feel confident about insights he deduces from the visualization since each circle consists from hundreds and thousands of complaints.

Pros and Cons of Visualization:

Pros:

* The visualization gives information that is not easy to get simply from the data, this means not only the amounts but the "feel" of the scaling as well.
* The arc chart allows the user to get specific information of a product while the tool-tip gives the user the exact amounts so we can have the right scaling.
* The packed circles are colored and thus the user can identify differences in product complaints between companies.

Cons:

* It takes time to understand and get used to the visualization.
* Exact comparison between companies, though possible, is hard and takes time.

Resources

* <https://github.com/d3/d3>
* Introduction to D3 - <http://dealloc.me/2011/06/24/d3-is-not-a-graphing-library/>
* D3 Tutorials - <http://alignedleft.com/tutorials/d3/>
* D3 API - <https://github.com/d3/d3/blob/master/API.md>
* Let’s Make a Bar Chart, III - <https://bost.ocks.org/mike/bar/3/>
* groped bar chart - <http://bl.ocks.org/mbostock/3887051>
* Line Chart - <https://bl.ocks.org/mbostock/3883245>
* Zoomable Circle Packing – <http://bl.ocks.org/mbostock/7607535>