

# Understanding Data: A Comparison of Information Visualization Tools and Techniques

Prashanth Vajjhala

**Abstract** - This paper seeks to evaluate data analysis from an information visualization point of view. A dataset from The Meteorological Society is used to evaluate two data visualization tools, and different techniques from these tools. The tools used for the visualization are IBM Many Eyes and Google Fusion Tables. The analysis spans several types of visualizations – from Bar Charts to Geospatial Mapping – to determine which works best for this dataset. Also considered is the usability of the software platform as well as the impact of the visualizations on the user.

**Index Terms** – Information Visualization, Summarization, Visualization Tools, Data Analysis

---

## 1. INTRODUCTION

The field of Information Visualization is constantly growing and various new visualization tools are now used by a large variety of people for various purposes. Today, it is not only scientists and specialists who take advantage of data visualization techniques to interpret complex data, but ordinary people as well. The availability and accessibility of useful visualization tools make it easier for the majority of people to build various visualizations and interpret them accordingly.

This paper describes and compares visualization techniques and outcomes produced by two such visualization tools that are available for free. These tools are Many Eyes and Fusion Tables. Many Eyes is a social visualization tool that allows people not only create and view visualizations but also introduces a social aspect where people can discuss, share and analyse their visual data with others. Fusion Tables helps in the visualization of geo-spatial data. The ability to collaborate on the web while using these tools adds a great value to the core visualization functionality these tools have to offer. While both the tools provide an ability to graph and visualize their own data, Many Eyes has a stronger emphasis on visual techniques which help to find patterns. The ability to visualize spatial data using Many Eyes is limited at present. Both the tools are intended to be accessed and used by people who do not require any extensive background in data processing or visualization experience.

Many Eyes is a freely available social visualization tool developed at IBM's Visual Communication Lab by Martin Wattenberg and Fernanda Viegas in Cambridge, MA. This application is designed to be an easy to use social tool for sharing visualizations in a collaborative environment. The basic functionality of Many Eyes is to upload data, create visualizations based on selected and uploaded data sets, and to share comments and experiences on both the data sets and resulting visualizations with others. Many Eyes uses a wide variety of graphing techniques to visualize data such as world maps, bar graphs, histogram, pie charts, tree-maps and many others. The idea behind Many Eyes is to "democratize"[1] visualization techniques by allowing broad and easy accessibility. This tool does not compete with high end scientific tools that are used in the scientific and data processing communities, since Many Eyes tries to strike a balance between providing powerful visualization techniques and a non-technical audience.

Fusion Tables is an experimental data visualization web application to gather, visualize, and share larger data tables built by the Google Research team. One of the biggest features of Fusion Tables is the ability to visualize large tabular data online. Fusion Tables also offers API for developers instantly. Many types of techniques such as charts, maps, network graphs are offered for visualization purposes. Another feature is the ability to merge data

with other people's data as well as publicly available data for better visualizations.

This paper will analyse and compare creating visualizations and comparing these tools in detail, concentrating on the creation of visualizations. Visualizations using each tool are created separately while concentrating on major functionalities and features that each tool offers. Then the paper is subdivided into 5 (4.1-4.5) sections in which these tools are compared side by side on the following aspects: Functionality, Data handling, Graphing features and visualizations, User Experience and Web Technologies, Collaboration Techniques and features and Usability. In this paper, the major advantages and disadvantages of generating visualizations by using the tools are analysed and possible future improvements are discussed.

## 2. OVERVIEW OF MANY EYES

"Many eyes" is an online collaboration community built by IBM's Visual Communication lab. The website allows users to share data and visualizations for collaborative data analysis. Users' main functions on the website include the following:

- Upload new data
- Create visualizations
- Post discussion comments to either displayed data or created visualizations.

All visualizations created within the Many Eyes website are public and any users can view and comment on them.

Exploring the visualization page allows the user to view and search for all visualizations created within the web site. By default, the user has the ability to sort the visualization by either a date or a rating. Within the website, a user also has the capability to search for a specific visualization. While the selected visualization is being displayed, the following actions are available:

- Analyse the visualization using interactive image
- View data set behind the visualization
- Comment and rate the visualization
- Share the visualization with others
- Add the visualization to a specific topic hub
- Add the visualization to the watch list to keep track of what other users will say in the future about the visualization or the data behind it.

The ability to post and view comments behind a visualization or a data set is the most important feature of "Many eyes". While discussing a specific visualization, users have the ability to give possible explanations for data trends which can lead to a deeper data analysis.

Another feature available in "Many Eyes" is the ability to upload a new data set. This feature is only allowed for registered users. Users have the ability to upload data in Excel or tab delimited format. Once the data is uploaded, the user has the ability to create new

visualizations. However, there are some limitations for deleting a newly created dataset. As long as no discussions have been created for the newly created dataset and visualization, the user who created it can delete it. On the other hand, if a discussion has been initiated, only the site's administrator has the power to delete the data.

### 3. OVERVIEW OF GOOGLE FUSION TABLES

Similar to Many Eyes, users have the capability to either upload new data for visualization, or use publicly available data. Users can upload spreadsheets, delimited text files, and Keyhole Markup Language files, or use a file from Google Spreadsheet, or create a new Fusion Table. Fusion Table has two looks currently. While the Classic look is more traditional, the new look offers several new features such as viewing the data in the form of card layout or row layout. The types of visualization available are line, bar, scatter, timeline, map, pie and intensity map.

The data when visualized on a map is interactive. Clicking on any of the points on the map will display relevant information about the specific data point. Users have the ability to decide what information to showcase on the map in this manner. Once the visualization is created, users can view the map as a terrain or from a satellite view. The style of the map can also be configured, where there are options to select different markers, as well as change the colours of the lines etc.

Heatmaps use the Maps API Heatmap layer, so they pull data into the browser before displaying it. This is unlike Fusion Tables feature maps that render map tiles on the server. Because of this, only the first 1,000 rows are mapped. Finally, the generated visualizations can be published anywhere. The Publish dialog includes the code to produce the visualization. Unlike Many Eyes, people using Fusion Tables discuss about any questions on Stackoverflow, or on the Fusion Tables API Community Group. There are no features to comment on datasets or visualizations created unless they are specifically shared or asked to discuss about.

### 4. COMPARISON OF MANY EYES AND FUSION TABLES

#### 4.1 Functionality

The purpose of Many Eyes is to navigate the continuously growing collections of visualizations, data sets, and comments, to browse visualizations created by others collaborate on them and learn new aspects of either custom or publicly available datasets. On the other hand, the purpose of Fusion Tables is to easily visualize larger data sets, geo-spatial data analysis, and utilize the results for embedding in web pages.

Many Eyes and Fusion Tables inherit their design from a number of participation-oriented web sites. The core functionality on the sites is to upload data, and construct visualizations. The data set browsing page of Many Eyes shows a table with the latest data sets that have been uploaded to the site. The table also displays metadata about each data set: keywords, source, the contributor's username, size in bytes and number of rows, date of contribution and links to existing visualizations with that data set. Every day the home page of Many Eyes website features visualizations that typically reflect latest events or good models of visualization usage. All the visualizations and data sets have an attached discussion forum where users can share comments and links to other web pages. Users can browse for public data through Fusion Tables that can be searched using keywords. Details about a publicly available dataset can be viewed which include details about the description of the dataset, the creators, its numeric ID, among others. Discussions concerning the dataset can be viewed directly from the table.

#### 4.2 Data Handling

Uploading a data set in Many Eyes, is different from Fusion Tables. Many Eyes is more interactive and flexible, extracting the

data and building the table with what seems to be an artificial intuition. While Many Eyes requires data to be pasted on the webpage, Fusion Tables has upload options which is more user friendly.

For creating visualizations in this case, data from the Meteoritical Society about meteorite landings was used. The data consisted of the following fields: Name, Type of meteorite, Class, Mass, Year, ID, Fell/found, Latitude and Longitude. For better visualization purposes, missing data points were not included in the analysis. As there were more than 30,000 points to visualize, it was felt that removal of missing records wouldn't affect the outcome significantly in this case.

The data was uploaded on to Many Eyes using the Upload Data page, and Many Eyes recognized the content. As was mentioned earlier, Many Eyes recognizes data with an amazing precision and displays it properly. After the data was verified, metadata needs to be entered so that the dataset and visualization can be easily recognized once uploaded. Finally, a page is displayed describing the dataset. [Figure 1]

Data can be easily uploaded to Google Fusion Tables using the "Create" function available on the website. Fusion Tables can also be created from Google Drive, which was done in this case. The data was uploaded in the form of a .csv file. Once the file is uploaded, it is specified by the user that the column names are in the 1st row. Depending on the row number specified, the rows before that are ignored. The name of the table is then given, as well as options to export the table, whom to attribute it to, attribution link, and the description of the data.

Since users can upload their data to both sites they benefit from the fact that they can visualize their own data. Many Eyes also brings a collective benefit: other users can augment the uploaded data by adding additional, related information. There's also an opportunity for research to understand user demand of the types of data that people visualize the most.

Both sites face interesting constraints on their design. Since data is uploaded to the site by end users, the data model provided by both competitors must be easily understandable, with a format that is appropriate for non-programmers. At the same time, the format must be flexible enough to express the data structures used by visualizations. Many Eyes is impressive as it handles this task with a better approach, offering users visualizations such as tree-maps and other complex graph layout algorithms. In addition to constraints on the data model, both sites also handle changes to data. Data sets and their visualizations are to be commented on and discussed. Traditional visualization methods such as bar charts, pie, scatter plots, are available on Fusion Tables, but the impressive aspect is the ability to visualize maps easily.

Another aspect of data handling is redistribution of dataset that is often necessary in preparing for visualization. Both sites use table structure as a model. Models define data types of columns to be either textual or numeric, with Fusion Tables being able to handle other data types – such as Date/Time and Location. In this research it is found that Many Eyes does not handle Date data type correctly. Both sites also support data that comes in the form of unstructured text. Each data set, whether a table or unstructured text, is associated with a collection of metadata. Some metadata, such as a (required) title, the source of the data and a paragraph-length description is provided by the user. Other metadata is automatically set by the systems, such as creation date.

Both sites include a considerable amount of educational materials to aid in the process of data preparation and uploading. Metadata attached to the datasets is then used in the visualization procedure to describe graphs and define their attributes.

#### 4.3 Comparison of Graphing Features

Fusion Tables offers users the following visualization techniques: map, intensity map, line, bar, pie, scatter plot, and timeline charts. Fusion Tables allows users to change the colours of the lines, the view of the map, which is a necessary feature to keep all the graphs from looking the same.

Unlike Fusion Tables, Many Eyes offers a variety of visualization types such as traditional charts, tree-maps, world maps, block, bubble, and network diagrams. All of the visualizations are interactive. Many Eyes offers many visualization types, divided into 6 categories:

- Text analysis – offering visualizations: tag clouds and word trees
- Comparison of set of values: bar charts, block histograms, and bubble charts
- Data relationship: matrix charts, network diagrams and scatterplots
- Parts of the whole: pie charts, tree-maps and tree-map for comparison
- Maps: country maps and world map
- Changes of data over time: line graphs, stack graphs and stack graphs for categories.

All of Many Eyes’ visualizations are interactive, allowing the user to query exact numbers, zoom, change graph parameters, select different category to visualize. The resulting graphs look much nicer on Many Eyes, and are also easier to read.

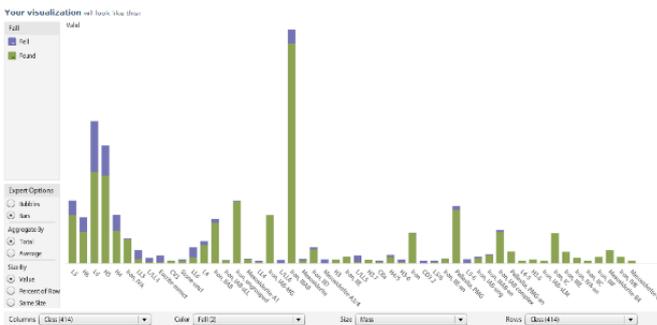
*The data in Many Eyes:*

First, a “bar chart” was used to visualize the data. Many Eyes understood the columns to be better presented in a bar chart form and gave options to change the visualization parameters.



(Figure 2 in the Appendix is an image in higher resolution)

One of the most common types of real world data is simply a list of labelled numbers. Such lists are often easily displayed via standard bar charts or histograms, but these methods can run into trouble for highly skewed distributions. One of the unique visualization types in Many Eyes is a Matrix chart. A matrix chart summarizes a multidimensional data set in a grid. The matrix chart divides the screen into a grid. Rows represent the values in one text column and columns represent another text column. Each cell then shows a circle or bar that represents the value for its row/column combination.

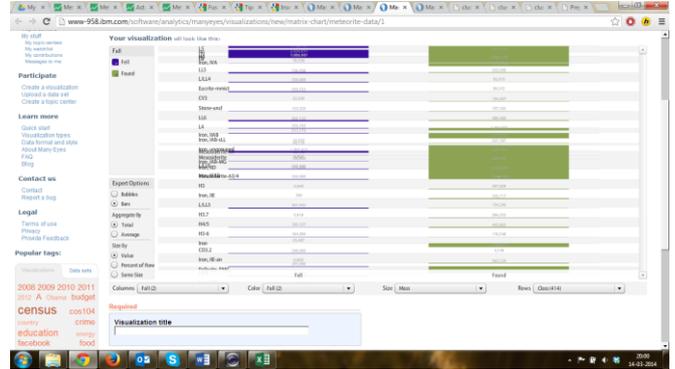


(Figure 3 in the Appendix is an image in higher resolution)

While this form of visualization for a large dataset is generally not a good option, in the case of this dataset, some clear indications can be

made. In the chart depicted, the x-axis consists of the Class of the meteorite. For each of the classes, the bars represent the meteorites that fell, and that were found. Clicking over any of the bars displays information of the Type of the meteorite among other details.

The Matrix chart has options to customize the view of the rows, columns etc. to better visualize the data, and gather any insights.



(Figure 4 in the Appendix is an image in higher resolution)

By selecting the categories to be graphed the changes in the visualization are clearly noticeable.

For data that has hierarchical structure Many Eyes offers “stack graph for categories” visualization. Clicking on elements in the tree control filters the stack graph to show only time series at that level of hierarchy.

For unstructured text Many Eyes offers “Tag Clouds” visualization, which can quickly give the user an overview of the items in the text. It seems that Many Eyes is able to measure the frequency of two word tags in the text and able to dynamically filter the tag cloud by entering query strings in a text box. Tag clouds seem to be an attraction to a set of users, whose interest is primarily in textual data instead of numbers.

To visualize a dataset Fusion Tables and Many Eyes have to match data to the visualization technique. For Many Eyes this is a more difficult task, since the number of techniques is larger. As expected the same data cannot be displayed by all visualizations. The same data will be displayed from different perspectives if two different graphing techniques are selected. For example, a tree map requires a hierarchy to be defined in the data, while a bar graph only requires a set of textual and numeric columns – one to define the size of the bar and another one for the label.

Although both the tools provide graph visualizations of a user uploaded file they differ in the graphing algorithms. Many Eyes uses a force directed graph drawing algorithm in a zoom able user interface. [1] The layout algorithms are implemented in java and run on the client machine, requiring less input from the potentially overloaded server, thus producing layouts in a matter of one or two seconds. Although many different data formats for graphs exist, Many Eyes uses “simple edge table”. [1] Although not the most efficient storage format it proves to be both highly understandable and suitable with table-oriented design.

The main difference between Many Eyes and Fusion Tables in graphing features is that, instead of providing static business graphics, Many Eyes offers a number of interactive visualizations of user’s data. This interactivity allows users to drill down into details, view the data from different perspectives and generally makes the visualizations fun to use.

*The data in Fusion Tables:*

As mentioned earlier, Google Fusion Tables has a number of traditional visualization options. However, the one particular type of visualization, which is unique is the map visualization. Though Many Eyes has an option to visualize maps, it is limited currently, and requires region names, while Fusion Tables works well with either names, addresses, or coordinates data. This dataset has the

longitude and latitude details of the places where the meteorites landed, and thus Fusion Tables is used to visualize.



(Figure 5 in the Appendix is an image in higher resolution)

There are two different layouts to the map, one is the simple terrain view, and the other, a satellite view. The views in the map can be configured from the options. Clicking on any of the icons on the map will display details about the icon. In the case of this dataset, the details displayed are the Name, Type, Class, Mass, Year, and whether the meteorite fell or was found.

Another form of visualization is a heat map. In the case of this dataset, the visualization shows the meteorite landings with regions of high intensity in a darker colour.



(Figure 6 in the Appendix is an image in higher resolution)

Fusion Tables also has a timeline graph, which isn't available on Many Eyes.



(Figure 7 in the Appendix is an image in higher resolution)

In terms of capabilities, Many Eyes is a clear winner. Better User interface, more flexibility, better user communication and more graphing options – all point in favour of Many Eyes. The only advantage that Fusion Tables has over Many Eyes in terms of visualization is the ability to visualize geo-spatial data easily, and also time series data.

#### 4.4 Web Technologies and User Experience

Visualizations in Fusion Tables and Many Eyes are delivered using different technologies, and thus they offer different user experiences while using these tools. Many Eyes is built on Java while Fusion Tables uses SQL. The capabilities of the two tools and produced visualizations greatly depend on the power and capabilities of two technologies. The advantage with SQL is that the processing times are much quicker when compared to Java, when working with large datasets. Many Eyes is built with the more stable and robust Java technology that has been proven to deliver solid and reliable applications. One of the weak points of using Java, is that the regular updates of Java can cause security issues.

#### 4.5 User Oriented Approach

Both the tools make visualization a social activity. Besides uploading and visualizing the data, a huge part of the sites' functionality lies in users' ability to collaborate on their work. This broadens horizons of visualization and brings out the "human" aspect of human-computer interaction. Both the sites have forums where users can share their experiences. Fusion Tables has a large support community, including a Google group and also large number of users on Stackoverflow where users discuss various aspects of the tool. An interesting feature of Fusion Tables is the code that generates the visualization is available and can be reproduced. Many Eyes also features a functionality that eases embedding the visualization in various blogs and forums.

A significant drawback on Many Eyes is that once a dataset has been uploaded, it cannot be edited. However, Fusion Tables has many options to edit data. Users can merge existing datasets with new datasets, create additional rows or columns, among many other features to play with the data.

## 5. CONCLUSION

In this paper the visualizations generated by Many Eyes and Google Fusion Tables are compared and discussed. Both the tools have their pros and cons. While Fusion Tables provides nice visualizations and ways to share them, there are no dynamic visualization options, nor any new visualization techniques that can be implemented. On one hand it allows for easier sharing in terms of the API, but on the other hand it greatly limits user interactions with visualizations. Fusion Tables also needs to incorporate more techniques such as Bubble Charts that are available on Many Eyes.

Many Eyes is more organized and dynamic. Many Eyes enables users to interact with data sets, create new visualizations and modify current ones. It also provides different types of visualizations such as tree maps, bubble charts, network diagrams and many others. This makes Many Eyes tool very appealing for users who are interested in a variety of solid dynamic visualizations that can be customized and manipulated to meet business needs. But Many Eyes has its negative sides too. As discussed in this paper it is based on the Java technology with the increasing number of updates, security issues arise for many users. Java applets are viewed as an aging technology and look less appealing for advanced users. In comparison with Fusion Tables, interactive visualizations are much harder to export and share than static images. There are several steps that Many Eyes can take to increase its popularity. Some of them are incorporating latest technologies such as Python and Flash in addition to Java that can still be used to process data on the backend. As stated earlier, a significant drawback is that time series data cannot be effectively visualized by Many Eyes. It would be wonderful if dynamic visualizations could be transformed on demand to the forms that are easy to share and export. At the same time Many Eyes can benefit by implementing techniques that would allow reading real time data from external sources and creating live visualizations. This would open endless possibilities for Many Eyes and its users making this tool very valuable in multiple areas. As stated earlier, a significant drawback is that time series data cannot be effectively visualized by Many Eyes.

Both Many Eyes and Fusion Tables are highly effective data visualization tools, but there is a significant scope for improvement. Today, there are a lot of new visualization techniques that keep emerging, and are more effective than traditional visualization methods. It would be interesting to observe how many techniques are incorporated into the two tools over a period of time, and note if the popularity changes.

## REFERENCES

1. Viégas, F., Wattenberg, M., van Ham, F., Kriss, J., and McKeon, M. 2007. Many Eyes: A Site for Visualization at Internet Scale. In IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 6, pp. 1121-1128.
2. Many Eyes (Software) (2007). <http://www-958.ibm.com/software/analytics/manyeyes/>; IBM
3. Fusion Tables (Software) (2011). <https://support.google.com/fusiontables/answer/2571232?hl=en>; Google
4. Meteorites data- <http://www.visualizing.org/datasets/meteorite-landings>
5. Radmila Berger,; Iosif Berger,; Irina Ilyaguyeva,; Vera Gabidova “Analysis and Comparison of Data Visualization using two Popular Tools: Many Eyes and Swivel”

## APPENDIX

Figure 1: Data upload description on Many Eyes

The screenshot shows the Many Eyes interface for a dataset titled "Data sets : Meteorite Data". The page is uploaded by Pisharath Vajhala and created on Mar 12 2014. The data source is The Meteoritical Society, and the description is "A subset of the Meteorite landings data, with no missing values". The tags are "meteorite landings".

The table below displays the data for 14 meteorites:

	Name	Type	Class	Mass	Fall	Year	ID	Latitude	Longitude
1	Aachen	Valid	L5	21	Fell	1880	1	50.775	6.08333
2	Aehus	Valid	H6	720	Fell	1951	2	55.18333	10.23333
3	Abze	Valid	EH4	107000	Fell	1952	6	54.21667	113
4	Acapulco	Valid	Acapulcoite	1914	Fell	1976	10	16.88333	-89.9
5	Achiras	Valid	L6	780	Fell	1902	370	-33.16667	-64.95
6	Adhi Ket	Valid	EH4	4239	Fell	1919	379	32.1	71.8
7	Aizhi-Hogdo (stone)	Valid	L1,3-5	910	Fell	1949	390	44.83333	95.16667
8	Agen	Valid	H5	30000	Fell	1814	392	44.21667	0.61667
9	Aguada	Valid	L6	1620	Fell	1930	390	-31.6	-65.23333
10	Agula Blanca	Valid	L	1440	Fell	1920	417	-30.85557	-64.55
11	Aicon el Atrouss	Valid	Diogenite-pm	1000	Fell	1974	423	16.39006	-5.57028
12	AA7	Valid	L6	24000	Fell	1925	424	19.08333	8.38333
13	Akaba	Valid	L6	779	Fell	1949	426	29.51667	35.05
14	Akberpur	Valid	H4	1800	Fell	1838	427	29.71667	77.95



Figure 4:

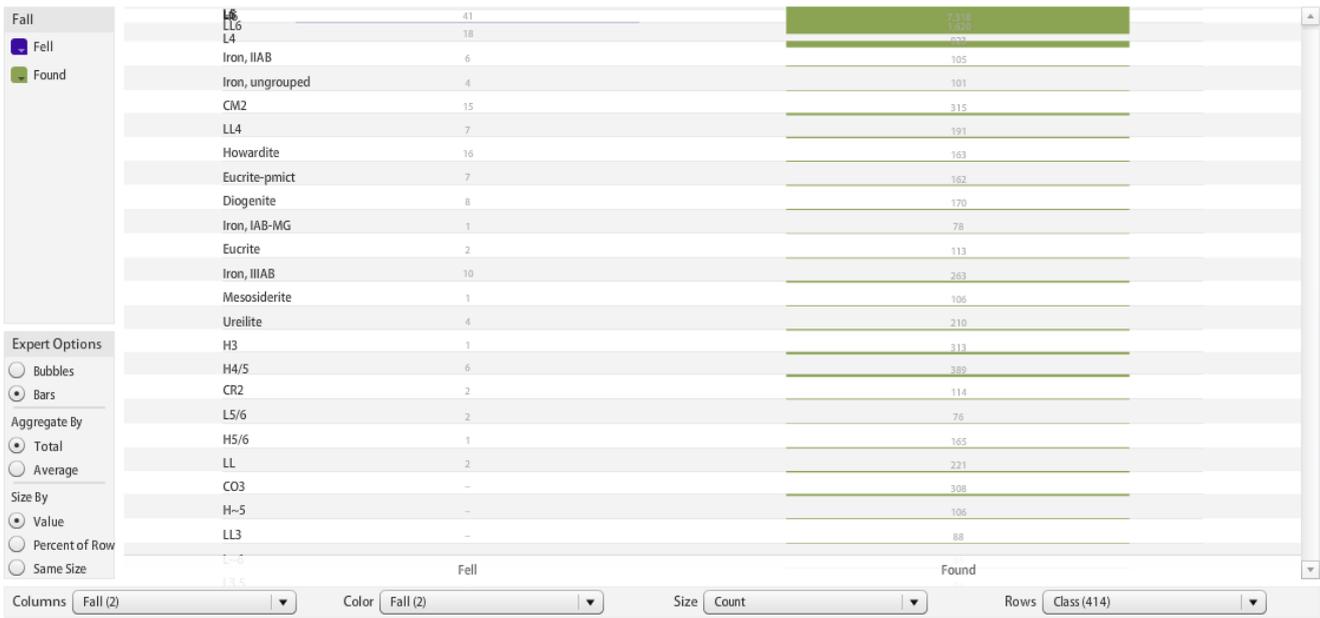


Figure 5: Map chart from Fusion Tables



Figure 6: Heat map from Fusion Tables

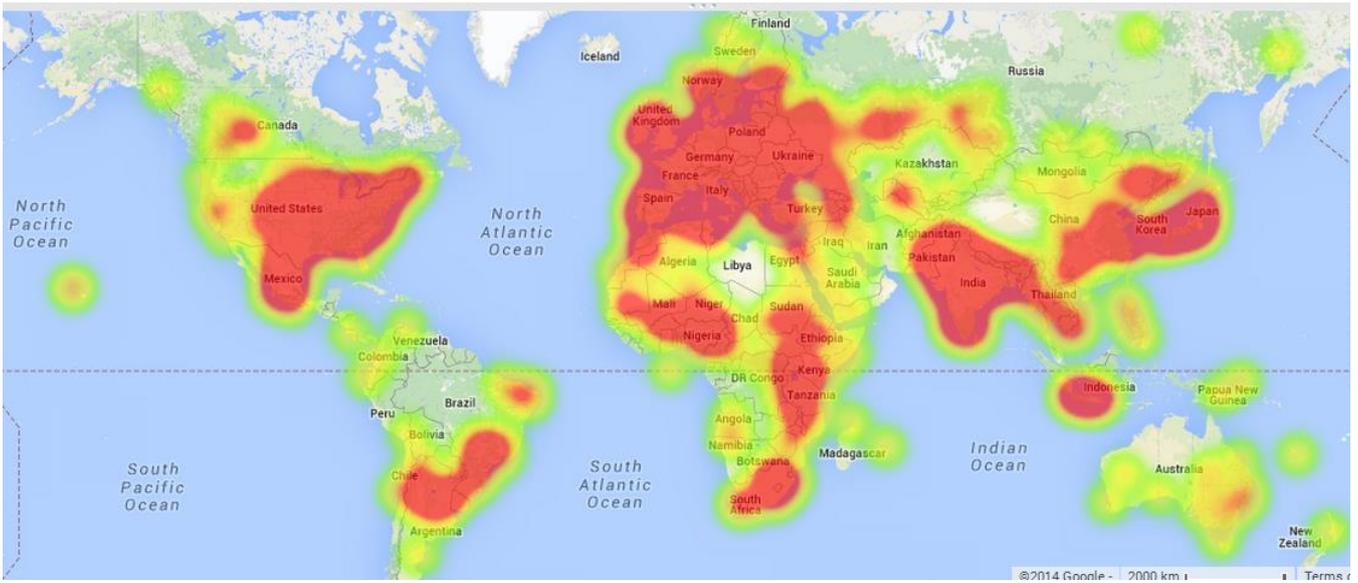


Figure 7: Timeline graph form Fusion Tables

