

Homework: Visual Search and Interaction with Weapons Datasets

Due: December 7th, 2015 12pm PT

1. Overview

In this assignment you will take your Apache Solr index constructed from Weapons data that your team collected to build a visual and interactive search interface for the data using the Data-Driven Documents framework (<http://www.d3js.org/>).

[Overview](#) [Examples](#) [Documentation](#) [Source](#)



and visualize them using the D3.js, Banana from LucidWorks, and OKFN FacetView.

Over the course of the semester, you made contributions by developing novel approaches for deduplication of arbitrary document features; for crawling and ingesting and approaches for link-based and content-based ranking of those science data and features and you have leveraged open source, and hopefully interesting technologies including Apache Nutch and Apache Tika, Apache Solr, GeoTopicParser, Stanford NER, Apache cTAKES, Tesseract, and other systems. The second assignment left you with a rich Apache Solr/Lucene based catalog of weapons that you used to answer a set of challenge questions. Consider yourself in nationally recognized and amazing company.

In this assignment, you will leverage your Apache Solr index and the D3.js data visualization technology (<http://d3js.org/>) to interact and visualize your search engine data. You will leverage D3.js and the geolocation information generated from GeoTopicParser along with time information in the data to build visual and interactive interfaces to your data. You will also leverage the Banana (<http://github.com/LucidWorks/Banana/>) and/or Kibana (<https://www.elastic.co/products/kibana>) technologies depending on whether you used Solr and/or Elasticsearch to construct an interactive Weapons “dashboard” that will be in turn used to provide visual answers to your challenge questions and topics focused on temporal and spatial aspects of the data as specified in Assignment #2. Finally you will build a FacetView (<http://github.com/chrismattmann/facetview/>) based search interface for forms/fielded, faceted and free-text search.

You should still have a copy of the your Solr index from assignment #2 to start the assignment.

2. Objective

The objective of this homework is to leverage your Weapons data, to construct a novel data visualization capability using D3.js, the Banana/Kibana Analytics Dashboard and FacetView. These packages form a novel data visualization technology, taking the approach that all data in a visualization should be available as soon as possible, that a user should be able to interact with the visualizations produced of the data as it is generated and after, and that the data behind visualizations should be interrogatable by the user. The objective of this assignment is to also expose you to real technologies and to cause you to consider design and UI/UX objectives related to search as we discussed in class, and the implications of the visual and interactive elements of search with the inverted index and with information retrieval.

To connect these DataViz capabilities to your Solr index, you will need to leverage Solr's REST-based web service, and likely its JSON response handler. You may either construct a specific type of Solr ResponseWriter, leverage Solr's XSLT-based transformer writer, or create a specific type of data loader in D3.js in order to load your Weapons dataset and its information into D3, Banana, and FacetView.

You are required to construct a web dashboard that uses D3.js and illustrates 6 of the examples from Mike Bostock's gallery (<https://github.com/mbostock/d3/wiki/Gallery>) using your Weapons data. In addition to your D3 dashboard, you will deploy Banana/Kibana, and also FacetView and then combine them into a single unified web interface for interacting with your Weapons data.

Your interactive web dashboard should allow you to answer a set of challenge questions which amount to generating specific views on your D3-based web page that show the data from your search index temporally and spatially.

3. Tasks

1. Construct a web front end ideally using AngularJS, or some pre identified web framework that leverages D3.js, Banana/Kibana, and FacetView to create an analytics dashboard and interactive search interface to your Weapons data.
 - a. The interface should directly leverage D3.js, Banana/Kibana, and FacetView to create the Analytics Dashboard.
 - b. The web interface should incorporate at least six D3 widgets identified in Mike Bostock's gallery: <https://github.com/mbostock/d3/wiki/Gallery>
 - c. The web interface should include the ability to provide a "Map" to your data using geocoded information from GeoTopicParser that should have been created in your assignment #2. To this end you can leverage Khooshe, a technology developed by JPL in the DARPA XDATA and Memex programs (<https://github.com/MBoustani/Khooshe/>).

- d. Your web interface components should communicate with your Weapons Solr/ElasticSearch REST end-point and call its JSON response handler, or you should develop your own end point (see task #2).
2. Provide web-based REST access to your Solr/ElasticSearch data in a compatible way with D3.js
 - a. You will need to ensure that D3.js can load your data from your Solr/ElasticSearch index. You may be able to get the REST endpoint and JSON response writer in Solr (or equivalent in ElasticSearch) to access the data directly from D3.js.
 - b. If you cannot get the existing Solr end point working, develop a new REST-based end point for Solr that simplifies the JSON output so that you can load it directly into D3.js.
 - c. You may need to develop an approach for loading data into D3.js. See this page from Mike Bostock (one of D3's creators, along with Jeff Heer) for more information on loading data into D3: <https://github.com/mbostock/d3/wiki/Requests>
3. Develop capabilities in your Analytics Dashboard to satisfy the following four challenge questions.
 - a. What time-based trends exist in Gun ads? Can you correlate temporal and spatial properties with buyers? For example can you identify based on ad time-window and/or based on geospatial area places where people try and purchase guns on behalf of someone unauthorized to purchase them?
 - b. Can you identify similar firearms image types (e.g., shotguns) that are sold in the same region and time? Does this indicate influx related to stolen goods?
 - c. When a shipment of bulk firearms is stolen, the rate of ads and images may indicate an increase in sales of that particular make/model – can you identify these?
 - d. Can you identify ads and/or weapons images that are posted by persons, whom are underage or in which the weapons are de-identified (by type and/or serial number, etc.)
 - e. Can you identify ads and/or images that relate to the unlawful transfer, sale, and possession of explosives, WMD devices, and precursors ?

4. Assignment Setup

4.1 Group Formation

Please keep the same groups as for your assignment #2. If you have any questions please contact:

Mohit Bagde
bagde@usc.edu

Divydeep Agarwal
divydeea@usc.edu

Komal Dhawan
komaldha@usc.edu

Use subject: CS 572: Team Details

4.2 Dataset

Please start with your data that you have prepared in assignment #2. If you would like additional data from Amazon S3, please let the graders know and we will coordinate its delivery via amazon S3 buckets and read-only keys to **one member of your group**.

4.3 Downloading and Installing D3.js, Banana, and FacetView

Download D3 from <http://d3js.org/>

You can find some step-by-step tutorials online for how to structure your D3.js project, for example:

<http://alignedleft.com/tutorials/d3/setup>
<https://www.dashingd3js.com/d3js-first-steps>
<https://www.npmjs.org/package/d3> (if you use NPM)
<http://bost.ocks.org/mike/d3/workshop>

Also see this great tutorial from Mike Bostock (D3's primary creator) on his blog about making a Bubble Map in D3: <http://bost.ocks.org/mike/bubble-map/>

Download Banana from <http://github.com/LucidWorks/Banana/> and install per these instructions:

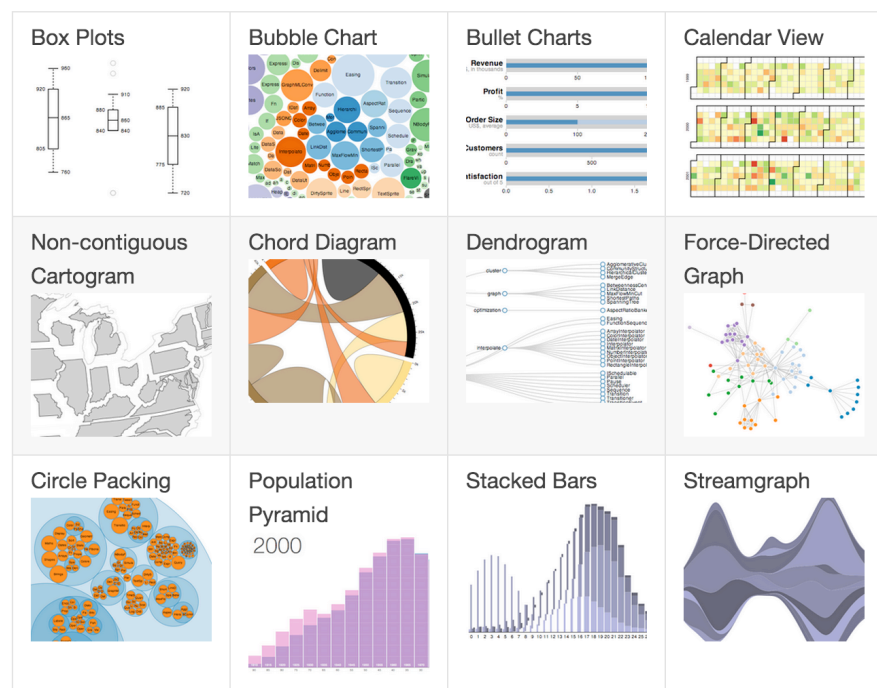
<https://docs.lucidworks.com/display/SiLK/Banana+Installation>

If you used Elasticsearch download Kibana (<https://www.elastic.co/products/kibana>) and install per these instructions:

<https://www.elastic.co/downloads/kibana>

Download FacetView from <https://github.com/chrismattmann/facetview> and install per the README.md file.

Visual Index



4.4 Constructing your Dashboard Web Framework

Ideally to construct your D3 analytics gallery, you will need to use a Web Framework or leverage some capability to access your Solr/ElasticSearch-based search index over the web and over the REST protocol. To do so, you will need to leverage Javascript (for data loading), REST (for access to your search index in Solr), and potentially a web framework for pulling it all together. You are free to use any web framework that you feel comfortable with, you must report which framework in your team's README file. You are also free to decide whether you need to construct a web service in the particular web framework that you choose to reformat the Solr/ElasticSearch JSON output so that you can load it into D3, or if you want, to add javascript libraries to load directly into D3 and to build your dashboard. **Note that loading your Solr data into Banana and FacetView is completely straightforward – both of them can talk to Solr directly, but note you will need to solve Cross Site Scripting issues to make this happen if you don't put all of the services onto the same web server and container.**

Start small initially – load only a subset of the data and attempt to visualize them in D3 and in Banana and FacetView. Test how responsive the framework is. You may need to develop an approach for subsetting the full dataset, and/or for summarizing it depending on how much data you have.

To demonstrate your Analytics Dashboard we ask that you prepare a simple video and upload it either to YouTube and/or Vimeo and provide a link to the video in

your README. The video should demonstrate the use of your dashboard, and how you are answering each of the four challenge questions.

4.5 Submitting patches and contributing to open source

Please note that if you contribute patches to D3, Banana, Kibana and FacetView on Github (via pull requests), and if you contribute to Nutch or Tika or Solr or GeoTopicParser or projects from prior assignments, you will have the opportunity to earn extra credit in a similar fashion to the prior assignment.

5. Report

Write a short 4-page report describing your observations. How easy to use was D3? How easy to use was Banana or Kibana? How easy was it to use FacetView? What was the hardest part, loading data, or visualizing it? Do NOT simply provide advantages and disadvantages from a quick Google search. You are required to think critically about this portion of the report and sincerely provide your feedback.

Describe in detail how your map answers the challenge questions presented in Task #3. Specify how to use the map to visualize the results.

Please also note that the graders will be given great flexibility to add/subtract from various areas of the report based on your submitted work.

6. Submission Guidelines

This assignment is to be submitted *electronically, by 12pm PT* on the specified due date, via Gmail csci572fall2015@gmail.com. Use the subject line: CSCI 572: Mattmann: Fall 2015: DataVis Homework: <Your Lastname>: <Your Firstname>. So if your name was Lord Voldemort, you would submit an email to csci572fall2015@gmail.com with the subject “CSCI 572: Mattmann: Fall 2015: DataVis Homework: Voldemort: Lord” (no quotes). **Please note only one submission per team.**

- All source code is expected to be commented, to compile, and to run. If you created a web service, or if you have web framework code, then you should also include those source files that you added, if any. Do **not** submit *.class files. We will compile your program from submitted source.
- Also prepare a `readme.txt` containing any notes you'd like to submit.
- Do **not** include Solr's deployment in your submission. We already have these.
- Please provide a link to the video illustrating your Analytics Dashboard and the answers to the challenge questions. If you have your DataViz demo up and running on a public URL, please provide it.
- If you have used any external libraries other than D3, Banana/Kibana and/or FacetView please note in your `readme.txt` a detailed explanation of how to use these libraries when evaluating your code.

- Save your report as a PDF file (Lastname_Firstname_DATAVIS.pdf) and include it in your submission.
- Compress all of the above into a single zip archive and name it according to the following filename convention:

<lastname>_<firstname>_CSCI572_HW_DATAVIS.zip

Use only standard zip format. Do **not** use other formats such as zipx, rar, ace, etc.

Important Note:

- Make sure that you have attached the file when submitting. Failure to do so will be treated as non-submission.
- Successful submission will be indicated in the assignment's submission history. We advise that you check to verify the timestamp, and download and double check your zip file for good measure.
- Again, please note, only **one submission per team**. Designate someone to submit.

6.1 Late Assignment Policy

- -10% if submitted within the first 24 hours
- -15% for each additional 24 hours or part thereof