

CSC-395: Final Project Proposal

Grinnell College, Department of Computer Science

Logan Goldberg

Zachary J. Susag

Peiyun Zhang

April 17th, 2017

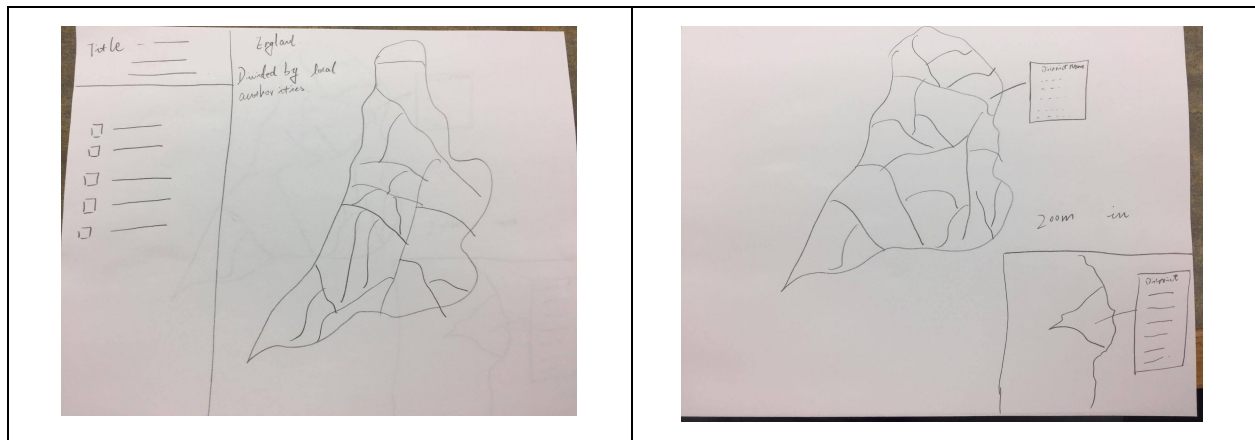
Purpose:

Many explorations and investigations have been conducted from a statistics/data visualization perspective for the surprising US election happened recently. Researchers found that there were very high correlations between the voting result and the demographic structure in a certain area. Inspired by these studies, we decided to investigate another surprising political event, Brexit, using data visualization so that we can have a better understanding whether people within different socio-economic level and race/ethnicity background influenced their decisions on Brexit.

In particular, we would like to see if a British citizen's religion, education, ethnicity, tenure, age, locale and many other demographic categories affected their vote in the Brexit referendum. To each question, there are a spectrum of answers as no one statistic on its own will explain a citizen's voting behavior. Therefore a data visualization is required to permit an intersectional analysis across all statistics to see what set of factors might have contributed to a specific cluster of citizens to vote in a certain manner. Moreover, we anticipate this data visualization will reveal many divisions across the plethora of demographic categories and possibly bring to light the structural differences that may have swayed a citizen's vote.

Design:

Our visualization will provide a geographic layout of the whole of England (Scotland and Northern Island are not available), broken up into districts (UK equivalent to counties). To the far left, we will have checkboxes or drop down menus to allow for a user to select which summary demographic they would like to display across the whole of England. The map itself will allow for hovering and clicking on individual districts, and filtering across regions. In particular, hovering over a district will show a few quick statistics for that district and clicking will allow a zoom-in on that district, showing all demarcated statistics. A color-blind friendly color scheme will be used for each statistic that is being shown.



Techniques:

1. **Where will the data for your visualization come from? Look closely at every dataset you plan to use to make sure it will actually meet your needs.**

We will collect our data from the following three sources:

1) EU Referendum result data - This dataset gives us the statistics of the voting result for each local authorities (geographical unit in UK). These statistics included region, local authority names, and the actual number and percentage of people who voted for remain and leave. The data comes from The Electoral Commission of the United Kingdom, located at <http://www.electoralcommission.org.uk>.

2) Demographic data - This dataset provides the statistics of the demographic breakdown within all of the local authorities, including age, gender, race and other socio-economic status indicators. However, the data for Scotland is missing here because they use their own system and statistics. Additionally, we have not found any terribly recent demographic data so we are relying on the data from the 2011 census. The data comes from the Office for National Statistics of the United Kingdom, located at http://www.nomisweb.co.uk/census/2011/key_statistics and https://www.nomisweb.co.uk/census/2011/quick_statistics.

3) Shapefile - The shapefile is provided by James Trimble, the creator of UK Data Explorer. The shapefile is only of England, which is okay due to the lack of data from Scotland and Northern Ireland that we found. The TopoJSON file is located at <http://ukdataexplorer.com/census/england/topo.json>.

2. What kinds of computations or transformations will you need to get your data into the visualization?

One benefit of the data that we have collected is the amount of statistics located within. There is a combination of raw values as well as percentages. In regards to computations we will need to use the percentages and values from the demographic data to compute what proportion of the population, in a given area, voted yes or no in leaving the European Union. Essentially, we will need to combine the two datasets that we have into a single “dataset” which will allow us to see if certain demographics leaned towards voting a certain way in the EU Referendum. We will be relying heavily on Square’s crossfilter to allow the users of the visualization to turn certain demographics on or off. This will require reading in the appropriate data, storing them in dimensions, and then implementing the UI to allow for users to select which data they would like to see.

3. Are there any D3 or data processing techniques you will need that we have not covered? If you have an example of a technique you would like to use, please include a link with your proposal.

One technique that we will most likely be using is to be able to zoom in on a certain region from the TopoJSON file. The creator of this transition, which we found here, <https://bl.ocks.org/mbostock/4699541>, calls this “Zoom to Bounding Box” and will allow us to see what is happening in a certain district instead of the entirety of England. Otherwise, the techniques that we will be using are ones that we have seen before, such as crossfilter and creating a geographic visualization. Something that we have not fully decided on is whether to include mini visualizations such as pie charts

when a certain district is clicked. If so, it would be interesting to know if there are any built in plugins for such a visualization or if we would have to design a pie chart from scratch.