Life in Hampton Roads Survey Visualization

Mohan Krishna Kodali, Ujwal Manjunath, and Andrew A. Schaefer

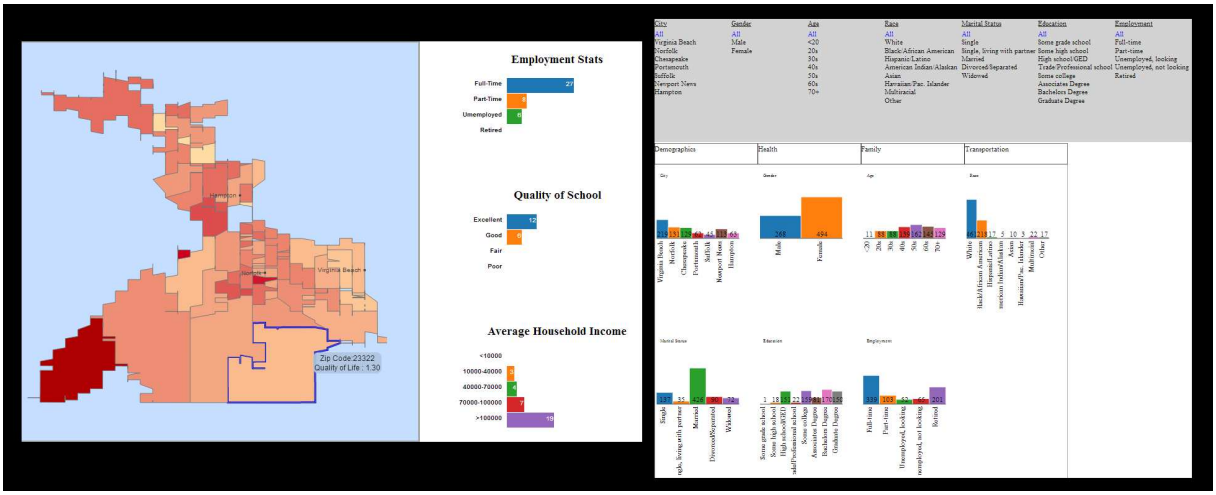


Fig. 1. Two visualizations of Life in Hampton Roads Survey data.

Abstract—This project focusses on attitudes and perception of citizens regarding topics of local interest such as transportation and traffic, health, crime and other issues. We present a map of Hampton roads with zip code borders and for each zip code we pull up statistics about crime, traffic and demographics using bar charts.

1 Introduction

The Social Science Research Center (SSRC) at Old Dominion University conducted the third annual Life in Hampton Roads Telephone survey in 2012. This survey consisted of a variety of questions asked to respondents which were based on quality of life, transportation, local and state government, political issues, environmental issues, health and education, housing, neighborhood issues and crime, military life and basic demographic information. The survey was programmed by SSRC using a computer assisted telephone interviewing (CATI) system.

Our project involves creating an interesting and meaningful visualization of this survey data. In our efforts, we utilized a number of visualization and data cleaning tools and techniques.

2 Related Work

Visualization is a rich field and diverse field, with branches in scientific visualization, visualizing information through its physical representation, and information visualization, visualizing abstract/non-physical elements. Early works like Kosslyn's discussion of chart and graph types and elements[3] provide the backdrop for designing information visualizations. More recent work, including Jeffrey Heer's Tour through the Visualization Zoo[2] catalog a wide diversity of visualization techniques that exist for visualizing different data sets.

The ever-increasing prevalence of the internet has seen the expansion of static visualization into interactive visualization, paving the way for visualization tools such as d3[1]. Along with the sophistication of visualization toolkits, comes the need to understand how to develop better visualizations. Ben Shneiderman approaches this question by identifying common tasks users perform on data[5]. Segel and Heer take a different approach and examine the types of visualization employed and attempt to classify them, while evaluating their strengths and weaknesses[4].

3 Data

While the survey has been conducted for 3 consecutive years, the number and types of questions has varied significantly from year to year. As such, we chose to focus on a single year to avoid the confusion of seemingly related questions that may or may not be collecting the same information. We used the latest survey, which was conducted in 2012. We received the data in SPSS (.sav) format and is weighted to adjust for demographic differences between the survey respondents and the general Hampton roads population as well as phone usage and city of residence. We converted the format of data file into csv-format. There were a total of 762 respondents extending over 61 questions.

The data file was formatted with question numbers and response values, to interpret these values we also received a supplementary file with the questions asked in the survey and their corresponding response options. All the questions are of multiple choices comprising of 2 or more options. Most of the questions are single questions while some of them have multiple questions within them. Each question has 3 properties: "min", "max", "l". Property "min" indicates minimum number of answers required for the question. Property "max" indicates the maximum number of answers that can be given to the question. Property "l" indicates the length of the option.

Mapping data was acquired from shapefiles provided by the US

- *Mohan Krishna Kodali, Old Dominion University, Norfolk, VA 23508. mkoda001@odu.edu*
- *Ujwal Manjunath, Old Dominion University, Norfolk, VA 23508. umanj001@odu.edu*
- *Andrew A. Schaefer, Old Dominion University, Norfolk, VA 23508. aschaefe@cs.odu.edu*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 27 September 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

Census Bureau[1] and Natural Earth Data[2]. Shapefiles from the Census Bureau provided geographic data for all zip codes, while the shapefiles from Natural Earth Data contain geographic data for States and Cities. Combining these two files into topoJson format we were able to create a custom map of the Hampton Roads area.

4 Tools

Data cleaning was done using Microsoft Excel. Initial exploration of the data was perfomed in Tableau[3] while our final product was developed in javascript using the d3 toolkit developed by Mike Bostock[4].

To convert the shapefiles into topojson format for use with d3, we utilized Mike Bostock's tutorial[5]. This requires the ogr2ogr tool to filter the shapefile to retrieve only those shapes of interest into GeoJson format, and the topojson tool to convert and merge the filtered GeoJson files into a single topoJson file that d3 can consume.

Listing 1. ogr2ogr command for filtering Virginia cities from the populated places shapefile

```
ogr2ogr -f GeoJSON -where "ADM0_A3 IN ('USA')
AND ADM1NAME = 'Virginia'" places.json
ne_10m_populated_places.shp
```

Listing 2. topojson command for merging the GeoJson files

```
topojson --id-property=name,postal
-p name=NAME,name=name -o hr.json
states.json places.json zip10.json
```

Once we had the final topojson file we were able to create our desired map of the Hampton Roads area using d3.

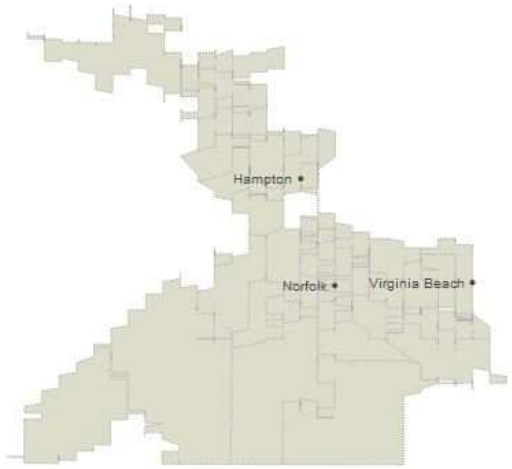
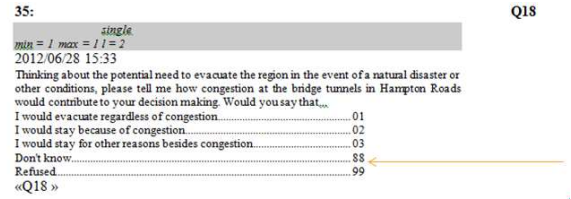


Fig. 2. Map of Hampton Roads by zip code.

5 Data Cleaning

We used Microsoft Excel for cleaning the data. In Excel, we used 'Filters' feature to identify the unwanted data and removed it from the file. Every question had the options: "Don't know" and "Refused to answer", we removed them because they do not contribute to our visualization.

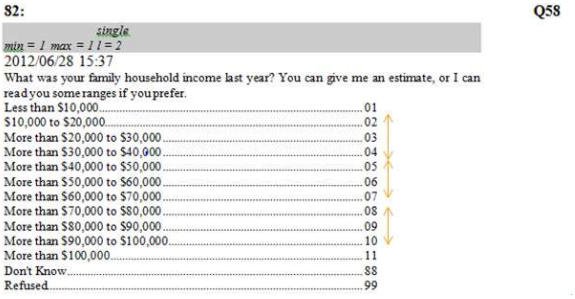
Next, we merged some of the options in the questions in order to make our visualization more effective. For example the average household income consisted of 11 options. We decided to merge these into


35: Q18

single
min = 1 max = 1 l = 2
2012/06/28 15:33
Thinking about the potential need to evacuate the region in the event of a natural disaster or other conditions, please tell me how congestion at the bridge tunnels in Hampton Roads would contribute to your decision making. Would you say that...
I would evacuate regardless of congestion....01
I would stay because of congestion....02
I would stay for other reasons besides congestion....03
Don't know....88
Refused....99
«Q18 »

Fig. 3. Example of Response values with unwanted data.

4 options roughly creating a low, low-middle, high-middle and high income levels.


82: Q58

single
min = 1 max = 11 = 2
2012/06/28 15:37
What was your family household income last year? You can give me an estimate, or I can read you some ranges if you prefer.
Less than $10,000....01
$10,000 to $20,000....02
More than $20,000 to $30,000....03
More than $30,000 to $40,000....04
More than $40,000 to $50,000....05
More than $50,000 to $60,000....06
More than $60,000 to $70,000....07
More than $70,000 to $80,000....08
More than $80,000 to $90,000....09
More than $90,000 to $100,000....10
More than $100,000....11
Don't Know....88
Refused....99

Fig. 4. Example of response value requiring merging.

Other challenge we faced, zip codes data in the data file is not complete. We didn't get boundaries of the regions when we plotted the map using that data. So, we searched for zip code data on the net and got some shape files. The files we got didn't have the zip codes directly for Hampton roads region, so we drilled down to get the required data. We used Mike Bostocks tutorial to convert shape file and get the TopoJson file. Next, we plotted our map using the TopoJson.

6 Approach

Our initial approach to this project was to identify some interesting facets of the dataset to create a mini-narrative that would follow the 'Martini Glass' structure: first highlighting the interesting facets to the user, while introducing them to the control of the system, eventually opening up to allow them to explore the data for themselves. In this way, we had hoped to create an author-driven narrative, emphasising the details of interest.

However, exploration of the dataset led away from this approach, as there were not any interesting or unexpected discoveries to be made. We switched our approach to develop more reader-driven interfaces where they would have free reign to explore the data in pursuit of answering the prescribed question of the visualization.

7 System Description

Our final product consists of two distinct visualizations, each attempting to answer a different question. Both of which are discussed in detail below.

7.1 "Where Shall I Stay?" Maps

The first visualization maps some of the survey responses by city and zip code, in an effort to allow users to answer the question "Where shall I stay?". This tool is targeted towards individuals who may be new to the Hampton Roads area or just interested in moving to another area within Hampton Roads. We looked at the data and questions and identified key relationships between them. We made some relationships between questions in sections like safety, transportation, and standard of living with respect to zip codes. Then we created a file containing unique zip codes and created a map of Hampton roads with zip code boundaries.

[1] http://www.census.gov/cgi-bin/geo/shapefiles2012/layers.cgi
[2] http://www.naturalearthdata.com/downloads/
[3] http://www.tableausoftware.com/
[4] http://d3js.org/
[5] http://bost.ocks.org/mike/map/

Table 1. Column Metadata Format and Sample data

Column Name	Display Name	Category
FIPS	City	Demographics
Q1	Health	Health
Q8	Type of School Children Attend	Family
Q17	Traffic Concern	Transportation
Q51	Gender	Demographics

Table 2. Value Metadata Format and Sample data

Column Name	Value	Display Name
FIPS	51710	Norfolk
Q1	2	Good
Q8	1	Public
Q17	5	Not at all concerned
Q51	1	Male

Next we took the selected data from the original file and put them into separate files. For safety, we took the crime rate and overall satisfaction with police, neighborhood and school safety; for transportation we took average commute time and overall concern of public with traffic; for standard of living we took quality of life, employment status, and average household income.

We provide tabs at the top of the map for Standard of Living, Safety and Transportation, clicking a tap loads the respective data and colors the map based on the primary attribute evaluated. The secondary attributes are displayed in bar charts to the right of the map. As the

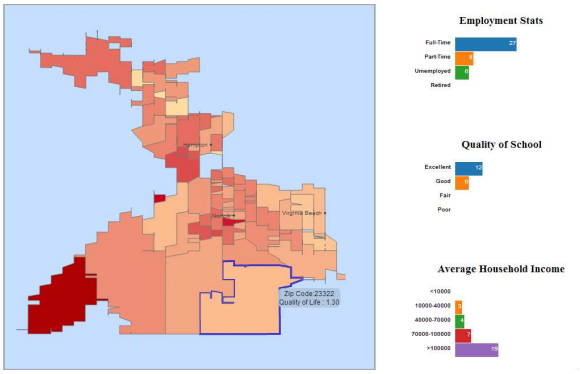


Fig. 5. Map and supplementary bar charts.

user selects a particular zip code, it becomes outlined in the map, and the secondary bar charts transition to display the responses from residents in that zip code only. Hovering over a zip code creates a tooltip that lists the zip code and the corresponding mapped value. This same aggregation is provided at the city level, so that a user may select a city, each zip code within that city becomes outlined on the map, and the secondary charts reflect responses from all residents within the selected city. 5.

7.2 "Like me" Explorer

The second visualization charts a variety of survey responses and allows filtering by demographics. The idea behind this tool is to allow users to answer the question "Who's like me?" it is targeted towards current residents of Hampton Roads who are interested in the responses of other residents. We created a few question sets of interest: Demographics, Health, Family, and Transportation. For each category we identified related questions. Two support this tool, two metadata file were created: the first maps column names to a display name, and category(Table 1); the second maps response values by column to a display name(Table 2).

Fig. 6. Demographic categories and values.

For this visualization we list the demographic categories and their response values in columns at the top (Figure 6), with the main display area below (Figure 7). The display area is tabbed, to allow the user to

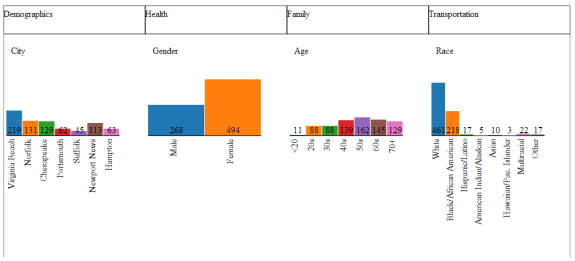


Fig. 7. Tabbed display panel.

select the question set of interest; as each tab is selected the content is updated to display the corresponding graphs. The user is free to filter up and down based on different demographics. As a demographic value is selected the displayed graphs transition to reflect the answers provided by the respondents that meet the demographic restrictions.

8 Future Work

As the system proposed uses only LIHR Survey data 2012, in the future we would extend the current system to utilize past and subsequent survey data and provide a visualization with respect to time.

We would also like to deploy maps into the "Like Me" tool which would also apply filters on the maps and display information on maps.

We had planned to add information about each zip code/city selected in the left over real estate on the extreme right end of "where shall I stay" tool, due to time constraints this data was information was never prepared and integrated, it could be easily done.

In the future we will allow multiple selections in both tool. In the "Like me" tool we could allow multiple filters for example to allow users to look up information on all white males in the age range 20-50. In the "Where shall I stay" tool where we could allow users to select multiple zip codes/ cities.

9 Conclusion

In this project, we demonstrate the crime rate, traffic and demographic statistics in Hampton Roads. This helps a person who is new to Hampton Roads know how safe the place is; the overall quality of life in the area; and traffic congestion over the Hampton Roads area.

By sorting out the unique zip codes from the data file, we plotted different regions covered by Hampton roads based on zip code boundaries. We thought bar charts would be a good option to show the statistics, so we implemented bar charts along with transitions and integrated them with the map using mouse click. We also look at a general overview of survey responses and allow filtering the responses by demographic values.

Our visualization provides a base for future enhancements. With data available, people can expand the visualization and give a much broader outlook of the area by looking into aspects like health, Light rail, election voting, and more.

10 Final Thoughts

We faced problems with the data. The data is incomplete, for example, data is given about the average commute time but the distance of the commute is not given, so we cannot have a clear understanding about the traffic congestion. Other example, no data is given about places to which people travel regularly for work/school by which we can know the percentage of them using Light rail and the reason why they want it to be extended? This is the reason why we dropped the idea on showing information about Light rail.

We learnt different things related to d3 like drawing a map of Hampton roads with zip codes. Coloring the map using a certain aspect of the data like we used crime rate data for coloring the map of crime statistics. Next, we learnt how to use "div" element to align the map, bar charts so that the visualization looks organized. We got to know on how to bind data into bar charts and insert transitions into them. Moving along, we read some tutorials about handling click events, so when a zip code is clicked, respective statistics are shown using the bar charts. We learnt how to handle the mouse hover property so whenever mouse pointer is hovered on the map, statistics about that zip code are displayed.

10.1 Contribution of each group member

Mohan Krishna Kodali Worked on Gathering Data and Cleaning up of data. Scheduled routine group meeting throughout the week and sent out updates about each meeting to all the members and worked on Preparing Project 1 PowerPoint and Report.

Ujwal Manjunath Involved in data cleaning, plugged data into tableau to hoping to find something interesting to visualize. Worked with d3.js library and developed the "where shall I stay" tool and also contributed to Project 1 PowerPoint.

Andrew Schaefer Worked on Gathering data, data cleaning and plugged data into tableau to hoping to find something interesting to visualize. Worked with d3.js library and developed the "Like me" tool and also contributed to Project 1 PowerPoint and Report.

Acknowledgments

We thank Tancy Vandecar-Burdin for the data and we thank Michele C. Weigle for her valuable suggestions and support given to us throughout the project. We thank Mike Bostok for providing us with the d3 library and some excellent example of visualizations and we thank Cynthia Brewer, Mark Harrower and The Pennsylvania State University for providing support for color brewer.

References

[1] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.

[2] J. Heer, M. Bostock, and V. Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, 53(6):59–67, 2010.

[3] S. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3:185–226, 1989.

[4] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.

[5] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.