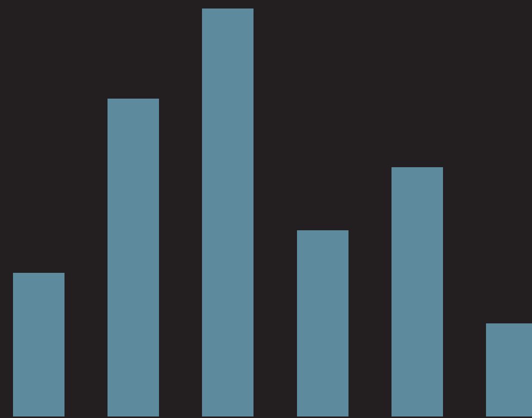
Water-Quality 
Interactive Data Visualization

Master's Project by Lauren Langley

in partial fulfillment of a Master's of Science in Digital Media
Georgia Institute of Technology
May 2012

Committee Chair
Carl DiSalvo

Committee Members
Ian Bogost
Janet Murray

Acknowledgements

The Water-Quality Interactive Data Visualization project could not have been possible without support of the following individuals:

Committee – Carl DiSalvo, Ian Bogost, Janet Murray
Upper Chattahoochee Riverkeeper - Sally Bethea, Jason Ulseth, Mike Meyer, Ryan Schultz, Jessica Preston
Comrades - Ben Chapman, Andrew Golaszewski, Joel Linderman, Allan Martell, Blacki Milhose, Kim Nguyen

Thank you all for your time, interest, and thoughtful advice during the creation of this project.

Table of Contents

"Information is the most powerful tool for combating the water crisis" [2]

- Peter Gleick, president of the Pacific Institute

Fig. 1: The Chattahoochee River (http://dangerpop.com/wp-content/themes/thesis_182/custom/rotator/3.jpg)

Abstract

Environmental science visualizations have the potential to translate quantitative data into representations that can be analyzed and used for legal, advocacy, and educational purposes. However, science-driven organizations lack the ability to collect, process, consolidate and distribute laboratory data freely and efficiently with generic software solutions [8]. The water-quality interactive data visualization project is a customized system that was created to address these issues by offering a comprehensive data management and visualization solution for an environmental advocacy organization.

Introduction

The Chattahoochee River is the most heavily-used water resource in Georgia [1]. The quality of water is a result of the complex interaction of natural and human influences on land and water. Metropolitan Atlanta, the largest and fastest growing metro area in the Southeast, is in the river basin's headwaters. The basin's growing population presents challenges to balancing human and ecosystem needs for water of sufficient quantity and quality [6].

The natural and human influences on water introduce dissolved solids, chemicals, and bacteria into waterways and pose a potential risk to human health, aquatic life, and fish-eating wildlife. Water quality standards have been put in place to protect water from such pollutants. Indicators of water contamination, like dissolved oxygen and E. coli, are measured and reported to the public by environmental advocacy organizations like the Upper Chattahoochee Riverkeeper and governmental agencies such as the U.S. Geological Survey (USGS), the United States Environmental Protection Agency (EPA), and the Centers for Disease Control and Prevention (CDC). Several of these entities advocate for environmental protection policies and actively educate the public about influences and effects on the environment.

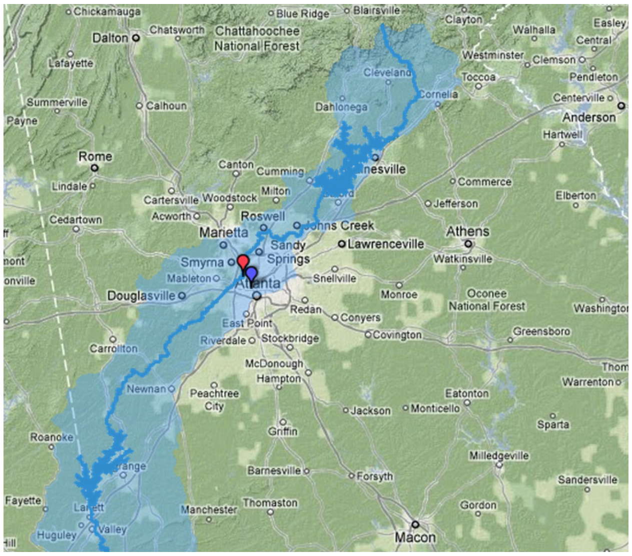


Fig. 2: The Chattahoochee River basin
(http://www.chattahoochee.org)

Fig. 3: Fly fisherman in the Chattahoochee River
(http://deepsouthflyanglers.com/wp-content/uploads/2012/01/Garner10.jpg)

The use of quantitative data is instrumental in supporting these policies and educational endeavors, as the data can be used to illustrate the health status of an ecosystem. For example, the detection or increase of a bacteria level in a waterway provides evidence that urban or industrial activities have detrimentally affected the health of the water system [5].

One way in which water-quality measurements can effectively be used to support policy and education is to visually display to the public the data collected and reported by organizations who monitor the water. I have developed my thesis project with these issues in mind. The remainder of this document outlines the process of creating the project, from research and design, to implementation and evaluation.

Research

Research for this project began in June 2011 after determining that I would be developing a web-based interactive visualization that would display environmental science data. The research consisted of contextual, technical, and design research.

Contextual Research

I chose to build a visualization that displays quantitative data about the local environment because I wanted the project to have the potential to impact the Atlanta community in a valuable way, by informing the community about the quality of the water that surrounds it and directly impacts its activities.

In order to do this I needed raw data, so I began researching reliable and active sources of data that I could use to create the visualization. I settled upon water-quality data for the Chattahoochee River because it was both publicly available and a significant indication of the health of the environment.

I found that I could have accessed and implemented data publicly available from sources like the USGS, the EPA, and the CDC. However, I wanted the data visualization to be used by, and useful for, the Atlanta community directly on an ongoing basis. So, I approached the Upper Chattahoochee Riverkeeper and proposed to build them a data visualization tool that they could maintain themselves in exchange for access to their data and water monitoring processes.

The Upper Chattahoochee Riverkeeper (UCR) is a nonprofit organization dedicated to protecting and preserving the river for the people, fish and wildlife that depend upon them, with active advocacy, education, research, communication, cooperation, water monitoring and legal actions. The organization accepted my proposal and offered data and access to their Neighborhood Water Watch monitoring program.

UCR's Neighborhood Water Watch monitoring program was created in 2010 as a collaborative program between UCR and neighborhood groups in the watershed with the goal of improving water quality in urban streams and protecting human health in the surrounding communities. Staff members and volunteers take weekly water samples from 18 stream sites to be analyzed for E. coli bacteria at UCR's EPA-approved lab. Elevated bacteria levels are a common indicator of sewage leaks and spills into waterways and the data can indicate the number and type of pathogens present in the water [6].

Fig. 4: Ansley Stream at the Dell, Neighborhood Water Watch monitoring site. (GIS map provided by City of Atlanta)

I had the privilege of experiencing the Neighborhood Water Watch program first-hand when I collected samples from three sites and completed the lab analysis on those samples in UCR's lab on October 20, 2011.

My experience collecting and analyzing the actual data I would be utilizing in the visualization project was invaluable. Through the process, I was able to visit the physical locations that the data is collected and witness how the water data is analyzed and reported.

The contextual research conducted prepared me to build a visualization tool that I could understand from the perspective of an administrator and a user. Throughout the development of the project, I continued to communicate with UCR to gain perspective on which of the data parameters collected were the most important to represent and why.

Fig. 5-7: UCR Intern Ryan Schultz collecting samples from Neighborhood Water Watch sites.

Fig. 8-11: UCR Intern Ryan Schultz and myself, conducting E. coli tests at UCR's lab.

Technical Research

Upon initial research, I discovered that there were several technologies that could be employed to implement the thesis project. My goal was for the technologies to match my skillset, while producing the desired interactivity and design quality. Additionally, after partnering with UCR, I was determined to build a tool that was streamlined with their existing protocol for entering data and could be maintained by the staff who has limited technical computer skills.

Some of the scripting technologies I considered for the web-based visualization tool were JavaScript, jQuery, HTML5, WebGL, Processing, SVG, AVS, and OpenDX. There were also a few JavaScript-based toolkits that I explored as well, including Polymaps (http://polymaps.org/), Protovis (http://vis.stanford.edu/protovis/), and JavaScript InfoVis Toolkit (http://thejit.org/). After discovering that UCR maintained their data in Microsoft Excel spreadsheets and Google Docs spreadsheets, I investigated the publicly available technologies that Google had to offer.

The technology stack that fit the needs of the project and was ultimately implemented was a combination of Google Docs, Google Visualization API, Google Maps JavaScript API Version 3, JavaScript, jQuery, HTML5, and CSS. The most significant benefit of this stack was the querying of data from the publicly published Google Docs spreadsheets with the Google Visualization API. This novel technological approach eliminated the need for traditional database management that uses PHP and SQL languages. This alone makes it possible for UCR to maintain this tool without the need for a web developer.

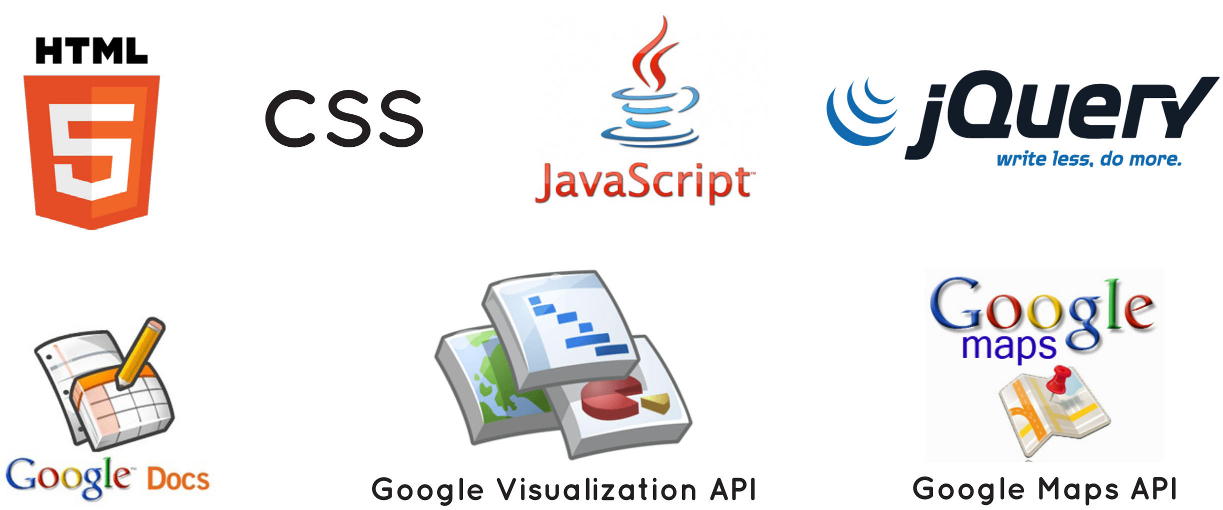


Fig. 12: Technologies implemented for project

Fig. 13: Flow Motion, Melbourne Water (http://www.melbournewater.com.au/applications/flowmotion/)

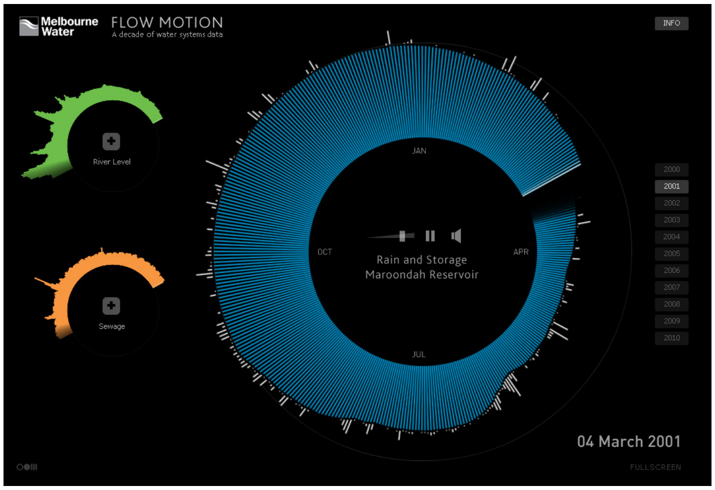
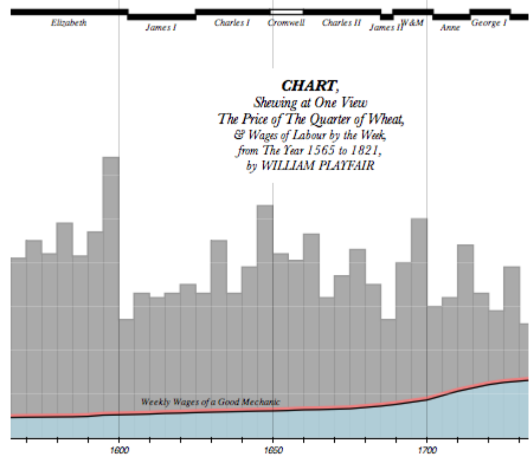


Fig. 14: Focus + Context chart, Protovis example (http://mbostock.github.com/protovis/ex/)



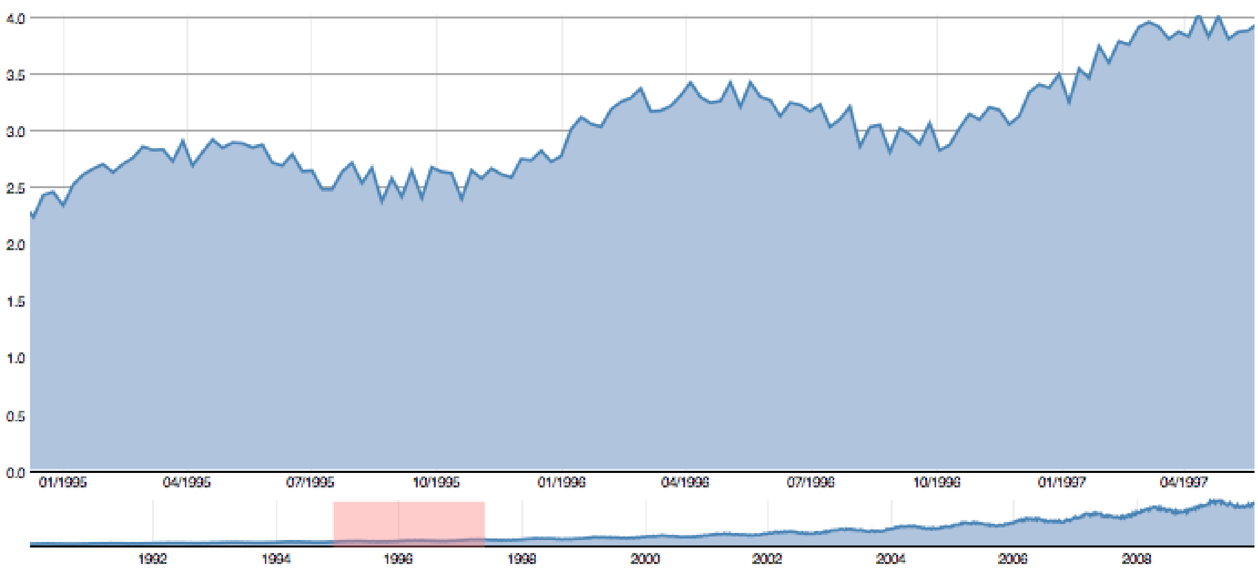


Fig. 15: Playfair's Wheat chart, Protovis example (http://mbostock.github.com/protovis/ex/)

Design Research

The visual style and information design decisions I made were based upon extensive design research. I reviewed several visual artifacts, in print and digital form, as well as read data visualization books, in order to inform the visual design of the project.

I did my best to follow Brian Suda's following advice, "Next time you sit down to illustrate some data, just remember that your main goal is to tell a story. This doesn't preclude a beautiful design, but if it's not comprehensible, it's not useful. You need to choose the right tool for the job: sometimes, simplicity is what's needed." [6].

In addition to this design research, I also received design input from UCR. For example, they requested that line charts not be used to represent certain parameters like E. coli because it would mislead the user to think that the levels gradually changed from week to week when in reality the levels from each collection are to be thought of as very separate results due to different influences on the water.

Existing Artifacts

There are a number of existing water data visualization systems that have been designed for communities, using similar technologies that I have explored. I would like to introduce a few of them and discuss their successes and opportunities for improvement.

MonRiverQUEST

Mon River Quest is an award-winning comprehensive water quality monitoring program established by the West Virginia Water Research Institute at West Virginia University [3, 7]. Several data parameters are displayed on a map interface and users may define and submit unique data queries for 16 monitored water sites.

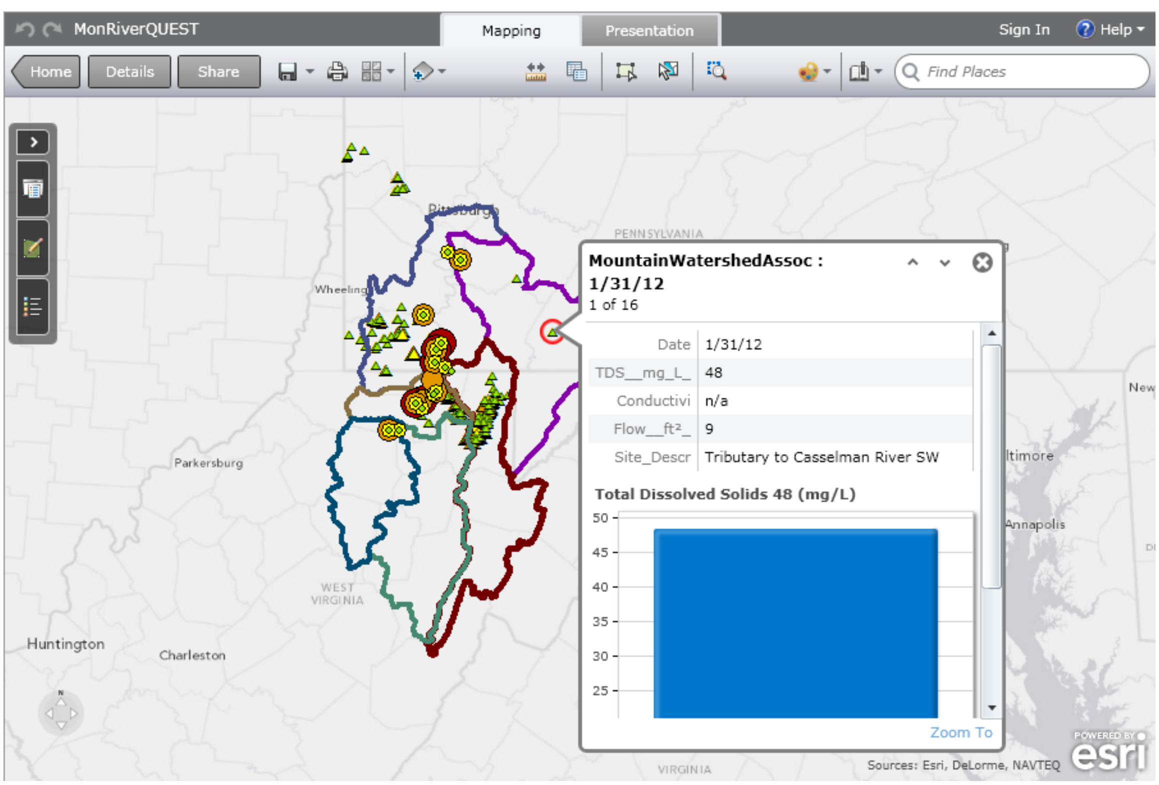


Fig. 16: Map interface for MonRiverQuest (http://www.monriverquest.com/map.cfm)

Mon River Quest successfully allows users to display very detailed information about a particular collection site in a one-page interface. However, the site is slow to respond to user input, the map animation is not smooth, there is not a way to easily compare data along a timeline, and it is difficult to visually identify a particular site on the map.

Adopt-A-Stream

Georgia's Adopt-A-Stream website displays the data for a large number of water monitoring programs throughout the state, including UCR's Neighborhood Water Watch. UCR sends their weekly data to Adopt-A-Stream for them to upload to this site. The UCR staff is very familiar with viewing their data on this existing artifact.

The site successfully shows comprehensive data parameters for individual collection sites in one place, but the visualization of that data could be improved. For instance, the default chart for E. coli is represented by lines, which UCR believes misrepresents the data as mentioned previously. Another issue with this site is the frustration that comes from clicking through an unreasonable number of pages in order to access the data for a specific site.

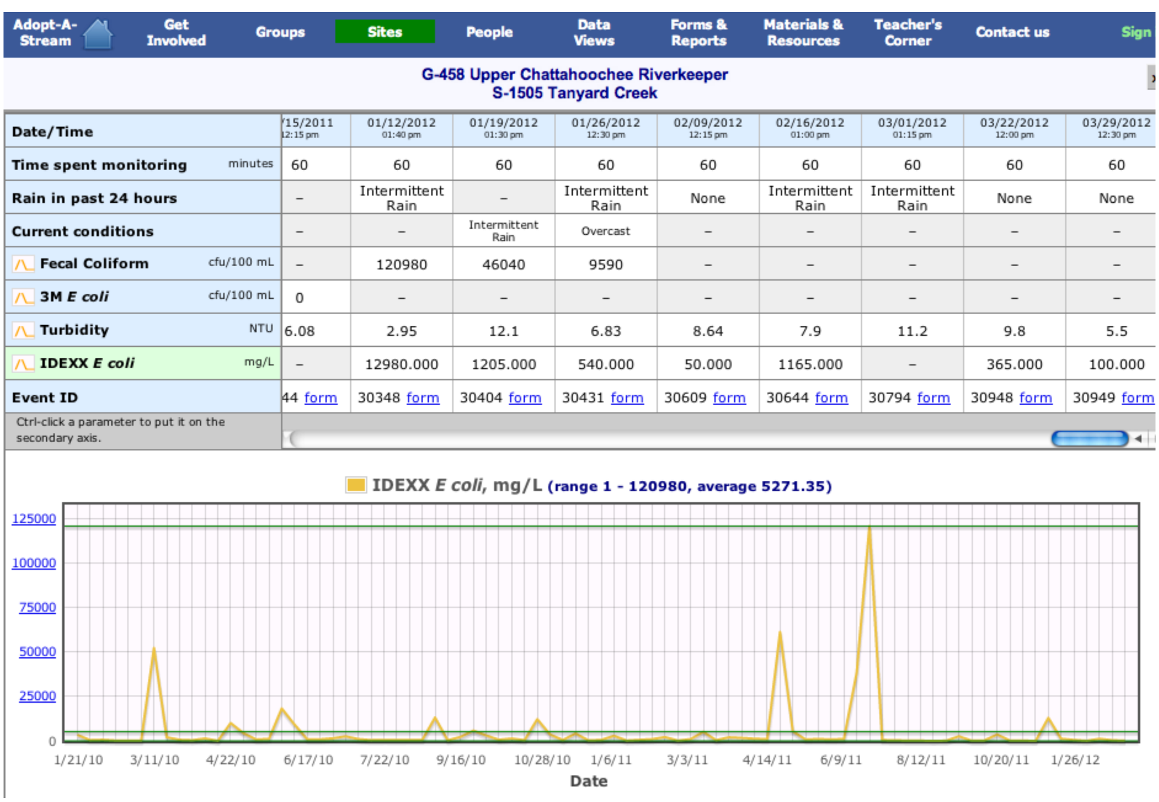


Fig. 17: Data visualization for Tanyard Creek, one of the sites monitored by UCR.
(http://georgiaadoptastream.org/db/Sites.asp?SiteID=1505)

Existing Neighborhood Water Watch Tool

Students at Georgia State University previously built UCR a tool to manage and display water-quality data for the Neighborhood Water Watch program. UCR uses this tool each week currently, but will eventually phase out its use when I deliver the system I built for them.

UCR uses this tool by entering data for every monitored site once a week into a spreadsheet. The results for the most recent E. coli levels are displayed on a Google Map. Each site is visible in a list view and as markers on the map.

Although this tool solves the basic issue of displaying water-quality data for the Neighborhood Water Watch sites, it has the potential to be improved upon greatly. One glaring area for improvement is to have an archive of data for each site as well as the ability to display more than one parameter.

This existing artifact in particular motivated me to design a complete customized visualization tool for UCR that would also extend the use of charts as well as interactivity.

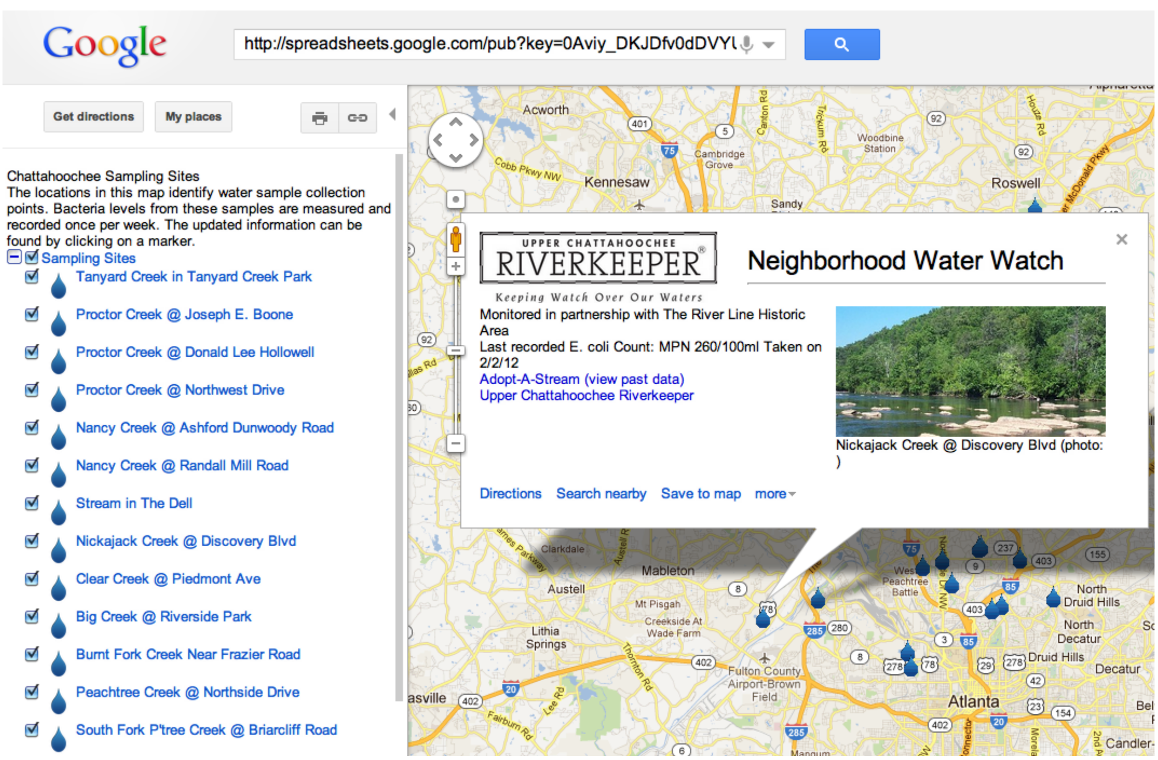


Fig. 18: UCR's existing data management tool (https://spreadsheets.google.com/spreadsheet/pub?hl=en_US&hl=en_US&key=0Aviy_DKJDfv0dDVYUnR5dm56OThxZElSRlZoZ2FjZGc&output=txt&gid=0&range=kml_output&time1=4101881)

Design + Implementation

Keeping in mind the findings revealed during the contextual, technical, and design research, I laid out the following set of design principles and features to refer to during the development of the project.

Design Principles + Features

I. LEGIBILITY OF DATA:

a. Analyzed data from water samples can be entered by UCR staff member.

b. An archive of data is accessible.

c. Units of measurement are accurate and are transparent to users.

d. Visualization displays data comparison in way that users can easily identify differences and similarities.

e. Data can be filtered in varying degrees of granularity.

f. Data can be interacted with by mouse clicks and drags.

II. COMMUNITY ACCESSIBILITY:

a. Visualization is hosted on public website.

b. Locations of water monitoring sites are identified clearly on a map and can be interacted with.

c. Instructions of how users can interact with visualization are clearly displayed on page.

d. Community is defined by data (local residents, politicians, and environmental activists may all value data sets, or the display of data, differently).

III. MANAGEABLE TECHNOLOGY:

a. Staff can enter and edit data and use tool without advanced technical skills (consider API for pulling data from existing database that UCR currently uses, or through GoogleDocs as database).

b. Water monitoring sites can be added/deleted/edited easily by staff.

c. Data can be imported and exported easily.

Design + Development of Actual Artifact

The design of the water-quality interactive data visualization project began in September 2011, development began in January 2012, and the entire project was complete in April 2012.

Initial wireframes were created in September 2011, which outlined the flow of the system, complete with proposed technologies and a visual design mockup.

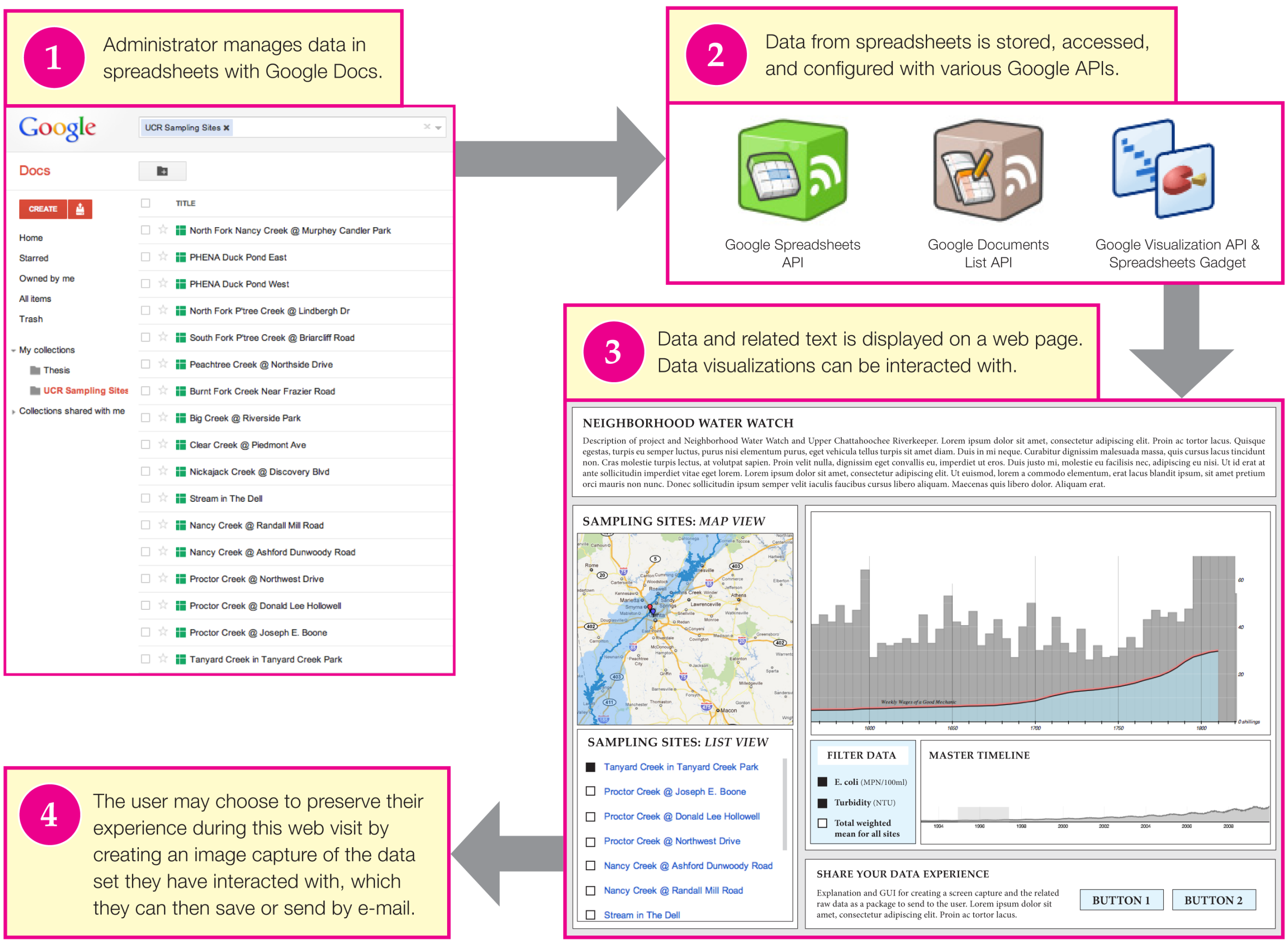


Fig. 19: Initial wireframe for the project

A majority of the features defined in the initial wireframe were implemented, but a few features were eliminated or altered and specific technologies shifted as the development progressed. For example, the need for Google Spreadsheets API and Google Documents API was no longer necessary to achieve the desired back end functionality. Also, parameter filtering methods were abandoned in favor of individual charts for each data parameter.

Big Creek @ Riverside Park S-1946

File Edit View Insert Format Data Tools Help   Last edit was made 5 days ago by nwwucr

Site	Date	Latitude	Longitude	Rainfall (inches)	Turbidity (NTU)	E. coli (MPN/100mL)
Big Creek @ Riverside Park	11/4/10	34.0061	-84.3495	3.50		100
Big Creek @ Riverside Park	1/6/11	34.0061	-84.3495			100
Big Creek @ Riverside Park	1/20/11	34.0061	-84.3495			1970
Big Creek @ Riverside Park	2/3/11	34.0061	-84.3495			1600
Big Creek @ Riverside Park	2/8/11	34.0061	-84.3495		16.70	
Big Creek @ Riverside Park	3/3/11	34.0061	-84.3495			740
Big Creek @ Riverside Park	3/31/11	34.0061	-84.3495			1730
Big Creek @ Riverside Park	6/2/11	34.0061	-84.3495			100
Big Creek @ Riverside Park	9/15/11	34.0061	-84.3495			410
Big Creek @ Riverside Park	9/29/11	34.0061	-84.3495		6.50	375
Big Creek @ Riverside Park	10/13/11	34.0061	-84.3495		7.30	155
Big Creek @ Riverside Park	10/27/11	34.0061	-84.3495		3.52	100
Big Creek @ Riverside Park	11/10/11	34.0061	-84.3495		4.05	1
Big Creek @ Riverside Park	1/5/12	34.0061	-84.3495		5.49	
Big Creek @ Riverside Park	2/2/12	34.0061	-84.3495		5.72	155
Big Creek @ Riverside Park	4/6/12	34.0061	-84.3495	2.00	4.60	200

Add 1 more rows at bottom.

Fig. 20: View of Google Docs spreadsheet for Big Creek @ Riverside Park collection site. Spreadsheet populates data visualization in real-time upon UCR staff entering data.

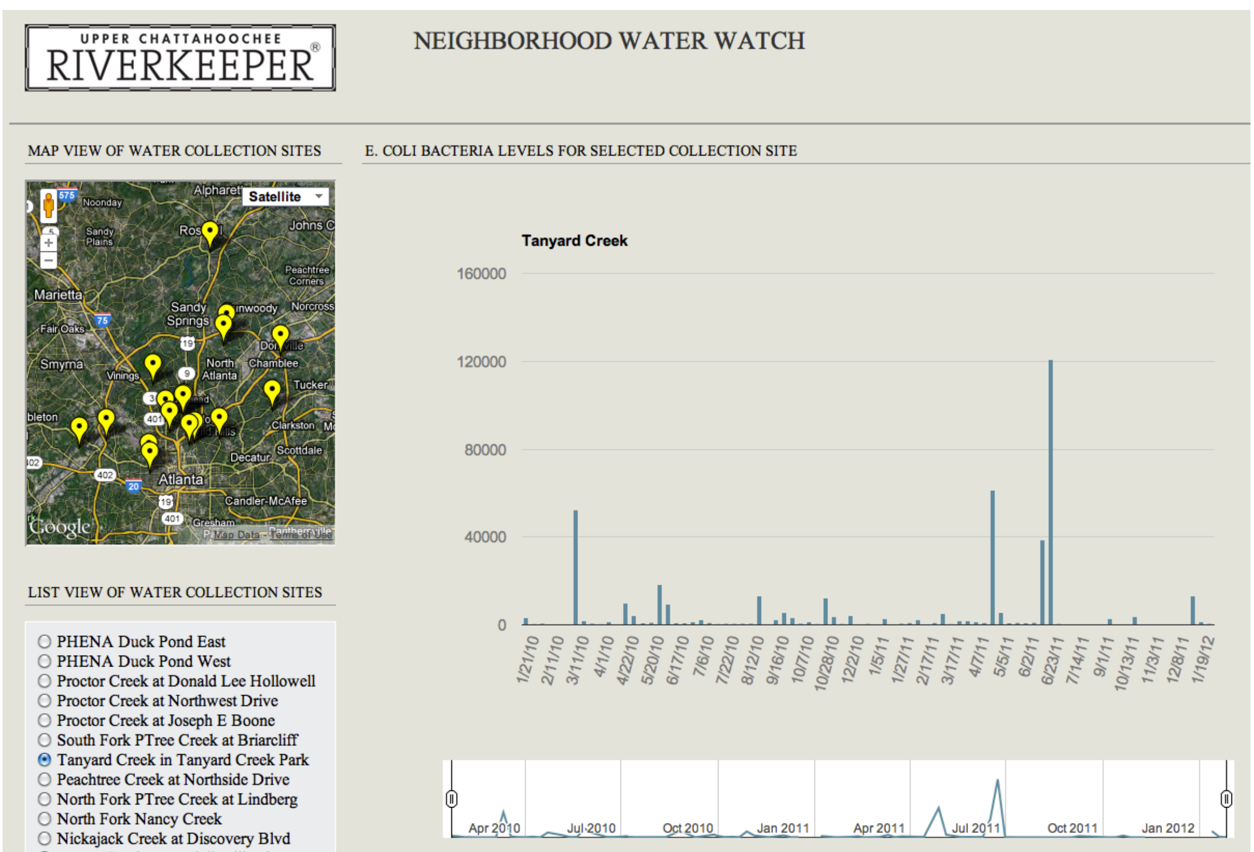


Fig. 21: First draft of the data visualization site.

The project design and technical functionality was iterated upon from the first draft in February 2012 until the final presentation of the artifact in April 2012.

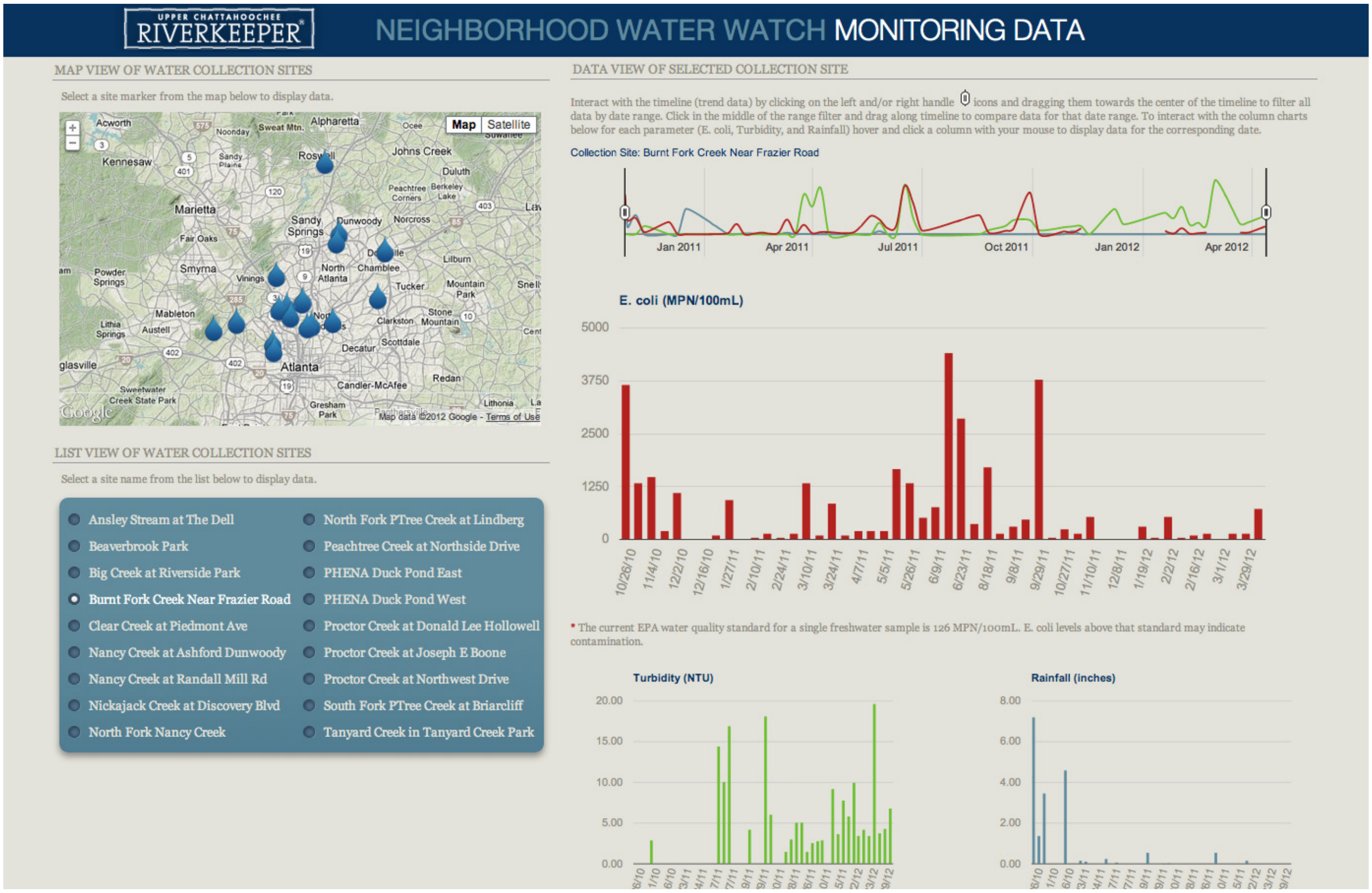


Fig. 22: Final iteration of the data visualization site.

Feature + Interaction Overview

The final digital artifact is a publicly accessible web-based interactive data visualization tool that displays water-quality data collected and analyzed by the Upper Chattahoochee Riverkeeper's Neighborhood Water Watch program. UCR staff enters their data into a Google Docs spreadsheet that populates data on the visualization site. The site allows users to select any of the 18 Neighborhood Water Watch collection sites from a map and list view and displays the corresponding data in a multiple-chart view. The user can then interact with the data, representing E. coli, turbidity, and rainfall levels, by manipulating a timeline slider. The results of a user's interactions displays a unique visualization that is useful to them for professional or educational purposes.

User Evaluation + Feedback

I had the opportunity to conduct formal Georgia Institute of Technology IRB-approved user evaluation sessions with four staff members of UCR in April 2012. Each user session involved the participant using the tool to enter data and interact with the visualizations for the collection site of their choice. The sessions lasted from 20 minutes to one hour and informed design decisions implemented in the final iteration of the project.

After using the tool during the evaluation session, Jason, their technical programs director, whom I worked the closest with throughout this project, commented "It is exactly what we needed". Other critical feedback received during this process has been translated into design improvements such as more distinct chart colors and a more organized list of collection sites.

Future Extensions

Many future extensions have been suggested by UCR, digital media students and faculty, in addition to features that have been on my wishlist for this project. Here is a short list of future extensions that may be implemented:

- Newly monitored Neighborhood Water Watch collection site.
- Chart for parameter called Specific Conductivity, which will soon be measured at each collection site.
- Feature that allows users to export data as an image and spreadsheet.
- Feature that allows users to compare visualizations of multiple sites on one screen.
- Map overlay representing which areas are affected by high E. coli and indicator parameters originating from each collection site.

Immediate future plans for the project include the implementation of this tool by UCR and incorporation of the visualization page on their official website. The organization has also named me as a partner on an EPA grant they have submitted which would allocate funds for this project to be implemented. Lastly, I will package the project as an open source tool targeting other environmental advocacy groups who wish to maintain and share their data.

Conclusion

Water-quality standards have been put in place to ensure the safety of the water we consume and rely upon for recreational and other activities. Data collected by monitoring organizations is available to their constituents and the public, but often the data is displayed under a veil of individual numbers without proper context or on websites that are difficult to navigate.

Through extensive research, design, and development, I have delivered a tool to an environmental advocacy organization that will use it as an administrative and informational data resource in order to promote the health of a local ecosystem.

References

Works Cited

[1] Couch, C., Hopkins, E., Hardy, P. *Influences of environmental settings on aquatic ecosystems in the Apalachicola-Chattahoochee-Flint River basin*. USGS Water Resources Investigations Report 95-4278. Available online at http://pubs.usgs.gov/wri/wrir95-4278/. Accessed April 19, 2012.

[2] Google Fusion Tables: Alon Halevy and Peter Gleick. Available online at http://www.youtube.com/watch?v=9f7LwXjGCSA&feature=player_embedded. Accessed January 16, 2012.

[3] MonRiverQUEST. Available online at http://www.monriverquest.com/map.cfm. Accessed April 19, 2012.

[4] Suda, Brian. .net magazine, November 2010.

[5] United States Geological Survey (USGS). *A Primer on Water Quality*. Available online at http://pubs.usgs.gov/fs/fs-027-01/. Accessed April 12, 2012

[6] Upper Chattahoochee Riverkeeper. Available online at http://www.chattahoochee.org. Accessed April 12, 2012.

[7] *W.Va. Water Research Institute Wins Regional Award*. Available online at http://www.wboy.com/story/16686847/2012/02/06/west-virginia. Accessed April 19, 2012.

[2] *Waters Corporation : Waters NuGenesis 8 with Lab Execution Technologies Improves IT Connection between Lab and Enterprise*... Available online at http://www.4-traders.com/WATERS-CORPORATION-14845/news/WATERS-CORPORATION-Waters-NuGenesis-8-with-Lab-Execution-Technologies-Improves-IT-Connection-between-14210112/.
Accessed March 12, 2012.

Works Consulted

[1] Aufreiter, Michael. Ask Ken: Visual Knowledge Browser. Quasipartikel Labs. http://askken.heroku.com/

[2] Barringer, Matt. “Powering the Earth”. CNN. http://edition.cnn.com/SPECIALS/2009/environment/energy/

[3] Belmonte, Nicolas Garcia. JavaScript InfoVis Toolkit. http://thejit.org/

[4] Bergen, Joseph. What is Your Water Footprint? http://www.josephbergen.com/viz/water/

[5] Centers for Disease Control and Prevention (CDC). http://www.cdc.gov

[6] Russell, Karl, and Charles Duhigg. “Clean Water Act Violations: The Enforcement Record”. The New York Times. September 13, 2009.

[7] Few, Stephen. Now You See It: Simple Visualization Techniques for Quantitative Analysis. 1st ed. Analytics Press, 2009.

[8] Georgia Adopt-A-Stream. http://georgiaadoptastream.com/

[9] Google Maps JavaScript API Version 3 Reference. https://developers.google.com/maps/documentation/javascript/reference

[10] Google Visualization API Reference. https://developers.google.com/chart/interactive/docs/reference

[11] jQuery Documentation. http://docs.jquery.com/Main_Page

[12] NASA Ames Ecological Forecasting Lab. http://ecocast.arc.nasa.gov/

[13] Ojalvo, Holly Epstein. "Teaching with Infographics | Science and Health". The New York Times. 25 August 2010. http://learning.blogs.nytimes.com/2010/08/25/teaching-with-infographics-science-and-health/#more-30521

[14] Polymaps. http://polymaps.org

[15] Sanglard, H., Nägel, H. "An End User Oriented Platform for Scientific Visualization", presented at IEEE Conference on Information Visualization (IV '97), 1997. 1093-9547/97

[16] Stanford Visualization Group. Protovis. http://vis.stanford.edu/protovis/

[17] Tufte, Edward Rolf. Envisioning Information. 3rd ed. Cheshire: Graphics Press, 1992.

[18] United States Environmental Protection Agency (EPA). http://epa.gov

[19] Upper Chattahoochee Riverkeeper. http://www.chattahoochee.org/

[20] Ward, Matthew, Georges Grinstein, and Daniel Keim. Interactive Data Visualization. A K Peters Ltd, 2010.

[21] Woolman, Matt. Digital Information Graphics. New York: Watson-Guptill Publications. 2002.

[22] Yao, Nathan. Visualize This: The FlowingData Guide to Design, Visualization, and Statistics. Wiley, 2011.

Lauren Langley
www.laurenlangley.com

Georgia Institute of Technology
www.gatech.edu