

# **DATA INTELLIGENCE FOR IMPROVED WATER RESOURCE MANAGEMENT**

by

**Mark Ziman**

Dr. Martin Doyle, Adviser

April 29, 2016

Masters project submitted in partial fulfillment of the  
requirements for the Master of Environmental Management degree in  
the Nicholas School of the Environment of  
Duke University

## **EXECUTIVE SUMMARY**

Technological enhancements have decreased the cost of data collection, increased our ability to share data, and expanded our insights concluded from data. These modern abilities, commonly described as big data, are rapidly affecting decision making methodologies across the world. With the increased amount of data present in the 21st century, we are not limited by quantity of information, but rather by our ability to deduce sensible intelligence from the massive amounts and different types of information present. To harness the power of data we must first understand what data we have, how we collect it, and how we can standardize and integrate it. Then we can apply analytical tools to transform the data to information, to knowledge and, finally, to informed decision making.

This research project is an investigation into how the water sector is actively working to integrate big data capabilities into managerial processes in the United States. The content of this report is two-fold. First, the current state of water resources data technologies, trends, initiatives, and opportunities are analyzed and recommendations for advancement are provided. Second, the development of a proof of concept water data application is presented to demonstrate how the water sector can use data to improve managerial decision making.

Water resource management has historically been a data-driven discipline with consistent measurements of water quantity and quality, as those measurements are of concern for environmental and anthropogenic needs. However, mainly due to funding constraints, the water sector has been slow compared to other industries to adopt big data capabilities. Today, water managers' eagerness to adjust systematics is made apparent through their development of initiatives and products to harness the value of big data to improve resource management. The primary example of this is the Open Water Data Initiative, a top-down collaborative approach to create an "open water web" by transforming data management from a one-to-one producer-to-user scheme to a many-to-many scheme. Throughout federal agencies, this initiative is spreading best management practices, including web service machine-to-machine communication and standardized schemas such as Water ML 2.0. In both the private and public sector, products have been developed to serve the data needs of a growing water market.

The availability of water data is inherently connected to regulations that determine who collects data, how data is collected, and where data is housed. The Safe Drinking Water Act and the Clean Water Act are the two primary laws that determine the water quality data landscape of the nation. The stipulations

of these acts present an opportunity to aggregate publically available water quality data, and use it to gain a higher resolution focus of the state of water quality in the nation.

Identification and segmentation of the various opportunities presented by big data enables more effective implementation of the practices. My research presents a series of recommendations to address these opportunities. Firstly, user needs should be better defined so projects can be designed to fulfill specific goals and have a higher probability of producing a sizable impact. To further harness the possibilities presented by big data, all available data should be aggregated. Sensor technology, citizen science data, and automated metering infrastructures are three examples of recently developed data types that could be used to increase the amount of water quality data available. Standardized schemas should be used to enable integrations of available data sources. Finally, analytical tools should be employed to use the available information and translate it into actionable intelligence in decision making processes.

As a model for how available, yet fragmented, data may be organized, aggregated, analyzed, and visualized to add value to a specific purpose, the Water Quality Risk Assessment Tool was developed and is presented in the report. The tool was built for the Duke Nicholas Institute of Environmental Policy Solutions. It is a proof of concept map-based web application that summarizes where, when, and to what extent water quality is out of compliance or trending out of compliance for investors and credit rating agencies. In its current form, the tool uses dissolved oxygen, pH, temperature, turbidity, and specific conductance data from the Water Quality Portal and presents a summary dashboard for the state of Colorado. This tool is designed to be used as a stepping stone for an institution to scale the project to a larger service area with measureable value for its users. It is accessible at [https://mark-ziman.shinyapps.io/WQRAT\\_MZ/](https://mark-ziman.shinyapps.io/WQRAT_MZ/).

The contents of this report assess the strengths, weaknesses, and opportunities for big data capabilities to improve water resource management. This comprehensive review provides fundamental insights for water managers and water investors to understand the water data framework and capitalize on the modern opportunities for advancement presented by big data.

# TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	i
INTRODUCTION .....	1
Flint Michigan, A Case Example .....	1
Research Goals.....	2
DATA INTELLIGENCE OF THE 21 <sup>ST</sup> CENTURY .....	3
Modern Big Data Capabilities .....	3
Big Data in the Context of Water Resource Management .....	4
<i>Types of Water Data</i> .....	5
<i>Data Sharing Methodologies</i> .....	6
<i>Sharing Initiatives</i> .....	7
<i>Data Service Case Examples</i> .....	9
WATER QUALITY GOVERNANCE AND MANAGEMENT.....	12
The Safe Drinking Water Act.....	13
Clean Water Act .....	15
OPPORTUNITIES FOR IMPROVED WATER RESOURCE MANAGEMENT .....	17
Better Definition of User Needs.....	17
Increased Sample Frequency .....	17
More Data Integration .....	18
Conclusion of Recommendations .....	19
THE WATER QUALITY RISK ASSESSMENT TOOL .....	20
Introduction .....	20
<i>What Is Water Risk and Who Does It Affect?</i> .....	20
<i>Comparable Water Risk Assessment Tools</i> .....	21
<i>Objectives of the Water Risk Assessment Tool</i> .....	22
Methods.....	23
<i>Data</i> .....	23
<i>Application Processes</i> .....	25
Results.....	28
Discussion.....	36
CONCLUSION.....	37
APPENDICIES .....	38

A.	North Carolina Community Surface Water (and GWUDI) System Monitoring Requirements .....	38
B.	Nutrient Sensor Challenge Participants .....	39
C.	Water Risk Assessment Tool Code .....	40
	<i>Python Code to Collect Data with Web Services .....</i>	<i>40</i>
	<i>R Code to Clean Data (Preparation for Shiny Application).....</i>	<i>41</i>
	<i>R Code to Run Shiny Application (UI and Server Included).....</i>	<i>56</i>
	REFERENCES.....	64

# INTRODUCTION

## Flint Michigan, A Case Example

Flint, Michigan was declared a state of emergency separately by city, state, and federal governments between December 2015 and January 2016.<sup>1, 2</sup> After switching from Lake Huron water supplied by the Detroit Sewer and Water District to locally treated and distributed Flint River water, tap water's lead concentration increased to harmful levels. The city of Flint, with administrative oversight by state officials, did not add necessary anticorrosion agents. Consequently, lead was leached from a portion of the city's distribution pipes, and residents were exposed to toxic concentrations of lead in the drinking water.

The public health crisis in Flint was intrinsically linked to water quality data management. Flint lacked a modern database of lead water lines; records of the lead pipes in the distribution network were stored on about 45,000 index cards in filing cabinets at Flint's public utility building.<sup>3,4</sup> The city was in the process of transitioning the data to electronic spreadsheets, but only 25% had been digitized. Water managers were unable to simply, yet accurately determine where consumer taps were most at risk from traveling furthest through potentially leaching lead pipes. Some reports also claim that Flint water managers employed questionable data cleaning methodologies that omitted outlier sample measurements.<sup>5</sup> Consequently, Flint officials inaccurately concluded that lead concentrations were below the federal threshold of 15 parts per billion of lead and were thereby not required to alert their customers and implement an anti-corrosion plan. Later, Flint water management officials were dismissive of the residents' and outside research groups' concerns because the outside groups' evidence was incongruent with the officials' data.<sup>6</sup>

The water quality sampling and processing methodologies did not properly reflect actual contaminant levels. With faulty data practices, Flint water managers did not change treatment methods and consequently put customers at risk.

Problems similar to Flint are present across the country. Since the Flint water crisis, many other cities have raised comparable questions about their water quality management.<sup>7</sup> Many utilities are struggling to upkeep their systems while dealing with budget constraints.<sup>8</sup> This is a particularly difficult challenge to address at a time when many infrastructure components are deteriorating and reaching the end of their anticipated lifetime. Water utilities are eager to identify ways to fund and implement updates within their management system. Modern big data capabilities, such as sensor technologies, integrated relational databases, and advanced analytics, offer effective means for public water systems to more effectively manage their infrastructure.

## **Research Goals**

With the increased amount of data present in the 21<sup>st</sup> century, we are not limited by the quantity of information, but rather our ability to deduce sensible intelligence from the massive amounts and types of information present.<sup>9</sup> Like in most sectors, there is great interest in water sector to utilize the potential of big data to effectively improve water resource management. In order to apply big data capabilities to water resource management, it is necessary to understand the landscape of water data management—who produces the data, where and how the data are stored, how the data are transferred, how the data are analyzed, who uses the data, and how the data are used. This research project is an investigation into water related big data and advanced data analytics in the United States. The purpose of this report is two-fold.

First, the current state of water resources data is explored to identify strengths, weaknesses, and areas of potential improvement. This entails an extensive review of literature about data technologies, trends, initiatives, and opportunities. The product of this research is a summary of big data capabilities and the identification of water quality as the optimal aspect of water resources to first apply big data capabilities for maximum return on investments. An overview of water quality regulation as it relates to data management is included to provide context for current water quality information management. Finally, a series of recommendations for increased utilization of sensor technologies and analytical tools are provided to help guide water professionals toward the most effect steps of management improvement.

Secondly, a water quality risk assessment tool is developed as a proof of concept to demonstrate how raw data for selected water quality parameters can be transformed into actionable intelligence. The tool is designed to provide municipal bond investors and credit rating agencies with quick, yet effective summaries of water quality related risk on a map-based interface. The tool uses automated web services to collect water quality data from multiple sources for the state of Colorado. The application has an interface which allows users to easily query the database and see where, when, and to what extent water quality measurements have been out of compliance.

## **DATA INTELLIGENCE OF THE 21<sup>ST</sup> CENTURY**

Today, data measurements are more likely to be collected by smart phones and satellites than by hydrologists and geologists.<sup>10</sup> Data are no longer solely collected by scientists with an intentional purpose; rather, many types of data are produced by many sources, some operating twenty-four hours a day, seven days a week. In order to apply the power of data analytics to management challenges, we must first understand what data we have, how it's collected, and how it can be integrated. Only after having a thorough understanding of the data landscape can we apply analytical tools process the data and transform it from data to information to knowledge to decision making.

To address this transformation, I conducted an extensive review of the water management sector as it relates to data management with a focus on water quality data. Research identified data providers, data users, and intermediaries in the process of data production and sharing. Data collection and reporting incentives and requirements were identified. Industry trends for increasing data management capabilities were also recognized. This research helps paint a picture of the water sector's current data management situation and understand where and how initiatives can use data to improve water resource management.

### **Modern Big Data Capabilities**

Modern technology enables organizations to collect and process information at an unprecedented rate. Readily available software and hardware permit quick and cheap data management. The catchall term to describe this capability is "big data." Capabilities provided by big data are rapidly altering operational methodologies and decision making processes. Big data is best summarized by the five Vs: volume, velocity, variety, veracity, and value.<sup>11</sup> These terms encapsulate the functional differences and opportunities between big data and traditional data.

Volume refers to the amount of data generated, processed, and stored. The first gigabyte hard drive was built in 1980 and weighed over 500 pounds. Today, Google Inc. processes an average of 20 petabytes of data per day (1 petabyte =  $10^6$  gigabytes). The cost of data storage has dramatically decreased to where a terabyte of storage can be purchased today for about \$100. For most organizations, data volume limits are nonexistent. For example, in water resource management, state departments of health and environmental quality have the ability to electronically store all utilities' water quality compliance reports.

Velocity refers to the speed data is transferred and processed. Information can now be moved at nearly an instantaneous rate. This enables the cleaning, transmitting, storing and processing of data all in one motion rather than in discrete, non-continuous segments. For water utilities, real-time and online



monitoring systems have improved intake protection, control operations, security, and providing information to customers.<sup>12</sup>

Variety refers to the many types of data available. Historically, data was structured into neatly organized databases, but today big data technologies manage and analyze unstructured data such as internet searches, social media conversations, or satellite imagery. Integrating relevant datasets can provide a more comprehensive and informative picture of a system. In water resource management, there are efforts to integrate all types of water data from across different departments into one, continuous data stream.

Veracity refers to the accuracy of data. The rise of non-traditional data sources and increased data sharing raise issues of quality control. Some organizations are hesitant to share data for fear that not all intricacies may be included in metadata. Veracity causes a risk of over-engineering models and concluding trends from inaccurate data sources. In water management, there are initiatives to use citizen science produced data, but there is controversy about the quality of the data and how an organization may be held liable for incorporating inaccurate data into faulty management decisions.

Value refers to the actionable intelligence gained from data. Because the power of conclusions are partially determined by the volume of evidence which supports them, we gain more decisive insight from big data capabilities. Advanced models and algorithms “mine” massive amounts of data and produce information which informs decision making. In water management, the National Flood Interoperability Experiment processed water quantity gauges in near real time in order to provide local emergency response with flood information services with the potential to avoid catastrophes.<sup>13</sup>

## **Big Data in the Context of Water Resource Management**

Water management has historically been a data rich subject. As a critical resource for agriculture, transportation, and health, attributes of water are scrupulously measured. The Nile River’s depth, for example, has been measured for at least 5000 years to manage flood plain agricultural systems.<sup>14</sup> Measured data is used to make decisions about how to best partition and move water between its many uses. For example, based on snowpack, streamflow, and storage reservoir measurements, the Northern Colorado Water Conservancy District determines how to supply municipal, agricultural, and industrial water usage.<sup>15</sup> Due to climate change, growing populations, and increasing water demands, water resource management is becoming increasingly important. With the advent of new technology, water, a dynamic and ever changing resource, can now be sampled at higher frequency and accuracy. Big data

capabilities, volume, velocity, veracity, variety, and value, provide an opportunity to improve our understanding of water resources, and manage it better.

### ***Types of Water Data***

It is efficient to address water data in two stages. *Primary* data refers to the collection of raw data, normally water quality or water quantity measurements. These may be measurements of streams or pipes collected by technicians or sensors (automated sampling technologies). Water quality refers to the chemical, physical, and biological characteristics of water.<sup>16</sup> It is a measure of the condition of water usually in reference to the requirements of some ecological process or anthropogenic purpose. Water quality is not a single measurement; rather it is a latent factor related to hundreds of water characteristics. On the contrast, water quantity refers to a volume of water or a rate at which a volume of water is moving downstream.<sup>17</sup> Water quantity data is frequently linked to other aspects of water management such as water rights or the fiscal value of water. Historic records of water quantity are more consistent than historic records of water quality.<sup>18</sup> Both quality and quantity measurements are evolving from time intensive sampling methodologies requiring discrete samples by technicians to automated, more frequent processes performed by sensors. Today the US Geological Survey (USGS) has 1,908 sample locations across the country that measure water quality and transmit data on 15 to 60 minute fixed intervals with the use of automated recording equipment.<sup>19</sup> Sensor measurements are beneficial because they significantly increase the amount of data available for water quality, an attributed that can quickly change and have drastic anthropogenic and environmental consequences.

The primary objective of water quality and quantity monitoring is to characterize variability. Therefore, it is important that individual measurements are suitably comparable across time and space. This requires a standardization of measurement practices. Although every measurement is somewhat unique, specialized techniques and technologies are designed to minimize variability and favor comparability.

Water is measured through many different methods and is recorded in many different formats. For example, staff gauges are used to measure river depths; sensors are used to measure nitrogen content; and satellites are used to measure groundwater volume. Depending on water characteristics and resource availability, water is measured differently among regions. The central dogma is that the more data that can be collected, the better secondary data processes can function.

*Secondary* data is information derived from direct hydraulic measurements or sensors. Unlike standardized primary data measurements, secondary data tends to be more customized to specific

circumstances and desires. For example, models, methods of taking primary data from a variety of sources and compiling them to estimate another variable, can use historic precipitation and stream flow measurement networks to predict water availability for a given municipality. Other models determined how releasing water from a reservoir would provide for fish habitat downstream. Models are frequently connected to regulatory and institutional frameworks, such as water rights, instream flow requirements, and environmental permitting.

There is much more variety among derived water data than primary water data. Some are small scale, functional for a local watershed. Others are larger, applicable to regions like the Colorado River watershed. Attributes of the secondary data are dependent on the application's sponsoring entity, mission, and budget priorities. Therefore, it is frequently easier to share primary data than secondary data.

### ***Data Sharing Methodologies***

Water resource data is notorious for its fragmentation—data is produced by many entities and stored in many locations.<sup>20</sup> To overcome this fragmentation, frequently, a given organization will use data provided by another, or multiple other, organizations to gain insight on a desired subject. However, data sharing is a resource-intensive task and generally regarded as a roadblock toward big data capabilities. Consequently, data sharing is a “hot subject” for water data management and data sharing systematics are evolving from one-to-one, to one-to-many, to many-to-many methodologies.

In the simplest sense, data is generated by one entity and used for a single purpose. This could be an academic research study or a contracted consulting project. This is referred to as one-to-one data sharing. Data are also generated by one entity and provided to many users for many purposes. This one-to-many data sharing model is well represented by the USGS National Water Information System (NWIS), which provides stream flow data for almost 10,000 sites across the country.<sup>21</sup> While all information is collected in-house by USGS, this data is used by many entities, including utilities, academic researchers, and private industry. Historically, the one-to-one and one-to-many methodologies dominated the water data landscape, but now, due to technological enhancements associated with big data, water data is shifting to a many-to-many sharing framework. This carries opportunity of high resolution data networks that lead to more precise managerial decisions.

The transformation from a one-to-one to a one-to-many to a many-to many data sharing framework requires a series of technological implications.<sup>22</sup> First, managing real time data is systematically different

than managing discrete sample data. Generally, real time samples are produced every fifteen minutes and uploaded every couple of hours. Consequently, database storage capacity requirements are orders of magnitude greater and the bandwidth required to upload and download real time data is significantly larger. Modern data is traditionally warehoused in cloud-based systems where rather than storing data in a local directory; information is physically stored in an offsite location that can be accessed remotely. Cloud-based computing services are becoming a standard; analysts can process data remotely rather than first downloading data and then processing locally. To download and upload information water managers are commonly using web services, a machine-to-machine communication system that, when properly performed, is faster and easier than having a person download and upload information.<sup>23</sup> Web services allow for information sharing to be more automated and performed frequently keeping datasets up to date. Among these technological advancements, the largest roadblock to data sharing is inconsistent schemas, or formatting.<sup>24</sup> Data of different schemas cannot be compiled or separated without intermediate processing steps. The actions required to transform data into a consistent format are quite resource intensive<sup>25</sup> and a consequential deterrent to data sharing. An efficient data sharing network requires a standardized schema.

When sharing data, there is an increased risk of problems caused by unknown collection methods or unknown data intricacies.<sup>26</sup> Data collection typically includes a series of assumptions specific to a project mission. Those assumptions may skew results when those data are applied to a different purpose. Quality control issues are greater in secondary data than in primary data.

All datasets should include detailed metadata, descriptions of data fields and collection methods. However, metadata rarely documents all critical attributes of the data. A human element of understanding is lost when data is transferred. Quality control issues incentivizes water managers to use government produced data, such as USGS water quantity and flow gauges because the data is upheld to high standards of quality control. Quality control questions raised by sharing data have liability implications for data users.

### ***Sharing Initiatives***

The water sector is ambitiously pursuing enhanced data management, but it is limited by data sharing restrictions. Because an increased data sharing infrastructure would enable big data capabilities and increased actionable intelligence, multiple initiatives have recently arisen to create an “open water web.”<sup>27</sup>

The Open Water Data Initiative (OWDI) is a top-down collaborative effort to make water data more integrated and accessible to suit modern data availability needs. The OWDI aims to connect currently fragmented water information into a national data framework by leveraging existing systems and tools to underpin innovation, modeling, data sharing and solution development. The initiative was officially launched in summer 2014 by the Federal Geographic Data Committee and the Advisory Committee on Water Information.<sup>28</sup> The conceptual framework of the OWDI includes water data cataloging, providing water data as a service, enriching water data, and fostering a community for water data and tools. To adhere with the OWDI, federal agencies, in addition to state governments and local municipalities, are working together to make their data more available and easier to synthesize. The Integrated Water Resource Science and Services (IRSS) is the OWDI’s business model for collaboration of federal agencies with complementary missions in water science, observation, management, and prediction.<sup>29,30</sup> Initial IRSS members include US National Oceanic and Atmospheric Association (NOAA), USGS, and the US Army Corps of Engineers. The goals of OWDI are valiant, but the initiative is not funded, and progress has been slow.

<b>Open Water Web</b>			
<b>water data Catalog</b>	<b>water data <i>As a Service</i></b>	<b>Enriching Water Data</b>	<b>water data and tools <i>Marketplace</i></b>
Find source data	Consensus standards	Include routing	Community exercise of tools & data
Create water & climate themes	Visualization and delivery	Coupling with models	Data usage tracking
Recruit/engage partners	Catalog and serve	Grounded to geofabric	Community-built extensions (eg map)

**Figure 1.** A summary of the components of the open water web; data cataloguing, servicing, enriching, for a market purpose. The open water web serves as the fundamental framework of the OWDI.<sup>31</sup>

The OWDI is a recently launched, official federal initiative, but the idea of an open water data network has been postulated by a decades-old grassroots movement. The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) is a non-profit research organization, incorporated in 2001, which aims to strengthen multidisciplinary collaboration and advance water research, education, and training. CUAHSI has helped inspire the water resource management sector to actively pursue integrated water data. CUAHSI hosts the Water Data Center, a database of about 100 data sources, 400 million observations of 2,500 measured properties, at 1.2 million sample locations. The data are accessed by the HydroClient online application or WaterOneFlow web services in the WaterML 2.0 (water markup

language version 2.0) schema. CUAHSI data services are not large or advanced enough to effectively provide valuable, shared water data for the entire nation. However, they provide a valuable, small scale model for how the water resource sector could aim to integrate.

There have been many efforts to standardize the format of water data, an aspect many professionals regard as the biggest barrier to data sharing. In 2013, the Open Geospatial Consortium, an international data standards development community, released the WaterML 2.0, a standardized data schema designed specifically for continuous monitoring data<sup>32</sup>. The schema prescribes vocabularies for categorical quality assertions, mediums sampled, observation processes, and interpolation methods.<sup>33</sup> The WaterML 2.0 data schema is the most widely accepted, standardized water data schema within the water resource sector.

### ***Data Service Case Examples***

Data services are a growing segment of the water sector. Both public and private initiatives have developed data-centric tools that enable enhanced water management. These tools commonly attempt to aggregate fragmented data sources to gain comprehensive and actionable insight. Examples of data services are described to provide specific examples of how knowledge can be improved by combining data, not subdividing them.

In 2007, the US Environmental Protection Agency (EPA) released the Water Quality Exchange (WQX), a platform for sharing water quality data between various data providers. The WQX was built upon the Environmental Sampling Analysis and Results (ESAR) standards developed by the National Water Quality Monitoring Council. It has become the de-facto standard for communicating discretely sampled water quality data. The WQX led to further integration between federal agencies by allowing the USGS, United States Department of Agriculture (USDA), and EPA to share water quality sampling data through the Water Quality Portal. The WQX, and the Water Quality Portal, are the largest publically available water quality databases in the nation and their data can be readily accessed through web services. However, the WQX was not designed to manage continuous data generated by water quality sensors—the most modern and valuable of water data producers. The EPA is currently working to modify the WQX to properly host continuously generated data.<sup>34</sup> Although the WQX is large, it is far from a fully integrated national water dataset. There exist many publically available water quality data sources, which are not included in the WQX. Furthermore, navigating the application's interface is not intuitive and collected data requires thorough post-processing. Interviewed water utilities reported that they do not use the WQX; instead, they prefer to get raw data from the EPA and combine it in house with their own datasets.

Another federal government water data service is the USGS' National Ground-Water Monitoring Network (NGWMN), a compilation of selected groundwater monitoring wells from federal, state, and local groundwater monitoring networks across the nation.<sup>35, 36</sup> The associated NGWMN data portal is a web-based mapping application which provides access to water levels, water quality, lithology, and well construction. NGWMN helps establish baseline conditions and long-term trends in water levels and quality. The participation in the network was voluntary and the barrier of entry was set relatively low, but the USGS attempted to maintain an adequate level of data quality for reuse and comparability.

In addition to intra-agency data integration initiatives, there are inter-agency data integration initiatives. The National Water Center (NWC), a National Weather Service facility opened in 2014, is a flagship of the OWDI and IRSS, and supported by CUAHSI. The NWC serves as a central operating center for interagency collaboration on data capabilities, models and decision-support tools. From September 2014 to August 2015, the NWC hosted the National Flood Interoperability Experiment, a model capable of near-real time surface water flood characterization at a resolution which facilitates decision making within the emergency management community.<sup>37</sup> The National Flood Interoperability Experiment uses web services to integrate geospatial hydrologic framework, high resolution hydrologic forecasting, flood emergency response planning areas, and real time river channel information.<sup>38</sup>

However, not all data integration initiatives are sponsored by federal agencies. In 2009, the Western States Water Council initiated the Water Data Exchange (WaDE) to address long-term issues related to water data availability for western states. The objective of WaDE is to allow access to water data from state agencies within the Western Governors Association, the U.S. Department of Energy, the Department of Energy National Labs, and the Western Federal Agency Support Team.<sup>39</sup> WaDE framework uses web services to allow users to retrieve desired datasets directly from a state's database, rather than transferring all state data to a central repository. This modern framework is efficient and should become standard methodology for sharing water data. WaDE currently accesses water rights data from give states, but the interface is difficult to navigate and its current value is questionable.

The private sector has also introduced initiatives to add value to the water sector through data integration services. Water Sage, first developed by Ponderosa Advisors LLC. In 2013, is a map-based web portal that allows users to easily search and view water rights.<sup>40</sup> The application is built on large databases which include all water rights data from multiple states' agencies, such as the Department of Water Resources. The system intuitively and quickly automates search processes which typically require a specialized water

attorney hours to perform. In western states where water rights are a contentious issue, Water Sage brings transparency to complex management structures.

Similarly, Kisters<sup>41</sup> and Aquatic Informatics<sup>42</sup> are companies which provide water specialized data management softwares. Kisters' product WISKI (Water Information Systems KISTERS) and Aquatic Informatics' product AQUARIUS both automate data collection, analyses, visualization, and publishing, to enable water managers to work with their data quicker and support better decisions. These softwares are not a platform for sharing or providing data; rather, they are a platform for an entity to more efficiently work with their own data. Kisters and Aquatic Informatics are the largest companies which provide data management software products, but many others exist, such as Data Concourse<sup>43</sup>, Data Forensics<sup>44</sup>, Accelerated Technology Laboratories<sup>45</sup>, and Promium<sup>46</sup>. The amount of companies providing environmental data management products indicates the demand in the market for increased data management capabilities.

A captivating new use-case for metadata management of water data is emerging.<sup>47</sup> An increase of data available due to technological advancements such as real time sensor monitoring networks in addition to a transition one-to-many to many-to-many data sharing systems provide new opportunities for advancement. The prolific amount of water data available, and its importance to government and private entities managing anthropogenic and environmental systems, provides incentives to use available data to its best possible use. Initiatives to integrate and better use data are present in both the public and private sector. With currently available tools, and many more in development, the water manager can use data-informed intelligence to optimize the many decision making processes related to water resource management.



## WATER QUALITY GOVERNANCE AND MANAGEMENT

Water quality management is a data-intensive aspect of water resource management. Water quality is comprised of hundreds of chemical, physical, biological, and radiological characteristics measured on the order of minutes to years.<sup>48</sup> The USGS has over 1000 sensors which measure specific conductance about every 15 minutes<sup>49</sup>; the North Carolina Department of Environmental Quality requires community surface water systems to routinely measure lead concentration every three years.<sup>50</sup> Water quality components are consistently monitored by water managers because poor water quality can lead to devastating public health and environmental consequences, such as lead contamination and eutrophication. In contamination spills, like the unplanned discharge of 10,000 gallons of latex into the Potomac River in September 2015, water managers increase monitoring of water quality characteristics to accordingly alter treatment practices.<sup>51</sup> Big data capabilities provide a significant opportunity to streamline data management practices so that managers can more efficiently and effectively translate primary water quality measurements into actionable intelligence.

Water quality monitoring is specific and resolute. Water quality analyses relate to a purpose, as summarized in Table 1.

**Table 1.** Types of water quality analyses sorted by purpose. Table summarizes specific purposes of water quality analysis.<sup>52</sup>

Type of Water Quality Analysis by Purpose				
Constituent	Sanitary	Industrial	Geochemical	Agricultural
Bacillus coli, streptococci	x			
Biochemical oxygen demand	x			
Chlorine, hydrogen ion	x			
Dissolved oxygen, water temperature	x			
Oxygen consumed	x			
Odor	x			
Organic and albuminoid nitrogen (N)	x			
Inorganic nitrogen as NH <sub>4</sub> , NO <sub>3</sub> , NO <sub>2</sub>	x			
Heavy metals (Cu, Pb, Sb, Zn)	x			
Color	x	x		
Suspended matter, loss on ignition	x	x		
Total and dissolved loss on ignition	x	x	x	
Turbidity	x	x	x	
Total hardness	x	x	x	x
Alkalinity, acidity, carbon dioxide	x	x	x	x
Cations (Al, Ca, Fe, K, Li, Mg, Mn, Na, Si)		x	x	x
Anions (Br, Cl, CO <sub>3</sub> , HCO <sub>3</sub> , I, NO <sub>3</sub> , SO <sub>4</sub> )		x	x	x
Boron				x
Range in number of determinations	16 - 30	8 - 27	6 - 24	5 - 21

Throughout the 1800s, water quality assessment techniques were first developed and water quality monitoring and treatment became standard practices of municipal water providers.<sup>53</sup> In 1924, the USGS published the first national water quality survey which included about 200 rivers and 10 lakes. Due to notable water quality issues such as the Cuyahoga River fires, the federal government passed a series of laws to improve water quality management, including the 1956 the Federal Water Pollution Control Act, the 1969 National Environmental Protection Act, the 1970 creation of the EPA, the 1972 Clean Water Act, the 1974 Safe Drinking Water Act, and the 1976 Resource Conservation and Recovery Act.

Today, the United States' water quality management methodologies are primarily based on laws passed in the 1960s and 1970s. Although regulations and policies are regularly updated, it is worth noting that the framework for federal water quality management was developed well before big data capabilities were largely available.

The type, volume, and availability of public water systems' water quality information are governed by regulation. Compliance with regulations determine how big data analytics can be applied to water quality information to gain insight from large amounts of water quality data. The most impactful federal laws related to water quality are the Safe Water Drinking Act and the Clean Water Act.

### **The Safe Drinking Water Act**

The Safe Drinking Water Act (SDWA) is the principal federal law which intends to ensure safe drinking water for the public by protecting against both naturally-occurring and man-made contaminants.<sup>54, 55</sup> Originally, the SDWA applied primarily to drinking water treatment facility processes. Amendments in 1996 increased SDWA's scope to include source water protection, operator training, water system improvement funding, and public information correspondence. More than 170,000 public water systems (including community water systems, non-transient non-community water systems, and transient non-community water systems<sup>56</sup>) in the US must comply with drinking water standards authorized by the SDWA. The SDWA policy, regulation, and enforcement are developed and implemented at both federal and state levels.

At a federal level, the SWDA authorizes the US EPA to set national standards for drinking water based on scientific conclusions, available technologies, and costs. The US EPA's National Primary Drinking Water Regulations set mandatory and enforceable maximum contaminant levels currently for 77 contaminants; seven microorganisms, four disinfectants, three disinfection byproducts, 16 inorganic chemicals, 43 organic chemicals, and four radionuclides contaminants.<sup>57,58</sup> Additionally, the US EPA establishes National

Secondary Drinking Water Regulations, a set of 15 non-mandatory water quality standards related to aesthetic, cosmetic, and technical effects that do not pose a risk to human health.<sup>59, 60</sup> The US EPA also provides guidance, assistance, and public information about drinking water quality to states and water systems.

Typically, the US EPA does not directly oversee water systems. All states, except for Wyoming and the District of Columbia, have met a series of minimum water quality regulation standards and qualified for *primacy*, the authority to implement the SDWA within their jurisdictions.<sup>61</sup> Most states enforce the SDWA through a department dedicated to public health and environmental quality (ex: the North Carolina Department of Environmental Quality or the Colorado Department of Public Health). The US EPA has the authority to audit state departments, but most regulatory enforcement happens within state jurisdictions. Regulations vary state to state, but all are required to meet the US EPA National Primary Drinking Water Regulations.<sup>62</sup>

Public water system water quality standards are intricate requirements which typically require significant human resources for compliance.<sup>63</sup> Sampling, analyzing, and reporting protocol are specific for each contaminant.<sup>64, 65</sup> Depending on the contaminant, daily, monthly, or annual monitoring may be required. Some reports are submitted as electronic database uploads to a state web portal; others are mailed paper reports. For a given contaminant, raw data may be provided or summary statistics may suffice. For a summary of North Carolina community surface water monitoring requirements, see Appendix A. Under federal regulation, public water systems are also required to publish annual Consumer Confidence Reports that summarize water quality information for the public.<sup>66</sup>

The state public health and environmental quality department reviews water quality reports for regulatory compliance.<sup>67</sup> Reported water quality data is all public information, but may not be easily accessible or manipulated into a format available for further analysis. Therefore, it is difficult to translate the available water quality data into actionable intelligence that can benefit managerial decisions. For example, in North Carolina the Safe Drinking Water Information System Drinking Water Watch is a North Carolina Department of Environmental Quality online database which houses all public water system reported parameters.<sup>68</sup> Although this system publicly accessible, it is difficult to navigate the interface, query information, understand what information is important, and export the information to a secondary system.

Not all information collected by public water systems is reported to state public health and environmental quality departments for SWDA compliance. Frequently, drinking water treatment centers will collect more data than that reported to the state. For example, although not required by the Colorado Department of Public Health, Denver Water collects algae samples at local reservoirs to monitor eutrophication.<sup>69</sup> The amount of data collected but not reported is specific for each water quality metric. Under the Freedom of Information Act, all public water treatment center data is available to the public to an extent, but this information may not be available in a format to which one can easily apply big data analytics.<sup>70</sup> The format of provided data is a critical component of data analytics as it is much less resource intensive to integrate data provided by web services in a standard schema rather than a pdf summary, such as an annual report.

## **Clean Water Act**

The Clean Water Act (CWA) “establishes the basic structure for regulating discharges of pollutants into the water of the United States by regulating quality standards for surface waters.”<sup>71</sup> The clean water act is an extensive piece of legislation with many implications on environmental management, from delineating geographic federal jurisdiction to developing total maximum daily load programs.

Section 402 of the Clean Water Act has direct impacts on surface water quantity and water quality monitoring data. Section 402 authorizes the National Pollutant Discharge Elimination System (NPDES) permit program to regulate point source pollutant discharges into waters of the United States.<sup>72</sup> The US Army Corps of Engineers and the EPA administer the program.<sup>73</sup> Similar to SDWA implementation, federal agencies set minimum regulatory standards and authorize state departments to perform many of the permitting, administrative, and enforcement aspects of the NPDES program.

Per CWA regulations, industrial wastewater, municipal wastewater, and storm water discharges are required to obtain a 402 permit.<sup>74</sup> Generally, utilities’ wastewater treatment facilities are required to have a 402 permit to discharge effluent, but this varies by region, utility customer base size, and type of permit (general or individual).<sup>75</sup>

NPDES permits are specific to the discharging entity (or entities) and the water body receiving effluent.<sup>76</sup>

<sup>77</sup> For example, a NPDES permit for a chemical plant discharging into small river will be very different from a municipal utility discharging into an oceanic harbor. A NPDES permit will generally specify acceptable levels of pollutants, and may require best management practices. These are commonly known as water quality-based effluent limitations and technology-based effluent limitations. To comply with the permit,

the permittee is required to sample discharge and report water quality data to the overseeing agency. Similar to the SDWA compliance records, the NPDES water quality data is public information, but may not be easily accessible.

Clean Water Act NPDES program compliance requires production of water quality data, but associated data is difficult to aggregate. NPDES permit compliance related water quality data may be reported in different formats (pdf, electronic database, etc.) and schemas depending on the region, permittee, and project. Furthermore, this research did not identify a central repository for NPDES program data. Therefore, it would likely be difficult to acquire and integrate water quality data associated with the NPDES program. Due to these limitations, NPDES program data is not an ideal place to search for water quality data which could easily lend itself to the capabilities of big data and supply improved data intelligence.

## **OPPORTUNITIES FOR IMPROVED WATER RESOURCE MANAGEMENT**

The capabilities of big data provide many opportunities for improved water resource management. Identification and segmentation of the various opportunities enables more effective implementation. The following is a summary of the highest potential opportunities identified by this research.

### **Better Definition of User Needs**

Although water professionals frequently express their desire for improved data intelligence,<sup>78</sup> they are rarely able to define specifically what they want. My research was unable to identify a specific benchmark or an “ideal” data gathering or processing scenario for federal or state agencies. The drive for more data and increased data intelligence frequently lacks specific objectives for which metrics can be applied in order to measure program success and financial benefits of increased data intelligence.

In order to identify successful water data ventures, it is critical to identify specific parties with specific data needs. Data intelligence initiatives may undergo difficulty gaining traction within a public water system which traditionally relies on capital expenditure projects related to physical infrastructure such as pipes and reservoirs. Commonly, funds are dedicated to one-time capital costs, but not to ongoing operations and maintenance, such as improving data management practices.<sup>79</sup> To effectively address this challenge, water organizations should specifically define their data goals with precise estimates of returns to the organization. Developing goals that are also measurable would further increase likelihood of funding data projects. With clear proposals, data-related projects can be more effectively compared with traditional infrastructure projects.

### **Increased Sample Frequency**

Modern technologies are providing opportunities for increased data variety and volume. To access all intelligence presented by big data, new sources of data should be incorporated into water resource management. Sensor technology, citizen science data, and automated metering infrastructures are three examples of initiatives to increase the amount of water quality data available.

The water sector is highly anticipating improved in-situ technologies, ranging from individual high-resolution nutrient sensors to water quality sensor networks.<sup>80</sup> Sensors are placed through watersheds to give downstream users and indication of upstream water quality. This is important for reacting to spills as well as standard water treatment operations. Water treatment facilities are able to refine their operations when they better know the characteristics of the water entering their system. The USGS

operates approximately 1,900 real time water quality sensors across the country,<sup>81</sup> which most typically are capable of measuring dissolved oxygen, pH, electric conductance, temperature, and turbidity.

The logistical ease of new, real time water quality sensors reduces cost per data point by orders of magnitude.<sup>82</sup> If a developer is able to break through the price limitation, there will be a significant influx of sensors across the country (currently, a nutrient sensor cost about \$15,000, but the Nutrient Sensor Challenge seeks sensors for less than \$5,000). A list of sensor technology companies is provided in Appendix B.

Citizen science, also known as crowd-sourced science, is non-traditional science conducted through the collaboration of public and professional scientists.<sup>83</sup> One reason citizen science has gained popularity in recent decades is because technology has enabled easier transfer of information. Citizen science data has inherent issues of precision, accuracy, and standardization of methodology and consequently, it is difficult to use the data for regulatory purposes.<sup>84</sup> However, citizen science data can provide insight into a situation that may be difficult to otherwise gain data from. For example, citizen science has been used to report flooding water and even helped calibrate satellite data.<sup>85</sup> Water quality and environmental monitoring is a hotspot of citizen science. For example, Colorado River Watch is a community driven organization where primarily students monitor water quality and other indicators of watershed health in order to inform decision makers about the condition of Colorado's waters.<sup>86</sup> The meteorological Phenomena Identification Near the Ground (mPING) is a NOAA application that crowd-sources more than 600,000 observations in order to verify weather models.<sup>87</sup>

Water quantity meter technology is becoming "smart" by having the capabilities of storing and uploading data to the cloud.<sup>88</sup> When strategically spread throughout a water utility, they work together as an advanced metering infrastructure (AMI) system.<sup>89</sup> The massive volume of measurements and frequent uploads of an AMI system enables "real time" monitoring of a water network. With a real time perspective of a utility, managers can better control water movements and refine system efficiency. AMI systems also help managers identify non-revenue water and optimize billing relations with customers.

### **More Data Integration**

Water management, and its data, is highly fragmented.<sup>90</sup> In the United States, there are over 50,000 public water utilities, hundreds of regulatory agencies, and hundreds of non-profit and private entities that store water related data.<sup>91</sup> Water data also exists in many formats, specific to the type of data (ex:

water quality or water quantity) and the organization. Consequently, water data exists in many forms and in many locations.

Integrating data is a resource intensive task. It requires significant time to convert data so that data variables and units match before combining datasets. This is a significant bottleneck and a critical step in aggregating data.<sup>92</sup> Original production of data in a standard schema greatly reduces the difficulty of later integrating datasets.

The ability to glean intelligence increases with increased data volume. By aggregating data from various sources in real time, water managers can make more effective and efficient decisions. For example, with a single warehouse of water quality data, managers could better implement emergency monitoring plans and track water quality changes in due to a spill or unplanned release. With a higher resolution view of the upstream watershed, facilities could refine treatment practices.

### **Conclusion of Recommendations**

The series of recommendations provide should be used to direct the water professionals toward effective implementation of big data capabilities for improved water resource management. By better defining user needs, projects can be designed to fulfill specific goals and have a higher probability of producing a sizable impact. By increasing sampling frequency and data integration, conclusions made from data will have more statistical power and be able to better benefit decision making.

However, all data will be underutilized without sufficient, analytical based tools to harness the opportunities presented by the data.<sup>93</sup> It is important to take the information available and translate it into actionable intelligence in decision making processes. An example to identifying a challenge, assembling data, and analyzing data to conclude actionable intelligence is present in the Water Risk Assessment Tool section.



# THE WATER QUALITY RISK ASSESSMENT TOOL

## Introduction

As previously described, in order to create value out of data, the information has to be properly organized, aggregated, analyzed, visualized to serve a specific purpose. Built for the Duke Nicholas Institute of Environmental Policy Solutions, the Water Quality Risk Assessment Tool is a proof of concept that converts raw data into actionable intelligence to determine where, when, and to what extent water quality is out of compliance or trending out of compliance. This tool is designed to be used as a stepping stone for another institution, such as CUAHSI or an OWDI agency, to scale the project to a larger service area with measureable value for its users. The Water Quality Risk Assessment Tool serves as a model for how the water sector can use available, yet fragmented, data to produce insight to better manage aspect of water resources. With time, water resource management could be improved by using harnessing the power of all available data and transforming it into actionable intelligence.

### *What Is Water Risk and Who Does It Affect?*

The confluence of population growth, climate change, and finite water resources are creating increasingly profound water risk. The World Economic Forum 2016 Global Risk Report highlighted this issue by rating water crisis as the ninth most likely and third most impactful risk facing countries and industries in the next decade.<sup>94</sup> Physical, reputational, regulatory water risks have direct implications for corporations, water utilities, and investors.<sup>95</sup> Physical risks are those that directly affect operations, such as water supply scarcity or contaminated water. Reputational risks are those that affect an organization's image, such as impacts to ecological resources or socio-economic dynamics. Regulatory risks are changes in government oversight that affect the costs of permitting, treatment, or reporting standards.

The most common at risk entities are municipal water systems. Risks include supply security, demand management, asset management, water quality, energy use and generation, and rates risks.<sup>96</sup> These risks are mainly due to their massive infrastructure and a common need for substantial renovation. It is estimated that over 80 percent of the American population relies on public water supplies.<sup>97</sup> Of that, 90 percent of the supply is delivered by more than 53,000 state and municipal water utilities.<sup>98</sup> However, these water systems are aging and deteriorating. The American Society of Civil Engineers recently gave American's water and waste water a D rating.<sup>99</sup> The EPA estimated that drinking water utilities face a total financing gap of \$334.8 billion over the next 20 years.<sup>100</sup> To fund the renovations, municipal water systems seek private capital.

Municipal bond packages are used to fund large projects, such as maintaining drinking water and waste water infrastructure. Municipal bonds related to water infrastructure are either special purpose bonds or general obligation bonds, like California's \$7.4 billion Water Quality, Supply, and Infrastructure Improvement Act of 2014.<sup>101</sup> In 2009, public utility bonds totaled \$300 billion in daily trading value—ten percent of the total municipal bond market.<sup>102</sup> When investors buy municipal bonds, they take on a risk that a city will not pay back their owed money (coupon) to the bond holder, resulting in a default. The risk associated with an investment should be reflected in the market price and potential yields of the investment.

The problem is that although we face increased water risk, municipal credit rating agencies, bond investors, and even water utilities have been slow to incorporate risk analysis into their financial plans. "One of the key limitations for assessing water risk is the lack of broadly applicable yet meaningful tools, such as quantitative frameworks regarding water scarcity, planning tools to address issues of drought, and qualitative assessments to determine whether companies examine water risk and reflect that risk in their business strategy or use it to strategize growth management. Investors understand that the biggest barriers to assessing water risk in their portfolios are lack of data and financial impact of water restrictions on an industry."<sup>103</sup>

### ***Comparable Water Risk Assessment Tools***

Interest in water risk assessment has increased in the past decade. To assess water related risk a small assortment of tools were developed, all within the past five years. These tools indicate the market's desire for a water risk assessment service. Each tool provides a valuable service; however each has some type of short coming which we hope to address.

The World Resource Institute (WRI) developed the Aqueduct Water Risk Atlas (Aqueduct), a publicly available map-based web application that indicates water-related risk.<sup>104</sup> The tool's framework is built on twelve water risk indicators; baseline water stress, inter-annual variability, seasonal variability, flood occurrence, drought severity, upstream storage, groundwater stress, return flow ratio, upstream protected land, media coverage, access to water, and threatened amphibians. Aqueduct impressively models water risk globally, however, the data resolution is course. The water risk indicators resolution's range from 30 meters to entire countries.<sup>105</sup> Consequently, it is difficult for decision makers to productively apply Aqueduct water risk information to small geographic scales, such as cities. Furthermore, the tool only provides metrics, not raw data, so it is difficult to gain insight on a specific aspect of interest.

Bloomberg LP and the Natural Capital Declaration released a Water Risk Valuation Tool (WRVT) in 2015.<sup>106</sup> The WRVT enables analysts to incorporate water risks into valuations of gold and copper mining companies. The tool builds on Bloomberg's Carbon Risk Valuation Tool by integrating the WRI's Aqueduct water stress indicators to determine changes in revenue and costs. The WRVT illustrates how the investment community is recognizing critical natural capital factors, however, similarly to WRI Aqueduct, the tool has coarse data resolution.

Ceres, a non-profit organization which advocates sustainable business practices, have published many materials on water risk. In 2010, Ceres published a model to investors to ascertain water risk scores for electric and water utilities.<sup>107</sup> Ceres also developed Aqua Gauge, an Excel-based tool and associated methodology that allows investors to interpret and evaluate a company's water management activities.<sup>108</sup> Both of the Ceres water risk assessment tools do not present any data; rather they present methodologies for investors to interpret data.

Ponderosa Advisors LLC.'s Water Sage is a map-based web portal that allows users to easily search and view water rights.<sup>109</sup> The tool has gained popularity among the Montana Bankers Association to help value agricultural properties. Without properly assessing water rights, banks risk acquiring foreclosed farm or ranch lands without the necessary water rights that underpin the lands' value.<sup>110</sup> Lenders use Water Sage to determine if borrowers have sufficient water rights to supply their agricultural processes.

### ***Objectives of the Water Risk Assessment Tool***

The Nicholas Institute of Environmental Policy Solutions' Water Policy Program at Duke University (Nicholas Institute) is an interdisciplinary effort focused on long-term viability of water quantity and quality, with regards to water infrastructure, water efficiency, and protection and restoration.<sup>111</sup> The Nicholas Institute would like to bring transparency to the underlying water risks present in the United States.

Their goal is to develop a tool which synthesizes data related to water risk to enable investors, and others, to quickly assess water risk at fine resolutions. At its current stage, the Nicholas Institute is developing a prototype (or proof of concept) application to acquire external funding for further development. The tool is designed to be a map-based web application with two main objectives; (1) aggregate raw data sources related to water risk and (2) analyze and visualize that data to communicate risk to the user by flagging areas with a history on non-compliance.

The tool will include information about the following water risk indicators:

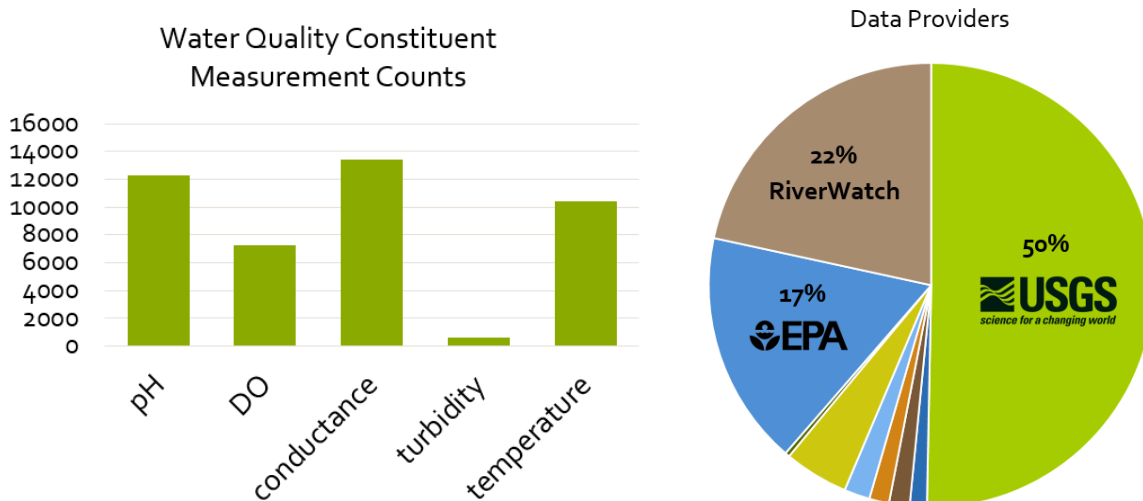
Supply	Quality
<ul style="list-style-type: none"><li>• freshwater use</li><li>• ground water</li><li>• rights, access, litigation</li><li>• storage capacity</li><li>• surface water</li></ul>	<ul style="list-style-type: none"><li>• dissolved oxygen</li><li>• pH</li><li>• temperature</li><li>• turbidity</li><li>• electric conductance</li></ul>
Climate	Stressors
<ul style="list-style-type: none"><li>• extreme event frequency</li><li>• drought severity index</li><li>• precipitation</li><li>• sea level rise</li><li>• snowpack</li><li>• temperature</li></ul>	<ul style="list-style-type: none"><li>• endangered aquatic species</li><li>• fracking</li><li>• population change</li><li>• water quality permits</li></ul>

## Methods

On behalf of the Nicholas Institute, I developed the water quality portion of the tool. This required first researching water quality management methodologies, identifying data sources, and determining which constituents to include in the water risk assessment tool. The developed Water Quality Risk Assessment Tool automatically aggregates raw data from public web sources, process the data, and presents the information in an intuitive design that displays raw data and summary statistics that indicate where, when, and to what extent sampling locations are out of compliance with water quality regulations.

### *Data*

The Water Quality Portal was chosen as this project's data source because it is the largest water quality data source in the United States. Hosted collaboratively by the EPA, USGS, and National Water Quality Monitoring Council, the Water Quality Portal integrates publicly available water quality data from the USGS National Water Information System (NWIS), the EPA STOrage and RETrieval (STORET) data warehouse, and the USDA Sustaining The Earth's Watersheds – Agricultural Research Database System (STEWARDS).<sup>112</sup> Within the contributing datasets are water quality monitoring data collected by over 400 water resource management groups across the country including states, tribes, watershed groups, federal agencies, volunteer groups, and universities. As of July 2015, the Water Quality Portal accesses over 265 million results from over 2.2 million monitoring locations. In this proof of concept, data was collect and present only for the state of Colorado. This included approximately 44,000 water quality samples, described in Figure 2 and Table 2.



**Figure 2.** The number of water quality measurements presented in The Water Quality Risk Assessment Tool for Colorado (left). The sampling organization which provided the water quality data (right).

**Table 2.** The number of water quality measurements presented in The Water Quality Risk Assessment Tool for Colorado, organized by data provider and constituent type.

		Sampling Count					Total	Percentage of Total
		pH	Dissolved Oxygen	Specific Conductance	Turbulence	Temperature		
Data Provider	USGS Colorado Water Science Center	3588		11402	452	6716	22158	50%
	EPA National Aquatic Resources Survey	216	165			165	546	1%
	Colorado Dept. of Public Health & Environment	1914	1825	1849		1916	7504	17%
	Southern Ute Tribe	132	132	132	132	132	660	1%
	Adams Rib Ranch	312				321	633	1%
	The Rivers of Colorado Water Watch Network (RiverWatch)	4753	4757				9510	22%
	Clear Creek Watershed Foundation (CCWF)	412	412				824	2%
	CBS Operations Inc.	988				1069	2057	5%
	Gunnison Basin & Grand Valley Selenium Task Force			56		92	148	0%
	<b>Total</b>	<b>12315</b>	<b>7291</b>	<b>13439</b>	<b>584</b>	<b>10411</b>	<b>44040</b>	<b>100%</b>
<b>Percentage of Total</b>	<b>28%</b>	<b>17%</b>	<b>31%</b>	<b>1%</b>	<b>24%</b>	<b>100%</b>		

Water quality is comprised of dozens of characteristics—the EPA’s National Primary Drinking Water Regulations require public water systems to monitor 77 contaminants.<sup>113</sup> To simplistically summarize water quality, I used dissolved oxygen, pH, specific conductance, temperature, and turbidity. These are general indicators of water quality as they correlate with many other facets of water quality. These five

constituents are also the most prolific types of water quality measurements because they are frequently measured in-situ in real time with sensors, rather than by discrete samples that are later processed in laboratories.

Dissolved oxygen refers to the concentration of gaseous oxygen incorporated into water. Dissolved oxygen increases with turbulence and decreases with temperature. Dissolved oxygen is essential aerobic aquatic life growth and reproduction.

pH is a negative logarithmic measurement of hydrogen ion concentration in water. Even minor changes in pH alter the states of many chemicals and thereby change their solubility, transport, and bioavailability. pH affects community composition and biological processes, such as reproduction, growth, disease, and death, because it is a physiological stressor. EPA suggests a range of 6.5 to 9 for optimal freshwater conditions.

Specific conductance is a measure of ionic strength, the concentration of ionic charge in water, as electrical conductance, normalized with reference to temperature. Ionic strength relates to a large array of potential freshwater problems because aquatic organisms generally prefer waters with specific ions and a specific ionic strength range. Changes specific conductance measurements can affect community composition and require changes water treatment processes.

Temperature is the concentration of thermal energy in water. Atmospheric and hydraulic processes and the structure and function of aquatic systems influence the thermal regime of a water system. Temperature is frequently altered by anthropogenic processes and is monitored in effluent discharges.

Turbidity is the measurement of relative clarity of water. It is a proxy measurement for the amount of sediment in water, such as dissolved organic and inorganic, algae, plankton, clay, and silt. Highly turbid waters can represent a health concern for aquatic communities and drinking water. Turbidity limits are common regulations for effluent discharges.

### ***Application Processes***

The developed water quality risk assessment tool consist of two fundamental parts; (1) back end scripts that download, clean, and process desired water quality data from the Water Quality Portal and (2) front end scripts which present preprocessed data in an interactive user interface. Application code is presented in Appendix C.

The Water Quality Portal provides web services where information can be queried using a RESTlike (REpresentational State Transfer) technique.<sup>114</sup> Water Quality Portal hosts two base Uniform Resource Locators (URLs); a URL for station information and a URL for result information. Python code with packages urllib, os, and csv, were used to scrape desired content from the Water Quality Portal and store information locally. A comma separated variable (csv) file with desired counties' Federal Information Processing Standard (FIPS) codes and names was first created to denote which counties' water quality information should be downloaded. In the current version of the tool, all Colorado counties, with the exception of Larimer and Phillips counties, were included. An iterative loop was used to download data for each of the five desired constituents; dissolved oxygen, pH, specific conductance, temperature and turbidity. A secondary iterative loop was used to download station data for each of the counties referenced in the csv file. The FIPS codes were used to define the county code in the station URL. Data was downloaded as a non-zipped csv for each constituent. A URL open function and web read function were used to collect data from the Water Quality Portal. This data was then written to a csv labeled by county name (from county csv file) and constituent name. This process was repeated in another secondary for loop for results data per county and per constituent. In this version, the URL specifies collection of results with a sample date from year 2005 and later. By running the web scraping python code, a series of csv files were written in a specified location. The csv files were specific to result or station, constituent, and county.

Collected data was cleaned and processed using the R statistical analysis software package.<sup>115</sup> First, new data frames were created to hold station and results information for each of the constituents. This was performed by initially creating empty data frames for each constituent with desired variable types (columns) to later house the data. An iterative loop was then used to read in station and results data into R for each constituent and county. The results data was then filtered by omitting locations with less than ten samples. This was performed to reduce noise in the dataset so that only locations where data trends can be determined are used. Next, station and results data were merged by their monitoring location identifier. As this process was repeated for each county, data was bound by row together into one data frame. These data frames were then renamed to reflect each constituent name. Once completed, the process produced a new data frame for each constituent containing results and station information for all sample sites within desired counties with ten or more observations.

A summary statistic was next calculated for each sample location and added to each constituent data frame. This was performed by creating a subset data frame for each sample location, per constituent,

calculating the sample statistic, and merging the subset data frame sample statistic with the original constituent data frame. For dissolved oxygen, the summary statistic was a compliance ratio which represents the frequency of samples at a given location less than 3 mg/L (Colorado Department of Public Health and Environmental Water Quality Control Division regulations).<sup>116</sup> For pH, the summary statistic was a compliance ratio which represents the frequency of samples at a given location less than 6.5 or greater than 9 (the EPA's suggestion for optimal freshwater conditions).<sup>117</sup> For temperature, the summary statistic is an average of temperature measurements for samples at a given location. For turbidity, the summary static is a ratio which represents the frequency of samples at a given location more than 50 NTUs (the threshold used by states for regulation classification and discharge limitations for both natural and anthropogenic systems).<sup>118</sup> For specific conductance, the summary statistic is a ratio of specific conductance measurements for samples at a given location less than 400 or greater than 1350  $\mu\text{S}/\text{cm}$  (the thresholds that USGS considers the South Platte River to be significantly affected by anthropogenic activities).<sup>119</sup> The data frames for each constituent, with summary statistics, were used in the map-based visualization of the data.

The data was next processed for each constituent by producing new data frames that could be used to graph scatter plots of the measured results. For each constituent, a new data frame was created for every unique monitoring location identifier. From this, an extensible time series (xts) data frame was created using the xts R package so that row headers are dates and the data are result measure values.<sup>120</sup> Results measured were separated and filter by whether they comply or do not comply with standards previously described (ex: pH is in compliance if between 6.5 and 9).

To identify data trends, average measurement per month per year were calculated for each unique monitoring location identifier for each constituent. An average deviation per month measurement was used to account for sub-seasonal variation within the data. Average measurements per month were also calculated. Deviance from the average per month was calculated by subtracting average measurements per month per year from average monthly results. Deviance values were stored in data frames per each constituent and monitoring location identifier. This data was accessed when plotting bar graphs of data trends.

The application user interface was built using the shiny R package,<sup>121</sup> which consists of user interface and server functions. A fluidpage user interface format with a "space lab" shiny theme<sup>122</sup> was used so that the application would automatically adjust for various window frame sizes. The user interface delineates a series of rows and columns to hold the data map, results graph, trends graph, and descriptive



information. Event reactive radio buttons present the user options to select a water quality constituent type (dissolved oxygen, pH, specific conductance, temperature, and turbidity). A reactive event “view data” button allows the user to graph result and trends plots for a selected sample location on the map.

The mapping feature was created using the leaflet R package.<sup>123</sup> The map plots data determined by the constituent type radio buttons. Point color symbology was a function of the data summary statistic. Point size was a function the amount of observations at a sample location. When a sample location was selected, a popup including monitoring location name, monitoring location identifier, data provider, and summary statistic was displayed.

The data results scatter plot was created using the dygraph R package.<sup>124</sup> This graphing function was initialized by the “view data” reactive event button and uses the map’s selected sample location to link to the desired xts data frame based on the monitoring location identifier. The dygraph plots measured results and colors the data based on whether the data is within recommended compliance bounds. A range selector was included to allow the user to inspect a specific date range.

The trends results graph was created using the ggplot2 R package.<sup>125</sup> This graphing function was also initialized by the “view data” reactive event button and uses the map’s selected sample location to link to the desired data frame based on the monitoring location identifier. The plot’s bars represent the average of a month’s result for a given year deviance from the average of a month’s result for all years.

## **Results**

The resulting application enables a user to intuitively interact with data gathered from the water quality portal. In this proof of concept, a user can search and view data from the Division 1 Colorado Water District for samples collected after 2004. The Water Quality Risk Assessment Tool can be accessed at [https://mark-ziman.shinyapps.io/WQRAT\\_MZ/](https://mark-ziman.shinyapps.io/WQRAT_MZ/).

https://mark-ziman.shinyapps.io/WQRAT\_MZ/

## Water Quality Risk Assessment Tool

*proof of concept for Colorado, USA*

This application visualizes water quality measurements and trends for dissolved oxygen, pH, specific conductance, temperature, and turbidity data from the Water Quality Portal, including EPA's STORET, USGS' NWIS, and USDA's STEWARDS databases. This proof of concept version includes water quality data for Colorado, USA.


Created by Mark Ziman, master of environmental management candidate at Duke University's Nicholas School of the Environment.

**Select constituent to map**

- dissolved oxygen
- pH
- specific conductance
- temperature
- turbidity

Click below for selected sample location results and trends.

[View Data](#)



**Monitoring Location Name:** CHERRY CREEK AT DENVER, CO.  
**Monitoring Location Identifier:** USGS-06713500  
**Data Provider:** USGS Colorado Water Science Center  
**Out of Compliance Ratio:** 0

Legend:  
 0.0 - 0.2  
 0.2 - 0.4  
 0.4 - 0.6  
 0.6 - 0.8  
 0.8 - 1.0

Leaflet | © OpenStreetMap contributors, CC-BY-SA


pH is a negative logarithmic measurement of hydrogen ion concentration in water. Even minor changes in pH alter the states of many chemicals and thereby change their solubility, transport, and bioavailability. pH affects community composition and biological processes, such as reproduction, growth, disease, and death, because it is a physiological stressor.

**Summary Map:**  
US EPA water quality criteria for pH in freshwater suggest a range of 6.5 to 9.

This map visualizes the frequency of pH measurement per sample location outside EPA recommendations.

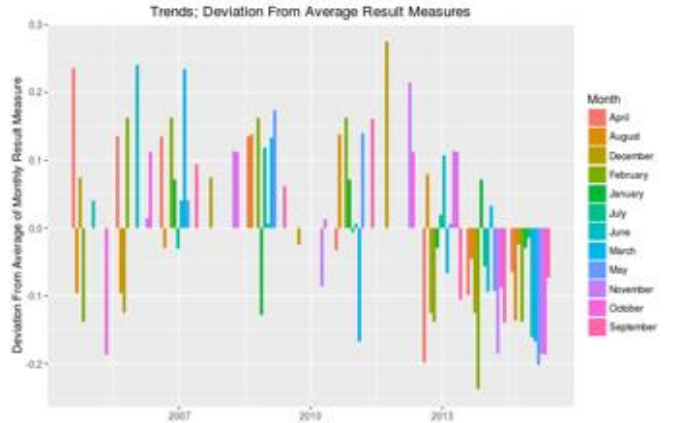
### pH measurements at USGS-06713500

Apr, 2007: In\_Comp: 8.6

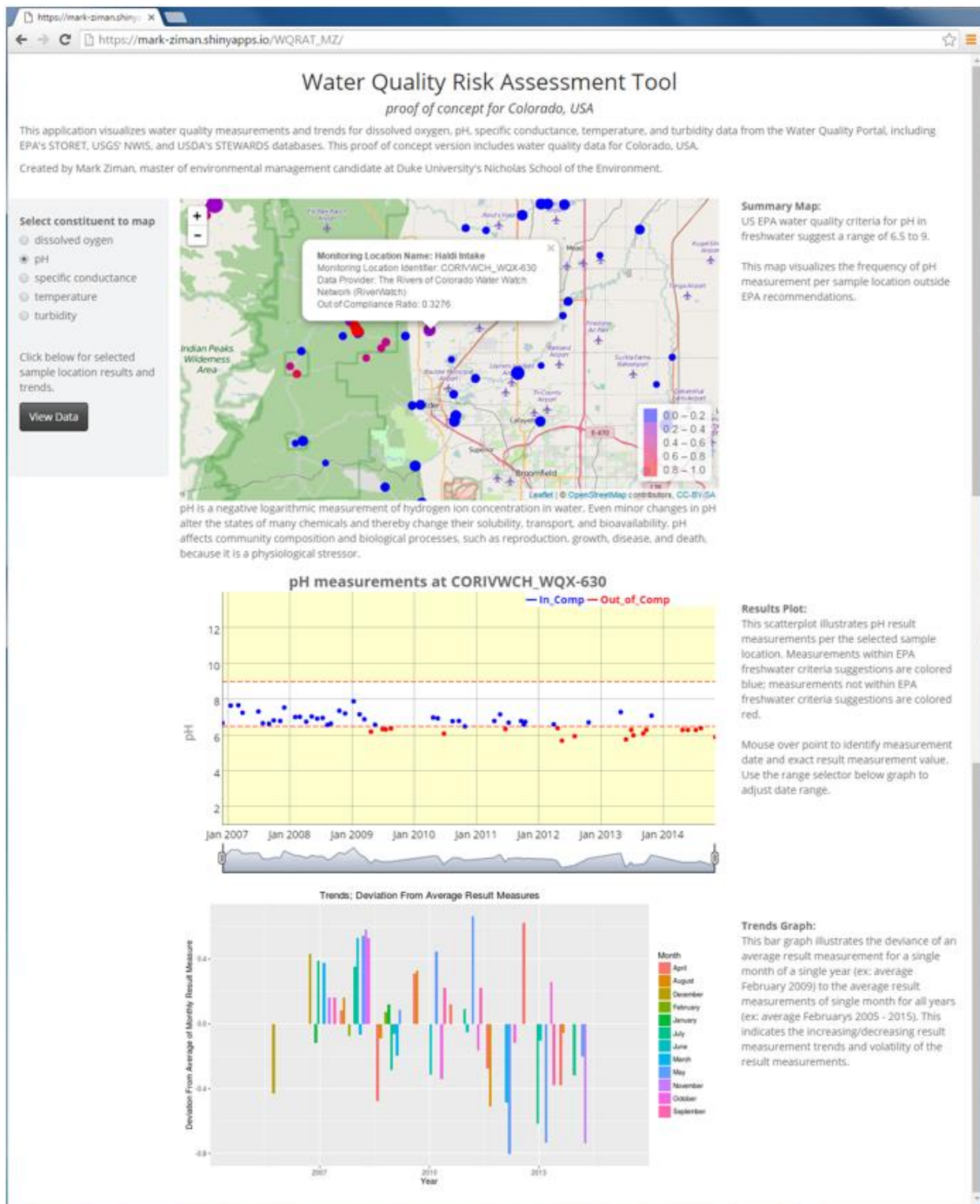


Mouse over point to identify measurement date and exact result measurement value. Use the range selector below graph to adjust date range.

### Trends: Deviation From Average Result Measures

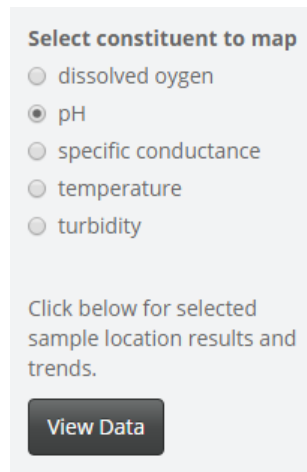


**Trends Graph:**  
This bar graph illustrates the deviance of an average result measurement for a single month of a single year (ex: average February 2009) to the average result measurements of single month for all years (ex: average Februarys 2005 - 2015). This indicates the increasing/decreasing result measurement trends and volatility of the result measurements.



**Figure 3.** Two examples of the Water Quality Risk Assessment Tool. Presented here is the entire interface, including radio button selections, map-based results, graphed results, result trends, and descriptive texts.

By selecting a constituent type from the radio buttons, data is plotted on the central map. By looking at the color of the sample location, a user can see the location's summary statistic, an indication of water quality risk. For example, for pH, sample points are colored by how frequently samples are not within EPA suggested values for optimal freshwater conditions. This, in a sense, flags risk areas with a history of water quality non-compliance. By looking at the size of a sample location, a user can see the number of samples taken at that location. By clicking on a sample location, a pop up informs the user of the sample location's name, identifier, data source, and summary statistic. These functionalities allow the user to quickly and easily search for water quality data from a geographic viewpoint.



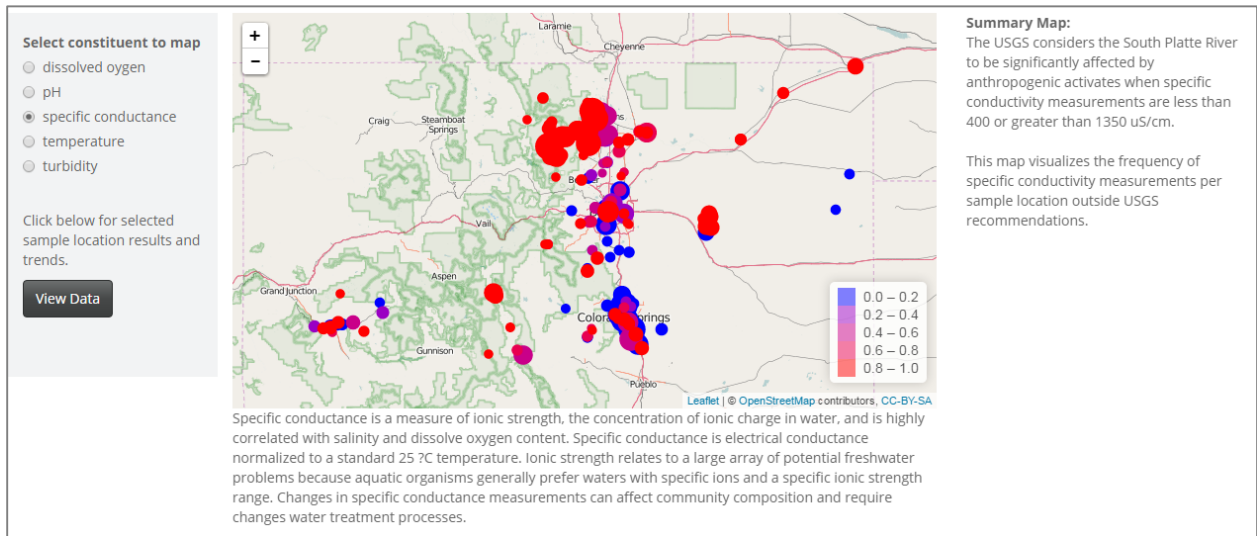
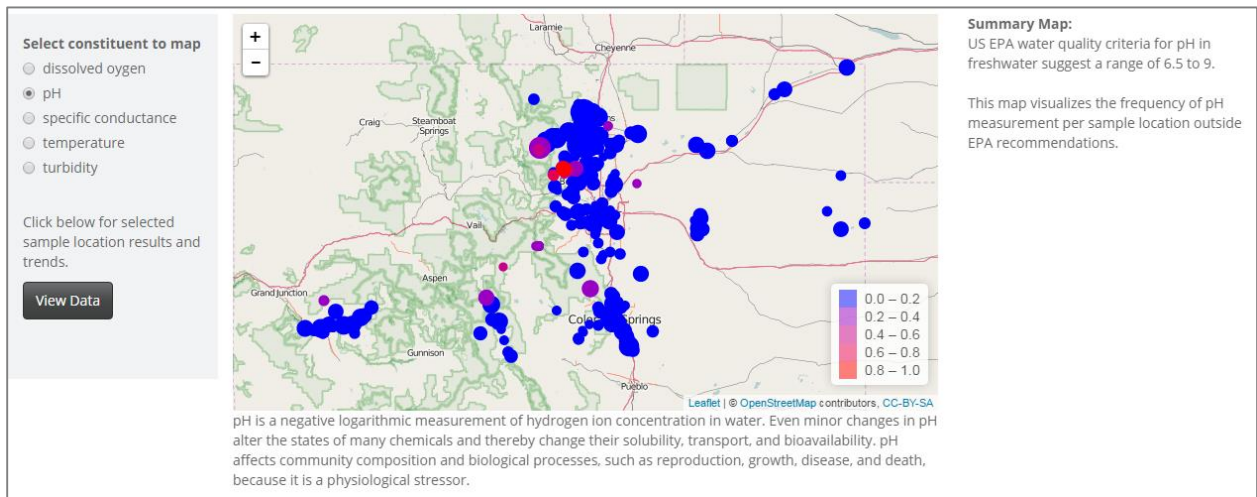
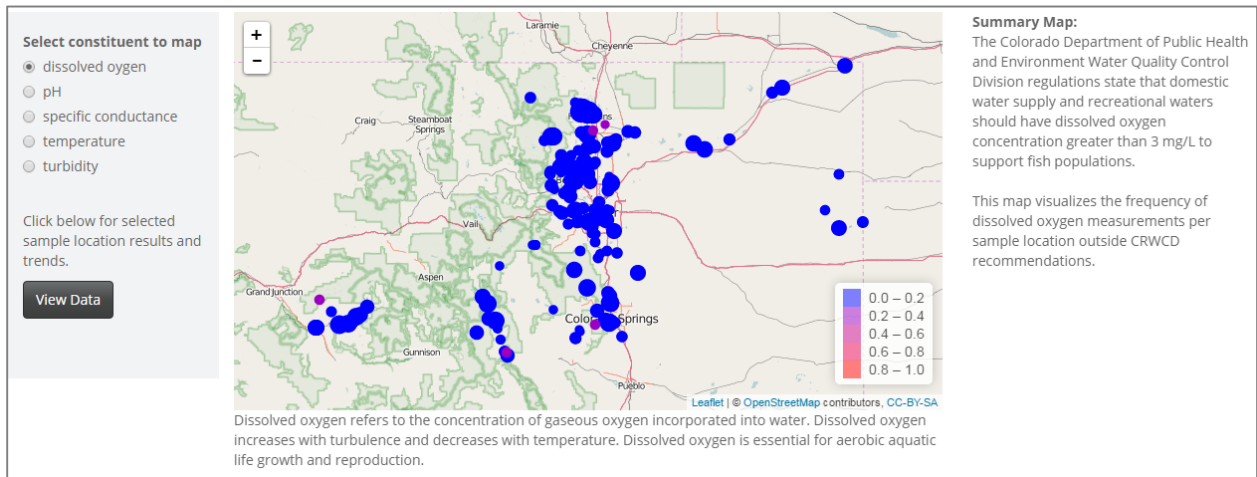
Select constituent to map

- dissolved oxygen
- pH
- specific conductance
- temperature
- turbidity

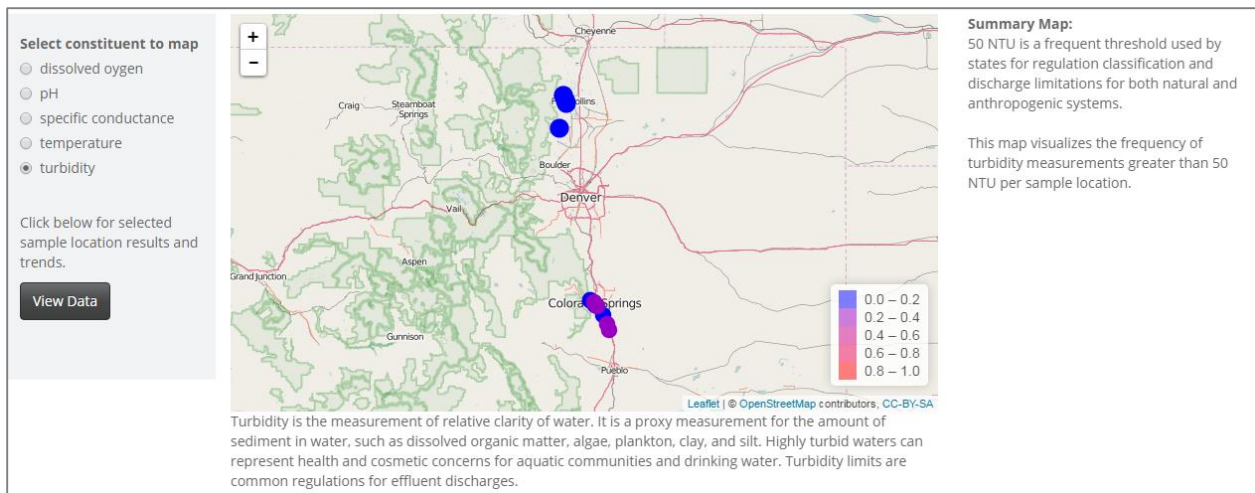
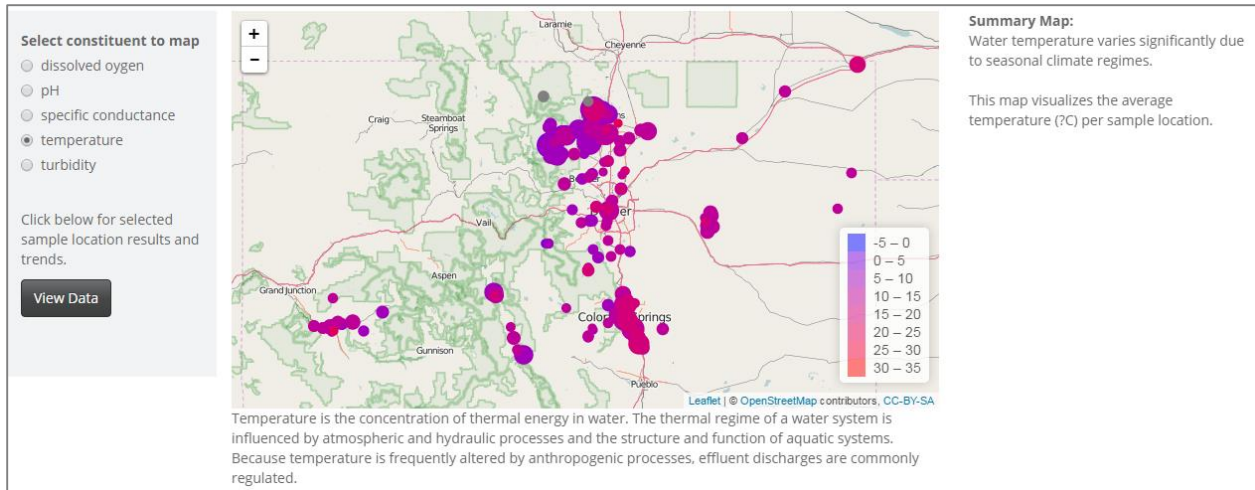
Click below for selected sample location results and trends.

[View Data](#)

**Figure 4.** Radio buttons of water quality constituent types. Selecting a constituent type plots associated data on the map. Also included here is the “View Data” button that produces results and trends graphs for any selected location.



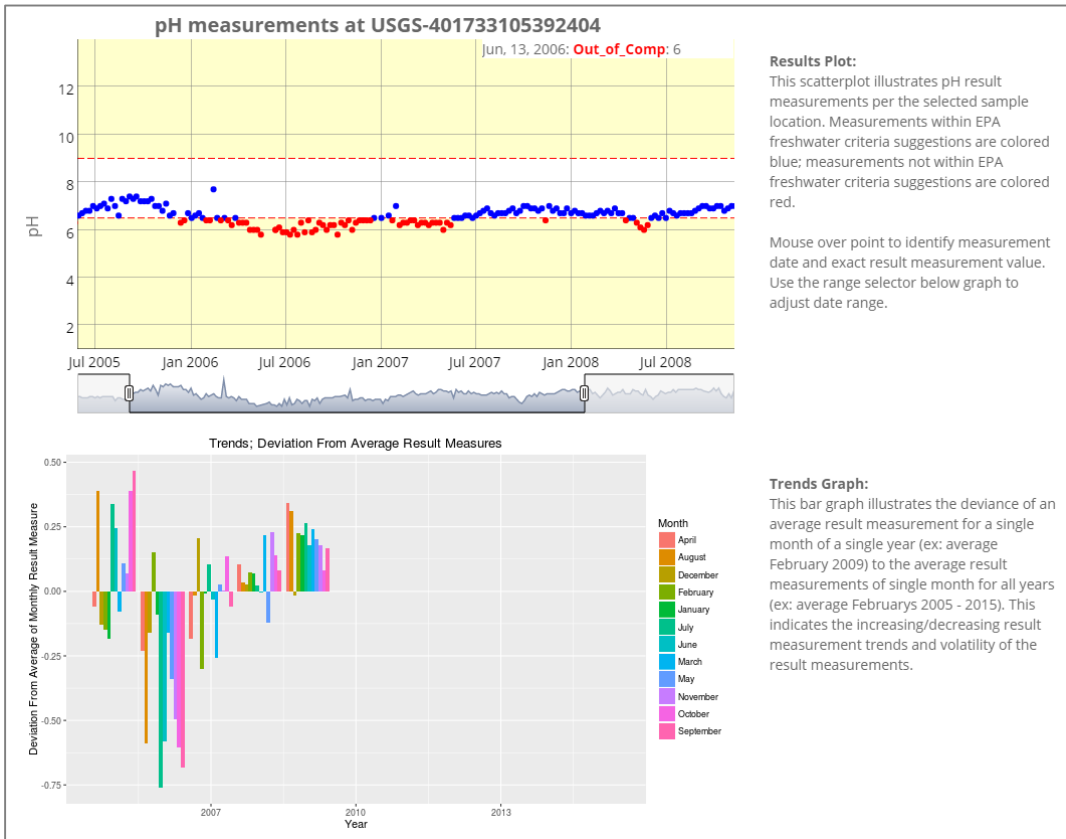
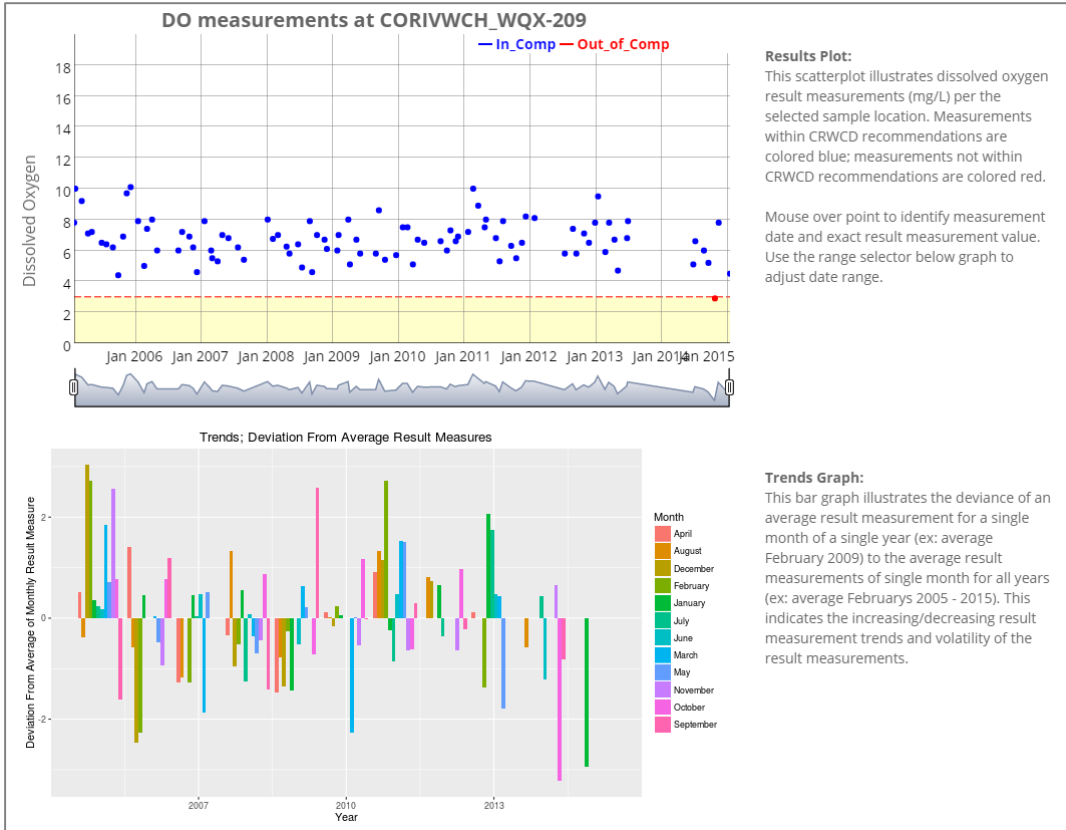


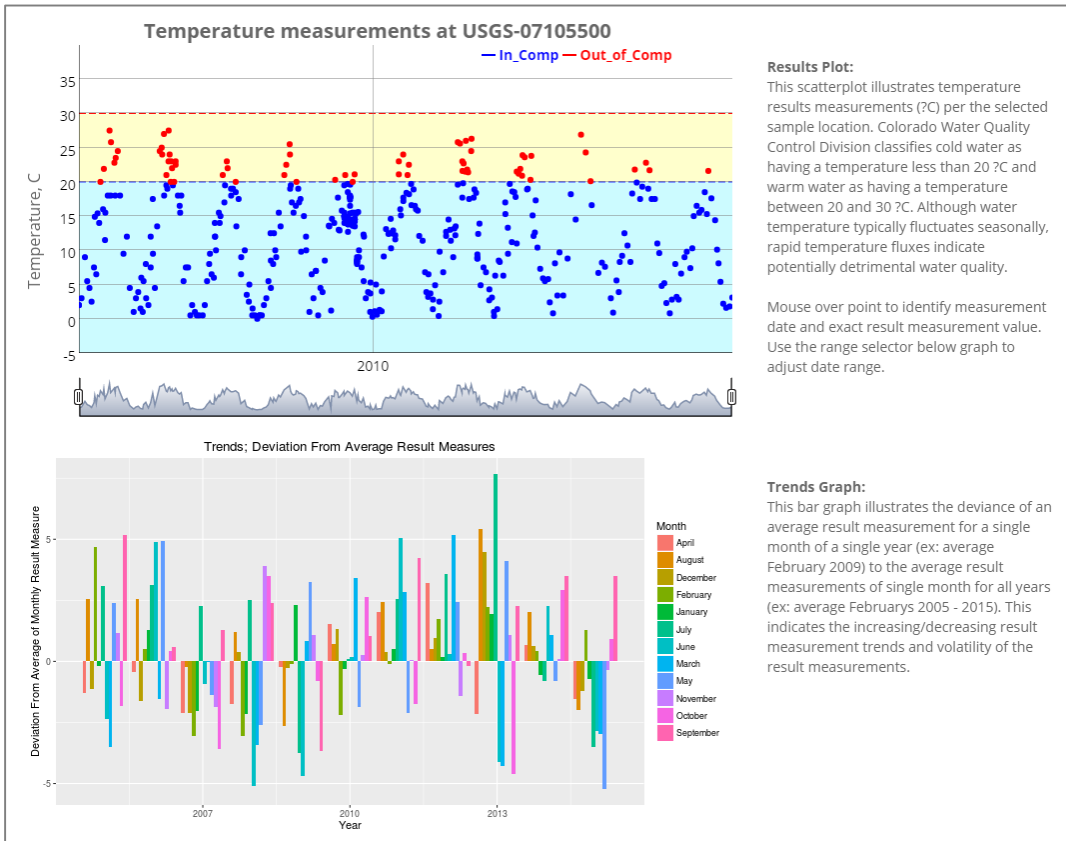
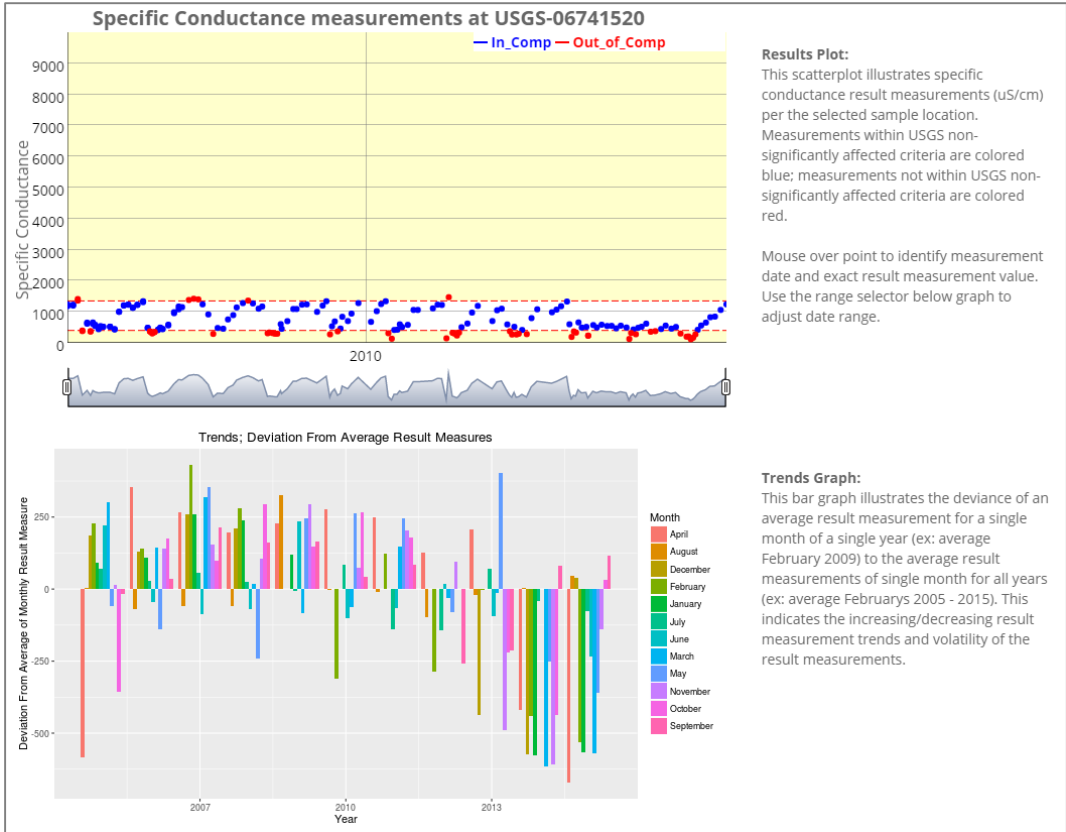


**Figure 5.** Maps of the five constituents (dissolved oxygen, pH, temperature, turbidity, specific conductance) measured results at sample locations within Colorado. Constituent type is selected, and shown here, to the left of the map. Symbology is provided to the right of the map. Constituent summaries are provided below the map.

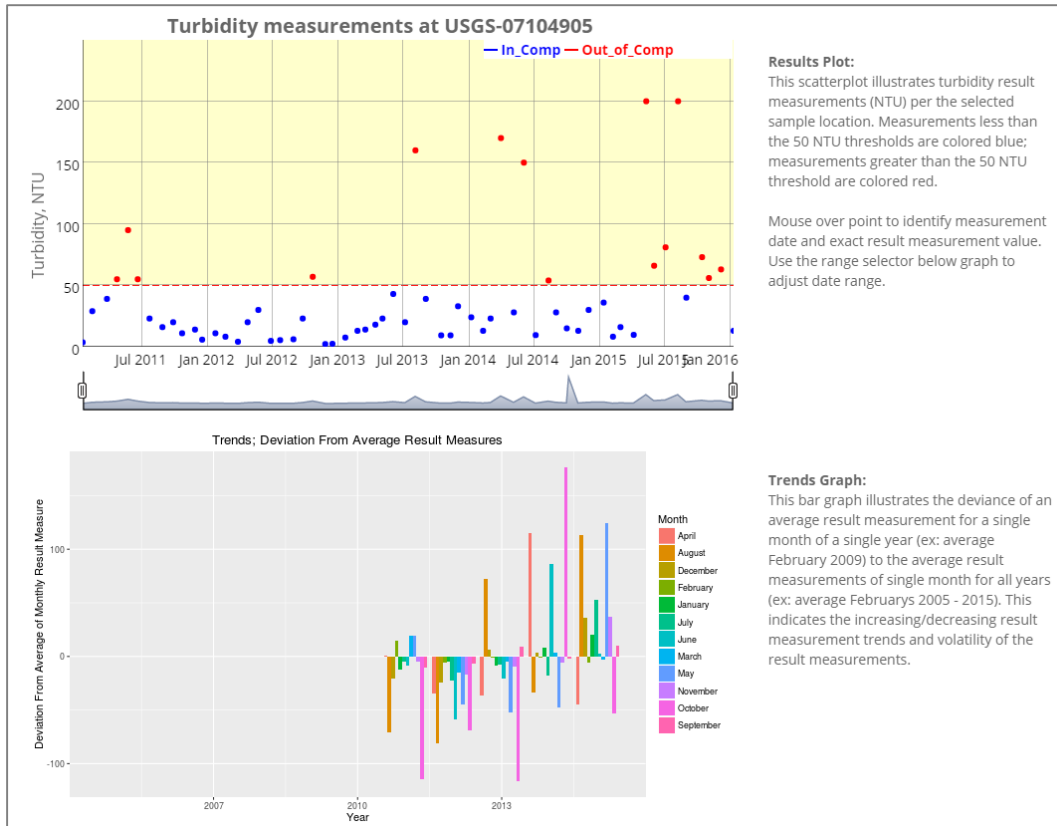
The values graph illustrates data results across time. This shows the user when and by how much measured results have fluctuated. Multiple color types and shaded regions are used to help call attention to results which indicator water risk. User can zoom in by using the date range selector at bottom of graph. By mousing over point, result measured value is displayed.

The trends graph also illustrates data results across time, within an emphasis on directionality and extent of variance. For example, the trends graph enables the user to quickly see if measured results are increasing or decreasing over time. The user can also assess how volatile results are by the length which variance extends. The trends also remove monthly variation from the dataset.









**Figure 6.** Example results and trends graphs for each of the five constituents (dissolved oxygen, pH, temperature, turbidity, specific conductance). Descriptions of plots are provided to right of the plots.

## Discussion

The water quality risk assessment tool enables a user to easily search and review water quality data related to risk assessment. The application is built upon an intuitive interface that uses a map-based projection of water quality data. With just a few clicks, a user is able to query for a constituent, and assess water quality with respect to geographic location and sample size. The tool also summarizes water quality measurements by presenting raw data measurements and deviations from average measurements. A user can intuitively identify regions of water quality non-compliance and gather raw data about those areas.

This tool presents an original and unique method of quickly searching and analyzing water quality data. My research identified no other applications which similarly aggregate and present publically available data. More possibilities are given to the user with the ability to interact with the information, rather than just having a static summary. In its current stage, the tool begins to address the challenges of big data by providing a better way to glean knowledge from a raw dataset.

## **CONCLUSION**

The growth of data and development of technologies provides value in almost every domain of society. In water resource management, big data capabilities can be used to prevent man-made crisis such as contamination in drinking water supplies to environmental ailments such as eutrophication in reservoirs. It can help the nation use water more efficiently and decrease overall consumption. Furthermore, data can be used to integrate managerial processes and identify inefficiencies in systems. When used correctly, data can be used to create actionable intelligence to enhance decision making. Faced with issues of growing populations and increasing risk of drought, water agencies and industries recognize the importance of optimally using data.

Understanding the data framework of water resource management is a prerequisite for applying the capabilities of big data. The materials provided in this report summarize the strengths, weaknesses, and opportunities associated with water data. By better defining user needs, integrating fragmented data sources, and using analytical tools, we can improve water resource management.

The Water Quality Risk Assessment Tool is an example of how data can be used to benefit decision making. The tool is a model for how agencies and private developers can create value by identifying specific data sources, processing data with analytics, and visualizing data to address a specific user need. Furthermore, the tool shows how data value is increased with the user can interact with the information in an intuitive, functional interface. Although just a proof of concept in its current form, the tool illustrates how fragmented data sources can be used to decipher important information. There exist many more opportunities in water resource management to use big data to enable decision makers.

Big data capabilities will be most quickly and effectively applied to the water resource sector if manager and investors recognize the opportunities of big data and dedicate resources to its development.

## APPENDICIES

### A. North Carolina Community Surface Water (and GWUDI) System Monitoring Requirements

A summary of community surface water (and GWUDI) system monitoring requirements for North Carolina.<sup>126</sup> Routine Measurement Frequencies provided are for when a community system is in compliance. Measurement frequencies have stipulated increases (not provided here) when a routine measurement is non-compliant.

Constituent	Routine Measurement Frequency
Alkalinity	monthly
Alkalinity	monthly
Asbestos	9 years
Bacteriological	monthly
Bromate	monthly
Bromide	monthly
Chloramine	monthly - quarterly
Chlorine	monthly - quarterly
Chlorite	daily - monthly
Chlorine Dioxide	daily
Disinfectants	-
Haloacetic Acids	quarterly
Inorganic Chemicals	annually
Lead and Copper	annually
Magnesium	monthly
Nitrate	quarterly
Nitrite	-
Pesticides / Synthetic Organic Chemicals	quarterly
Radionuclides	quarterly
Specific Ultraviolet Absorbance	monthly
Total Organic Carbon	monthly
Trihalomethane	quarterly
Volatile Organic Chemicals	quarterly

## **B. Nutrient Sensor Challenge Participants**

A list of sensor technology organizations that were participants in the Nutrient Sensor Challenge. These entities are all in the business of producing cheaper, more functional real-time nutrient sensors.

### **Finalists**

- 1 Decagon Devices, Inc.
- 2 National Oceanography Centre
- 3 Real Tech
- 4 Sea-Bird Coastal
- 5 SYSTEA S.p.A.
- 6 T.E. Laboratories & Dublin City University

### **Contestants**

- 7 Aquisure
- 8 ASA Analytics
- 9 Ayyeka
- 10 Blue Legacy International
- 11 CleanGrow, Ltd.
- 12 Decagon Devices, Inc.
- 13 Environmental Monitoring Solutions, Ltd.
- 14 Franklin Thompson
- 15 Geekchitecture
- 16 JAL Engineering
- 17 Katsujinken Foundation
- 18 Lumense, Inc.
- 19 National Oceanography Centre
- 20 Open Photonics Inc.
- 21 RATES
- 22 Real Tech
- 23 Sea-Bird Coastal
- 24 SRI International, Marine & Space Sensing Laboratory
- 25 SubChem Sensor Systems, Inc.
- 26 SUNY Binghamton
- 27 SYSTEA S.p.A.
- 28 T.E. Laboratories & Dublin City University
- 29 Translume, Inc.
- 30 Turner Designs
- 31 UCSD Biodynamics Lab
- 32 University of Illinois / MoboSense, LLC.
- 33 USDA Ag Research Service
- 34 Water Canary
- 35 YSI, Inc.

## C. Water Risk Assessment Tool Code

### *Python Code to Collect Data with Web Services*

```
# This is a script used to scrape Water Quality Portal
# it loops through all CO counties AND constituents and writes data to it

# does not include Phillips County, Larimer County

# Mark Ziman
# 4/3/2016

import urllib
import os
import csv

# Create list with variable names (for constructing url)
var = ['pH', 'Dissolved%20oxygen%20(DO)', 'Specific%20conductance', 'Turbidity',
'Temperature,%20water']

# define filePath
filePath = "Z:\MP\WQRAT\Data"

# Loop through Big 5 constituents and extract station/result data for each one
for j in var:
    characteristicName = j

    # Open list with all CO Counties names
    # coded for all CO counties (_CoCounties.csv)
    d = open('Z:\MP\WQRAT\Data\_CoCounties.csv')
    csv_d = csv.reader(d)

    # STATION data
    for i in csv_d:
        FIPS = i[0]
        CountyName = i[1]

        # Construct the URL
        countyFIPS = FIPS ## the county FIPS
        countycode = 'US%3A08%3A' + countyFIPS ## USA, Colorado state

        # for STATION data
        my_url = 'http://www.waterqualitydata.us/Station/search?' + \
            'countycode=' + countycode + '&' + \
            'characteristicName=' + characteristicName + '&' + \
            'mimeType=csv' + '&' + \
            'zip=no'

        # Open URL, GET data
        my_web = urllib.urlopen(my_url)
        htmlSource = my_web.read()
        my_web.close()
        print "station data scraped for " + characteristicName + " for " + CountyName

        # Write data to csv
        outfile = open(filePath + '\\\' + characteristicName + '_' + CountyName +
'_Station' + '.csv', 'w')
        outfile.write(htmlSource)
        outfile.close()
        print "station data written for " + characteristicName + " for " + CountyName

    d.close()
```

```

# Open list with all CO Counties names
# coded for all CO counties (_CoCounties.csv)
d = open('Z:\MP\WQRAT\Data\_CoCounties.csv')
csv_d = csv.reader(d)

# RESULT data
for i in csv_d:
    FIPS = i[0]
    CountyName = i[1]

    # Construct the URL
    countyFIPS = FIPS ## the county FIPS
    countycode = 'US%3A08%3A' + countyFIPS ## USA, Colorado state

    # for RESULT data
    # note: start date is MM-DD-YYYY format
    my_url = 'http://www.waterqualitydata.us/Result/search?' + \
        'countycode=' + countycode + '&' + \
        'characteristicName=' + characteristicName + '&' + \
        'startDateLo=' + "01-01-2005" + '&' + \
        'mimeType=csv' + '&' + \
        'zip=no'

    # Open URL, GET data
    my_web = urllib.urlopen(my_url)
    htmlSource = my_web.read()
    my_web.close()
    print "result scraped for " + characteristicName + " for " + CountyName

    # Write data to csv
    outfile = open(filePath + '\\\ ' + characteristicName + '_' + CountyName +
'_Result' + '.csv', 'w')
    outfile.write(htmlSource)
    outfile.close()
    print "result written for " + characteristicName + " for " + CountyName

d.close()

```

### ***R Code to Clean Data (Preparation for Shiny Application)***

```

### Mark Ziman
### 4-5-2016

### clean all big 5 data:
###   selects county (all CO counties)
###   filters by >10 occurances
###   calculate compliance ratio
###   create xts dataframe for dygraph
###   create trends dataframe for trends bargraph

library(leaflet)
library(magrittr)
library(plyr)
library(RColorBrewer)
library(xts)
library(date)
library(chron)
library(lubridate)
require(zoo)

##### define data file directory (when csv files will be pulled from) #####

```

```

dat.dir <- "C:/Users/Mark/Documents/Duke/MP/WQRAT/Data"

##### get data (for >10 obs) #####

# list all county names
CoCount <-
c("Adams", "Alamosa", "Arapahoe", "Archuleta", "Baca", "Bent", "Boulder", "Broomfield", "Chaff
ee", "Cheyenne",

"ClearCreek", "Conejos", "Costilla", "Crowley", "Custer", "Delta", "Denver", "Dolores", "Dougl
as", "Eagle",

"Elbert", "ElPaso", "Fremont", "Garfield", "Gilpin", "Grand", "Gunnison", "Hinsdale", "Huerfan
o", "Jackson",

"Jefferson", "Kiowa", "KitCarson", "LaPlata", "Lake", "LasAnimas", "Lincoln", "Logan", "Mesa",

"Mineral", "Moffat", "Montezuma", "Montrose", "Morgan", "Otero", "Ouray", "Park", "Pitkin", "Pr
owers",

"Pueblo", "RioBlanco", "RioGrande", "Routt", "Saguache", "SanJuan", "SanMiguel", "Sedgwick", "
Summit",
          "Teller", "Washington", "Weld", "Yuma")

# create map data frame
map.df.headers <- c("MonitoringLocationIdentifier",
                  "MonitoringLocationName",
                  "MonitoringLocationTypeName",
                  "HUCEightDigitCode",
                  "DrainageAreaMeasure.MeasureValue",
                  "DrainageAreaMeasure.MeasureUnitCode",
                  "LatitudeMeasure",
                  "LongitudeMeasure",
                  "HorizontalCoordinateReferenceSystemDatumName",
                  "StateCode",
                  "CountyCode",
                  "freq",
                  "OrganizationFormalName",
                  "ActivityStartDate",
                  "CharacteristicName",
                  "ResultMeasureValue",
                  "ResultMeasure.MeasureUnitCode",
                  "ProviderName")

# create data frames for big 5
### this is probably unnecessary:
map.df.pH <- as.data.frame(matrix(nrow=0,ncol=18)) # pH
colnames(map.df.pH) <- map.df.headers
map.df.temp <- as.data.frame(matrix(nrow=0,ncol=18)) # temperature
colnames(map.df.temp) <- map.df.headers
map.df.turb <- as.data.frame(matrix(nrow=0,ncol=18)) # turbidity
colnames(map.df.turb) <- map.df.headers
map.df.spcn <- as.data.frame(matrix(nrow=0,ncol=18)) # specific conductivity
colnames(map.df.spcn) <- map.df.headers
map.df.DO <- as.data.frame(matrix(nrow=0,ncol=18)) # DO
colnames(map.df.DO) <- map.df.headers

# loop through list big5 and fill the data frames
big5 <- c("pH", "Temperature,%20water", "Turbidity", "Specific%20conductance",
"Dissolved%20oxygen%20(DO)")

shmy <- 0

```

```

for (j in 1:5){
  shmy[j] <- big5[j]
  print(shmy[j])

  map.df <- as.data.frame(matrix(nrow=0,ncol=18))      # for inter-loop functions
  colnames(map.df) <- map.df.headers

  usq <- 0
  for(i in 1:24){
    usq[i] <- CoCount[i]

    # read in station and results
    if(file.info(paste0("C:/Users/Mark/Documents/Duke/MP/WQRAT/Data", shmy[j], "_",
usq[i], "_Station.csv"))$size == 0){
      next
    }
    else
      sttn.df <- read.csv(paste0("C:/Users/Mark/Documents/Duke/MP/WQRAT/Data",
shmy[j], "_", usq[i], "_Station.csv"), header=T)

    if(file.info(paste0("C:/Users/Mark/Documents/Duke/MP/WQRAT/Data", shmy[j], "_",
usq[i], "_Result.csv"))$size == 0){
      next
    }
    else
      rslt.df <- read.csv(paste0("C:/Users/Mark/Documents/Duke/MP/WQRAT/Data",
shmy[j], "_", usq[i], "_Result.csv"), header=T)

    # grab only monitoring locations with occurrence >10
    # generate data frame of frequencies of MonitoringLocationIdentifier
    tab <- count(rslt.df, 'MonitoringLocationIdentifier')
    # create the bins which to break frequency data by ==== 10
    bins <- c(0, 10, Inf)
    # create column "cutoff" based off freq and bins
    tab$cutoff <- (findInterval(tab$freq, bins, all.inside = TRUE)) - 1
    # go through each row and determine if a value is zero
    row_sub = apply(tab, 1, function(row) all(row !=0))
    # update original table based on subset
    tab <- tab[row_sub,]
    # merge subset table with original data to apply frequency threshold
    rslt.df_10 <- merge(tab[, c("MonitoringLocationIdentifier",
                              "freq")],
                       rslt.df[, c("MonitoringLocationIdentifier",
                              "OrganizationFormalName",
                              "ActivityStartDate",
                              "CharacteristicName",
                              "ResultMeasureValue",
                              "ResultMeasure.MeasureUnitCode",
                              "ProviderName")])

    rslt.df_10$MonitoringLocationIdentifier <-
as.character(rslt.df_10$MonitoringLocationIdentifier)

    # make date a date
    rslt.df_10$ActivityStartDate <- as.Date(rslt.df_10$ActivityStartDate)

    # merge station and results data
    df.dat <- merge(sttn.df[, c("MonitoringLocationIdentifier",
                              "MonitoringLocationName",
                              "MonitoringLocationTypeName",
                              "HUCEightDigitCode",
                              "DrainageAreaMeasure.MeasureValue",
                              "DrainageAreaMeasure.MeasureUnitCode",
                              "LatitudeMeasure",

```



```

        "LongitudeMeasure",
        "HorizontalCoordinateReferenceSystemDatumName",
        "StateCode",
        "CountyCode"]],
    rslt.df_10[, c("MonitoringLocationIdentifier",
        "freq",
        "OrganizationFormalName",
        "ActivityStartDate",
        "CharacteristicName",
        "ResultMeasureValue",
        "ResultMeasure.MeasureUnitCode",
        "ProviderName")]])

# add new county data to map.df
map.df <- rbind(map.df, df.dat)

# print progress
print(paste("usq loop completed for", usq[i]))
} # end usq loop

##### work each constituent data #####

if(shmy[j]=="pH"){
  # add compliance ratio
  Unique.id <- unique(map.df$MonitoringLocationIdentifier) # unique id
  #Create dataframe
  foo <- as.data.frame(matrix(nrow=0, ncol=2))
  colnames(foo) <- c("MonitoringLocationIdentifier", "ComplianceRatio")
  #Loop through -->
  for (i in 1:(length(Unique.id))){
    zt <- subset(map.df, map.df$MonitoringLocationIdentifier==Unique.id[i]) # subset
    CompRat <- length(zt$ResultMeasureValue[zt$ResultMeasureValue<6.5 |
zt$ResultMeasureValue > 9 ])/length(zt$ResultMeasureValue)
    #do math
    foo[i, 1] <- as.character(Unique.id[i])
    foo[i, 2] <- CompRat #fill data frame
  }
  map.df <- merge(map.df[, c("MonitoringLocationIdentifier",
        "MonitoringLocationName",
        "MonitoringLocationTypeName",
        "HUCEightDigitCode",
        "DrainageAreaMeasure.MeasureValue",
        "DrainageAreaMeasure.MeasureUnitCode",
        "LatitudeMeasure",
        "LongitudeMeasure",
        "HorizontalCoordinateReferenceSystemDatumName",
        "StateCode",
        "CountyCode",
        "freq",
        "OrganizationFormalName",
        "ActivityStartDate",
        "CharacteristicName",
        "ResultMeasureValue",
        "ResultMeasure.MeasureUnitCode",
        "ProviderName")],
    foo[, c("MonitoringLocationIdentifier",
        "ComplianceRatio")])

# sort by compliance ratio
map.df <- map.df[order(map.df$ComplianceRatio),]
# rename map as pH
map.df.pH <- map.df
print(paste0("map.df created for ", shmy[j]))

```

```

}
else if(shmy[j]=="Temperature,%20water"){
  Unique.id <- unique(map.df$MonitoringLocationIdentifier) # unique id
  # Create dataframe
  foo <- as.data.frame(matrix(nrow=0, ncol=2))
  colnames(foo) <- c("MonitoringLocationIdentifier", "ComplianceRatio")
  # Loop through -->
  for (i in 1:(length(Unique.id))){
    zt <- subset(map.df, map.df$MonitoringLocationIdentifier==Unique.id[i]) # subset
    CompRat <- mean(zt$ResultMeasureValue)
    # ^ do math
    foo[i, 1] <- as.character(Unique.id[i])
    foo[i, 2] <- CompRat #fill data frame
  }
  map.df <- merge(map.df[, c("MonitoringLocationIdentifier",
                            "MonitoringLocationName",
                            "MonitoringLocationTypeName",
                            "HUCEightDigitCode",
                            "DrainageAreaMeasure.MeasureValue",
                            "DrainageAreaMeasure.MeasureUnitCode",
                            "LatitudeMeasure",
                            "LongitudeMeasure",
                            "HorizontalCoordinateReferenceSystemDatumName",
                            "StateCode",
                            "CountyCode",
                            "freq",
                            "OrganizationFormalName",
                            "ActivityStartDate",
                            "CharacteristicName",
                            "ResultMeasureValue",
                            "ResultMeasure.MeasureUnitCode",
                            "ProviderName")],
                 foo[, c("MonitoringLocationIdentifier",
                         "ComplianceRatio")])
  # sort by compliance ratio
  map.df <- map.df[order(map.df$ComplianceRatio),]
  # rename map as temp
  map.df.temp <- map.df
  print(paste0("map.df created for ", shmy[j]))
}
else if(shmy[j]=="Turbidity"){
  # add compliance ratio
  Unique.id <- unique(map.df$MonitoringLocationIdentifier) # unique id
  #Create dataframe
  foo <- as.data.frame(matrix(nrow=0, ncol=2))
  colnames(foo) <- c("MonitoringLocationIdentifier", "ComplianceRatio")
  #Loop through -->
  for (i in 1:(length(Unique.id))){
    zt <- subset(map.df, map.df$MonitoringLocationIdentifier==Unique.id[i]) # subset
    CompRat <- length(zt$ResultMeasureValue[zt$ResultMeasureValue<0 |
zt$ResultMeasureValue > 50 ])/length(zt$ResultMeasureValue)
    #do math
    foo[i, 1] <- as.character(Unique.id[i])
    foo[i, 2] <- CompRat #fill data frame
  }
  map.df <- merge(map.df[, c("MonitoringLocationIdentifier",
                            "MonitoringLocationName",
                            "MonitoringLocationTypeName",
                            "HUCEightDigitCode",
                            "DrainageAreaMeasure.MeasureValue",
                            "DrainageAreaMeasure.MeasureUnitCode",
                            "LatitudeMeasure",
                            "LongitudeMeasure",

```

```

        "HorizontalCoordinateReferenceSystemDatumName",
        "StateCode",
        "CountyCode",
        "freq",
        "OrganizationFormalName",
        "ActivityStartDate",
        "CharacteristicName",
        "ResultMeasureValue",
        "ResultMeasure.MeasureUnitCode",
        "ProviderName"]],
        foo[, c("MonitoringLocationIdentifier",
               "ComplianceRatio")])
# sort by compliance ratio
map.df <- map.df[order(map.df$ComplianceRatio),]
# rename map as turb
map.df.turb <- map.df
print(paste0("map.df created for ", shmy[j]))
}
else if(shmy[j]=="Specific%20conductance"){
# add compliance ratio
Unique.id <- unique(map.df$MonitoringLocationIdentifier) # unique id
#Create dataframe
foo <- as.data.frame(matrix(nrow=0, ncol=2))
colnames(foo) <- c("MonitoringLocationIdentifier", "ComplianceRatio")
#Loop through -->
for (i in 1:(length(Unique.id))){
  zt <- subset(map.df, map.df$MonitoringLocationIdentifier==Unique.id[i]) # subset
  CompRat <- length(zt$ResultMeasureValue[zt$ResultMeasureValue<200 |
zt$ResultMeasureValue > 1000 ])/length(zt$ResultMeasureValue)
  #do math
  foo[i, 1] <- as.character(Unique.id[i])
  foo[i, 2] <- CompRat #fill data frame
}
map.df <- merge(map.df[, c("MonitoringLocationIdentifier",
                          "MonitoringLocationName",
                          "MonitoringLocationTypeName",
                          "HUCEightDigitCode",
                          "DrainageAreaMeasure.MeasureValue",
                          "DrainageAreaMeasure.MeasureUnitCode",
                          "LatitudeMeasure",
                          "LongitudeMeasure",
                          "HorizontalCoordinateReferenceSystemDatumName",
                          "StateCode",
                          "CountyCode",
                          "freq",
                          "OrganizationFormalName",
                          "ActivityStartDate",
                          "CharacteristicName",
                          "ResultMeasureValue",
                          "ResultMeasure.MeasureUnitCode",
                          "ProviderName")],
               foo[, c("MonitoringLocationIdentifier",
                       "ComplianceRatio")])
# sort by compliance ratio
map.df <- map.df[order(map.df$ComplianceRatio),]
# rename map as spec con
map.df.spcn <- map.df
print(paste0("map.df created for ", shmy[j]))
}
else{ # dissolved oxygen
# add compliance ratio
Unique.id <- unique(map.df$MonitoringLocationIdentifier) # unique id
#Create dataframe

```

```

foo <- as.data.frame(matrix(nrow=0, ncol=2))
colnames(foo) <- c("MonitoringLocationIdentifier", "ComplianceRatio")
#Loop through -->
for (i in 1:(length(Unique.id))){
  zt <- subset(map.df, map.df$MonitoringLocationIdentifier==Unique.id[i]) # subset
  CompRat <- length(zt$ResultMeasureValue[zt$ResultMeasureValue<3 |
zt$ResultMeasureValue > 20 ])/length(zt$ResultMeasureValue)
  #do math
  foo[i, 1] <- as.character(Unique.id[i])
  foo[i, 2] <- CompRat #fill data frame
}
map.df <- merge(map.df[, c("MonitoringLocationIdentifier",
"MonitoringLocationName",
"MonitoringLocationTypeName",
"HUCEightDigitCode",
"DrainageAreaMeasure.MeasureValue",
"DrainageAreaMeasure.MeasureUnitCode",
"LatitudeMeasure",
"LongitudeMeasure",
"HorizontalCoordinateReferenceSystemDatumName",
"StateCode",
"CountyCode",
"freq",
"OrganizationFormalName",
"ActivityStartDate",
"CharacteristicName",
"ResultMeasureValue",
"ResultMeasure.MeasureUnitCode",
"ProviderName")],
foo[, c("MonitoringLocationIdentifier",
"ComplianceRatio")])
# sort by compliance ratio
map.df <- map.df[order(map.df$ComplianceRatio),]
# rename map as dissolved oxygen
map.df.DO <- map.df
print(paste0("map.df created for ", shmy[j]))
}
} # end shmy loop

##### make dygraph xts and trends graph data
#####

# create default bardat (def.bardat)
mos <- c(1,2,3,4,5,6,7,8,9,10,11,12)
momos <- rep(mos, 11)
yrs <- c(rep(2005,12), rep(2006,12), rep(2007,12), rep(2008,12), rep(2009,12),
rep(2010,12),
rep(2011,12), rep(2012,12), rep(2013,12), rep(2014,12), rep(2015,12))
def.bardat <- as.data.frame(cbind(momos, yrs))
colnames(def.bardat) <- c("month", "year")
def.bardat$yearmonth <- paste0(def.bardat$year, "-", def.bardat$month)
def.bardat$date <- as.yearmon(def.bardat$yearmonth)
def.bardat$date2 <- as.character(def.bardat$date)

# make xts and trend files for pH
print(paste0("begin: create xts and trends data for map.df.pH"))
my.uniq.id <- unique(map.df.pH$MonitoringLocationIdentifier) # unique id
for(v in 1:(length(my.uniq.id))){
  # make subset for each location ID
  st <- subset(map.df.pH, map.df.pH$MonitoringLocationIdentifier==my.uniq.id[v])

  # make xts file
  my.xtsA <- xts(x = st$ResultMeasureValue,

```

```

        order.by = st$ActivityStartDate)
my.xtsB <- my.xtsA
my.xts <- cbind(my.xtsA, my.xtsB)
colnames(my.xts) <- c("In_Comp", "Out_of_Comp")
my.xts$In_Comp[my.xts$In_Comp < 6.5 | my.xts$In_Comp > 9 ] <- NA
my.xts$Out_of_Comp[my.xts$Out_of_Comp >= 6.5 & my.xts$Out_of_Comp <= 9 ] <- NA
assign(paste0("my.xts_pH_", my.uni.id[v]), my.xts)

# make bardat
data <- st
# add month and year columns
data$dtstr <- as.character((data$ActivityStartDate))
data$yr <- sapply(strsplit(data$dtstr, "-") , "[", 1)
data$mon <- sapply(strsplit(data$dtstr, "-") , "[", 2)
data$mon <- as.numeric(data$mon)
# calculate average by month over time period
avemo <- as.data.frame(matrix(nrow=0, ncol=2))
colnames(avemo) <- c("month", "aveph")
for (i in 1:12){
  zt <- subset(data, mon==i)
  meanph <- mean(zt$ResultMeasureValue, na.rm=T)
  avemo[i,1] <- i
  avemo[i,2] <- meanph
}
# bind back to dataframe
datamean <- merge(data, avemo, by.x="mon", by.y="month", all.x = TRUE)
# calculate difference
datamean$difffh <- datamean$ResultMeasureValue - datamean$aveph
# create date month/year
datamean$yearmonth <- paste0(datamean$yr, "-", datamean$mon)
datamean$date <- as.yearmon(datamean$yearmonth)
# calculate ave of difference by year
unique.date <- unique(datamean$date)
yearmoaveph <- as.data.frame(matrix(nrow=0, ncol=2))
colnames(yearmoaveph) <- c("date", "avediff")
for (i in 1:length(unique.date)){
  zt <- subset(datamean, yearmonth == unique.date[i])
  aveyrmo <- mean(zt$difffh, na.rm=T)
  yearmoaveph[i,1] <- as.character(zt$date[1])
  yearmoaveph[i,2] <- aveyrmo
}
yearmoaveph$date <- as.yearmon(yearmoaveph$date)
# order data
bardat <- yearmoaveph[with(yearmoaveph, order(date)), ]
bardat$date <- as.character(bardat$date)
bardat$month <- substring(bardat$date, 1, 3)
bardat$year <- substring(bardat$date, 5, 8)
bardat <- merge(def.bardat,
               bardat,
               by.x = "date2", by.y = "date",
               all.x=T)
bardat2 <- bardat[with(bardat, order(year.x)), ]
# redefine month column from numbers to names
for(k in 1:nrow(bardat2)){
  if(bardat2[k,2] == 1){
    bardat2[k,9] <- "January"
  }
  else if(bardat2[k,2] == 2){
    bardat2[k,9] <- "February"
  }
  else if(bardat2[k,2] == 3){
    bardat2[k,9] <- "March"
  }
}

```

```

else if(bardat2[k,2] == 4){
  bardat2[k,9] <- "April"
}
else if(bardat2[k,2] == 5){
  bardat2[k,9] <- "May"
}
else if(bardat2[k,2] == 6){
  bardat2[k,9] <- "June"
}
else if(bardat2[k,2] == 7){
  bardat2[k,9] <- "July"
}
else if(bardat2[k,2] == 8){
  bardat2[k,9] <- "August"
}
else if(bardat2[k,2] == 9){
  bardat2[k,9] <- "September"
}
else if(bardat2[k,2] == 10){
  bardat2[k,9] <- "October"
}
else if(bardat2[k,2] == 11){
  bardat2[k,9] <- "November"
}
else{
  bardat2[k,9] <- "December"
}
} # end loop to rename month from number to name
assign(paste0("bardat_pH_", my.uni.q.id[v]), bardat2)
}
print(paste0("completed: created xts and trends data for map.df.pH"))

# make xts and trends files for temp
print(paste0("begin: create xts and trends data for map.df.temp"))
my.uni.q.id <- unique(map.df.temp$MonitoringLocationIdentifier) # unique id
for(v in 1:(length(my.uni.q.id))){
  # make subset for each location ID
  st <- subset(map.df.temp, map.df.temp$MonitoringLocationIdentifier==my.uni.q.id[v])

  # make xts file
  my.xtsA <- xts(x = st$ResultMeasureValue,
                order.by = st$ActivityStartDate)
  my.xtsB <- my.xtsA
  my.xts <- cbind(my.xtsA, my.xtsB)
  colnames(my.xts) <- c("In_Comp", "Out_of_Comp")
  my.xts$In_Comp[my.xts$In_Comp >= 20] <- NA
  my.xts$Out_of_Comp[my.xts$Out_of_Comp < 20] <- NA
  assign(paste0("my.xts_temp_", my.uni.q.id[v]), my.xts)

  # make bardat
  data <- st
  # add month and year columns
  data$dtstr <- as.character((data$ActivityStartDate))
  data$yr <- sapply(strsplit(data$dtstr, "-") , "[", 1)
  data$mon <- sapply(strsplit(data$dtstr, "-") , "[", 2)
  data$mon <- as.numeric(data$mon)
  # calculate average by month over time period
  avemo <- as.data.frame(matrix(nrow=0,ncol=2))
  colnames(avemo) <- c("month", "aveph")
  for (i in 1:12){
    zt <- subset(data, mon==i)
    meanph <- mean(zt$ResultMeasureValue, na.rm=T)
    avemo[i,1] <- i
  }
}

```

```

    avemo[i,2] <- meanph
  }
# bind back to dataframe
datamean <- merge(data, avemo, by.x="mon", by.y="month", all.x = TRUE)
# calculate difference
datamean$difffph <- datamean$ResultMeasureValue - datamean$aveph
# create date month/year
datamean$yearmonth <- paste0(datamean$yr,"-",datamean$mon)
datamean$date <- as.yearmon(datamean$yearmonth)
# calculate ave of difference by year
unique.date <- unique(datamean$date)
yearmoaveph <- as.data.frame(matrix(nrow=0,ncol=2))
colnames(yearmoaveph) <- c("date","avediff")
for (i in 1:length(unique.date)){
  zt <- subset(datamean, yearmonth == unique.date[i])
  aveyrmo <- mean(zt$difffph, na.rm=T)
  yearmoaveph[i,1] <- as.character(zt$date[1])
  yearmoaveph[i,2] <- aveyrmo
}
yearmoaveph$date <- as.yearmon(yearmoaveph$date)
# order data
bardat <- yearmoaveph[with(yearmoaveph, order(date)), ]
bardat$date <- as.character(bardat$date)
bardat$month <- substring(bardat$date, 1, 3)
bardat$year <- substring(bardat$date, 5, 8)
bardat <- merge(def.bardat,
                bardat,
                by.x = "date2", by.y = "date",
                all.x=T)
bardat2 <- bardat[with(bardat, order(year.x)), ]
# redefine month column from numbers to names
for(k in 1:nrow(bardat2)){
  if(bardat2[k,2] == 1){
    bardat2[k,9] <- "January"
  }
  else if(bardat2[k,2] == 2){
    bardat2[k,9] <- "February"
  }
  else if(bardat2[k,2] == 3){
    bardat2[k,9] <- "March"
  }
  else if(bardat2[k,2] == 4){
    bardat2[k,9] <- "April"
  }
  else if(bardat2[k,2] == 5){
    bardat2[k,9] <- "May"
  }
  else if(bardat2[k,2] == 6){
    bardat2[k,9] <- "June"
  }
  else if(bardat2[k,2] == 7){
    bardat2[k,9] <- "July"
  }
  else if(bardat2[k,2] == 8){
    bardat2[k,9] <- "August"
  }
  else if(bardat2[k,2] == 9){
    bardat2[k,9] <- "September"
  }
  else if(bardat2[k,2] == 10){
    bardat2[k,9] <- "October"
  }
  else if(bardat2[k,2] == 11){

```

```

        bardat2[k,9] <- "November"
    }
    else{
        bardat2[k,9] <- "December"
    }
} # end loop to rename month from number to name
assign(paste0("bardat_temp_", my.uniq.id[v]), bardat2)
}
print(paste0("completed: created xts and trends data for map.df.temp"))

# make xts and trends files for turb
print(paste0("begin: create xts and trendsdata for map.df.turb"))
my.uniq.id <- unique(map.df.turb$MonitoringLocationIdentifier) # unique id
for(v in 1:(length(my.uniq.id))){
    # make subset for each location ID
    st <- subset(map.df.turb, map.df.turb$MonitoringLocationIdentifier==my.uniq.id[v])

    # make xts file
    my.xtsA <- xts(x = st$ResultMeasureValue,
                  order.by = st$ActivityStartDate)
    my.xtsB <- my.xtsA
    my.xts <- cbind(my.xtsA, my.xtsB)
    colnames(my.xts) <- c("In_Comp", "Out_of_Comp")
    my.xts$In_Comp[my.xts$In_Comp < 0 | my.xts$In_Comp > 50 ] <- NA
    my.xts$Out_of_Comp[my.xts$Out_of_Comp >= 0 & my.xts$Out_of_Comp <= 50 ] <- NA
    assign(paste0("my.xts_turb_", my.uniq.id[v]), my.xts)

    # make bardat
    data <- st
    # add month and year columns
    data$dtstr <- as.character((data$ActivityStartDate))
    data$yr <- sapply(strsplit(data$dtstr, "-") , "[", 1)
    data$mon <- sapply(strsplit(data$dtstr, "-") , "[", 2)
    data$mon <- as.numeric(data$mon)
    # calculate average by month over time period
    avemo <- as.data.frame(matrix(nrow=0,ncol=2))
    colnames(avemo) <- c("month", "aveph")
    for (i in 1:12){
        zt <- subset(data, mon==i)
        meanph <- mean(zt$ResultMeasureValue, na.rm=T)
        avemo[i,1] <- i
        avemo[i,2] <- meanph
    }
    # bind back to dataframe
    datamean <- merge(data, avemo, by.x="mon", by.y="month", all.x = TRUE)
    # calculate difference
    datamean$diffph <- datamean$ResultMeasureValue - datamean$aveph
    # create date month/year
    datamean$yearmonth <- paste0(datamean$yr, "-", datamean$mon)
    datamean$date <- as.yearmon(datamean$yearmonth)
    # calculate ave of difference by year
    unique.date <- unique(datamean$date)
    yearmoaveph <- as.data.frame(matrix(nrow=0,ncol=2))
    colnames(yearmoaveph) <- c("date", "avediff")
    for (i in 1:length(unique.date)){
        zt <- subset(datamean, yearmonth == unique.date[i])
        aveyrmo <- mean(zt$diffph, na.rm=T)
        yearmoaveph[i,1] <- as.character(zt$date[1])
        yearmoaveph[i,2] <- aveyrmo
    }
    yearmoaveph$date <- as.yearmon(yearmoaveph$date)
    # order data
    bardat <- yearmoaveph[with(yearmoaveph, order(date)), ]
}

```



```

bardat$date <- as.character(bardat$date)
bardat$month <- substring(bardat$date, 1, 3)
bardat$year <- substring(bardat$date, 5, 8)
bardat <- merge(def.bardat,
               bardat,
               by.x = "date2", by.y = "date",
               all.x=T)
bardat2 <- bardat[with(bardat, order(year.x)), ]
# redefine month column from numbers to names
for(k in 1:nrow(bardat2)){
  if(bardat2[k,2] == 1){
    bardat2[k,9] <- "January"
  }
  else if(bardat2[k,2] == 2){
    bardat2[k,9] <- "February"
  }
  else if(bardat2[k,2] == 3){
    bardat2[k,9] <- "March"
  }
  else if(bardat2[k,2] == 4){
    bardat2[k,9] <- "April"
  }
  else if(bardat2[k,2] == 5){
    bardat2[k,9] <- "May"
  }
  else if(bardat2[k,2] == 6){
    bardat2[k,9] <- "June"
  }
  else if(bardat2[k,2] == 7){
    bardat2[k,9] <- "July"
  }
  else if(bardat2[k,2] == 8){
    bardat2[k,9] <- "August"
  }
  else if(bardat2[k,2] == 9){
    bardat2[k,9] <- "September"
  }
  else if(bardat2[k,2] == 10){
    bardat2[k,9] <- "October"
  }
  else if(bardat2[k,2] == 11){
    bardat2[k,9] <- "November"
  }
  else{
    bardat2[k,9] <- "December"
  }
} # end loop to rename month from number to name
assign(paste0("bardat_turb_", my.uniq.id[v]), bardat2)
}
print(paste0("completed: created xts data for map.df.turb"))

# make xts and trends files for spcn
print(paste0("begin: create xts and trends data for map.df.spcn"))
my.uniq.id <- unique(map.df.spcn$MonitoringLocationIdentifier) # unique id
for(v in 1:(length(my.uniq.id))){
  # make subset for each location ID
  st <- subset(map.df.spcn, map.df.spcn$MonitoringLocationIdentifier==my.uniq.id[v])

  # make xts file
  my.xtsA <- xts(x = st$ResultMeasureValue,
                order.by = st$ActivityStartDate)
  my.xtsB <- my.xtsA
  my.xts <- cbind(my.xtsA, my.xtsB)
}

```

```

colnames(my.xts) <- c("In_Comp","Out_of_Comp")
my.xts$In_Comp[my.xts$In_Comp < 400 | my.xts$In_Comp > 1350 ] <- NA
my.xts$Out_of_Comp[my.xts$Out_of_Comp >= 400 & my.xts$Out_of_Comp <= 1350 ] <- NA
assign(paste0("my.xts_spcn_", my.uni.q.id[v]), my.xts)

# make bardat
data <- st
# add month and year columns
data$dtstr <- as.character((data$ActivityStartDate))
data$yr <- sapply(strsplit(data$dtstr, "-") , "[", 1)
data$mon <- sapply(strsplit(data$dtstr, "-") , "[", 2)
data$mon <- as.numeric(data$mon)
# calculate average by month over time period
avemo <- as.data.frame(matrix(nrow=0,ncol=2))
colnames(avemo) <- c("month","aveph")
for (i in 1:12){
  zt <- subset(data, mon==i)
  meanph <- mean(zt$ResultMeasureValue, na.rm=T)
  avemo[i,1] <- i
  avemo[i,2] <- meanph
}
# bind back to dataframe
datamean <- merge(data, avemo, by.x="mon", by.y="month", all.x = TRUE)
# calculate difference
datamean$diffph <- datamean$ResultMeasureValue - datamean$aveph
# create date month/year
datamean$yearmonth <- paste0(datamean$yr,"-",datamean$mon)
datamean$date <- as.yearmon(datamean$yearmonth)
# calculate ave of difference by year
unique.date <- unique(datamean$date)
yearmoaveph <- as.data.frame(matrix(nrow=0,ncol=2))
colnames(yearmoaveph) <- c("date","avediff")
for (i in 1:length(unique.date)){
  zt <- subset(datamean, yearmonth == unique.date[i])
  aveyrmo <- mean(zt$diffph, na.rm=T)
  yearmoaveph[i,1] <- as.character(zt$date[1])
  yearmoaveph[i,2] <- aveyrmo
}
yearmoaveph$date <- as.yearmon(yearmoaveph$date)
# order data
bardat <- yearmoaveph[with(yearmoaveph, order(date)), ]
bardat$date <- as.character(bardat$date)
bardat$month <- substring(bardat$date, 1, 3)
bardat$year <- substring(bardat$date, 5, 8)
bardat <- merge(def.bardat,
               bardat,
               by.x = "date2", by.y = "date",
               all.x=T)
bardat2 <- bardat[with(bardat, order(year.x)), ]
# redefine month column from numbers to names
for(k in 1:nrow(bardat2)){
  if(bardat2[k,2] == 1){
    bardat2[k,9] <- "January"
  }
  else if(bardat2[k,2] == 2){
    bardat2[k,9] <- "February"
  }
  else if(bardat2[k,2] == 3){
    bardat2[k,9] <- "March"
  }
  else if(bardat2[k,2] == 4){
    bardat2[k,9] <- "April"
  }
}

```

```

else if(bardat2[k,2] == 5){
  bardat2[k,9] <- "May"
}
else if(bardat2[k,2] == 6){
  bardat2[k,9] <- "June"
}
else if(bardat2[k,2] == 7){
  bardat2[k,9] <- "July"
}
else if(bardat2[k,2] == 8){
  bardat2[k,9] <- "August"
}
else if(bardat2[k,2] == 9){
  bardat2[k,9] <- "September"
}
else if(bardat2[k,2] == 10){
  bardat2[k,9] <- "October"
}
else if(bardat2[k,2] == 11){
  bardat2[k,9] <- "November"
}
else{
  bardat2[k,9] <- "December"
}
} # end loop to rename month from number to name
assign(paste0("bardat_spcn_", my.uniq.id[v]), bardat2)
}
print(paste0("completed: created xts and trends data for map.df.spcn"))

# make xts and trends files for DO
print(paste0("begin: create xts and trends data for map.df.DO"))
my.uniq.id <- unique(map.df.DO$MonitoringLocationIdentifier) # unique id
for(v in 1:(length(my.uniq.id))){
  # make subset for each location ID
  st <- subset(map.df.DO, map.df.DO$MonitoringLocationIdentifier==my.uniq.id[v])

  # make xts file
  my.xtsA <- xts(x = st$ResultMeasureValue,
                order.by = st$ActivityStartDate)
  my.xtsB <- my.xtsA
  my.xts <- cbind(my.xtsA, my.xtsB)
  colnames(my.xts) <- c("In_Comp", "Out_of_Comp")
  my.xts$In_Comp[my.xts$In_Comp < 3 ] <- NA
  my.xts$Out_of_Comp[my.xts$Out_of_Comp >= 3 ] <- NA
  assign(paste0("my.xts_DO_", my.uniq.id[v]), my.xts)

  # make bardat
  data <- st
  # add month and year columns
  data$dtstr <- as.character((data$ActivityStartDate))
  data$yr <- sapply(strsplit(data$dtstr, "-") , "[", 1)
  data$mon <- sapply(strsplit(data$dtstr, "-") , "[", 2)
  data$mon <- as.numeric(data$mon)
  # calculate average by month over time period
  avemo <- as.data.frame(matrix(nrow=0, ncol=2))
  colnames(avemo) <- c("month", "aveph")
  for (i in 1:12){
    zt <- subset(data, mon==i)
    meanph <- mean(zt$ResultMeasureValue, na.rm=T)
    avemo[i,1] <- i
    avemo[i,2] <- meanph
  }
}
# bind back to dataframe

```

```

datamean <- merge(data, avemo, by.x="mon", by.y="month", all.x = TRUE)
# calculate difference
datamean$difffph <- datamean$ResultMeasureValue - datamean$aveph
# create date month/year
datamean$yearmonth <- paste0(datamean$yr,"-",datamean$mon)
datamean$date <- as.yearmon(datamean$yearmonth)
# calculate ave of difference by year
unique.date <- unique(datamean$date)
yearmoaveph <- as.data.frame(matrix(nrow=0,ncol=2))
colnames(yearmoaveph) <- c("date","avediff")
for (i in 1:length(unique.date)){
  zt <- subset(datamean, yearmonth == unique.date[i])
  aveyrmo <- mean(zt$difffph, na.rm=T)
  yearmoaveph[i,1] <- as.character(zt$date[1])
  yearmoaveph[i,2] <- aveyrmo
}
yearmoaveph$date <- as.yearmon(yearmoaveph$date)
# order data
bardat <- yearmoaveph[with(yearmoaveph, order(date)), ]
bardat$date <- as.character(bardat$date)
bardat$month <- substring(bardat$date, 1, 3)
bardat$year <- substring(bardat$date, 5, 8)
bardat <- merge(def.bardat,
               bardat,
               by.x = "date2", by.y = "date",
               all.x=T)
bardat2 <- bardat[with(bardat, order(year.x)), ]
# redefine month column from numbers to names
for(k in 1:nrow(bardat2)){
  if(bardat2[k,2] == 1){
    bardat2[k,9] <- "January"
  }
  else if(bardat2[k,2] == 2){
    bardat2[k,9] <- "February"
  }
  else if(bardat2[k,2] == 3){
    bardat2[k,9] <- "March"
  }
  else if(bardat2[k,2] == 4){
    bardat2[k,9] <- "April"
  }
  else if(bardat2[k,2] == 5){
    bardat2[k,9] <- "May"
  }
  else if(bardat2[k,2] == 6){
    bardat2[k,9] <- "June"
  }
  else if(bardat2[k,2] == 7){
    bardat2[k,9] <- "July"
  }
  else if(bardat2[k,2] == 8){
    bardat2[k,9] <- "August"
  }
  else if(bardat2[k,2] == 9){
    bardat2[k,9] <- "September"
  }
  else if(bardat2[k,2] == 10){
    bardat2[k,9] <- "October"
  }
  else if(bardat2[k,2] == 11){
    bardat2[k,9] <- "November"
  }
  else{

```

```

        bardat2[k,9] <- "December"
    }
} # end loop to rename month from number to name
assign(paste0("bardat_DO_", my.uniq.id[v]), bardat2)
}
print(paste0("completed: created xts and trends data for map.df.DO"))

```

### ***R Code to Run Shiny Application (UI and Server Included)***

```

### Mark Ziman
### WQRAT
### 4-5-2016
##### user interface #####

library(shiny)
library(shinythemes)
library(shinydashboard)
library(htmlwidgets)
library(dygraphs)
library(leaflet)
library(graphics)
library(ggplot2)

m <- leaflet() %>% addTiles()

ui <- (fluidPage(
  theme = shinytheme("spacelab"),
  h2("Water Quality Risk Assessment Tool", align="center"),
  h4(em("proof of concept for Colorado, USA"), align="center"),
  fluidRow(
    column(12,
      p("This application visualizes water quality measurements and trends for
dissolved oxygen, pH, specific conductance, temperature,
and turbidity data from the Water Quality Portal, including EPA's STORET,
USGS' NWIS, and USDA's STEWARDS databases. This proof
of concept version includes water quality data for Colorado, USA."),
      p("Created by Mark Ziman, master of environmental management candidate at
Duke University's Nicholas School of the Environment."),
      br()
    ) # end column
  ), # end headers fluidrow

  fluidRow(
    column(2,
      style="background-color: #F2F3F4",
      br(),
      radioButtons("var_layer",
        label = "Select constituent to map",
        c("dissolved oxygen" = "DO",
          "pH" = "pH",
          "specific conductance" = "specCon",
          "temperature" = "temp",
          "turbidity" = "turb"),
        selected = "DO"
      ), # end radioButtons
      br(),
      p("Click below for selected sample location results and trends."),
      actionButton("getdata", "View Data"),
      br(),
      br(),
      br(),
      br()
    )
  )
)

```

```

), # end column
column(7,
  leafletOutput("map")
),
column(3,
  strong("Summary Map:"),
  htmlOutput("rightMap")
) # end column
), # end fluidRow

fluidRow(
  column(2,
    p("")
  ), # end column
  column(7,
    textOutput("belowMap")
  ) # end column
), # end fluid Row

fluidRow(
  column(2,
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br()
  ), # end column
  column(7,
    br(),
    dygraphOutput("chart", width="100%", height=400)
  ), #End Column
  column(3,
    br(),
    br(),
    br(),
    strong("Results Plot:"),
    htmlOutput("rightResult"),
    br()
  ) # end column
), # end fluidRow
fluidRow(
  column(2,
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br(),
    br()
  ), # end column
  column(7,
    br(),
    plotOutput("plt"),
    br()
  )

```

```

),#End Column
column(3,
  br(),
  br(),
  br(),
  strong("Trends Graph:"),
  p("This bar graph illustrates the deviance of an average result measurement
    for a single month of a single year (ex: average February 2009) to the
    average result measurements of single month for all years (ex: average
    Februarys 2005 - 2015). This indicates the increasing/decreasing
result
    measurement trends and volatility of the result measurements.")
  ) # end column
) # end fluidrow
) # end fluidPage
) # end ui
### Mark Ziman
### WQRAT
### 4-5-2016
##### server #####

library(shiny)
library(shinythemes)
library(shinydashboard)
library(htmlwidgets)
library(dygraphs)
library(leaflet)
library(graphics)
library(ggplot2)

load('AllDat.RData')
m <- leaflet() %>% addTiles()

server <- function(input, output) {

  # radio button function
  map.dat <- eventReactive(input$var_layer, {
    if(input$var_layer == "pH"){
      map.df.pH}
    else if(input$var_layer == "specCon"){
      map.df.spcn}
    else if(input$var_layer == "turb"){
      map.df.turb}
    else if(input$var_layer == "temp"){
      map.df.temp}
    else
      map.df.DO
  })

  # define popup
  my.pops <- eventReactive(input$var_layer, {
    if(input$var_layer == "pH" | input$var_layer == "specCon" |input$var_layer ==
"turb" |input$var_layer == "DO"){
      map.df <- map.dat()
      paste0("<b>", "Monitoring Location Name: ",
map.df$MonitoringLocationName,"</b><br>",
          "Monitoring Location Identifier: ", map.df$MonitoringLocationIdentifier,
"<br>",
          "Data Provider: ", map.df$OrganizationFormalName, "<br>",
          "Out of Compliance Ratio: ", round(map.df$ComplianceRatio, 4))
    }
    else{
      map.df <- map.dat()

```

```

    paste0("<b>", "Monitoring Location Name: ",
map.df$MonitoringLocationName,"</b><br>",
        "Monitoring Location Identifier: ", map.df$MonitoringLocationIdentifier,
"<br>",
        "Data Provider: ", map.df$OrganizationFormalName, "<br>",
        "Average Temperature: ", round(map.df$ComplianceRatio, 1))
  }
})
# define colors
my.cols <- eventReactive(input$var_layer,{
  if(input$var_layer == "pH" | input$var_layer == "specCon" |input$var_layer ==
"turb" |input$var_layer == "DO"){
    mycols <- c("blue", "red")
    colorBin(palette = palette(mycols), domain = c(0,1), bins=7)
  }
  else{
    mycols <- c("blue", "red")
    colorBin(palette = palette(mycols), domain = c(-5,35), bins=7)
  }
})

# draw map
output$map <- renderLeaflet({
  # radio buttons output
  map.df <- map.dat()
  col <- my.cols()
  mypop <- my.pops()
  setView(m, lng = -105.3600, lat = 40.0900, zoom = 9)
  addCircleMarkers(m,
    map.df$LongitudeMeasure, map.df$LatitudeMeasure,
    color = col(map.df$ComplianceRatio),
    radius=(log(map.df$freq)*2),
    stroke = FALSE,
    fillOpacity = 1,
    popup = mypop,
    layerId=map.df$MonitoringLocationIdentifier) %>%
  addLegend(m, position="bottomright", pal=col)
})

# right of map text
output$rightMap <- renderUI({
  map.df <- map.dat()
  if(as.character(map.df$CharacteristicName[1])=="pH"){
    HTML(paste("US EPA water quality criteria for pH in freshwater suggest a range
of 6.5 to 9.", " ", "This map visualizes the frequency of pH measurement per sample
location outside EPA recommendations.", sep="<br/>"))
  }
  else if(as.character(map.df$CharacteristicName[1])=="Temperature, water"){
    HTML(paste("Water temperature varies significantly due to seasonal climate
regimes.", " ", "This map visualizes the average temperature (?C) per sample location.",
sep="<br/>"))
  }
  else if(as.character(map.df$CharacteristicName[1])=="Turbidity"){
    HTML(paste("50 NTU is a frequent threshold used by states for regulation
classification and discharge limitations for both natural and anthropogenic
systems.", " ", "This map visualizes the frequency of turbidity measurements greater
than 50 NTU per sample location.", sep="<br/>"))
  }
  else if(as.character(map.df$CharacteristicName[1])=="Specific conductance"){
    HTML(paste("The USGS considers the South Platte River to be significantly
affected by anthropogenic activates when specific conductivity measurements are less
than 400 or greater than 1350 uS/cm.", " ", "This map visualizes the frequency of

```



```

specific conductivity measurements per sample location outside USGS recommendations.",
sep="<br/>"))
}
else{
  HTML(paste("The Colorado Department of Public Health and Environment Water
Quality Control Division regulations state that domestic water supply and recreational
waters should have dissolved oxygen concentration greater than 3 mg/L to support fish
populations.", " ", "This map visualizes the frequency of dissolved oxygen measurements
per sample location outside CRWCD recommendations.", sep="<br/>"))
}
}) # end right of map text

# below of map text
output$belowMap <- renderText({
  map.df <- map.dat()
  if(as.character(map.df$CharacteristicName[1])=="pH"){
    paste("pH is a negative logarithmic measurement of hydrogen ion concentration in
water. Even minor changes in pH alter the states of many chemicals and thereby change
their solubility, transport, and bioavailability. pH affects community composition
and biological processes, such as reproduction, growth, disease, and death, because it
is a physiological stressor.")
  }
  else if(as.character(map.df$CharacteristicName[1])=="Temperature, water"){
    paste("Temperature is the concentration of thermal energy in water. The thermal
regime of a water system is influenced by atmospheric and hydraulic processes and the
structure and function of aquatic systems. Because temperature is frequently altered
by anthropogenic processes, effluent discharges are commonly regulated.")
  }
  else if(as.character(map.df$CharacteristicName[1])=="Turbidity"){
    paste("Turbidity is the measurement of relative clarity of water. It is a proxy
measurement for the amount of sediment in water, such as dissolved organic matter,
algae, plankton, clay, and silt. Highly turbid waters can represent health and
cosmetic concerns for aquatic communities and drinking water. Turbidity limits are
common regulations for effluent discharges.")
  }
  else if(as.character(map.df$CharacteristicName[1])=="Specific conductance"){
    paste("Specific conductance is a measure of ionic strength, the concentration of
ionic charge in water, and is highly correlated with salinity and dissolve oxygen
content. Specific conductance is electrical conductance normalized to a standard 25
°C temperature. Ionic strength relates to a large array of potential freshwater
problems because aquatic organisms generally prefer waters with specific ions and a
specific ionic strength range. Changes in specific conductance measurements can
affect community composition and require changes water treatment processes.")
  }
  else{
    paste("Dissolved oxygen refers to the concentration of gaseous oxygen
incorporated into water. Dissolved oxygen increases with turbulence and decreases
with temperature. Dissolved oxygen is essential for aerobic aquatic life growth and
reproduction.")
  }
}) # end below of map text

# reactivity function
locID <- eventReactive(input$getdata,{
  input$map_marker_click$id # get "id" from the clicked marker
})

# draw dygraph with xts with reactivity
output$chart <- renderDygraph({
  if(!is.null(locID())){
    map.df <- map.dat()
    id <- locID()
    if(as.character(map.df$CharacteristicName[1])=="pH"){

```

```

my.xts <- get(paste0("my.xts_pH_", id))
dygraph(my.xts, main = paste0("pH measurements at ", id)) %>%
  dySeries("In_Comp", strokeWidth = 0, color = "blue", drawPoints = TRUE,
pointSize = 3) %>%
  dySeries("Out_of_Comp", strokeWidth = 0, color="red", drawPoints = TRUE,
pointSize = 3) %>%
  dyAxis("y", label = "pH", valueRange = c(1, 14)) %>%
  dyRangeSelector() %>%
  dyLimit(6.5, color = "red") %>%
  dyLimit(9, color = "red") %>%
  dyShading(from = 0, to = 6.5, axis = "y", color = "#FFFFCC") %>%
  dyShading(from = 9, to = 14, axis = "y", color = "#FFFFCC")
} # end if pH
else if(as.character(map.df$CharacteristicName[1])=="Temperature, water"){
my.xts <- get(paste0("my.xts_temp_", id))
dygraph(my.xts, main = paste0("Temperature measurements at ", id)) %>%
  dySeries("In_Comp", strokeWidth = 0, color = "blue", drawPoints = TRUE,
pointSize = 3) %>%
  dySeries("Out_of_Comp", strokeWidth = 0, color="red", drawPoints = TRUE,
pointSize = 3) %>%
  dyAxis("y", label = "Temperature, C", valueRange = c(-5, 40)) %>%
  dyRangeSelector() %>%
  dyLimit(20, color = "blue") %>%
  dyLimit(30, color = "red") %>%
  dyShading(from = -5, to = 20, axis = "y", color = "#CCFBFF") %>%
  dyShading(from = 20, to = 30, axis = "y", color = "#FFFFCC")
} # end if temp
else if(as.character(map.df$CharacteristicName[1])=="Turbidity"){
my.xts <- get(paste0("my.xts_turb_", id))
dygraph(my.xts, main = paste0("Turbidity measurements at ", id)) %>%
  dySeries("In_Comp", strokeWidth = 0, color = "blue", drawPoints = TRUE,
pointSize = 3) %>%
  dySeries("Out_of_Comp", strokeWidth = 0, color="red", drawPoints = TRUE,
pointSize = 3) %>%
  dyAxis("y", label = "Turbidity, NTU", valueRange = c(0, 250)) %>%
  dyRangeSelector() %>%
  dyLimit(50, color = "red") %>%
  dyShading(from = 50, to = 250, axis = "y", color = "#FFFFCC")
} # end if turb
else if(as.character(map.df$CharacteristicName[1])=="Specific conductance"){
my.xts <- get(paste0("my.xts_spcn_", id))
dygraph(my.xts, main = paste0("Specific Conductance measurements at ", id))
%>%
  dySeries("In_Comp", strokeWidth = 0, color = "blue", drawPoints = TRUE,
pointSize = 3) %>%
  dySeries("Out_of_Comp", strokeWidth = 0, color="red", drawPoints = TRUE,
pointSize = 3) %>%
  dyAxis("y", label = "Specific Conductance", valueRange = c(0, 10000)) %>%
  dyRangeSelector() %>%
  dyLimit(400, color = "red") %>%
  dyLimit(1350, color = "red") %>%
  dyShading(from = 0, to = 400, axis = "y", color = "#FFFFCC") %>%
  dyShading(from = 1350, to = 10000, axis = "y", color = "#FFFFCC")
} # end if spcn
else{
my.xts <- get(paste0("my.xts_DO_", id))
dygraph(my.xts, main = paste0("DO measurements at ", id)) %>%
  dySeries("In_Comp", strokeWidth = 0, color = "blue", drawPoints = TRUE,
pointSize = 3) %>%
  dySeries("Out_of_Comp", strokeWidth = 0, color="red", drawPoints = TRUE,
pointSize = 3) %>%
  dyAxis("y", label = "Dissolved Oxygen", valueRange = c(0, 20)) %>%
  dyRangeSelector() %>%

```

```

        dyLimit(3, color = "red") %>%
        dyShading(from = 0, to = 3, axis = "y", color = "#FFFFCC")
    } # end else DO
  } # end if !is.null
}) # end dygraph

# right of result text
output$rightResult <- renderUI({
  map.df <- map.dat()
  if(as.character(map.df$CharacteristicName[1])=="pH"){
    HTML(paste("This scatterplot illustrates pH result measurements per the selected
sample location. Measurements within EPA freshwater criteria suggestions are colored
blue; measurements not within EPA freshwater criteria suggestions are colored red.", "
", "Mouse over point to identify measurement date and exact result measurement value.
Use the range selector below graph to adjust date range.", sep="<br/>"))
  }
  else if(as.character(map.df$CharacteristicName[1])=="Temperature, water"){
    HTML(paste("This scatterplot illustrates temperature results measurements (?C)
per the selected sample location. Colorado Water Quality Control Division classifies
cold water as having a temperature less than 20 ?C and warm water as having a
temperature between 20 and 30 ?C. Although water temperature typically fluctuates
seasonally, rapid temperature fluxes indicate potentially detrimental water
quality.", " ", "Mouse over point to identify measurement date and exact result
measurement value. Use the range selector below graph to adjust date range.",
sep="<br/>"))
  }
  else if(as.character(map.df$CharacteristicName[1])=="Turbidity"){
    HTML(paste("This scatterplot illustrates turbidity result measurements (NTU) per
the selected sample location. Measurements less than the 50 NTU thresholds are
colored blue; measurements greater than the 50 NTU threshold are colored red.", "
", "Mouse over point to identify measurement date and exact result measurement value.
Use the range selector below graph to adjust date range.", sep="<br/>"))
  }
  else if(as.character(map.df$CharacteristicName[1])=="Specific conductance"){
    HTML(paste("This scatterplot illustrates specific conductance result
measurements (uS/cm) per the selected sample location. Measurements within USGS non-
significantly affected criteria are colored blue; measurements not within USGS non-
significantly affected criteria are colored red.", " ", "Mouse over point to identify
measurement date and exact result measurement value. Use the range selector below
graph to adjust date range.", sep="<br/>"))
  }
  else{
    HTML(paste("This scatterplot illustrates dissolved oxygen result measurements
(mg/L) per the selected sample location. Measurements within CRWCD recommendations
are colored blue; measurements not within CRWCD recommendations are colored red.", "
", "Mouse over point to identify measurement date and exact result measurement value.
Use the range selector below graph to adjust date range.", sep="<br/>"))
  }
}) # end right of result text

# draw trend graph --- with reactivity
output$plt <- renderPlot({
  if(!is.null(locID())){
    map.df <- map.dat()
    id <- locID()
    if(as.character(map.df$CharacteristicName[1])=="pH"){
      bardat <- get(paste0("bardat_pH_", id))
      Month <- as.factor(bardat[,9])
      ggplot(bardat, aes(year.x, avediff, fill=Month)) +
        geom_bar(stat="identity", position = "dodge") +
        ggtitle("Trends; Deviation From Average Result Measures") +
        xlab("Year") +
        ylab("Deviation From Average of Monthly Result Measure")
    }
  }
})

```

```

} # end pH loop
else if(as.character(map.df$CharacteristicName[1])=="Temperature, water"){
  bardat <- get(paste0("bardat_temp_", id))
  Month <- as.factor(bardat[,9])
  ggplot(bardat, aes(year.x, avediff, fill=Month)) +
    geom_bar(stat="identity", position = "dodge") +
    ggtitle("Trends; Deviation From Average Result Measures") +
    xlab("Year") +
    ylab("Deviation From Average of Monthly Result Measure")
} # end temp loop
else if(as.character(map.df$CharacteristicName[1])=="Turbidity"){
  bardat <- get(paste0("bardat_turb_", id))
  Month <- as.factor(bardat[,9])
  ggplot(bardat, aes(year.x, avediff, fill=Month)) +
    geom_bar(stat="identity", position = "dodge") +
    ggtitle("Trends; Deviation From Average Result Measures") +
    xlab("Year") +
    ylab("Deviation From Average of Monthly Result Measure")
} # end turb loop
else if(as.character(map.df$CharacteristicName[1])=="Specific conductance"){
  bardat <- get(paste0("bardat_spcn_", id))
  Month <- as.factor(bardat[,9])
  ggplot(bardat, aes(year.x, avediff, fill=Month)) +
    geom_bar(stat="identity", position = "dodge") +
    ggtitle("Trends; Deviation From Average Result Measures") +
    xlab("Year") +
    ylab("Deviation From Average of Monthly Result Measure")
} # end spcn loop
else{
  bardat <- get(paste0("bardat_DO_", id))
  Month <- as.factor(bardat[,9])
  ggplot(bardat, aes(year.x, avediff, fill=Month)) +
    geom_bar(stat="identity", position = "dodge") +
    ggtitle("Trends; Deviation From Average Result Measures") +
    xlab("Year") +
    ylab("Deviation From Average of Monthly Result Measure")
} # end DO loop
} # end if loop
}) # end trend graph
} # end server

```

## REFERENCES

- <sup>1</sup> "State of Emergency Declared in the City of Flint." December 14, 2015. Accessed April 19, 2016. <https://www.cityofflint.com/state-of-emergency/>.
- <sup>2</sup> Southall, Ashley. "State of Emergency Declared Over Man-Made Water Disaster in Michigan City." *The New York Times*, January 17, 2016. Accessed April 19, 2016. [http://www.nytimes.com/2016/01/18/us/obama-flint-michigan-water-fema-emergency-disaster.html?\\_r=0](http://www.nytimes.com/2016/01/18/us/obama-flint-michigan-water-fema-emergency-disaster.html?_r=0).
- <sup>3</sup> Smith, Lindsey. "After ignoring and trying to discredit people in Flint, the state was forced to face the problem." *Michigan Radio*, December 16, 2015. Accessed April 19, 2016. <http://michiganradio.org/post/after-ignoring-and-trying-discredit-people-flint-state-was-forced-face-problem#stream/0>.
- <sup>4</sup> Fonger, Ron. "Flint data on lead water lines stored on 45,000 index cards." *MLive*, October 1, 2015. Accessed April 19, 2016. [http://www.mlive.com/news/flint/index.ssf/2015/10/flint\\_official\\_says\\_data\\_on\\_lo.html](http://www.mlive.com/news/flint/index.ssf/2015/10/flint_official_says_data_on_lo.html).
- <sup>5</sup> Brush, Mark. "Expert says Michigan officials changed a Flint lead report to avoid federal action." *Michigan Radio*, November 5, 2015. Accessed April 19, 2016. <http://michiganradio.org/post/expert-says-michigan-officials-changed-flint-lead-report-avoid-federal-action#stream/0>.
- <sup>6</sup> Bosman, Julie, Monica Davey, and Mitch Smith. "As Water Problems Grew, Officials Belittled Complaints From Flint." *The New York Times*, January 20, 2016. Accessed April 19, 2016. [http://www.nytimes.com/2016/01/21/us/flint-michigan-lead-water-crisis.html?\\_r=0](http://www.nytimes.com/2016/01/21/us/flint-michigan-lead-water-crisis.html?_r=0).
- <sup>7</sup> "Beyond Flint: In The South, Another Water Crisis Has Been Unfolding For Years." *National Public Radio, North Carolina Public Radio*, February 6, 2016. Accessed April 19, 2016. [http://www.npr.org/2016/02/06/465702398/beyond-flint-in-the-south-another-water-crisis-has-been-unfolding-for-years?utm\\_source=facebook.com&utm\\_medium=social&utm\\_campaign=npr&utm\\_term=nprnews&utm\\_content=20160206](http://www.npr.org/2016/02/06/465702398/beyond-flint-in-the-south-another-water-crisis-has-been-unfolding-for-years?utm_source=facebook.com&utm_medium=social&utm_campaign=npr&utm_term=nprnews&utm_content=20160206).
- <sup>8</sup> "2013 Report Card For America's Infrastructure; Drinking Water." *American Society of Civil Engineers*, 2013. Accessed April 19, 2016. <http://www.infrastructurereportcard.org/a/documents/Drinking-Water.pdf>.
- <sup>9</sup> "Data Intelligence for 21<sup>st</sup> Century Water Management: A Report From the 2015 Aspen-Nicholas Water Forum." *The Aspen Institute and The Nicholas Institute for Environmental Policy Solutions*, 2015. Accessed April 19, 2016. [https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report\\_1.pdf](https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report_1.pdf).
- <sup>10</sup> "Data Intelligence for 21<sup>st</sup> Century Water Management: A Report From the 2015 Aspen-Nicholas Water Forum." *The Aspen Institute and The Nicholas Institute for Environmental Policy Solutions*, 2015. Accessed April 19, 2016. [https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report\\_1.pdf](https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report_1.pdf).
- <sup>11</sup> Marr, Bernard. *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. Chichester: John Wiley & Sons, 2015.
- <sup>12</sup> van der Gaag, Bram, Jurgen Volz. "Real-time on-line monitoring of contaminants in water: developing a research strategy from utility experiences and needs." *Water Research Foundation*, May 16, 2008. Accessed April 19, 2016. <http://www.waterrf.org/publicreportlibrary/4025.pdf>.
- <sup>13</sup> Vance, Tiffany, Nazila Merati, Chaowei Yang, May Yuan. *Cloud Computing in Ocean and Atmospheric Sciences*. London: Elsevier Inc., 2016.
- <sup>14</sup> Bell, Barbara. "The Oldest Records of the Nile Floods." *The Geographical Journal* 136, no. 4 (December 1970): 569-73.
- <sup>15</sup> Northern Colorado Water Conservancy District. "C-BT Project Water Quota Increased to 70 Percent." *Northern Colorado Water Conservancy District News Release*, April 14, 2016. Accessed April 19, 2016. <http://www.northernwater.org/docs/Home%20Page/WhatsNewFinalPDFs/2016%20C-BT%20Quota%20News%20Release.pdf>.
- <sup>16</sup> "A Primer on Water Quality." *US Geological Survey*, March 2001. Accessed April 19, 2016. <http://pubs.usgs.gov/fs/fs-027-01/pdf/FS-027-01.pdf>.
- <sup>17</sup> Hornberger, George, Jeffrey Raffensperger, Patricia Wiberg, Keith Eshleman. *Elements of Physical Hydrology*. Baltimore: The Johns Hopkins University Press. 1998.
- <sup>18</sup> Myers, Donna. "History of Water Quality Monitoring and Assessment in the United States." *American Water Resources Association*, October 6, 2015. [https://members.awra.org/AWRA/Publications/AWRA\\_Webinars.aspx](https://members.awra.org/AWRA/Publications/AWRA_Webinars.aspx).
- <sup>19</sup> "USGS Water-Quality Data for the Nation." *USGS National Water Information System: Web Interface*. Accessed April 19, 2016. <http://waterdata.usgs.gov/nwis/qw>.

- 
- <sup>20</sup> "Open Water Data Initiative Proposal and Charge for Discussion." *Advisory Committee on Water Information*. Accessed April 19, 2016. [http://acwi.gov/spatial/open\\_water\\_data\\_charge\\_to\\_fgdc\\_june23\\_2014.pdf](http://acwi.gov/spatial/open_water_data_charge_to_fgdc_june23_2014.pdf).
- <sup>21</sup> "USGS Surface-Water Data for the Nation." *USGS National Water Information System: Web Interface*. Accessed April 19, 2016. <http://waterdata.usgs.gov/nwis/sw>.
- <sup>22</sup> "Continuous Monitoring Data Sharing Strategy." *Michael Baker International, LimnoTech, and MapTech, Inc.*, September 21, 2015. EPA Contract EP-C-12-052 Order No. 0005.
- <sup>23</sup> Goodall, Jonathan, Jeffery Horsburgh, Timothy Whiteaker, David Maidment, Ilya Zaslavsky. "A first approach to web services for the National Water Information System." *Environmental Modelling & Software*, 23, no. 4 (April 2008): 404-411.
- <sup>24</sup> "Open Water Data Initiative Overview." *Advisory Committee on Water Information*. Accessed April 19, 2016. <http://acwi.gov/spatial/owdi/>.
- <sup>25</sup> Goodall, Jonathan, Jeffery Horsburgh, Timothy Whiteaker, David Maidment, Ilya Zaslavsky. "A first approach to web services for the National Water Information System." *Environmental Modelling & Software*, 23, no. 4 (April 2008): 404-411.
- <sup>26</sup> Larsen, Sara, Stuart Hamilton, Jessica Lucido, Bradley Garner, and Dwane Young. "Supporting Diverse Data Providers in the Open Water Data Initiative: Communicating Water Data Quality and Fitness of Use." *Journal of the American Water Resources Association (JAWRA)*, (2016): 1-14.
- <sup>27</sup> "Open Water Data Initiative Overview." *Advisory Committee on Water Information*. Accessed April 19, 2016. <http://acwi.gov/spatial/owdi/>.
- <sup>28</sup> "Open Water Data Initiative Proposal and Charge for Discussion." *Advisory Committee on Water Information*. Accessed April 19, 2016. [http://acwi.gov/spatial/open\\_water\\_data\\_charge\\_to\\_fgdc\\_june23\\_2014.pdf](http://acwi.gov/spatial/open_water_data_charge_to_fgdc_june23_2014.pdf).
- <sup>29</sup> "Integrated Water Resources Science and Services (IWRSS). *US Army Corps of Engineers*. Accessed April 19, 2016. <http://www.hec.usace.army.mil/misc/IWRSS/>.
- <sup>30</sup> "Information Sheet: Integrated Water Resources Science and Services (IWRSS)" *National Weather Service; National Oceanic and Atmospheric Association*. Accessed April 19, 2016. [http://www.nws.noaa.gov/oh/nwc/IWRSS%20Information%20Sheet\\_Nov%205%20%202012%20PDF.pdf](http://www.nws.noaa.gov/oh/nwc/IWRSS%20Information%20Sheet_Nov%205%20%202012%20PDF.pdf).
- <sup>31</sup> "Open Water Data Initiative Overview." *Advisory Committee on Water Information*. Accessed April 19, 2016. <http://acwi.gov/spatial/owdi/>.
- <sup>32</sup> "OGC WaterML 2.0: Part 1- Timeseries." *Open Geospatial Consortium*, January 16, 2012. OGC 10-126r2. Accessed April 19, 2016. [http://www.waterml2.org/10-126r2\\_10-126r2\\_WaterML\\_2.0\\_Candidate\\_Standard.pdf](http://www.waterml2.org/10-126r2_10-126r2_WaterML_2.0_Candidate_Standard.pdf).
- <sup>33</sup> Larsen, Sara, Stuart Hamilton, Jessica Lucido, Bradley Garner, and Dwane Young. "Supporting Diverse Data Providers in the Open Water Data Initiative: Communicating Water Data Quality and Fitness of Use." *Journal of the American Water Resources Association (JAWRA)*, (2016): 1-14.
- <sup>34</sup> "Continuous Monitoring Data Sharing Strategy." *Michael Baker International, LimnoTech, and MapTech, Inc.*, September 21, 2015. EPA Contract EP-C-12-052 Order No. 0005.
- <sup>35</sup> "National Ground-Water Monitoring Network." *Advisory Committee on Water Information*. Accessed April 19, 2016. <http://cida.usgs.gov/ngwmn/>.
- <sup>36</sup> "A National Framework for Ground-Water Monitoring in the United States." *The Subcommittee on Ground Water of the Advisory Committee on Water Information*. July 2013. Accessed April 19, 2016. [http://acwi.gov/sogw/ngwmn\\_framework\\_report\\_july2013.pdf](http://acwi.gov/sogw/ngwmn_framework_report_july2013.pdf).
- <sup>37</sup> "National Flood Interoperability Experiment. National Weather Service. Accessed April 19, 2016. <http://www.caee.utexas.edu/prof/maidment/giswr2014/Synopsis/GISWRSynopsis12.pdf>.
- <sup>38</sup> Clark, Edward. "National Flood Interoperability Experiment: Leveraging National Hydrography Dataset to explore the future of flood forecasting." *US Geological Survey National Hydrography Dataset*, May 21, 2015. Accessed April 19, 2016. [http://nhd.usgs.gov/documents/NFIEUseCase\\_UnderlyingTech.pdf](http://nhd.usgs.gov/documents/NFIEUseCase_UnderlyingTech.pdf).
- <sup>39</sup> "About WaDE." *Western States Water Council*. Accessed April 19, 2016. <http://www.westernstateswater.org/wade/about-wade/>.
- <sup>40</sup> "WaterSage." *WaterSage*. Accessed April 19, 2016. <http://www.watersage.com/>.
- <sup>41</sup> "Kisters." *Kisters*. Accessed April 19, 2016. <http://www.kisters.net/>.
- <sup>42</sup> "Aquatic Informatics." *Aquatic Informatics*. Accessed April 19, 2016. <http://aquaticinformatics.com/>.
- <sup>43</sup> "Concourse." *Concourse DB*. Accessed April 19, 2016. <http://concoursedb.com/>.
- <sup>44</sup> "Data Forensics." *Data Forensics*. Accessed April 19, 2016. <http://www.dataforensics.net/>.

- 
- <sup>45</sup> “Accelerated Technology Laboratories, Inc.” *Accelerated Technology laboratories, Inc.* Accessed April 19, 2016. <http://www.atlab.com/>.
- <sup>46</sup> “Promium.” *Promium*. Accessed April 19, 2016. <https://promium.com/>.
- <sup>47</sup> Larsen, Sara, Stuart Hamilton, Jessica Lucido, Bradley Garner, and Dwane Young. “Supporting Diverse Data Providers in the Open Water Data Initiative: Communicating Water Data Quality and Fitness of Use.” *Journal of the American Water Resources Association (JAWRA)*, (2016): 1-14.
- <sup>48</sup> “A Primer on Water Quality.” *US Geological Survey*, March 2001. Accessed April 19, 2016. <http://pubs.usgs.gov/fs/fs-027-01/pdf/FS-027-01.pdf>.
- <sup>49</sup> “USGS Water-Quality Data for the Nation.” *USGS National Water Information System: Web Interface*. Accessed April 19, 2016. <http://waterdata.usgs.gov/nwis/qw>.
- <sup>50</sup> “Community Surface Water (& GWUDI) Systems – Monitoring Requirements.” *North Carolina Department of Environmental Quality; Water Division*. Accessed April 19, 2016. <http://www.ncwater.org/files/pws/compliance/CWSSurfaceSamplingCharts.pdf>.
- <sup>51</sup> Hedgpeth, Dana. “10,000 gallons of latex spills into part of Potomac River in western Maryland.” *The Washington Post*, September 25, 2015. Accessed April 19, 2016. [https://www.washingtonpost.com/national/health-science/10000-gallons-of-latex-spills-into-part-of-potomac-river-in-western-maryland/2015/09/25/2059c1e0-6394-11e5-9757-e49273f05f65\\_story.html](https://www.washingtonpost.com/national/health-science/10000-gallons-of-latex-spills-into-part-of-potomac-river-in-western-maryland/2015/09/25/2059c1e0-6394-11e5-9757-e49273f05f65_story.html).
- <sup>52</sup> Myers, Donna. “History of Water Quality Monitoring and Assessment in the United States.” *American Water Resources Association*, October 6, 2015. [https://members.awra.org/AWRA/Publications/AWRA\\_Webinars.aspx](https://members.awra.org/AWRA/Publications/AWRA_Webinars.aspx).
- <sup>53</sup> Myers, Donna. “History of Water Quality Monitoring and Assessment in the United States.” *American Water Resources Association*, October 6, 2015. [https://members.awra.org/AWRA/Publications/AWRA\\_Webinars.aspx](https://members.awra.org/AWRA/Publications/AWRA_Webinars.aspx).
- <sup>54</sup> Safe Drinking Water Act (SDWA) § 42 U.S.C. 300f–300j–9.
- <sup>55</sup> “Understanding the Safe Drinking Water Act.” *US Environmental Protection Agency Office of Water*, June 2004. EPA 816-F04-030. Accessed April 19, 2016. <https://www.epa.gov/sites/production/files/2015-04/documents/epa816f04030.pdf>.
- <sup>56</sup> “Information about Public Water Systems.” *US Environmental Protection Agency*, December 3 2015. Accessed April 19, 2016. <https://www.epa.gov/dwreginfo/information-about-public-water-systems>.
- <sup>57</sup> Safe Drinking Water Act (SDWA) § 42 U.S.C. 300f–300j–9.
- <sup>58</sup> “Table of Regulated Drinking Water Contaminants.” *US Environmental Protection Agency*, February 18, 2016. Accessed April 19, 2016. <https://www.epa.gov/your-drinking-water/table-regulated-drinking-water-contaminants>.
- <sup>59</sup> Safe Drinking Water Act (SDWA) § 42 U.S.C. 300f–300j–9.
- <sup>60</sup> “Secondary Drinking Water Standards: Guidance for Nuisance Chemicals.” *US Environmental Protection Agency*, January 6, 2016. Accessed April 19, 2016. <https://www.epa.gov/dwstandardsregulations/secondary-drinking-water-standards-guidance-nuisance-chemicals>.
- <sup>61</sup> “Safe Drinking Water Act (SDWA): A Summary of the Act and Its Major Requirements.” *Federation of American Scientists*, February 5, 2014. Accessed April 19, 2015. <https://www.fas.org/sgp/crs/misc/RL31243.pdf>.
- <sup>62</sup> Safe Drinking Water Act (SDWA) § 42 U.S.C. 300f–300j–9.
- <sup>63</sup> Krenkel, Peter and Vladimir Novotny. *Water Quality Management*. New York: Academic Press, Inc. 2000.
- <sup>64</sup> “Table of Regulated Drinking Water Contaminants.” *US Environmental Protection Agency*, February 18, 2016. Accessed April 19, 2016. <https://www.epa.gov/your-drinking-water/table-regulated-drinking-water-contaminants>.
- <sup>65</sup> “Community Surface Water (& GWUDI) Systems – Monitoring Requirements.” *North Carolina Department of Environmental Quality; Water Division*. Accessed April 19, 2016. <http://www.ncwater.org/files/pws/compliance/CWSSurfaceSamplingCharts.pdf>.
- <sup>66</sup> Safe Drinking Water Act (SDWA) § 42 U.S.C. 300f–300j–9.
- <sup>67</sup> “Understanding the Safe Drinking Water Act.” *US Environmental Protection Agency Office of Water*, June 2004. EPA 816-F04-030. Accessed April 19, 2016. <https://www.epa.gov/sites/production/files/2015-04/documents/epa816f04030.pdf>.
- <sup>68</sup> “Drinking Water Watch.” *Safe Drinking Water Information System*. Accessed April 19, 2016. <https://www.pwss.enr.state.nc.us/NCDWW2/>.
- <sup>69</sup> “2014 Treated Water Quality Summary.” *Denver Water*, 2014. Accessed April 19, 2016. <http://www.denverwater.org/docs/assets/FCB8A422-155D-01CB-0C315C01BE75096C/TreatedWaterSummaryReport2014.pdf>.

- 
- <sup>70</sup> The Freedom of Information Act (FOIA), 5 U.S.C. § 552.
- <sup>71</sup> "Summary of the Clean Water Act." *US Environmental Protection Agency*, October 8, 2015. Accessed April 19, 2016. <http://www2.epa.gov/laws-regulations/summary-clean-water-act>.
- <sup>72</sup> Clean Water Act § 402(b), 33 U.S.C. § 1342(a)
- <sup>73</sup> Walton, Brett. "EPA Clean Water Rule Meets Political Pushback." *Circle of Blue*, August 6, 2015. Accessed April 19, 2016. <http://www.circleofblue.org/2015/world/epa-clean-water-rule-meets-political-pushback/>.
- <sup>74</sup> Clean Water Act § 404(b)(1), 33 U.S.C. § 1342(a)
- <sup>75</sup> Davison, Steven G. "General Permits Under Section 404 of the Clean Water Act." *Pace Environmental Law Review* 26, no. 1 (2009). Accessed April 19, 2016. <http://digitalcommons.pace.edu/cgi/viewcontent.cgi?article=1015&context=pehr>.
- <sup>76</sup> Gaba, Jeffrey M. "Generally Illegal: NPDES General Permits Under the Clean Water Act." *Harvard Environmental Law Review* 31 (2007). Accessed April 19, 2016. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1555771](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1555771).
- <sup>77</sup> "National Pollutant Discharge Elimination System (NPDES)." *US Environmental Protection Agency*, April 11, 2016. Accessed April 19, 2016. <https://www.epa.gov/npdes>.
- <sup>78</sup> "Data Intelligence for 21<sup>st</sup> Century Water Management: A Report From the 2015 Aspen-Nicholas Water Forum." *The Aspen Institute and The Nicholas Institute for Environmental Policy Solutions*, 2015. Accessed April 19, 2016. [https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report\\_1.pdf](https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report_1.pdf).
- <sup>79</sup> Annex, Patricia. "Urban Infrastructure Finance From Private Operators: What Have We Learned From Experience." *World Bank*, November 2006. WPS 4045. Accessed April 19, 2016. [https://books.google.com/books?hl=en&lr=&id=PedTAGlgflwC&oi=fnd&pg=PA2&dq=water+infrastructure+finance&ots=\\_fBuKRvfrj&sig=JvYnvvAapC2c60NBeyfgjKDDWsM#v=onepage&q=water%20infrastructure%20finance&f=false](https://books.google.com/books?hl=en&lr=&id=PedTAGlgflwC&oi=fnd&pg=PA2&dq=water+infrastructure+finance&ots=_fBuKRvfrj&sig=JvYnvvAapC2c60NBeyfgjKDDWsM#v=onepage&q=water%20infrastructure%20finance&f=false).
- <sup>80</sup> Larsen, Sara, Stuart Hamilton, Jessica Lucido, Bradley Garner, and Dwane Young. "Supporting Diverse Data Providers in the Open Water Data Initiative: Communicating Water Data Quality and Fitness of Use." *Journal of the American Water Resources Association (JAWRA)*, (2016): 1-14.
- <sup>81</sup> "USGS Water-Quality Data for the Nation." *USGS National Water Information System: Web Interface*. Accessed April 19, 2016. <http://waterdata.usgs.gov/nwis/qw>.
- <sup>82</sup> "Data Intelligence for 21<sup>st</sup> Century Water Management: A Report From the 2015 Aspen-Nicholas Water Forum." The Aspen Institute and The Nicholas Institute for Environmental Policy Solutions, 2015. Accessed April 19, 2016. [https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report\\_1.pdf](https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report_1.pdf).
- <sup>83</sup> Dickinson, Janis, Jennifer Shirk, David Bonter, Rick Bonney, Rhiannon Crain, Jason Martin, Tina Phillips, Karen Purcell. "The current state of citizen science as a tool for ecological research and public engagement." *The Ecological Society of America; Frontiers in Ecology and the Environment*, (2012), 10: 291-297.
- <sup>84</sup> Gustetic, Jenn, Le Shanley, Jay Benforado, Arianne Miller. "Designing a Citizen Science and Crowdsourcing Toolkit for the Federal Government." The White House, December 2, 2014. Accessed April 19, 2016. <https://www.whitehouse.gov/blog/2014/12/02/designing-citizen-science-and-crowdsourcing-toolkit-federal-government>
- <sup>85</sup> The GLOBE Program. Accessed April 19, 2016. <http://www.globe.gov/>.
- <sup>86</sup> River Watch of Colorado. Accessed April 19, 2016. <http://www.coloradoriverwatch.org/>.
- <sup>87</sup> mPING; crowdsourcing weather reports. Accessed April 19, 2016. <http://mping.nssl.noaa.gov/>.
- <sup>88</sup> "IBM Intelligent Water." *IBM*. Accessed April 19, 2016. <http://www-03.ibm.com/software/products/en/intelligentwater>
- <sup>89</sup> "AMI for the water industry: applying lessons from electric industry deployment." *West Monroe*, 2015. Accessed April 19, 2016. <http://www.westmonroepartners.com/Insights/White-Papers/AMI-for-the-Water-Industry>.
- <sup>90</sup> "Open Water Data Initiative Proposal and Charge for Discussion." *Advisory Committee on Water Information*. Accessed April 19, 2016. [http://acwi.gov/spatial/open\\_water\\_data\\_charge\\_to\\_fgdc\\_june23\\_2014.pdf](http://acwi.gov/spatial/open_water_data_charge_to_fgdc_june23_2014.pdf).
- <sup>91</sup> Speight, Vanessa. "Innovation in the water industry: barriers and opportunities for US and UK utilities." *Wiley Interdisciplinary Reviews: Water*, (March 11, 2015), 2-4: 301-313.
- <sup>92</sup> Larsen, Sara, Stuart Hamilton, Jessica Lucido, Bradley Garner, and Dwane Young. "Supporting Diverse Data Providers in the Open Water Data Initiative: Communicating Water Data Quality and Fitness of Use." *Journal of the American Water Resources Association (JAWRA)*, (2016): 1-14.
- <sup>93</sup> "Data Intelligence for 21<sup>st</sup> Century Water Management: A Report From the 2015 Aspen-Nicholas Water Forum." *The Aspen Institute and The Nicholas Institute for Environmental Policy Solutions*, 2015. Accessed April 19, 2016. [https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report\\_1.pdf](https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report_1.pdf).



- 
- <sup>94</sup> "The Global Risks Report 2016." World Economic Forum, 2016. <http://www3.weforum.org/docs/Media/TheGlobalRisksReport2016.pdf>.
- <sup>95</sup> "Water Scarcity & Climate Change: Growing Risks for Business and Investors." *Ceres and Pacific Institute*, February 2009. Accessed April 19, 2016. <http://www.ceres.org/resources/reports/water-scarcity-climate-change-risks-for-investors-2009>
- <sup>96</sup> "Disclosure Framework for Water & Sewer Enterprises." *Ceres*, April 2013. Accessed April 19, 2016. <http://www.ceres.org/resources/reports/disclosure-framework-for-water-sewer-enterprises>
- <sup>97</sup> Kenny, Joan, Nancy Barber, Susan Hutson, Kristin Linsey, John Lovelace, Molly Maupin. "USGS Circular 1344: Estimated use of water in the United States in 2005," *US Geological Survey*, 2009.
- <sup>98</sup> "2009 Drinking Water Report Card," *American Society of Civil Engineers*, 2009. Accessed April 19, 2016. <http://www.infrastructurereportcard.org/fact-sheet/drinking-water#conditions>.
- <sup>99</sup> "2013 Report Card For America's Infrastructure; Drinking Water." *American Society of Civil Engineers*, 2013. Accessed April 19, 2016. <http://www.infrastructurereportcard.org/a/documents/Drinking-Water.pdf>.
- <sup>100</sup> "2009 Drinking Water Infrastructure Needs Survey and Assessment." *US Environmental Protection Agency Fourth Report to Congress*, 2009. EPA 816-R-09-001. Accessed April 19, 2016. [http://water.epa.gov/infrastructure/drinkingwater/dwns/upload/2009\\_03\\_26\\_needssurvey\\_2007\\_report\\_needssurvey\\_2007.pdf](http://water.epa.gov/infrastructure/drinkingwater/dwns/upload/2009_03_26_needssurvey_2007_report_needssurvey_2007.pdf)
- <sup>101</sup> "Water Quality, Supply, and Infrastructure Improvement Act of 2014; Funding Summary." *The California Department of Water Resources*, September 2014. Accessed April 19, 2016. [http://www.water.ca.gov/waterbond/Regional\\_Bond\\_Summary\\_2014\\_Final\\_web\\_2.pdf](http://www.water.ca.gov/waterbond/Regional_Bond_Summary_2014_Final_web_2.pdf).
- <sup>102</sup> "Market Statistics." *Electronic Municipal Market Access*. Accessed April 19, 2016. <http://emma.msrb.org/Home/Index>.
- <sup>103</sup> "Data Intelligence for 21<sup>st</sup> Century Water Management: A Report From the 2015 Aspen-Nicholas Water Forum." *The Aspen Institute and The Nicholas Institute for Environmental Policy Solutions*, 2015. Accessed April 19, 2016. [https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report\\_1.pdf](https://nicholasinstitute.duke.edu/sites/default/files/publications/2015-water-forum-report_1.pdf).
- <sup>104</sup> "Aqueduct Global Maps 2.1: Constructing Decision-Relevant global Water Risk Indicators." *World Resources Institute*, April 2005. Accessed April 19, 2016. <http://www.wri.org/publication/aqueduct-global-maps-21-indicators>.
- <sup>105</sup> "Aqueduct Global Maps 2.1." *World Resources Institute*, April 2015. Accessed April 19, 2016. [http://www.wri.org/sites/default/files/Aqueduct\\_Global\\_Maps\\_2.1.pdf](http://www.wri.org/sites/default/files/Aqueduct_Global_Maps_2.1.pdf).
- <sup>106</sup> "The Water Risk Valuation Tool (WRVT)." *Bloomberg LP*, September 2015. Accessed April 19, 2016. [http://www.bloomberg.com/bcause/content/uploads/sites/6/2015/09/Bloomberg\\_WRVT\\_09162015\\_WEB.pdf](http://www.bloomberg.com/bcause/content/uploads/sites/6/2015/09/Bloomberg_WRVT_09162015_WEB.pdf)
- <sup>107</sup> "The Ripple Effect: Water Risk in the Municipal Bond Market." *Ceres and Water Asset Management*, October 2010. Accessed April 19, 2016. <http://www.ceres.org/resources/reports/water-bonds>.
- <sup>108</sup> "The Ceres Aqua Gauge: A Framework for 21<sup>st</sup> Century Water Risk Management." *Ceres*. Accessed April 19, 2016. <http://www.ceres.org/resources/reports/aqua-gauge>.
- <sup>109</sup> "Water Sage." *Ponderosa Advisors, LLC*. Accessed April 19, 2016. <http://www.watersage.com/>.
- <sup>110</sup> "Water Sage: Water Rights in Lending." *Montana Bankers Association*. Accessed April 19, 2016. <http://www.montanabankers.com/?page=WaterSage>.
- <sup>111</sup> "Nicholas Institute for Environmental Policy Solutions." *Duke University*. Accessed April 19, 2016. <https://nicholasinstitute.duke.edu/>.
- <sup>112</sup> "Water Quality Portal." *National Water Quality Monitoring Council*. Accessed April 19, 2016. <http://waterqualitydata.us/>.
- <sup>113</sup> "Table of Regulated Drinking Water Contaminants." *US Environmental Protection Agency*, February 18, 2016. Accessed April 19, 2016. <https://www.epa.gov/your-drinking-water/table-regulated-drinking-water-contaminants>.
- <sup>114</sup> "WQP Web Services Guide." The Water Quality Portal. Accessed April 19, 2016. [http://waterqualitydata.us/webservices\\_documentation/](http://waterqualitydata.us/webservices_documentation/).
- <sup>115</sup> R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree". Copyright (C) 2015 The R Foundation for Statistical Computing.
- <sup>116</sup> "Dissolved Oxygen Standard Literature Review." *Colorado River Water Conservation District*, April 4, 2008.
- <sup>117</sup> "Temperature. Water Quality Standards Criteria Summaries: A Compilation of State/Federal Criteria." *US Environmental Protection Agency*, September 1988. EPA 440 5-88-023.

- 
- <sup>118</sup> “Turbidity. Water Quality Standards Criteria summaries: A Compilation of State/Federal Criteria.” *US Environmental Protection Agency*, September 1988. 440 5-88 013.
- <sup>119</sup> Water Quality in the South Platte River Basin.” *US Geological Survey*, 1998. Circular 1167.
- <sup>120</sup> Jeffrey A. Ryan and Joshua M. Ulrich (2014). xts: eXtensible Time Series. R package version 0.9-7. <https://CRAN.R-project.org/package=xts>.
- <sup>121</sup> Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2016). shiny: Web Application Framework for R. R package version 0.13.0. <https://CRAN.R-project.org/package=shiny>.
- <sup>122</sup> Winston Chang (2015). shinythemes: Themes for Shiny. R package version 1.0.1. <https://CRAN.R-project.org/package=shinythemes>.
- <sup>123</sup> Joe Cheng and Yihui Xie (2015). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 1.0.0. <https://CRAN.R-project.org/package=leaflet>.
- <sup>124</sup> Dan Vanderkam, JJ Allaire, Jonathan Owen, Daniel Gromer, Petr Shevtsov and Benoit Thieurmel (2016). dygraphs: Interface to 'Dygraphs' Interactive Time Series Charting Library. R package version 0.7. <https://CRAN.R-project.org/package=dygraphs>.
- <sup>125</sup> H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
- <sup>126</sup> “Community Surface Water (& GWUDI) Systems – Monitoring Requirements.” *North Carolina Department of Environmental Quality; Water Division*. Accessed April 19, 2016. <http://www.ncwater.org/files/pws/compliance/CWSSurfaceSamplingCharts.pdf>.