

Open Hydrology Courseware Using the United States Geological Survey's

National Water Census Data Portal

Jacob Daniel Nelson

A project submitted to the faculty of  
Brigham Young University  
in partial fulfillment of the requirements for the degree of

Master of Science

Daniel P. Ames, Chair  
E. James Nelson  
A. Woodruff Miller

Department of Civil and Environmental Engineering

Brigham Young University

Copyright © 2017 Jacob Daniel Nelson

All Rights Reserved

## ABSTRACT

### Open Hydrology Courseware Using the United States Geological Survey's National Water Census Data Portal

Jacob Daniel Nelson  
Department of Civil and Engineering, BYU  
Master of Science

The United States Geological Survey (USGS) is the primary U.S. government agency for water data collection and dissemination. In this role, the USGS has recently created and deployed a National Water Census Data Portal (NWC-DP) which provides access to streamflow, evapotranspiration, precipitation, aquatic biology and other data at the national level. Recognizing the value of these data sets for hydrologic science education, this paper presents an effort to bridge the gap between pencil and paper-based hydrology curriculum and the USGS NWC-DP resource. Specifically, we have developed an R package, National Water Census Education (NWCEd), and five associated laboratory exercises that integrate R and web services based access to the NWC-DP datasets. Using custom functions built into the NWCEd, students are able to access unprecedented amounts of hydrologic data from the NWC-DP which can be applied to current hydrology curriculum and analyzed using new R software.

Keywords: hydrology, education, software, R, National Water Census Data Portal

## ACKNOWLEDGEMENTS

I wish to acknowledge the United States Geological Survey (USGS) for funding and participation in the development of this project under award number 200251-00001-176 (G16AP00110). I wish to thank David Blodgett of the USGS and Dr. Dan Ames of Brigham Young University for their guidance and mentorship throughout the development of this project. I would especially like to thank Dr. Dan Ames for providing me the opportunity to advance my knowledge in principles of hydrology and programming through working on this project. I would like to thank Melanie Latham for her help in editing the laboratory exercises. I would also like to thank Shawn Crawley, Bryce Anderson, Zhiyu Li, Xiaohui Qiao, and Sarva Pulla, who work in the BYU Hydroinformatics lab, for helping me understand programming development through using GitHub and for their encouragement and support. I would also like to thank my family for their support and encouragement while completing this project.

## TABLE OF CONTENTS

Abstract.....	ii
Table of Contents.....	iv
List of Tables .....	vi
List of Figures.....	viii
1 Introduction.....	1
2 Methods.....	5
2.1 Documentation .....	5
2.2 Custom NWCEd R-package Functions.....	6
2.3 Laboratory Exercises Development .....	9
2.4 Laboratory Content .....	10
3 Results.....	13
4 Conclusion and Discussion.....	23
5 Software Availability .....	25
References.....	27
Appendix A.....	31

## LIST OF TABLES

Table 1: NWCEd works in concert with functionalities provided by external R packages. ....	8
Table 2: The content of the laboratory exercises associated with NWCEd was summarized with the associated NWCEd functions. ....	11

## LIST OF FIGURES

Figure 1: The getNWCdata retrieves hydrologic data from USGS servers and returns an RStudio data table to the user. ....	6
Figure 2: NWCEd, external R packages, and data from NWC-DP is downloaded into a local instance of RStudio. ....	9
Figure 3: Lab 2 instructs users how to access and run statistics on streamflow data from NWC-DP.....	14
Figure 4: Histogram bandwidths are compared in Lab 3.....	15
Figure 5: Lab 4 instructs users how to visualize NWC-DP time series data to improve the understanding of the water budget concept. ....	16
Figure 6: Double-mass curve analysis is visualized in Lab 4 using NWC-DP data.....	17
Figure 7: Code has been provided to reproduce density curve plots, including the manual adjustment of labels and percentile lines. ....	19
Figure 8: The code for a histogram with positive and negative bars is provided to introduce an alternative way to visualize water storage. ....	20
Figure 9: Code is provided to replicate interactive maps in the user's local instance of RStudio. ....	21

## 1 INTRODUCTION

As the hydrologic scientific community continues to adopt new technology, there becomes a need for educators to incorporate new technology and techniques into their curriculum (Thompson et al. 2012; Wagener et al. 2012). In hydroinformatics courses in particular, there is great potential to use online tools and data to aid the educational experience (Popescu et al. 2012). In comparison to pencil-and-paper-based laboratory assignments, data-driven laboratory exercises with interactive modeling and visualization components can improve learning and better help prepare students to enter the profession (Sanchez et al. 2016). Standalone interactive technology based tools can also reduce the need for in-classroom or instructor-led instruction, thereby increasing learning (Ruddell and Wagener 2013).

While there exists a plethora of new software which could be implemented in university classrooms, research has shown that many educators struggle or are adverse to amending their curriculum. Several roadblocks can contribute to this problem, including “a lack of access to easily adoptable curriculum materials and a lack of time and training to learn constantly changing tools and methods” and that implementing new technology “should emphasize conceptual learning, and should be used to complement rather than replace lecture-based pedagogies” (Merwade and Ruddell 2012). Indeed, in hydrologic research, reuseable code and data should always be made available so that scientific results can be reproduced (Hutton et al.). Such availability of code and data is equally important in the classroom.

One specific challenge in technology-augmented learning is in the incorporation of web based data resources, such as those provided by the United States Geological Survey (USGS), into an introductory hydrology course. The USGS has made many attempts to make their data holdings accessible to the general public and to scientists through, for example, the National Map ([www.nationalmap.gov](http://www.nationalmap.gov)) (Simley and Carswell Jr 2009), South Florida Information Access Data Exchange ([http://www1.usgs.gov/csas/obis-usa/data\\_search\\_and\\_access.html](http://www1.usgs.gov/csas/obis-usa/data_search_and_access.html)), the Geo Data Portal (Blodgett 2011), and Biodiversity Information Serving Our Nation (BISON) (<https://bison.usgs.gov/#home>) (Hanken 2013). In 2001, the USGS released a public web interface for their National Water Information System (NWIS) called USGS Water Data for the Nation (<http://dx.doi.org/10.5066/F7P55KJN>). This website includes simple tutorials on navigating the website interface and methods for downloading hydrologic data. Many of the web-based data sources provided by the USGS have either examples or primers on how to use their tools, but no additional resources which could be integrated with or support current hydroscience curriculum.

The Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI), has made extensive efforts to increase the academic use of USGS and other water-related, web-based data sets by educators through various initiatives, webinars, and conferences (<https://www.cuahsi.org/Posts/Cyberseminars>) (CUAHSI 2012; Keim et al. 2014; Morsy et al. 2016; Tarboton et al. 2014; Tarboton et al. 2011). These efforts have generally focused on the CUAHSI Hydrologic Information System (Ames et al. 2009; Tarboton et al. 2009), which has included USGS NWIS datasets since its inception. CUAHSI resources include training modules that provide step-by-step instruction on how to access many datasets through CUAHSI's

HydroClient and how to create data visualizations in Microsoft Excel and RStudio (Brazil 2016a; Brazil 2016b).

Another data portal has been developed by the USGS called the National Water Census Data Portal (NWC-DP) (<https://cida.usgs.gov/nwc/>). This portal provides public access to large amounts of hydrologic data, collected through the USGS's National Water Census (NWC) research program (<https://water.usgs.gov/watercensus/water-budgets.html>). The NWC is responsible for the collection of precipitation, evapotranspiration, and stream runoff data. Along with NWC data, web-based tools are available in the portal to provide a means whereby the data can be analyzed. NWC-DP provides support for making informed decisions affecting water availability and usage (<https://water.usgs.gov/watercensus/>). The NWC-DP, however, does not have an educational component to instruct users how to access the associated tools and data.

Missing in these efforts to communicate USGS water resources data to educators and students, is an interactive, code-based set of educational materials that integrates scripting for data access, retrieval and analysis with basic hydrologic principles. In this paper, we present the development of the National Water Census Education R package (NWCEd) and corresponding open courseware educational laboratory exercises which serve to fill this gap.

The remainder of this paper is organized as follows. The methods section presents a description of the development of the NWCEd package and its associated functions. A discussion on the development of five educational laboratory exercises is provided, including established learning objectives for the materials. The Results section discusses how the NWCEd technology supports the content of each of the laboratory exercises. An analysis of the correspondence between laboratory content and the established learning objectives is also

included. The conclusion provides recommendations for future research with regards to the effectiveness of laboratory exercises in improving learning.

## 2 METHODS

The National Water Census Education R-package (NWCEd) and associated laboratory exercises serve as tools which introduce hydroscience students to USGS hydrologic time series datasets, processing and analysis tools, and hands-on learning activities. This section describes the NWCEd, its associated package components and functions, and external R-packages required for complete NWCEd functionality. Discussion is also provided on the creation of the laboratory exercises.

### 2.1 Documentation

The NWCEd R-package was developed in RStudio (RStudio-Team 2015) and contains several custom R functions, preloaded data tables, NWC-DP screenshot images, and HTML and R Markdown (.Rmd) files of the laboratory exercises. Much of the documentation was generated automatically using Roxygen2 (<https://cran.r-project.org/web/packages/roxygen2/vignettes/rd.html>), including seven Markdown (.Rd) files describing the custom-made R functions and the NAMESPACE file which contains the list of functions made available through the package. The documentation also includes a manually created README file which provides instructions for installing the NWCEd package, laboratory content, required external R packages, links for viewing the laboratory exercises in a web

browser, and licensing information (<http://water.usgs.gov/software/help/notice/>). Complete NWCEd documentation and source code is available at <https://github.com/NWCEd/NWCEd>.

## 2.2 Custom NWCEd R-package Functions

One of the primary objectives of the NWCEd package is to allow users to access hydrologic time series datasets from the NWC-DP in RStudio. To do this, we created custom functions to retrieve the respective datasets from a USGS THREDDs server (<http://doi.org/10.5065/D6N014KG>). The process is outlined in Figure 1 below.

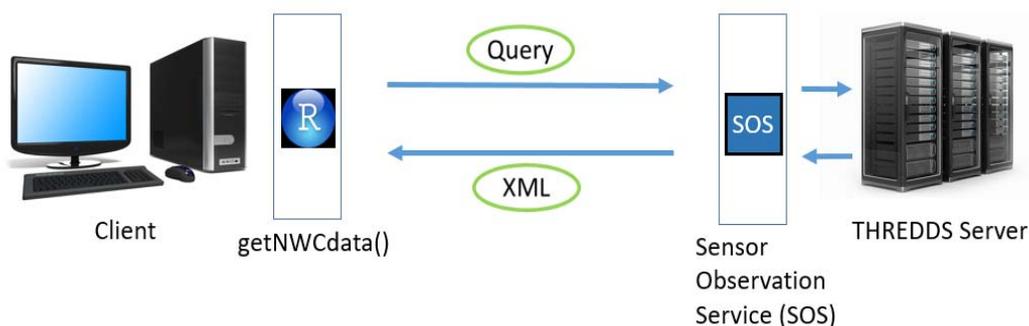


Figure 1: The `getNWCdata` retrieves hydrologic data from USGS servers and returns an RStudio data table to the user.

The user builds a query for a desired hydrologic dataset based on a user-selected hydrologic unit code (HUC) using the `getNWCdata` function. Using an HTTP protocol, the query is sent to a Sensor Observation Service (SOS) web service (Jirka et al. 2012). The SOS web service retrieves the desired dataset and returns it to the user formatted as XML text (Erl 2004). The `getNWCdata` function then parses the XML file into an RStudio data table view.

Another function, `getNWCWatershed`, extracts watershed geometry information from a USGS server in a similar way as the `getNWCdata` function. A query is built using the

getNWCWatershed function and is sent using an http protocol. The query interacts with the web feature service (WFS) (Peng and Zhang 2004). The WFS interacts with GeoServer to obtain a polygon feature associated with the specified HUC. A GeoJSON text file (Butler et al. 2008) containing the watershed boundary coordinates is then returned to the user and converted to a map representation for visualization. To improve visualization, the polygon fill option is turned off, subsequently displaying only the boundary of the watershed. The boundary is then laid over a base map provided by the external Leaflet R package (Graul 2016) which allows users to zoom in and out, as well as pan around the area of their desired watershed.

A Log-Pearson Type III function was developed for the purpose of teaching concepts related to hydrologic probability distributions and return periods. This function, called Lp3, implements the method described by Oregon State University (<http://streamflow.engr.oregonstate.edu/analysis/floodfreq/#log>). The Lp3 function receives two arguments: a user-named variable containing previously downloaded NWC-DP data, and a dataset type such as “prcp” for precipitation or “streamflow” for streamflow data. The Lp3 function then returns a Log-Pearson Type III distribution plot. Other functionality has been provided in the NWCEd package including the annualize function which converts time series datasets from daily time series to annual time series, and the getNWISsite which verifies the location of an NWIS gage.

Table 1 lists other external open source R packages used in this project together with each specific, respective function. A description of the application of these packages in the project is provided later under the Laboratory Exercises Development section.

Table 1: NWCEd works in concert with functionalities provided by external R packages.

<b>Package Name</b>	<b>Functionality Incorporated</b>
datasets	Preloaded datasets
devtools	Install NWCEd from GitHub
dplyr	Data manipulation
foreach	Looping
ggplot2	Plotting
gridExtra	Grid graphics
iterators	Iterating
leaflet	Interactive map
psych	Summary statistics
scales	Graphical aesthetics
stats	Statistics functions

Each of these packages are hosted on the Comprehensive R Archive Network (CRAN) website (R-Core-Team 2016). The packages are downloaded from CRAN into an instance of RStudio using default RStudio functions. The NWCEd package is downloaded from GitHub into an instance of RStudio using the functionality of the devtools package as listed in the table above. An illustration of the relationship between the local instance of RStudio and GitHub, CRAN, and the NWC-DP is provided as Figure 2 below.

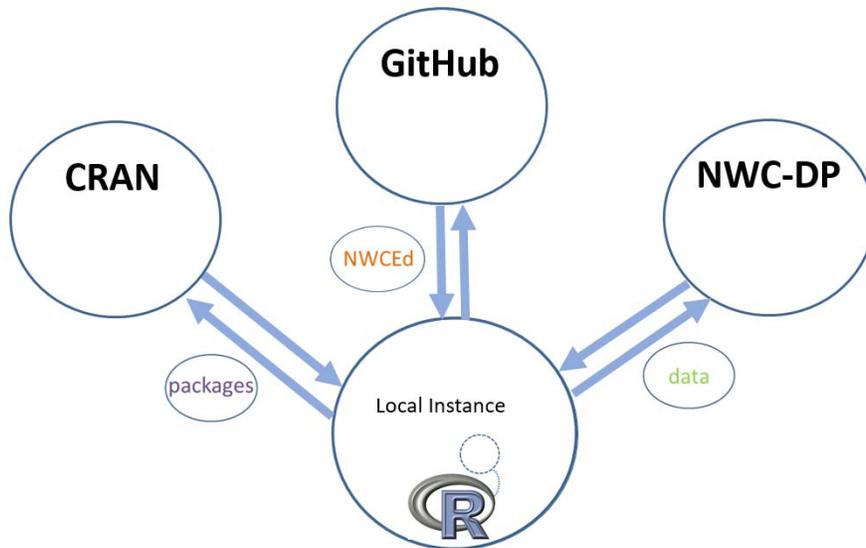


Figure 2: NWCEd, external R packages, and data from NWC-DP is downloaded into a local instance of RStudio.

### 2.3 Laboratory Exercises Development

A core goal of this work was to develop a set of laboratory exercises that expose the National Water Census data and web services to hydrology students and educators. These exercises were developed in RStudio. The raw content for these labs was written in html, R Markdown, and Javascript in order to achieve the desired interactive functionality. The knitr package was then used to read each of the languages and convert the contents into interactive html pages (Xie 2016).

The labs have many shared, built-in features, including layout, image display, and interactive buttons. The specific layout of the lab documents comes from an R Markdown document template. The template used for the NWCEd labs was an html template. The “cerulean” theme was selected which corresponds to the blue font of the labs. An interactive table of contents is also a common layout feature. This feature allows the user to click on the

desired section within a given lab and automatically be relocated within the document to the specified location.

Each of the five labs are set up to instruct students on how to use the NWC-DP. This is done by including several screenshots of the NWC-DP as part of the content. Relative paths for these images were coded in R Markdown into the raw document, and knitr was used to generate the images into the html documents with varying formatting styles.

Another feature of the labs is the ability to toggle code blocks, hints, and answers to questions on and off using Javascript code. A simple function was included in the raw document which defaults the toggle settings to “none”, meaning that the information is not shown. The user can then click the blue toggle buttons to reveal the hidden information.

The plotting capabilities of the ggplot2 package (Wickham 2009) were used extensively throughout the development of the third, fourth, and fifth labs. The library was used to produce boxplots, histograms, probability and fitted distributions, time series, and double-mass curve plots. Knitr was then used to convert ggplot2 code, written in R Markdown, into the desired plots embedded into the html pages. More information regarding the use of the plots will be given in the Results section.

## **2.4 Laboratory Content**

We selected content for the labs to coincide with topics discussed in the commonly used college textbook, *Hydrology – Water Quantity and Quality Control* (Wanielista et al. 1997) while incorporating the use of RStudio and data from the NWC-DP. The hydrology concepts discussed in the labs include locating metadata for hydrologic datasets and tools provided by the NWC-DP, graphing analysis methods for hydrologic datasets, water balances, double-mass curve analyses, and Log-Pearson Type III distributions. Other concepts mentioned include common

statistics terms such as range, skew, and kurtosis and an introduction to simple R functions in RStudio. Table 2 below provides a comprehensive summary of the content contained in each of the labs.

Table 2: The content of the laboratory exercises associated with NWCEd was summarized with the associated NWCEd functions.

Lab #	Material Description	NWCEd Functions
Lab 1	<ul style="list-style-type: none"> <li>• Accessing the NWC-DP</li> <li>• NWC-DP Water Budget tool</li> <li>• Introduction to HUC's</li> </ul>	
Lab 2	<ul style="list-style-type: none"> <li>• Accessing the NWC-DP</li> <li>• NWC-DP Water Streamflow Stats tool</li> <li>• Downloading stats results</li> </ul>	
Lab 3	<ul style="list-style-type: none"> <li>• Accessing ET/ precipitation data from NWC-DP</li> <li>• Analyzing and plotting box plots, histograms, and density curve plots in R</li> <li>• Writing exercises</li> </ul>	<ul style="list-style-type: none"> <li>• getNWCData()</li> <li>• annualize()</li> </ul>
Lab 4	<ul style="list-style-type: none"> <li>• Analyzing time series and bar plots in R using NWC-DP data</li> <li>• Water balances</li> <li>• Double-mass curve analysis</li> </ul>	<ul style="list-style-type: none"> <li>• getNWCWatershed()</li> <li>• getNWCData()</li> <li>• annualize()</li> </ul>
Lab 5	<ul style="list-style-type: none"> <li>• Using a Log Pearson Type III model to analyze NWC-DP data in Excel</li> <li>• Using a Log Pearson Type III model to analyze NWC-DP data in Rstudio</li> <li>• Writing exercises</li> </ul>	<ul style="list-style-type: none"> <li>• getNWCData()</li> <li>• Lp3()</li> </ul>

The content for each of the labs was developed to support three learning objectives for the hydroscience students. The first learning objective is understanding how to access hydrologic datasets and associated metadata from the NWC-DP. The second objective is learning how to analyze datasets from the NWC-DP using both numerical and graphical methods in R. The third objective is to practice effective communication of data analysis through the use of graphs. An assessment of the correspondence between learning objectives and the provided lab content are provided in the Results section.

### **3 RESULTS**

The NWCEd package functions and external R-package functions serve to support the hydrology concepts presented in the NWCEd laboratory exercises. This section describes the relationship between the lab content and the technology incorporated into the labs. Also included in this section is an analysis of the effectiveness of the lab content in meeting the established learning objectives as described previously.

Lab 1 and Lab 2 were designed specifically to instruct students how to navigate through the NWC-DP to access USGS hydrologic datasets. Step-by-step instructions were laid out in Lab 1 and Lab 2 for choosing the desired datasets according to a user-specified HUC or a user-specified USGS stream gage. Lab 1 walks the user through NWC-DP features while Lab 2 teaches the user how to access associated metadata, how to access statistical tools for running analyses in the portal, and how to download the results into Microsoft Excel as shown in Figure 3 below. There is no interactive R technology incorporated into these two labs.



Figure 5: La Pine gage location link.



Figure 6: Webpage displaying metadata.

Move your cursor over the plotted data. You will notice that actual data points are displayed below the plot similarly to data plots generated using the Water Budget tool. There are two identical tables below the streamflow data plot, each with unchecked boxes and a list of statistic group names. See Figure 7 below.

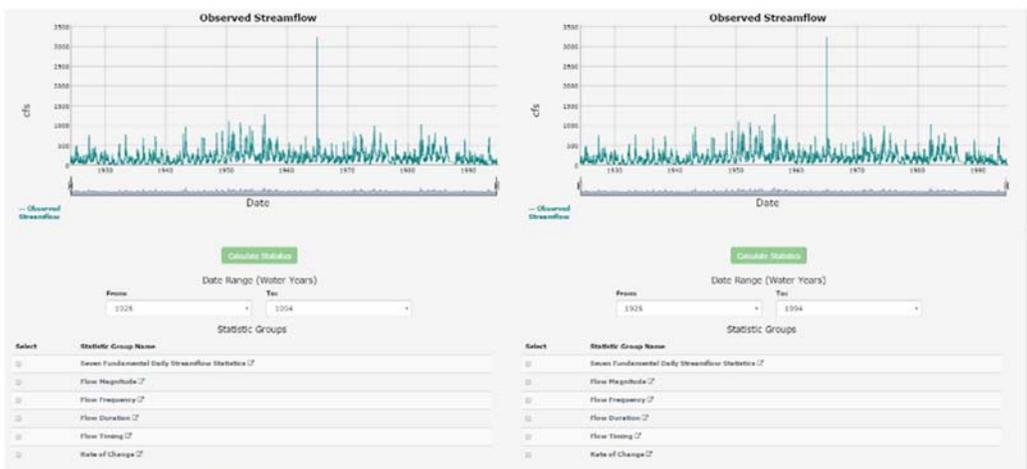


Figure 7: Statistic group tables link.

Figure 3: Lab 2 instructs users how to access and run statistics on streamflow data from the NWC-DP.

While the first two labs do not have any supporting R code, the other three labs are heavily supported through R technology. These labs assume students have a working knowledge of R and RStudio. Lab 3 focuses on graphical methods in R for analyzing hydrologic data. The graphical methods discussed include box plots, histograms, and density plots.

Box plots were discussed to teach students how to identify outliers in datasets, as well as how to analyze the spread of the data. The cars dataset from the built-in datasets R package was also used in connection with ggplot2 to provide a box plot with non-uniform spread. This allowed students to more easily visualize the attributes of the dataset. In addition, ggplot2 was

used to provide labels to the box plot which corresponded to the content discussed in the lab. The R package gridExtra was applied during the box plot section of Lab 3 to compare box plots side by side (Baptiste 2016).

A discussion on using histograms to analyze hydrologic data was also presented in the content for Lab 3. The ggplot2 and gridExtra packages were used to compare several histograms of evapotranspiration data side by side, each with varying band widths as shown in Figure 4 below.

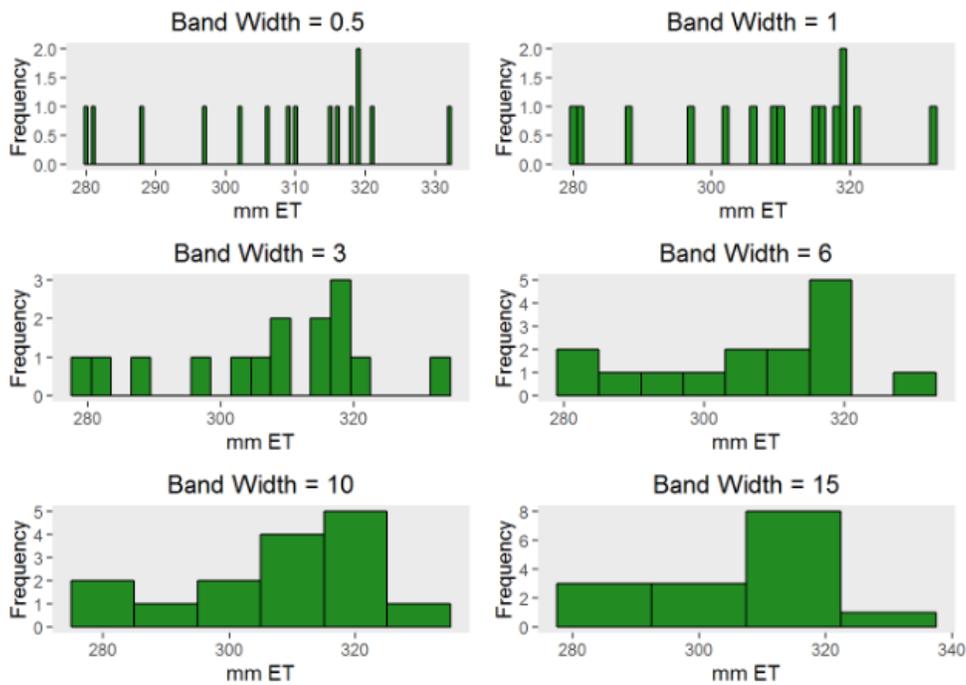


Figure 4: Histogram bandwidths are compared in Lab 3.

With all of the histograms plotted side by side and effectively labeled, students are able to see the effect that band width has on histograms. This example is expected to help students make more informed decisions about how they represent their data.

Lab 4 content teaches general principles related to water balances. An interactive map is used to support the water balance concept by allowing the students to view the physical nature of their watershed with the `getNWCWatershed` function from the `NWCED` package and the `leaflet` package as described previously. Time series plots produced using `ggplot2` allow students to view actual datasets from the NWC-DP and track the current state of the water balance for their selected watershed. Figure 5 below contains a time series plot from Lab 4 containing evapotranspiration, precipitation, streamflow, and calculated water storage levels.

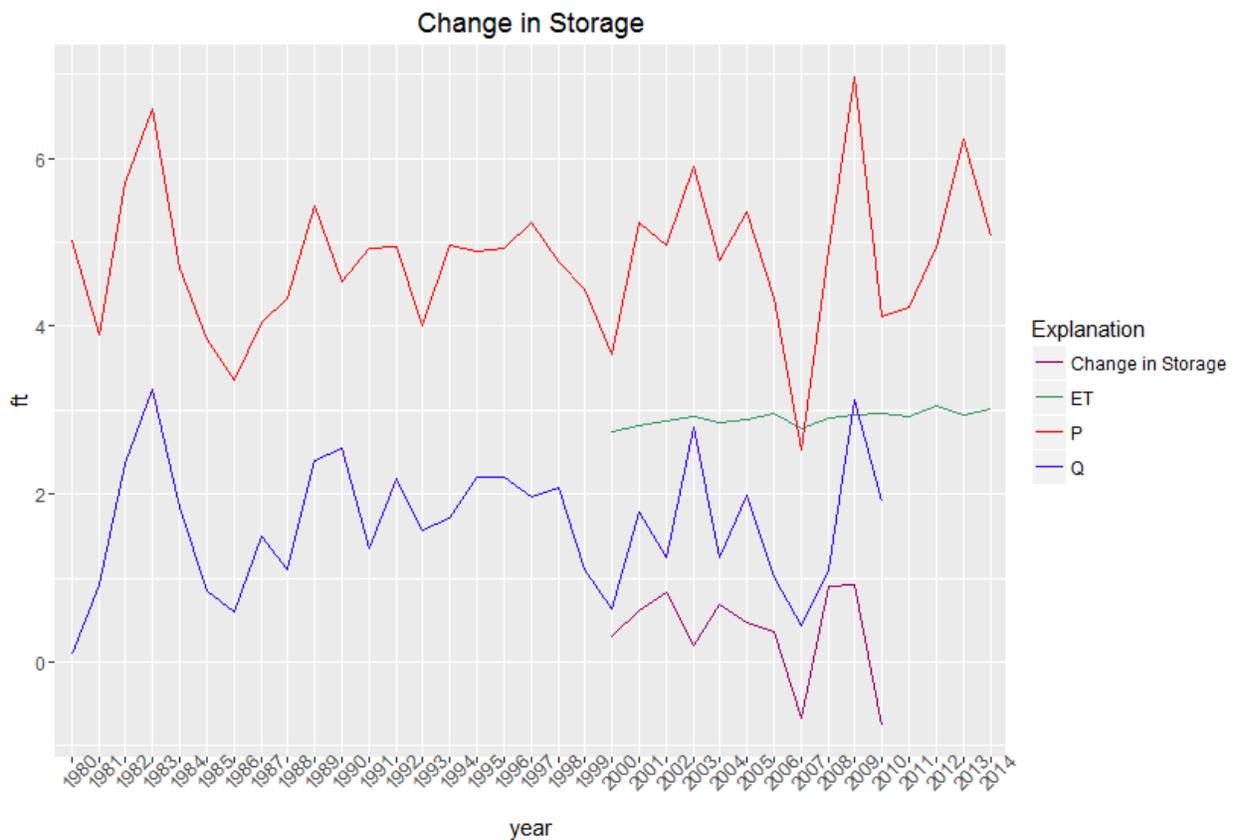


Figure 5: Lab 4 instructs users how to visualize NWC-DP time series data to improve the understanding of the water budget concept.

Lab 4 also uses the dplyr package (Wickham 2016) in tandem with ggplot2 to produce a color-filled histogram with bars above the axis referring to positive storage and bars below the axis referring to a loss in storage. By using different plot types, students' awareness and understanding of a water balance can be increased, as well as their methods to communicate water balance information in the future.

Numerical analyses are introduced in Lab 4 with the use of double-mass curves. To support the concept, an equation and its variables are presented to educate students on how the analysis is to be done numerically. A series of plotted curves are also used to graphically describe the process. The series of plotted curves is presented below in Figure 6.

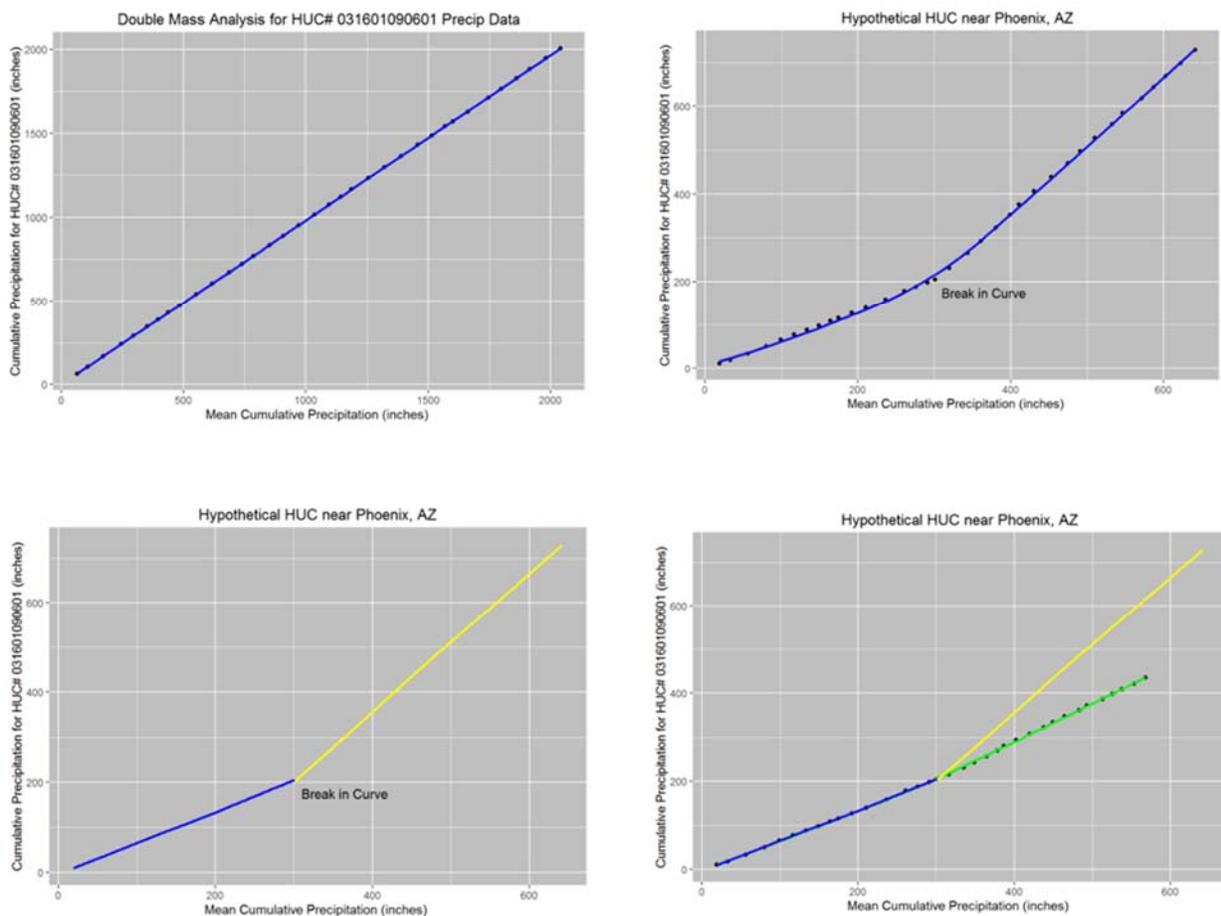


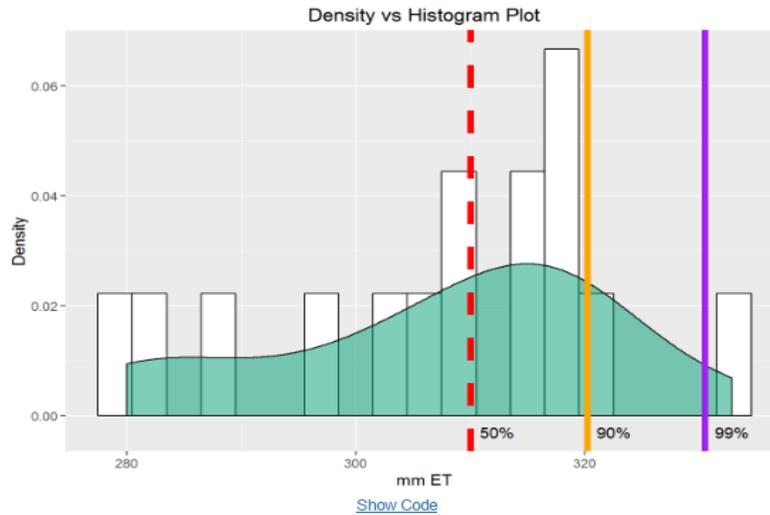
Figure 6: Double-mass curve analysis is visualized in Lab 4 using NWC-DP data.

The use of different colors helps students better visualize the adjustment they have made to their dataset to correct for error. Showing only the final, corrected line would not adequately help students recognize the significance of the adjustment to the data. Plotting the curves in sequence along with both the uncorrected and corrected data together improve the understanding of the method used to perform the adjustment as well as the significance of the change made.

Lab 5 incorporates Microsoft Excel and its PivotTable feature in a step by step tutorial on calculating a Log-Pearson Type III distribution for a hydrologic dataset from the NWC-DP. The datasets from the NWC-DP are very large, which makes the steps for performing the distribution challenging. After walking the students through the lengthy process, the lab then shows students an alternative method using the custom Lp3 function in the NWCEd package.

This process allows students to learn the individual steps of calculating a distribution before using R functions to perform the same calculation. This process also helps students see the benefit of using the Lp3 function in that students are able to download large datasets directly into their instance of RStudio and then perform the analysis, saving them time.

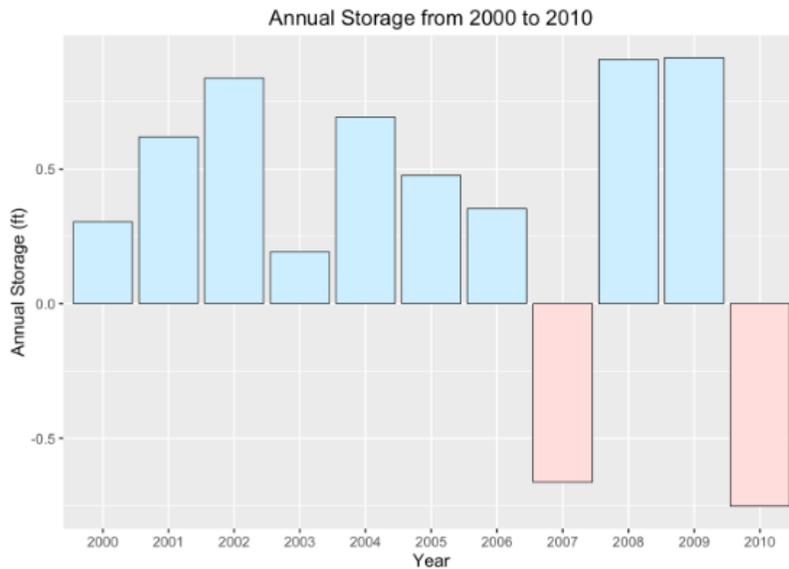
It is important to note that for each of the features of the labs described above, there is a viewable block of code which can be copied and pasted directly into an instance of RStudio. This allows for the reproduction of each graph, interactive map, and curve shown in the labs. These code blocks allow students to familiarize themselves with the components and functions used to produce the associated content as shown in Figure 7, Figure 8, and Figure 9 below.



```
library(ggplot2)
ggplot(annualgetetdata, aes(x = data)) + geom_histogram(aes(y = ..density..), binwidth = 3, colour = "black", fill = "white") + geom_density(alpha=.5, fill="#009E73") + geom_vline(aes(xintercept = 310.00), color="red", linetype = "dashed", size = 2) + ggtitle("Density vs Histogram Plot") + geom_vline(aes(xintercept = 320.20), color="orange", linetype = "solid", size = 2) + geom_vline(aes(xintercept = 330.46), color="purple", linetype = "solid", size = 2) + annotate("text", x=312.3, y = -.003, label = "50%") + annotate("text", x=322.5, y=-.003, label = "90%") + annotate("text", x=332.8, y=-.003, label = "99%") + xlab("mm ET") + ylab("Density")
```

This plot shows a density curve plotted on top of the histogram of the annual ET data. A vertical red dashed line indicates the median. As you can see, the hump of the curve is shifted to the right with the tail on the left. This much more clearly describes the distribution of the data. The 90th and 99th percentiles have been plotted with vertical orange and purple lines, respectively. The 90th percentile and 99th percentiles were found to be 320.20 and 330.46 mm ET, respectively. This means that in a given year for our designated watershed, there is a 10% chance of losing more than 320.2 mm of water to ET and a 1% chance of losing more than 330.46 mm water to ET. Lets take a look at the density plot for the annual precipitation dataset.

Figure 7: Code has been provided to reproduce density curve plots, including the manual adjustment of labels and percentile lines.



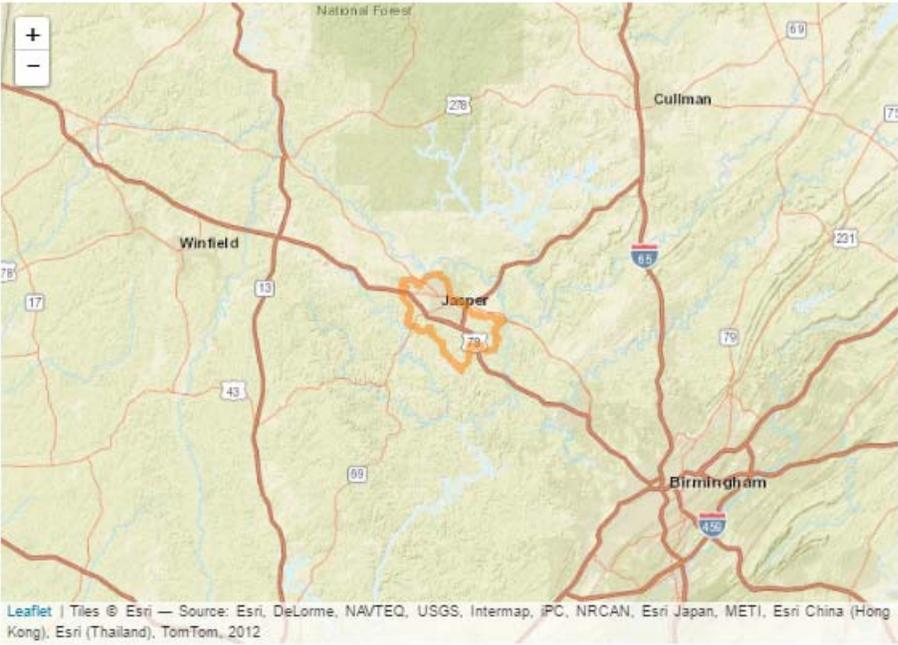
[Show Code](#)

```
library(dplyr)
NWdata<-getNWdata(huc="031501070601", local = FALSE)
NWwatershed<-getNWwatershed(huc="031501070601",local=FALSE)
areasqft<-NWwatershed$features[[1]]$properties$areaacres*43560
NWdata$et$data<-NWdata$et$data*0.0032808
annual_dataet<-annualize(NWdata$et, method=sum)
annual_dataet$group<- 'ET'
NWdata$prcp$data<-NWdata$prcp$data*0.0032808
annual_dataprcp<-annualize(NWdata$prcp, method=sum)
annual_dataprcp$group<- 'P'
NWdata$streamflow$data<-NWdata$streamflow$data*60*60*24/areasqft
annual_datastrf<-annualize(NWdata$streamflow,method=sum)
annual_datastrf$group<- 'Q'
annual_union<-merge(annual_dataet,annual_dataprcp,by.x="year",by.y="year")
annual_union<-merge(annual_union,annual_datastrf,by.x="year",by.y="year")
names(annual_union)<-c("year","ET","groupET","P","groupP","Q","groupQ")
annual_diff<-as.data.frame(annual_union$P - annual_union$ET - annual_union$Q)
names(annual_diff) = c('data')
annual_diff$group<- 'diff'
annual_diff$year<-annual_union$year

annual_diff = annual_diff%>%
  mutate(pos = data >= 0)

ggplot(annual_diff, aes(x=year, y=data, fill=pos)) + geom_bar(stat='identity', position = 'identity', color='black', size=0.
25)+ scale_fill_manual(values=c('#FFDDDD','#CCEEFF'), guide = FALSE) + xlab("Year") + ylab("Annual Storage (ft)") +
ggtitle("Annual Storage from 2000 to 2010")
```

Figure 8: The code for a histogram with positive and negative bars is provided to introduce an alternative way to visualize water storage.



[Show Code](#)

```

library(NWCed)
library(leaflet)
# Enter the huc12 to be analyzed like this:
huc12<-"031601090601"
# Get watershed geometry and put it on a map.
NWCwatershed<-getNWCwatershed(huc=huc12,local=FALSE)
coords<-NWCwatershed$features[[1]]$geometry$coordinates[[1]][[1]]
coords<-as.data.frame(matrix(unlist(coords),nrow=length(coords),byrow = T))
names(coords)<-c('lon','lat')
leaflet() %>%
  setView(lng = mean(coords$lon), lat = mean(coords$lat), zoom = 9) %>%
  addTiles("http://server.arcgisonline.com/ArcGIS/rest/services/World_Street_Map/MapServer/tile/{z}/{y}/{x}") %>%
  addWMSTiles(baseUrl = "https://cida.usgs.gov/nwc/geoserver/gwc/service/wms",
              options = WMSTileOptions(format = "image/png", transparent = TRUE, opacity = 1), attribution =
"Tiles &copy; Esri &mdash; Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kon
g), Esri (Thailand), TomTom, 2012") %>%
  addGeoJSON(NWCwatershed, weight = 5, color = "#ff7800", fill = FALSE)

```

Figure 9: Code is provided to replicate interactive maps in the user's local instance of RStudio.

Students who are permitted to study and produce their own code for the associated hydrology topics will be prepared for more advanced coursework related to hydrology and hydroinformatics.

We established three learning objectives for the content of the laboratory exercises.

Learning Objective 1 is for students to understand how to access the NWC-DP including the associated datasets and tools. Learning Objective 2 is for students to learn how to analyze NWC-

DP data using both numerical methods and graphical methods in R. Learning Objective 3 is for students to practice communicating the results of their analyses using graphs.

The R code and labs were posted to <https://github.com/NWCEd/NWCEd> and are available for download and use as public domain resources.

## 4 CONCLUSION AND DISCUSSION

The NWCEd R package and associated labs were developed to accomplish two objectives. The first was to promote hydrologic data from the NWC-DP. The second was to incorporate new R technology into current hydrology curriculum to better prepare students for future work in hydrology and hydroinformatics.

The NWCEd package contains custom functions which allow users to install hydrologic datasets directly from the NWC-DP into their instance of RStudio for viewing and analyzing. The NWCEd package works in tandem with several pre-existing packages hosted on CRAN. There are 5 laboratory exercises developed to teach students about hydrology concepts with hands-on examples using NWC-DP data in RStudio.

The NWCEd package and associated laboratory exercises are intended to be used as stand-alone material to support hydroscience courses, such as undergraduate hydrology and hydroinformatics. This design allows educators to have flexibility in selecting lab content that supports their current curriculum. Because the content is self-teaching, instructors do not need to be expert in programming in R. Additionally, the lab content is designed to be used outside of the classroom. This relieves the educator from needing special technology in the classroom.

There exist limitations associated with the laboratory content and NWCEd package. As mentioned previously in this paper, the materials do require a working knowledge of RStudio. It is recommended that educators and students seek open source instruction on downloading and

installation of R and RStudio (<https://www.rstudio.com/products/rstudio/download/>). Students must have access to a computer and internet in order to access and use the materials.

There is a need for further research to determine how effective these tools are at educating students on the NWC-DP, data acquisition through the NWC-DP and R, and improving current understanding of hydrology concepts through applying new technology. Continued research is necessary to determine the confidence of students in beginning hydrologic data analysis after completing the materials. It is also recommended that research be conducted to determine how the materials are received and used by hydrology and hydroinformatics educators.

Potential outreach activities for these tools include a CUAHSI webinar. We expect in the near future to post this material on the NWC-DP as well as CUAHSI's website. As an open source project, this material is open to the public to update and reuse and access through GitHub. The USGS is responsible for maintaining the code for the foreseeable future.

## **5 SOFTWARE AVAILABILITY**

NWCEd, Version 1.0, David Blodgett, Jake Nelson, Software is in Public Domain,  
Available starting November 15, 2016, Retrievable at <https://github.com/NWCEd/NWCEd>.

## REFERENCES

- Ames, D., J. Horsburgh, J. Goodall, T. Whiteaker, D. Tarboton & D. Maidment, Introducing the open source CUAHSI Hydrologic Information System desktop application (HIS Desktop). In: 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, 2009. p 4353-4359.
- Baptiste, A., Anton, Antonov, 2016. gridExtra: Miscellaneous Functions for "Grid" Graphics. 2.2.1 edn.
- Blodgett, D. L., 2011. Geo Data Portal Spatial Data Infrastructure Documentation.
- Brazil, L., Pollak, J., 2016a. Downloading Discharge and Precipitation Data from HydroClient and Interpreting it with Graphs in Excel. In: Consortium of Universities for the Advancement of Hydrologic Science, Inc. <http://serc.carleton.edu/hydromodules/steps/114767.html> Accessed November 21 2016.
- Brazil, L., Pollak, J., 2016b. Investigating Air Temperature with HydroClient and RStudio. In. <http://serc.carleton.edu/hydromodules/steps/120944.html> Accessed November 21 2016.
- Butler, H., M. Daly, A. Doyle, S. Gillies, T. Schaub & C. Schmidt, 2008. The GeoJSON format specification. Rapport technique:67.
- CUAHSI, CUAHSI Third Biennial Science Meeting Updates-" Fusing Science and Solutions" July 16-18, 2012-Boulder, CO. In: Workshop-Penn State University, 2012.
- Erl, T., 2004. Service-oriented architecture: a field guide to integrating XML and web services. Prentice Hall PTR.
- Graul, C., 2016. leafletR: Interactive Web-Maps Based on the Leaflet JavaScript Library. 0.4-0 edn.
- Hanken, J., 2013. Biodiversity online: toward a network integrated biocollections alliance. *BioScience* 63(10):789-790.

- Hutton, C., T. Wagener, J. Freer, D. Han, C. Duffy & B. Arheimer, Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*:n/a-n/a doi:10.1002/2016WR019285.
- Jirka, S., A. Bröring, P. Kjeld, J. Maidens & A. Wytzisk, 2012. A lightweight approach for the sensor observation service to share environmental data across Europe. *Transactions in GIS* 16(3):293-312.
- Keim, R. F., C. Kendall & A. Jefferson, 2014. The expanding utility of laser spectroscopy. *Eos, Transactions American Geophysical Union* 95(17):144-144.
- Merwade, V. & B. L. Ruddell, 2012. Moving university hydrology education forward with community-based geoinformatics, data and modeling resources. *Hydrol Earth Syst Sci* 16(8):2393-2404 doi:10.5194/hess-16-2393-2012.
- Morsy, M. M., J. L. Goodall, A. M. Castronova, P. Dash, B. Miles, V. Merwade, J. M. Sadler & D. G. Tarboton, 2016. Design and Implementation of Hydrologic Model Sharing Capabilities within the CUAHSI HydroShare System.
- Peng, Z.-R. & C. Zhang, 2004. The roles of geography markup language (GML), scalable vector graphics (SVG), and Web feature service (WFS) specifications in the development of Internet geographic information systems (GIS). *Journal of Geographical Systems* 6(2):95-116.
- Popescu, I., A. Jonoski & B. Bhattacharya, 2012. Experiences from online and classroom education in hydroinformatics. *Hydrol Earth Syst Sci* 16(11):3935-3944 doi:10.5194/hess-16-3935-2012.
- R-Core-Team, 2016. R: A Language and Environment for Statistical Computing. In. <https://www.R-project.org> Accessed November 23, 2016 2016.
- RStudio-Team, 2015. RStudio. 3.2.5 edn.
- Ruddell, B. L. & T. Wagener, 2013. Grand challenges for hydrology education in the 21st century. *Journal of Hydrologic Engineering* 20(1):A4014001.
- Sanchez, C. A., B. L. Ruddell, R. Schiesser & V. Merwade, 2016. Enhancing the T-shaped learning profile when teaching hydrology using data, modeling, and visualization activities. *Hydrol Earth Syst Sci* 20(3):1289-1299 doi:10.5194/hess-20-1289-2016.
- Simley, J. D. & W. J. Carswell Jr, 2009. The national map—hydrography. US Geological Survey Fact Sheet 3054(4).
- Tarboton, D., J. Horsburgh, D. Maidment, T. Whiteaker, I. Zaslavsky, M. Piasecki, J. Goodall, D. Valentine & T. Whitenack, Development of a community hydrologic information system. In: 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New

- Zealand and International Association for Mathematics and Computers in Simulation, 2009. p 988-994.
- Tarboton, D., R. Idaszak, J. Horsburgh, J. Heard, D. Ames, J. Goodall, L. Band, V. Merwade, A. Couch & J. Arrigo, HydroShare: advancing collaboration through hydrologic data and model sharing. In: International Environmental Modelling and Software Society (iEMSs) 7th International Congress on Environmental Modelling and Software, San Diego, California, USA [www.iemss.org/society/index/php/iemss-2014-proceedings](http://www.iemss.org/society/index/php/iemss-2014-proceedings), 2014.
- Tarboton, D. G., D. Maidment, I. Zaslavsky, D. Ames, J. Goodall, R. P. Hooper, J. Horsburgh, D. Valentine, T. Whiteaker & K. Schreuders, Data Interoperability in the Hydrologic Sciences. In: Environmental Information Management Conference, Santa Barbara, CA, USA, 2011.
- Thompson, S. E., I. Ngambeki, P. A. Troch, M. Sivapalan & D. Evangelou, 2012. Incorporating student-centered approaches into catchment hydrology teaching: a review and synthesis. *Hydrol Earth Syst Sci* 16(9):3263-3278 doi:10.5194/hess-16-3263-2012.
- Wagener, T., C. Kelleher, M. Weiler, B. McGlynn, M. Gooseff, L. Marshall, T. Meixner, K. McGuire, S. Gregg, P. Sharma & S. Zappe, 2012. It takes a community to raise a hydrologist: the Modular Curriculum for Hydrologic Advancement (MOCHA). *Hydrol Earth Syst Sci* 16(9):3405-3418 doi:10.5194/hess-16-3405-2012.
- Wanielista, M., R. Kersten & R. Eaglin, 1997. *Hydrology: Water quantity and quality control*. John Wiley and Sons.
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlan New York.
- Wickham, H., Francois, Romain, RStudio, 2016. *dplyr: A Grammar of Data Manipulation*. 0.5 edn.
- Xie, Y., 2016. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. 1.15.1 edn.

## APPENDIX A

Documentation: <https://github.com/NWCEd/NWCEd/blob/master/README.md>

Lab 1: [https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab\\_1.html](https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab_1.html)

Lab 2: [https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab\\_2.html](https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab_2.html)

Lab 3: [https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab\\_3.html](https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab_3.html)

Lab 4: [https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab\\_4.html](https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab_4.html)

Lab 5: [https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab\\_5.html](https://cdn.rawgit.com/NWCEd/NWCEd/master/inst/Lab_5.html)