Assignment 4: Water Quality in Rivers

Yutao Gong

OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis on water quality in rivers.

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single HTML file.
5. After Knitting, submit the completed exercise (HTML file) to the dropbox in Sakai. Add your last name into the file name (e.g., "A04_Chamberlin.html") prior to submission.

The completed exercise is due on 25 September 2019 at 9:00 am.

Setup

1. Verify your working directory is set to the R project file,
2. Load the tidyverse, dataRetrieval, cowplot, xts and dygraphs packages.
3. Set your ggplot theme (can be theme_classic or something else)

```
getwd()
```

```
## [1] "C:/Users/gongy/Documents/Hydrologic_Data_Analysis/Assignments"
```

```
packages <- c("tidyverse",
              "dataRetrieval",
              "cowplot",
              "xts",
              "dygraphs",
              "lubridate")
invisible(lapply(packages, library, character.only = TRUE))
```

```
## -- Attaching packages ------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
##
## ********************************************************
```

```
## Note: As of version 1.0.0, cowplot does not change the
```

```
##   default ggplot2 theme anymore. To recover the previous
```

```
##   behavior, execute:
##   theme_set(theme_cowplot())
```

```
## ********************************************************
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Registered S3 method overwritten by 'xts':
##   method     from
##   as.zoo.xts zoo
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:cowplot':
##
##     stamp
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
theme_set(theme_bw())
```

Hypoxia

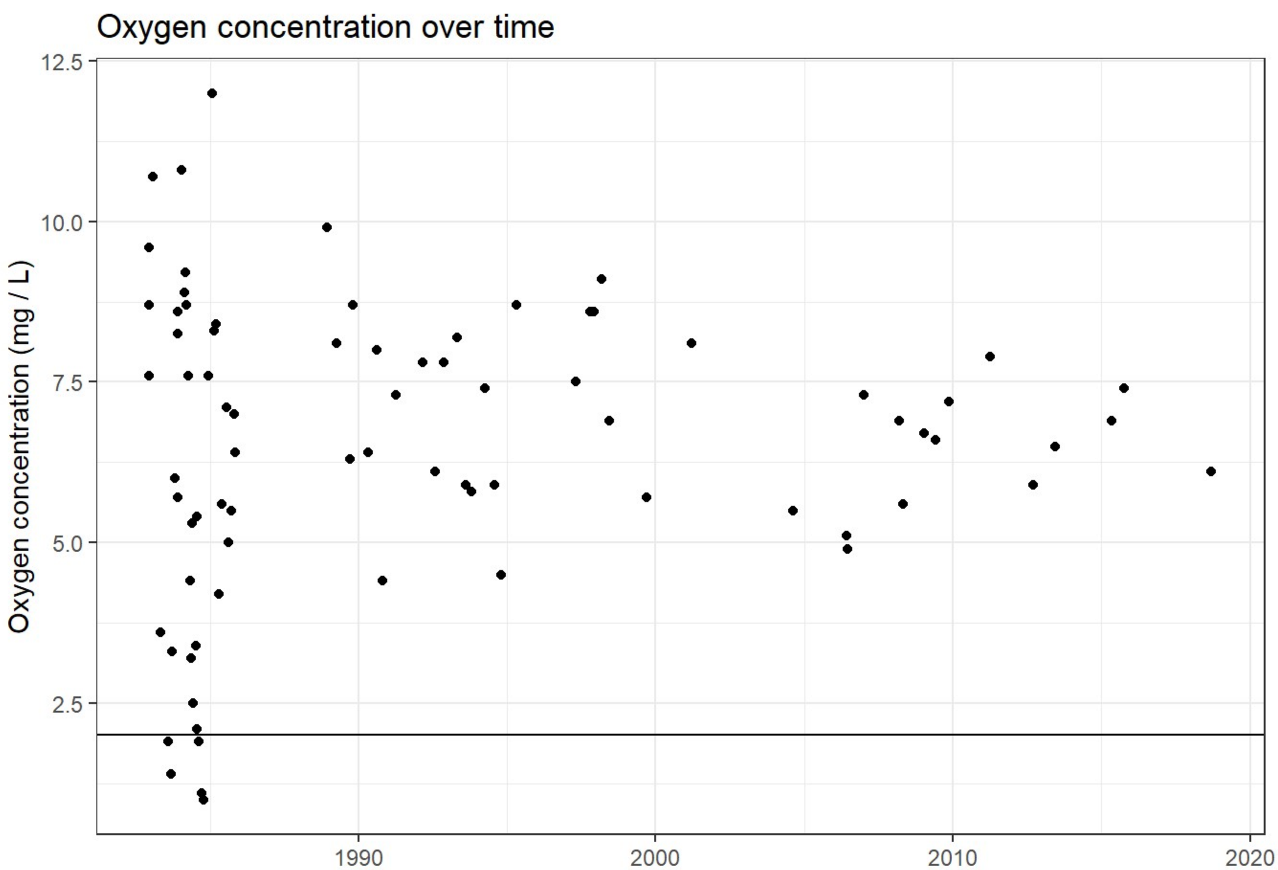
This assignment will look at another measure of water quality - oxygen concentration. Though not directly important for human health, oxygen in the water column is very important for aquatic life, and so is considered a measure of water quality. Hypoxia (low oxygen) has many different definitions. For this assignment, we will use 2 mg/L O_2 as our cut-off.

4. Import the oxygen water quality data from New Hope Creek at Blands (using `readNWISqw()` , site code `02097314` , parameter code `00300`). Make a data frame called `O2.dat` that includes only the Date and O_2 concentration values. Give your data frame understandable column names.

```
parameters <- parameterCdFile

dat.raw <- readNWISqw(siteNumbers = "02097314", #New Hope Creek at Blands
                     parameterCd = "00300", #oxygen concentration
                     startDate = "",
                     endDate = "")
O2.dat <- dat.raw %>%
  select(Date = sample_dt,
         Value = result_va) %>%
  group_by(Date) %>%
  summarize(Value = mean(Value))
```

5. Create a ggplot of oxygen concentrations over time. Include a horizonal line at 2 mg/l to show the hypoxia cutoff.

```
begin.date <- min(O2.dat$Date)
end.date <- max(O2.dat$Date)

Oxygen.plot <- ggplot(O2.dat, aes(x = Date)) +
  geom_point(aes(y = Value)) +
  lims(x = c(begin.date, end.date)) +
  labs(x = "", y = "Oxygen concentration (mg / L)", title = "Oxygen concentration over ti
me") +
  theme(legend.position = "top") +
  geom_hline(yintercept = 2)

print(Oxygen.plot)
```



6. What do you notice about the frequency of hypoxia overtime?

> Incidents of hypoxia only happened several times in the early 1990's, so its frequency definitely decreased over time.

Nutrients

7. Often times hypoxia is associated with high nutrient concentrations, because abundant nutrients promote biomass growth which increases respiration and depletes oxygen concentrations in the water (remember how oxygen concentrations were very low in the hypolimnion from the Physical Properties of Lakes week). Create a new data frame, called `nutrients.dat` with total nitrogen (parameter code `00600`) and total phosphorus (parameter code `00665`) data from the USGS. Your data frame should have 3 columns, `Date`, `TotalNitrogen_mgl-N`, and `TotalPhosphorus_mgl-P`.

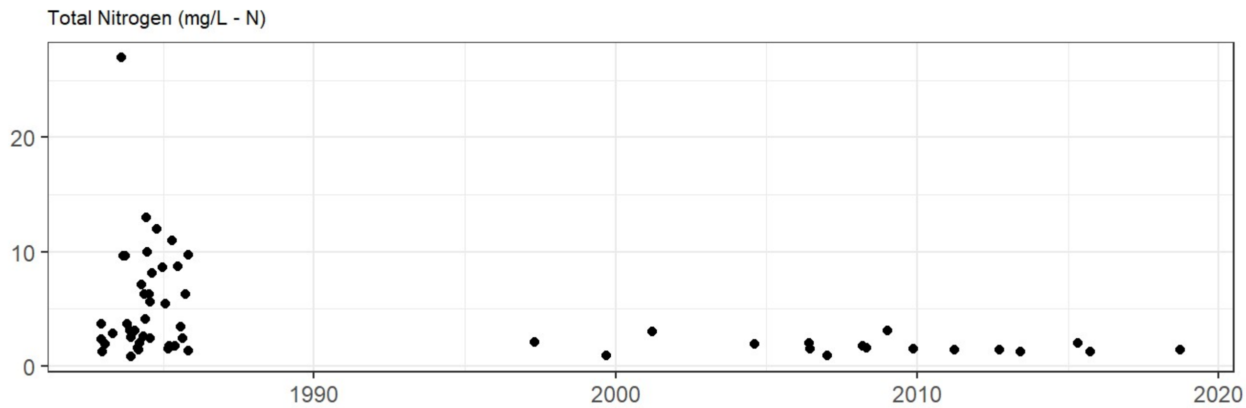
```
nutrients.raw <- readNWISqw(siteNumbers = "02097314", #New Hope Creek at Blands
                     parameterCd = c("00600",  #Total nitrogen
                                     "00665"),  #Total phosphorus
                     startDate = "",
                     endDate = "")

nutrients.dat <- nutrients.raw %>%
  select(Date = sample_dt,
         Parameter = parm_cd,
         Value = result_va) %>%
  group_by(Date, Parameter) %>%
  summarize(Value = mean(Value)) %>%
  spread(key = Parameter, value = Value)

names(nutrients.dat)[2:3] <- c("TotalNitrogen_mgl_N", "TotalPhosphorus_mgl_P")
```

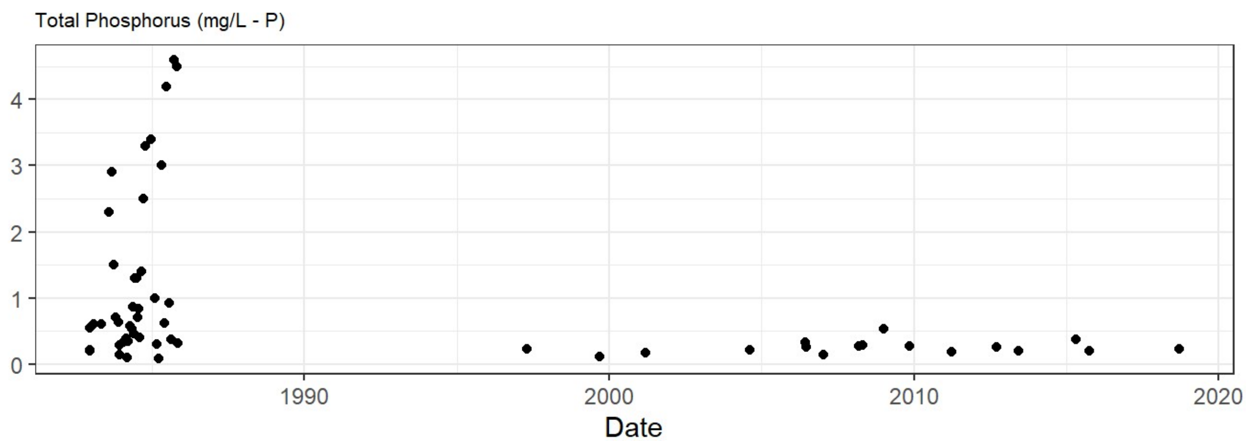
8. Create two ggplots stacked with cowplot that show nutrient concentrations over time.

```
begin.date <- min(nutrients.dat$Date)
end.date <- max(nutrients.dat$Date)

Nitrogen.plot <- ggplot(nutrients.dat, aes(x = Date)) +
  geom_point(aes(y = TotalNitrogen_mgl_N)) +
  lims(x = c(begin.date, end.date)) +
  labs(title = "Total Nitrogen (mg/L - N)", x = "", y = "") +
  theme(plot.title = element_text(size=8))

Phosphorus.plot <- ggplot(nutrients.dat, aes(x = Date)) +
  geom_point(aes(y = TotalPhosphorus_mgl_P)) +
  lims(x = c(begin.date, end.date)) +
  labs(title = "Total Phosphorus (mg/L - P)", y = "") +
  theme(plot.title = element_text(size=8))

nutrients.plot <- plot_grid(Nitrogen.plot,Phosphorus.plot,ncol = 1)
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
print(nutrients.plot)
```





9. What do these plots tell you about nutrient concentrations over time? How might this relate to your previous plot of hypoxia events?

> Over time the nutrient concentrations have decreased, for both nitrogen and phosphorus. As we know that hypoxia is associated with high nutrient concentrations (biomass consumes oxygen), this nutrient concentration trend is positively related to hypoxia events, further verifying the theory.

Discharge and Temperature

10. Turbulent flow in rivers mixes oxygen into the water column. As discharge decreases, water moves slower, and oxygen diffuses slower into the water from the atmosphere. Download and reformat the daily discharge data for New Hope Creek (function `readNWISdv()` , site `02097314` , parameter `00060`).

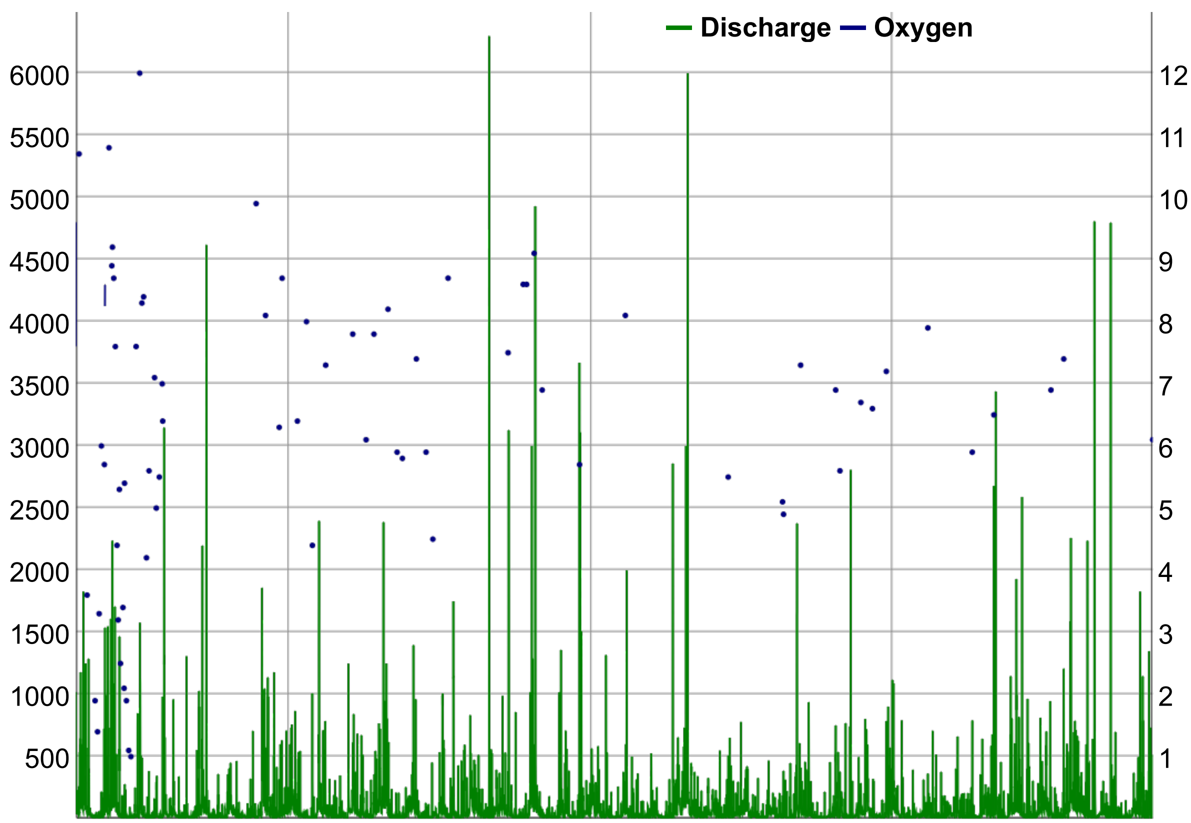
```
discharge.raw <- readNWISdv(site = "02097314", parameterCd = c("00060"), startDate = begi
n.date, endDate = end.date)
str(discharge.raw, give.attr = FALSE)
```

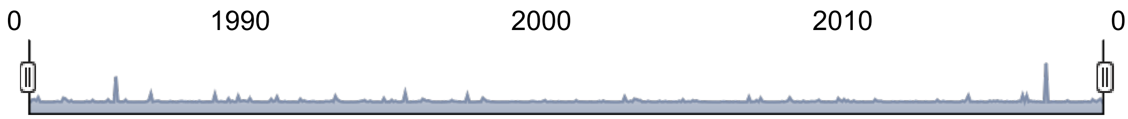
```
## 'data.frame':    13058 obs. of  5 variables:
##  $ agency_cd       : chr  "USGS" "USGS" "USGS" "USGS" ...
##  $ site_no         : chr  "02097314" "02097314" "02097314" "02097314" ...
##  $ Date            : Date, format: "1982-12-16" "1982-12-17" ...
##  $ X_00060_00003   : num  271 1020 502 157 106 86 68 58 53 50 ...
##  $ X_00060_00003_cd: chr  "A" "A" "A" "A" ...
```

```
discharge.dat <- discharge.raw %>%
  select(Date = Date, Discharge_cfs = X_00060_00003)
```

11. Create a dygraph of discharge and oxygen concentrations. You should have discharge on the y axis and oxygen concentration on the y2 axis. Set your y-axes limits so that they don't change as you zoom in and out. (hint: use `dyAxis(..., valueRange = c(0, 6500))` for discharge & `dyAxis(..., valueRange = c(0, 13))` for oxygen).

```
Oxygen <- with(O2.dat, xts(x = Value, order.by = Date))
Discharge <- with(discharge.dat, xts(x = Discharge_cfs, order.by = Date))

DyDat <- cbind(Oxygen, Discharge)

dygraph(DyDat) %>%
  dySeries("Oxygen", axis = "y2") %>%
  dyAxis(name = "y", label = "Discharge (cfs)", valueRange = c(0,6500)) %>%
  dyAxis(name = "y2", label = "Oxygen concentration (mg / L)", valueRange = c(0, 13)) %>%
  dyRangeSelector()
```





12. Do hypoxic events happen in the wet or the dry season? How many days after storm events (of at least 100 cfs) do they occur?

Based on the dygraph, hypoxic events happen in the dry season (when discharge is relatively on low levels). The closest interval between a storm event and a hypoxic event is that on Aug 12, 1984 the discharge is 237 cfs, and on Aug 22 the oxygen concentration is 1.9 mg/L. But other than that, it seems that usually there are several months between a storm event and a hypoxic event.

13. Temperature also impacts the solubility of oxygen in water. More oxygen can dissolve when temperatures are *lower*. How would you analyze the effects of temperature on dissolved oxygen concentrations? What data would you need? How would you separate the effects of temperature from the effects of flow?

I would need water temperature data of New Hope Creek for the same time period, and plot against dissolved oxygen concentration, to observe whether there seems to be correlations and trends. Moreover, I could run a regression to see whether temperature affects oxygen concentration. In order to separate the effects of temperature from those of flow, a regression model would come in handy, in that it will show effects on oxygen concentration for both factors, and we can also calculate whether temperature and flow have correlation within themselves.

14. What do you think is the biggest driver of hypoxia in New Hope Creek? What data do you use to support your conclusion?

I think high nutrient concentration is the biggest driver of hypoxia in New Hope Creek. The time period when concentration of N and P are high is almost exactly the same time when we observe hypoxia incidents, which suggests that these two factors are highly correlated. Also we did not observe apparent relationship between discharge level and oxygen concentration.

Reflection

15. What are 2-3 conclusions or summary points about river water quality you learned through your analysis?

> 1. High nutrient level leads to low level of oxygen.
> 2. This problem of high nutrient level in water has been improved, at least at New Hope Creek.

16. What data, visualizations, and/or models supported your conclusions from 15?

> The graph showing oxygen concentration trend along time.

17. Did hands-on data analysis impact your learning about water quality relative to a theory-based lesson? If so, how?

> Yes, by seeing the correlation between nutrient level and oxygen concentration on graphs, I feel that I would memorize them more vividly.

18. How did the real-world data compare with your expectations from theory?

> I was expecting a worsening trend when it comes to nutrient level in water. However, it seems that we are doing well in terms of controlling nutrient level in water.