Assignment 1: Introduction

Walker Grimshaw

OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis on introductory material.

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 (marked with >).
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "FILENAME") prior to submission.

The completed exercise is due on 2019-09-04 before class begins.

Course Setup

1. Post the link to your forked GitHub repository below. Your repo should include one or more commits and an edited README file.

    Link: https://github.com/wgrimshaw/Hydrologic_Data_Analysis

2. Complete the Consent Form in Sakai. You must choose to either opt in or out of the research study being conducted in our course.

Did you complete the form? (yes/no)

> Yes

Course Project

3. What are some topics in aquatic science that are particularly interesting to you?

    > ANSWER: I am most interested in the aspects of aquatic science that affect human systems. I am perhaps most interested in water quantity. Though water quality issues are also intriguing, such as the fate of contaminants in natural systems that affect human health and are targets of removal by drinking water treatment systems. This could be as simple as organic matter that can cause disinfection by-products or emerging contaminants, like pharmaceuticals or PFAS. Such contaminants may also be introduced by wastewater treatment systems not designed to remove emerging contaminants.

4. Are there specific people in class who you would specifically like to have on your team?

    > ANSWER: Nope.

5. Are there specific people in class who you would specifically *not* like to have on your team?

    > ANSWER: Rachel Bash. She is a fantastic teammate, but we are already in an MP group together and should probably branch out.

Data Visualization Exercises

6. Set up your work session. Check your working directory, load packages `tidyverse`, `dataRetrieval`, and `lubridate`. Set your ggplot theme as theme_classic (you may need to look up how to set your theme).

```r
# keep warnings from appearing in knitted pdf
knitr::opts_chunk$set(warning = FALSE)

getwd()
```

```
## [1] "C:/Users/walke/OneDrive/Documents/Duke/Courses/Fall_2019/Hydrologic_Data_Analysis/Assignments"
```

```r
library(tidyverse)
library(dataRetrieval)
library(lubridate)

## install viridis for later use
# install.packages("viridis")
library(viridis)

# set theme to theme_classic
theme_set(theme_classic())
```

7. Upload discharge data for the Eno River at site 02096500 for the same dates as we studied in class (2009-08-01 through 2019-07-31). Obtain data for discharge and gage height (you will need to look up these parameter codes). Rename the columns with informative titles. Imperial units can be retained (no need to change to metric).

```r
# Import data
# discharge code = 00060 cubic feet per second
# gage height code = 00065 feet
EnoDischarge <- readNWISdv(siteNumbers = "02096500", # selecting eno river
                     parameterCd = c("00060","00065"), # discharge (ft3/s)
                     startDate = "2009-08-01",
                     endDate = "2019-07-31")

#rename columns
names(EnoDischarge)[4:7] <- c("Discharge.CFS", "Discharge.Approval",
                              "Gage.Feet", "Gage.Approval")
```

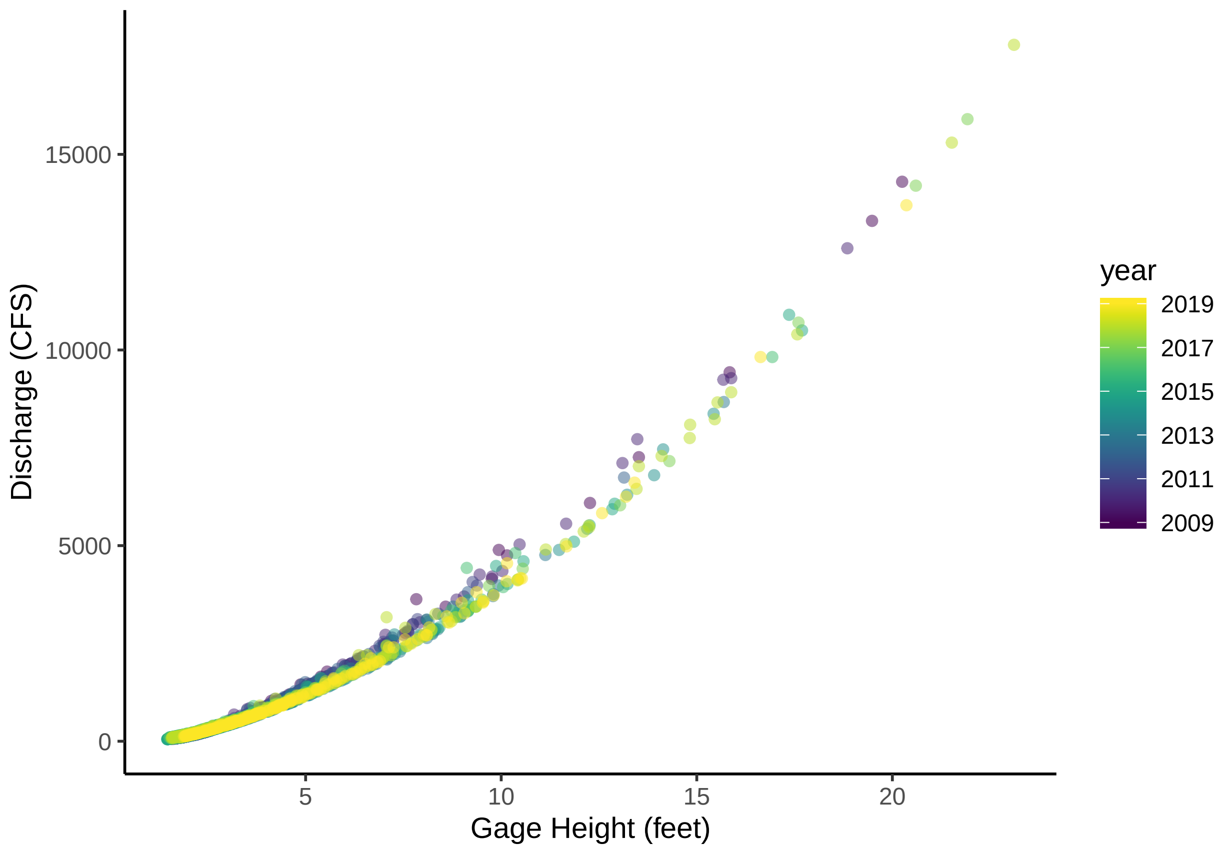
8. Add a "year" column to your data frame (hint: lubridate has a `year` function).

```r
## add year column using mutate function
EnoDischarge <- mutate(EnoDischarge, year = year(Date))
```

9. Create a ggplot of discharge vs. gage height, with gage height as the x axis. Color each point by year. Make the following edits to follow good data visualization practices:

- Edit axes with units
- Change color palette from ggplot default
- Make points 50 % transparent

```r
ggplot(data = EnoDischarge, aes(x = Gage.Feet, y = Discharge.CFS,
                                color = year)) +
  geom_point(alpha = 0.5) +
  labs(x = "Gage Height (feet)", y = "Discharge (CFS)") +
  scale_color_viridis()
```



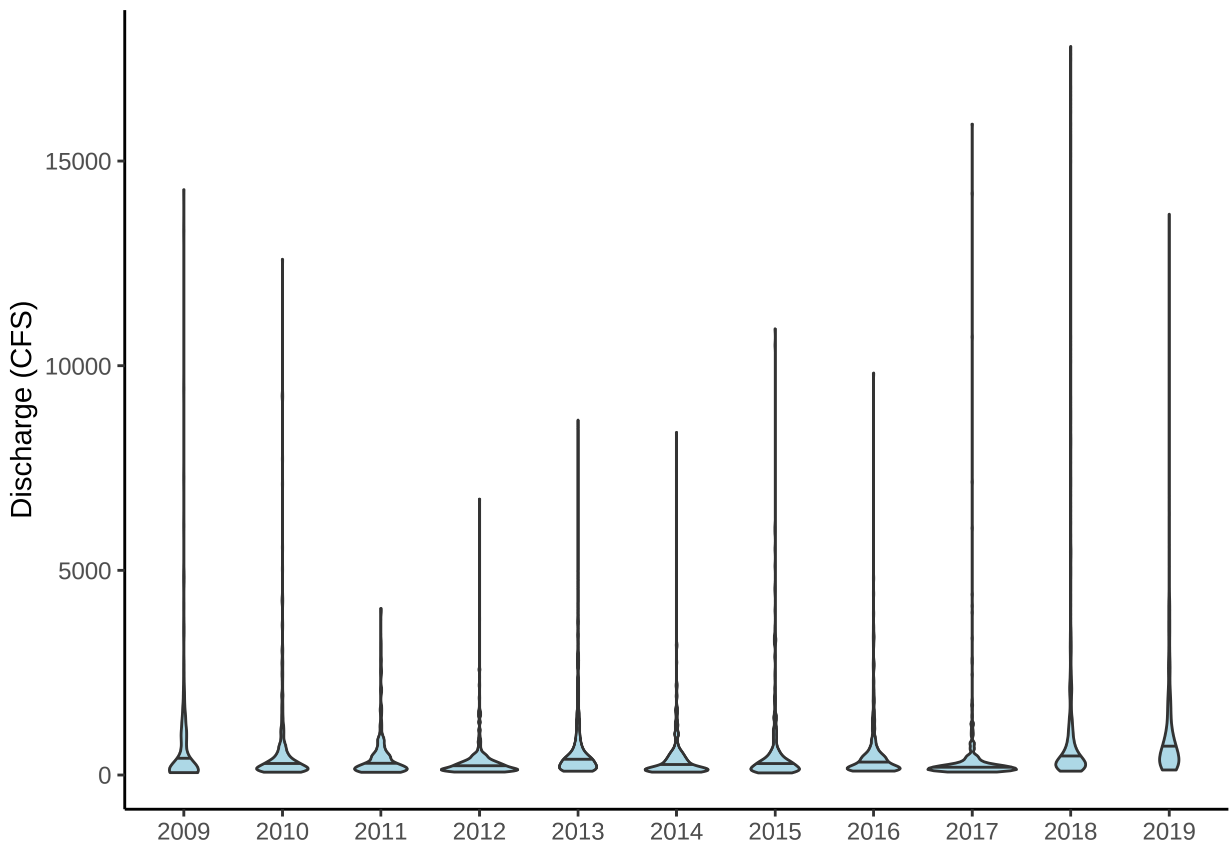
10. Interpret the graph you made. Write 2-3 sentences communicating the main takeaway points.

    ANSWER: At the Eno River site in question, there is an exponential relationship between gage height and discharge of the river. At the majority of measurement times, the discharge is below 5,000 CFS, but there are some discharge measurements nearly as high as 20,000 CFS. The figure may indicate more days with high discharge after 2014 than before 2014.

11. Create a ggplot violin plot of discharge, divided by year. (Hint: in your aesthetics, specify year as a factor rather than a continuous variable). Make the following edits to follow good data visualization practices:

- Remove x axis label
- Add a horizontal line at the 0.5 quantile within each violin (hint: draw_quantiles)

```
ggplot(data = EnoDischarge) +
  geom_violin(aes(y = Discharge.CFS, x = factor(year)),
              draw_quantiles = 0.5, fill = "light blue") +
  labs(x = NULL, y = "Discharge (CFS)")
```



12. Interpret the graph you made. Write 2-3 sentences communicating the main takeaway points.

    ANSWER: In all years, the discharge of the Eno River is less than approximately 500 CFS the majority of days. The maximum discharge, however, changes substantially from year to year, with a minimum in 2011 less than 5,000 CFS and a maximum in 2018 above 17,000 CFS.