Assignment 1: Introduction

Theo Cai

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/gyc4/Hydrologic_Data_Analysis.git

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'm interested in pollutants, their varying residence times in the water cycle, and their effects on wildlife + humans. Water quality in general is interesting to me, as well.

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

   ANSWER: It would be cool to work with Ethan Ready, but other than him I have no preferences. Maybe someone well-versed in R would be helpful for me, who's pretty rusty at it.

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

   ANSWER: Nope!

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).

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

```r
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
siteNumber <- "02096500"
parameterCd <- c("00060","00065")
statCd <- "00001"
startDate <- "2009-08-01"
endDate <- "2019-07-31"

discharge <- readNWISdv(siteNumber, parameterCd, startDate, endDate)

colnames(discharge)[colnames(discharge)=="X_00060_00003"] <- "Discharge"
colnames(discharge)[colnames(discharge)=="X_00065_00003"] <- "GageHeight"
```

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

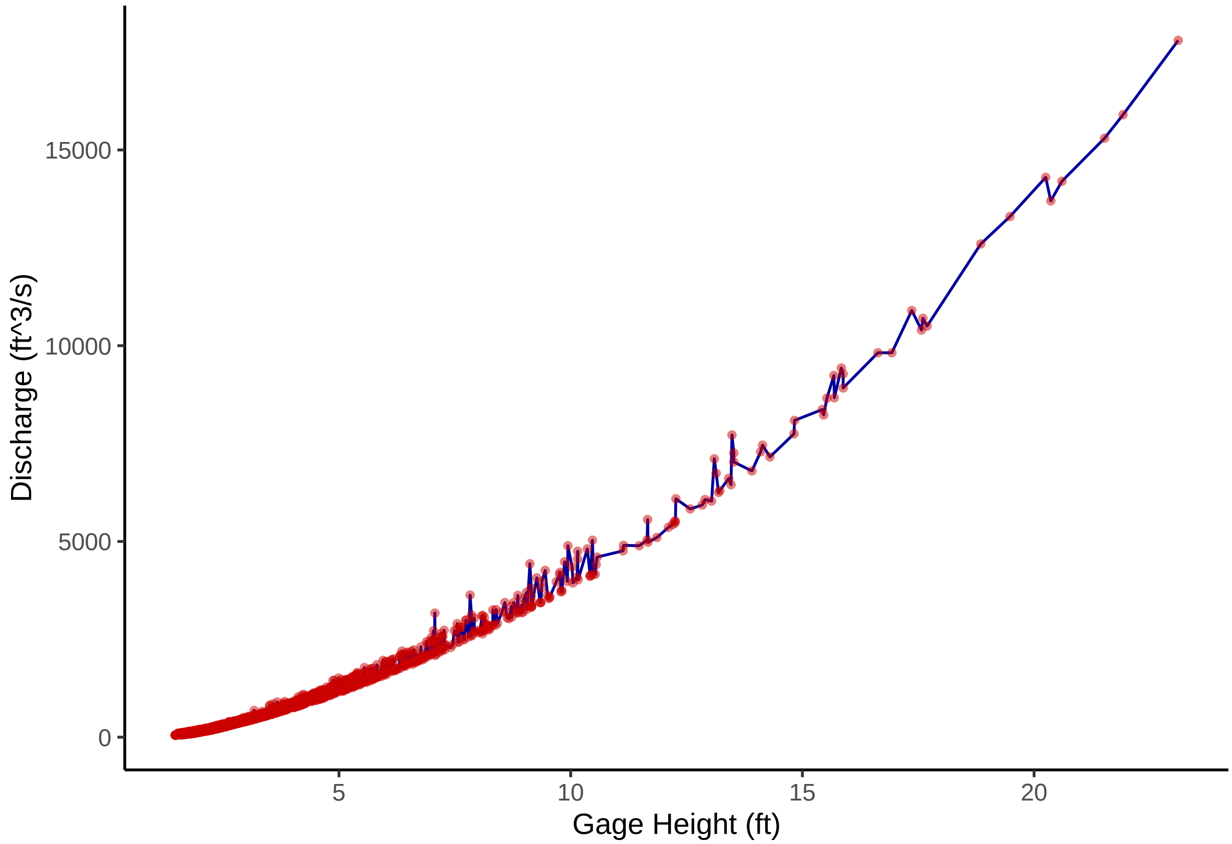
```r
discharge$year <- year(as.Date(discharge$Date, origin = '2009-08-01'))
```

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
library(ggplot2)

theme_set(theme_classic())

dischargePlot <-
  ggplot(discharge, aes(x = GageHeight, y = Discharge)) +
  geom_line(aes(group = 1), colour = "#000099") +
  geom_point(alpha = 0.50, size = 1, colour = "#CC0000") +
  xlab("Gage Height (ft)") +
  ylab("Discharge (ft^3/s)")
print(dischargePlot)
```

```
## Warning: Removed 8 rows containing missing values (geom_path).

## Warning: Removed 9 rows containing missing values (geom_point).
```



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

    ANSWER: It seems that, the lower the gage height, the more discharge measurements were taken, as well as the lower the measurement. Perhaps this is an indication that deeper water moves slower than water near the top of the Eno. This seems relatively intuitive.

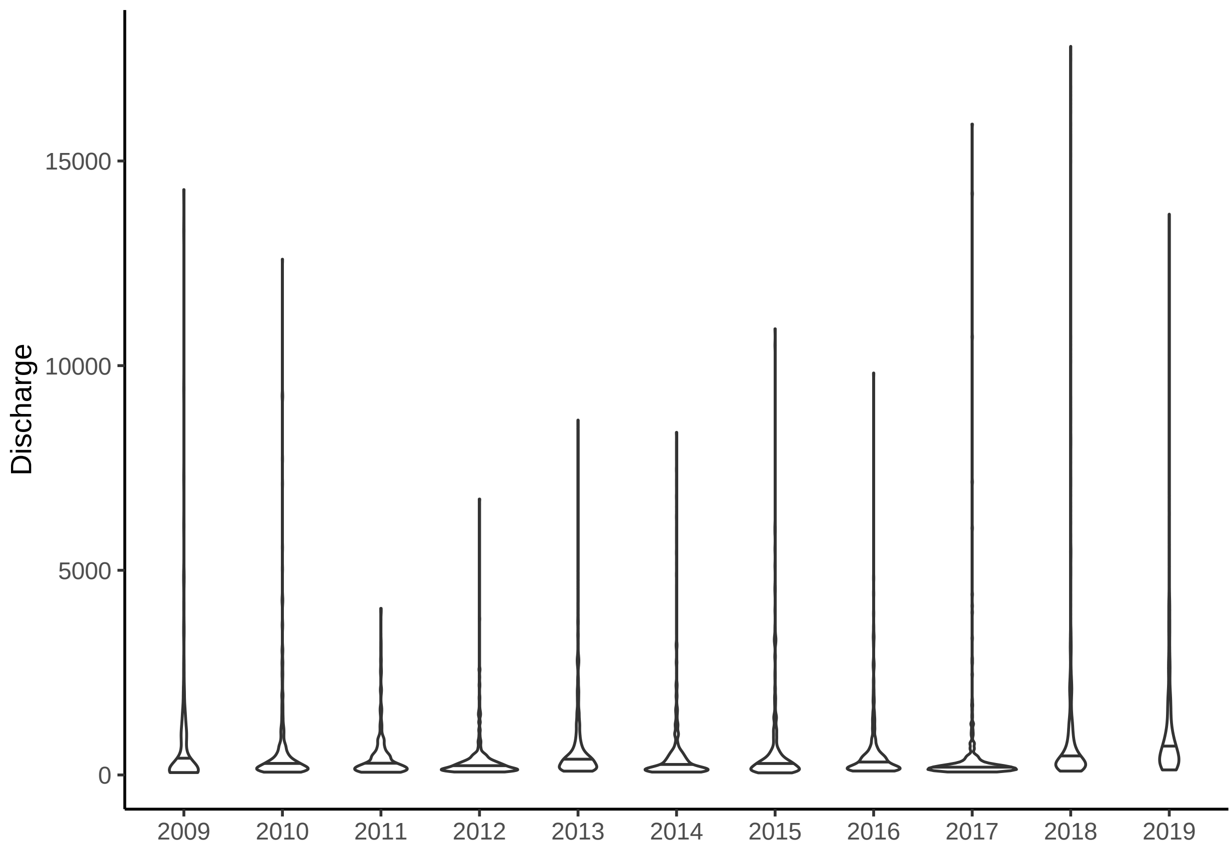
11. Create a ggplot violin plot of discharge, divided by year. (Hint: in your aesthetics, specify year as a factor rather than a continuous variable). Make the following edits to follow good data visualization practices:

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

```
library(ggplot2)
discharge$year <- as.factor(discharge$year)
dischargeviolinPlot <-
  ggplot(discharge, aes(x = year, y = Discharge)) +
  geom_violin(trim = TRUE, draw_quantiles = c(0.5)) +
  xlab(NULL)
print(dischargeviolinPlot)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_ydensity).

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values
```

```
## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties)): collapsing to
## unique 'x' values
```



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

    ANSWER: It seems the range of discharge is very great and maybe violin graphs are not the best way to visualize the data in its current state. Other than that, however, it seems that discharge averages and quartiles oscillate through the years - as evidenced by the violin plots cycling from stretched to compressed to stretched again as we move through the decade. Similarly, the highest discharge of each year does not remain the same, and seems to cycle.