Assignment 1: Introduction

Caroline Watson

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/cwatson1013/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: Some topics in aquatic science that particularly interest me are: limnology and hydrology, particularly anything dealing with water quality.

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

   ANSWER: Rachel Bash, Felipe Ray, Walker Grimshaw

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

   ANSWER: Gaby Garcia

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).

```
#getting my working directory and loading packages
getwd()
```

```
## [1] "/Users/carolinewatson/Documents/Fall 2019/Hydrologic_Data_Analysis/Assignments"
```

```
suppressMessages(library(tidyverse))
library(dataRetrieval)
library(lubridate)
library(RColorBrewer)

#settting ggplot theme to theme_classic
caroline_theme <- theme_classic()
theme_set(caroline_theme)
```

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).

```
#importing discharge data for Eno River site 02096500
EnoDischarge <- readNWISdv(siteNumbers = "02096500",
                     parameterCd = c("00060", "00065"),
                     startDate = "2009-08-01",
                     endDate = "2019-07-31")

#renaming columns
names(EnoDischarge)[4:7] <- c("Discharge", "Approval.Code.D", "Gage.Height", "Approval.Code.GH")
```

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

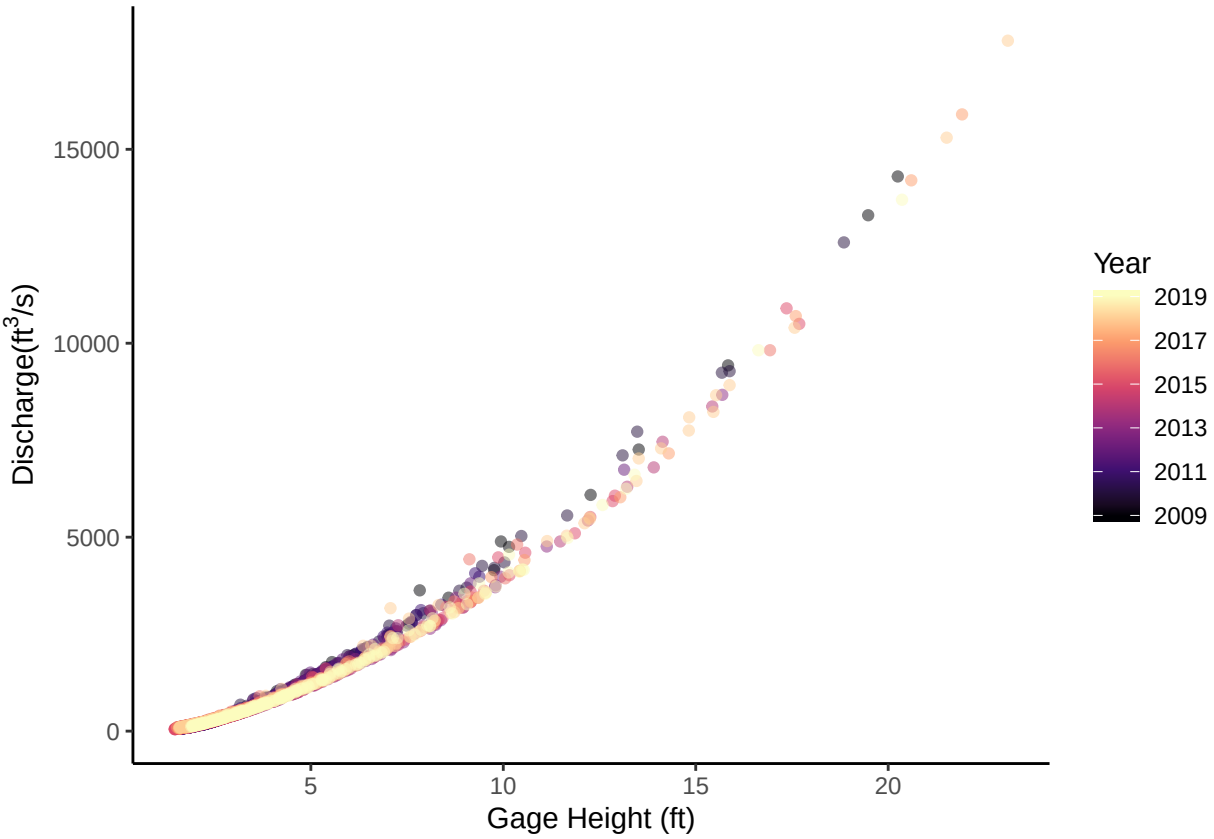
```
#add a "year" column to the data frame
EnoDischarge_summary <- EnoDischarge %>%
  mutate(Year = year(Date))
```

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

```
#creating ggplot of discharge vs. gage height
Eno_Plot <- ggplot(EnoDischarge_summary, aes(x = Gage.Height, y = Discharge, color = Year)) +
  geom_point(alpha = 0.5) +
  labs(x = expression(paste("Gage Height (ft)")),
       y = expression(paste("Discharge", "(ft"^{3}*"/s)"))) +
   scale_color_viridis_c(option = "magma")

print(Eno_Plot)
```



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

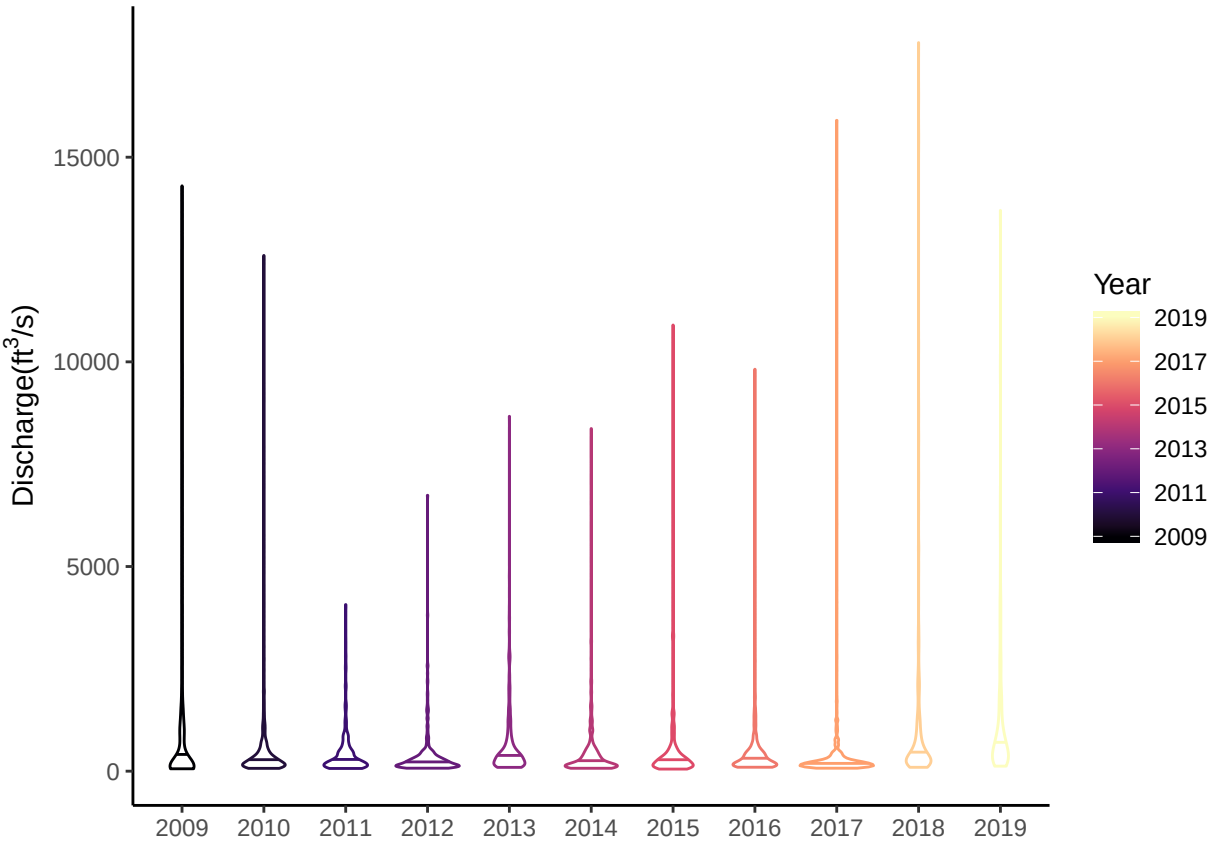
    ANSWER: This graph shows the relationship between gage height (ft) and discharge (cfs) at the Eno River for measurements taken between 2009 and 2019. From the plot, as gage height (ft) increases, discahrge (cfs) generally increases as well. Also, the highest discharge (cfs) value and highest gage height (ft) occured between 2017 and 2019.

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)

```
Eno_Plot_violin <- ggplot(EnoDischarge_summary) +
  geom_violin(draw_quantiles = c(0.5), aes(x = as.factor(Year), y = Discharge, color = Year)) +
  labs(x = "Year", y = expression(paste("Discharge", "(ft"^{3}*"/s)"))) +
  theme(axis.title.x=element_blank()) +
  scale_color_viridis_c(option = "magma")


print(Eno_Plot_violin)
```



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

    ANSWER: This graph shows the distribution of discharge measurements at the Eno River from 2009 to 2019. This graph also shows how varied hydrologic data is. From this plot, we can see that in 2011, discharge (cfs) was lower compared to other years, whereas in 2018, the highest flow (cfs) was recorded.