Assignment 1: Introduction

Rachel Bash

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/rachelbash/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 interested in hydrology and groundwater, specifically as it relates to the arid west.

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

   ANSWER: Caroline and Walker

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

   ANSWER: Lindsay (only because we have all the same classes together and are already in a project class together)

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).

```
getwd()
```

```
## [1] "C:/Users/19524/Documents/DUKE/Hydrologic Data Analytics/Hydrologic_Data_Analysis/Assignments"
```

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

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
EnoDischarge <- readNWISdv(siteNumbers = "02096500",
                     parameterCd = c("00060","00065"),
                     startDate = "2009-08-01",
                     endDate = "2019-07-31")
names(EnoDischarge)[4:7] <- c("Discharge", "Approval.Code", "Gage.Ht", "Approval.Code2")
```

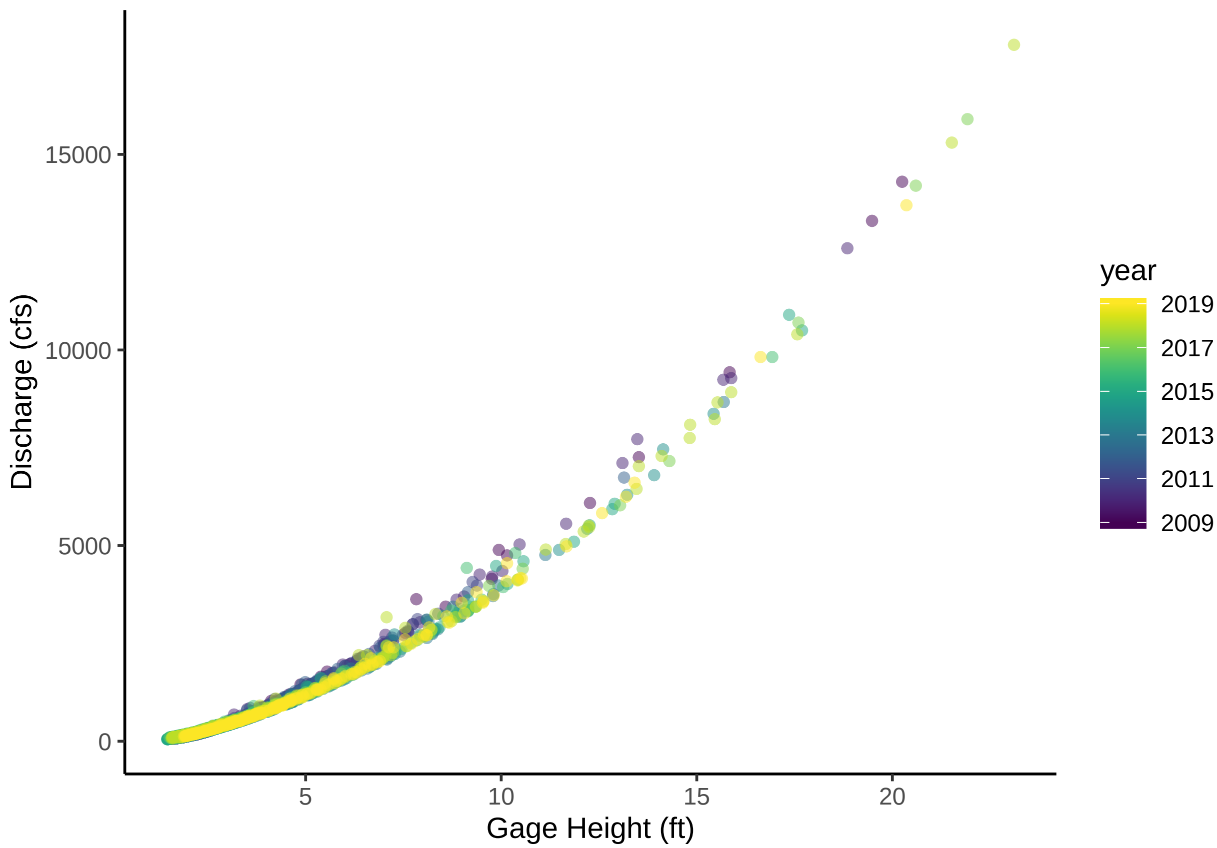
8. Add a "year" column to your data frame (hint: lubridate has a `year` function).

```r
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(EnoDischarge, aes(x=Gage.Ht, y=Discharge, color=year)) +
  geom_point(alpha=0.5) +
  scale_color_viridis(option="viridis") +
  labs(x= "Gage Height (ft)", y=
         "Discharge (cfs)")
```



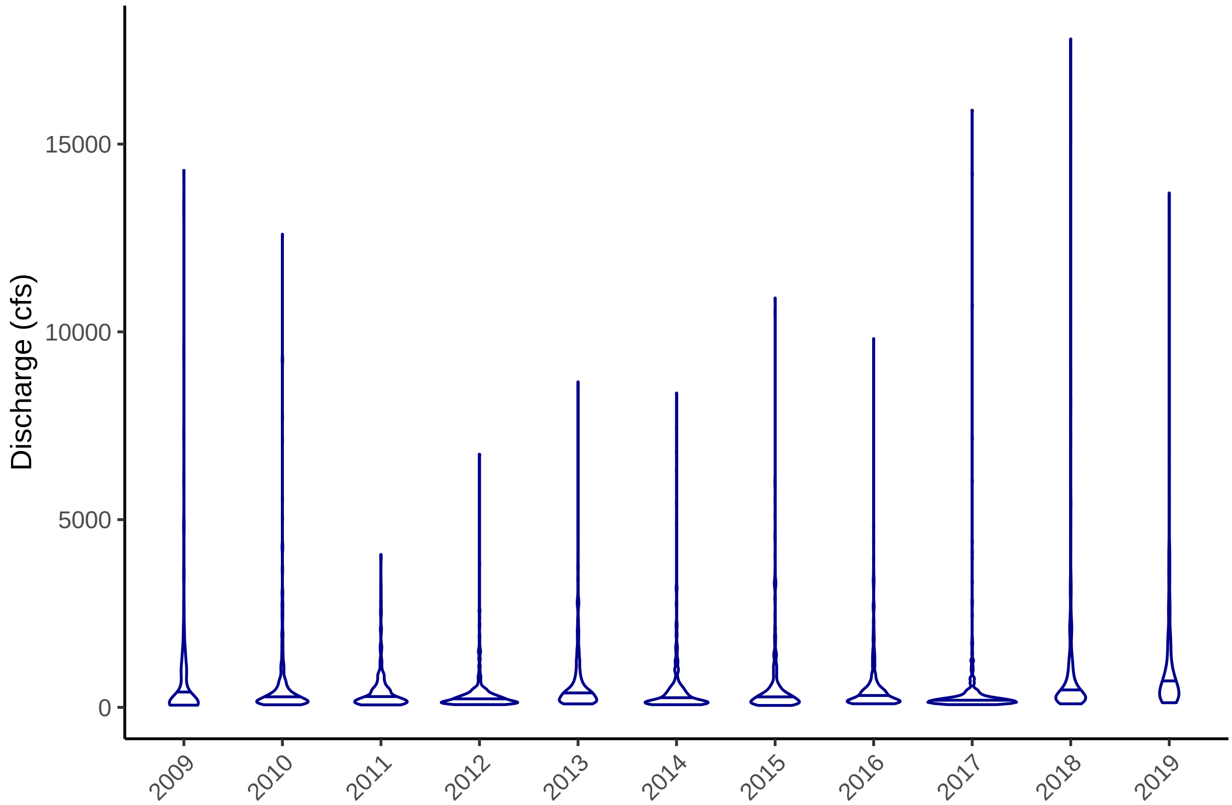
10. Interpret the graph you made. Write 2-3 sentences communicating the main takeaway points.

    ANSWER: Discharge and gage height have a strong positive relationship, meaning that discharge increases as gage height increases. Most data points occur in the lower range of gage height and discharge values, as there is a large concentration of points in the lower lefthand corner of the graph. There doesn't seem to be a difference in relationship by year, as all of points within various years follow the same curve.

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(EnoDischarge,aes(x=as.factor(year), y=Discharge)) +
  geom_violin(draw_quantiles = 0.5, color="darkblue") +
  labs(x="",y="Discharge (cfs)")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



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

    ANSWER: Average Discharge over the years stayed relatively constant, with the vast majority of discharge values hovering just above 0. There are a handful of extremely high discharge values within each year, with 2018 having the highest recorded discharge value, and 2011 having the lowest maximum recorded discharge value.