Assignment 1: Introduction

Keith Bollt

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/keithbollt/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: water policy, Freshwater fisheries (I fly fish), land use and its effects on hydrographs and stream conditions, saltwater fisheries, macroinvertebrates

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

   ANSWER: Any of the second year WRMs, just because I know them, and also know they are all motivated and hardworking students.

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

   ANSWER: Not yet!

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
getwd()
#install.packages(tidyverse)
#install.packages(dataRetrieval)
#install.packages(lubridate)
```

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

mytheme <- theme_set(theme_classic())
mytheme
```

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"),  # discharge (ft3/s), #gage height (feet)
                     startDate = "2009-08-01",
                     endDate = "2019-07-31")

names(EnoDischarge)[4:7] <- c("Discharge", "Approval Code1", "GageHeight", "Approval Code2")
```

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

```r
EnoDischarge$newcolumn <- year(EnoDischarge$Date)
EnoDischarge$newcolumn <- NULL    # I messed up
EnoDischarge$Year <- year(EnoDischarge$Date)
```

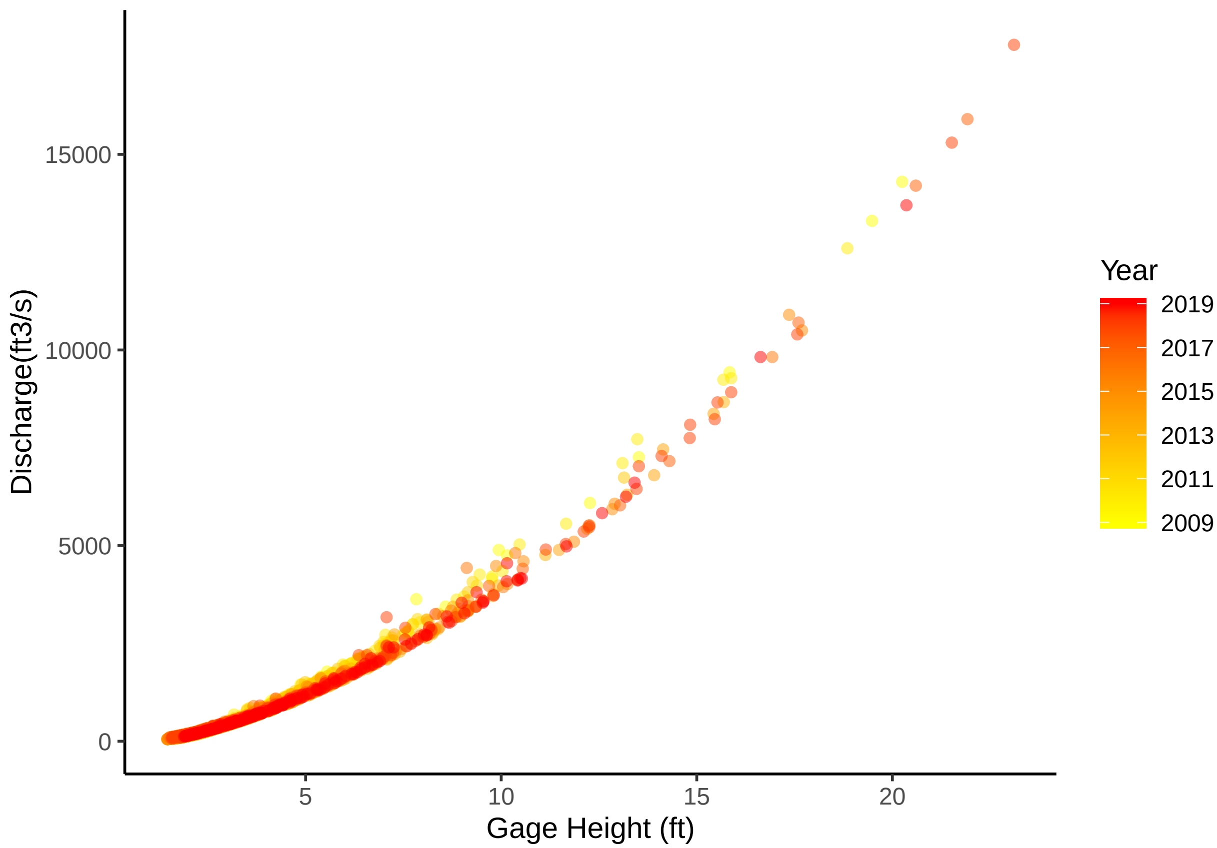
9. Create a ggplot of discharge vs. gage height, with gage height as the x axis. Color each point by year. Make the following edits to follow good data visualization practices:

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

```r
flowvsgage <- ggplot(EnoDischarge, aes(x= GageHeight, y = Discharge, color =Year))+
  geom_point(alpha = 0.5)+
  xlab("Gage Height (ft)")+
  ylab("Discharge(ft3/s)")+
  scale_color_gradient(low = "Yellow", high = "Red")


print(flowvsgage)
```

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



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

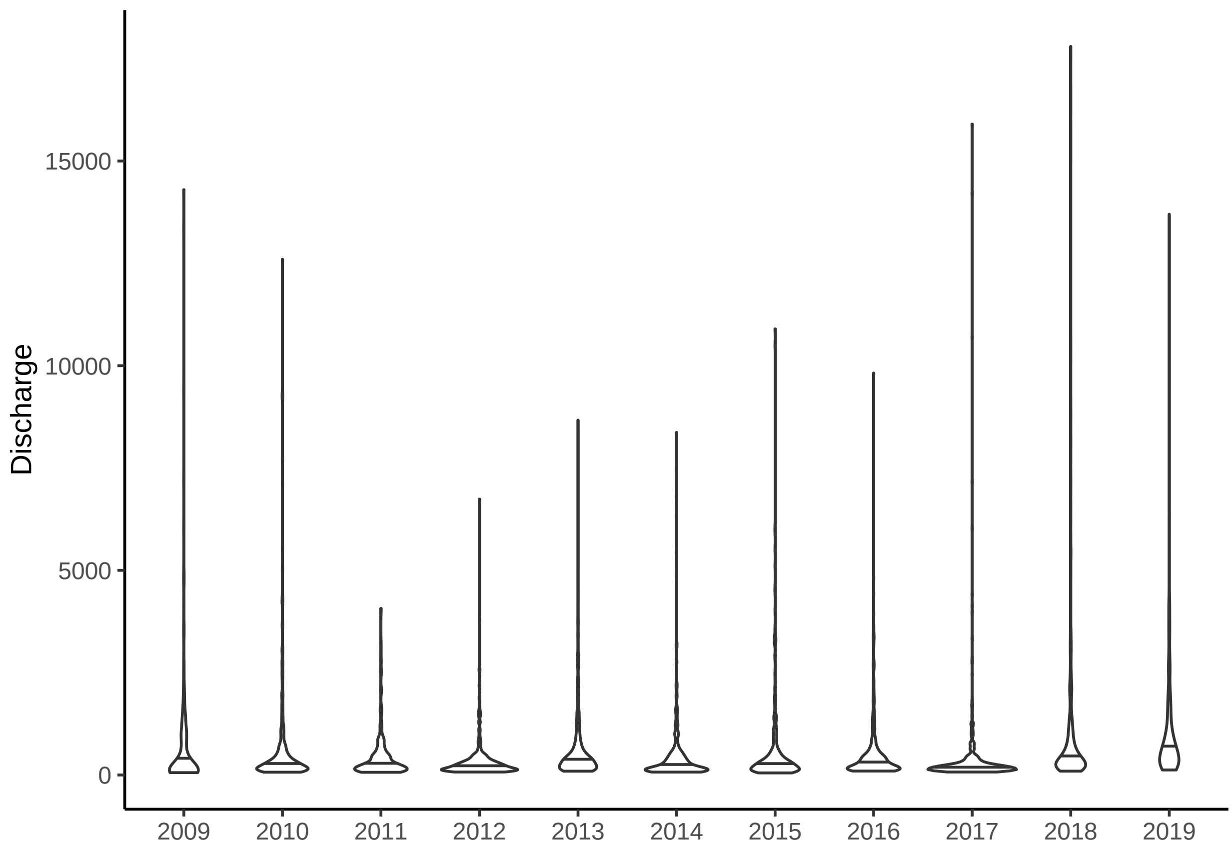
    ANSWER: There is a pretty strong correlation between gage height and discharge. This correlation is not quite linear, and this makes sense because stream beds are not uniform rectangular prisms. It also looks like maybe over time, a lower discharge is associated with a given gage height.

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)

```
worldssmallestviolin <- ggplot(EnoDischarge, aes(x = as.factor(Year), y = Discharge))+
  geom_violin(draw_quantiles = 0.5)+
  xlab(NULL)

print(worldssmallestviolin)
```



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

    ANSWER: Different years have different distributions of discharge. For example, 2017 had a larger percentage of low discharges than did 2018. Also, there is a wide variety ofmaximum yearly discharge. Basically, the hydrographs of the different years would look very different from one another.