Assignment 1: Introduction

Jake Greif

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/jakegreif/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'm interested in water quality issues. In particular, I'm interested in learning about human-induced water quality issues, and how they affect our water resoures.

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

   ANSWER: I would like to work with any of the other MEMs, but I know that you will intentionally disperse us among the groups. Hopefully there are enough of us so that there have to be multiple MEMs in some groups.

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

   ANSWER: No

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).

```
# Check Working Directory
getwd()
```

```
## [1] "/Users/jakegreif/Duke/Fall_2019/Hydrologic_Data_Analysis/Assignments"
```

```
# Install Packages
#install.packages(c("tidyverse", "dataRetrieval", "lubridate", "quantreg"))

# Load Packages
library(tidyverse)
library(dataRetrieval)
library(lubridate)
library(viridis)
library(quantreg)

# Set ggplot Theme
mytheme <- theme_classic()

theme_set(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).

```
# Obtain Stream Gage Data
siteNo <- "02096500"
pCode <- c("00060", "00065")
start.date <- "2009-08-01"
end.date <- "2019-07-31"

# Read Data Into Dataset
Haw <- readNWISdv(siteNumbers = siteNo,
                  parameterCd = pCode,
                  startDate = start.date,
                  endDate = end.date)

# Rename Columns
Haw <- renameNWISColumns(Haw)
```

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

```
# Add "Year" Column to Dataset
Haw$Year <- substring(Haw$Date, 1, 4)

class(Haw$Year)
```

```
## [1] "character"
```

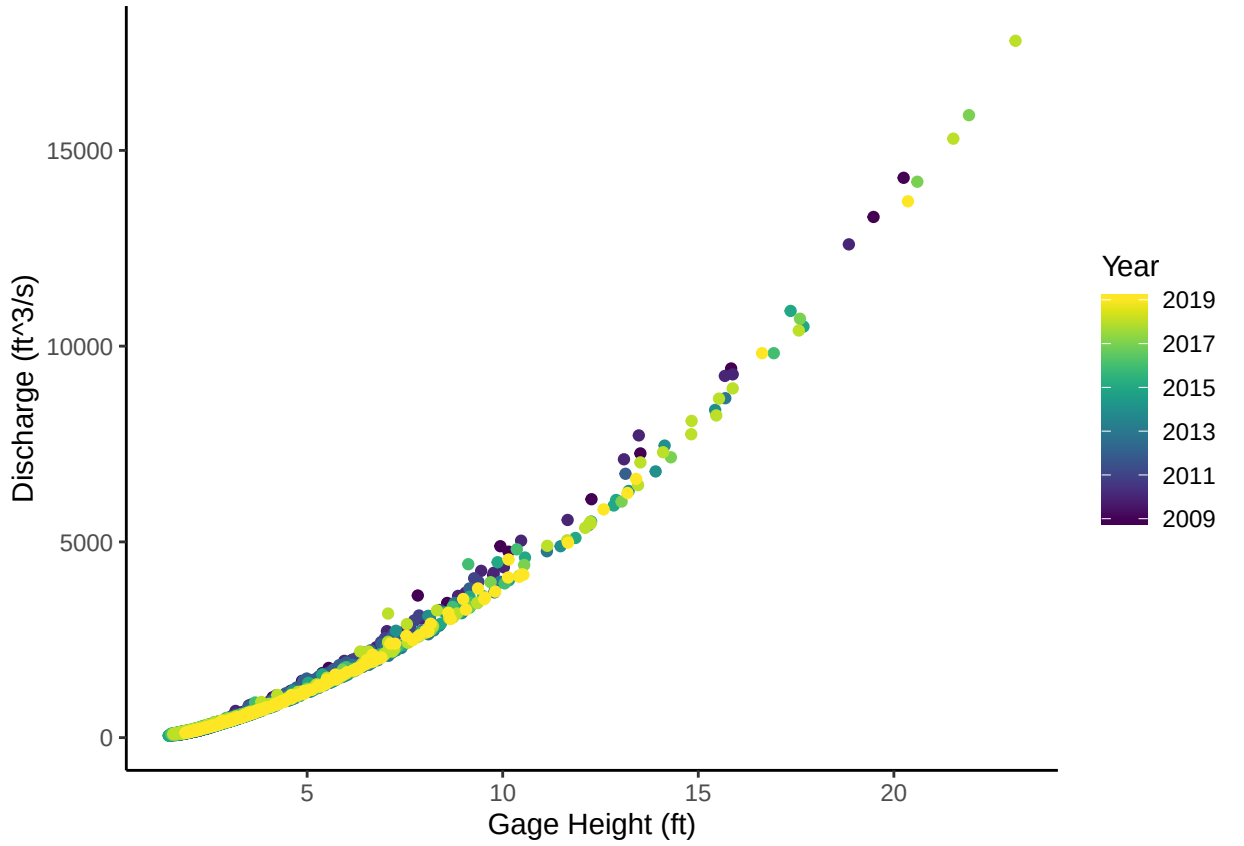
```
# Change Year to Numeric Class
Haw$Year <- as.numeric(Haw$Year)
```

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

```
# Discharge vs. Gage Height Scatterplot
ggplot(Haw, aes(x = GH, y = Flow, color = Year)) +
  geom_point() +
  scale_color_viridis() +
  ylab("Discharge (ft^3/s)") +
  xlab("Gage Height (ft)")
```

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



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

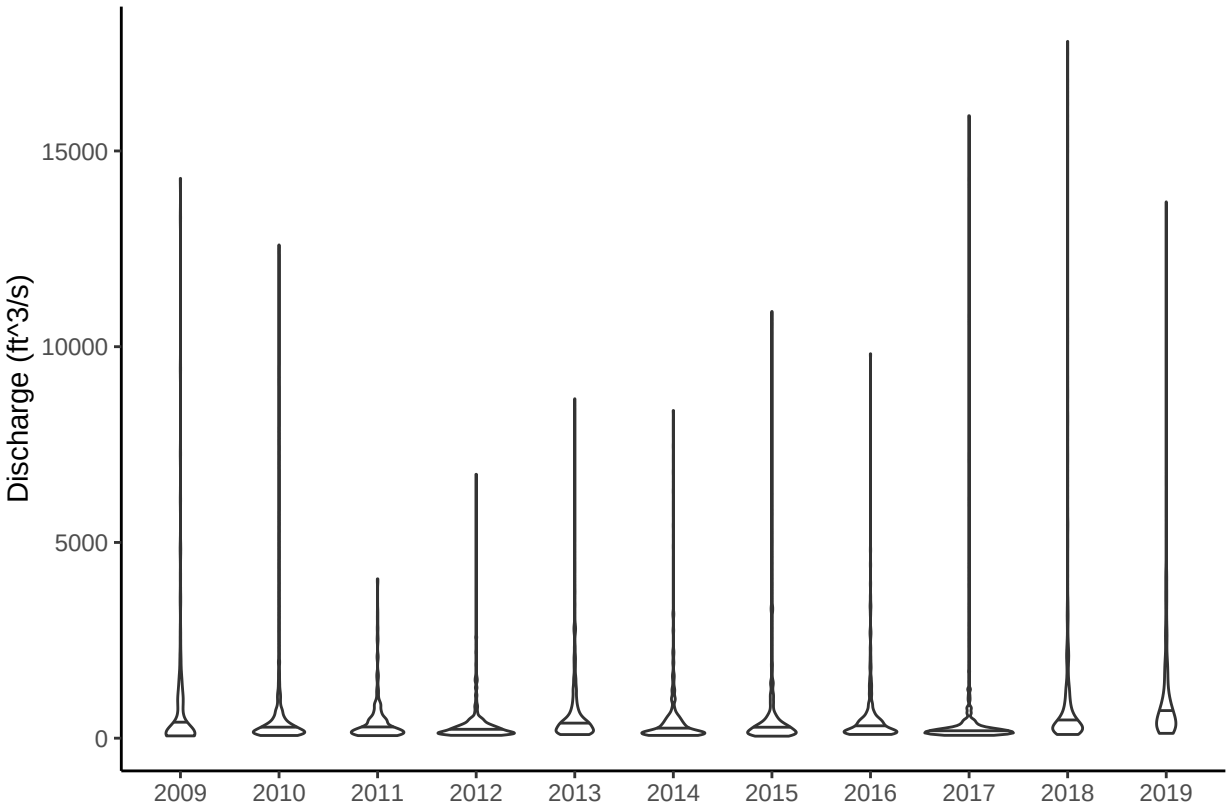
    ANSWER: The graph reveals that discharge and gage height are positively correlated. For the majority of the time period of the data, the discharge and gage height have been at or below 5,000 ft^3 and 10 ft, respectively. Though, the discharge and gage height are occasionally relatively high, which occurred more often during the second half of the period that the data represents.

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)

```r
# Discharge Violin Plot, by Year
ggplot(Haw, aes(y = Flow, x = factor(Year))) +
  geom_violin(draw_quantiles = 0.5) +
  scale_color_viridis() +
  ylab("Discharge (ft^3/s)") +
  xlab("")
```

```
## 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: Three of the four highest peak dischargers in the Haw river have occurred in the last three years. Additionally, the past two years have seen less stable discharge compared to previous years. 2012 and 2017 had the most stable discharges over the 10 year period, evident by the wide sections of each violin plot.