# Assignment 1: Introduction

*Gaby Garcia*

## 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.

   https://github.com/gdg12/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, I completed the form.

### Course Project

3. What are some topics in aquatic science that are particularly interesting to you?

   > ANSWER: Aquatic science is extremely interdisciplinary and there are many attractive topics. I'm extremely interested in honing my Master's Project study of hydrologic alterations to river and stream discharge, especially with the addition of insfrastructure development (specifically hydroelectric power). I'm also interested in environmental contaminant transport and its impact on water quality parameters such as dissolved oygen, conductivity, etc.

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

   > ANSWER: I am happy to have anyone on my team, but in the past I have worked well together with Felipe, Lindsay, Caroline, and Walker.

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

   > ANSWER: No, everyone seems very smart and capable!

### 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()
```

```
## [1] "/Users/gabrielagarcia/Desktop/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Assignments"
```

```r
setwd("~/Desktop/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Assignments")
library(tidyverse)
library(dataRetrieval)
library(lubridate)


gabytheme <- theme_classic(base_size = 16)+
  theme(plot.title=element_text(face="bold", size="20", color="hotpink4", hjust=0.5),
        axis.title=element_text(face="bold.italic", size="15", color="black"))
theme_set(gabytheme)
```

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
EnoDischargeAssignment1 <- readNWISdv(siteNumbers = "02096500",
                     parameterCd = c("00060","00065"),
                     startDate = "2009-08-01",
                     endDate = "2019-07-31")
```

##Print Site Info

```r
siteNumbers <- c("02096500")
siteINFO <- readNWISsite(siteNumbers)
print(siteINFO)
```

```
##   agency_cd  site_no                 station_nm site_tp_cd lat_va long_va
## 1      USGS 02096500 HAW RIVER AT HAW RIVER, NC         ST 360514  792158
##   dec_lat_va dec_long_va coord_meth_cd coord_acy_cd coord_datum_cd
## 1   36.08722   -79.36611             G            S          NAD83
##   dec_coord_datum_cd district_cd state_cd county_cd country_cd land_net_ds
## 1              NAD83          37       37       001         US          NA
##   map_nm map_scale_fc alt_va alt_meth_cd alt_acy_va alt_datum_cd   huc_cd
## 1 MEBANE           NA 471.69           R       0.01       NGVD29 03030002
##   basin_cd topo_cd                 instruments_cd construction_dt
## 1       07    <NA> NNNNYNNNNNNNNYNNNNNNNNNNNNNNNN              NA
##   inventory_dt drain_area_va contrib_drain_area_va tz_cd local_time_fg
## 1           NA           606                    NA   EST             Y
##   reliability_cd gw_file_cd nat_aqfr_cd aqfr_cd aqfr_type_cd well_depth_va
## 1           <NA>   NYNNNNNN        <NA>    <NA>         <NA>            NA
##   hole_depth_va depth_src_cd project_no
## 1            NA         <NA>         NA
```

##Use names function to get the names of an object

```r
names(EnoDischargeAssignment1)
```

```
## [1] "agency_cd"        "site_no"          "Date"
## [4] "X_00060_00003"    "X_00060_00003_cd" "X_00065_00003"
## [7] "X_00065_00003_cd"
```

##Information about the data frame attributes

```
names(attributes(EnoDischargeAssignment1))
```

```
## [1] "names"         "row.names"     "url"           "siteInfo"
## [5] "variableInfo"  "disclaimer"    "statisticInfo" "queryTime"
## [9] "class"
```

##The column names can be shortened and simplified using the renameNWISColumns function

```
EnoDischargeAssignment1Final<-renameNWISColumns(EnoDischargeAssignment1, p00060="Discharge",
                                                p_00060_00003_cd="Discharge_Qualification_Code",
p00065="GageHeight")
```

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

```
library(lubridate)
EnoDischargeAssignment1Final$Date<-as.Date(EnoDischargeAssignment1Final$Date,
                                           format="%m/%d/%y")
EnoDischargeAssignment1Final<-mutate(EnoDischargeAssignment1Final,
                                     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

```
library(viridis)
```

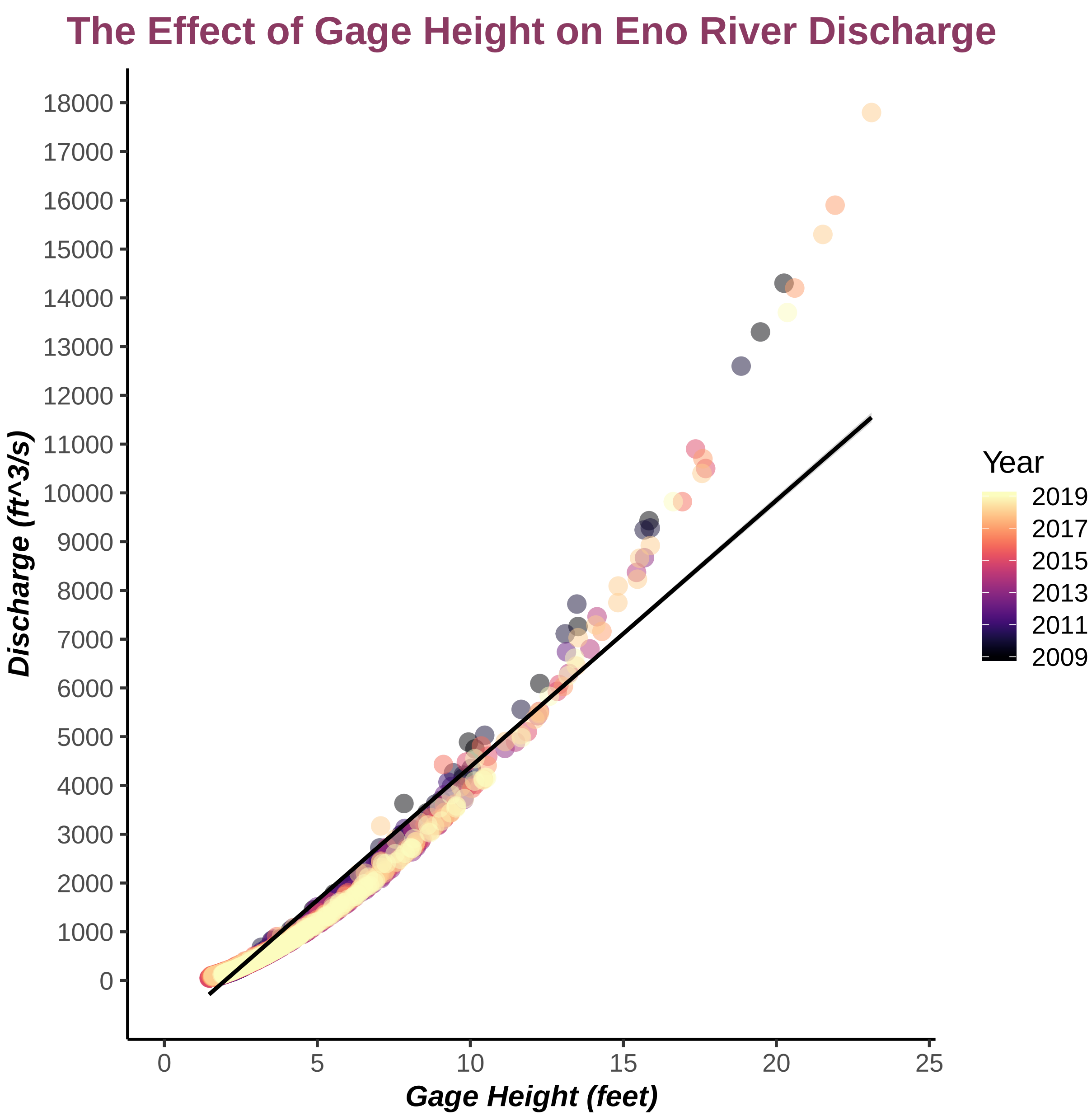
```
## Loading required package: viridisLite
```

```
library(RColorBrewer)
library(colormap)
library(wesanderson)
library(devtools)
```

```
## Loading required package: usethis
```

```
EnoDischargePlot<-ggplot(EnoDischargeAssignment1Final,
                         aes(x =GageHeight, y =Discharge, color=Year)) +
  geom_point(aes(x =GageHeight, y =Discharge), size=4, alpha=0.5)+
  geom_smooth(method="lm", color="black",
              SE=FALSE)+
  labs(title="The Effect of Gage Height on Eno River Discharge",
       x="Gage Height (feet)",
       y="Discharge (ft^3/s)")+
xlim(0, 24) +
  scale_color_viridis(option = "magma")+
  scale_y_continuous(breaks=c(0, 1000, 2000, 3000, 4000, 5000,
                              6000, 7000, 8000, 9000, 10000, 11000,
                              12000, 13000, 14000, 15000, 16000,
                              17000, 18000))

 print(EnoDischargePlot)
```




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

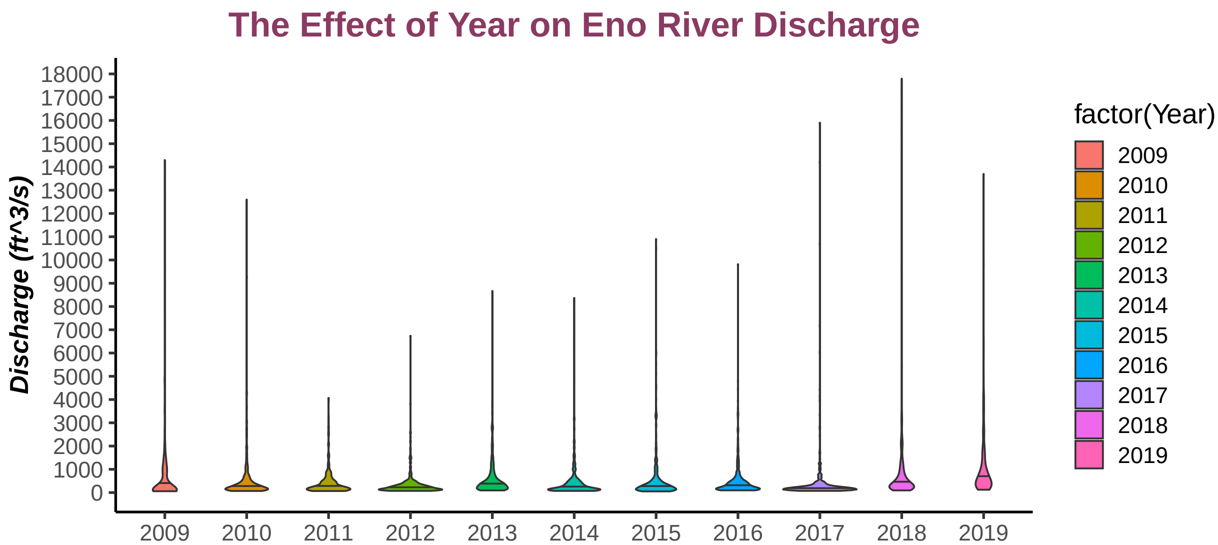
    ANSWER: Gage height is the elevation of the water surface at the streamgaging station, while discharge is the volume of water moving down a stream or river per unit of time at the stream gage. There is a positive correlation (although it is not a linear relationship) between gage height and river discharge.

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)

```
EnoDischargeViolin<-ggplot(EnoDischargeAssignment1Final,
                           aes(x =factor(Year), y =Discharge)) +
```

```
  geom_violin(draw_quantiles = c(0.5),aes(fill=factor(Year))) +
  labs(title="The Effect of Year on Eno River Discharge", y="Discharge (ft^3/s)")+ scale_y_continuous(b
                              4000, 5000, 6000, 7000, 8000, 9000,
                              10000, 11000, 12000, 13000, 14000,
                              15000, 16000, 17000, 18000))+
  theme(axis.title.x = element_blank())
print(EnoDischargeViolin)
```




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

    ANSWER Violin plots show the probability density of numeric data; the 0.5 quartile is the median. It appears the median discharges for all of the years have been similar and below 1000 ft^3/s. Where the violin plots is wide, there are a high density of points; where it is narrow, there are a low density of points. The year 2017 has the highest density of points with the least variability then the other years because most of the points are massed in the same region.