Assignment 1: Introduction

Gabi Richichi

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/grichichi/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 biogeochemistry and aquatic pollutant chemistry.

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

   ANSWER: other undergrads

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).

```
getwd()
```

```
## [1] "/Users/gabriellerichichi/Documents/5th year @ Duke/Stats/Hydrologic_Data_Analysis/Assignments"
```

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------- tidyverse 1.2.1
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------ tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dataRetrieval)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
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).

```
EnoDGH <- readNWISdv(siteNumbers = "02096500",
                     parameterCd = c("00060", "00065"),# discharge (ft3/s)
                     startDate = "2009-08-01",
                     endDate = "2019-07-31")

names(EnoDGH)[4:6] <- c("Discharge", "Approval.Code", "GageHeight")

attr(EnoDGH, "variableInfo")
```

```
##   variableCode           variableName              variableDescription
## 1        00060 Streamflow, ft&#179;/s Discharge, cubic feet per second
## 2        00065        Gage height, ft                Gage height, feet
##       valueType  unit options noDataValue
## 1 Derived Value ft3/s    Mean          NA
## 2 Derived Value    ft    Mean          NA
```

```
attr(EnoDGH, "siteInfo")
```

```
##                   station_nm  site_no agency_cd timeZoneOffset
## 1 HAW RIVER AT HAW RIVER, NC 02096500      USGS         -05:00
##   timeZoneAbbreviation dec_lat_va dec_lon_va       srs siteTypeCd    hucCd
## 1                  EST   36.08722  -79.36611 EPSG:4326         ST 03030002
##   stateCd countyCd network
## 1      37    37001    NWIS
```

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

```
#EnoDischarge$sampledate <- as.Date(EnoDischarge$sampledate, "%m/%d/%y") <- unnecessary because already

EnoDGH <- EnoDGH %>%
  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

```
library(viridis)
```

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

```
library(cowplot)
```

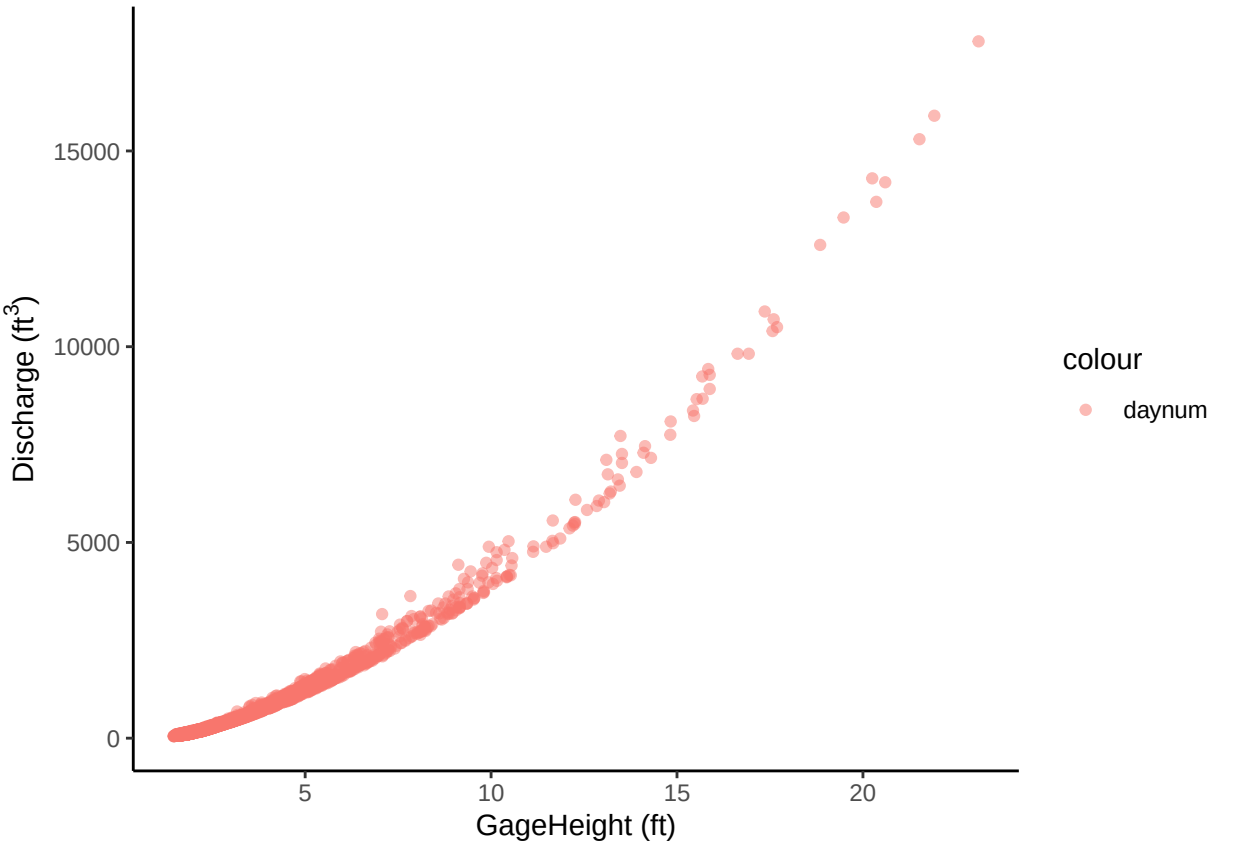
```
##
## ************************************************************
## Note: As of version 1.0.0, cowplot does not change the
##   default ggplot2 theme anymore. To recover the previous
##   behavior, execute:
##   theme_set(theme_cowplot())
## ************************************************************
##
## Attaching package: 'cowplot'
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
EnoPlot <-
   ggplot(EnoDGH, aes(x = GageHeight, y = Discharge, color = "daynum")) +
        #  geom_point(aes(color = Discharge)) +
         # scale_color_viridis(option = "D")
          geom_point(alpha = 0.5) +
         # scale_color_viridis(option = "D")
          labs(x = "GageHeight (ft)", y = expression("Discharge (ft"^3*")"))
print(EnoPlot)
```

```
## 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: As Gage Height increases, so does Discharge. The relationship is positive and slightly exponential. There is a greater number of data points at lower GageHeights, compared to at higher GageHeights.

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)

```
library(viridis)
library(hrbrthemes)
```

```
## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
##       Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
##       if Arial Narrow is not on your system, please see http://bit.ly/arialnarrow
```

```
library(ggplot2)
library(plotly)
```

```
##
## Attaching package: 'plotly'
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

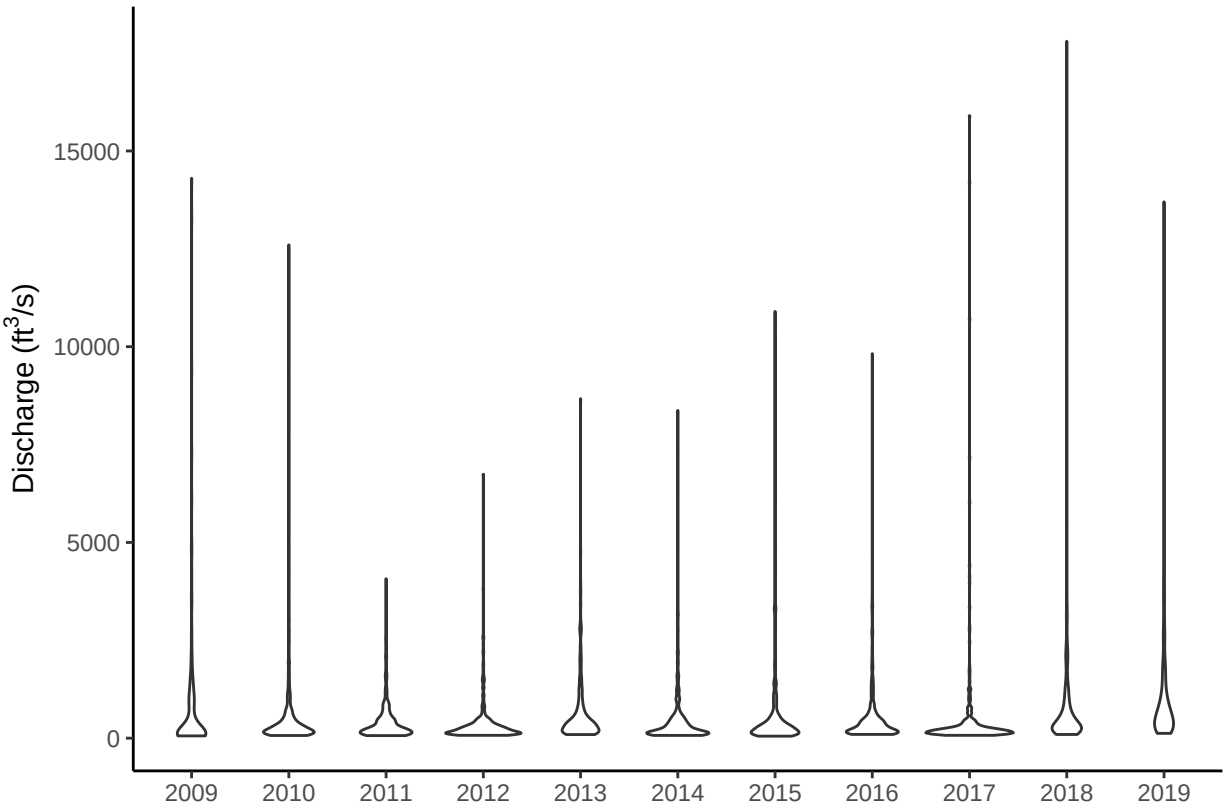
```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
EnoDGH$Year <- as.factor(EnoDGH$Year)
head(EnoDGH)
```

```
##   agency_cd  site_no       Date Discharge Approval.Code GageHeight
## 1      USGS 02096500 2009-08-01       186             A       2.13
## 2      USGS 02096500 2009-08-02       129             A       1.89
## 3      USGS 02096500 2009-08-03       123             A       1.85
## 4      USGS 02096500 2009-08-04       118             A       1.84
## 5      USGS 02096500 2009-08-05        84             A       1.66
## 6      USGS 02096500 2009-08-06       112             A       1.77
##   X_00065_00003_cd Year
## 1                A 2009
## 2                A 2009
## 3                A 2009
## 4                A 2009
## 5                A 2009
## 6                A 2009
```

```
EnoPlot2 <-
  ggplot(EnoDGH, aes(x = Year, y = Discharge)) +
  labs(x = "", y = expression("Discharge (ft"^3*"/s)")) +
  geom_violin()
  #draw_quantiles(0.5)
print(EnoPlot2)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_ydensity).
```



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

    ANSWER: Discharge decreases until 2011 and then increases thereafter. Perhaps an event transpired in 2011 that decreased discharge.