Assignment 1: Introduction

Nadia Swit

OVERVIEW

This exercise accompanies the lessons in Water Data Analytics 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 completing your assignment, fill out the assignment completion survey in Sakai.

Having trouble? See the assignment's answer key if you need a hint. Please try to complete the assignment without the key as much as possible - this is where the learning happens!

Target due date: 2022-01-18

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/nadiaswit/Water_Data_Analytics_2022.git

Data Visualization Exercises

2. Set up your work session. Check your working directory, load packages `tidyverse`, `dataRetrieval`, and `zoo`. Set your ggplot theme as theme_classic (you may need to look up how to set your theme).

```
#check working directory
getwd()
```

```
## [1] "C:/Users/nadsw/Desktop/SP22_Water_Data_Analytics/Water_Data_Analytics_2022"
```

```
#library
library(tidyverse)
library(dataRetrieval)
library(zoo)

#set theme
```

```
mytheme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right",
        legend.box.background = element_rect(color="black"), legend.background = element_blank(),
        plot.title=element_text(hjust = 0.5))
#makes font size bigger, puts legend on right, puts box around legend, centers title in middle
theme_set(mytheme) #setting for all subsequent plots
```

3. Upload discharge data for the Eno River at site 02096500 for the same dates as we studied in class (2012-01-01 through 2021-12-31). Obtain data for discharge. Rename the columns with informative titles, as we did in class.

```
#load discharge data for Eno river site
EnoDischarge <- readNWISdv(siteNumbers = "02096500",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "2012-01-01",
                     endDate = "2021-12-31")

#Rename columns
names(EnoDischarge)[4:5] <- c("Discharge_cfs", "Approval.Code")
```

4. Build a plot called EnoPlot2. Use the base plot we made in class and make the following changes:

   - Add a column to your data frame for discharge in meters cubed per second. hint: package dplyr in tidyverse includes a `mutate` function
   - Add a column in your data frame for a 30-day rolling mean of the metric discharge. (hint: package dplyr in tidyverse includes a `mutate` function. hint: package zoo includes a `rollmean` function)
   - Create two geom_line aesthetics, one for daily discharge (meters cubed per second) and one for rolling mean of discharge. Color these differently.
   - Update your ggplot theme. I suggest "classic." (hint: https://ggplot2.tidyverse.org/reference/ggtheme.html)
   - Update axis names
   - Change the y axis from a linear to a log10 axis (hint: google "ggplot logged axis")
   - Add a legend. (hint: Google "add legend two geom layers ggplot")

```
EnoPlot <-
  ggplot(EnoDischarge, aes(x = Date, y = Discharge_cfs)) +
        geom_line() +
        xlab("Year")

#Added new row for cms
EnoDischarge <- mutate(EnoDischarge, Discharge_cms = Discharge_cfs*0.0283)

#Added new row for 30-day rolling mean of metric discharge
EnoDischarge <- mutate(EnoDischarge, Discharge_rollmean = rollmean(Discharge_cms, 30, fill=NA)) #fillin

#From Kateri - added pipe to be done in one line
#EnoDischarge <- EnoDischarge %>%
 # mutate(Discharge_cms = Discharge_cfs*0.0283,
        # Discharge_rollmean = rollmean(Discharge_cms, 30, fill=NA)
```

```
#Create visualization
EnoPlot2 <- ggplot(data=EnoDischarge, aes(x= Date))+
  geom_line(aes(y=Discharge_cms, color = "Daily Discharge")) +
  geom_line(aes(y=Discharge_rollmean, color = "Rolling 30 Day Avg")) +
  scale_y_log10(name=expression("Discharge (m"^3*"/s)")) +
  labs(x="Date", title="Eno River Discharge \nSite 02096500") +
  scale_fill_manual(name="", values = c("Daily Discharge" = "blue")) +
  scale_fill_manual(name="", values = c("Rolling 30 Day Avg" = "green"))
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

```
EnoPlot2
```

```
## Warning: Removed 29 row(s) containing missing values (geom_path).
```
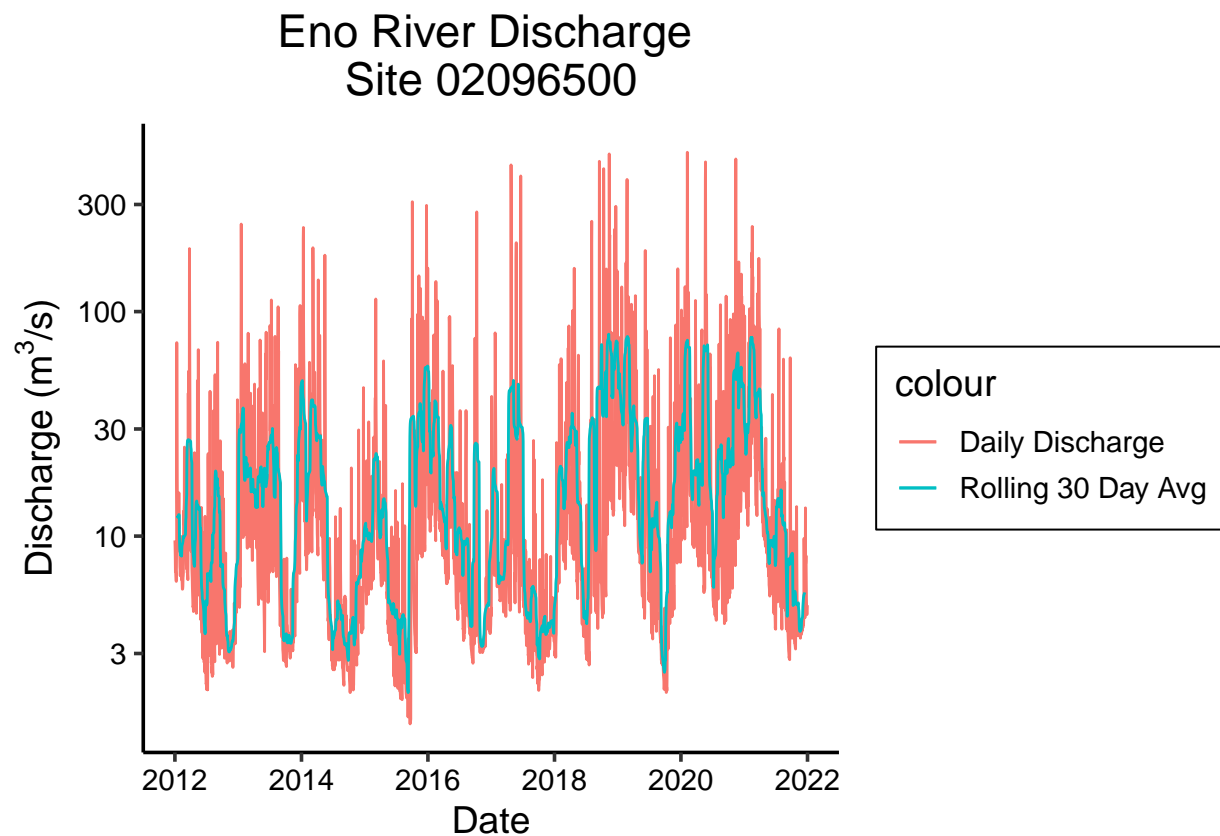

5. In what ways was the second plot a more effective visualization than the first?

   ANSWER: You can better visualize the difference in the daily discharge and the rolling 30 day average with the different colors. Likewise, making them the same units was helpful.

6. What portions of the coding were challenging for you?

ANSWER: The rolling 30-day discharge (I did not know how to fill the NA values) and the aesthetics for the two lines - still working on removing the title from legend.

7. Interpret the graph you made. What are the things you notice about within- and across-year variability, as well as the differences between daily values and 30-day rolling mean?

ANSWER: The average values are smaller in magnitude compared to daily values. It would also be helpful to have the months to better understand seasonality of the data; however, you can see that the discharge minimum is increasing over time, as well as the maximum.