# Assignment 3: Physical Properties of Rivers

*Yixin Wen*

## OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis on the physical properties of rivers.

### 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.
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., "Salk_A03_RiversPhysical.Rmd") prior to submission.

The completed exercise is due on 18 September 2019 at 9:00 am.

### Setup

1. Verify your working directory is set to the R project file,
2. Load the tidyverse, dataRetrieval, and cowplot packages
3. Set your ggplot theme (can be theme_classic or something else)
4. Import a data frame called "MysterySiteDischarge" from USGS gage site 03431700. Upload all discharge data for the entire period of record. Rename columns 4 and 5 as "Discharge" and "Approval.Code". DO NOT LOOK UP WHERE THIS SITE IS LOCATED.
5. Build a ggplot of discharge over the entire period of record.

```
getwd()
```

```
## [1] "/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/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(cowplot)
```

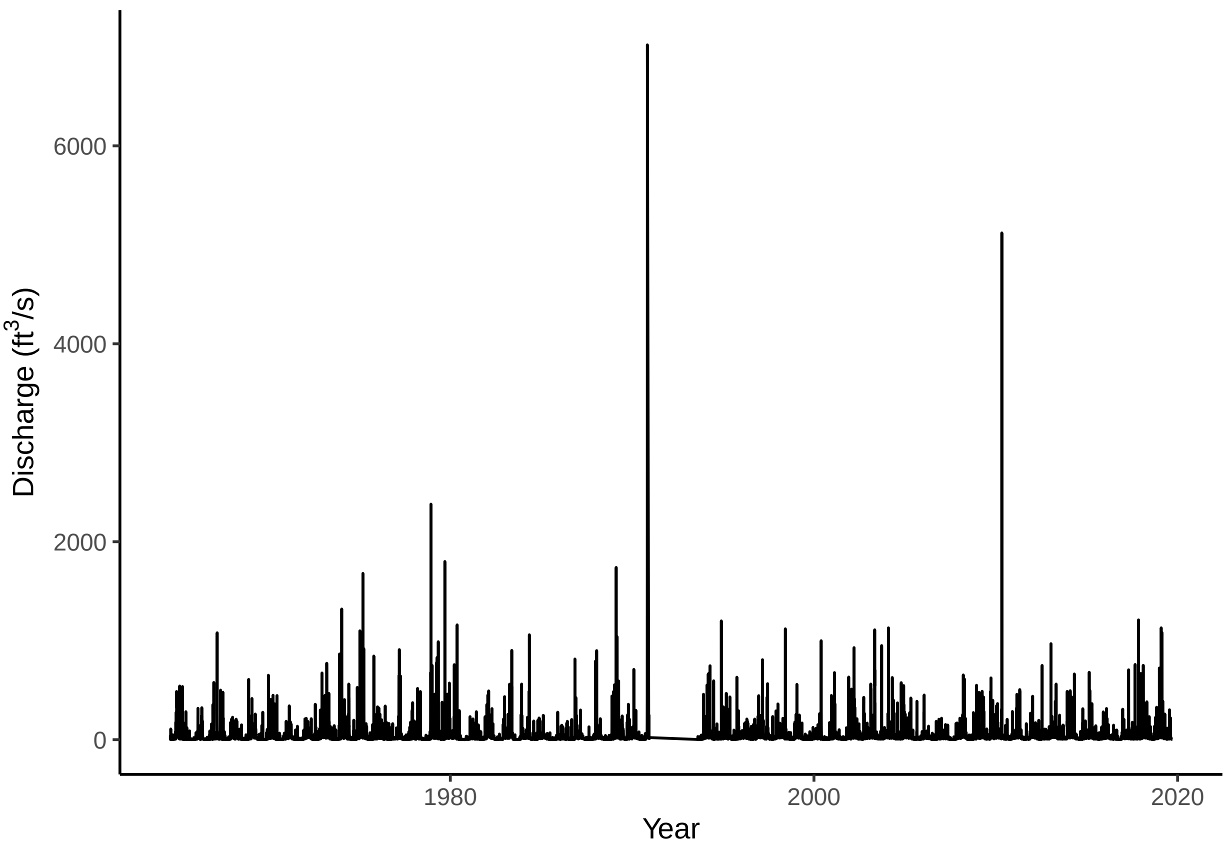
```
##
## ********************************************************

## 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())
```

```
## ********************************************************
```

```r
library(lubridate)
```

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

## The following object is masked from 'package:cowplot':
##
##     stamp

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

```r
theme_set(theme_classic())

MysterySiteDischarge <- readNWISdv(siteNumbers = "03431700",
                                   parameterCd = "00060",
                                   startDate = "",
                                   endDate = "")

names(MysterySiteDischarge)[4:5] <- c("Discharge", "Approval.Code")

MysterySitePlot <-
  ggplot(MysterySiteDischarge, aes(x = Date, y = Discharge)) +
  geom_line() +
  labs(x = "Year", y = expression("Discharge (ft"^3*"/s)"))
print(MysterySitePlot)
```

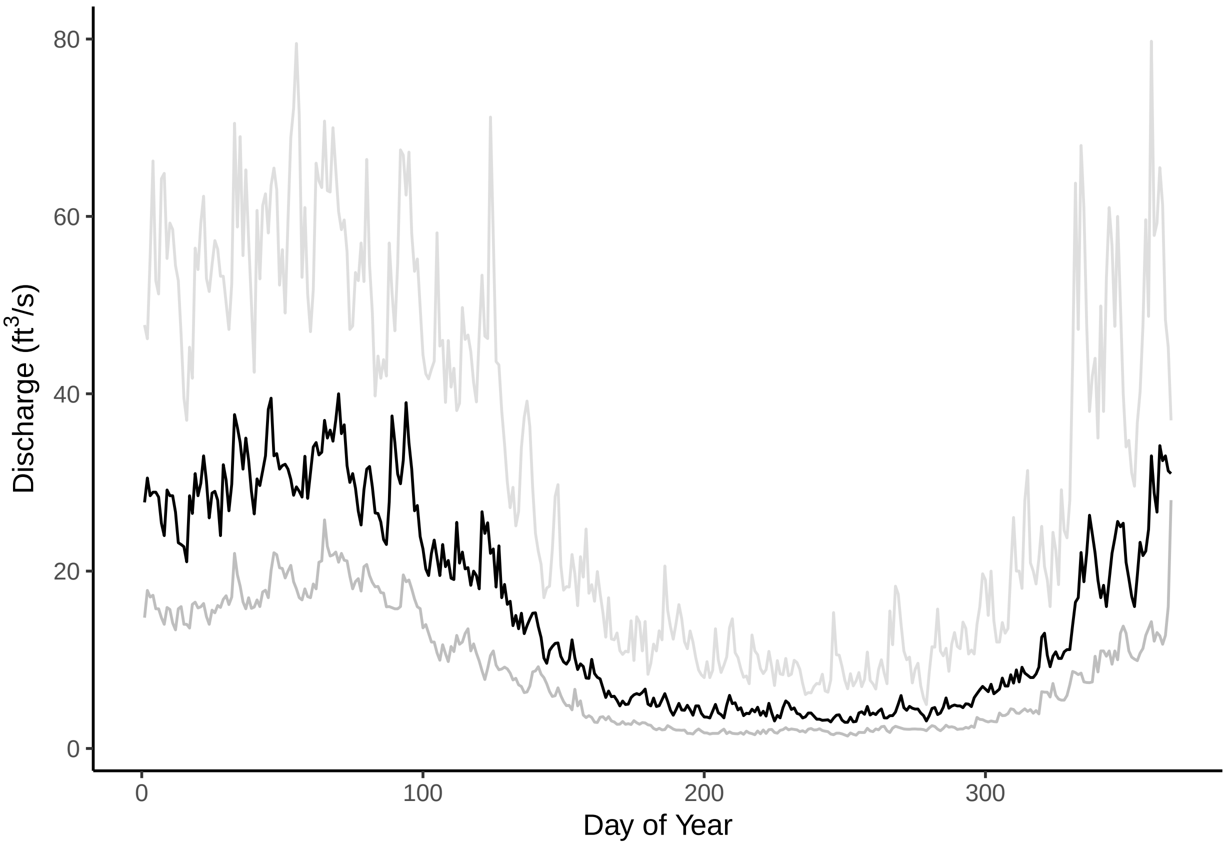



### Analyze seasonal patterns in discharge

5. Add a "Year" and "Day.of.Year" column to the data frame.
6. Create a new data frame called "MysterySiteDischarge.Pattern" that has columns for Day.of.Year, median discharge for a given day of year, 75th percentile discharge for a given day of year, and 25th percentile discharge for a given day of year. Hint: the summarise function includes `quantile`, wherein you must specify `probs` as a value between 0 and 1.
7. Create a plot of median, 75th quantile, and 25th quantile discharges against day of year. Median should be black, other lines should be gray.

```
MysterySiteDischarge <-
  MysterySiteDischarge %>%
  mutate(Year = year(Date), Day.of.Year = yday(Date))

MysterySiteDischarge.Pattern <-
  MysterySiteDischarge %>%
  group_by(Day.of.Year) %>%
  summarise( Q25th.Discharge = quantile(Discharge, 0.25),
             Median.Discharge  = quantile(Discharge, 0.50),
             Q75th.Discharge = quantile(Discharge, 0.75))

MysterySitePatternPlot <-
  ggplot(MysterySiteDischarge.Pattern, aes(x= Day.of.Year))+
      geom_line(aes(y = Q25th.Discharge), color = "gray")+
      geom_line(aes( y = Median.Discharge), color = "black")+
      geom_line(aes(y = Q75th.Discharge),alpha = 0.5, color = "gray")+
```

```
      labs(x = "Day of Year", y = expression("Discharge (ft"^3*"/s)"))
print(MysterySitePatternPlot)
```




8. What seasonal patterns do you see? What does this tell you about precipitation patterns and climate in the watershed?

   The discharge of river is larger in spring and winter, and decreases in summer and fall.The precipitation may be concentrated in spring and winter. In summer, the weather temperature may be high and the climate may be drought in summer, and evapotranspiration rates are higher, reducing surface run off.

## Create and analyze recurrence intervals

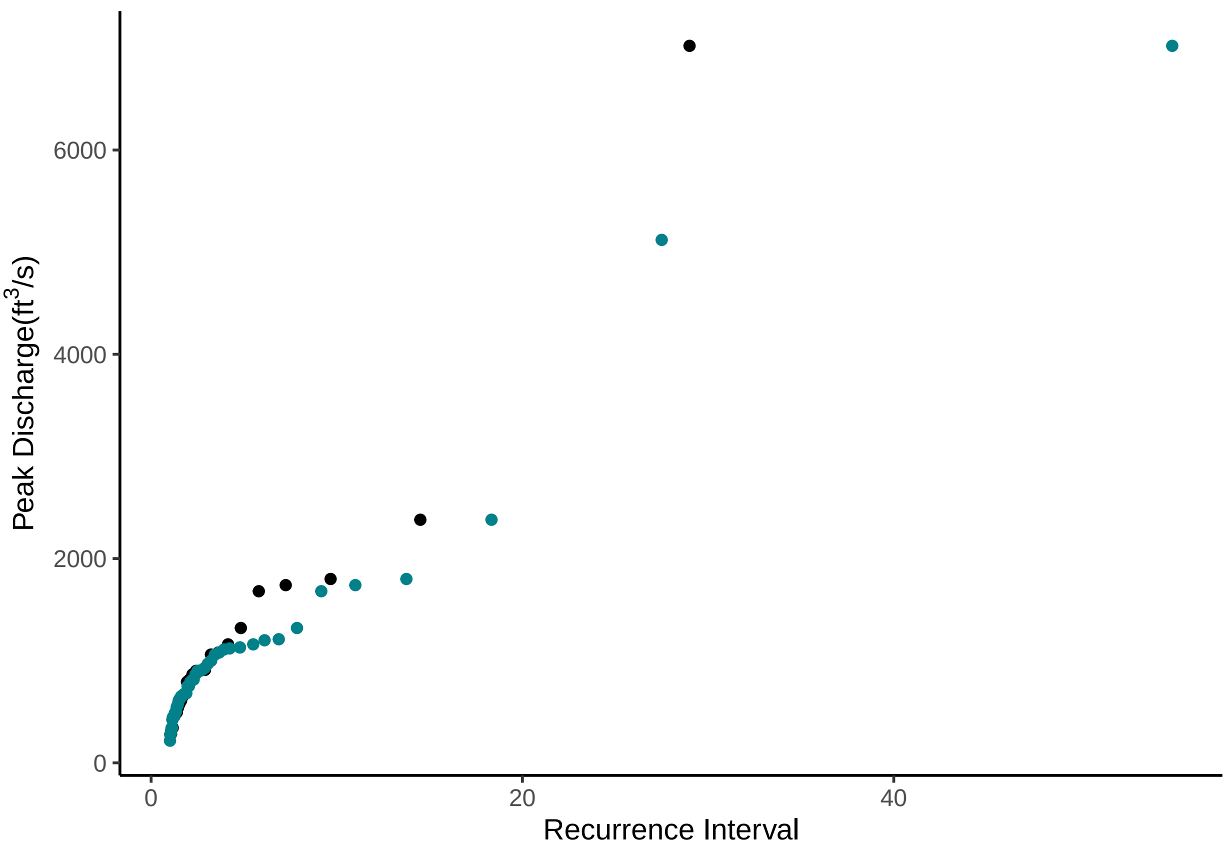
9. Create two separate data frames for MysterySite.Annual.30yr (first 30 years of record) and MysterySite.Annual.Full (all years of record). Use a pipe to create your new data frame(s) that includes the year, the peak discharge observed in that year, a ranking of peak discharges, the recurrence interval, and the exceedende probability.

10. Create a plot that displays the discharge vs. recurrence interval relationship for the two separate data frames (one set of points includes the values computed from the first 30 years of the record and the other set of points includes the values computed for all years of the record.

11. Create a model to predict the discharge for a 100-year flood for both sets of recurrence intervals.

```
MysterySite.Annual.30yr <-
  MysterySiteDischarge %>%
  filter(Year < 1994) %>%
  group_by(Year) %>%
```

```
  summarise(PeakDischarge = max(Discharge)) %>%
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(Year) + 1)/Rank,
         ExceedenceProbability = 1/RecurrenceInterval)

MysterySite.Annual.Full <-
  MysterySiteDischarge %>%
  group_by(Year) %>%
  summarise(PeakDischarge = max(Discharge)) %>%
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(Year) + 1)/Rank,
         ExceedenceProbability = 1/RecurrenceInterval)

MysterySiteRecurrencePlot <-
  ggplot(MysterySite.Annual.30yr, aes(x = RecurrenceInterval,
                                      y = PeakDischarge)) +
  geom_point() +
  geom_point(data = MysterySite.Annual.Full, color = "#02818a",
             aes(x = RecurrenceInterval, y = PeakDischarge))+
  labs(x = "Recurrence Interval", y = expression ("Peak Discharge(ft"^3*"/s)"))

print(MysterySiteRecurrencePlot)
```




```
MysterySite.RImodel.30 <- lm(data = MysterySite.Annual.30yr, PeakDischarge ~ RecurrenceInterval)
summary(MysterySite.RImodel.30)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ RecurrenceInterval, data = MysterySite.Annual.30yr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -973.89  -30.29   49.21  126.61  524.10
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          211.88      60.98   3.474  0.00181 **
## RecurrenceInterval   216.69       8.77  24.709  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 261.7 on 26 degrees of freedom
## Multiple R-squared:  0.9592, Adjusted R-squared:  0.9576
## F-statistic: 610.5 on 1 and 26 DF,  p-value: < 2.2e-16
```

```
MysterySite.RImodel.Full <- lm(data = MysterySite.Annual.Full, PeakDischarge ~ RecurrenceInterval)
summary(MysterySite.RImodel.Full)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ RecurrenceInterval, data = MysterySite.Annual.Full)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -441.3 -113.2   18.1  106.1 1181.5
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         415.781     36.224   11.48  7.2e-16 ***
## RecurrenceInterval  128.100      3.795   33.76  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 232.3 on 52 degrees of freedom
## Multiple R-squared:  0.9564, Adjusted R-squared:  0.9555
## F-statistic:  1139 on 1 and 52 DF,  p-value: < 2.2e-16
```

```
MysterySite.RImodel.30$coefficients[1] +
  MysterySite.RImodel.30$coefficients[2]*100
```

```
## (Intercept)
##     21880.9
```

```
MysterySite.RImodel.Full$coefficients[1] +
  MysterySite.RImodel.Full$coefficients[2]*100
```

```
## (Intercept)
##    13225.76
```

12. How did the recurrence interval plots and predictions of a 100-year flood differ among the two data frames? What does this tell you about the stationarity of discharge in this river?

    In the recurrence interval plot, the slope of first 30 years points is steeper than that of full years points. And the discharge prediction of first 30 years model is much greater than that of full years model.The discharge of this river is non-stationary since the 100-year flood predicted values of two models are quite different over different time period.

## Reflection

13. What are 2-3 conclusions or summary points about river discharge you learned through your analysis?

    1. Precipitation and climate change will impact seasonal changes of river discharge. 2. Rivers can be non-stationary because of the climate impact, the floods and droughts may occur which are outside the normal range.

14. What data, visualizations, and/or models supported your conclusions from 13?

    From the MysterySitePatternPlot it can be seen that the river discharge have a very obvious seasonal trend, which may caused by climate and precipitation impact. From 2 time-scale models, the prediction of 100-year flood varies greatly, which indicates the nonstationarity of river.

15. Did hands-on data analysis impact your learning about discharge relative to a theory-based lesson? If so, how?

    Hands-on data analysis can let me explore the theory on my own, and it can help me understand it better from examples. The only problem is it may not cover the whole theory from one or two examples, so it'll be better to summary main theories after hands-on data analysis.

16. How did the real-world data compare with your expectations from theory?

    Sometimes the real-world data may not in accordance with theory. For example, we learn from class that we may underpredict the discharge when using shorter periods of data, however in this example, the prediction of full-year model is less than first 30-year model.