Assignment 3: Physical Properties of Rivers

Keqi He

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] "C:/Users/keqi/Desktop/Hydrologic_Data_Analysis/Assignments"
```

```
# install.packages("tidyverse")
# install.packages("dataRetrieval")
# install.packages("cowplot")
# install.packages("lubridate")

library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------------------------------

## 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 ------------------------------------------------------------------------------------ tidy
## 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", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")
attr(MysterySiteDischarge, "variableInfo")
```

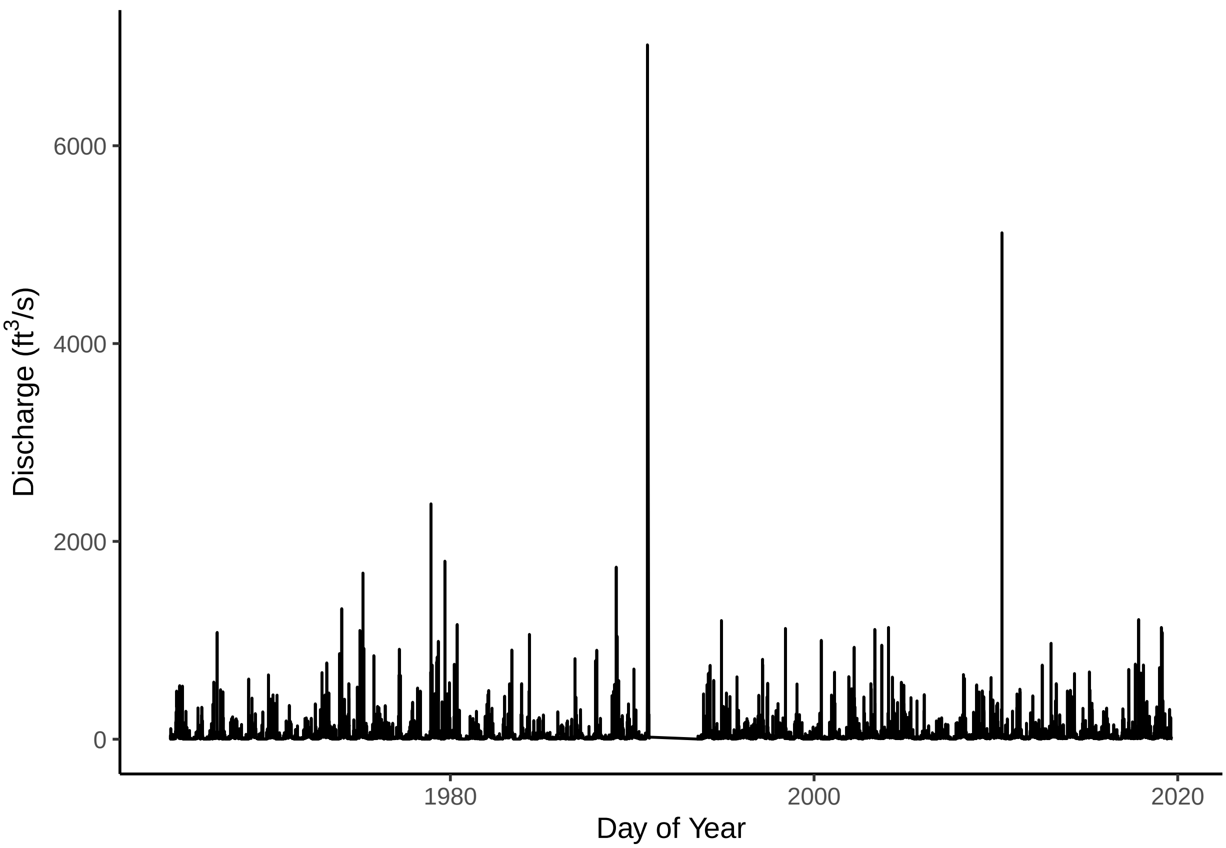
```
##   variableCode           variableName              variableDescription
## 1        00060 Streamflow, ft&#179;/s Discharge, cubic feet per second
##       valueType  unit options noDataValue
## 1 Derived Value ft3/s    Mean          NA
```

```r
attr(MysterySiteDischarge, "siteInfo")
```

```
##                                          station_nm  site_no agency_cd
## 1 RICHLAND CREEK AT CHARLOTTE AVE, AT NASHVILLE, TN 03431700      USGS
##   timeZoneOffset timeZoneAbbreviation dec_lat_va dec_lon_va       srs
## 1         -06:00                  CST   36.15122  -86.85427 EPSG:4326
##   siteTypeCd    hucCd stateCd countyCd network
## 1         ST 05130202      47    47037    NWIS
```

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

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



Analyze seasonal patterns in discharge

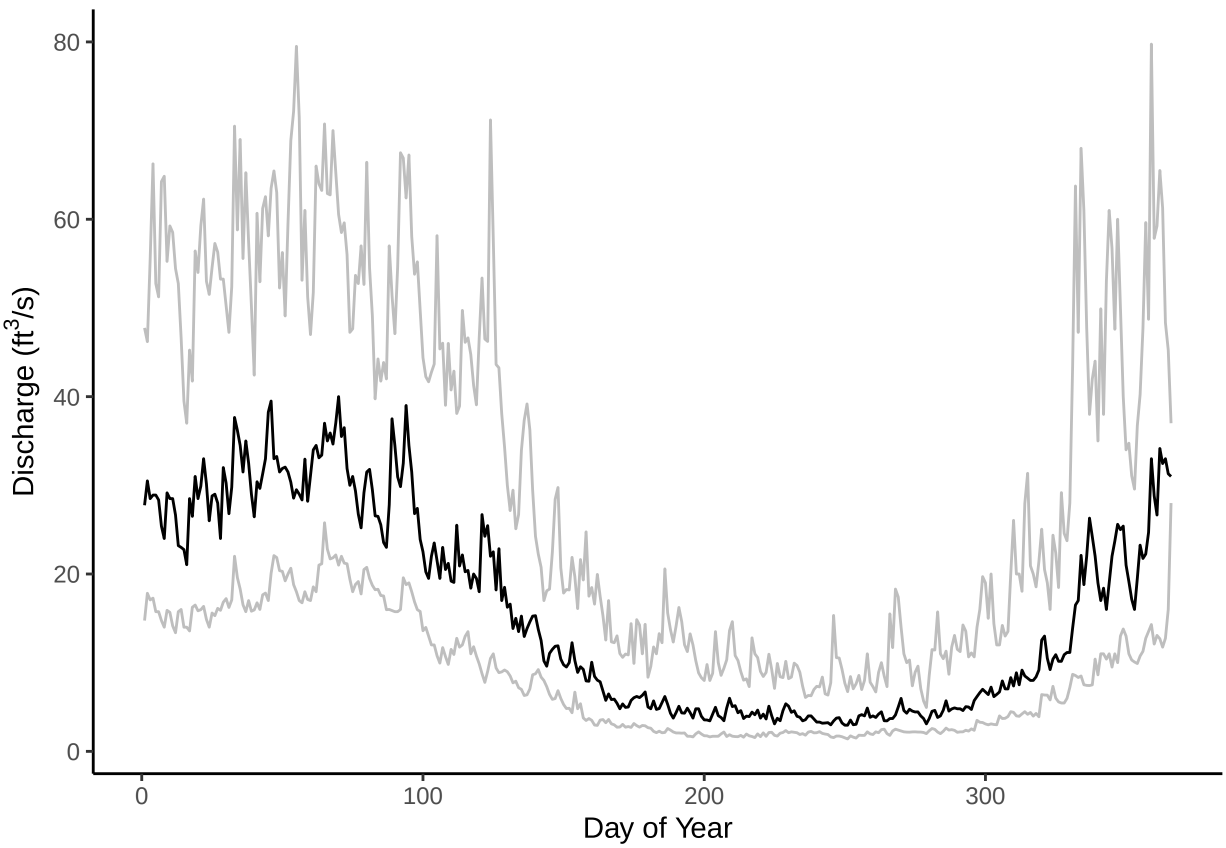
5. Add a "Year" and "Day.of.Year" column to the data frame.
6. Create a new data frame called "MysterySiteDischarge.Pattern" that has columns for Day.of.Year, median discharge for a given day of year, 75th percentile discharge for a given day of year, and 25th percentile discharge for a given day of 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(MedianDischarge = median(Discharge),
      Discharge_75 = quantile(Discharge, .75),
      Discharge_25 = quantile(Discharge, .25))

MysteryPatternPlot <-
  ggplot(MysterySiteDischarge.Pattern, aes(x = Day.of.Year)) +
  geom_line(aes(y = MedianDischarge)) +
  geom_line(aes(y = Discharge_75), color = "gray") +
```

```
  geom_line(aes(y = Discharge_25), color = "gray") +
  labs(x = "Day of Year", y = expression("Discharge (ft"^3*"/s)"))
print(MysteryPatternPlot)
```



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

   Discharge is much larger in winter and spring (November to April), while it is small in summer and fall. This watershed has different periods of precipitation and heat. In summer and fall, the temperature is high but the precipitation is small which will cause the watershed hot and dry. While in winter and spring, the temperature is low but the precipitation is large which will cause the watershed cold and wet.

Create and analyze recurrence intervals

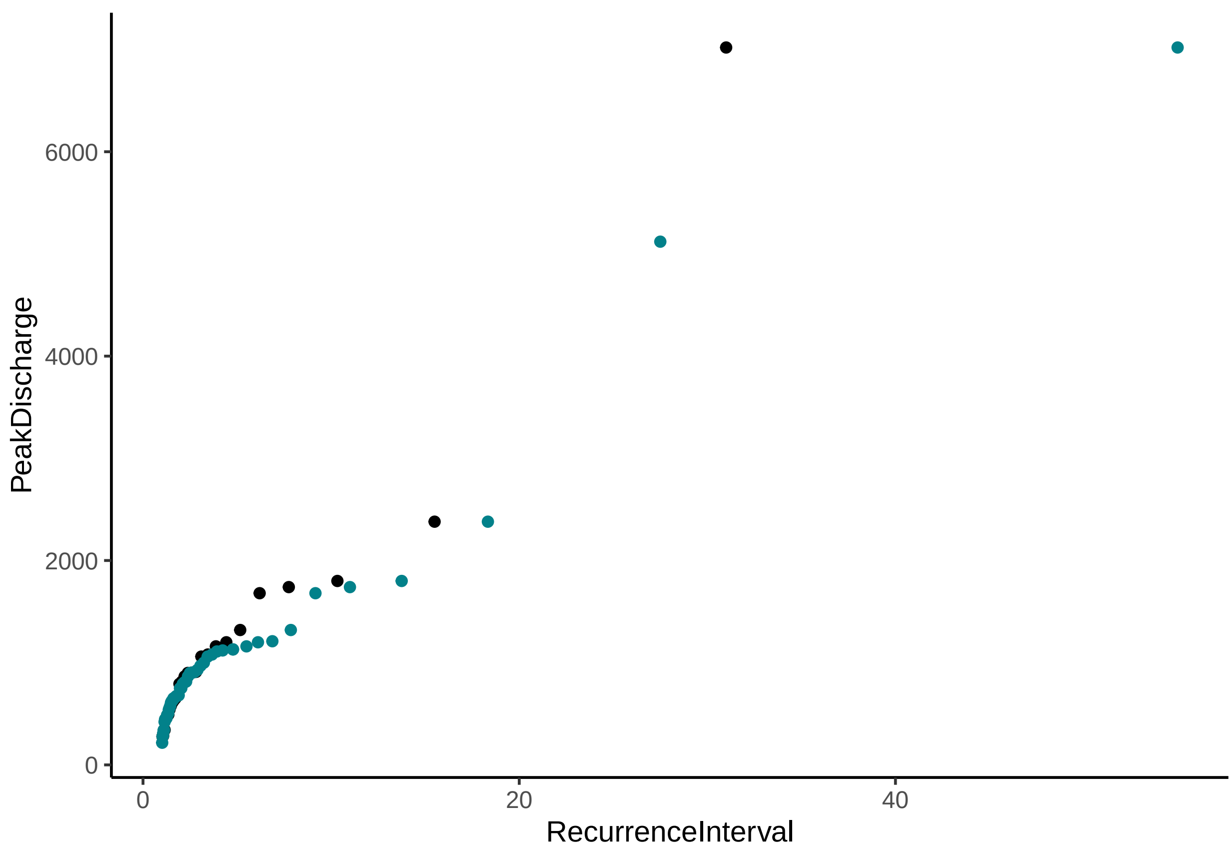
9. Create two separate data frames for MysterySite.Annual.30yr (first 30 years of record) and MysterySite.Annual.Full (all years of record). Use a pipe to create your new data frame(s) that includes the year, the peak discharge observed in that year, a ranking of peak discharges, the recurrence interval, 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 < 1996) %>%
  group_by(Year) %>%
  summarise(PeakDischarge = max(Discharge)) %>%
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(Year) + 1)/Rank,
         Probability = 1/RecurrenceInterval)

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

MysteryRecurrencePlot.Full <-
  ggplot(MysterySite.Annual.30yr, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() +
  geom_point(data = MysterySite.Annual.Full, color = "#02818a",
             aes(x = RecurrenceInterval, y = PeakDischarge))
print(MysteryRecurrencePlot.Full)
```




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

```
##
```

```
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = MysterySite.Annual.30yr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -974.12 -337.65   34.84  232.57 2908.00
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -69.87     185.73  -0.376     0.71
## log(RecurrenceInterval)  1217.79     147.16   8.275 5.26e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 673.9 on 28 degrees of freedom
## Multiple R-squared:  0.7098, Adjusted R-squared:  0.6994
## F-statistic: 68.48 on 1 and 28 DF,  p-value: 5.261e-09
```

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

```
##
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = MysterySite.Annual.Full)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -955.95 -236.29   41.91  210.67 2805.35
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -2.001    116.322  -0.017    0.986
## log(RecurrenceInterval) 1052.234     88.834  11.845   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 578.3 on 52 degrees of freedom
## Multiple R-squared:  0.7296, Adjusted R-squared:  0.7244
## F-statistic: 140.3 on 1 and 52 DF,  p-value: < 2.2e-16
```

```
Mystery.RImodel.30yr$coefficients[1] + Mystery.RImodel.30yr$coefficients[2]*log(100)
```

```
## (Intercept)
##    5538.257
```

```
Mystery.RImodel.Full$coefficients[1] + Mystery.RImodel.Full$coefficients[2]*log(100)
```

```
## (Intercept)
##    4843.717
```

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?

    For some specific peak discharge, the recurrence interval computed for all years of the record is smaller than that computed for the first 30 years of the record. So the prediction of a 100-year flood using the recurrence interval computed for all years of the record is smaller than the prediction of a 100-year flood using the recurrence interval computed for 30 years of the record. This means that compared with the first 30 years peak discharge, peak discharge happened these years is much smaller. Discharge in this river is not completely stationary.

Reflection

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

    Some rivers have seasonal cycle of discharge while some rivers do not have. Whether a river has seasonal cycle is based on its geographical environment and climatic conditions. We can create a model considering the relationship between recurrence interval and peak discharge to predict the peak discharge of a flood with a given exceedence probability.

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

    Data we used includes discharge, date. Visualizations includes ggplot. And we created linear function between peak discharge and log(recurrence interval) to predicit the peak discharge of a flood with a given exceedence probability.

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

    Yes. Hands-on data analysis can help me learn theory about discharge better. It is more vivid and memorable. Hands-on data analysis is like a procedure that we explore some laws of discharge instead of only receiving knowledge from teacher without any thinking.

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

    Real-world data is complicated and imperfect because there are lots of factors can affect the value of real-world data. However, in theory, there are some assumputions that can simplify the problems, like assuming that discharge is stationary when we use discharge records to predict the peak discharge of a flood with a given exceedence probability.