Assignment 3: Physical Properties of Rivers

Theo Cai

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.

```
#Session set up
getwd()
```

```
## [1] "Z:/Hydrologic_Data_Analysis"
```

```
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   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------------------- tid
## 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
```

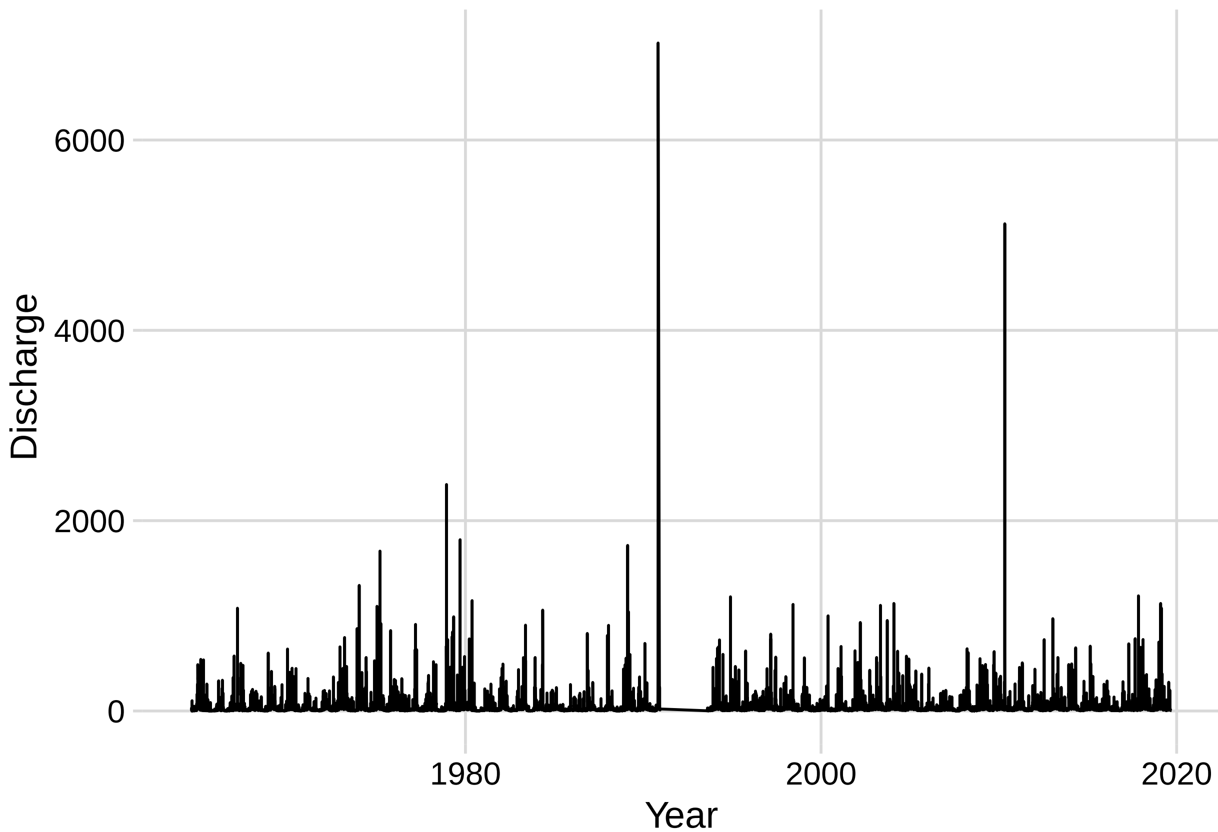
```r
theme_set(theme_minimal_grid())

#Import dataset
MysterySiteDischarge <- readNWISdv(siteNumbers = "03431700",
                                  parameterCd = "00060",
                                  startDate = "",
                                  endDate = "")
#Rename columns
names(MysterySiteDischarge)[4:5] <- c("Discharge", "Approval.Code")

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
#GGplot of entire period of record
MysteryPlot <-
  ggplot(MysterySiteDischarge, aes(x = Date, y = Discharge)) +
  geom_line() +
  xlab("Year")
print(MysteryPlot)
```

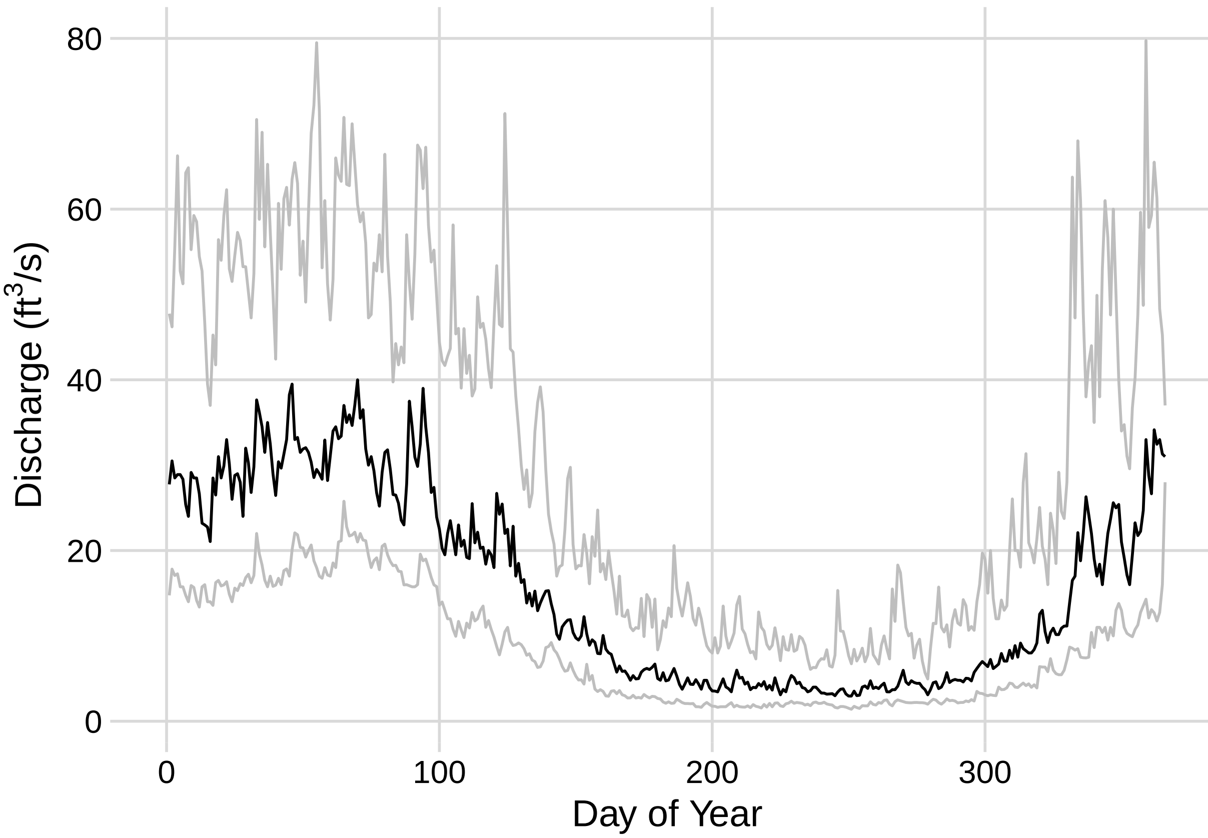


Analyze seasonal patterns in discharge

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.

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

#New Data Frame
MysterySiteDischarge.Pattern <-
  MysterySiteDischarge %>%
  group_by(Day.of.Year) %>%
  summarise(MedianDischarge = median(Discharge), SeventyFifthQ = quantile(Discharge, probs = 0.75),
            TwentyFifthQ = quantile(Discharge, probs = 0.25))
```

```r
#Create GGplots
CombinedPlot <-
  ggplot(MysterySiteDischarge.Pattern, aes(Day.of.Year)) +
  geom_line(aes(y = MedianDischarge)) +
  geom_line(aes(y = SeventyFifthQ), colour = "808080") +
  geom_line(aes(y = TwentyFifthQ), colour = "808080") +
  labs(x = "Day of Year", y = expression("Discharge (ft"^3*"/s)"))
print(CombinedPlot)
```



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

   I see high discharge from November all the way till April/May. This indicates to me that this location does not follow traditional temperate patterns of weather, as it seems it experiences the most precipitation in the winter months. I would guess that this watershed is in a hot region that sees little rainfall in the summer but a lot in the winter - as opposed to snow.

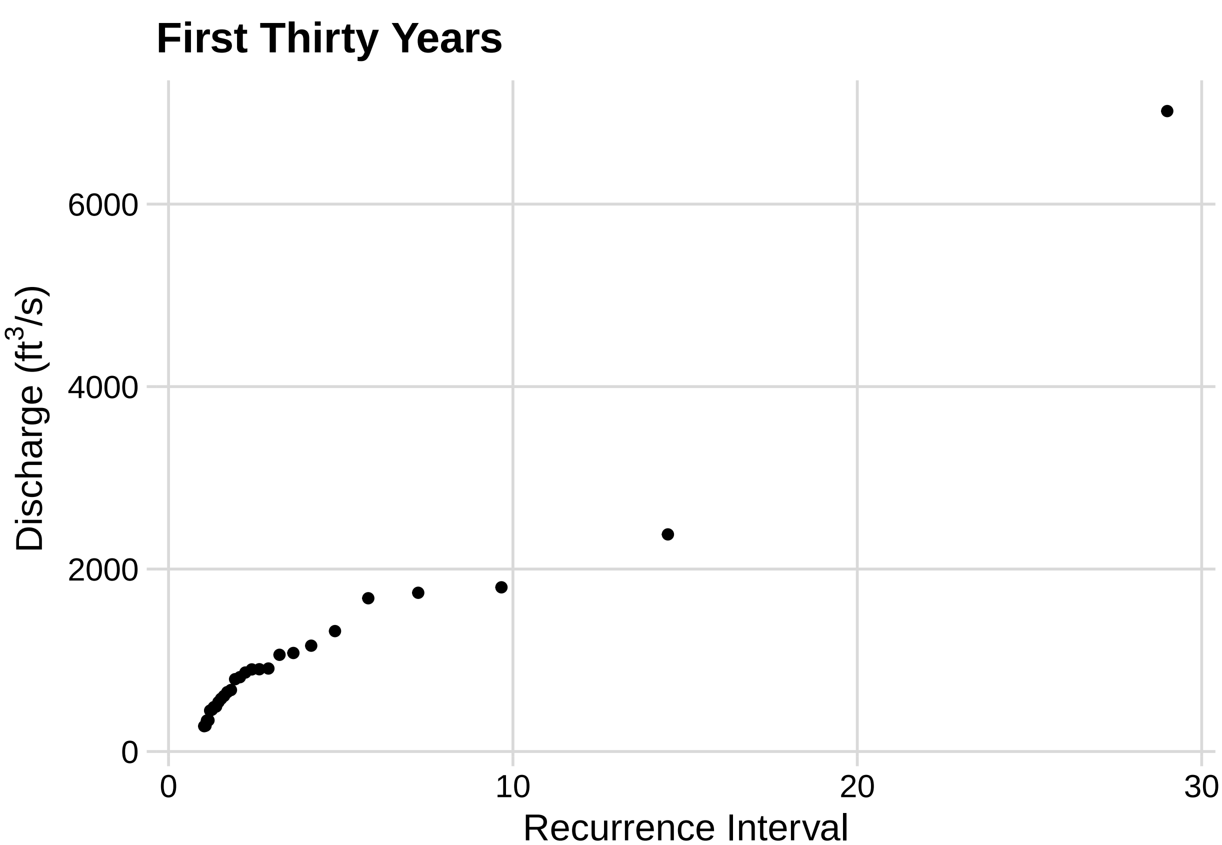
Create and analyze recurrence intervals

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 exceedence probability.

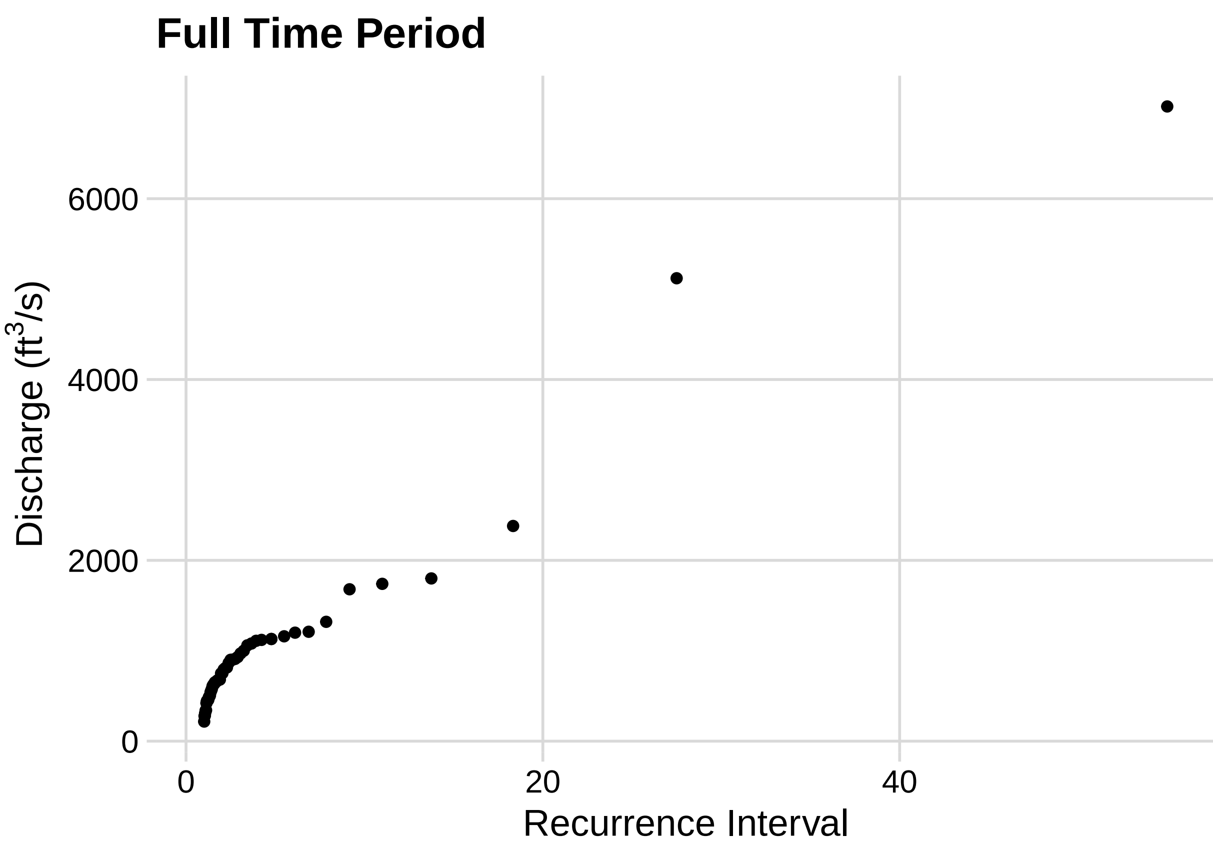
10. Create a plot that displays the discharge vs. recurrence interval relationship for the two separate data frames (one set of points includes the values computed from the first 30 years of the record and the other set of points includes the values computed for all years of the record.

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

```
#Make Date Frame
MysterySite.Annual.30yr <-
  MysterySiteDischarge %>%
  filter(Year < 1994) %>%
  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)

#Make plots
MysterySiteRecurrencePlot.30yr <-
  ggplot(MysterySite.Annual.30yr, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() +
  labs(x = "Recurrence Interval", y = expression("Discharge (ft"^3*"/s)"))
print(MysterySiteRecurrencePlot.30yr + ggtitle("First Thirty Years"))
```



```r
MysterySiteRecurrencePlot.Full <-
  ggplot(MysterySite.Annual.Full, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() +
  labs(x = "Recurrence Interval", y = expression("Discharge (ft"^3*"/s)"))
print(MysterySiteRecurrencePlot.Full + ggtitle("Full Time Period"))
```




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

```
##
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = MysterySite.Annual.30yr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -982.36 -366.95   42.38  247.09 2838.21
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -107.5      197.3  -0.545     0.59
## log(RecurrenceInterval)   1273.8      157.1   8.107 1.38e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 689.4 on 26 degrees of freedom
```

```
## Multiple R-squared:  0.7166, Adjusted R-squared:  0.7057
## F-statistic: 65.73 on 1 and 26 DF,  p-value: 1.377e-08
```

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

```
##
## 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
```

```r
#100-Year Recurrence
Mystery.30yr.RImodel$coefficients[1] + Mystery.30yr.RImodel$coefficients[2]*log(100)
```

```
## (Intercept)
##    5758.621
```

```r
Mystery.Full.RImodel$coefficients[1] + Mystery.Full.RImodel$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?

    The recurrence interval plot of the first thirty years of data definitely shows us that, historically, the curve of the recurrence log was less extreme. The slope of the increase of discharge per increase in recurrence interval used to be smaller than that of the full time period. Curiously, however, the 100-year flood discharge for the past is greater than that of the full period. This tells me that the stationarity of discharge in this river is fairly stable - perhaps there have been minor changes to the system, but in general there doesn't seem to have been any major landscape or environmental changes throughout the entire time period.

Reflection

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

    I found that river discharge isn't always as cut and dry as "snow melt in spring = high discharge around same time." Additionally, I found that it was possible for the discharge of 100-Year floods to decrease from the past to the present, indicating that either global warming isn't affecting certain sites as much as others, or that it is affecting different sites differently depending on geographical location.

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

CombinedPlot helped me realize conclusion #1, while all of the Recurrence plots/models helped me realize conclusion #2.

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

    Yes, it did! It forced me to organize and extrapolate the information for myself, rather than just giving it to me. It taught me to read and recognize patterns of a larger theory.

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

    The real-world data matched up a little bit. I didn't expect the site to be in Tennessee, but I guess I at least got right the fact that it's technically in the South.