Assignment 3: Physical Properties of Rivers

Tristen Townsend

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/Tristen/OneDrive - Duke University/Fall 2019/Hydrologic Data Analysis/Hydrologic_Data_Ana
```

```
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
```

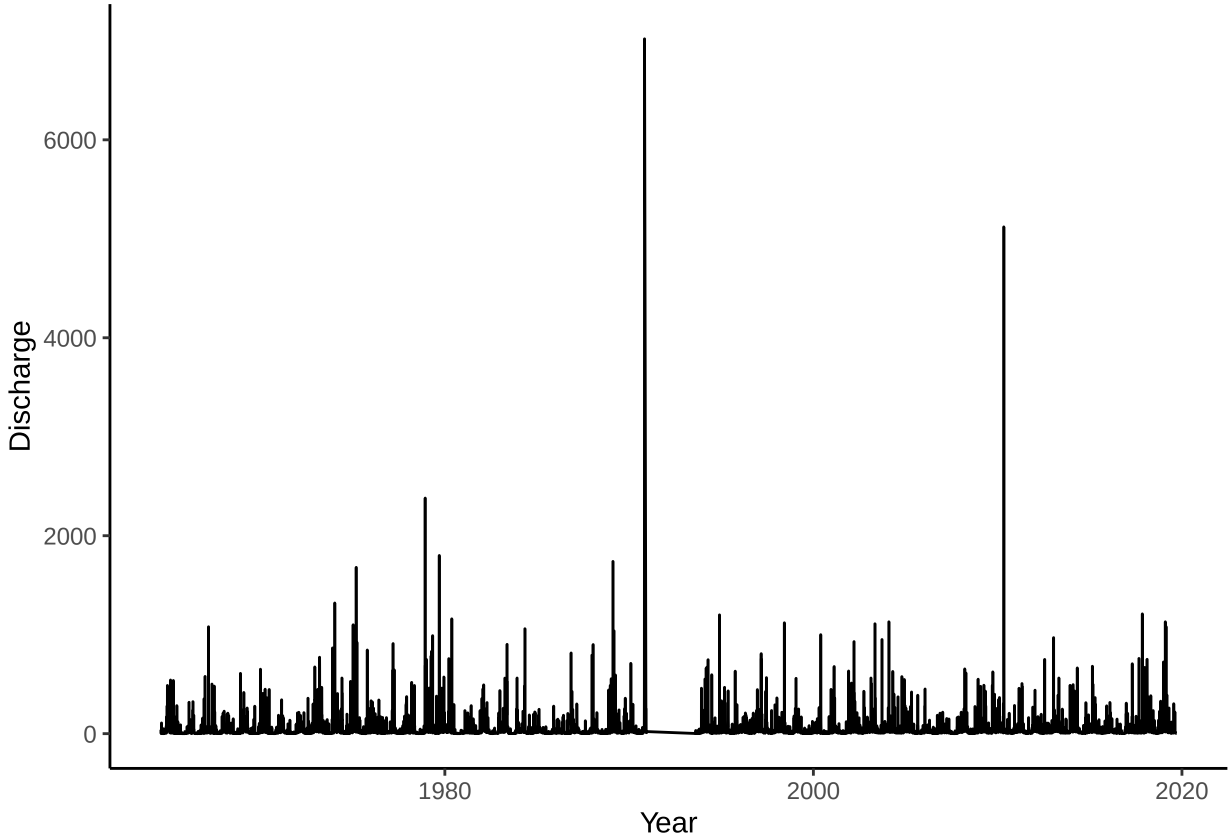
```
## The following object is masked from 'package:base':
##
##     date
```

```r
theme_set(theme_classic())

#Load data
MysterySiteDischarge <- readNWISdv(siteNumbers = "03431700",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")

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

#Plot graph of discharge vs time
MysterySitePlot <-
  ggplot(MysterySiteDischarge, aes(x = Date, y = Discharge)) +
         geom_line() +
         xlab("Year")
print(MysterySitePlot)
```



Analyze seasonal patterns in discharge

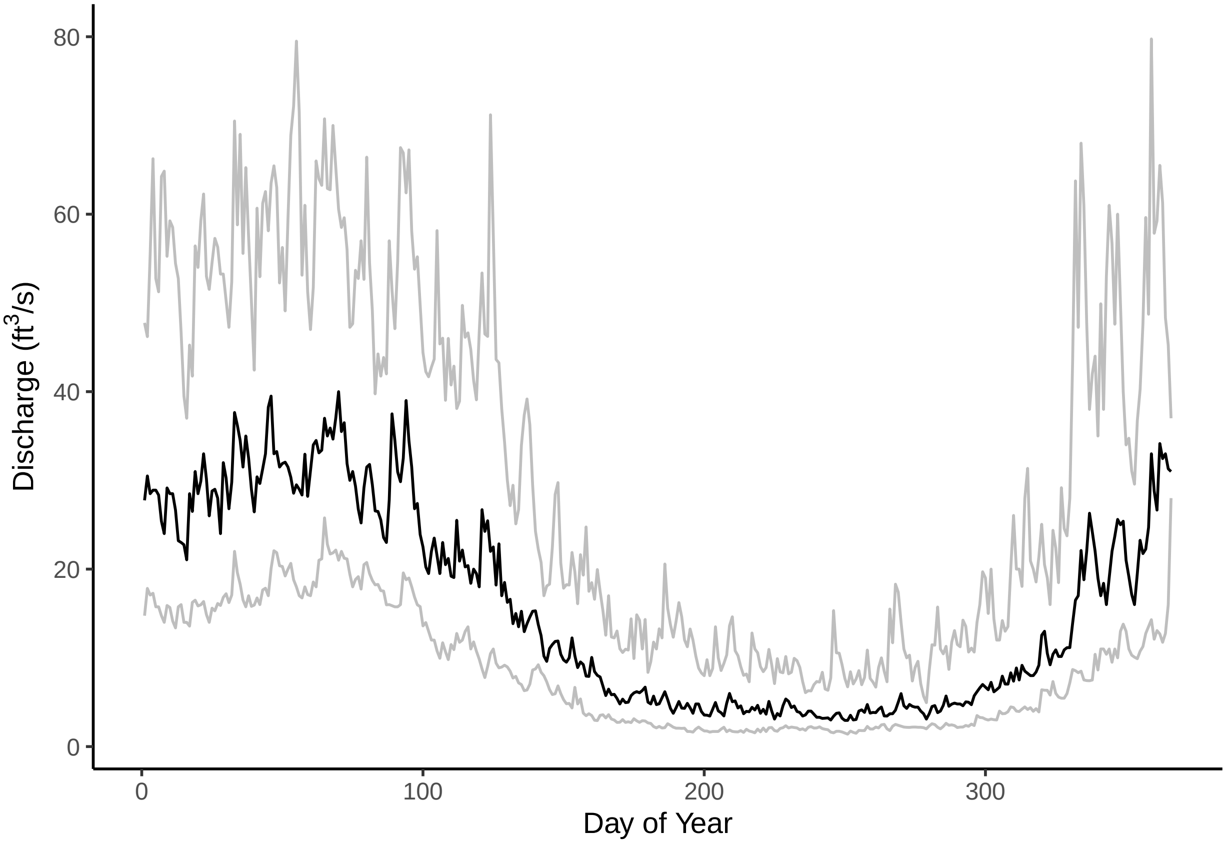
5. Add a "Year" and "Day.of.Year" column to the data frame.
6. Create a new data frame called "MysterySiteDischarge.Pattern" that has columns for Day.of.Year, median discharge for a given day of year, 75th percentile discharge for a given day of year, and 25th percentile discharge for a given day of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

```r
#Adding year and day of year columns
MysterySiteDischarge <-
  MysterySiteDischarge %>%
  mutate(Year = year(Date))

MysterySiteDischarge <-
  MysterySiteDischarge %>%
  mutate(DOY = yday(Date))

#Make new dataframe
MysterySiteDischarge.Pattern <-
  MysterySiteDischarge %>%
  group_by(DOY) %>%
  summarise(Median.Discharge = median(Discharge),
      SeventyFifthPercentile.Discharge = quantile(Discharge, probs=.75),
      TwentyFifthPercentile.Discharge = quantile(Discharge, probs=.25))

MysterySitePatternPlot <-
  ggplot(MysterySiteDischarge.Pattern, aes(x = DOY)) +
  geom_line(aes(y = Median.Discharge)) +
  geom_line(aes(y = SeventyFifthPercentile.Discharge), color = "gray") +
  geom_line(aes(y = TwentyFifthPercentile.Discharge), 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?

   It looks like it is the wet season between days 1-150 (January to May) as discharge is relatively higher then. Beginning around day 150 (June) up to around day 325 (late-November), discharge drops noticeably, then increases again after day 325 (into December and the new year). This indicates this watershed typically has its wet season during the winter and spring and experiences its dry season during summer and fall.

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.

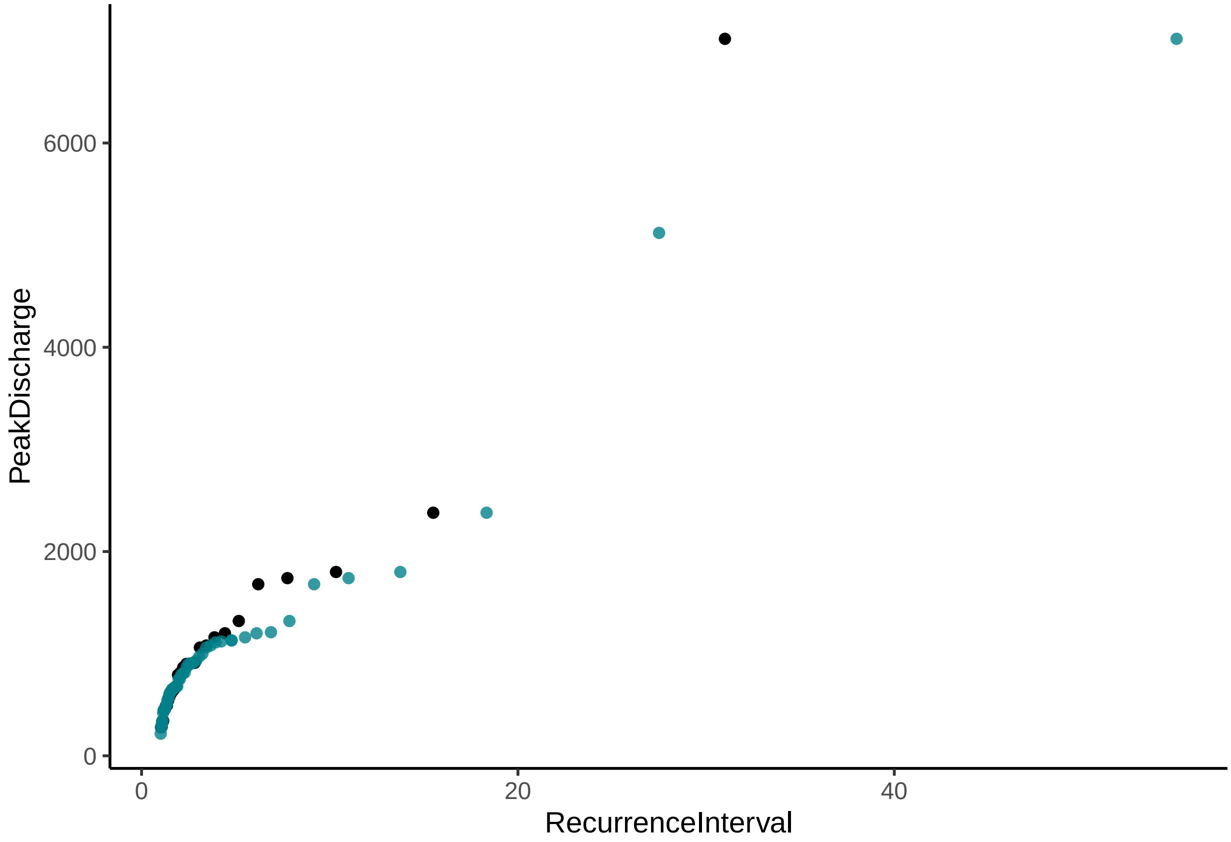
10. Create a plot that displays the discharge vs. recurrence interval relationship for the two separate data frames (one set of points includes the values computed from the first 30 years of the record and the other set of points includes the values computed for all years of the record.

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

```r
#Create dataframes
MysterySite.Annual.30yr <-
  MysterySiteDischarge %>%
  filter(Year < min(Year) + 32) %>%  #+32 to account for missing years
  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)

#Plotting discharge vs recurrence for both datasets
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),
             alpha = 0.8)
print(MysterySiteRecurrencePlot)
```



```
#Running both models
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
```

```
#Getting coefficients
Mystery.RImodel.30yr$coefficients # -69.9, 1217.8
```

```
##             (Intercept) log(RecurrenceInterval)
##               -69.87363              1217.79015
```

```
Mystery.RImodel.Full$coefficients # -2, 1052.2
```

```
##             (Intercept) log(RecurrenceInterval)
##               -2.000535             1052.234166
```

```
#100 year recurrence
Mystery.RImodel.30yr$coefficients[1] + Mystery.RImodel.30yr$coefficients[2]*log(100) #5538
```

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

```
Mystery.RImodel.Full$coefficients[1] + Mystery.RImodel.Full$coefficients[2]*log(100) #4843
```

```
## (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?

    Based on the first 30 years of data, the discharge for a 100-year flood was predicted to be 5538 cubic feet per second. Based on the full data of record, the discharge for a 100-year flood was predicted to be 4843 cubic feet per second. This illustrates the concept that “stationarity is dead”, because when including data from the most recent twenty-five years, the predicted discharge is much lower. This probably is due to more dry conditions in the area.

Reflection

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

    In many areas, we expect climate change to increase precipitation and result in increased discharge and flooding events, but this assignment illustrated that is not always the case (especially from my perspective, as I am from the East Coast).

The data record one chooses to use during an analysis can greatly impact the results, and in turn play a large role in how management is done.

Just because you include a larger set of data in your model, it does not always mean you have captured all variability.

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

    For this particular watershed, the 100-year flood event had lower discharge when including the full set of data. My first instinct was that the discharge should be higher, but it is likely this watershed is located somewhere in the West where climate change is resulting in less discharge over time (rather than more discharge like in the East Coast). So their 100-year flood event is actually less severe than it would've been in the past. This was supported by the models ran and the visualization of discharge vs recurrence for both datasets.

    This is supported by looking at the discharge predictions from both models. There is about 700 cubic feet per second difference in the discharge values. Management decisions could differ depending on that number (how much does an area choose to dedicate to flood budgets, what should be the design flood elevation, etc)

    This was supported by the models ran on both datasets. The model with 30 years of data only accounted for about 70% of the variance in the data. But even when we add in another 24 years of data (almost twice as much!), we still only account for 72% of the variance. This may indicate we should consider using another method, perhaps the moving window.

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

    A bit. Not knowing the location of the site actually made me re-think my expectations for the results of the analysis and pay closer attention to the results of my model.

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

    I think it was comparable - my expectations were more influenced based on my own personal experiences and climates I am familiar with. But I believe the results may match up with theory for climates I am less familiar with (though it would bit a bit awkward if this watershed was in the East Coast- then I'd say real world data totally didn't match my expectations!)