Assignment 3: Physical Properties of Rivers

Jackie Van Der Hout

OVERVIEW

This exercise accompanies the lessons in Water Data Analytics 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, check your PDF against the key and then submit your assignment completion survey at https://forms.gle/ydeD5axzCnaNzgss9

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: 2022-02-08

Setup

1. Verify your working directory is set to the R project file. Load the tidyverse, dataRetrieval, lubridate, and lfstat packages. Set your ggplot theme (can be theme_classic or something else).
2. Import a data frame called "MysterySiteDischarge" from USGS gage site 03431700. Import discharge data starting on 1964-10-01 and ending on 2021-09-30. Rename columns 4 and 5 as "Discharge" and "Approval.Code". DO NOT LOOK UP WHERE THIS SITE IS LOCATED.
3. Build a ggplot of discharge over the entire period of record.

```
#1
getwd()
```

```
## [1] "C:/Users/Jackie/Box/Classes Spring 2022/Water Data Analytics/Water_Data_Analytics_2022/Assignmen
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(dataRetrieval)
library(lubridate)
```

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

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

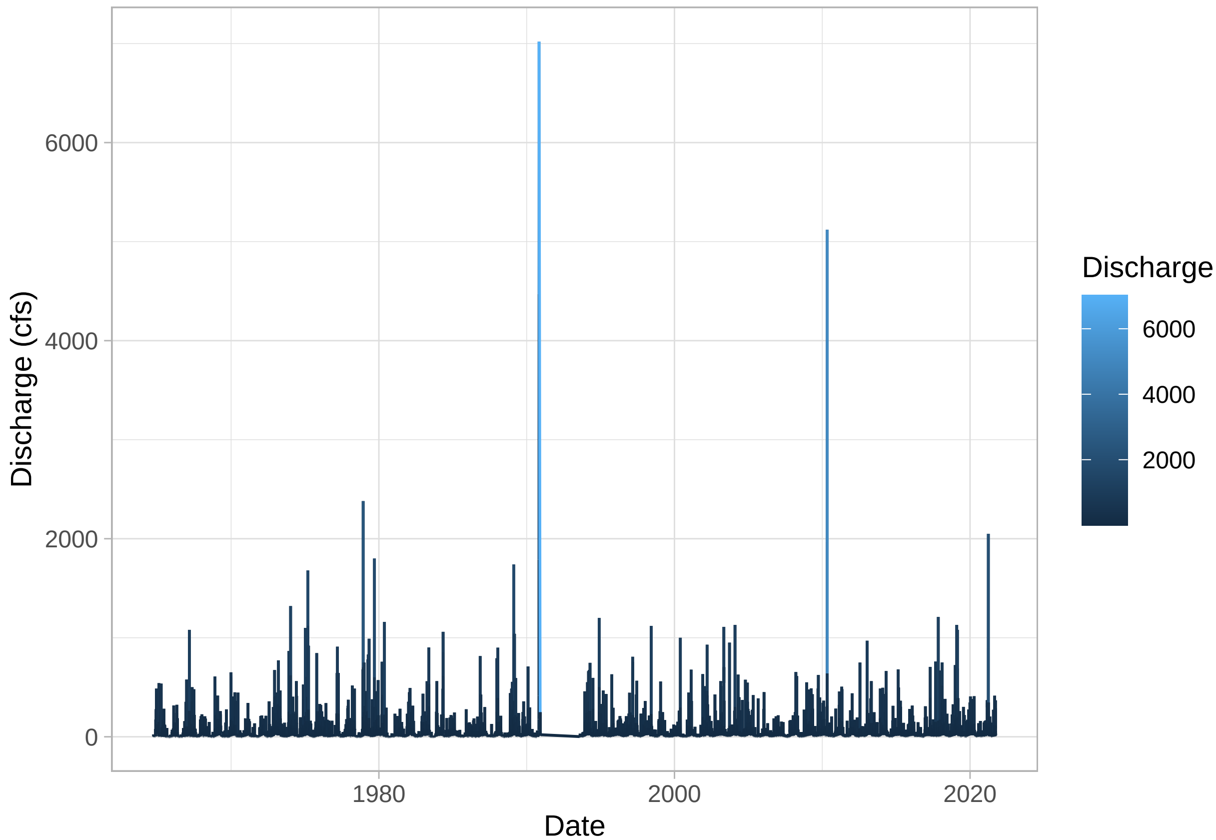
```r
library(lfstat)
```

```
## Loading required package: xts

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last

## Loading required package: lmom

## Loading required package: lattice
```

```r
theme_set(theme_light())

#2
MysterySiteDischarge <- readNWISdv(siteNumbers = "03431700",
                     parameterCd = c("00060", "00065"), #00060 discharge in cfs, 00065 is gage height
                     startDate = "1964-10-01",
                     endDate = "2021-09-30")

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

class(MysterySiteDischarge$Date) #checking that class is a date
```

```
## [1] "Date"
```

```r
#3
ggplot(MysterySiteDischarge, aes(x = Date, y = Discharge, color = Discharge)) +
  geom_line()+
  labs(x = "Date", y = "Discharge (cfs)")
```

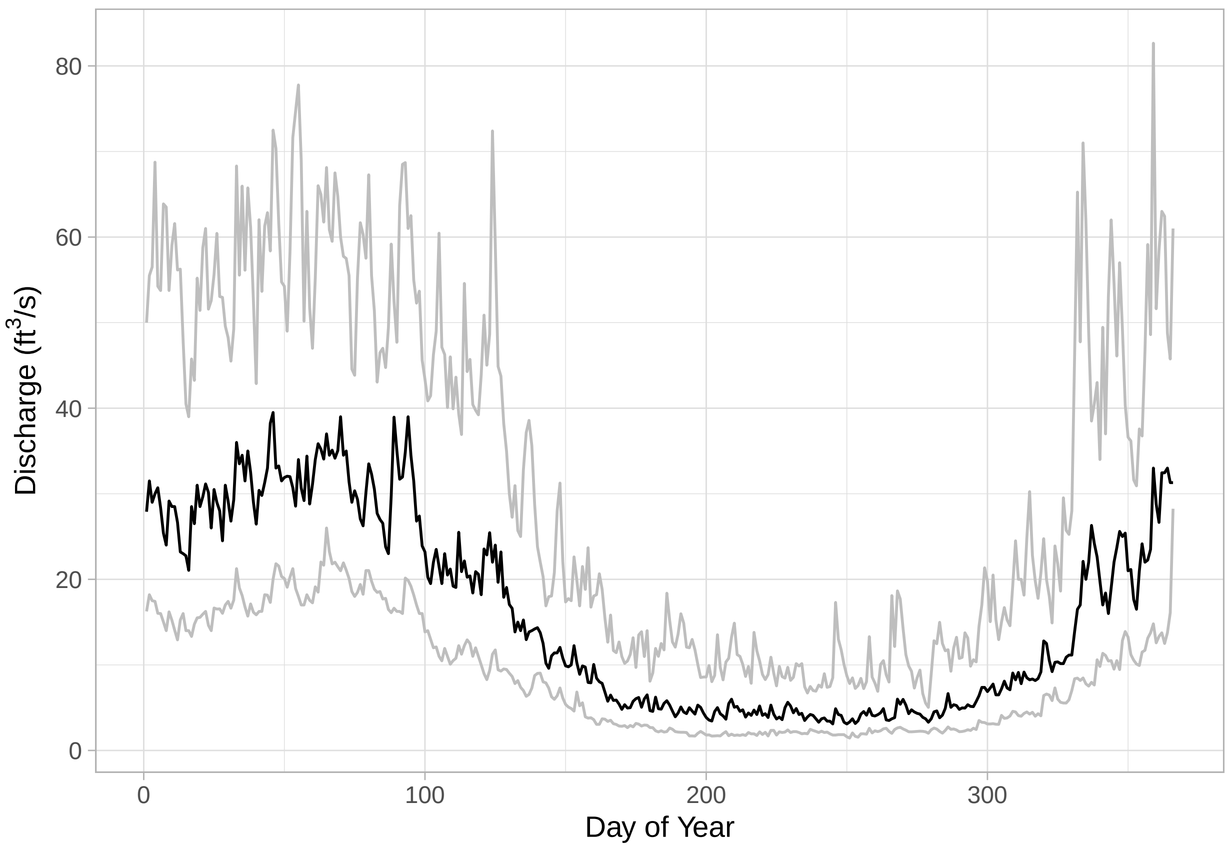


Analyze seasonal patterns in discharge

4. Add a "WaterYear" and "DayOfYear" column to the data frame. Hint: Use a pipe, and you will need both the lubridate and lfstat packages. Set WaterYear to numeric.
5. 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.
6. 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
#4
MysterySiteDischarge.time <- mutate(MysterySiteDischarge,
  DOY = yday(Date),
  WaterYear = water_year(Date, origin = "usgs")) #had to look up how to complete water_year function

#5
MysterySiteDischarge.Pattern <- MysterySiteDischarge.time %>%
  group_by(DOY) %>%
```

```
  summarise(Median.Discharge = median(Discharge, na.rm = TRUE),
            p75.Discharge = quantile(Discharge, 0.75, na.rm = TRUE),
            p25.Discharge = quantile(Discharge, 0.25, na.rm = TRUE))

#6
ggplot(MysterySiteDischarge.Pattern, aes(x = DOY))+
  geom_line(aes(y = p75.Discharge), color = "gray")+
  geom_line(aes(y = p25.Discharge), color = "gray")+
  geom_line(aes(y = Median.Discharge), color = "black")+
  labs(x = "Day of Year", y = expression("Discharge (ft"^3*"/s)"))
```



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

   This watershed has very seasonal discharge patterns, with high flows starting in the late fall / early winter and continuing through the early spring. The late spring and early summer appear to be dry with baseflows contributing to the majority of flows during that period of the year. High discharge events are pretty even spread throughout the wet season, indicating that they may be precipitation driven instead of snowmelt. If the precipitation were snowmelt, I would instead expect to see a steady peak in the spring as temperatures rise.

Create and analyze recurrence intervals

8. Create two separate data frames for MysterySite.Annual.30yr (first 30 years of record) and MysterySite.Annual.30yr (all years of record). Use a pipe to create your new data frame(s) that includes

the water year, the peak discharge observed in that year, a ranking of peak discharges, the recurrence interval, and the exceedende probability.

9. 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.

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

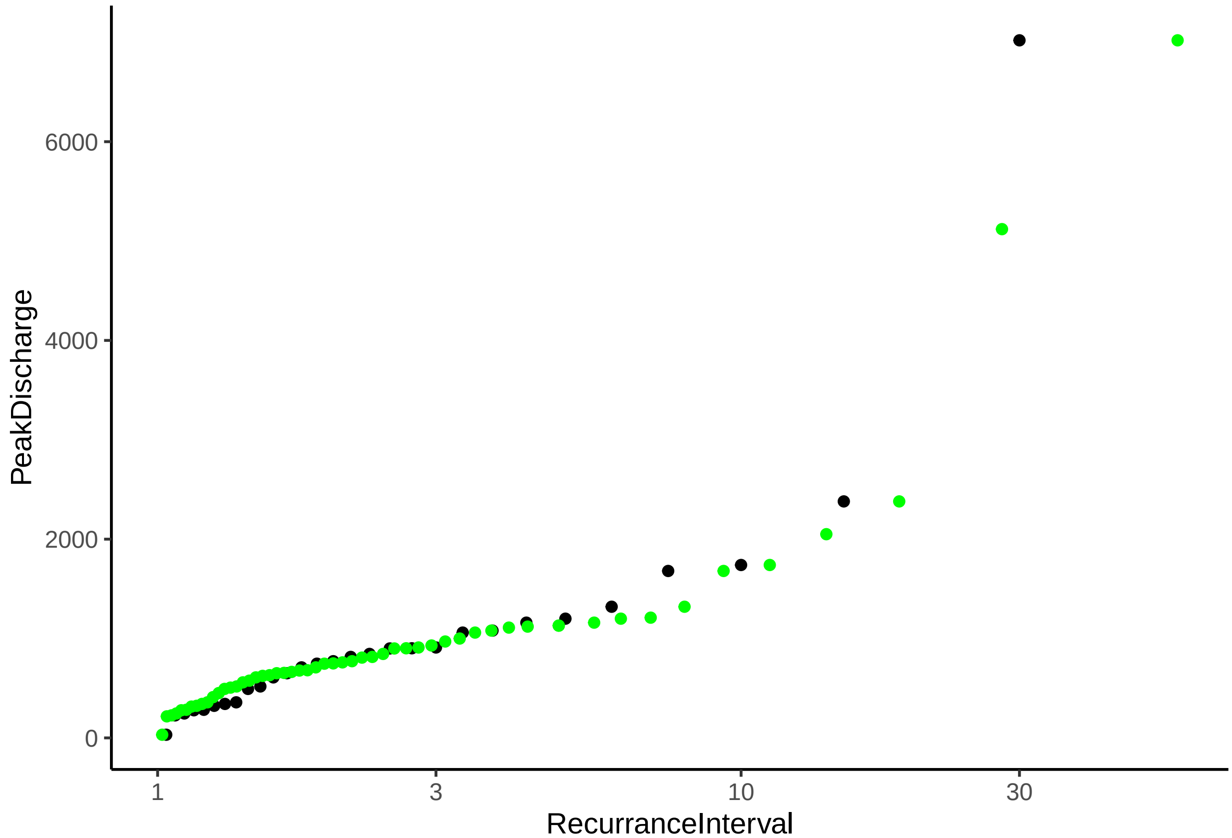
```
#8
class(MysterySiteDischarge.time$WaterYear) #class is factor
```

```
## [1] "factor"
```

```
MysterySiteDischarge.time$WaterYear <- as.numeric(as.character(MysterySiteDischarge.time$WaterYear))
#had to look at solutions for line 99

MysterySite.Annual.All <- MysterySiteDischarge.time %>%
  filter(WaterYear != "1965") %>% #all years of record
  group_by(WaterYear) %>%
  summarise(PeakDischarge = max(Discharge)) %>% #initially used mutate when it should have been summari
  mutate(RankDischarge = rank(-PeakDischarge),
         RecurranceInterval = (length(WaterYear)+1)/RankDischarge, #got help from Kateri on () placemen
         Exceedance = (1/RecurranceInterval))

summary(MysterySiteDischarge.time$WaterYear)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1965    1978    1994    1993    2008    2021
```

```
MysterySite.Annual.30yr <-  MysterySiteDischarge.time %>%
   filter(WaterYear < 1996) %>% #first 30 years of record
   filter(WaterYear != "1965") %>%
   group_by(WaterYear) %>%
   summarise(PeakDischarge = max(Discharge)) %>% #initially used mutate when it should have been summar
   mutate(RankDischarge = rank(-PeakDischarge),
         RecurranceInterval = (length(WaterYear)+1)/RankDischarge,
         Exceedance = (1/RecurranceInterval))

#9

RecurrancePlot <- ggplot(MysterySite.Annual.30yr, aes(x = RecurranceInterval, y = PeakDischarge))+
  geom_point()+
  geom_point(data = MysterySite.Annual.All, color = "green", aes(x = RecurranceInterval, y = PeakDischa
  scale_x_log10()+
  theme_classic()

RecurrancePlot
```




```
#had to look in example for how to combine two plots



#10 - 100 year flood prediction model

Mystery.RImodel.All <- lm(data = MysterySite.Annual.All, PeakDischarge ~ log10(RecurranceInterval))
summary(Mystery.RImodel.All)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log10(RecurranceInterval), data = MysterySite.Annual.All)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -864.93 -239.06   80.34  207.60 2755.47
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -37.38     112.14  -0.333     0.74
## log10(RecurranceInterval)  2460.78     197.01  12.491   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 562.9 on 53 degrees of freedom
## Multiple R-squared:  0.7464, Adjusted R-squared:  0.7417
```

```
## F-statistic:   156 on 1 and 53 DF,  p-value: < 2.2e-16
```

```
Mystery.RImodel.30 <- lm(data = MysterySite.Annual.30yr, PeakDischarge ~ log10(RecurranceInterval))
summary(Mystery.RImodel.30)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log10(RecurranceInterval), data = MysterySite.Annual.30yr)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -991.4 -305.3  123.5  201.9 2905.2
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -168.1      194.2  -0.865    0.394
## log10(RecurranceInterval)   2899.5      355.2   8.163  9.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 691.8 on 27 degrees of freedom
## Multiple R-squared:  0.7117, Adjusted R-squared:  0.701
## F-statistic: 66.64 on 1 and 27 DF,  p-value: 9.099e-09
```

```
Mystery.RImodel.All$coefficients[1] + Mystery.RImodel.All$coefficients[2]*log10(100) #4884.187
```

```
## (Intercept)
##    4884.187
```

```
Mystery.RImodel.30$coefficients[1] + Mystery.RImodel.30$coefficients[2]*log10(100) #5630.873
```

```
## (Intercept)
##    5630.873
```

11. 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 30 year model shows a 100 year flood event at approximately 5,630 cfs, whereas the full model shows the 100 year flood event at approximately 4,884 cfs. The first thirty years of record show slightly higher discharge 100 year events, indicating that the size of flood events may slightly be decreasing over time. I am curious if a paried T-test of model results from flood events at other recurrance intervals would show any significant trends, or if the difference between the 5,630 cfs and the 4,884 cfs 100 year events is small enought to not show any significant changes to the stationarity of the river. If I had to guess, I would say that the river is likeley remaining stationary over time.