Assignment 3: Physical Properties of Rivers

Keith Bollt

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] "Z:/Hydrologic_Data_Analysis2"
```

```
#install.packages("cowplot")
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 ---------------------------------------------------------------------------- tid
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

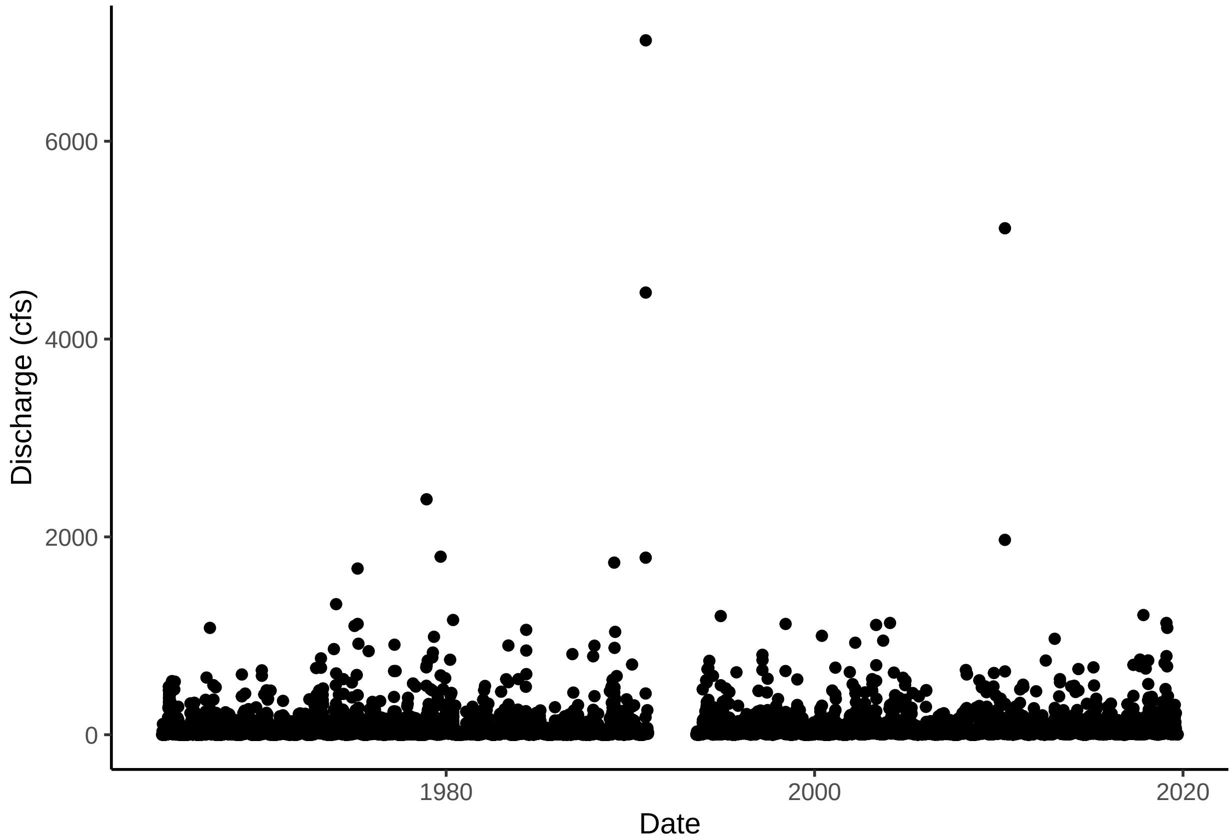
```
library(dataRetrieval)
library(lubridate)
```

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

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

```r
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())
## ********************************************************
##
## Attaching package: 'cowplot'
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```r
theme_set(theme_classic())

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

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

Mysterysite.dischargeplot <- ggplot(MysterySiteDischarge, aes(x= Date, y = Discharge))+
  geom_point()+
  labs(y = "Discharge (cfs)")

print(Mysterysite.dischargeplot)
```

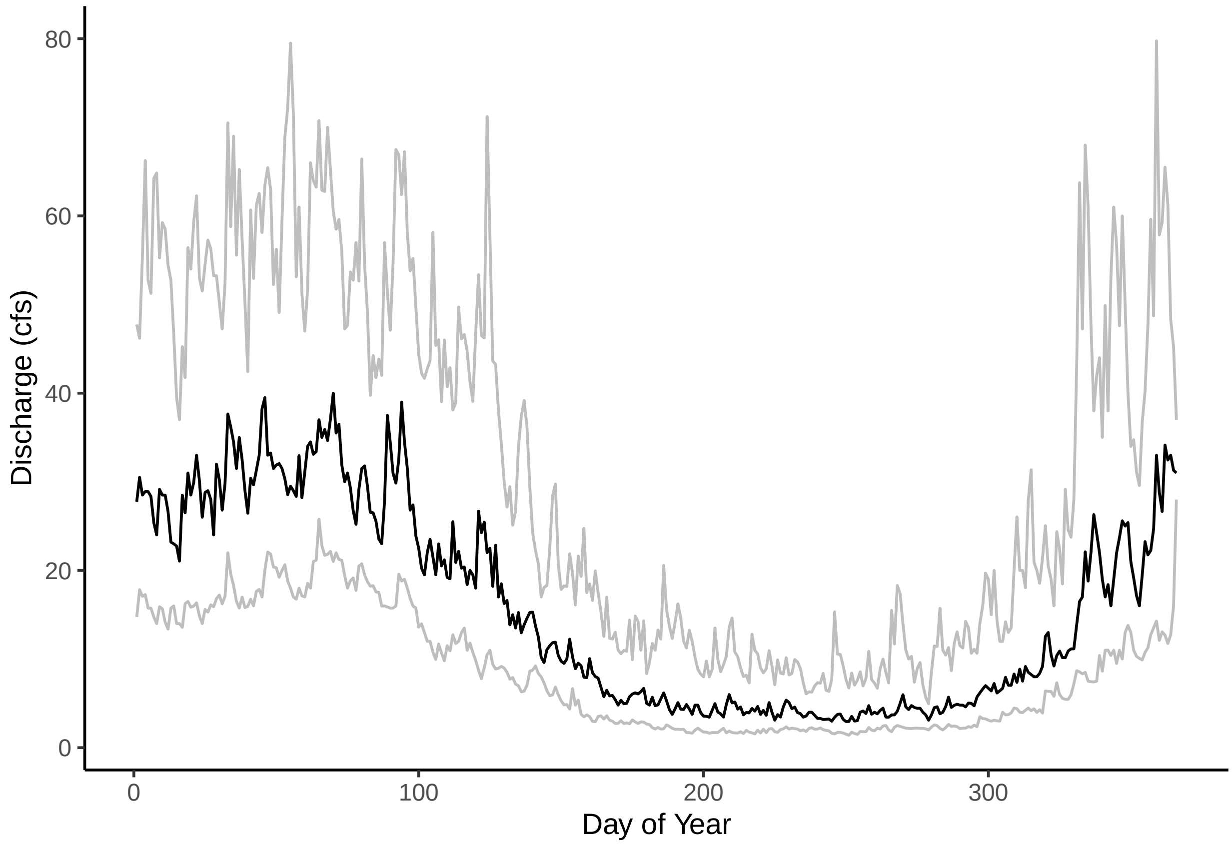


Analyze seasonal patterns in discharge

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(
    FirstQuartile.Discharge = quantile(Discharge, probs=(0.25)),
    Median.Discharge = quantile(Discharge, probs=(0.50)),
    ThirdQuartile.Discharge = quantile(Discharge, probs=(0.75)))

Pattern.plot <- ggplot(MysterySiteDischarge.Pattern)+
  geom_line(aes(x = Day.of.Year, y = FirstQuartile.Discharge), color = "gray")+
  geom_line(aes(x = Day.of.Year, y = Median.Discharge))+
```

```
  geom_line(aes(x = Day.of.Year, y = ThirdQuartile.Discharge), color = "gray")+
  labs(x = "Day of Year", y = "Discharge (cfs)")+
  theme(legend.position = "none")

print(Pattern.plot)
```



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

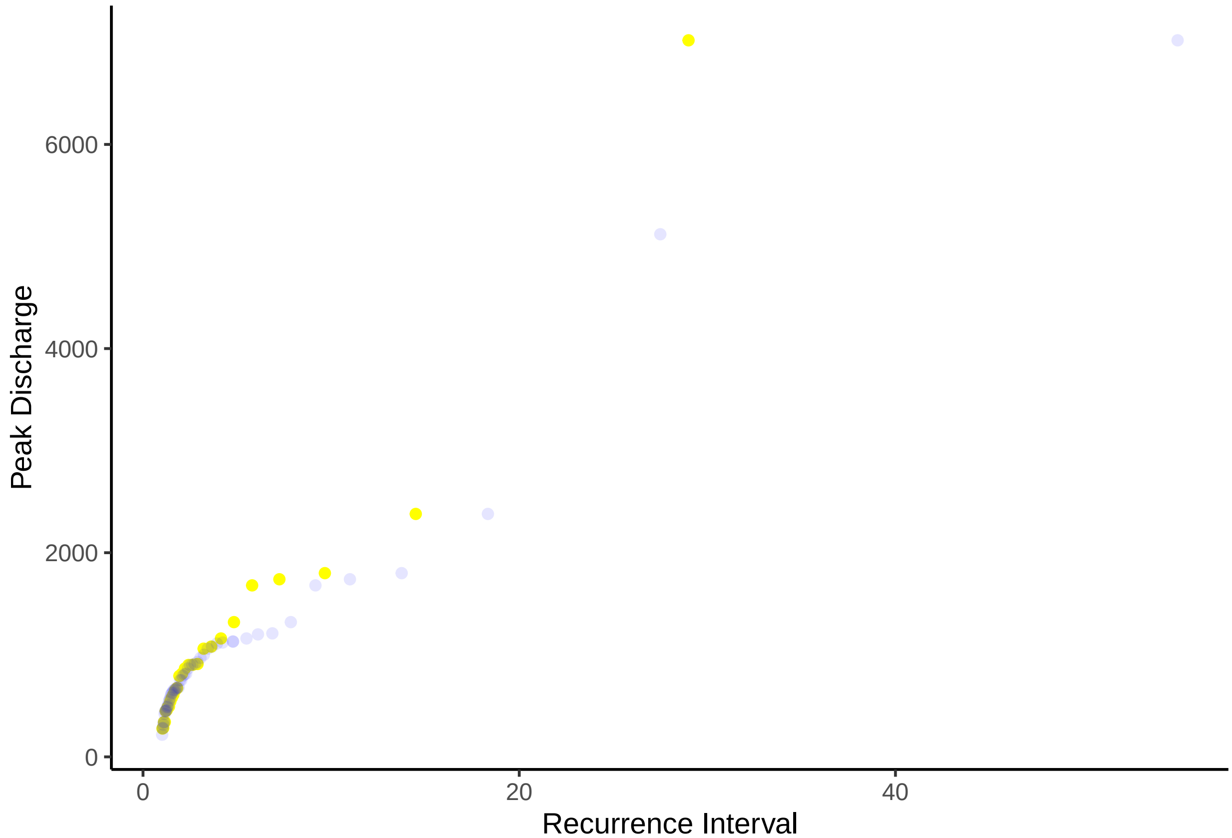
   There is a cyclical cycle to all three quantiles. Discharge is lowest in late summer, and also has little variation during this time. The discharge increases around November 1, and stays pretty high through April. The variation is also higher during this time. This tells us that this watershed sees dry summers, wet winters, large winter storms/snowmelt events, and few large summer storms. I would guess that this river is in a cold climate but not near high mountains.

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.

```r
#Creating the dataframes
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)

#Creating the plot
names(MysterySite.Annual.30yr)[2:5] <- c("PeakDischarge30", "Rank30", "RecurrenceInterval30",
                                          "Probability30")

MysterySitejoin <- full_join(MysterySite.Annual.30yr, MysterySite.Annual.Full, by = "Year")

MysterySiteRecurrence <- ggplot(MysterySitejoin)+
  geom_point(aes(x = RecurrenceInterval30, y = PeakDischarge30), color = "Yellow")+
  geom_point(aes(x = RecurrenceInterval, y = PeakDischarge), color = "Blue", alpha = 0.1)+
  labs(x = "Recurrence Interval", y = "Peak Discharge")

print(MysterySiteRecurrence)
```

```
## Warning: Removed 26 rows containing missing values (geom_point).
```




```r
#30 year data modelling a 100 year flood
MysterySite30.RImodel <- lm(data = MysterySitejoin, PeakDischarge30 ~ log(RecurrenceInterval30))
summary(MysterySite30.RImodel)
```

```
##
## Call:
## lm(formula = PeakDischarge30 ~ log(RecurrenceInterval30), data = MysterySitejoin)
##
## 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(RecurrenceInterval30)   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
##   (26 observations deleted due to missingness)
## 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
MysterySite30.RImodel$coefficients[1] + MysterySite30.RImodel$coefficients[2]*log(100)
```

```
## (Intercept)
```

```
##    5758.621
```

```r
#5758.6 cfs

#Full Dataset modelling a 100 year flood
MysterySiteFull.RImodel <- lm(data = MysterySitejoin, PeakDischarge ~ log(RecurrenceInterval))
summary(MysterySiteFull.RImodel)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = MysterySitejoin)
##
## 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
MysterySiteFull.RImodel$coefficients[1] + MysterySiteFull.RImodel$coefficients[2]*log(100)
```

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

```r
#4843.7 cfs
```

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 full dataset started above the 30 year plot, but was "passed" around a recurrence interval of 5 years. For larger recurrence intervals such as the prediction of a 100 year flood, the 30 year plot predicts higher flow rates than does the full dataset. This tells us that stationarity of discharge is probably not a great assumption for River X. If it was, we would see the 30 year plot and the full dataset plot almost mirror eachother. Instead, we see that fewer extreme flooding events are occurring over time.

Reflection

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

    There are some areas where rivers are actually seeing fewer extreme flooding events. A recurrence interval vs peak discharge graph that will predict a lower number of extreme flooding events may actually predict a higher number of smaller flooding events. It is hard to nail down exactly where a river is located based just on its hydrograph.

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

    My first conclusion is based off of the model I ran in question 11. I found that the full dataset predicted a lower 100 year flood than did the 30 year dataset. My second conclusion is based off of looking at the graph I generated in question 10. The full dataset graph (Blue) starts off above the 30 year dataset graph (Yellow), but is passed by it. My third conclusion is based off of chewing on where I think this river is located. I am trying to think where it has gotten drier but it is not mountainous. Probably somewhere west of the Mississippi and east of the Rockies.

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

    Sure. I got to visualize how stationarity of discharge isn't a good assumption by graphing on question 11. Really all my graphs illuminated the ideas we've been discussing in class about hydrographs.

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

    I would have guessed that the full dataset would have predicted a higher 100 year flood than the 30 year dataset. I suppose climate change is also increasing drought.