Assignment 3: Physical Properties of Rivers

Jake Greif

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

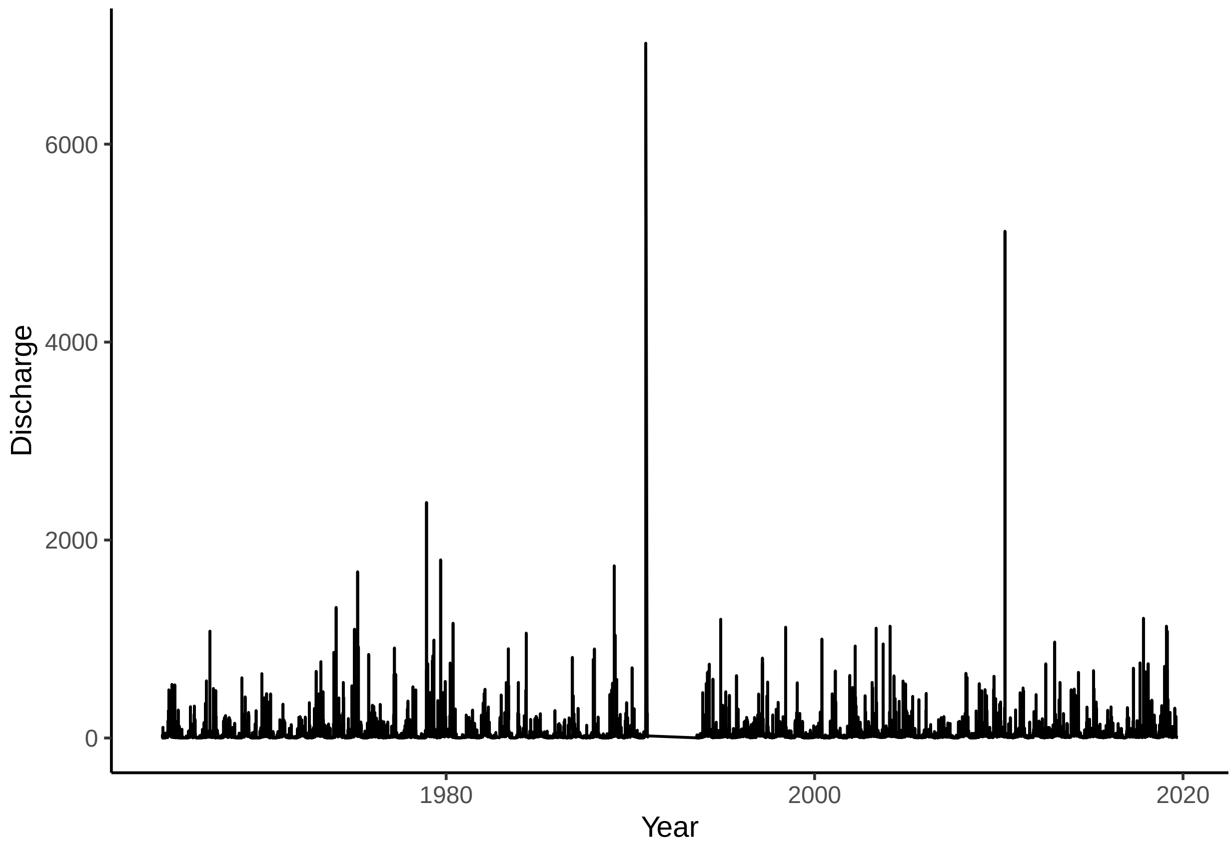
1. Verify your working directory is set to the R project file,
2. Load the tidyverse, dataRetrieval, and cowplot packages
3. Set your ggplot theme (can be theme_classic or something else)
4. Import a data frame called "MysterySiteDischarge" from USGS gage site 03431700. Upload all discharge data for 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.

```
# Check Working Directory
getwd()
```

```
## [1] "/Users/jakegreif/Duke/Fall_2019/Hydrologic_Data_Analysis"
```

```
# Load Packages
library(tidyverse)
library(dataRetrieval)
library(cowplot)
library(lubridate)

# Import Data
MysterySiteDischarge <- readNWISdv(siteNumbers = "03431700",
                     parameterCd = "00060",
                     startDate = "",
                     endDate = "")

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

# Set ggplot Theme
mytheme <- theme_classic()

theme_set(mytheme)
```

```r
# Discharge over entire period
MystPlot <-
  ggplot(MysterySiteDischarge, aes(x = Date, y = Discharge)) +
         geom_line() +
         xlab("Year")
print(MystPlot)
```

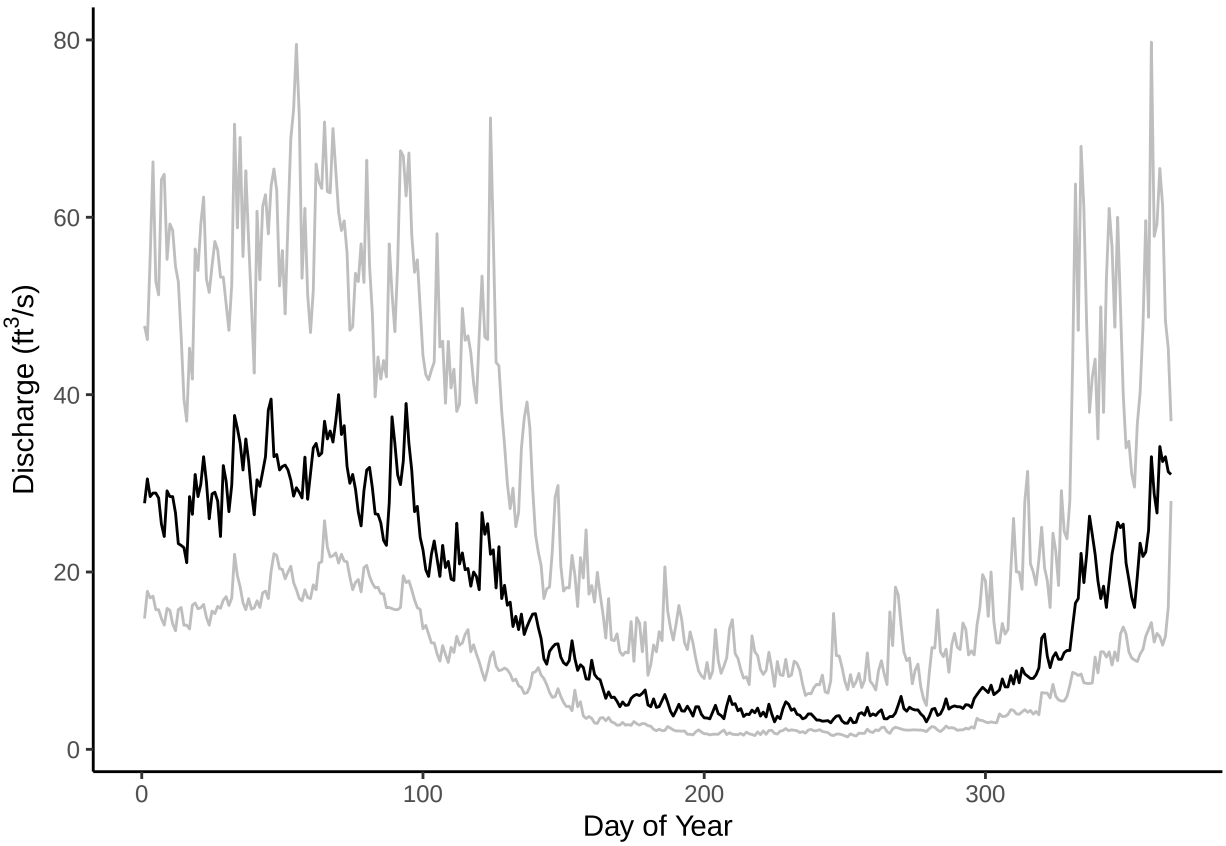


Analyze seasonal patterns in discharge

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
# Add Year column
MysterySiteDischarge <-
  MysterySiteDischarge %>%
  mutate(Year = year(Date))

# Add Day.of.Year column
MysterySiteDischarge <- mutate(MysterySiteDischarge,
                                  DOY = yday(Date))
```

```r
# Calculate median discharge, create columns for 25th/75th percentiles
MysterySiteDischarge.Pattern <- MysterySiteDischarge %>%
  group_by(DOY) %>%
  summarise(Median.Discharge = median(Discharge),
            Discharge.75th = quantile(Discharge, probs = 0.75),
            Discharge.25th = quantile(Discharge, probs = 0.25))

# Create plot of median, 75th/25th quantiles
MysteryPatternPlot <-
  ggplot(MysterySiteDischarge.Pattern, aes(x = DOY)) +
  geom_line(aes(y = Median.Discharge)) +
  geom_line(aes(y = Discharge.75th), color = "gray") +
  geom_line(aes(y = Discharge.25th), 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?

   The mystery river has relatively high flow during the winter months, and discharge is noticably lower throughout the summer and fall. The discharge only noticeably increases again around October/November. This river must be in a region that primarily has winter precipitation, and a dry summer.

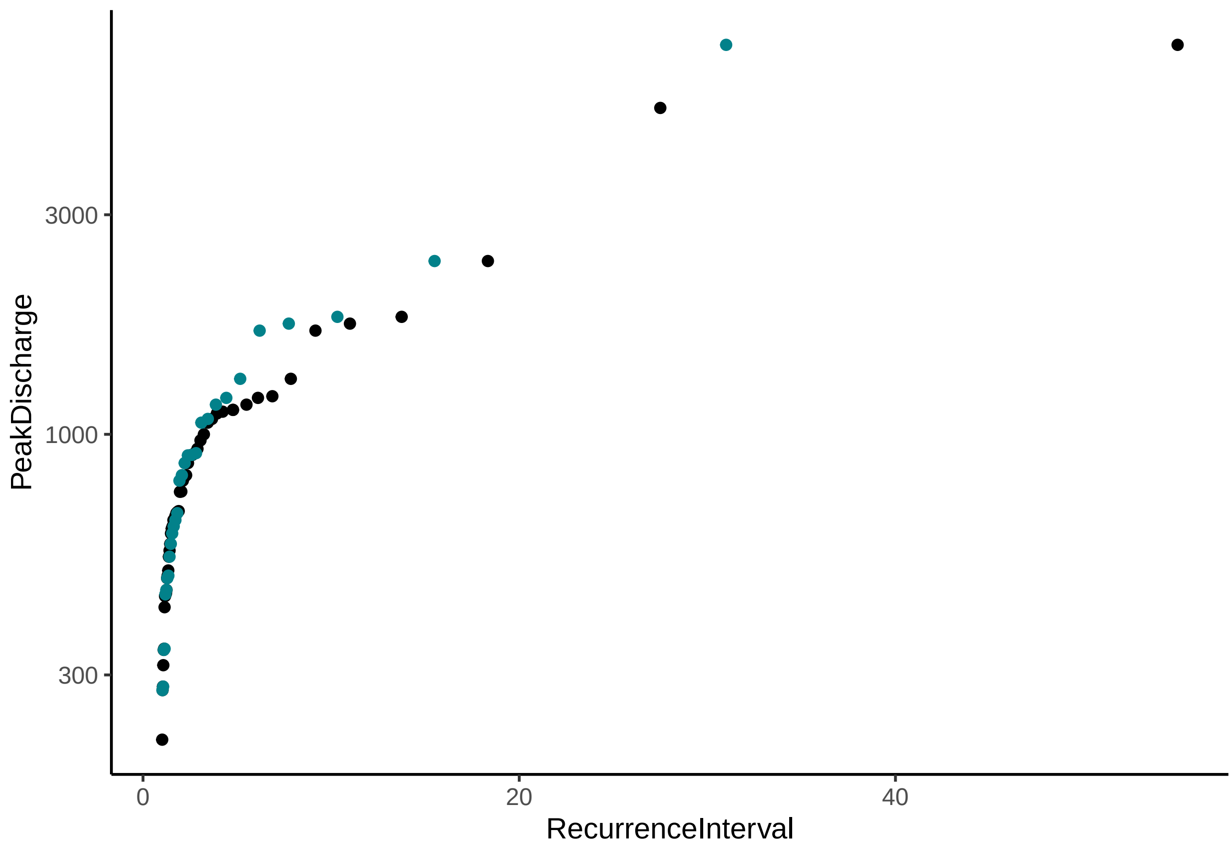
Create and analyze recurrence intervals

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.

```
# Create MysterySite.Annual.30yr data frame
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)

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

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.

```
# Create discharge vs. recurrence interval plots
MystRecurrencePlot.Combined <-
  ggplot(MysterySite.Annual.Full,
         aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() +
  geom_point(data = MysterySite.Annual.30yr, color = "#02818a",
             aes(x = RecurrenceInterval, y = PeakDischarge)) +
  scale_y_log10()
print(MystRecurrencePlot.Combined)
```



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

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

```
Myst.RImodel.30yr$coefficients
```

```
##             (Intercept) log(RecurrenceInterval)
```

```
##                -69.87363              1217.79015
```

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

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

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

```
Myst.RImodel.Full$coefficients
```

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

```
Myst.RImodel.Full$coefficients[1] + Myst.RImodel.Full$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?

    The 30 year recurrence interval plot is steeper and the 100-year flood prediction is greater than the recurrence interval and 100-year flood discharge derived from the full dataset. Overall, the first 30 years saw higher discharges compared to the full dataset, and therefore using only the first 30 years would cause one to over-estimate the discharge of certain flood events. This tells us that the discharge of the river is not stationary.

Reflection

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

    The timing and magnitude of discharge in rivers across the country are highly dependent on climate and precipitation, but both timing and magnitude do not stay constant. They change are changing with a changing climate and precipitation patterns, therefore the idea of stationarity does not apply well to many rivers.

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

    Observing the mean discharge on each day of the year over the course of data collection period made it easier to identify seasonal trends in discharge. Also, comapring the recurrence intervals of the first 30 years vs. the full data set tested the assumption of stationarity.

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

    Yes, but I think it would have been a little more difficult to grasp some concepts without a theory-based background. Being able to manipulate the data myself allowed me to get familiar with it at my own pace, and I was able to think about why the data is the way it is by comparing it to my prior theory-based lessons.

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

    Finding that the 30 year recurrence intervals did not match up with the full data recurrence intervals was technically surprising. However, when I initially learned the theory, I was taught then that stationarity was dead. So, while the results were what I expected, compared to traditional theory that assumes there is stationarity, these results are surprising.