

Assignment 3: Physical Properties of Rivers

Yutao Gong

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] "C:/Users/gongy/Documents/Hydrologic_Data_Analysis/Assignments"

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())

## *****

library(lubridate)

##
## Attaching package: 'lubridate'

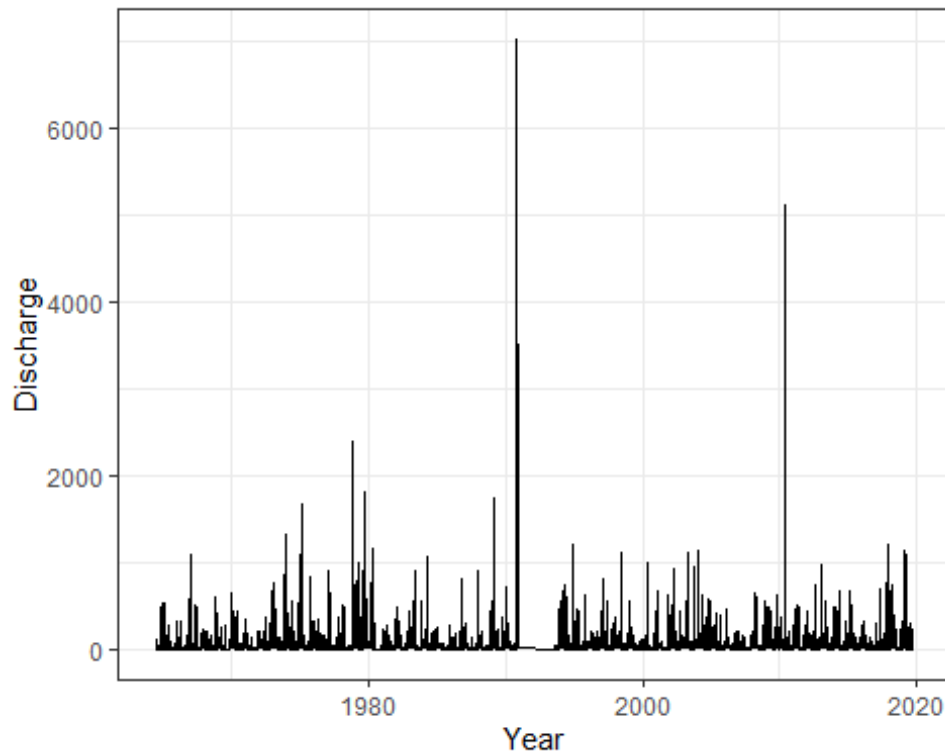
## The following object is masked from 'package:cowplot':
##
## stamp

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

theme_set(theme_bw())

MysterySiteDischarge <- readNWISdv(siteNumbers = "03431700",
                                   parameterCd = "00060", # discharge (ft3/s)
                                   startDate = "",
                                   endDate = "")
names(MysterySiteDischarge)[4:5] <- c("Discharge", "Approval.Code")
MysteryPlot <-
  ggplot(MysterySiteDischarge, aes(x = Date, y = Discharge)) +
    geom_line() +
    xlab("Year")
print(MysteryPlot)

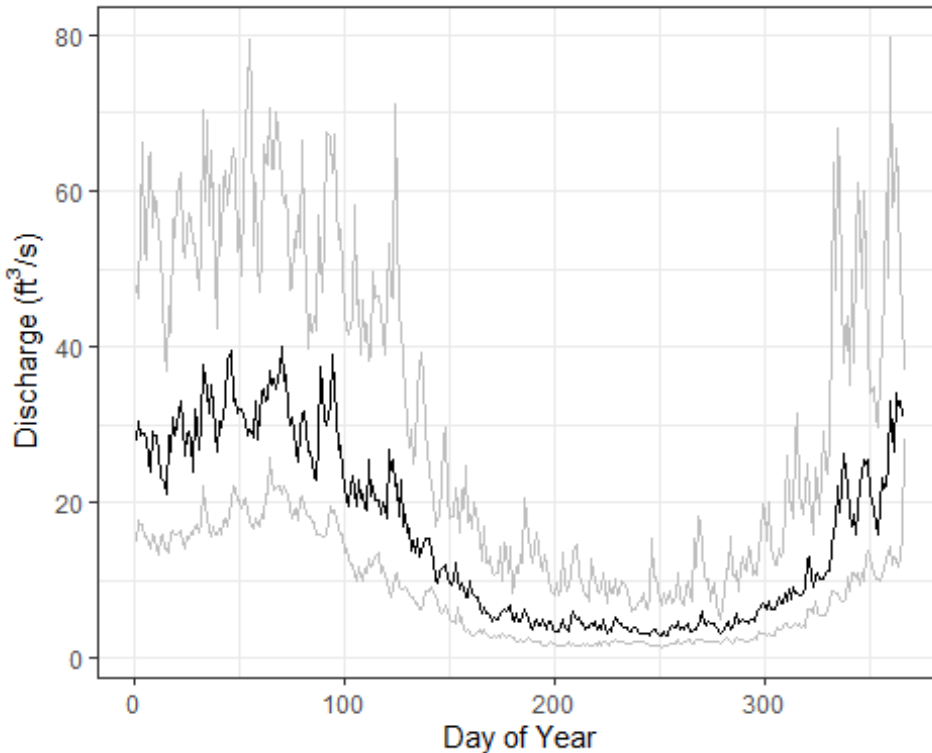
```



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.

```
MysterySiteDischarge <-
  MysterySiteDischarge %>%
  mutate(Year = year(Date), Day.of.Year = yday(Date))
MysterySiteDischarge.Pattern <- MysterySiteDischarge %>%
  group_by(Day.of.Year) %>%
  summarise(Median.Discharge = median(Discharge),
            S.Discharge = quantile(Discharge, 0.75),
            T.Discharge = quantile(Discharge, 0.25))
MysterySitePatternPlot <-
  ggplot(MysterySiteDischarge.Pattern, aes(x = Day.of.Year)) +
  geom_line(aes(y = Median.Discharge)) +
  geom_line(aes(y = S.Discharge), color = "gray") +
  geom_line(aes(y = T.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?

In general, the discharge level is higher in winter times and decreases as it goes from winter to spring and summer. Summertime actually displays the lowest discharge level. The 75th percentile displays more significant difference in discharge between summer and winter, compared to median and 25th percentile. This implies that the precipitation at this site concentrates in colder seasons, probably very little in summer. The climate is warm and dry in summer and cold and wet in winter (probably Mediterranean climate type)

Create and analyze recurrence intervals

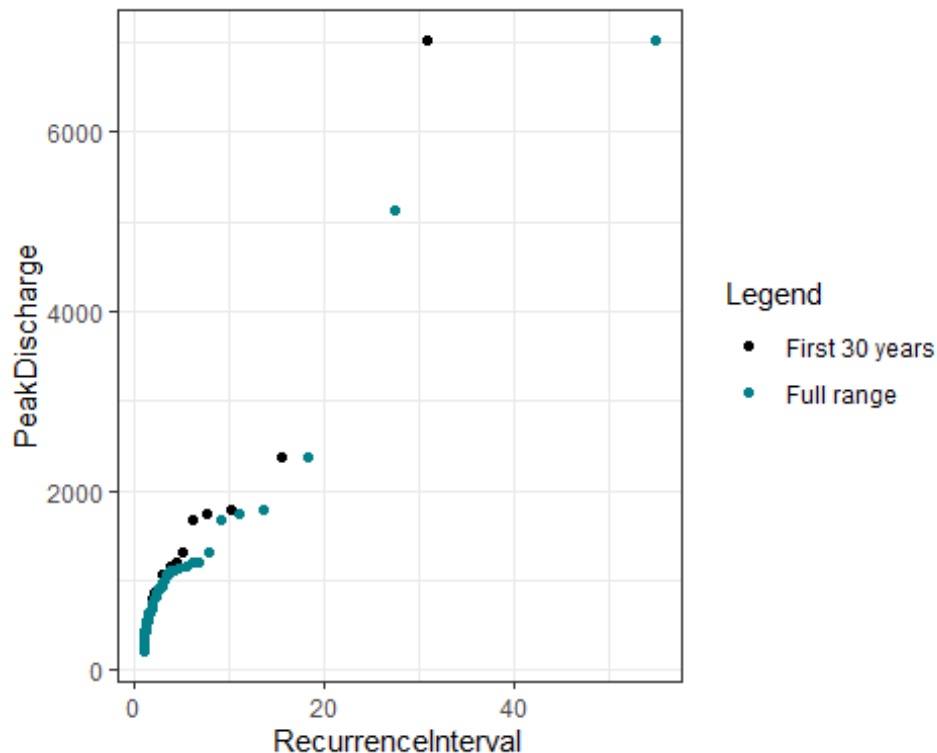
9. Create two separate data frames for `MysterySite.Annual.30yr` (first 30 years of record) and `MysterySite.Annual.Full` (all years of record). Use a pipe to create your new data frame(s) that includes the year, the peak discharge observed in that year, a ranking of peak discharges, the recurrence interval, 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.

```
MysterySite.Annual.30yr <-
  MysterySiteDischarge %>%
  filter(Year < 1996) %>% #included 1994 and 1995 because missing data from
  1991 and 1992
```

```

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)
MysteryRecurrencePlot <-
  ggplot(MysterySite.Annual.30yr, aes(x = RecurrenceInterval, y =
PeakDischarge, color="First 30 years")) +
  geom_point() +
  geom_point(data = MysterySite.Annual.Full,
            aes(x = RecurrenceInterval, y = PeakDischarge, color = "Full
range")) +
  scale_colour_manual("Legend", values = c("First 30 years" = "black", "Full
range" = "#02818a"))
print(MysteryRecurrencePlot)

```



```

#Model for first 30 years of record
MysterySite.RImodel.30yr <- lm(data = MysterySite.Annual.30yr, PeakDischarge
~ log(RecurrenceInterval))
summary(MysterySite.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

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

## (Intercept)
##      5538.257

#Model for all years of record
MysterySite.RImodel.Full <- lm(data = MysterySite.Annual.Full, PeakDischarge
~ log(RecurrenceInterval))
summary(MysterySite.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
```

```
MysterySite.RImodel.Full$coefficients[1] +  
MysterySite.RImodel.Full$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 recurrence interval plot of full data displays a gentler trend, i.e. peak discharge with longer recurrence intervals do not jump dramatically. The recurrence intervals deducted from first 30 year data are generally shorter than those deducted from full data, except from one outlier. However, this outlier has a high leverage which makes the prediction of the discharge of a 100-year flood based on first 30 year data higher than the prediction made based on full data. This tells us that the discharge in this river shows somewhat nonstationarity, because there has been outliers, which makes prediction of future events based on past data less reliable.

Reflection

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

1. Prediction of discharge recurrence intervals is very dependent on past data, and sometimes vulnerable to extreme events.
2. High discharge level is more affected by precipitation and weather compared to low discharge level.

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

For 1, the two models using first 30 years data and full data respectively supported my conclusion because they gave rather different predictions for discharge level of a 100-year flood. For 2, the visualization of 75th percentile, mean, 25th percentile of discharge vs day of year supported my conclusion, because we can see in seasons when discharge level is high, 75th percentile exceeds mean way more compared to seasons when discharge level is low.

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

Yes, by visualizing the data and building the models by my own, I was able to approach the knowledge as if I were solving for a puzzle. It helps me understand the knowledge more.

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

I was expecting the prediction of 100-year flood using full data will be higher than that using first 30 year data, because of the impact of climate change. However, I guess the outlier outweighs climate change :)