Assignment 3: Physical Properties of Rivers

Gaby Garcia

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.

```
setwd("/Users/gabrielagarcia/Desktop/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Assignments")

library(tidyverse)
library(dataRetrieval)
library(cowplot)
library(lubridate)
library(scales)
library(purrr)
library(broom)

gabytheme <- theme_bw(base_size = 14) +
  theme(plot.title=element_text(face="bold", size="20", color="IndianRed3", hjust=0.5),
        axis.title=element_text(size=13, color="black"),
axis.text = element_text(face="bold", size=10, color = "black"),
panel.background=element_rect(fill="white", color="darkblue"),
panel.border = element_rect(color = "black", size = 2),
legend.position = "top", legend.background = element_rect(fill="white", color="black"),
            legend.key = element_rect(fill="transparent", color="NA"))

theme_set(gabytheme)
```

Import MysterySiteDischarge Data Frame

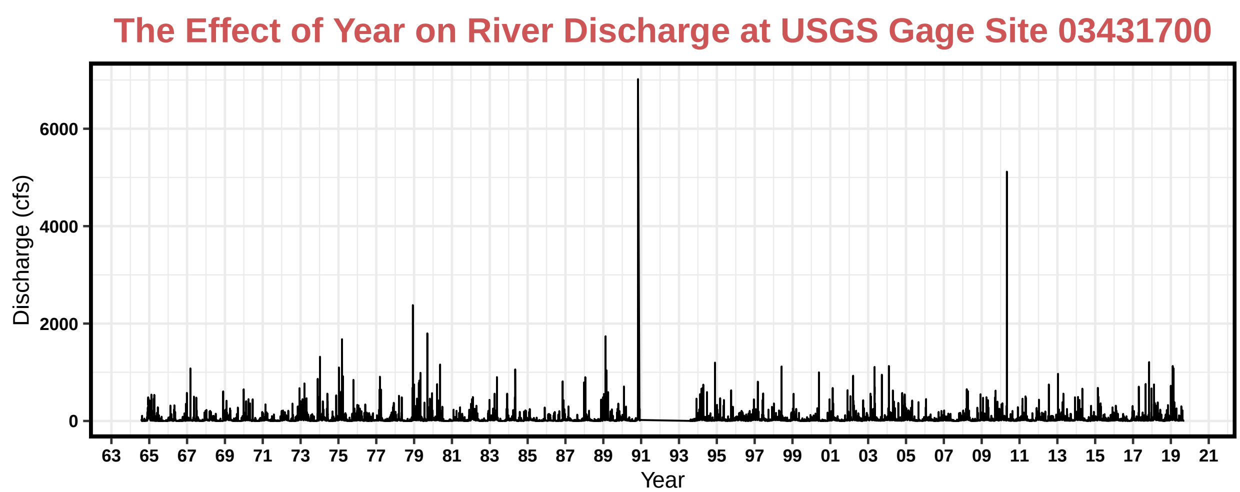
```
MysterySiteDischarge <- readNWISdv(siteNumbers = "03431700",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")
```

Rename Columns 4 and 5 as "Discharge" and "Approval.Code"

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

GGPlot of Discharge over Entire Record

```
library(scales)
MysteryPlot <-
  ggplot(MysterySiteDischarge, aes(x = Date, y = Discharge)) +
         geom_line() +
  labs(title="The Effect of Year on River Discharge at USGS Gage Site 03431700",
       x ="Year", y = "Discharge (cfs)")+
  scale_x_date(labels = date_format("%y"),
               breaks = date_breaks("2 year")) +
  gabytheme
print(MysteryPlot)
```



Analyze seasonal patterns in discharge

5. Add a "Year" and "Day.of.Year" column to the data frame.

Add Year Column

```
library(lubridate)
MysterySiteDischarge<-
MysterySiteDischarge%>%
  mutate(Year = year(Date))
```

Add Day of Year Column

```
MysterySiteDischarge<-
MysterySiteDischarge%>%
  mutate(Day.of.Year = yday(Date))
```

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.

Create new data frame

```
MysterySiteDischarge.Pattern<-MysterySiteDischarge%>%
   group_by(Day.of.Year) %>%
  mutate(qnt_25  = quantile(Discharge, probs= 0.25),
           qnt_75  = quantile(Discharge, probs= 0.75),
          MedianDischarge=median(Discharge)) %>%
   distinct(qnt_25 ,qnt_75, MedianDischarge)
```

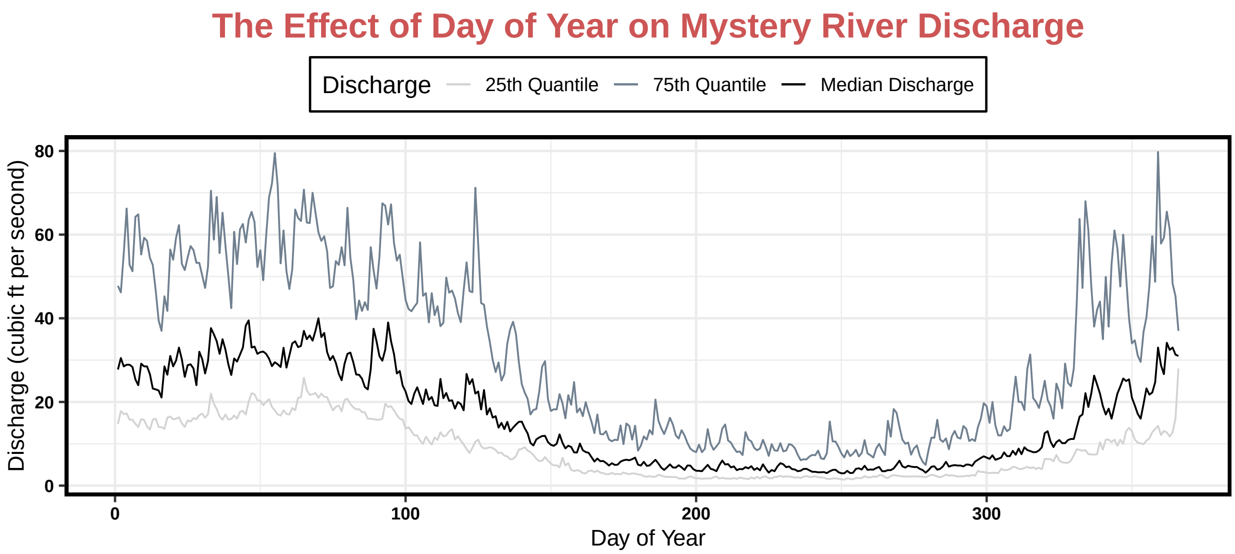
7. Create a plot of median, 75th quantile, and 25th quantile discharges against day of year. Median should be black, other lines should be gray. #make one graph and overlay lines

Create Median, 75th, and 25th Quartile Plot

```
DischargePlot <-
  ggplot(data=MysterySiteDischarge.Pattern) +
  geom_line(data=MysterySiteDischarge.Pattern,
            aes(x =Day.of.Year, y=MedianDischarge, color = "Median Discharge")) +
  geom_line(data =MysterySiteDischarge.Pattern,
            aes(x =Day.of.Year, y = qnt_25, color="25th Quantile")) +
  geom_line(data=MysterySiteDischarge.Pattern,
            aes(x =Day.of.Year, y = qnt_75, color = "75th Quantile"))+
  labs((fill = "Discharge Type"),
       title="The Effect of Day of Year on Mystery River Discharge",
       y ="Discharge (cubic ft per second)", x = "Day of Year")+
  gabytheme+
  scale_colour_manual(values=c('lightgrey', 'slategray','black'))+
guides(color = guide_legend(title = "Discharge"))



print(DischargePlot)
```



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

I can first deduce from my graph that this river is within a non-arid climate and potentially a climate with dry summers and wet/humid winters, as there are consistent precipitation events with high magnitudes until the late spring. If this river was within a semi-arid or arid climate (such as the Verde River in Arizona from Lesson 5), there would be far fewer precipitation events of great magnitude and longer stretches of baseflow along the hydrograph's period of record. In addition, the magnitude of river discharge amounts begin to gradually decline after around Day 125, which is in early May; the number and magnitude of significant precipitation events then begins to steadily rise around Day 300, or late October. Thus,the wet winter and dry summer climate displayed by this graph make me suspect that this river may be in the Mid-Atlantic region in the United States, such as Virginia where I'm from. Considering the median discharge curve, for any given day of the year, there is a higher range of maximum river discharge amounts (range from 25th quartile to median to 75th quartile) from January 1 (Day 1) to early May, then the range in maximum river discharge declines throughout the summer and increases again around October.

The type of precipitation can also have an impact on the hydrograph. The lag time from the precipitation event (time from max precipitation amount to peak discharge) is likely to be greater if the precipitation is snow rather than rain because snow takes time to melt before the water enters the river channel. When there is rapid melting of snow, the peak discharge could be high. While I can't deduce lag times from this graph because consecutive data points aren't necessarily from the same precipitation event, I think these precipitation events are rain and not snow based on the Bitteroot River Hydrograph created in Lesson 5; colder or mountainous climate zones would potentially have hydrographs with a few large peak discharge events centered around the early spring when snowmelt typically occurs.

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.

Create Data Frame for First 30 Years of Record

```
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)
```

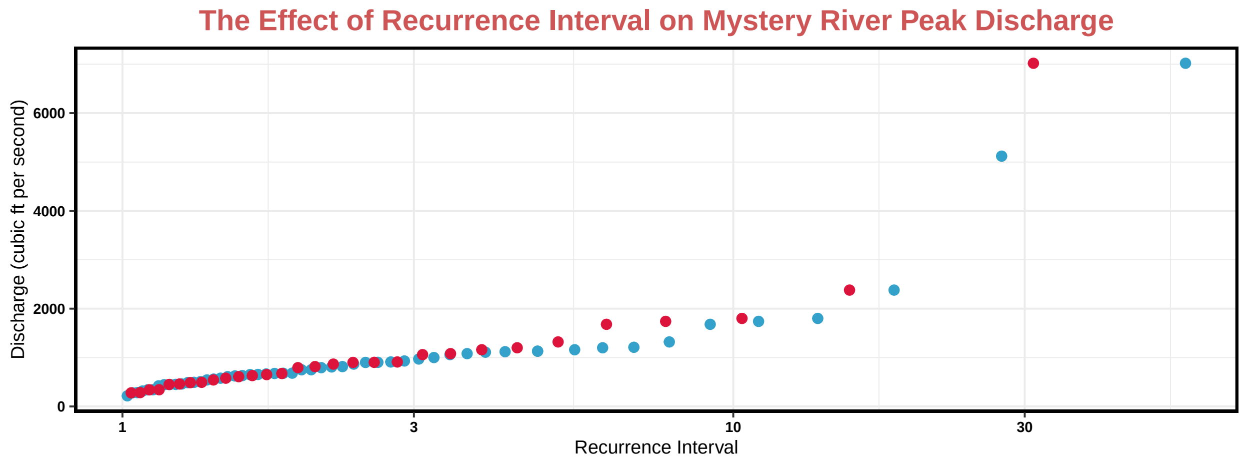
Create Data Frame for All Years of Record

```
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.

Discharge vs. Recurrence Plot

```
DischargevsRecurrencePlot<-
  ggplot(MysterySite.Annual.Full, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point(size=3, color="#33A1C9") + #blue
  geom_point(data = MysterySite.Annual.30yr, color = "#DC143C",#crimson
             aes(x = RecurrenceInterval, y = PeakDischarge), size=3) +
  gabytheme+
    scale_x_log10() +
  labs(title="The Effect of Recurrence Interval on Mystery River Peak Discharge",
       y ="Discharge (cubic ft per second)", x = "Recurrence Interval")
print(DischargevsRecurrencePlot)
```



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

30 Year Recurrence Interval Model

```
Mystery30YrMod<- lm(data = MysterySite.Annual.30yr, PeakDischarge ~ log(RecurrenceInterval))
summary(Mystery30YrMod)
```

```
##
## 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 Site 30 Year Model Coefficients

```
Mystery30YrMod$coefficients
```

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

What is the discharge for a 100-year flood in this system?

```
Mystery30YrMod$coefficients[1] + Mystery30YrMod$coefficients[2]*log(100)
```

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

Full Period of Record Mystery Site Model

```
MysteryFullMod<- lm(data = MysterySite.Annual.Full, PeakDischarge ~ log(RecurrenceInterval))
summary(MysteryFullMod)
```

```
##
## 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
```

Mystery Site Full Model Coefficients

```
MysteryFullMod$coefficients
```

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

What is the discharge for a 100-year flood in this system?

```
MysteryFullMod$coefficients[1] + MysteryFullMod$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 above shows there is stationarity of discharge in the Mystery River. While there is variability in precipitation and river discharge, the variability is oscillating around a relatively constant average value and has a relatively constant spread. Depending on the size of the period of record, the same precipitation event in the 30 year period (crimson curve) and the full period of record (blue curve) could have different recurrence intervals. For example, Question 5's Graph tells me that there was a significant precipitation event in late 1990. I will use this precipitation event to explain this question.

    Depending on the period of record we are considering, this 1990 precipitation event within the 30 year period (1964-1995) or the full 54 year period (1964-2019) could have different recurrence intervals because the 1990 event within the 30 year period would have a lower recurrence interval than the 1990 event within the full period of record. This is because there is a smaller probability of this event happening in 30 years than 54 years. This is visible on the above graph (the two points above 7000 cfs). In addiiton, if I look at the data frames, this 1990 event has a recurrence interval of 31 and a probability of 0.03 in the 30 year period; in the complete 54-year period, the 1990 event has a recurrence interval of 55 and a probability of 0.02.The effect of the 1990 storm event is less statistically significant in the full period of record with a higher recurrence interval and more statistically significant 30 year period with a lower recurrence interval, which is why the river discharge for the 30 year period of record is 5538.257 cfs while the discharge for the full 54 period of record is 4843.717 cfs. Recurrence intervals don't shift much with a stationary river. If this river was nonstationary, I would expect a greater range in the probabilities and recurrence intervals between the 1990 event within the 30 year period and the 1990 event within the 54 year period.

Reflection

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

    The location of the river, climate, and weather patterns strongly affect river discharge. For instance, if we are considering a river in the Southeast climate with consistent precipitation in the form of rain but very few snow events, I would expect a high frequency of peaks over time. If we are considering a river in a cooler mountainous area such as Montana with less precipitation, I would expect this river's hydrograph to have less frequent but higher magnitude peaks centered around the springtime, due to snow melt events entering the river, with a noticeable lag time(time between the precipitation event and the peak discharge).

I also learned about the concept of hydrologic stationarity and nonstationarity. Stationarity means that the statistical properties of hydrologic variables (such as river discharge) in future time periods will be similar to past time periods. For a given streamflow-gaging station, stationarity of annual peak discharge requires that all of the data represent a consistent hydrologic regime within the same variable climate system. If the original climatic system has definitively shifted such that a new climactic system exists (for example, due to anthropogenic climate change or infrastructure additions such as dams), appropriate statistical analysis of hydrologic characteristics might require restricting the analysis to the data's period of record that now currently represents the new system.

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

    The MysterySiteDischarge data frame and the resulting plot I constructed in Question 5 support my first conclusion; upon inspection of the effect of year on river discharge, there is obvious cyclical seasonality present, with the magnitude of river discharge at its highest before the start of the calendar year then decreasing in the early summer until the fall. The cycle then repeats. The plot I constructed in Question 10 supports my stationarity conclusion because the two curves for the 30 year period of record and the full period of record were extremely similar and displayed stationarity.

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

    Yes, it did. Analyzing the Green River data and plots in Lesson 6 allowed me to visualize the effect of implementing infrastructure such as flood control dams, and constructing the recurrence interval models for 100 year and 500 year flood events let me see how the magnitude of river discharge was affected by the dams. For this specific assignment, Question 7 allowed me to study the data range of the 25th percentile, median, and 75th percentile, and see how the range is affected by seasonality (range decreased during the late spring through summer months).

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

    Based on our theoretical class discussions, I expected the magnitude of discharge to increase with a larger period of record. However, with this MysterySiteDischarge data set, our curve and model for 30 year time period slightly overpredicts the data compared to the curve and model for the 54 year period of record when discharge is above 1000 cfs. Basically, I did not expect the river discharge to decrease with a larger period of record.