6: Physical Properties of Rive

Hydrologic Data Analysis | Kateri Salk

Fall 2019

Lesson Objectives

1. Compute recurrence intervals for stream discharge
2. Analyze the effects of watershed disturbance on recurrence intervals and interpret results against the concept of stationarity
3. Communicate findings with peers through oral, visual, and written modes

Opening Discussion

How is climate change impacting discharge in streams and rivers? What are the new and ongoing challenges faced by these impacts in watersheds? 1. Salk River, very seasonal before 2015 ,after 2015, more variable, impact flood control 2. mountainous region, impact magnitude, timing of the flood, impact farmer, change regular pattern, potential uses 3. unpredicable precipitation, more effort on civil engineering control on watersheds

Session Set Up

```
getwd()
```

```
## [1] "/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Lessons"
```

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


theme_set(theme_classic())
```

Recurrence Intervals and Exceededence Probability

A **recurrence interval** is the past recurrence of an event, in this case a peak annual discharge measurement of a given magnitude. The value of a recurrence interval corresponds to the average number of years between discharge of a given magnitude. Typically the minimum amount of years required to construct a recurrence interval is 10, but 30 is more robust. A recurrence interval, T, is calculated as:

> the more years we have, more information we get, like weird events, we can get more suitable model use.

$T = (n + 1)/m$

where n is the number of years and m is the ranking of an event within the observed period. We add one to n because we are computing the recurrence interval for a discharge event of a given magnitude *or greater*.

Similarly, we can calculate an **exceedence probability**, or the probability of encountering a discharge event of a given magnitude or greater in any given year:

$P = 1/T = m/(n + 1)$

This is where the terms "100-year flood" and similar are derived. Remember this is a probability based on past occurrence, not an accurate forecast of how often we will see that event happening. When current patterns of discharge differ from past patterns, we observe **nonstationary** behavior. Nonstationarity results in events that occur more or less frequency than predicted based on the exceedence probability.

Has Eno River dicharge displayed stationary behavior over the period of record?

Let's import discharge data for the Eno River near Durham for all available dates.

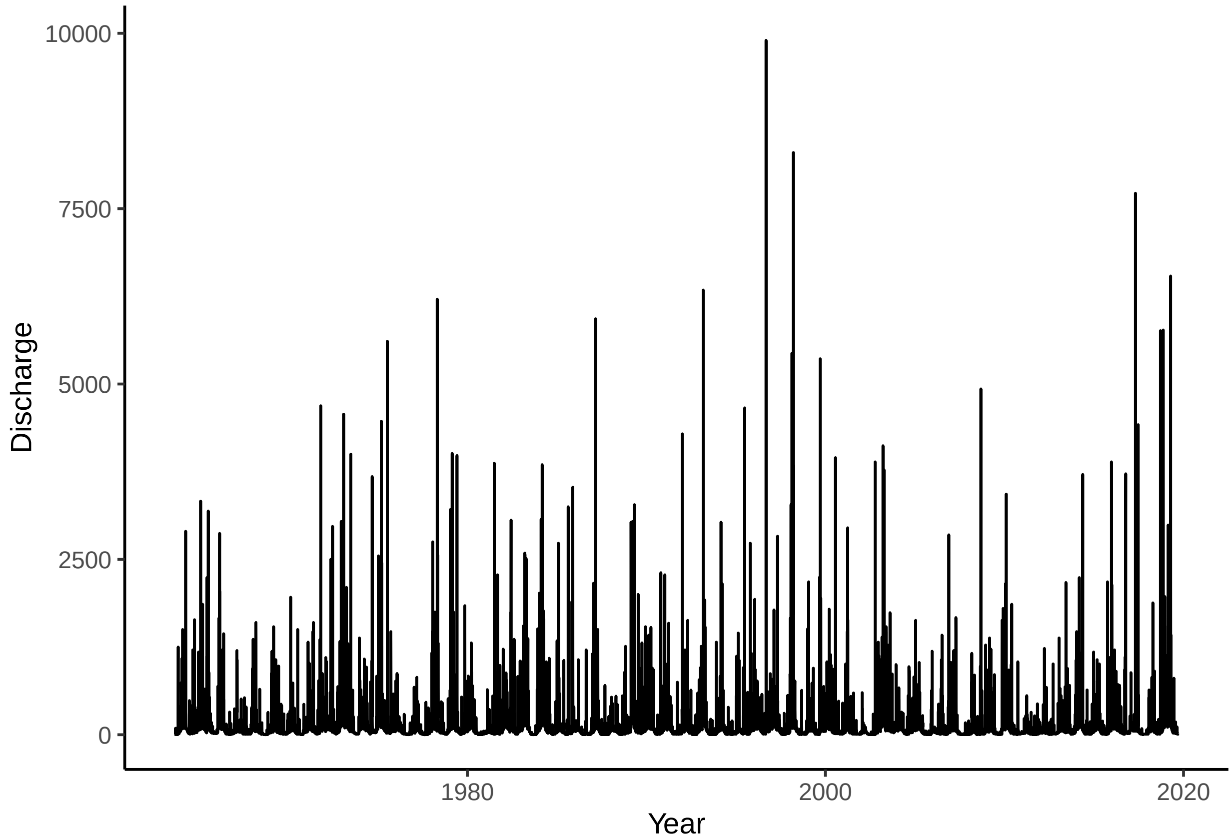
```
EnoDischarge <- readNWISdv(siteNumbers = "02085070",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")

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

attr(EnoDischarge, "variableInfo")
```

```
##   variableCode           variableName              variableDescription
## 1        00060 Streamflow, ft&#179;/s Discharge, cubic feet per second
##       valueType  unit options noDataValue
## 1 Derived Value ft3/s    Mean          NA
```

```
attr(EnoDischarge, "siteInfo")
```

```
##                   station_nm  site_no agency_cd timeZoneOffset
## 1 ENO RIVER NEAR DURHAM, NC 02085070      USGS         -05:00
##   timeZoneAbbreviation dec_lat_va dec_lon_va       srs siteTypeCd    hucCd
## 1                  EST   36.07222  -78.90778 EPSG:4326         ST 03020201
##   stateCd countyCd network
## 1      37    37063    NWIS
```

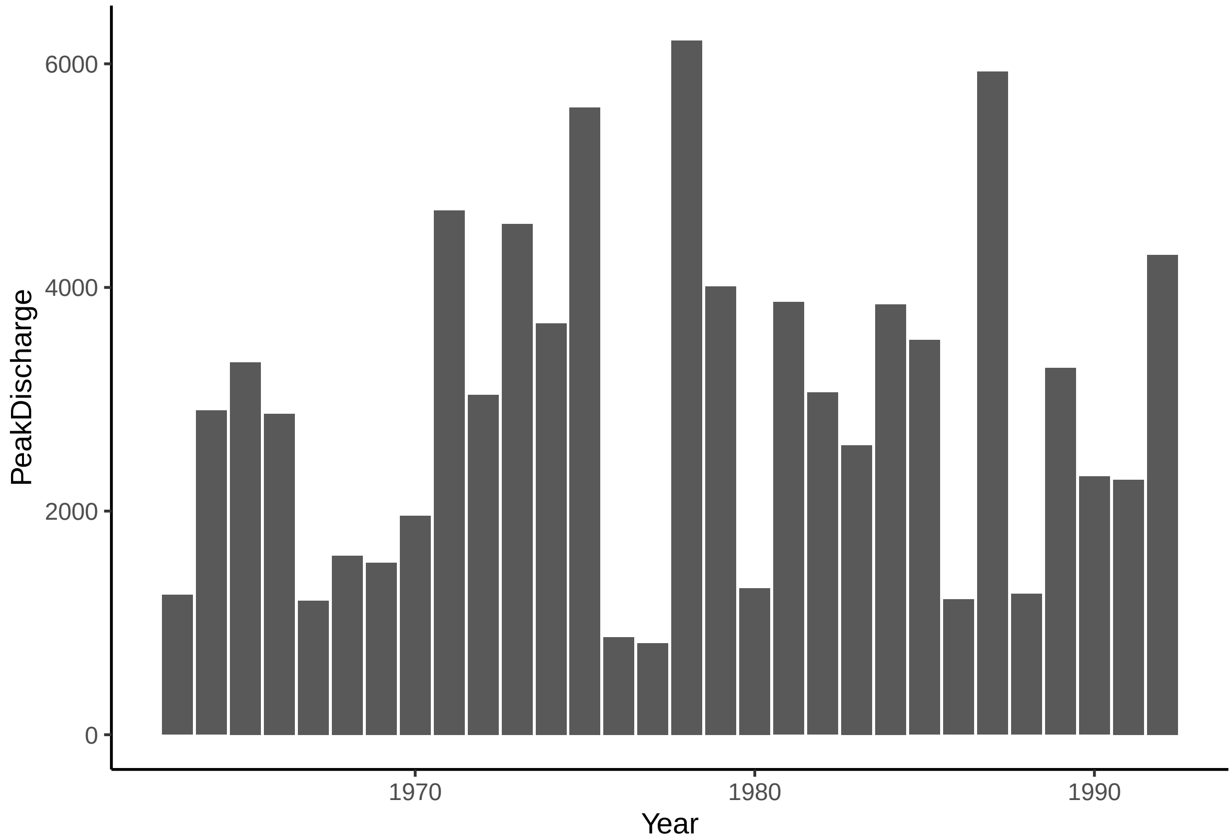
```
# Build a ggplot
EnoPlot <-
  ggplot(EnoDischarge, aes(x = Date, y = Discharge)) +
         geom_line() +
         xlab("Year")
print(EnoPlot)
```



the highest one near 2000, the smallest is near 1980, overall is very variable. no real consistent pattern.

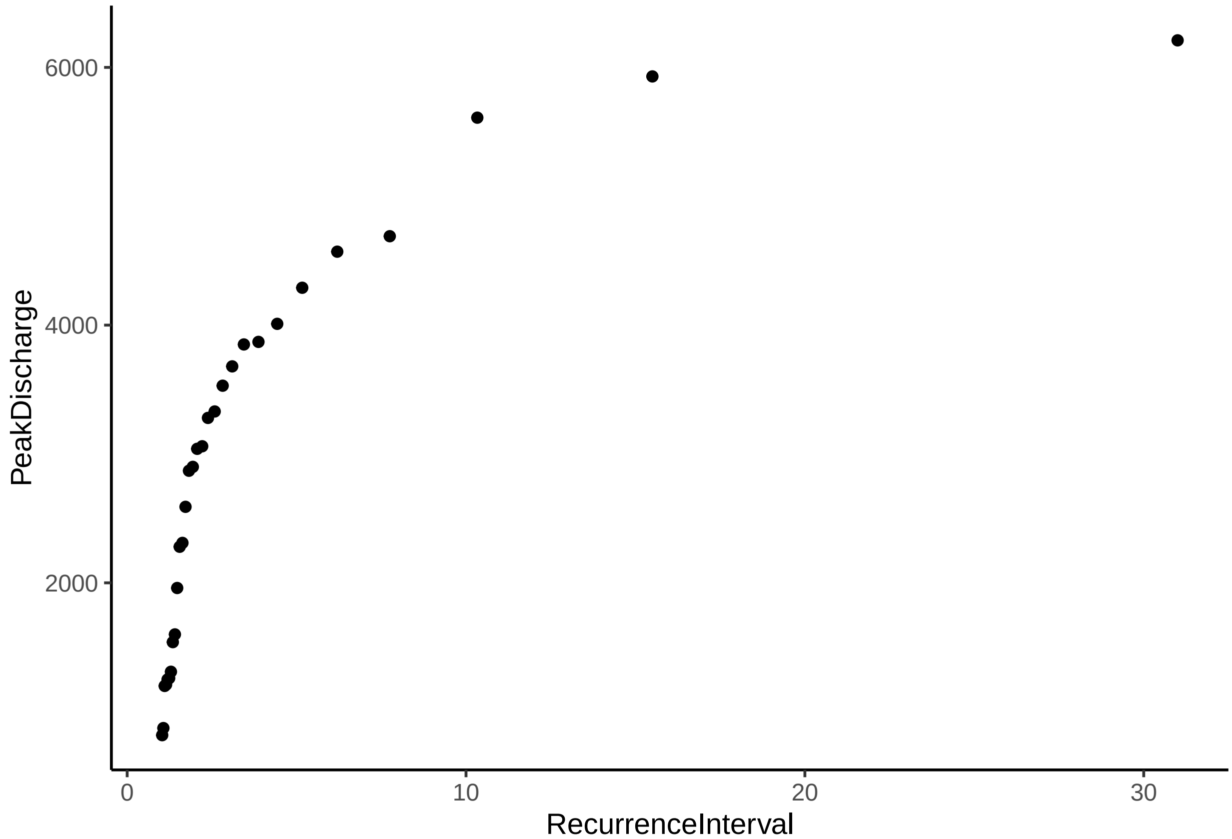
We can then compute recurrence intervals based on the first 30 years of data.

```
EnoDischarge <-
  EnoDischarge %>%
  mutate(Year = year(Date))

EnoRecurrence <-
  EnoDischarge %>%
  filter(Year < 1993) %>%
  group_by(Year) %>%
  summarise(PeakDischarge = max(Discharge)) %>% # 30 numbers
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(Year) + 1)/Rank,
         Probability = 1/RecurrenceInterval)

EnoPeakPlot <-
  ggplot(EnoRecurrence, aes(x = Year, y = PeakDischarge)) +
         geom_bar(stat = "identity") +
         xlab("Year")
print(EnoPeakPlot)
```



Let's display and model the relationship between peak annual disharge and recurrence interval. We can use the statistical model to compute discharge for recurrence intervals that occur above the 30-year mark.

```
EnoRecurrencePlot <-
  ggplot(EnoRecurrence, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() #+
  #scale_x_log10()
print(EnoRecurrencePlot)
```




```r
Eno.RImodel <- lm(data = EnoRecurrence, PeakDischarge ~ log(RecurrenceInterval))
summary(Eno.RImodel) # large R-squared and small p-value indicates that log is a good model
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = EnoRecurrence)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1048.9  -320.3   116.3   358.6   500.7
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1332.5      112.3   11.86 1.95e-12 ***
## log(RecurrenceInterval)   1725.8       89.0   19.39  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 407.6 on 28 degrees of freedom
## Multiple R-squared:  0.9307, Adjusted R-squared:  0.9282
## F-statistic: 376.1 on 1 and 28 DF,  p-value: < 2.2e-16
```

```r
#What is the discharge for a 100-year flood in this system? a 500-year flood?
# the 1st coeff is the estimate y intercept of the model. the 2nd coeff is the estimate log(RecurrenceI
Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log(100)
```

```
## (Intercept)
```

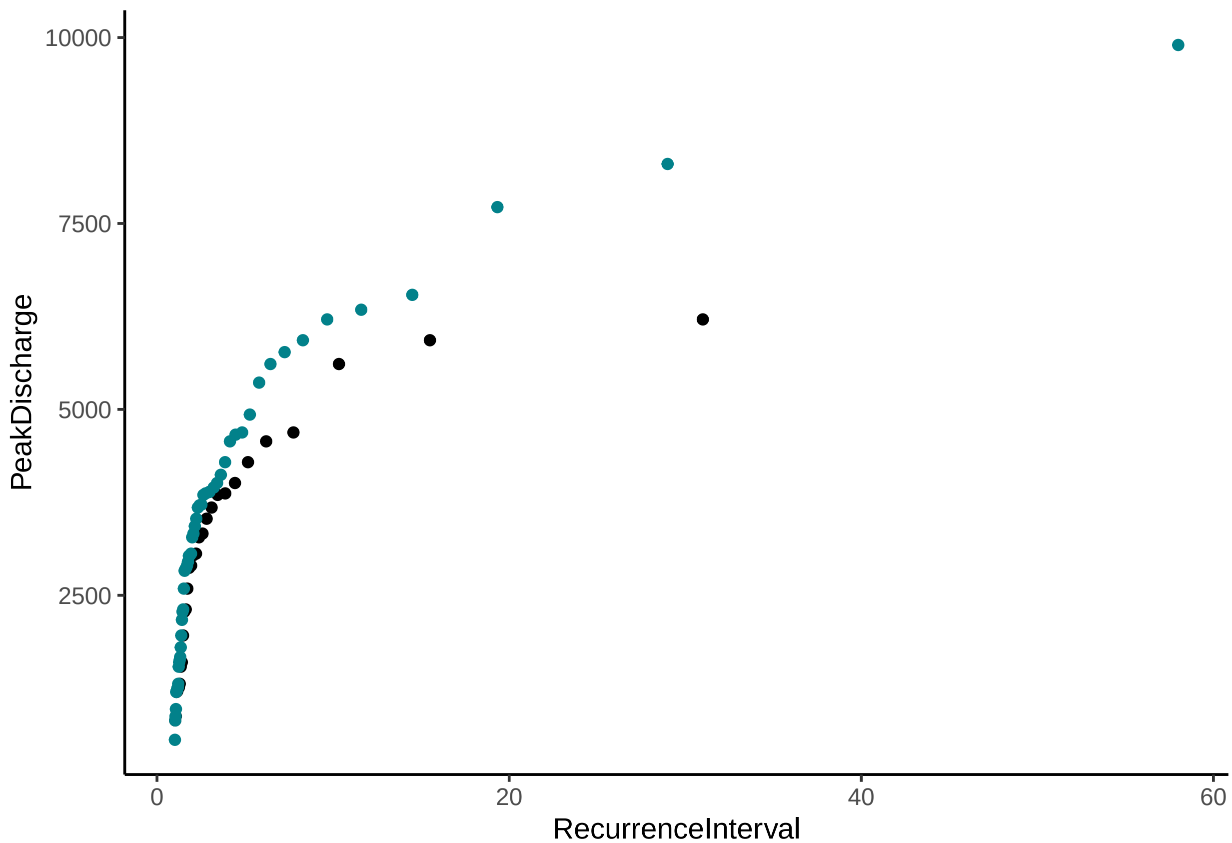
```
## 	9280.152
```

```r
Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log(500)
```

```
## (Intercept)
## 	12057.74
```

What if we were to build a recurrence interval model for the entire period of record? How would this compare to the 30-year construction?

```r
EnoRecurrence.Full <-
  EnoDischarge %>%
  group_by(Year) %>%
  summarise(PeakDischarge = max(Discharge)) %>%
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(Year) + 1)/Rank,
         Probability = 1/RecurrenceInterval)

EnoRecurrencePlot.Full <-
  ggplot(EnoRecurrence, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() +
  geom_point(data = EnoRecurrence.Full, color = "#02818a",
             aes(x = RecurrenceInterval, y = PeakDischarge))
print(EnoRecurrencePlot.Full)
```




```r
# deeper slope

Eno.RImodel.Full <- lm(data = EnoRecurrence.Full, PeakDischarge ~ log(RecurrenceInterval))
```

```r
summary(Eno.RImodel.Full)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = EnoRecurrence.Full)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -831.15 -281.84   51.82  325.57  500.85
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1350.28      69.17   19.52   <2e-16 ***
## log(RecurrenceInterval)  2177.56      52.68   41.34   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 353.8 on 55 degrees of freedom
## Multiple R-squared:  0.9688, Adjusted R-squared:  0.9683
## F-statistic:  1709 on 1 and 55 DF,  p-value: < 2.2e-16
```

```r
Eno.RImodel.Full$coefficients
```

```
##             (Intercept) log(RecurrenceInterval)
##                1350.282                2177.559
```

```r
Eno.RImodel$coefficients
```

```
##             (Intercept) log(RecurrenceInterval)
##                1332.484                1725.814
```

```r
Eno.RImodel.Full$coefficients[1] + Eno.RImodel.Full$coefficients[2]*log(100)
```

```
## (Intercept)
##    11378.31
```

```r
Eno.RImodel.Full$coefficients[1] + Eno.RImodel.Full$coefficients[2]*log(500)
```

```
## (Intercept)
##    14882.96
```

```r
Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log(100)
```

```
## (Intercept)
##    9280.152
```

```r
Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log(500)
```

```
## (Intercept)
##    12057.74
```

```r
# we usually underpredict the discharge based on the current data
```

What differences did you see for the recurrence intervals built under different periods of record? How would your prediction of flood events differ if you were to use these models for forecasting purposes?

> under more years, we can get more information and get more suitable model. if we use smaller period of data, the prediction discharge would be smaller, as we usually underpredict the discharge based on the smaller periods.

What would you recommend for a watershed manager seeking to build the most accurate recurrence interval model for the Eno River?

1. use the longer records for your stream, looking for other similar river if they have same trends.
2. use the newest data to analysis, use the more recent data.

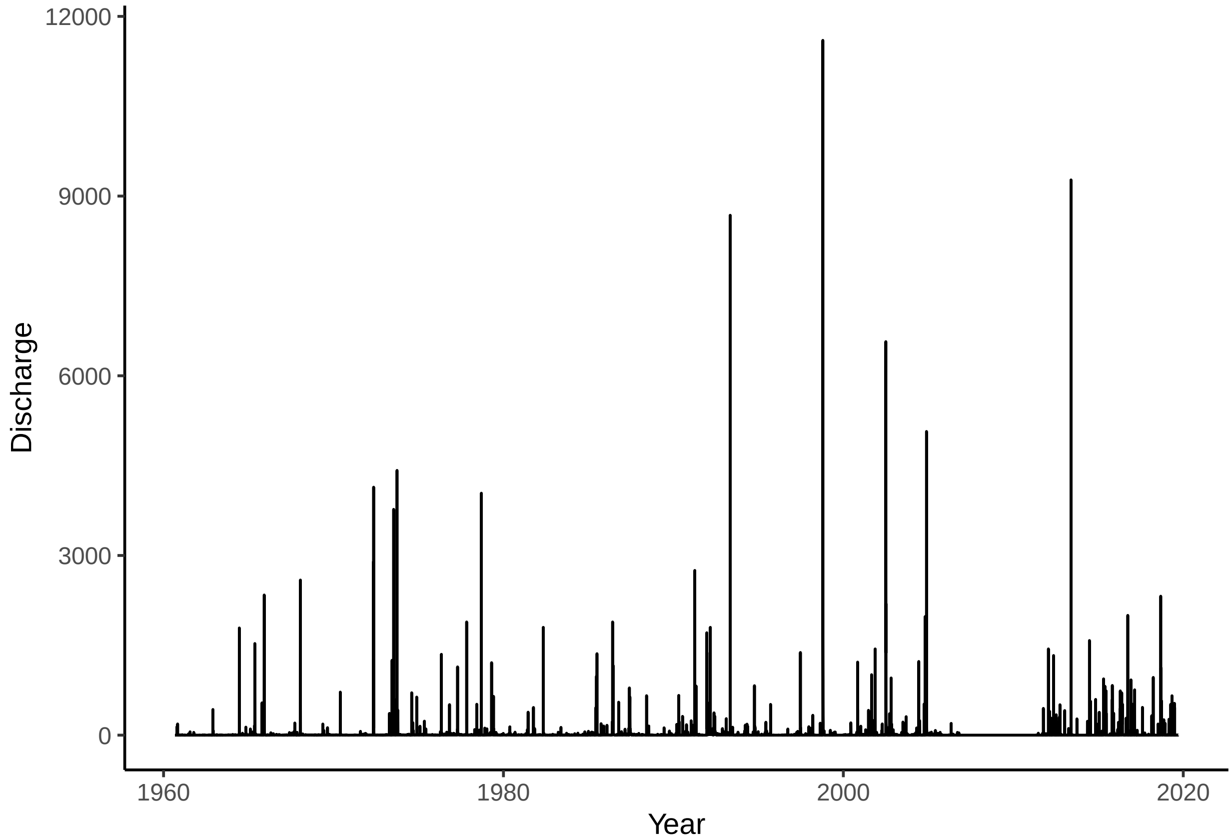
Examining the effects of urbanization on discharge

Salado Creek is located in San Antonio, Texas, an area that has been rapidly urbanizing over the course of the last several decades (http://worldpopulationreview.com/us-cities/san-antonio-population/#byPopulation). Using the code above, compute recurrence intervals for Salado Creek for the first 30 years of record and for the most recent 30 years of record. How do the graphs and models differ? How do your predictions of 100- and 500-year floods differ?

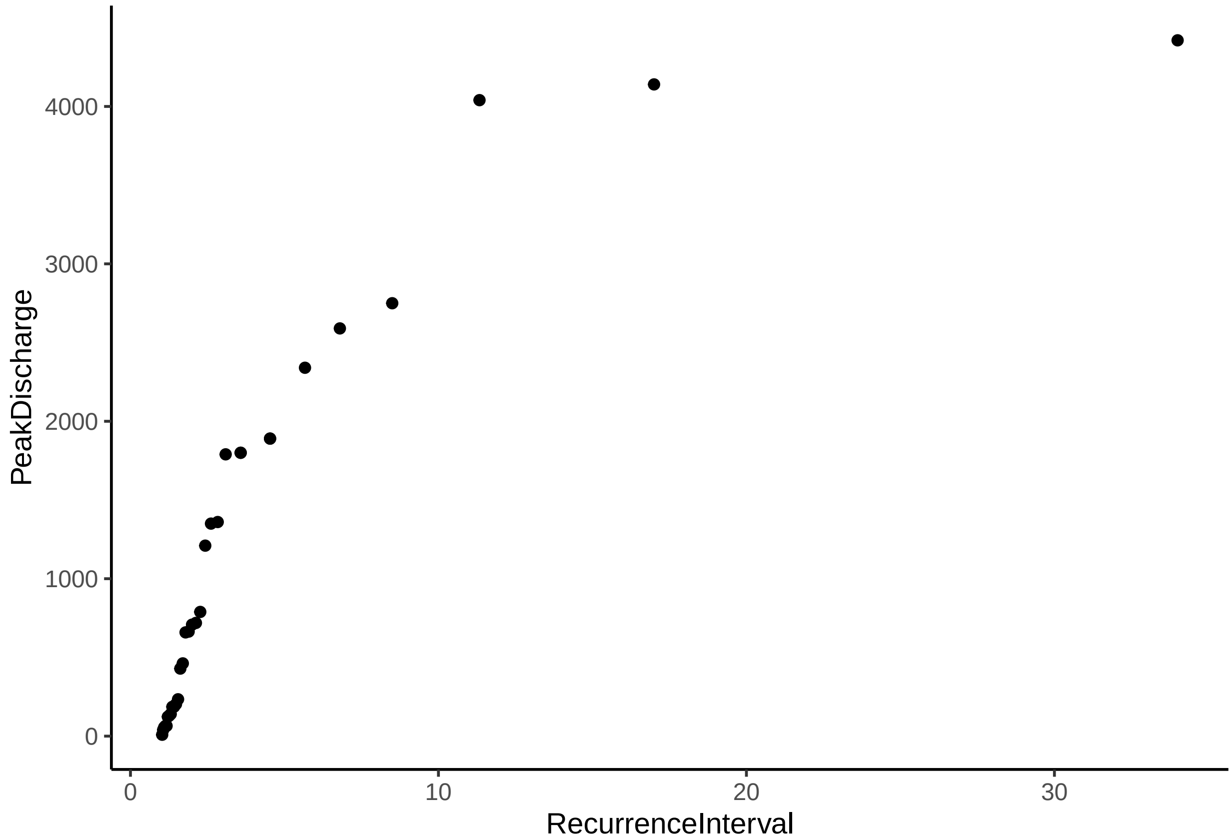
```
# Import data
SaladoDischarge <- readNWISdv(siteNumbers = "08178700",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")
names(SaladoDischarge)[4:5] <- c("Discharge", "Approval.Code")
attr(SaladoDischarge, "siteInfo")
```

```
##                               station_nm  site_no agency_cd timeZoneOffset
## 1 Salado Ck at Loop 410, San Antonio, TX 08178700      USGS         -06:00
##   timeZoneAbbreviation dec_lat_va dec_lon_va       srs siteTypeCd    hucCd
## 1                  CST   29.51606  -98.43113 EPSG:4326         ST 12100301
##   stateCd countyCd network
## 1      48    48029    NWIS
```

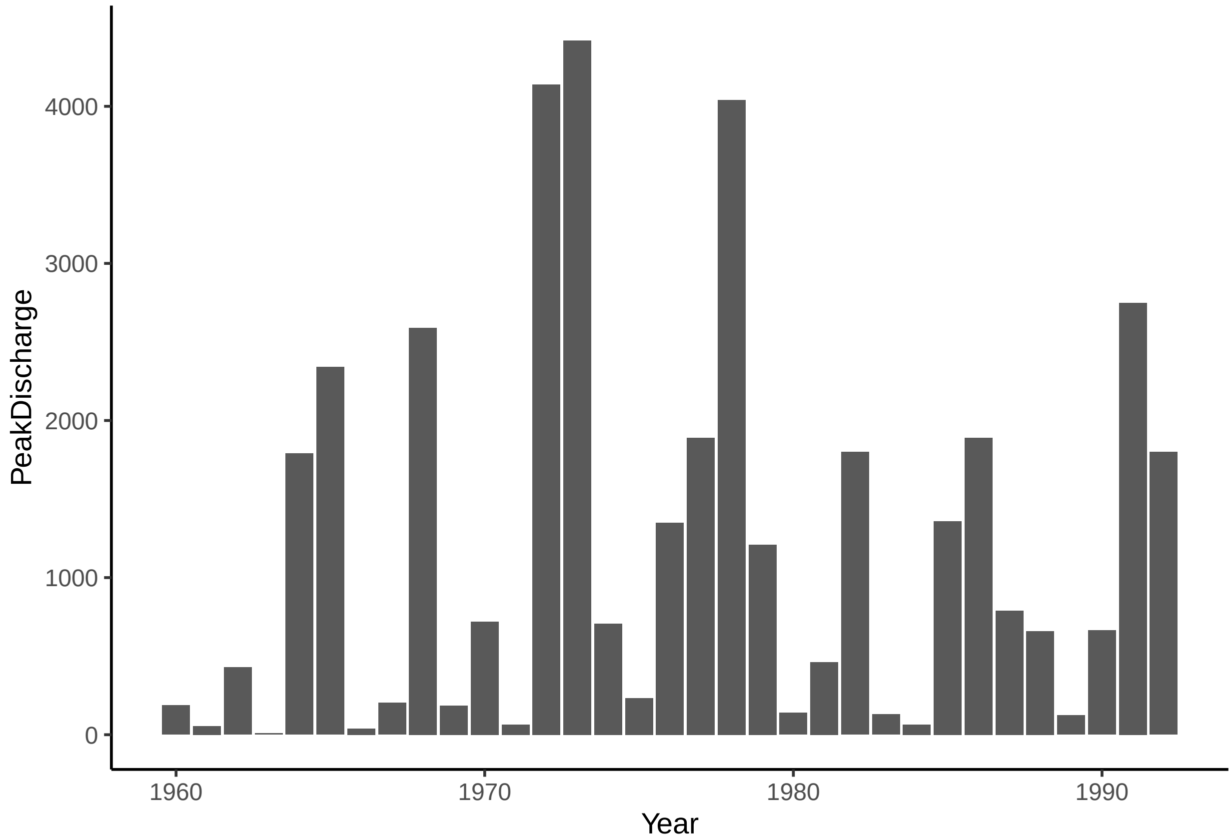
```
SaladoPlot <-
  ggplot(SaladoDischarge, aes(x = Date, y = Discharge)) +
         geom_line() +
         xlab("Year")
print(SaladoPlot)
```




```r
# add more code here:
SaladoDischarge <-
  SaladoDischarge %>%
  mutate(Year = year(Date))

# for the first 30 years
SaladoRecurrence_first30 <-
  SaladoDischarge %>%
  filter(Year < 1993) %>%
  group_by(Year) %>%
  summarise(PeakDischarge = max(Discharge)) %>% # 30 numbers
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(Year) + 1)/Rank,
         Probability = 1/RecurrenceInterval)

SaladoRecurrencePlot <-
  ggplot(SaladoRecurrence_first30, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() #+
  #scale_x_log10()
print(SaladoRecurrencePlot)
```




```
SaladoPeakPlot_first30 <-
  ggplot(SaladoRecurrence_first30, aes(x = Year, y = PeakDischarge)) +
         geom_bar(stat = "identity") +
         xlab("Year")
print(SaladoPeakPlot_first30)
```




```r
Salado_first30.RImodel <- lm(data = SaladoRecurrence_first30, PeakDischarge ~ log(RecurrenceInterval))
summary(Salado_first30.RImodel) # large R-squared and small p-value indicates that log is a good model
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = SaladoRecurrence_first30)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -519.56 -117.79   -7.76  113.05  698.97
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -191.45      54.13  -3.537   0.0013 **
## log(RecurrenceInterval)  1455.05      42.61  34.149   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 206.8 on 31 degrees of freedom
## Multiple R-squared:  0.9741, Adjusted R-squared:  0.9733
## F-statistic:  1166 on 1 and 31 DF,  p-value: < 2.2e-16
```

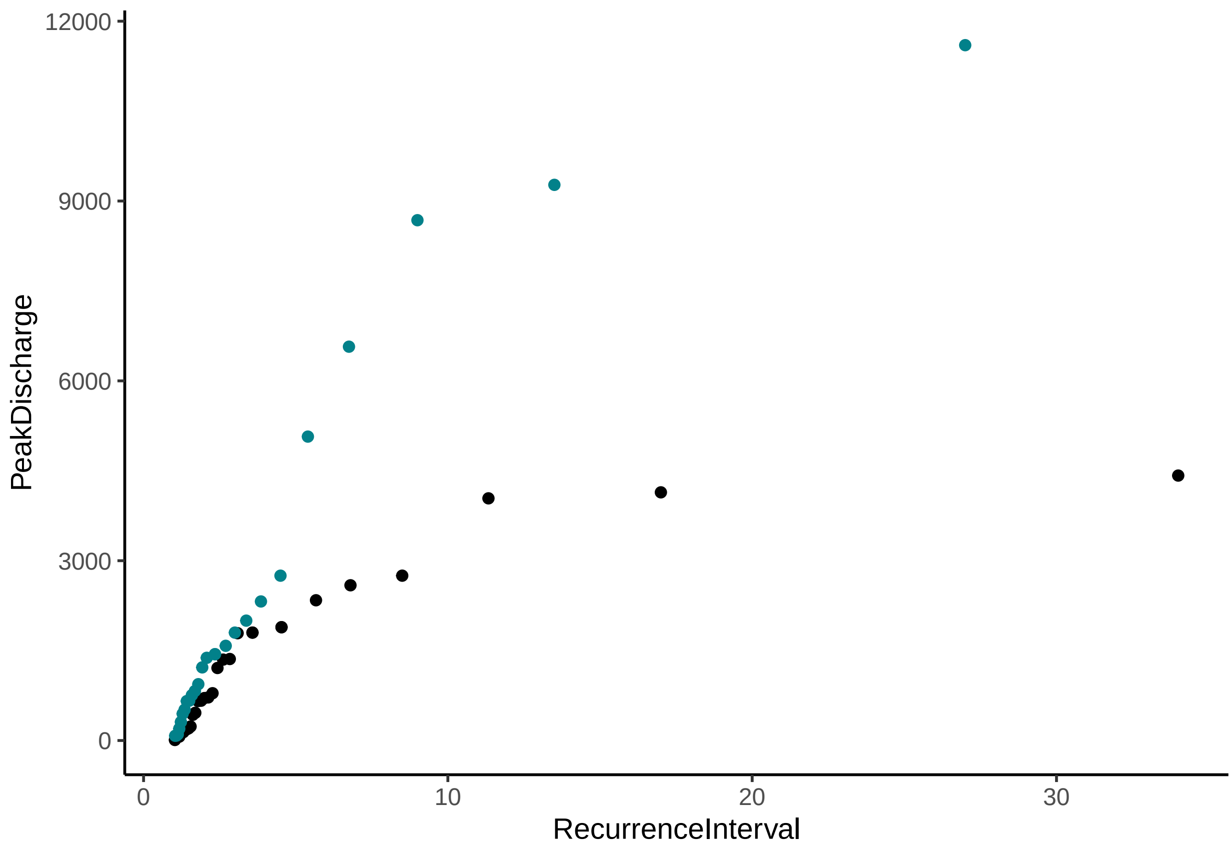
```r
Salado_first30.RImodel$coefficients[1] + Salado_first30.RImodel$coefficients[2]*log(100)
```

```
## (Intercept)
##    6509.281
```

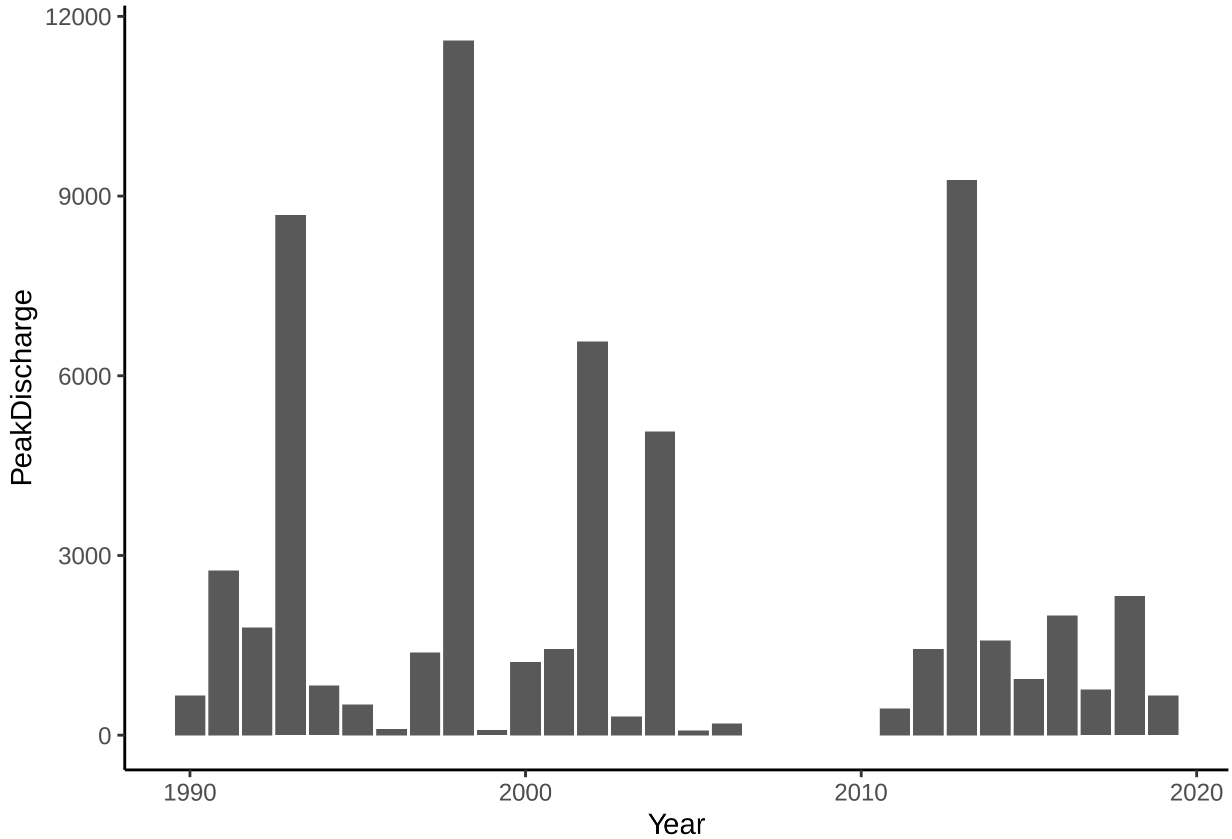
```r
Salado_first30.RImodel$coefficients[1] + Salado_first30.RImodel$coefficients[2]*log(500)
```

```
## (Intercept)
##    8851.088
```

```r
# for the most latest 30 year
SaladoRecurrence_latest30 <-
  SaladoDischarge %>%
  filter(Year > 1989) %>%
  group_by(Year) %>%
  summarise(PeakDischarge = max(Discharge)) %>% # 30 numbers
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(Year) + 1)/Rank,
         Probability = 1/RecurrenceInterval)

SaladoRecurrencePlot <-
  ggplot(SaladoRecurrence_first30, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point()+
  geom_point(data = SaladoRecurrence_latest30, color = "#02818a",
             aes(x = RecurrenceInterval, y = PeakDischarge))
print(SaladoRecurrencePlot)
```




```r
SaladoPeakPlot_latest30 <-
  ggplot(SaladoRecurrence_latest30, aes(x = Year, y = PeakDischarge)) +
         geom_bar(stat = "identity") +
         xlab("Year")
print(SaladoPeakPlot_latest30)
```




```r
Salado_latest30.RImodel <- lm(data = SaladoRecurrence_latest30, PeakDischarge ~ log(RecurrenceInterval))
summary(Salado_latest30.RImodel) # large R-squared and small p-value indicates that log is a good model
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log(RecurrenceInterval), data = SaladoRecurrence_latest30)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1699.7  -565.0   124.3   642.0  1727.6
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -981.0      262.5  -3.737  0.00102 **
## log(RecurrenceInterval)   3610.6      210.3  17.168 5.55e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 881 on 24 degrees of freedom
## Multiple R-squared:  0.9247, Adjusted R-squared:  0.9216
## F-statistic: 294.7 on 1 and 24 DF,  p-value: 5.552e-15
```

```r
Salado_first30.RImodel$coefficients[1] + Salado_first30.RImodel$coefficients[2]*log(100)
```

```
## (Intercept)
##    6509.281
```

```r
Salado_first30.RImodel$coefficients[1] + Salado_first30.RImodel$coefficients[2]*log(500)
```

```
## (Intercept)
##    8851.088
```

```r
Salado_latest30.RImodel$coefficients[1] + Salado_latest30.RImodel$coefficients[2]*log(100)
```

```
## (Intercept)
##    15646.61
```

```r
Salado_latest30.RImodel$coefficients[1] + Salado_latest30.RImodel$coefficients[2]*log(500)
```

```
## (Intercept)
##     21457.7
```

Examining the effects of dam construction on recurrence intervals

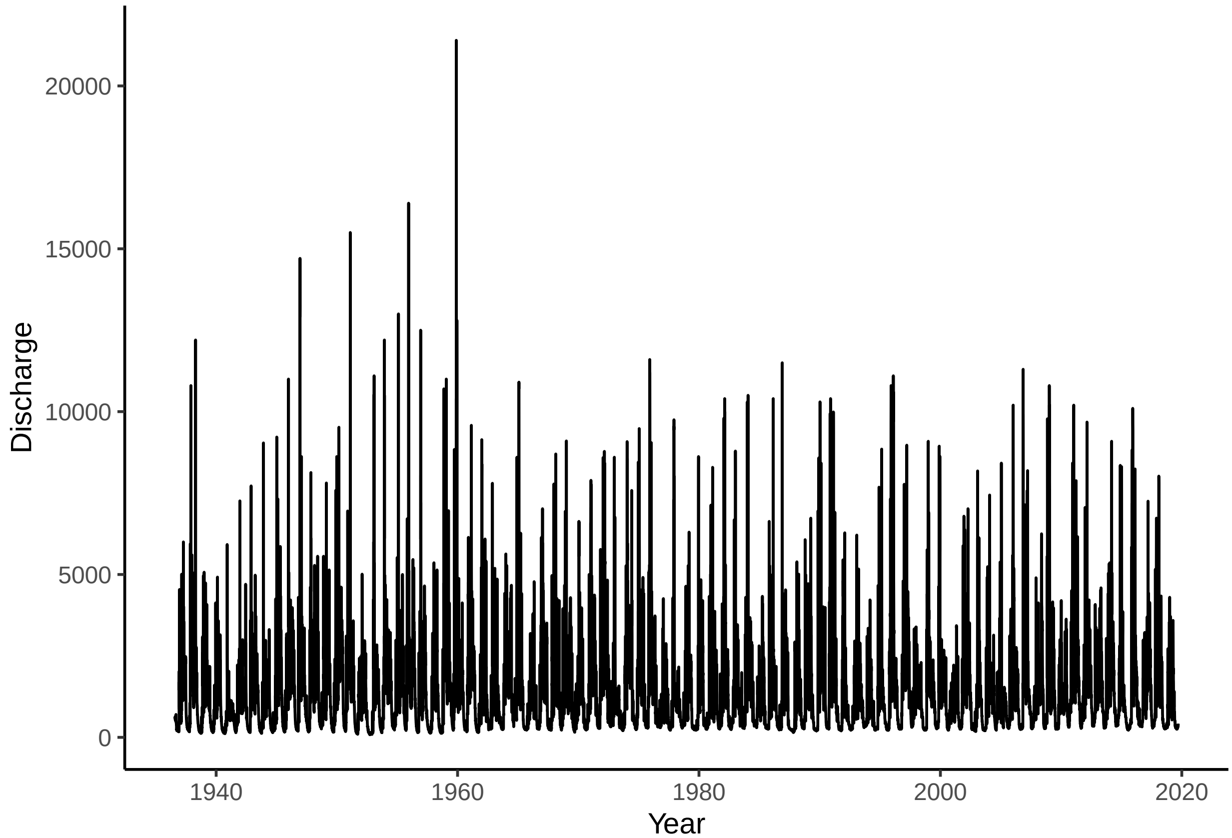
The stream gage in the Green River near Auburn, Washington, is located directly downstream of the Howard A. Hanson Dam. The dam was built in 1961 for flood control purposes, and the reservoir now provides water supply to the city of Tacoma. How have peak discharges changed since the construction of the dam?

Using the code above, construct recurrence intervals for the periods before and after the construction of the dam. How do the graphs and models differ? How do your predictions of 100- and 500-year floods differ?

```r
GreenDischarge <- readNWISdv(siteNumbers = "12113000",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")
names(GreenDischarge)[4:5] <- c("Discharge", "Approval.Code")
attr(GreenDischarge, "siteInfo")
```

```
##                      station_nm  site_no agency_cd timeZoneOffset
## 1 GREEN RIVER NEAR AUBURN, WA 12113000      USGS         -08:00
##   timeZoneAbbreviation dec_lat_va dec_lon_va       srs siteTypeCd    hucCd
## 1                  PST   47.31232   -122.204 EPSG:4326         ST 17110013
##   stateCd countyCd network
## 1      53    53033    NWIS
```

```r
GreenPlot <-
  ggplot(GreenDischarge, aes(x = Date, y = Discharge)) +
         geom_line() +
         xlab("Year")
print(GreenPlot)
```




```r
# add more code here:
```

Closing Discussion

This week we focused on discharge as a physical property of a stream or river. How might you use your knowledge of discharge to inform other physical processes occurring in rivers?