4: Physical Properties of Rivers

Water Data Analytics | Kateri Salk

Spring 2022

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?

- Higher variability in flow (flashy)
- Shifts in peak discharge, timing of snowmelt
- Types of discharge change
- Changes in evapotranspiration due to temperature increases

Session Set Up

```
getwd()
```

```
## [1] "/Users/ataliefischer/Desktop/WDA/Water_Data_Analytics_2022/Lessons"
```

```
# install.packages("lfstat") #low flow statistics

library(tidyverse)
library(dataRetrieval)
library(lubridate)
library(lfstat)


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:

$T = (n + 1)/m$

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:

$P = 1/T$

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.

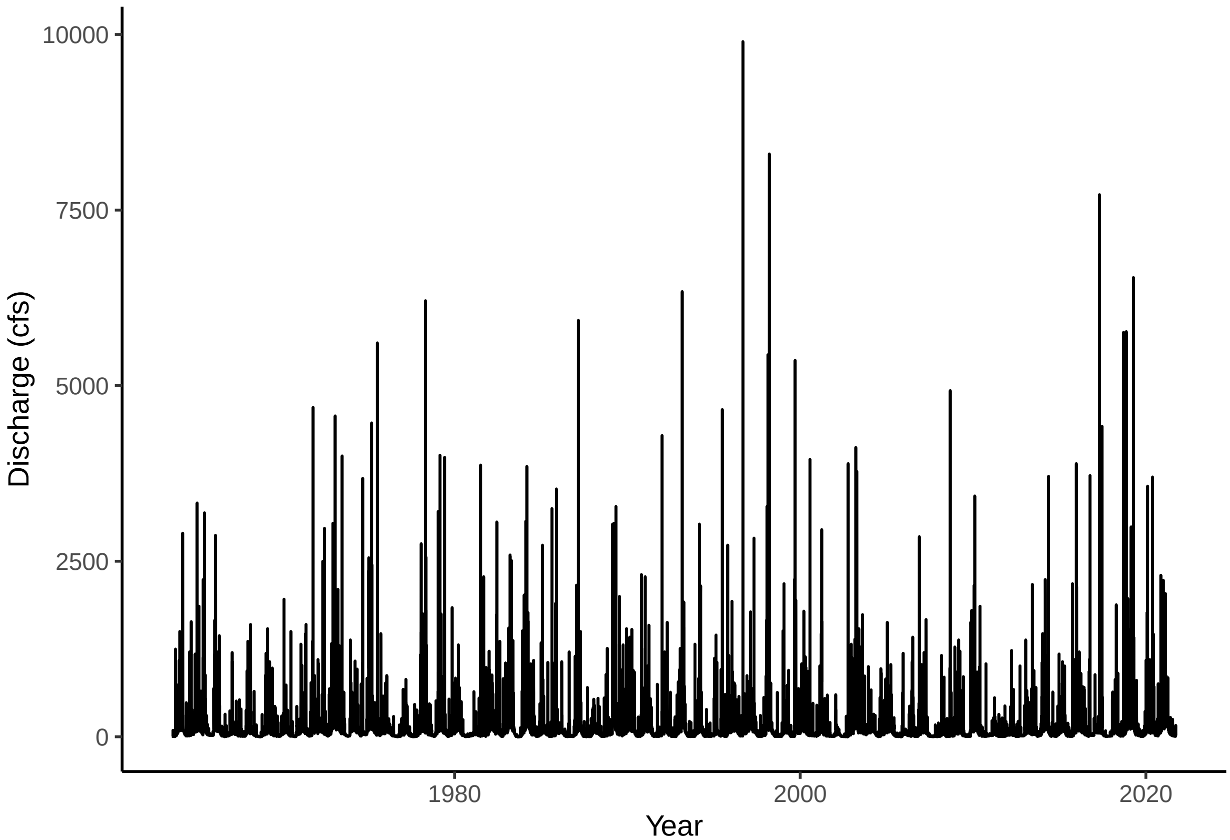
```
EnoDischarge <- readNWISdv(siteNumbers = "02085070",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "2021-09-30") # "the water year" starts on the first day of October and

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

attr(EnoDischarge, "variableInfo") # mean daily discharge
```

```
##   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") # geographic context of the gage
```

```
##                  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
```

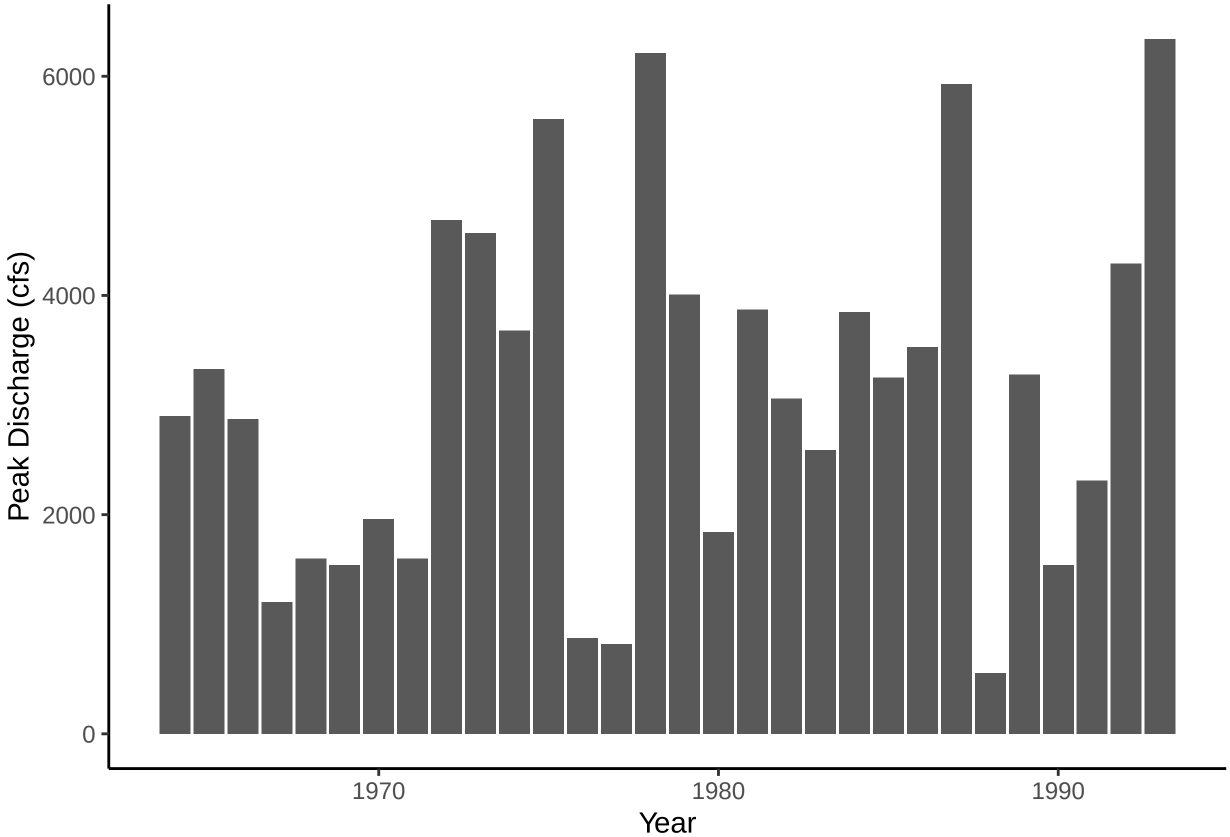
```
# Build a ggplot
ggplot(EnoDischarge, aes(x = Date, y = Discharge)) +
  geom_line() +
  labs(x = "Year", y = "Discharge (cfs)")
```




```
#event-based discharge peaks. some seasonablity but not as much as in snowpack affected rivers.
```

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

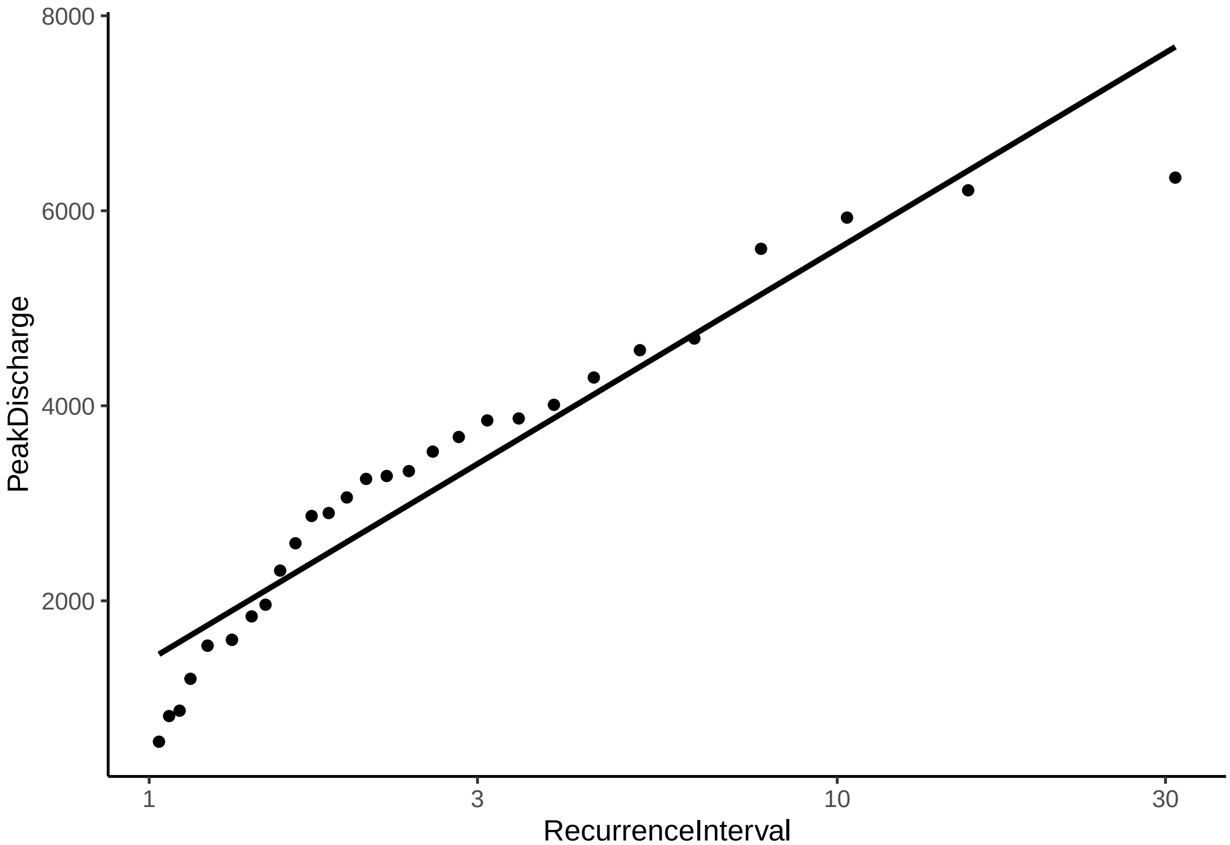
```
EnoDischarge <- EnoDischarge %>%
  mutate(Year = year(Date),
         WaterYear = water_year(Date, origin = "usgs")) %>%
  filter(WaterYear != "1963")
# different regions of the world have different water years. this is why you need to specify the origin

# Water Year is a factor. We want to re-classify as numeric.
EnoDischarge$WaterYear <- as.numeric(as.character(EnoDischarge$WaterYear))
# need to specify that a factor is a character first before you change it to numeric. if you try to cha

# Calculate Recurrence Intervals for 30 years. group_by water year to analyze the whole water year at t
EnoRecurrence <-
  EnoDischarge %>%
  filter(WaterYear < 1994) %>%
  group_by(WaterYear) %>%
  summarise(PeakDischarge = max(Discharge)) %>%
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(WaterYear) + 1)/Rank,
         Probability = 1/RecurrenceInterval)

ggplot(EnoRecurrence, aes(x = WaterYear, y = PeakDischarge)) +
  geom_bar(stat = "identity") +
  labs(x = "Year", y = "Peak Discharge (cfs)")
```



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.

```
ggplot(EnoRecurrence, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() +
  scale_x_log10() +
  geom_smooth(method = "lm", color = "black", se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
#lm=linear model, se = standard error

Eno.RImodel <- lm(data = EnoRecurrence, PeakDischarge ~ log10(RecurrenceInterval))
summary(Eno.RImodel)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log10(RecurrenceInterval), data = EnoRecurrence)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1341.1  -209.5   153.9   389.5   528.6
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1391.8      130.9   10.63 2.43e-11 ***
## log10(RecurrenceInterval)   4217.1      238.8   17.66  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 474.9 on 28 degrees of freedom
## Multiple R-squared:  0.9176, Adjusted R-squared:  0.9147
## F-statistic: 311.9 on 1 and 28 DF,  p-value: < 2.2e-16
```

```
#What is the discharge for a 100-year flood in this system? a 500-year flood?
Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log10(100)
```

```
## (Intercept)
```

```
##    9826.082
```

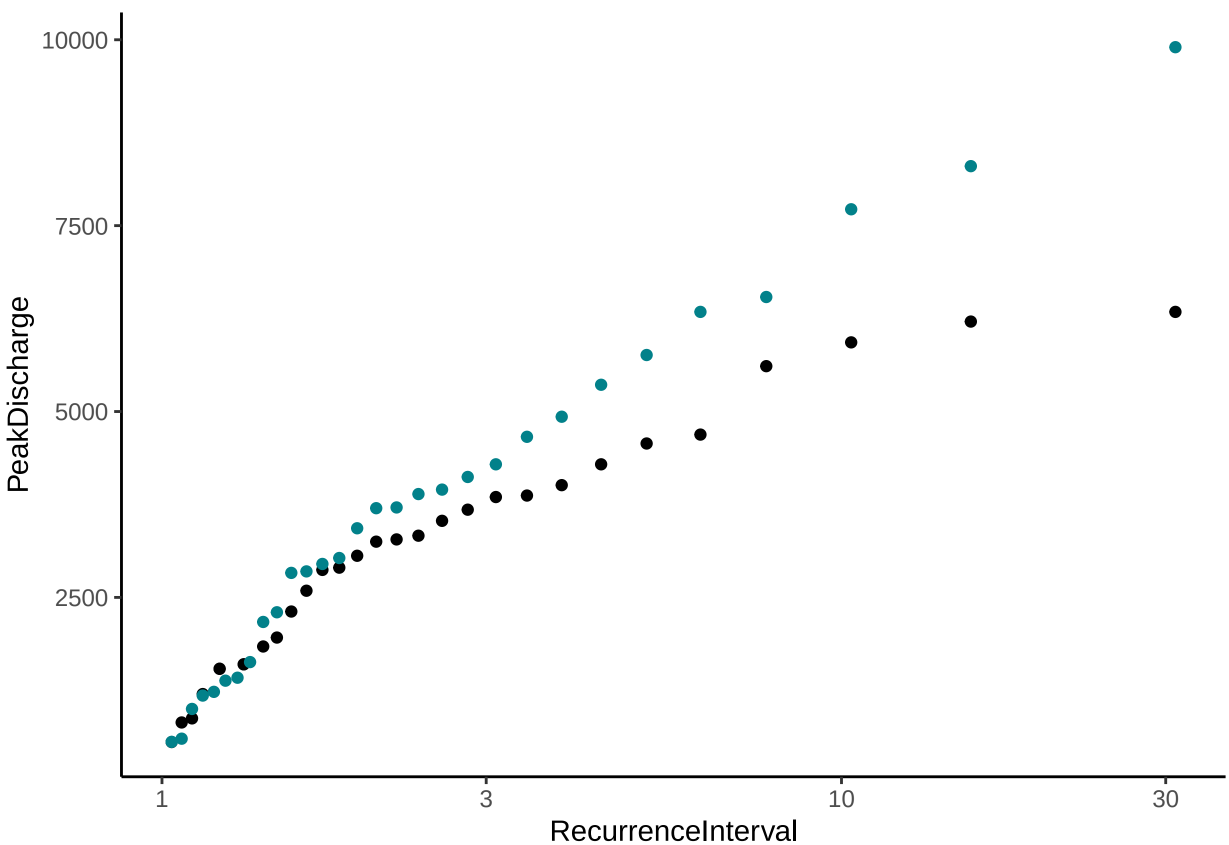
```
Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log10(500)
```

```
## (Intercept)
##    12773.73
```

```
# using summary statistics...intercept (coeff. 1) + slope (coeff. 2) * recurrence interval
```

What if we were to build a recurrence interval model for the most recent 30 years? How would this compare to the early period recurrence interval?

```
EnoRecurrence.Late <-
  EnoDischarge %>%
  filter(WaterYear >= 1992) %>%
  group_by(WaterYear) %>%
  summarise(PeakDischarge = max(Discharge)) %>%
  mutate(Rank = rank(-PeakDischarge),
         RecurrenceInterval = (length(WaterYear) + 1)/Rank,
         Probability = 1/RecurrenceInterval)

ggplot(EnoRecurrence, aes(x = RecurrenceInterval, y = PeakDischarge)) +
  geom_point() +
  geom_point(data = EnoRecurrence.Late, color = "#02818a",
             aes(x = RecurrenceInterval, y = PeakDischarge)) +
  scale_x_log10()
```




```
# less frequent events are higher. lowest discharges overlap. discharges at higher recurrence intervals
```

```r
Eno.RImodel.Late <- lm(data = EnoRecurrence.Late, PeakDischarge ~ log10(RecurrenceInterval))
summary(Eno.RImodel.Late)
```

```
##
## Call:
## lm(formula = PeakDischarge ~ log10(RecurrenceInterval), data = EnoRecurrence.Late)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -714.2 -338.5  112.5  242.1  577.6
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1131.2      104.3   10.85 1.54e-11 ***
## log10(RecurrenceInterval)   6315.5      190.2   33.20  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 378.4 on 28 degrees of freedom
## Multiple R-squared:  0.9752, Adjusted R-squared:  0.9743
## F-statistic:  1102 on 1 and 28 DF,  p-value: < 2.2e-16
```

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

```
##               (Intercept) log10(RecurrenceInterval)
##                  1131.245                  6315.535
```

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

```
##               (Intercept) log10(RecurrenceInterval)
##                  1391.809                  4217.137
```

```r
#slope in later years has greatly increased!

Eno.RImodel.Late$coefficients[1] + Eno.RImodel.Late$coefficients[2]*log10(100)
```

```
## (Intercept)
##    13762.31
```

```r
Eno.RImodel.Late$coefficients[1] + Eno.RImodel.Late$coefficients[2]*log10(500)
```

```
## (Intercept)
##    18176.68
```

```r
#100 and 500 year floods are much greater than before

Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log10(100)
```

```
## (Intercept)
##    9826.082
```

```r
Eno.RImodel$coefficients[1] + Eno.RImodel$coefficients[2]*log10(500)
```

```
## (Intercept)
##    12773.73
```

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?

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

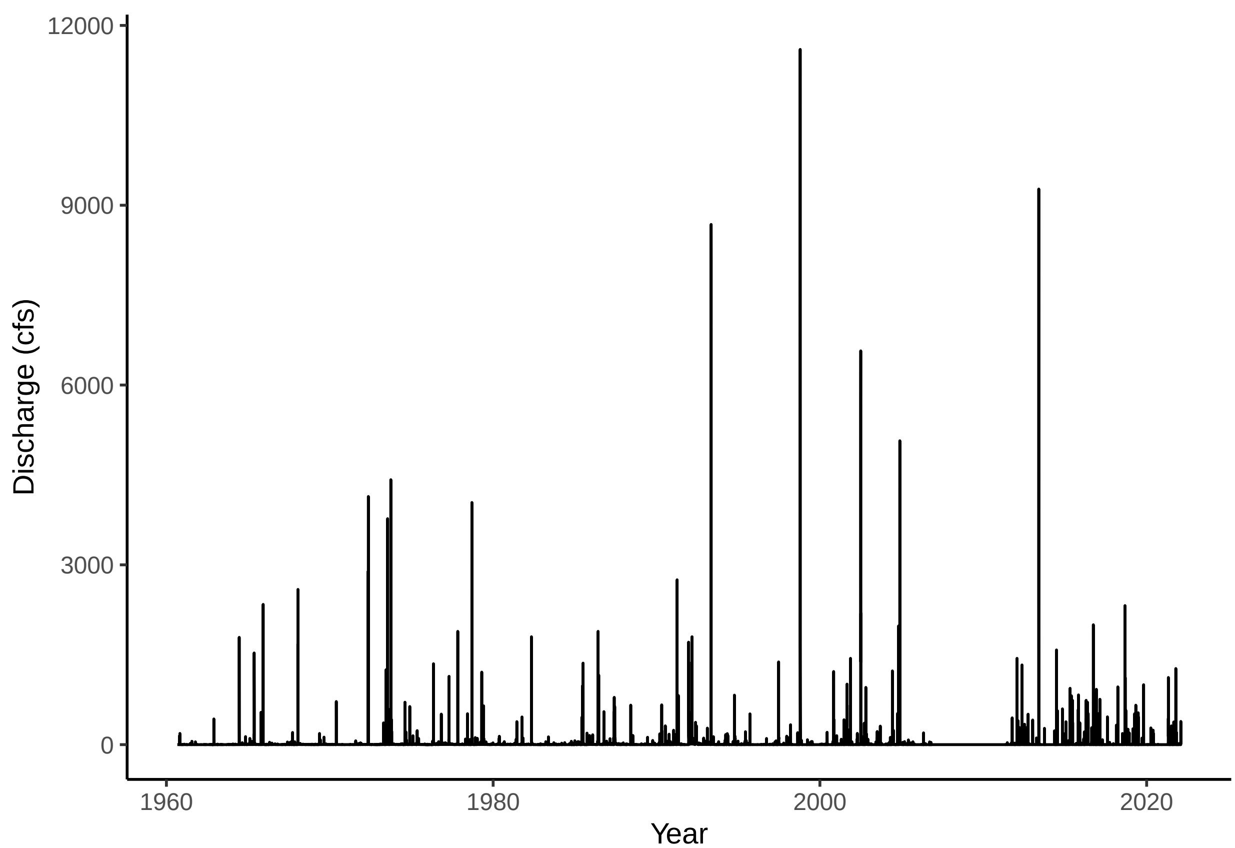
Examining the effects of urbanization on discharge

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). Is this system exhibiting stationarity?

```
# Import data
SaladoDischarge <- readNWISdv(siteNumbers = "08178700",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "")
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
```

```
ggplot(SaladoDischarge, aes(x = Date, y = Discharge)) +
  geom_line() +
  labs(x = "Year", y = "Discharge (cfs)")
```



```
# flashy, event based discharge (random high, fast peaks that are not seasonal). more small-medium size
```

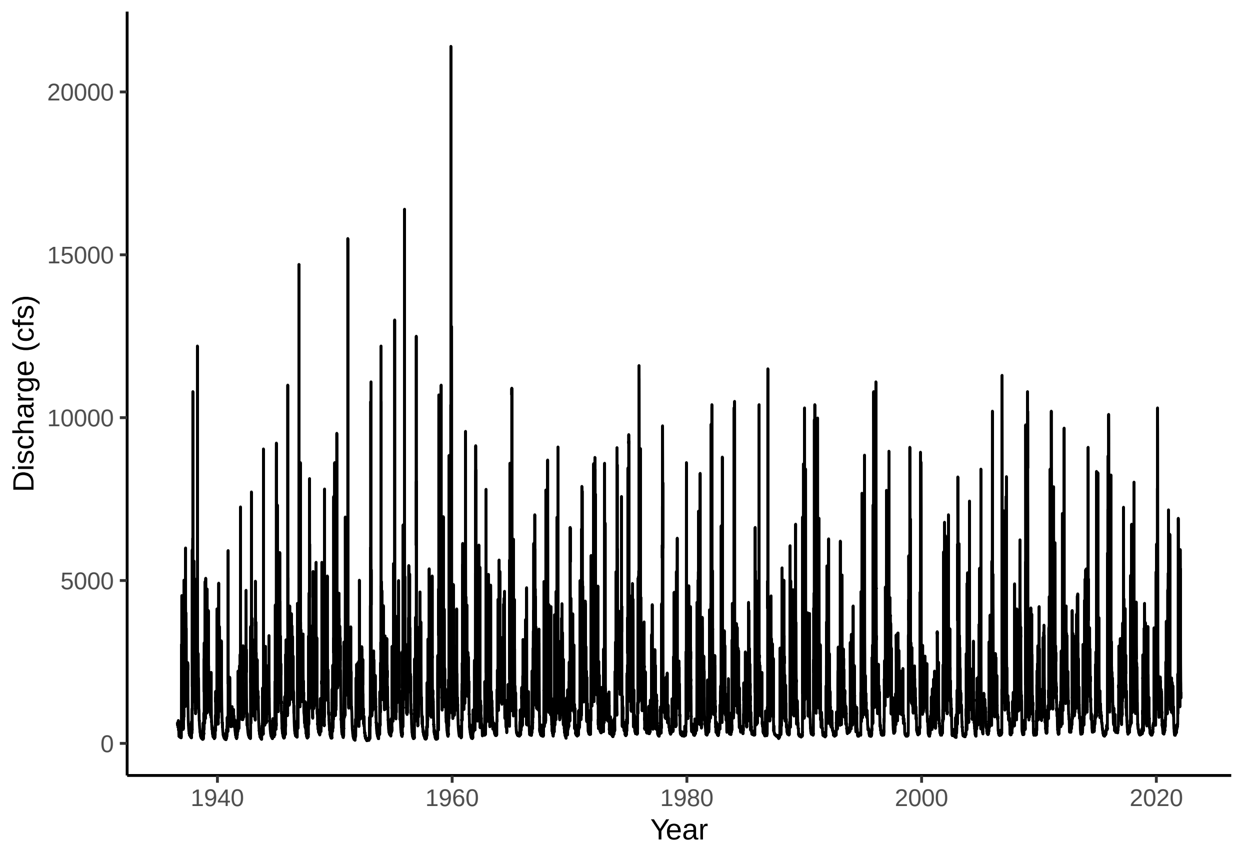
Examining the effects of dam construction on recurrence intervals

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?

```
GreenDischarge <- readNWISdv(siteNumbers = "12113000",
                         parameterCd = "00060", # discharge (ft3/s)
                         startDate = "")
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
```

```
ggplot(GreenDischarge, aes(x = Date, y = Discharge)) +
  geom_line() +
  labs(x = "Year", y = "Discharge (cfs)")
```




```
# dam was built for flood control! you can see the drop in discharge.
```

Bonus content: Flow Duration Curves and Low Flow Statistics

Flow-duration curves can be generated from daily discharge data, similar to how we calculated recurrence intervals for annual data.

$P = 100 * (m/(n + 1))$

where P is the exceedance probability, m is the ranking of all daily mean flows in the period of record (at least 10 years), and n is the total number of daily mean flows.

We focused today on recurrence intervals, which use peak flow statistics. On the other end of the discharge gradient are low flow statistics, most commonly estimated by 7Q2 and 7Q10 metrics (7-day, 2-year and 10-year annual low flow statistics). These can be used to evaluate drought conditions and are another metric for evaluating stationarity in rivers and streams.

See the USGS description of these statistics here:(Calculating Flow-Duration and Low-Flow Frequency Statistics at Streamflow-Gaging Stations)[https://pubs.usgs.gov/sir/2008/5126/section3.html]