2: R and Data Visualization Boot Camp

Hydrologic Data Analysis | Kateri Salk

Fall 2019

Lesson Objectives

1. Analyze seasonal and interannual characteristics of stream discharge
2. Compare discharge patterns in different regions of the United States
3. Communicate findings with peers through oral, visual, and written modes

Opening Discussion

What are the physical properties of streams and rivers? What variables might be of interest?

Session Set Up

```
getwd()
```

```
## [1] "/Users/katerisalk/Box Sync/Courses/Hydrologic Data Analysis/Lessons"
```

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

theme_set(theme_classic())
```

Discharge graphing 1. Seasonal cycles for Eno River (02096500), Verde River in Arizona (09504000; largest perennial stream in AZ), Bitterroot River in Montana (14299800), Sauk River in Minnesota (05270500), Nehalem River in Oregon (14299800). a. look at time series b. create DOY vs. discharge graph. Black points are mean discharge for a given day of year, individual dates are filled in by color (year).

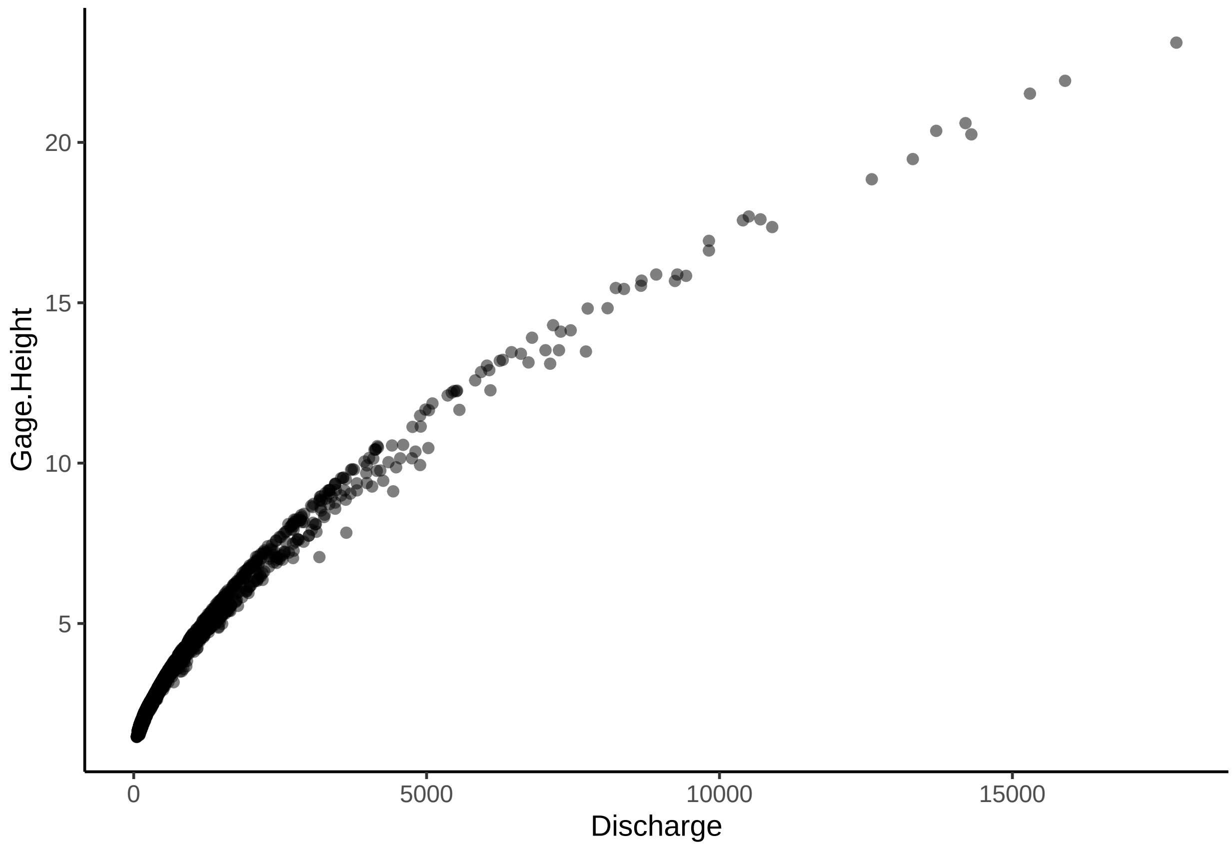
Discharge

One of the most important physical characteristics of a stream or river is **discharge**, the volume of water moving through the stream over a given amount of time. Discharge can be measured directly by measuring the velocity of flow in several spots in a stream and multiplying the flow velocity over the cross-sectional area of the stream. However, this method is effort-intensive. An easier way to approximate discharge is by developing a **rating curve** for a stream at a given sampling point. To develop a rating curve, a series of measurements of discharge and stream stage must be made together to develop a mathematical relationship. Then, subsequent discharge measurements can be calculated from a measurement of stream stage, measured by a gage that takes readings remotely.

```
EnoDischarge.ratingcurve <- readNWISdv(siteNumbers = "02096500",
                     parameterCd = c("00060", "00065"),
                     startDate = "2009-08-01",
                     endDate = "2019-07-31")

names(EnoDischarge.ratingcurve)[4:7] <- c("Discharge", "D.Approval.Code", "Gage.Height", "GH.Approval.C
```

```
DvsGHplot <-
  ggplot(EnoDischarge.ratingcurve, aes(x = Discharge, y = Gage.Height)) +
  geom_point(alpha = 0.5)
print(DvsGHplot)
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```



Data import

We will again be using the dataRetrieval package, which allows us to access USGS hydrologic data. We will be using the hydrologic data capabilities of this package today, but the package also comes with additional functionality to access water quality data from the Water Quality Portal.

dataRetrieval Vignette: https://cran.r-project.org/web/packages/dataRetrieval/vignettes/dataRetrieval.html

The function `whatNWISdata` allows us to access metadata about a site, including what variables have been monitored at that site, start and end dates, and the number of samples for a given variable.

We will be analyzing data from the following rivers/streams today:

- Eno River in North Carolina
- Verde River in Arizona (the largest perennial stream in AZ)
- Bitterroot River in Montana
- Sauk River in Minnesota (Dr. Salk's hometown)
- Nehalem River in Oregon

```r
EnoSummary <- whatNWISdata(siteNumbers = "02096500")
VerdeSummary <- whatNWISdata(siteNumbers = "09504000")
BitterrootSummary <- whatNWISdata(siteNumbers = "12344000")
SaukSummary <- whatNWISdata(siteNumbers = "05270500")
NehalemSummary <- whatNWISdata(siteNumbers = "14299800")
```

What data are available from the sites? Pay particular attention to the parameter codes and statistic codes.

What is the period of record for mean daily discharge at each of these sites?

Eno River:

Verde River:

Bitterroot River:

Sauk River:

Nehalem River:

The function `readNWISdv` allows us to access daily hydrologic data, which include discharge, gage height, temperature, precipitation, and pH. The function `readNWISqw` allows us to access water quality data.

We will access the last 10 years of data for our purposes today. Alternatively, you could code `startDate = ""` to indicate you want to access all available data for the entire period of record. Then, you could filter out the years you don't want to analyze.

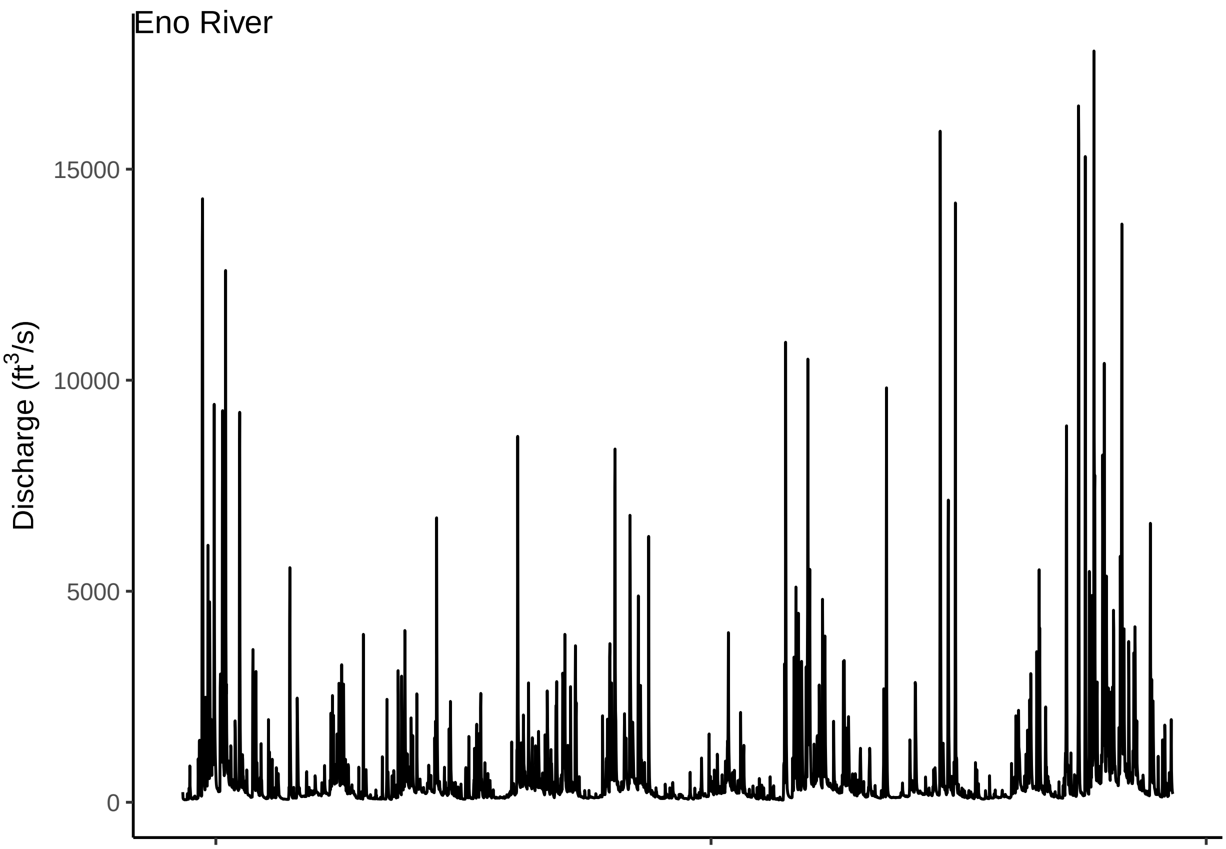
```r
# Import data
EnoDischarge <- readNWISdv(siteNumbers = "02096500",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "2009-09-01",
                     endDate = "2019-08-31")

VerdeDischarge <- readNWISdv(siteNumbers = "09504000",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "2009-09-01",
                     endDate = "2019-08-31")

BitterrootDischarge <- readNWISdv(siteNumbers = "12344000",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "2009-09-01",
                     endDate = "2019-08-31")

SaukDischarge <- readNWISdv(siteNumbers = "05270500",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "2009-09-01",
                     endDate = "2019-08-31")

NehalemDischarge <- readNWISdv(siteNumbers = "14299800",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "2009-09-01",
                     endDate = "2019-08-31")
```

Data Wrangling

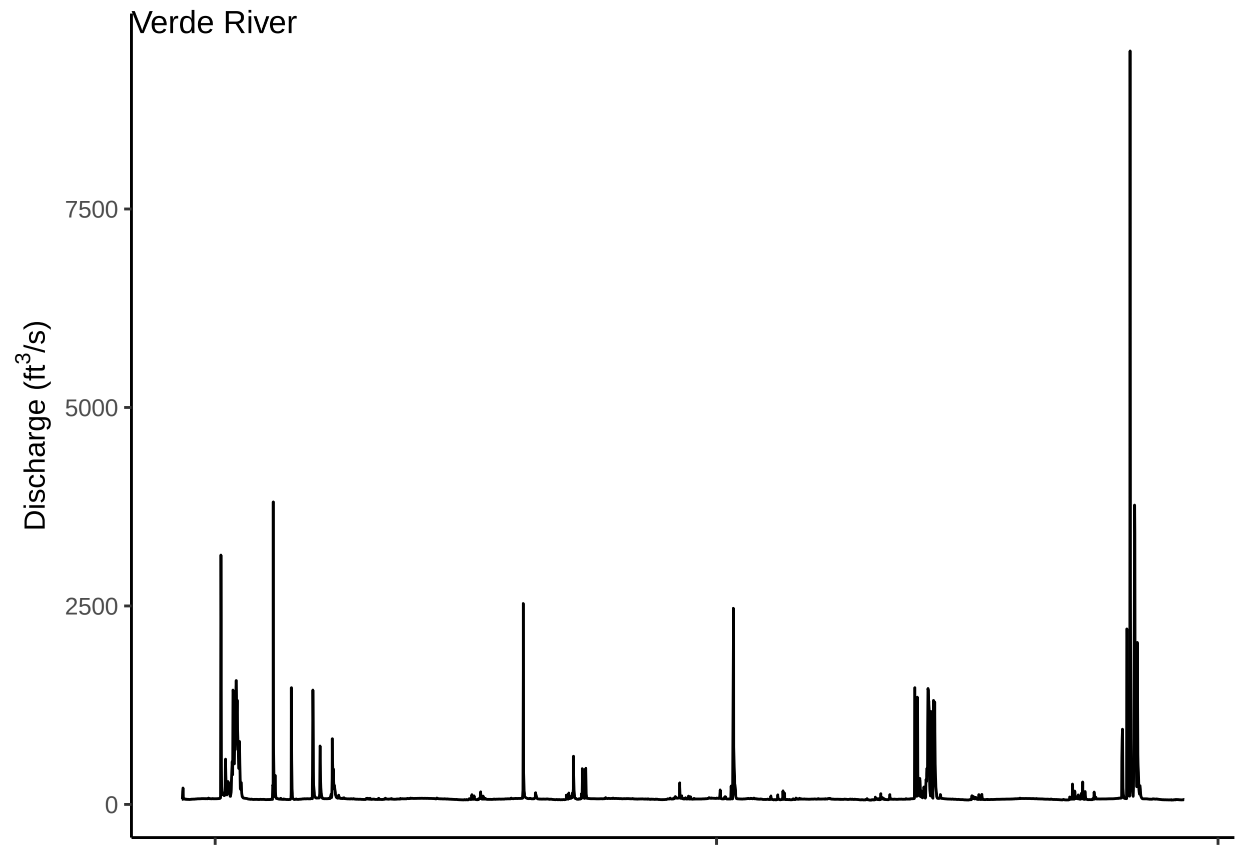
```r
# Renaming columns (one method of multiple)
names(EnoDischarge)[4:5] <- c("Eno.Discharge", "Eno.Approval.Code")
names(VerdeDischarge)[4:5] <- c("Verde.Discharge", "Verde.Approval.Code")
names(BitterrootDischarge)[4:5] <- c("Bitterroot.Discharge", "Bitterroot.Approval.Code")
names(SaukDischarge)[4:5] <- c("Sauk.Discharge", "Sauk.Approval.Code")
names(NehalemDischarge)[4:5] <- c("Nehalem.Discharge", "Nehalem.Approval.Code")

# Remove agency_cd column from each data frame
EnoDischarge <- select(EnoDischarge, -agency_cd)
VerdeDischarge <- select(VerdeDischarge, -agency_cd)
BitterrootDischarge <- select(BitterrootDischarge, -agency_cd)
SaukDischarge <- select(SaukDischarge, -agency_cd)
NehalemDischarge <- select(NehalemDischarge, -agency_cd)

# Join data frames
CombinedDischarge <- full_join(EnoDischarge, VerdeDischarge, by = "Date") %>%
  full_join(., BitterrootDischarge, by = "Date") %>%
  full_join(., SaukDischarge, by = "Date") %>%
  full_join(., NehalemDischarge, by = "Date")
```

Data Visualization

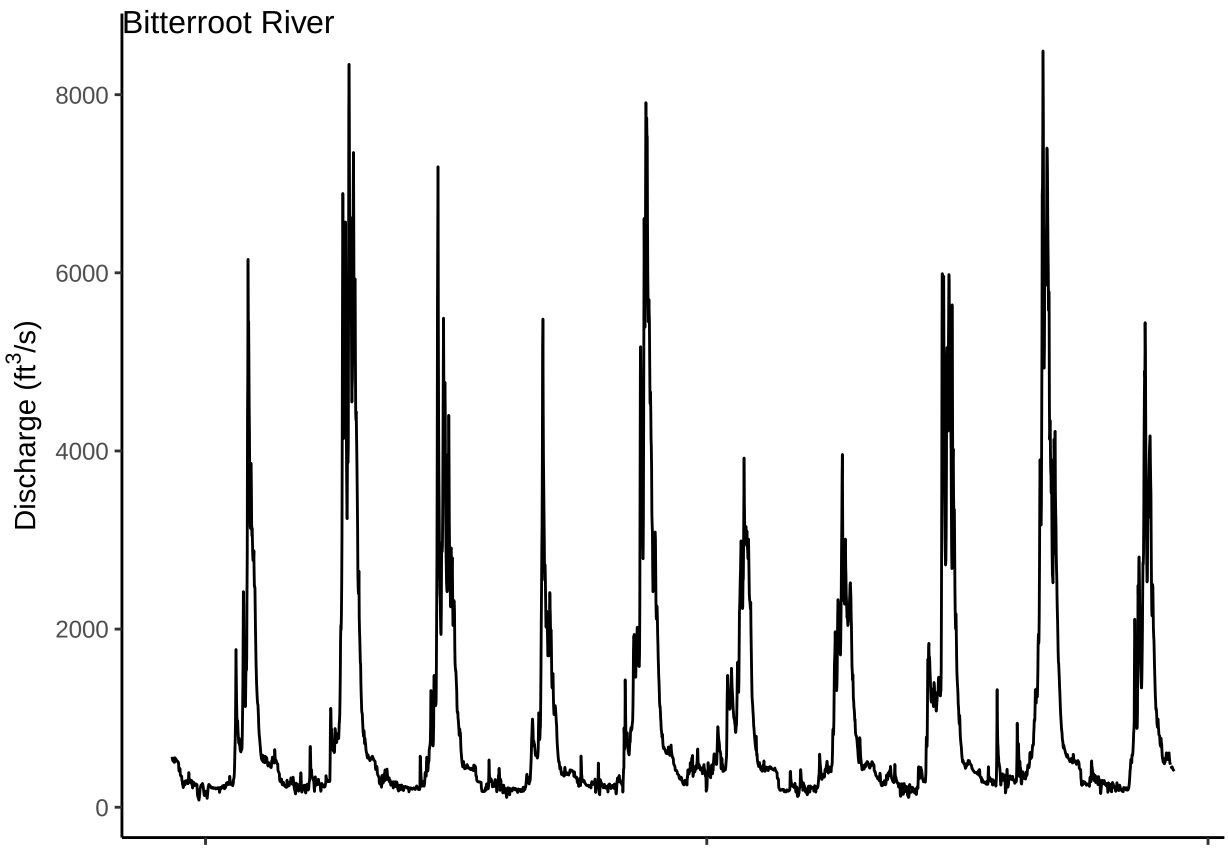
```r
# Build a ggplot
EnoPlot <-
  ggplot(CombinedDischarge, aes(x = Date, y = Eno.Discharge)) +
  geom_line() +
  ggtitle("Eno River") +
  labs(x = "", y = expression("Discharge (ft"^3*"/s)")) +
  theme(plot.title = element_text(margin = margin(b = -10), size = 12),
        axis.text.x = element_blank(), axis.title.x = element_blank())
print(EnoPlot)
```



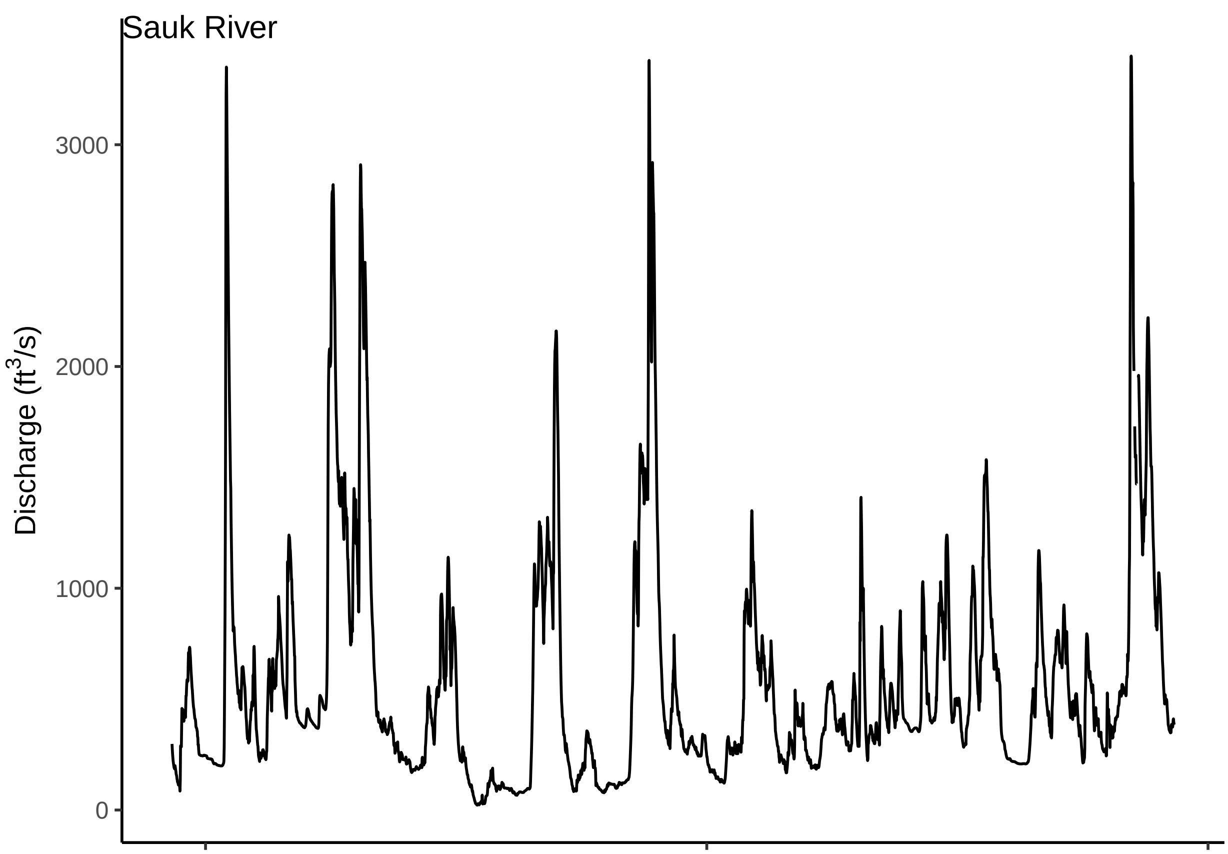

```r
VerdePlot <-
  ggplot(CombinedDischarge, aes(x = Date, y = Verde.Discharge)) +
  geom_line() +
  ggtitle("Verde River") +
  labs(x = "", y = expression("Discharge (ft"^3*"/s)")) +
  theme(plot.title = element_text(margin = margin(b = -10), size = 12),
        axis.text.x = element_blank(), axis.title.x = element_blank())
print(VerdePlot)
```



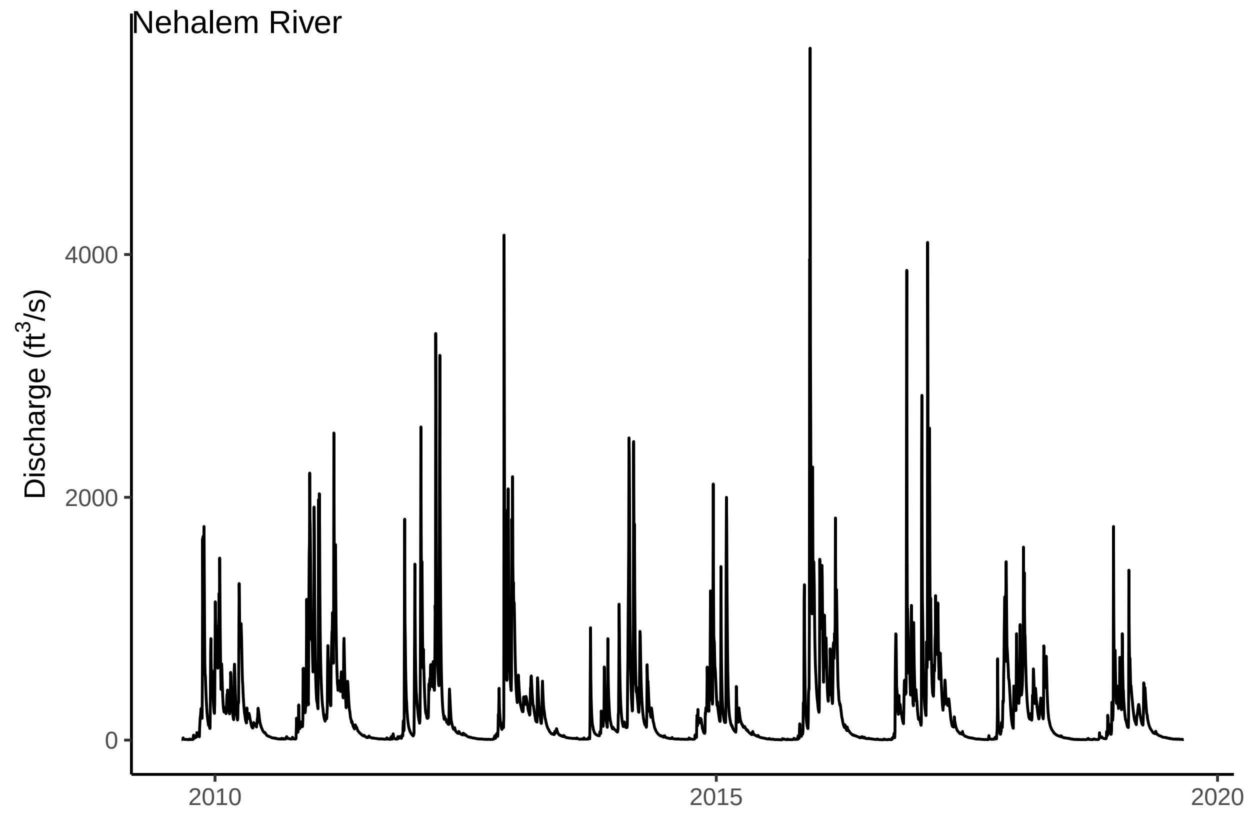

```r
BitterrootPlot <-
  ggplot(CombinedDischarge, aes(x = Date, y = Bitterroot.Discharge)) +
  geom_line() +
  ggtitle("Bitterroot River") +
  labs(x = "", y = expression("Discharge (ft"^3*"/s)")) +
  theme(plot.title = element_text(margin = margin(b = -10), size = 12),
        axis.text.x = element_blank(), axis.title.x = element_blank())
print(BitterrootPlot)
```



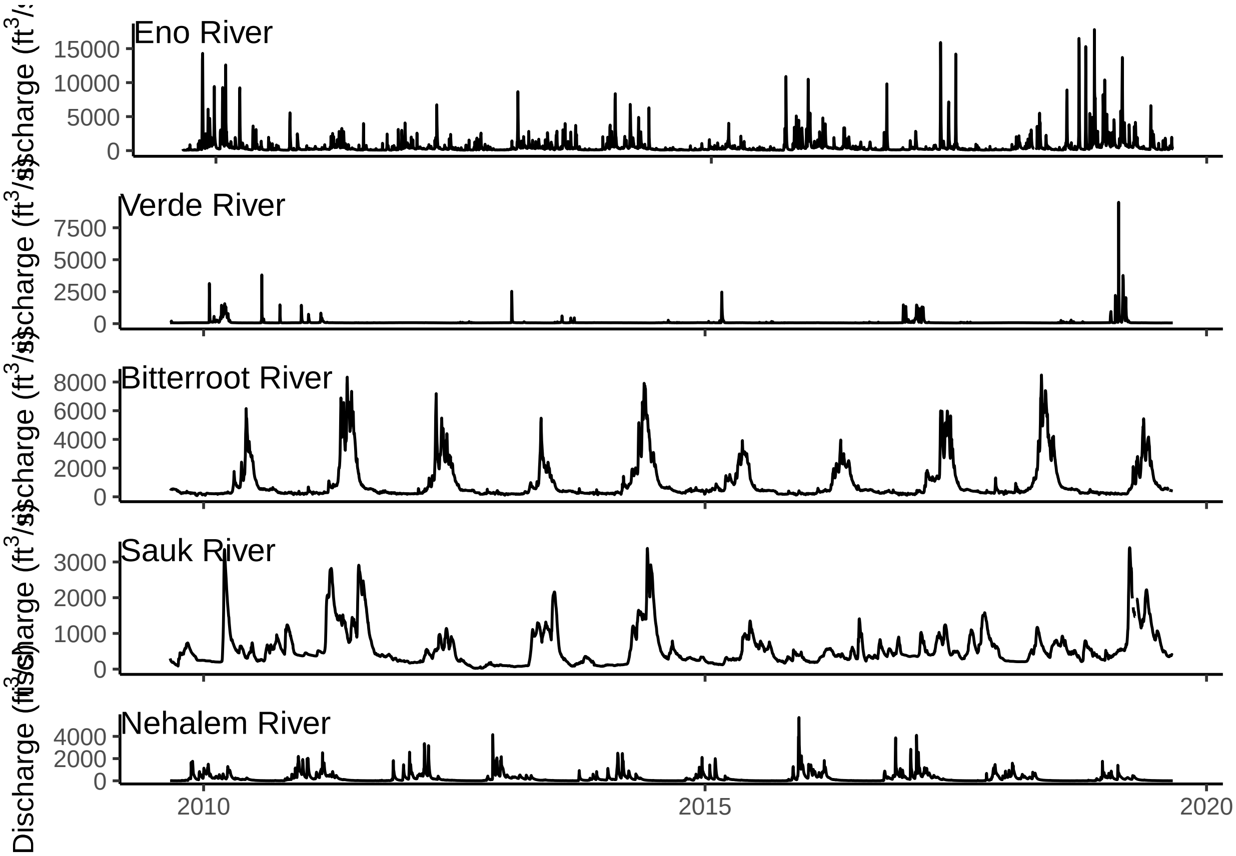

```r
SaukPlot <-
  ggplot(CombinedDischarge, aes(x = Date, y = Sauk.Discharge)) +
  geom_line() +
  ggtitle("Sauk River") +
  labs(x = "", y = expression("Discharge (ft"^3*"/s)")) +
  theme(plot.title = element_text(margin = margin(b = -10), size = 12),
        axis.text.x = element_blank(), axis.title.x = element_blank())
print(SaukPlot)
```




```r
NehalemPlot <-
  ggplot(CombinedDischarge, aes(x = Date, y = Nehalem.Discharge)) +
  geom_line() +
  ggtitle("Nehalem River") +
  labs(x = "", y = expression("Discharge (ft"^3*"/s)")) +
  theme(plot.title = element_text(margin = margin(b = -10), size = 12))
print(NehalemPlot)
```




```
CombinedPlot <-
  plot_grid(EnoPlot, VerdePlot, BitterrootPlot, SaukPlot, NehalemPlot,
            ncol = 1)
print(CombinedPlot)
```



How does the magnitude of discharge compare for these rivers?

Which of these rivers have a seasonal cycle of discharge? How might you explain the presence and timing of seasonal cycles based on climatic conditions?

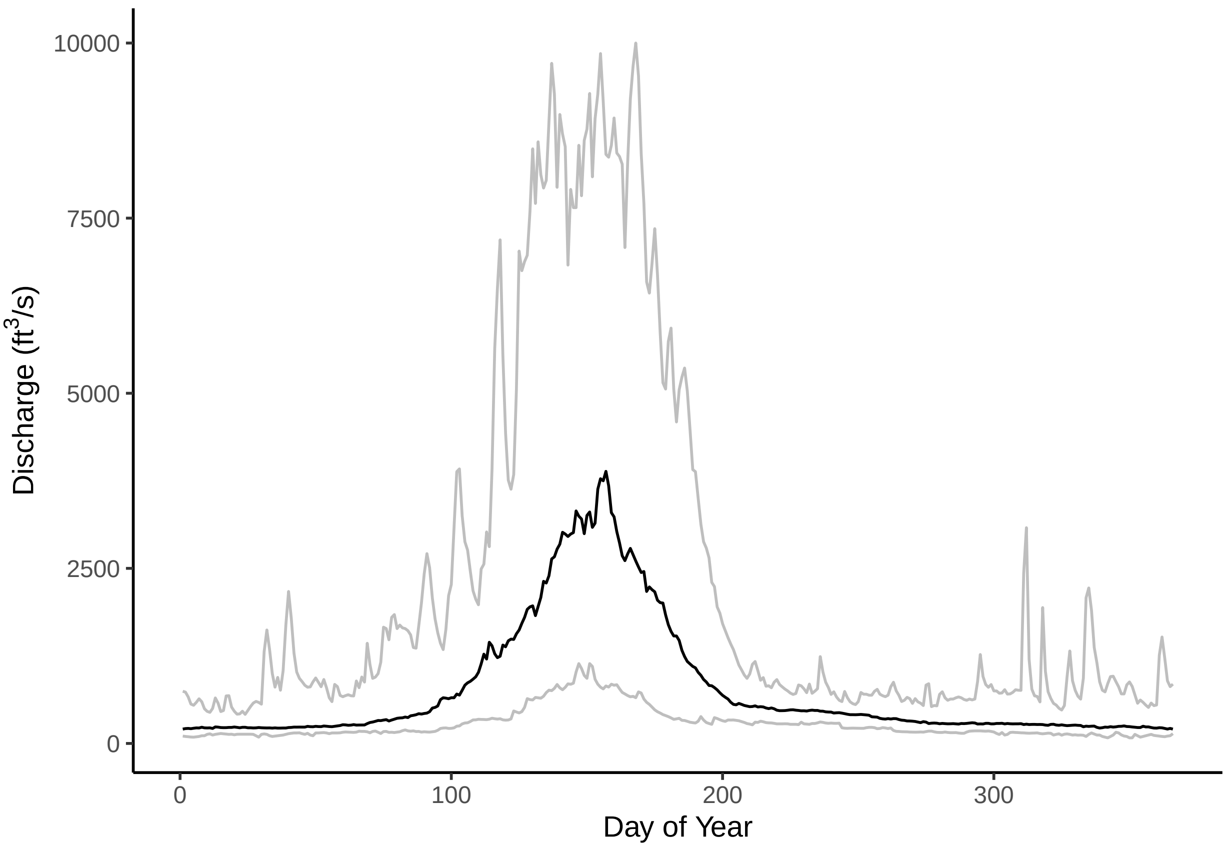
Hypothetical question: If you wanted to create a ggplot using facets instead of `plot_grid`, how would you go about wrangling your dataset differently?

Defining "typical" discharge pattern

Let's take the Bitterroot River, which displays a clear seasonal cycle. Let's define the median discharge on a given day of year and overlay this onto the range experienced. We will take a longer period of record this time.

```
BitterrootDischarge.Historical <- readNWISdv(siteNumbers = "12344000",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "1969-09-01",
                     endDate = "2019-08-31")

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

BitterrootDischarge.Historical <- mutate(BitterrootDischarge.Historical,
```

```
                                         DOY = yday(Date))

BitterrootDischarge.Pattern <- BitterrootDischarge.Historical %>%
  group_by(DOY) %>%
  summarise(Median.Discharge = median(Discharge),
            Max.Discharge = max(Discharge),
            Min.Discharge = min(Discharge))

BitterrootPatternPlot <-
  ggplot(BitterrootDischarge.Pattern, aes(x = DOY)) +
  geom_line(aes(y = Median.Discharge)) +
  geom_line(aes(y = Max.Discharge), color = "gray") +
  geom_line(aes(y = Min.Discharge), color = "gray") +
  labs(x = "Day of Year", y = expression("Discharge (ft"^3*"/s)"))
print(BitterrootPatternPlot)
```



Challenge: Edit the code above so that rather than plotting the minimum and maximum discharge, you are plotting the 95 % confidence interval of the discharge on a given date. Hint: look up the formula for confidence interval.

Closing Discussion

What are your impressions of seasonal and interannual discharge for our example rivers? How might you go about interpreting discharge data for any given river?