7: Water Quality in Rivers

Hydrologic Data Analysis | Cathy Chamberlin

Fall 2019

Lesson Objectives

1. Analyze inorganic aspects of water quality following a watershed disturbance.
2. Compare water quality data to federal standards.
3. Communicate findings with peers through oral, visual, and written modes

Opening Discussion

What is water quality? What are some of the things it encompasses?

Session Set Up

```
getwd()
```

```
## [1] "C:/Users/cac118/Box Sync/Hydrologic Data Analysis/Lessons"
```

```
packages <- c("dataRetrieval", "tidyverse", "cowplot", "lubridate")
invisible(suppressPackageStartupMessages(lapply(packages,
                                                library,
                                                character.only = TRUE)))


theme_set(theme_classic())
```

Specific Conductance and pH

pH is a measure of the acidity of water. Most natural rivers have pH close to 7 (neutral), but depending on the geology of the watershed they drain, pH can be higher or lower. Most aquatic life has difficulty living in pH less than 6. In areas high in limestone, such as much of the southern Appalachian mountain range, pH tends to be between 7.5 and 8.

Specific conductance is a measure of water's ability to conduct electricity. Because conductivity depends on the number of ions in solution, specific conductance is an inexpensive estimate of how many salts are in solution. Specific conductance is dependant on the *total* ionic strength of the water, and cannot be directly converted to concentrations of any one ion. In general though, saltier water will have higher specific conductance. Units for specific conductance in freshwater are microsiemens per centimeter at 25°C.

Why are these important measures of water quality?

What types of disturbances can impact these parameters?

How does mining impact specific conductance and pH?

Let's import data from two sites in West Virginia. One site, Twelvepole Creek, experienced comparatively less mining in its water shed than the other, the Kanawha river. We will read in all of the data, but will start by looking at just data from before 1977.

```
dat.raw <- readNWISqw(siteNumbers = c("03201300","03206790"), #Kanawha river & Twelvepole Creek
                      parameterCd = c("00095", "00400"), # Specific Conductance & pH
                      startDate = "",
                      endDate = "")

str(dat.raw, give.attr = FALSE)
```

```
## 'data.frame':    1378 obs. of  34 variables:
##  $ agency_cd                 : chr  "USGS" "USGS" "USGS" "USGS" ...
##  $ site_no                   : chr  "03206790" "03206790" "03206790" "03206790" ...
##  $ sample_dt                 : Date, format: "1969-09-09" "1969-12-09" ...
##  $ sample_tm                 : chr  "11:00" "10:45" "10:45" "13:10" ...
##  $ sample_end_dt             : Date, format: NA NA ...
##  $ sample_end_tm             : chr  NA NA NA NA ...
##  $ sample_start_time_datum_cd: chr  "EDT" "EST" "EST" "EST" ...
##  $ tm_datum_rlbty_cd         : chr  "T" "T" "T" "T" ...
##  $ coll_ent_cd               : chr  "USGS-WRD" "USGS-WRD" "USGS-WRD" "USGS-WRD" ...
##  $ medium_cd                 : chr  "WS" "WS" "WS" "WS" ...
##  $ project_cd                : chr  NA NA NA NA ...
##  $ aqfr_cd                   : chr  NA NA NA NA ...
##  $ tu_id                     : chr  NA NA NA NA ...
##  $ body_part_id              : chr  NA NA NA NA ...
##  $ hyd_cond_cd               : chr  "A" "A" "A" "A" ...
##  $ samp_type_cd              : chr  "9" "9" "9" "9" ...
##  $ hyd_event_cd              : chr  "9" "9" "9" "9" ...
##  $ sample_lab_cm_tx          : chr  NA NA NA NA ...
##  $ parm_cd                   : chr  "00095" "00095" "00400" "00095" ...
##  $ remark_cd                 : chr  NA NA NA NA ...
##  $ result_va                 : num  118 90 7.2 48 6.8 75 6.8 65 7 65 ...
##  $ val_qual_tx               : chr  NA NA NA NA ...
##  $ meth_cd                   : chr  NA NA NA NA ...
##  $ dqi_cd                    : chr  "A" "A" "A" "A" ...
##  $ rpt_lev_va                : logi  NA NA NA NA NA NA ...
##  $ rpt_lev_cd                : chr  NA NA NA NA ...
##  $ lab_std_va                : logi  NA NA NA NA NA NA ...
##  $ prep_set_no               : logi  NA NA NA NA NA NA ...
##  $ prep_dt                   : logi  NA NA NA NA NA NA ...
##  $ anl_set_no                : logi  NA NA NA NA NA NA ...
##  $ anl_dt                    : logi  NA NA NA NA NA NA ...
##  $ result_lab_cm_tx          : chr  NA NA NA NA ...
##  $ anl_ent_cd                : chr  NA NA NA NA ...
##  $ startDateTime             : POSIXct, format: "1969-09-09 11:00:00" "1969-12-09 10:45:00" ...
```

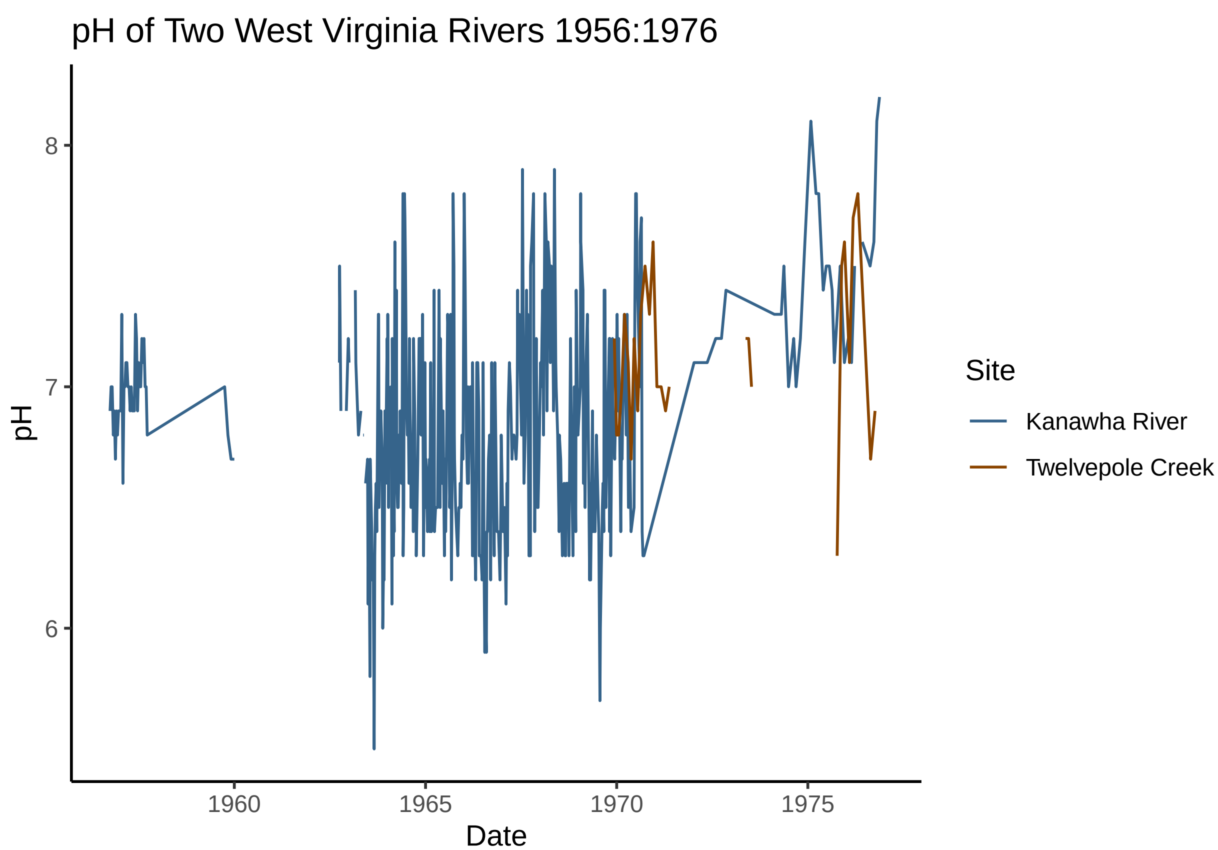
```
dat <- dat.raw %>%
  select(Site = site_no,
         Date = sample_dt,
         Parameter = parm_cd,
         Value = result_va) %>%
  group_by(Date, Parameter, Site) %>%
```

```
  summarize(Value = mean(Value)) %>%
  spread(key = Parameter, value = Value) %>%
  rename(SpC = '00095', pH = '00400') %>%
  mutate(Site = case_when(Site == "03201300" ~ "Kanawha River",
                          Site == "03206790" ~ "Twelvepole Creek"))

dat.early <- dat %>%
  filter(Date < "1977-01-01")
```

Now let's compare the pH and specific conductance of the two rivers.

```
pH.plot <- ggplot(dat.early, aes(x = Date, y = pH, color = Site)) +
  geom_line() +
  scale_color_manual(values = c("steelblue4", "darkorange4")) +
  ggtitle("pH of Two West Virginia Rivers 1956:1976")

print(pH.plot)
```

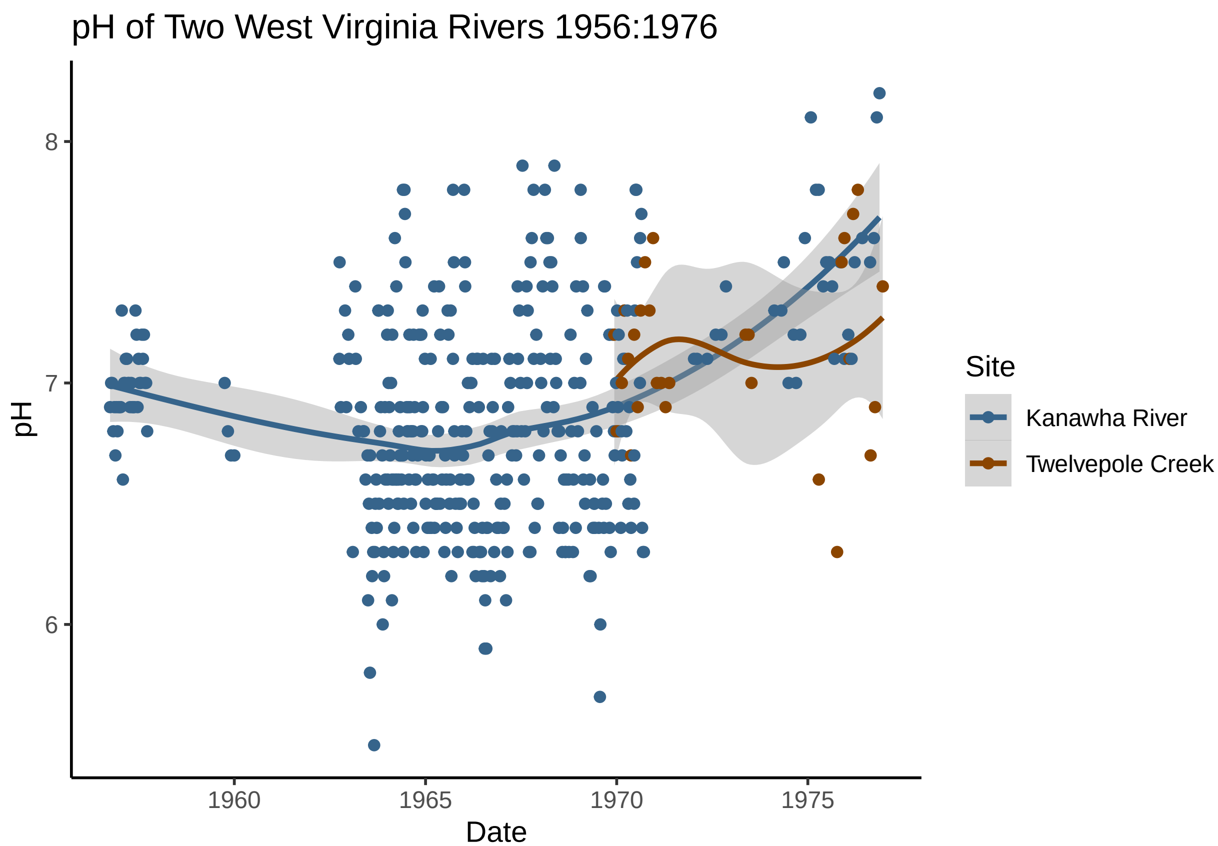
```
## Warning: Removed 1 rows containing missing values (geom_path).
```



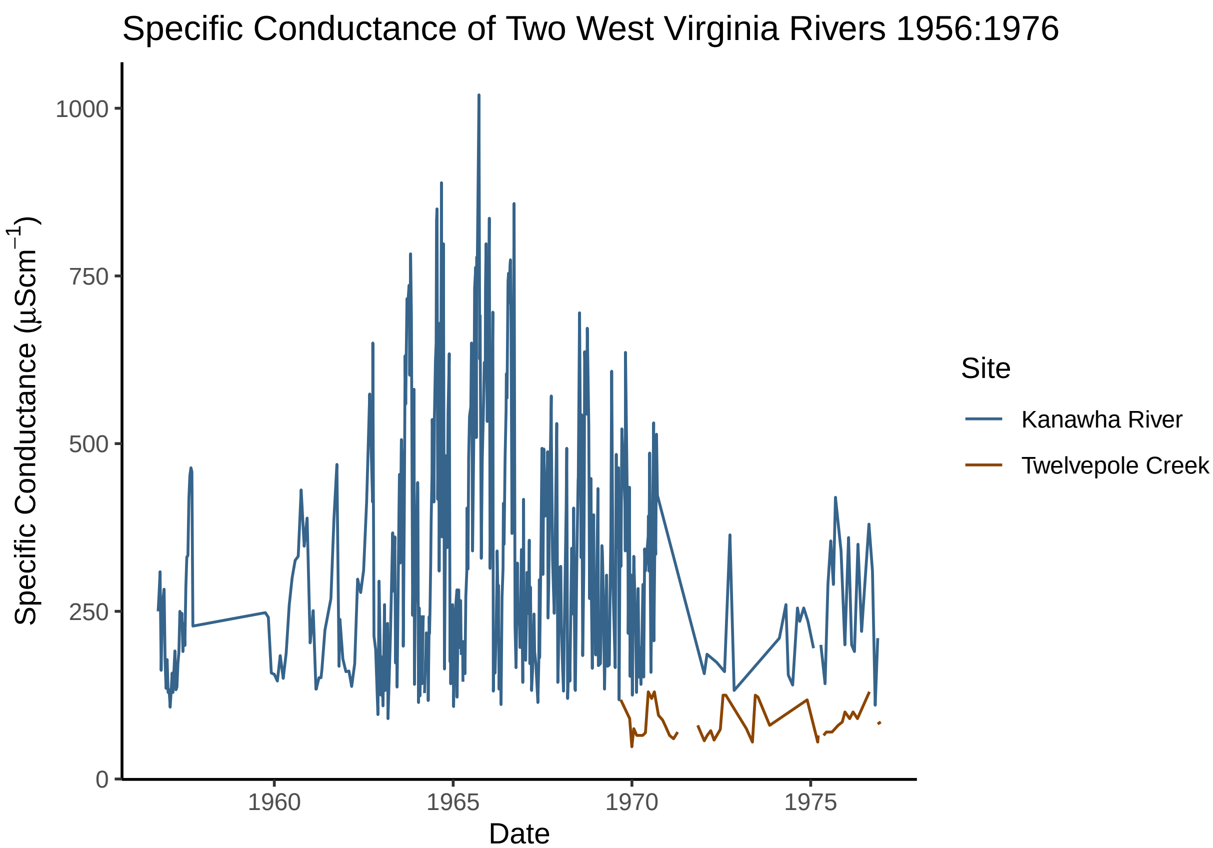

```
pH.plot.2 <- ggplot(dat.early, aes(x = Date, y = pH, color = Site)) +
  geom_smooth() +
  geom_point() +
  scale_color_manual(values = c("steelblue4", "darkorange4")) +
    ggtitle("pH of Two West Virginia Rivers 1956:1976")
```

```r
print(pH.plot.2)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 63 rows containing non-finite values (stat_smooth).

## Warning: Removed 63 rows containing missing values (geom_point).
```




```r
SpC.plot <- ggplot(dat.early, aes(x = Date, y = SpC, color = Site)) +
  geom_line() +
  labs(y = expression("Specific Conductance ("*mu*S * cm**-1*")")) +
  scale_color_manual(values = c("steelblue4", "darkorange4")) +
    ggtitle("Specific Conductance of Two West Virginia Rivers 1956:1976")


print(SpC.plot)
```



What differences do you see?

Why might mining be causing the changes you see?

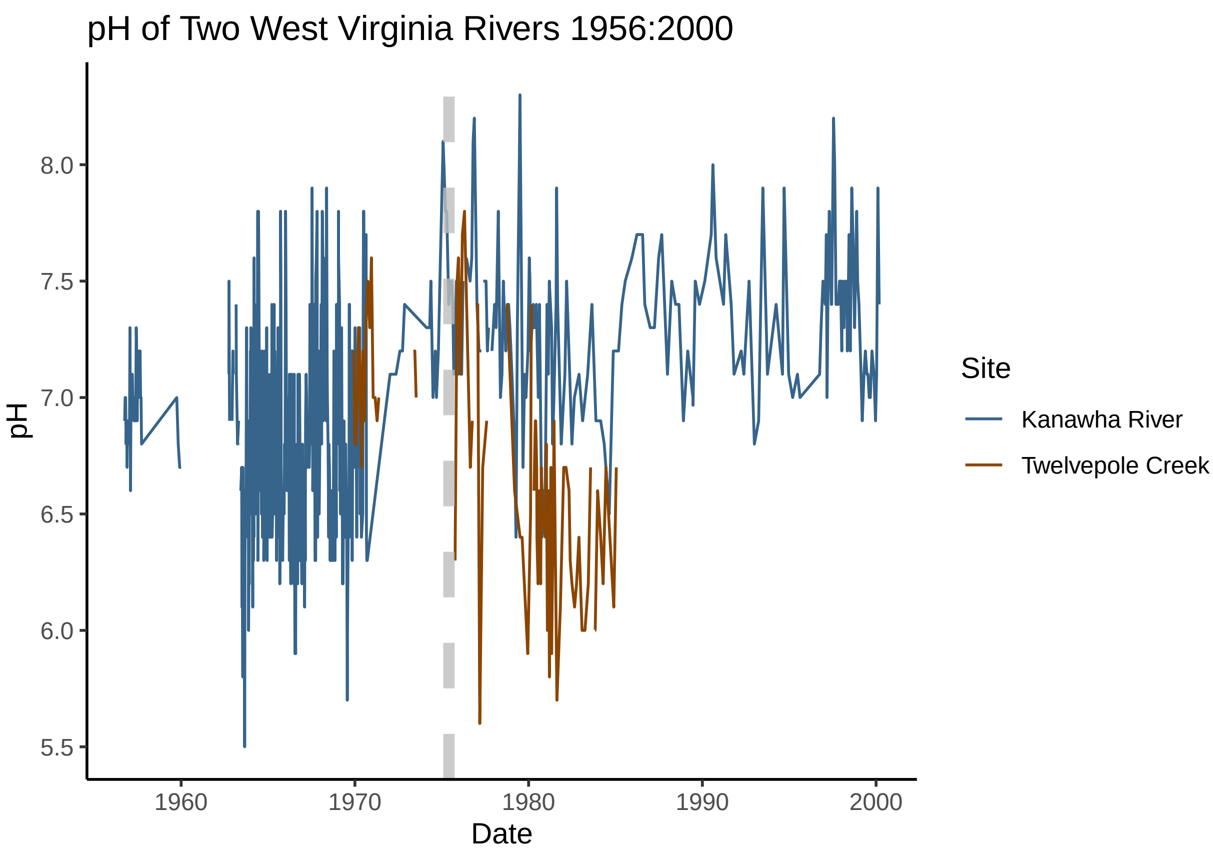
What is the data frequency of these water quality samples? Why don't we have daily values? How does this inform your interpretation of the data?

What can legislative protections do for water quality?

In 1977 the Surface Mining Control and Reclamation Act (SMCRA) was passed. Take a look at how water quality changed after 1977.

```
pH.plot <- ggplot(dat, aes(x = Date, y = pH, color = Site)) +
  geom_line() +
  geom_vline(xintercept = 1977, color = "grey", alpha = 0.8, lwd = 2, lty = 2) +
  scale_color_manual(values = c("steelblue4", "darkorange4")) +
  ggtitle("pH of Two West Virginia Rivers 1956:2000")

print(pH.plot)
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```

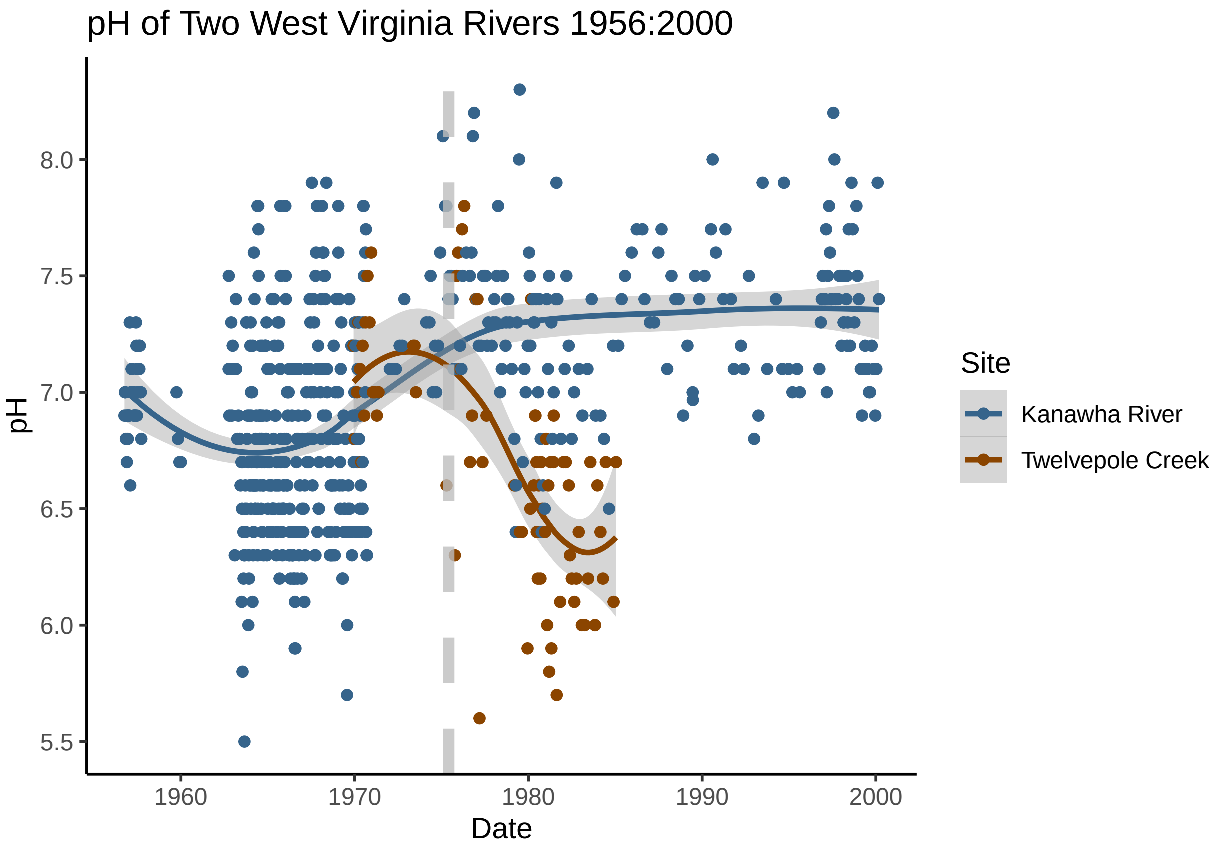



```
pH.plot.2 <- ggplot(dat, aes(x = Date, y = pH, color = Site)) +
  geom_smooth() +
  geom_point() +
  geom_vline(xintercept = 1977, color = "grey", alpha = 0.8, lwd = 2, lty = 2) +
  scale_color_manual(values = c("steelblue4", "darkorange4")) +
    ggtitle("pH of Two West Virginia Rivers 1956:2000")


print(pH.plot.2)
```

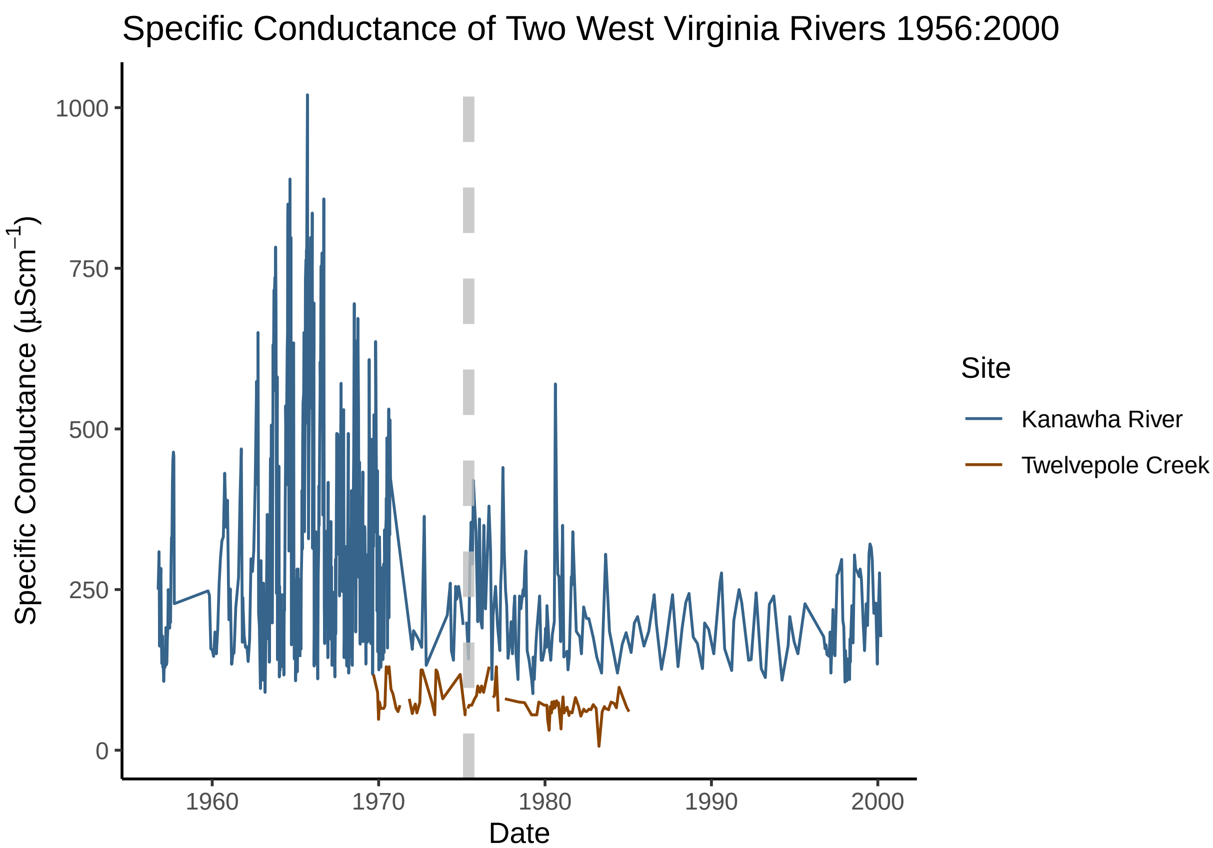
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 68 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 68 rows containing missing values (geom_point).
```




```
SpC.plot <- ggplot(dat, aes(x = Date, y = SpC, color = Site)) +
  geom_line() +
  geom_vline(xintercept = 1977, color = "grey", alpha = 0.8, lwd = 2, lty = 2) +
  labs(y = expression("Specific Conductance ("*mu*S * cm**-1*")")) +
  scale_color_manual(values = c("steelblue4", "darkorange4")) +
    ggtitle("Specific Conductance of Two West Virginia Rivers 1956:2000")


print(SpC.plot)
```



Discuss with your group about the differences you see.

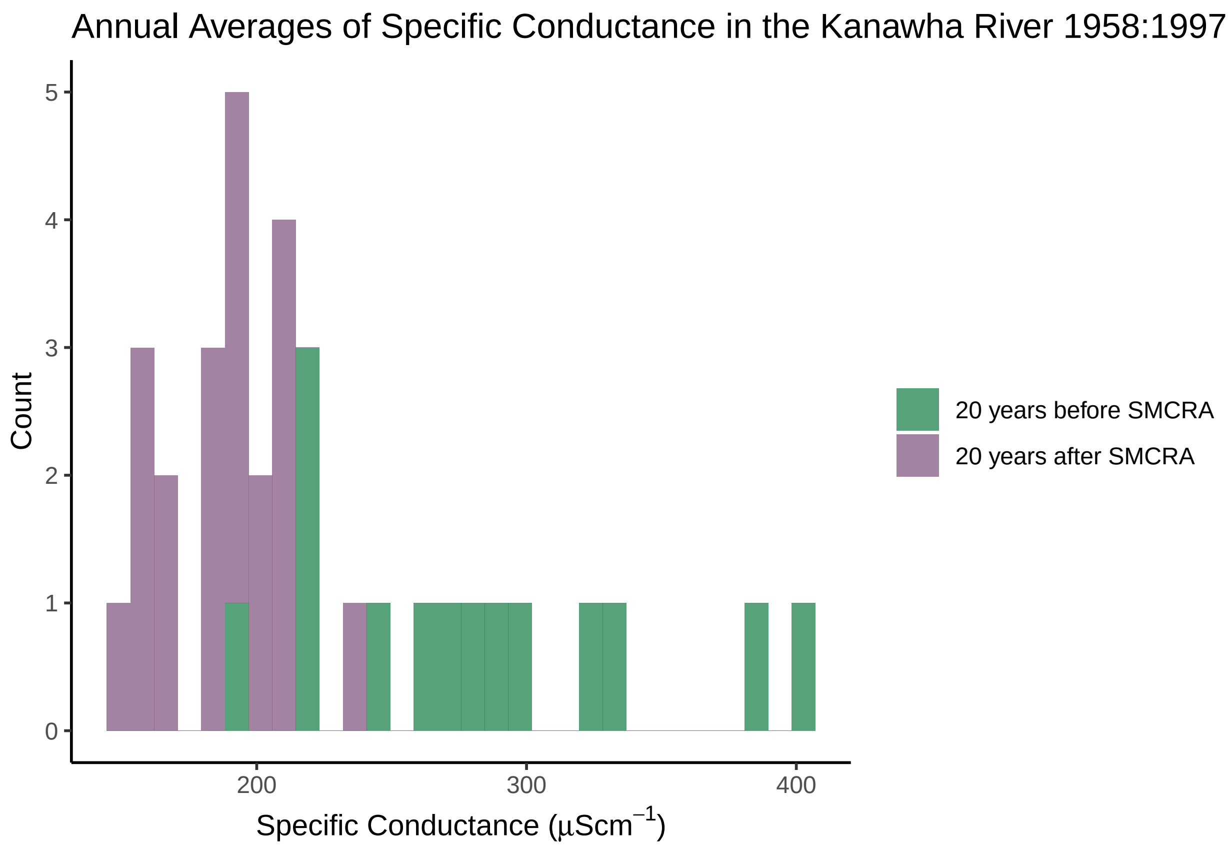
Were remediation efforts successful?

Let's test whether or not the yearly average concentrations in the Kanawha River are significantly different before and after 1977.

```
Kanawha.yearly.averages <- dat %>%
  filter(Site == "Kanawha River") %>%
  group_by(Year = year(Date)) %>%
  select(-Site, -Date) %>%
  summarize_all(mean) %>%
  mutate(bin = case_when(Year %in% c(1958:1977) ~ "Pre",
                         Year %in% c(1978:1997) ~ "Post"))

Kanawha.early <- Kanawha.yearly.averages %>% filter(Year %in% c(1958:1977))
Kanawha.late <- Kanawha.yearly.averages %>% filter(Year %in% c(1978:1997))


histograms <- ggplot(Kanawha.yearly.averages %>% filter(!is.na(bin)),
       aes(x = SpC, fill = bin)) +
  geom_histogram(alpha = 0.8) +
  ggtitle("Annual Averages of Specific Conductance in the Kanawha River 1958:1997") +
  labs(x = expression("Specific Conductance ("*mu*S * cm**-1*")"),
```

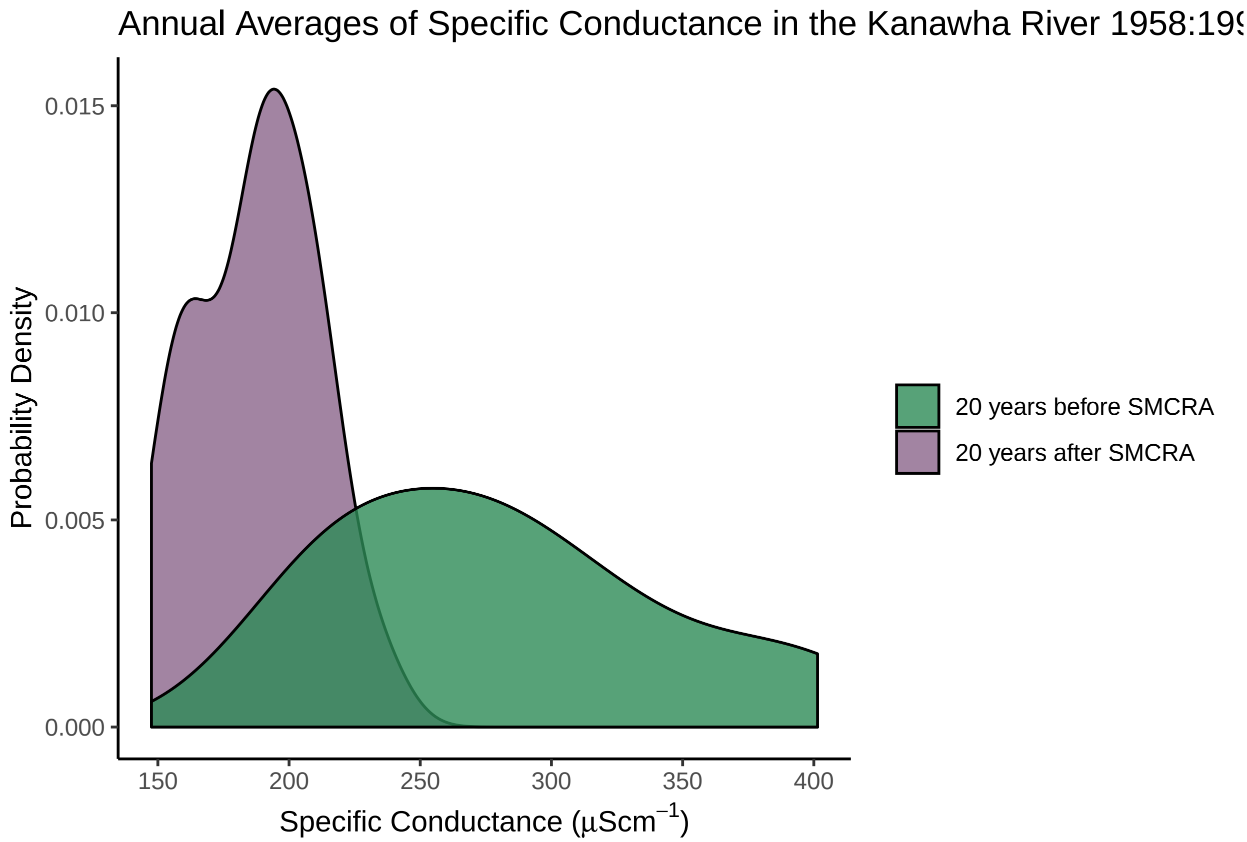
```r
       y = "Count") +
  scale_fill_manual(values = c("plum4", "seagreen4"),
                    name="",
                    breaks=c("Pre", "Post"),
                    labels=c("20 years before SMCRA", "20 years after SMCRA"))

print(histograms)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 3 rows containing non-finite values (stat_bin).
```

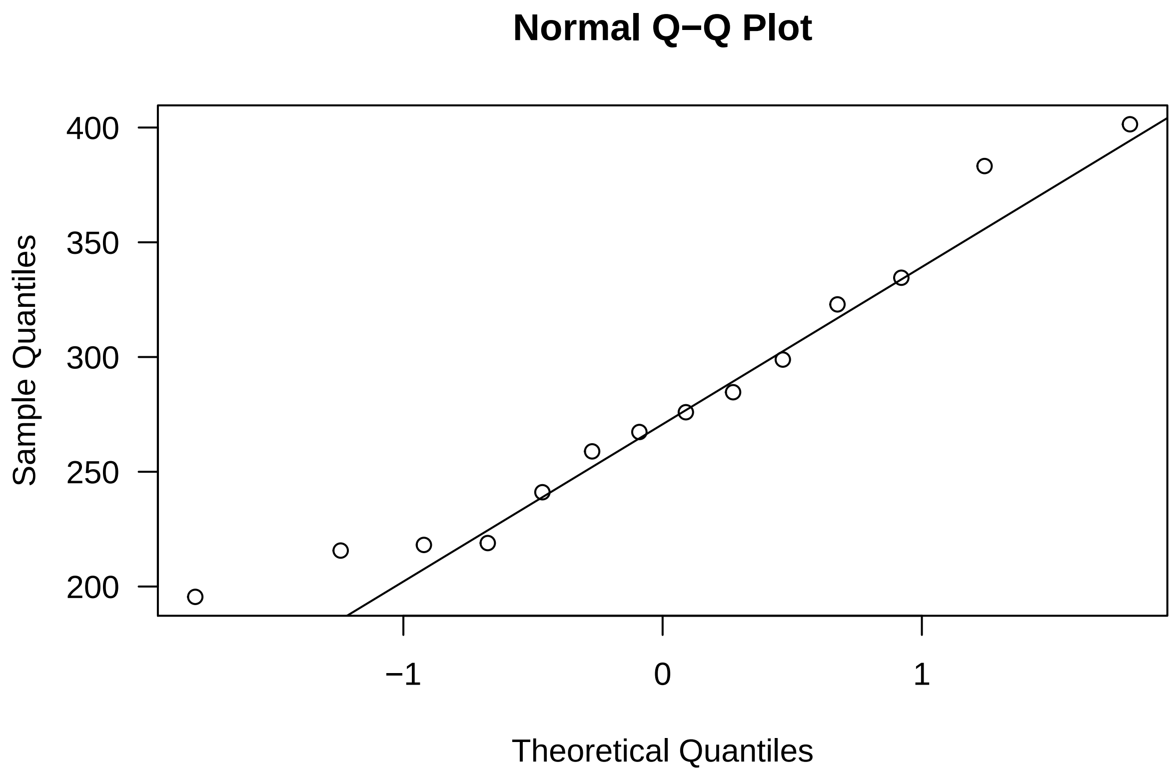



```r
densityplots <- ggplot(Kanawha.yearly.averages %>% filter(!is.na(bin)),
       aes(x = SpC, fill = bin)) +
  geom_density(alpha = 0.8) +
  ggtitle("Annual Averages of Specific Conductance in the Kanawha River 1958:1997") +
  labs(x = expression("Specific Conductance ("*mu*S * cm**-1*")"),
       y = "Probability Density") +
  scale_fill_manual(values = c("plum4", "seagreen4"),
                    name="",
                    breaks=c("Pre", "Post"),
                    labels=c("20 years before SMCRA", "20 years after SMCRA"))

print(densityplots)
```

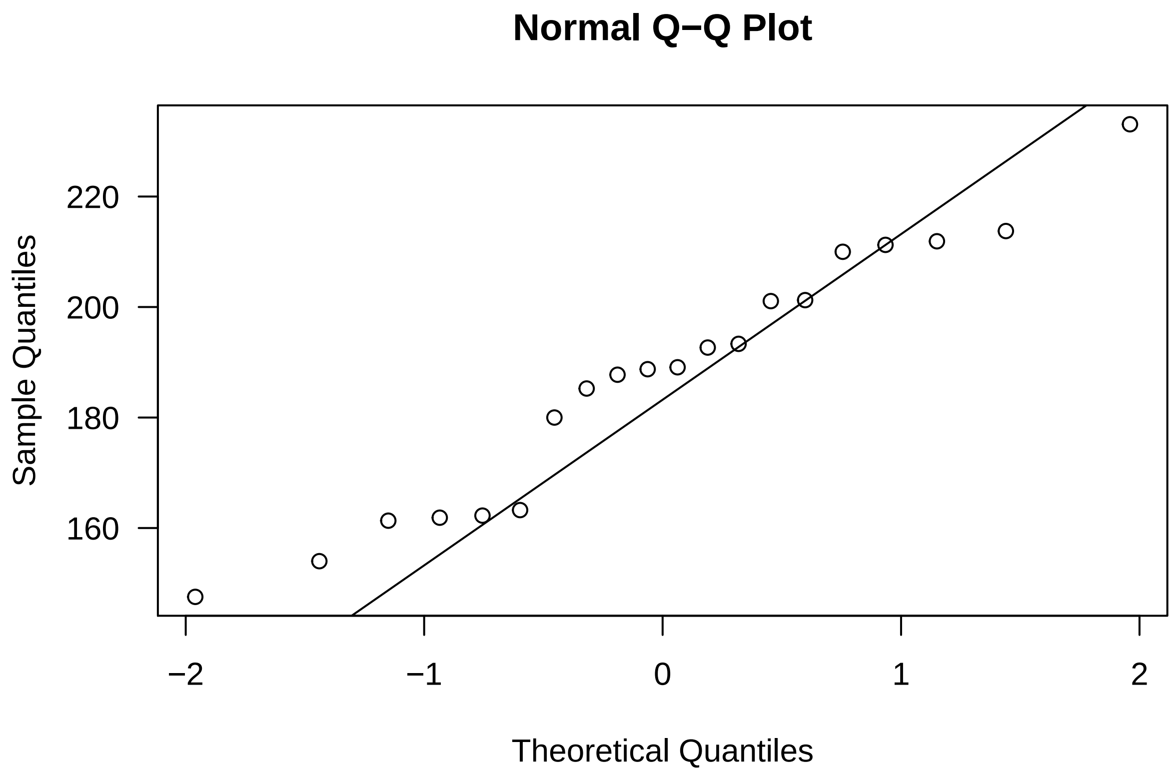
```
## Warning: Removed 3 rows containing non-finite values (stat_density).
```




```
x <- (Kanawha.yearly.averages %>% filter(bin == "Pre"))$SpC
y <- (Kanawha.yearly.averages %>% filter(bin == "Post"))$SpC

qqnorm(x, las = 1); qqline(x)
```




```r
qqnorm(y, las = 1); qqline(y)
```




```r
cat("Length x:", length(x), "\n", "Length y:", length(y))
```

```
## Length x: 17
##  Length y: 20
```

```r
t.test(x, y)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 5.2419, df = 15.489, p-value = 8.958e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   54.88884 129.76659
## sample estimates:
## mean of x mean of y
##  279.7976  187.4699
```

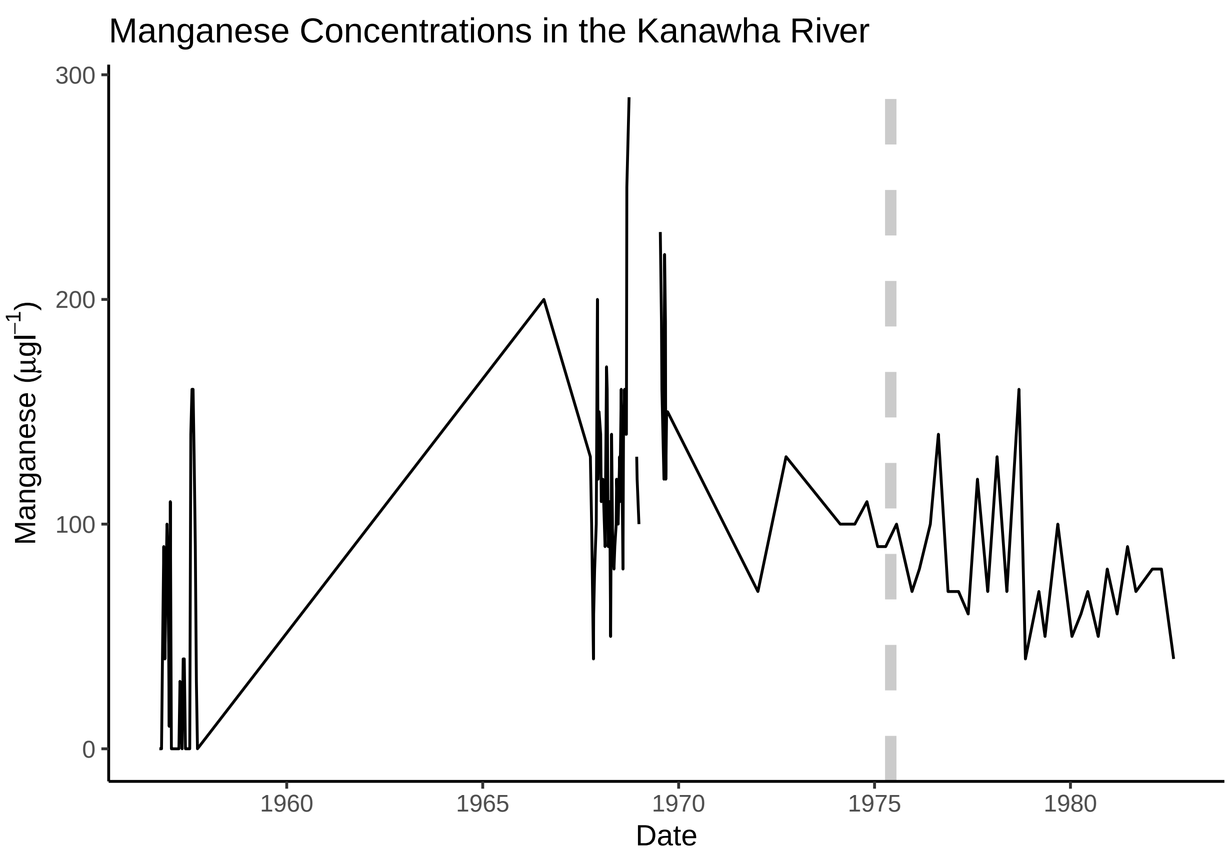
How would you communicate these results to a manager?

What can you say with data visualization that you can't say with just the statistical tests?

How have individual ions changed over the this timeperiod?

Let's look at the Manganese data for the Kanawha river.

```
Manganese.raw <- readNWISqw('03201300', parameterCd = '01055')

Manganese.dat <- Manganese.raw %>%
 select(Date = sample_dt, Manganese_ugL = result_va)


Manganese.plot <- ggplot(Manganese.dat, aes(x = Date, y = Manganese_ugL)) +
  geom_line() +
  geom_vline(xintercept = 1977, color = "grey", alpha = 0.8, lwd = 2, lty = 2) +
  labs(y = expression("Manganese ("*mu*g*l^-1*")")) +
  ggtitle("Manganese Concentrations in the Kanawha River")

print(Manganese.plot)
```



Try plotting the manganese data with points instead of lines. Which do you prefer? Why?

EPA drinking water standards for Manganese are 0.05 mg/L. How does the Kanawha stack up?

Take some time to look at other ions as well. Some suggestions are sulfate (`00945`), nitrate (`71851`), and iron (`71885`), or use the `whatNWISdata()` function to find others! EPA drinking water regulations can be found here:

https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations. Pay close attention to units! Information about parameter codes can be found in the dataframe `parameterCdFile`.

Closing Discussion

Based on the data you've looked at today, what are some of the water quality impacts of mining? What other disturbances might cause similar changes?

References

Water Quality in the Kanawha-New River Basin, West Virginia, Virginia, and North Carolina, 1996–98 (USGS): https://pubs.usgs.gov/circ/circ1204/pdf/section2.pdf

National Primary Drinking Water Regulations (EPA): https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations