8: Water Quality in Rivers

Water Data Analytics | Kateri Salk

Spring 2022

Lesson Objectives

1. Gather river water quality data using the dataRetrieval package
2. Analyze seasonal and flow-related patterns in microbiological and chemical water quality indicators

Session Set Up

Water Quality Data Processing

Microbiological Indicators

Bacteria are present in almost every stream on earth. Most bacteria are harmless, but some are dangerous to human health. In particular, coliform bacteria are present in the gut of mammals and can enter the environment through excrement. Typical large sources of bacterial contamination include wastewater treatment plants, animal manure lagoons, and septic systems. There are multiple ways that bacteria can be measured for water quality.

1. **Total coliforms**. Coliform bacteria are for the most part harmless microorganisms. If only total coliform is detected in the water, an environmental rather than human-derived source is unlikely. The way these are measured are by taking a known volume of water, cultivating it in the lab, and counting the number of colonies that grow. Units for this metric are colony forming units (cfu) per 100 ml.
2. **Fecal coliforms**. Fecal coliforms are coliforms that are specifically associated with warm-blooded animals, and can be grown at higher temperatures than other coliforms. Fecal coliforms are a more precise way of estimating sewage inputs, though they are only an indicator species and not necessarily harmful. The most well known species in this group is Escherichia coli (E. coli). Units are cfu per 100 ml.
3. **Fecal streptocci**. Fecal streptocci are a different type of bacteria that are also associated with human and animal digestive tracts. Streptocci are more widespread than coliforms. Units are cfu per 100 ml.

The EPA has water quality criteria for bacteria for both drinking water and recreation.

Specific Conductance and Freshwater Salinization

Freshwater salinization is an increasingly common phenomenon, caused by salt pollution from anthropogenic sources (road de-icers, irrigation, sewage), accelerated weathering, and increased impervious surface cover. A great paper on this topic comes from Kaushal et al. (2018).

In addition to the concentration of individual ions, one can also measure conductivity of water. Conductivity is the ability of water to conduct an electrical current and is directly proportional to the total concentration of dissolved ions.

Example: Neuse River, NC

The Neuse River has bacterial and chemical indicators of water quality going back through the 1970s. We will explore water quality indicators in this river and the temporal and flow-related sources of variability in those indicators.

Parameters of interest:

00060: flow, cfs 31625: fecal coliform 0.7 micron, CFU/100 mL 31673: fecal streptococci #/100 mL 90095: specific conductance uS/cm 00915: calcium, filtered mg/L 00925: magnesium, filtered mg/L 00930: sodium,filtered mg/L

```r
NeuseParams <- whatNWISdata(siteNumbers = "02089500")
```

```
## Warning: 1 failed to parse.
```

```r
NeuseWQ <- readWQPqw(siteNumbers = "USGS-02089500", # Neuse River at Kinston, NC
                     parameterCd = c("00915", # calcium, filtered mg/L
                                     "00925", # magnesium, filtered mg/L
                                     "00930", # sodium,filtered mg/L
                                     "90095",  # Specific conductance, uS/cm
                                     "31625",  # Fecal coliform, CFU/100 mL
                                     "31673"),  # Fecal streptocci, CFU/100 mL
                     startDate = "1976-01-01",
                     endDate = "")

unique(NeuseWQ$CharacteristicName)
```

```
## [1] "Fecal Coliform"                  "Sodium"
## [3] "Magnesium"                       "Calcium"
## [5] "Fecal Streptococcus Group Bacteria" "Specific conductance"
```

```r
NeuseFlow <- readNWISdv(siteNumbers = "02089500",
                        parameterCd = "00060", # discharge (cfs)
                        startDate = "1976-01-01",
                        endDate = "")

NeuseWQ_processed <- NeuseWQ %>%
  select(MonitoringLocationIdentifier, ActivityStartDate, HydrologicCondition,
         CharacteristicName, ResultMeasureValue) %>%
  mutate(Variable = case_when(CharacteristicName == "Fecal Coliform" ~ "FecalColiform_cfu.100mL",
                              CharacteristicName == "Fecal Streptococcus Group Bacteria" ~ "FecalStrep_
                              CharacteristicName == "Specific conductance" ~ "SpC_uS.cm",
                              CharacteristicName == "Calcium" ~ "Calcium_mg.L",
                              CharacteristicName == "Magnesium" ~ "Magnesium_mg.L",
                              CharacteristicName == "Sodium" ~ "Sodium_mg.L")) %>%
  select(-CharacteristicName) %>%
  pivot_wider(names_from = "Variable", values_from = "ResultMeasureValue")

NeuseFlow_processed <- NeuseFlow %>%
  select(ActivityStartDate = Date, Flow_cfs = X_00060_00003) %>%
  filter(ActivityStartDate >= min(NeuseWQ_processed$ActivityStartDate) &
           ActivityStartDate <= max(NeuseWQ_processed$ActivityStartDate))

NeuseWQFlow_processed <- NeuseWQ_processed %>%
  full_join(., NeuseFlow_processed) %>%
  mutate(Year = year(ActivityStartDate),
         Month = month(ActivityStartDate),
         DOY = yday(ActivityStartDate),)
```

```
## Joining, by = "ActivityStartDate"
```

```
NeuseWQFlow_long <- NeuseWQFlow_processed %>%
  pivot_longer(cols = c("FecalColiform_cfu.100mL", "FecalStrep_cfu.100mL", "SpC_uS.cm",
                        "Sodium_mg.L", "Magnesium_mg.L", "Calcium_mg.L", "Flow_cfs"),
               names_to = "Variable", values_to = "Value") %>%
  drop_na(Value)

NeuseFlow_summary <- NeuseWQFlow_processed %>%
  group_by(DOY) %>%
  summarise(median_Flow_cfs = median(Flow_cfs, na.rm = TRUE),
            p25_Flow_cfs = quantile(Flow_cfs, 0.25, na.rm = TRUE),
            p75_Flow_cfs = quantile(Flow_cfs, 0.75, na.rm = TRUE))
```

Seasonal and flow-related changes in water quality

Period of record time series

```
ggplot(NeuseWQFlow_long, aes(x = ActivityStartDate, y = Value)) +
  geom_line() +
  facet_wrap(vars(Variable), nrow = 7, scales = "free_y") +
  scale_y_log10() +
  labs(x = "Date")
```
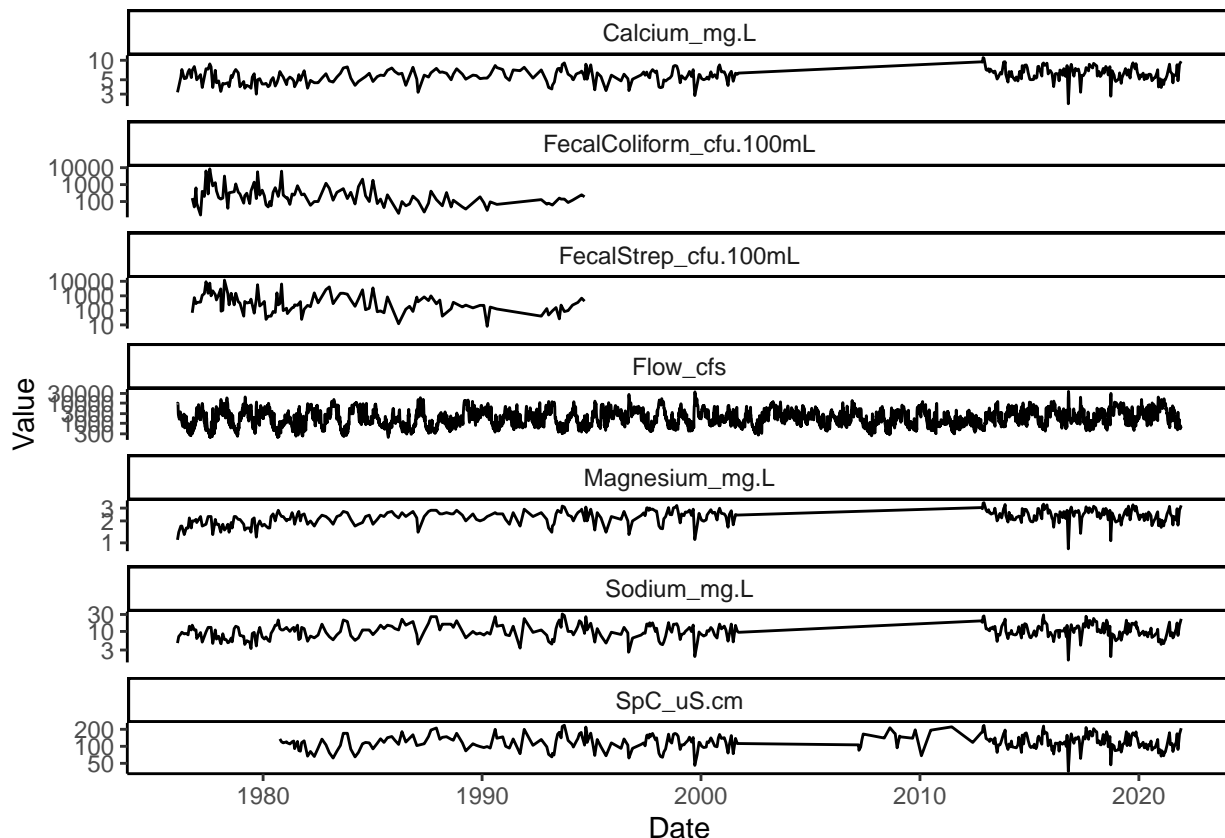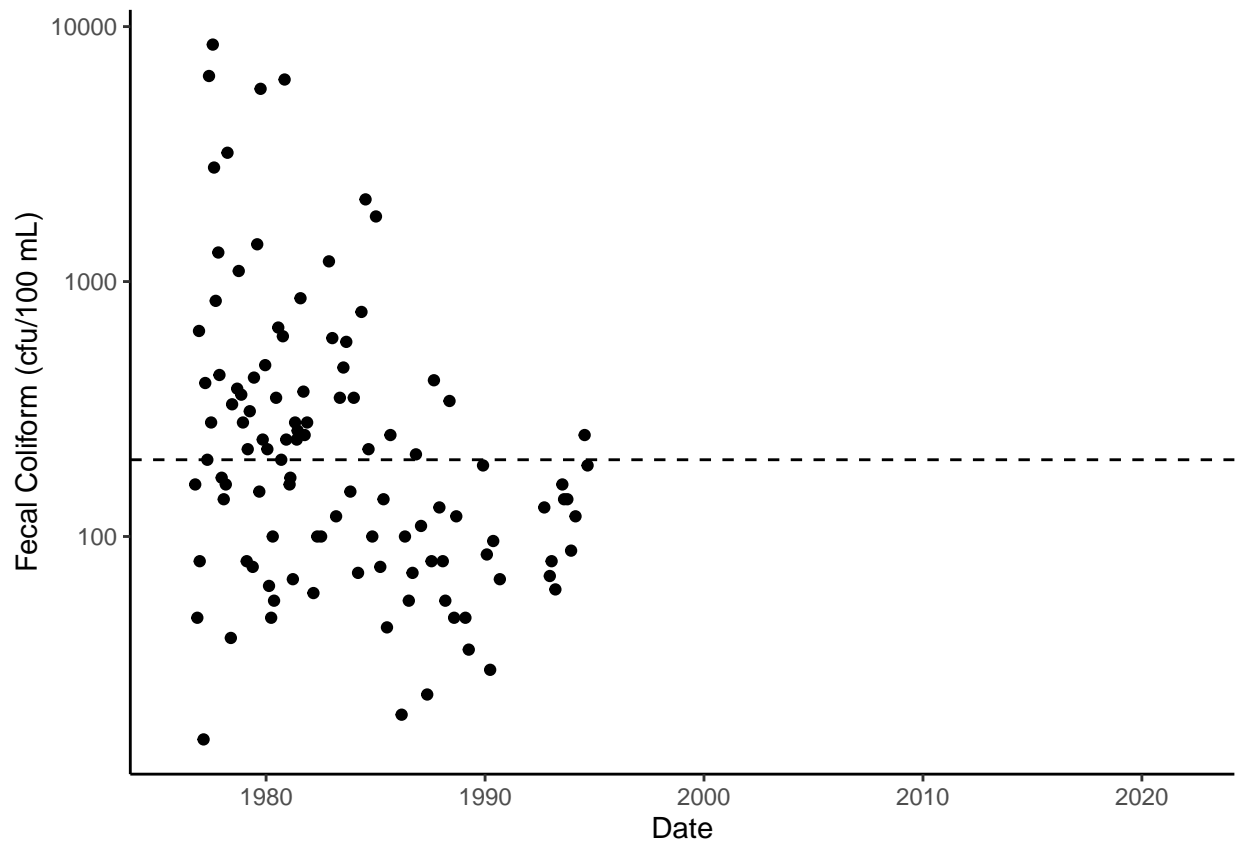


```
ggplot(NeuseWQFlow_processed, aes(x = ActivityStartDate, y = FecalColiform_cfu.100mL)) +
  geom_point() +
  geom_hline(yintercept = 200, lty = 2) + # 1986 EPA recreational criteria for freshwater
  scale_y_log10() +
```
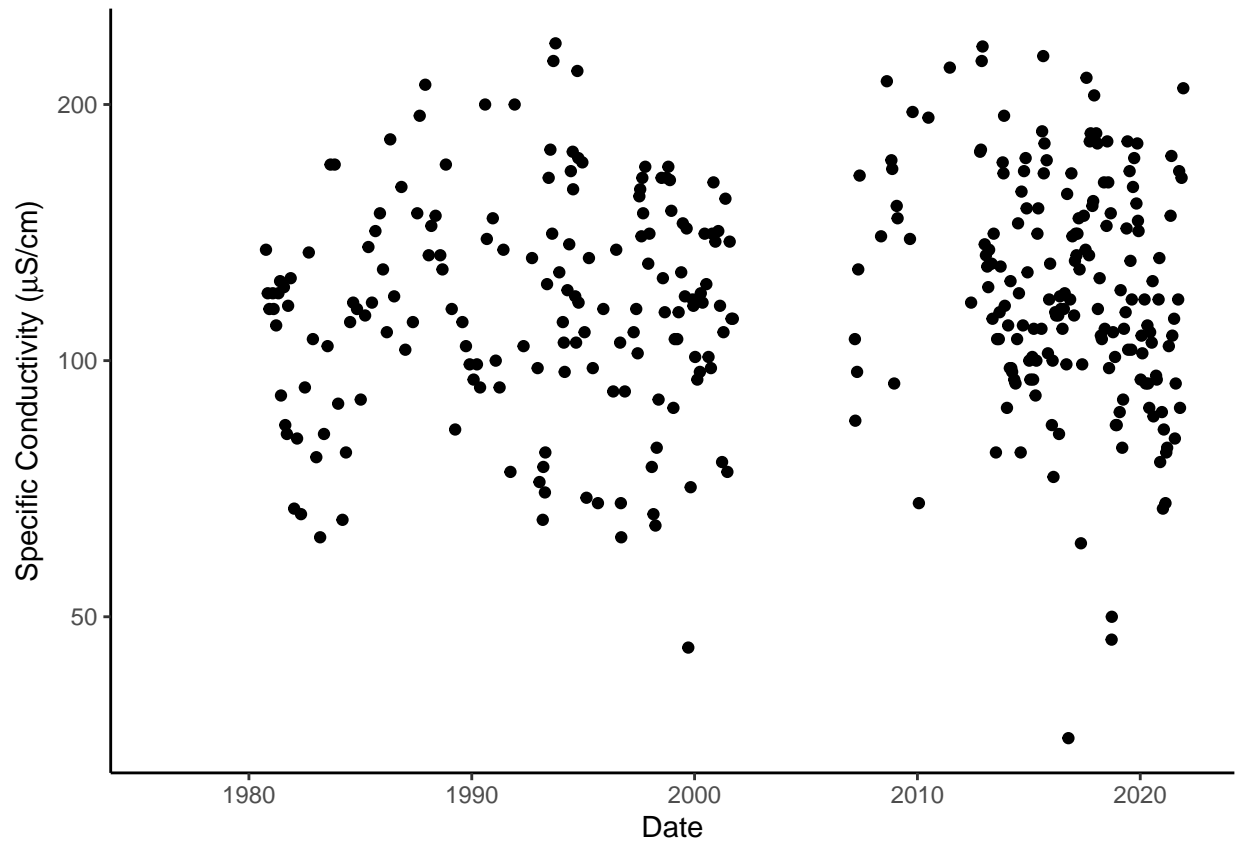
```
labs(x = "Date", y = "Fecal Coliform (cfu/100 mL)")
```

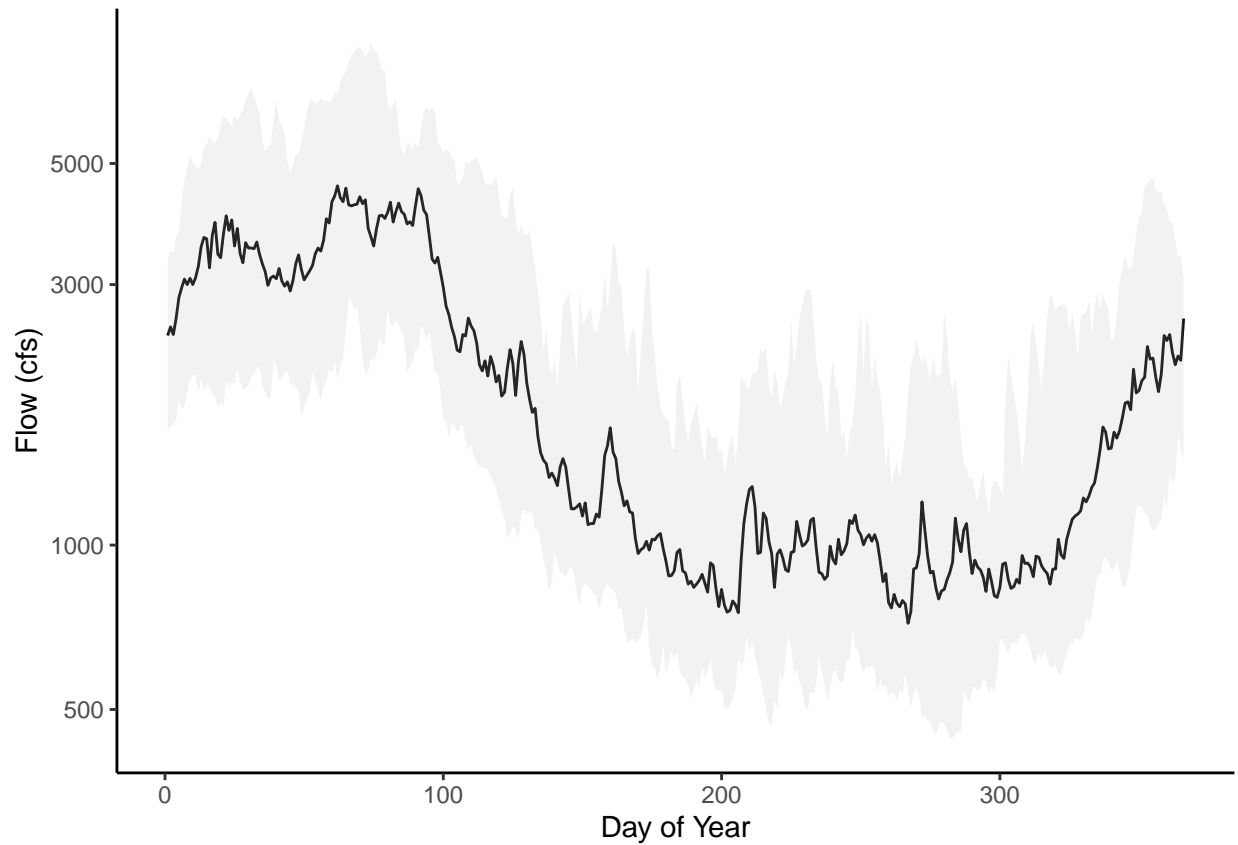## Warning: Removed 16636 rows containing missing values (geom_point).


```
ggplot(NeuseWQFlow_processed, aes(x = ActivityStartDate, y = SpC_uS.cm)) +
  geom_point() +
  scale_y_log10() +
  labs(x = "Date", y = expression("Specific Conductivity ("*mu*"S/cm)"))
```

Warning: Removed 16393 rows containing missing values (geom_point).

What is the relationship between flow and the water quality indicators?

```
ggplot(NeuseFlow_summary, aes(x = DOY)) +
  geom_line(aes(y = median_Flow_cfs)) +
  geom_ribbon(aes(ymin = p25_Flow_cfs, ymax = p75_Flow_cfs), fill = "gray", alpha = 0.2) +
  #geom_point(data = NeuseWQFlow_processed, aes(y = FecalColiform_cfu.100mL)) +
  #geom_point(data = NeuseWQFlow_processed, aes(y = SpC_uS.cm)) +
  scale_y_log10() +
  labs(x = "Day of Year", y = "Flow (cfs)")
```
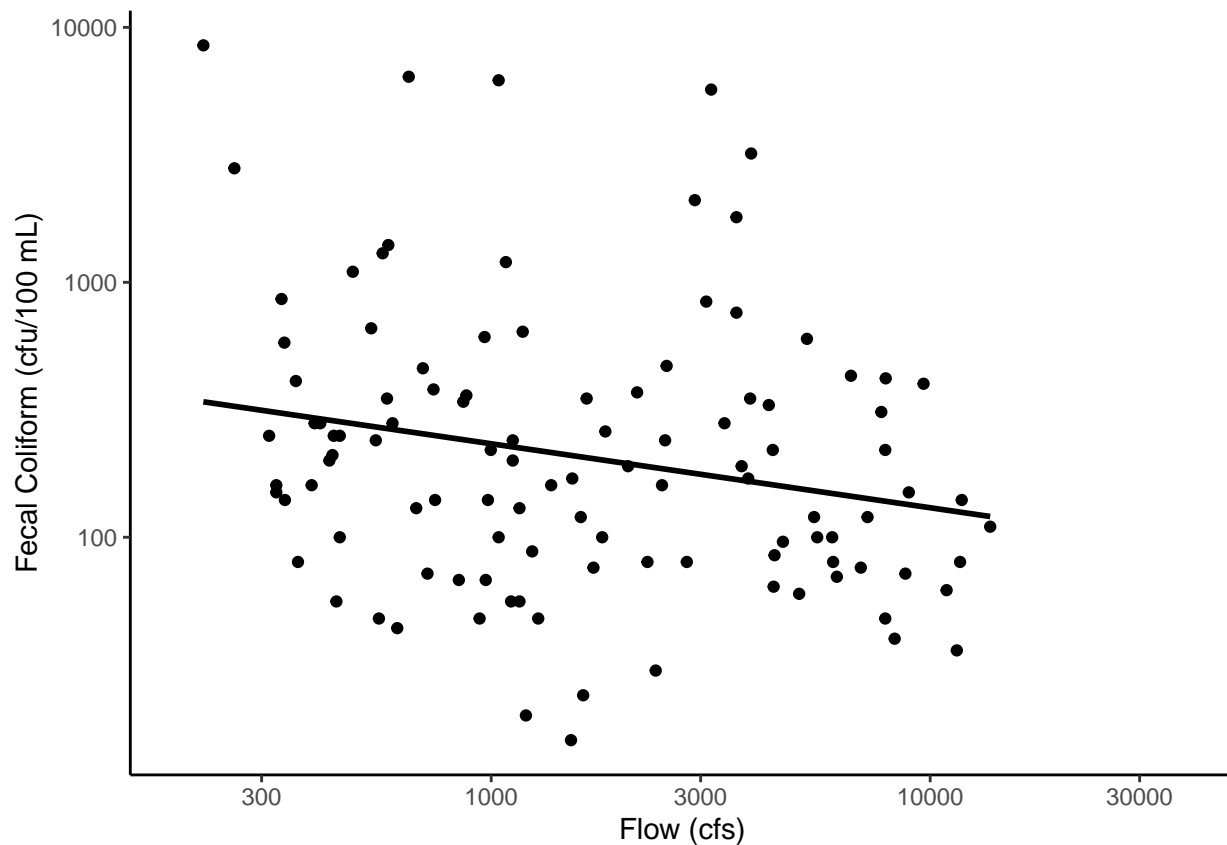
```
ggplot(NeuseWQFlow_processed, aes(x = Flow_cfs, y = FecalColiform_cfu.100mL)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  scale_x_log10() +
  scale_y_log10() +
  labs(x = "Flow (cfs)", y = "Fecal Coliform (cfu/100 mL)")
```

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

Warning: Removed 16636 rows containing non-finite values (stat_smooth).

## Warning: Removed 16636 rows containing missing values (geom_point).

```
summary(lm(data = NeuseWQFlow_processed, log10(FecalColiform_cfu.100mL) ~ log10(Flow_cfs)))
```

```
##
## Call:
## lm(formula = log10(FecalColiform_cfu.100mL) ~ log10(Flow_cfs),
##     data = NeuseWQFlow_processed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11695 -0.32395 -0.05878  0.27845  1.51493
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.1198     0.3465   9.004 9.09e-15 ***
## log10(Flow_cfs)  -0.2510     0.1067  -2.352   0.0205 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5338 on 107 degrees of freedom
##   (16636 observations deleted due to missingness)
## Multiple R-squared:  0.04916,    Adjusted R-squared:  0.04027
## F-statistic: 5.532 on 1 and 107 DF,  p-value: 0.02051
```

```
ggplot(NeuseWQFlow_processed, aes(x = Flow_cfs, y = SpC_uS.cm)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  scale_x_log10() +
```

```
  scale_y_log10() +
  labs(x = "Flow (cfs)", y = expression("Specific Conductivity ("*mu*"S/cm)"))
```

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

Warning: Removed 16393 rows containing non-finite values (stat_smooth).

Warning: Removed 16393 rows containing missing values (geom_point).
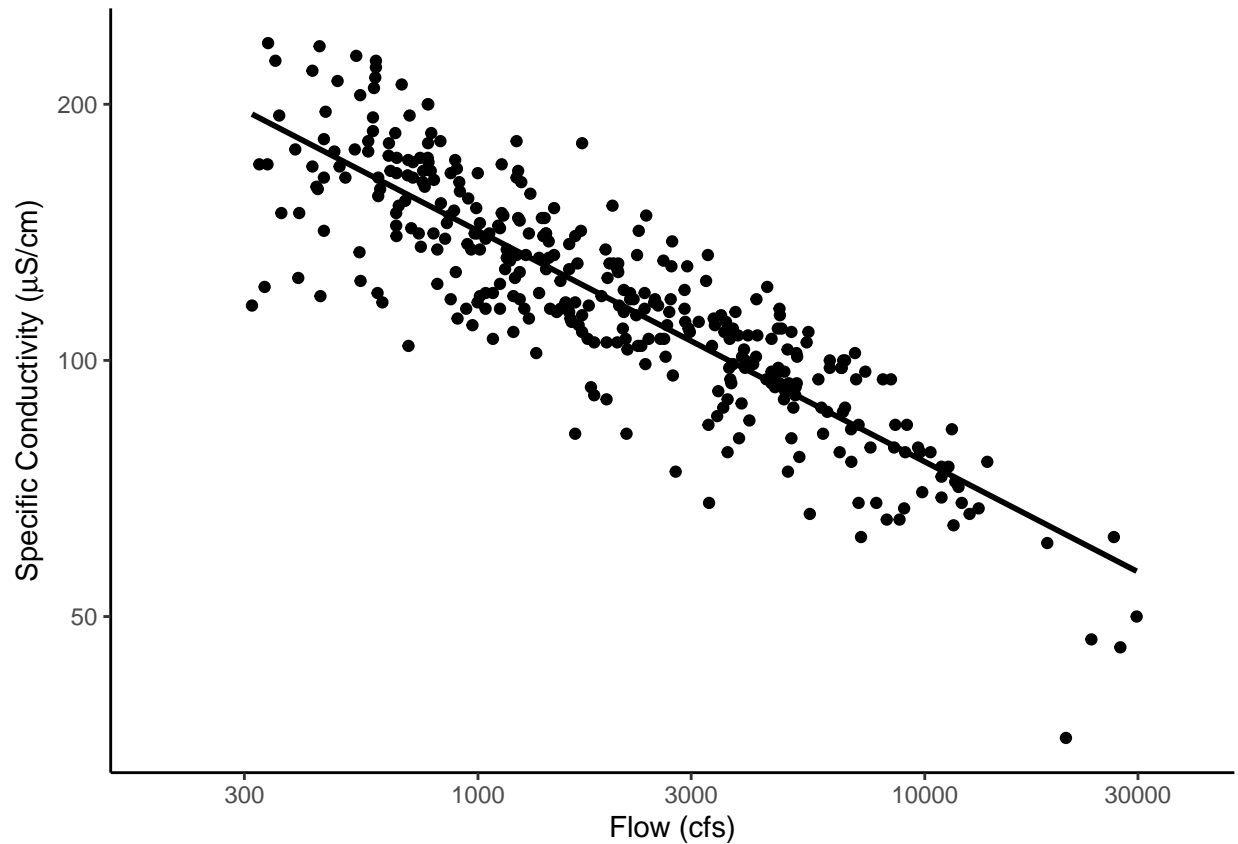

```
summary(lm(data = NeuseWQFlow_processed, log10(SpC_uS.cm) ~ log10(Flow_cfs)))
```

```
##
## Call:
## lm(formula = log10(SpC_uS.cm) ~ log10(Flow_cfs), data = NeuseWQFlow_processed)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.238863 -0.039181  0.007637  0.041361  0.166288
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.966123   0.027420  108.17   <2e-16 ***
## log10(Flow_cfs) -0.271308   0.008251  -32.88   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06591 on 350 degrees of freedom
```

```
##   (16393 observations deleted due to missingness)
## Multiple R-squared:  0.7554, Adjusted R-squared:  0.7547
## F-statistic:  1081 on 1 and 350 DF,  p-value: < 2.2e-16
```

Exercise:

1. Reorder the factor levels of HydrologicCondition in an order that makes hydrologic sense.
2. Create 2 plots that display the ranges of fecal coliform and specific conductivity by hydrologic condition: 2a. Create a boxplot without displaying outliers. 2b. On top of the boxplot, create a `geom_jitter` to show each of the sampled points. Adjust the size, width, and transparency to your liking. 2c. Relabel axes.
3. How do fecal coliform and specific conductivity vary as a function of hydrologic condition?