

Advanced Ecology: Communities and Ecosystems  
PCB5443-U01-1241

Exam 1 - Project EDDIE Chesapeake Bay

Yuri Souza

February 09, 2024

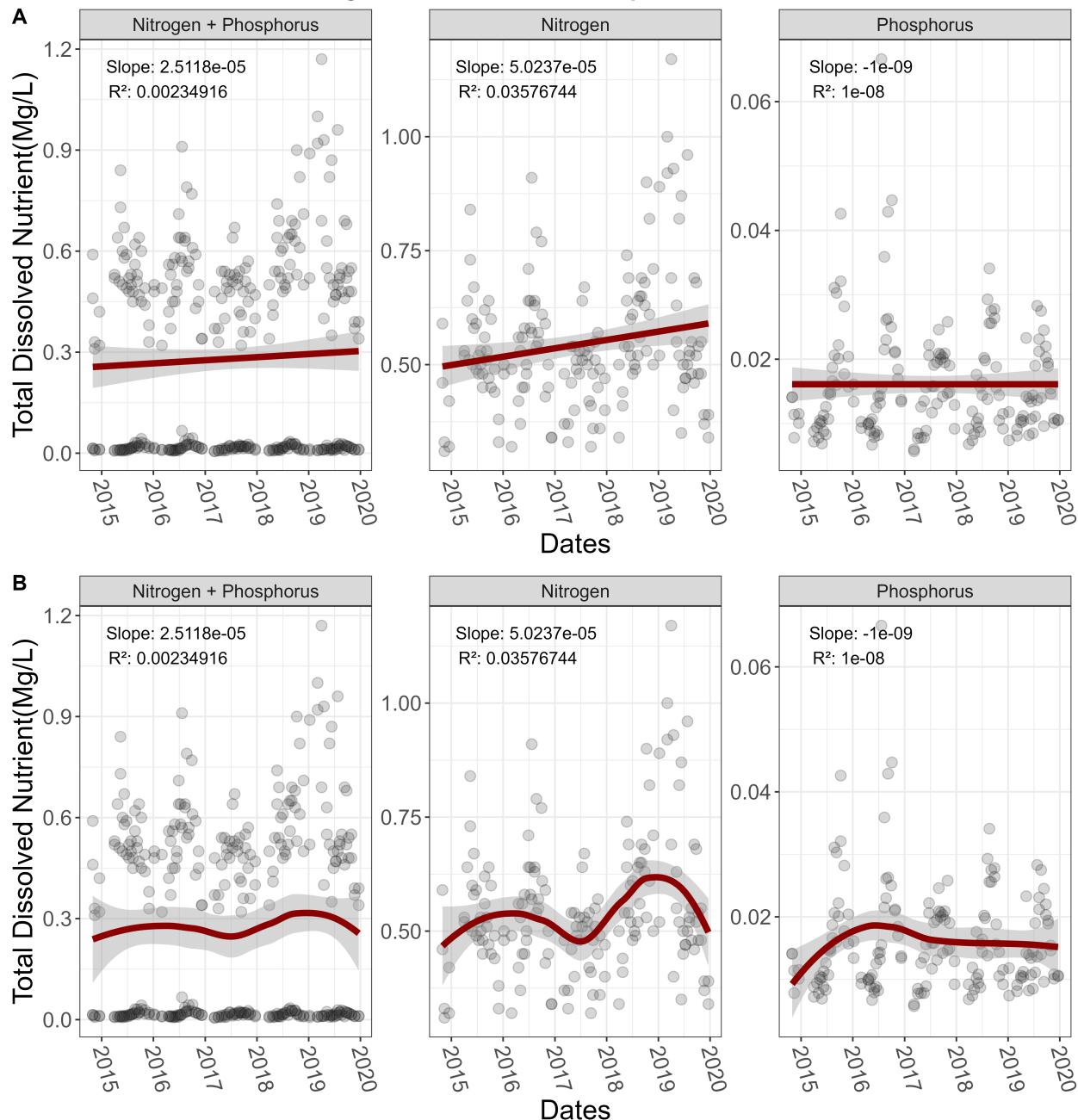
All the data and code are available [here](#)

## Part I: Comparing Nutrient Concentrations in Watersheds (75 points).

### 1) A plot of nutrient concentrations vs time using R or Excel, but try using R (5 points)

**A:** Graphs showing the linear (A) and locally estimated scatterplot smoothing (B) trend of total dissolved nutrient (nitrogen+phosphorus) and nitrogen and phosphorus individually for **Harris Creek-Choptank River (HUC12 20600050506)**.

**Figure 1. Harris Creek-Choptank River**



## 2) A prediction about how you expect nutrient concentrations to change and why (5 points)

**A:** The Choptank River is a tributary of the Chesapeake Bay, which is an estuary ecosystem that goes through a transition between freshwater and saltwater. The bay, as well as the river, is situated and bordered by agricultural and urban areas and also experiences human activities such as boats and cruises moving in and out of the bay.

Under natural stressors and dynamic conditions, over time, one would expect the nutrient concentration to change more according to the ocean's current across the year, which can bring nutrients from the open sea or mix the water that is already present, thus moving nutrients from the bottom of the river to the top. This process is influenced by seasonal and climatic events, but on a larger temporal scale, it would remain natural and balanced.

In the anthropogenic scenario, the proximity of the river to human environments makes it vulnerable to practices like bad agriculture management through soil nutrient fertilization, urban pollution, and sewage outfalls into the river. These activities can lead to an increase in the concentration of nitrogen, phosphorus, and other macronutrients in the water over time. The situation can get worse with the introduction of exotic species due to boat, cruise, and ship activity because these species might assimilate the available nutrients at a higher rate than the native ones, thus leading to negative trophic cascade feedback. This, in turn, can alter the nutrient dynamic and their concentration in the water.

A third prediction is that human activities may alter nutrient concentrations over time, but ecosystem resilience or mitigation efforts would eventually restore them to normal levels.

## 3) The equation of the fitted line, R<sup>2</sup>-value, the slope (5 points)

**A:** The R<sup>2</sup>-value and Slope were both acquired by fitting a *linear model* using the *lm()* function. with the response **variable** being *MeasureValue* and **predictor** the time *SampleDates*. The chunk below shows the function I created to extract only the R<sup>2</sup>-value and Slope from the model since the model output had more information than asked by the question. The function was applied to *total dissolve nutrients* and *individually to nitrogen and phosphorus* (as shown in the figure of question 1). For more details, take a look at the full code included in the last pages of this document.

```
# function to extract the R2-value and Slope form a lm() model

model_coeff <- function(x) {
  model_summary_np <- data_fltd_ttl |>
    dplyr::group_by(model = 1) |>
    dplyr::filter(Parameter == x) |>
    dplyr::do({
      model <- lm(MeasureValue ~ SampleDates, data = .)
      dplyr::tibble(
        slope = broom::tidy(model)$estimate[2],
        r_squared = broom::glance(model)$r.squared
      )
    }) |>
    dplyr::ungroup() |>
    dplyr::select(!1)

  return(model_summary_np)
}
```

**4) Description of Results: How is nutrient concentration changing over time? (5 points)**

**A:** Over time, there has been a slightly positive trend in the *linear fit* of total dissolved nitrogen and phosphorus together. However, the slope of their concentration shows a slight decrease (Slope =  $2.5118^{-05}$ ,  $R^2 = 0.02$ ). It appears that both patterns are driven by nitrogen alone (Slope =  $5.0237^{-05}$ ,  $R^2 = 0.03$ ) rather than phosphorus (Slope =  $-1^{-09}$ ,  $R^2 = 1^{-08}$ ). These inconsistencies among the graphs and the values can probably be better explained by fitting a *locally estimated scatterplot smoothing*, which is a non-parametric and nonlinear fit. This analysis shows that nitrogen has been experiencing years with peaks of increase (2016 and 2019, respectively), followed by a decrease shortly after. On the other hand, phosphorus had an increase after 2015 and remained almost constant until 2020.

Both approaches suggest that nutrient concentration may not vary in a linear fashion, but instead may experience fluctuations over time. However, these changes are minimal and cannot be attributed to the nutrient itself, as evidenced by the very low  $R^2$  value.

**5) Interpretation of Results: What could explain these changes (refer to information learned from lecture and class readings)? (5 points)**

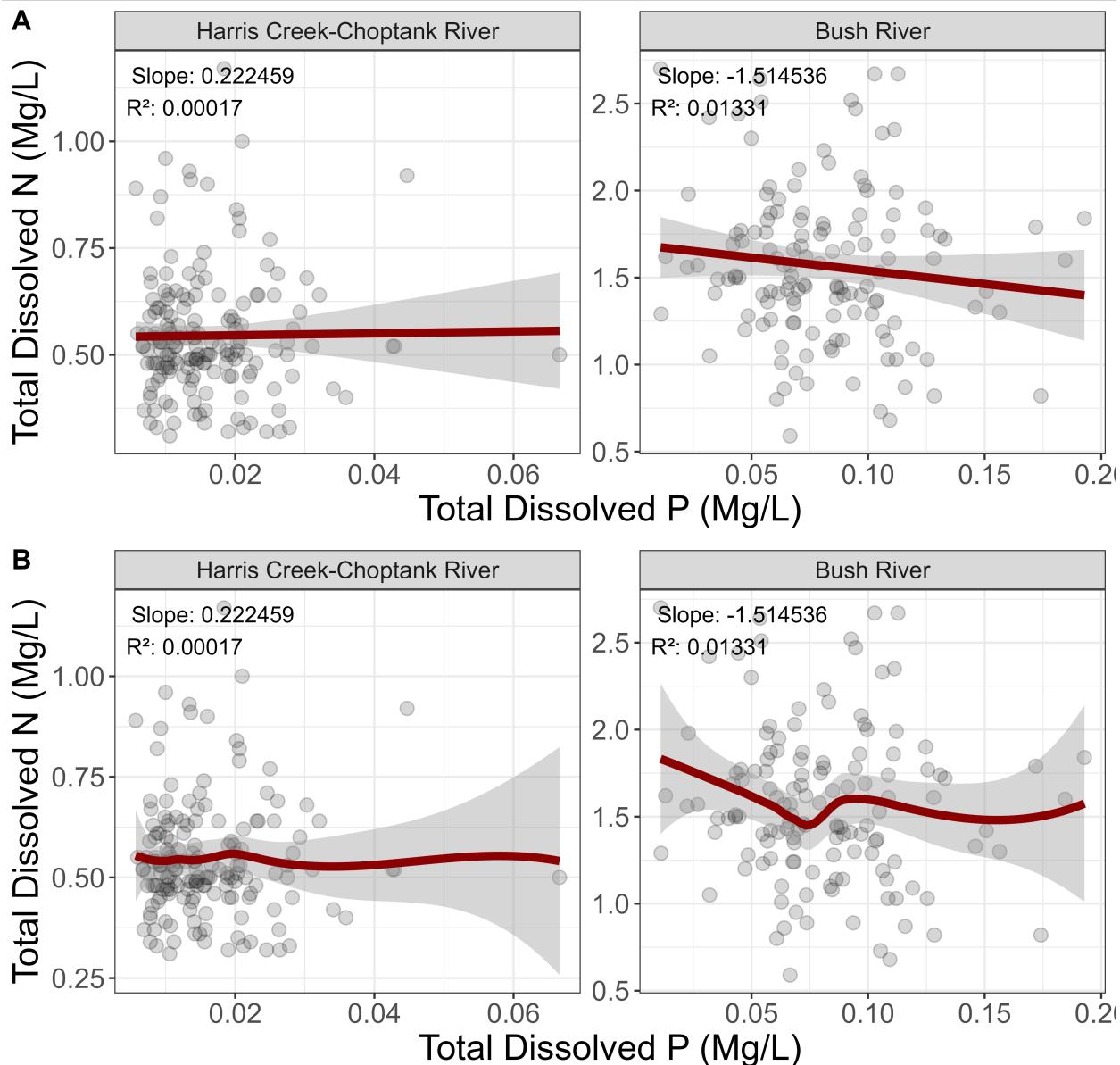
**A:** Harris Creek is situated at the Chesapeake Bay that has been part of a long-term conservation plan aimed at restoring oysters. Various restoration efforts have been implemented in Harris Creek, such as planting oyster spat (young oysters) in designated reefs, constructing artificial reefs, and improving water quality to support oyster populations. These initiatives contribute to the overall health and resilience of the Chesapeake Bay ecosystem, which has been threatened by human activities (Bruce et al., 2021). Although there were no significant change in the total dissolved nutrient concentration based on the *slope* and  $R^2$  values, we do see that nitrogen fluctuates over time (Figure 1B). If we take only nitrogen into consideration, we can rationalize that the re-introduced oysters might potentially play their function in mitigating the excessive nitrogen via trophic interactions and benthic denitrification. Excessive nitrogen loading can stimulate algal growth, which may lead to harmful algal blooms (Bruce et al., 2021). On the other hand, oysters are known to remove nitrogen from the water by consuming microscopic algae and organic matter present in the water. This nitrogen is either stocked in their tissues or released back into the environment in a different molecular form, such as ammonia or organic nitrogen. Additionally, oyster byproducts accumulate in sediment and support microbial communities which further remove nitrogen from water through denitrification (Humphries et al., 2016). In this sense, we can hypothesize that the practices in Chesapeake Bay have been effective in controlling nutrient concentrations (Bruce et al., 2021).

## Part II: Assessing Relative Nitrogen and Phosphorus Limitation among Watersheds (25 points).

- 1) Two plots of N vs P concentrations from two watersheds using R or Excel, but try using R (5 points)

A: Graphs showing the linear (A) and smooth (B) trend of total dissolved Nitrogen vs Phosphorus for Harris Creek-Choptank River (HUC12 20600050506) and Bush River (020600030106).

**Figure 2. Harris Creek-Choptank River & Bush River**



2) A prediction about which nutrient limitation you expect and why (5 points)

3) The equation of the fitted line, R<sup>2</sup>-value, the slope (5 points)

**A:** The R<sup>2</sup>-value and Slope were both acquired by fitting a *linear model* using the *lm()* function. with the response **variable** being the *Total Dissolved Nitrogen* and **predictor** the *Total Dissolved Phosphorus*. The chunk below shows the function I created to extract only the R<sup>2</sup>-value and Slope from the model since the model output had more information than asked by the question. The function was applied to check the relationship of *Total Dissolved Nitrogen* and *Total Dissolved Phosphorus* for both **Harris Creek-Choptank River** and **Bush River** (as shown in the figure of question 2). For more details, take a look at the full code included in the last pages of this document.

```
# function to extract the R2-value and Slope form a lm() model

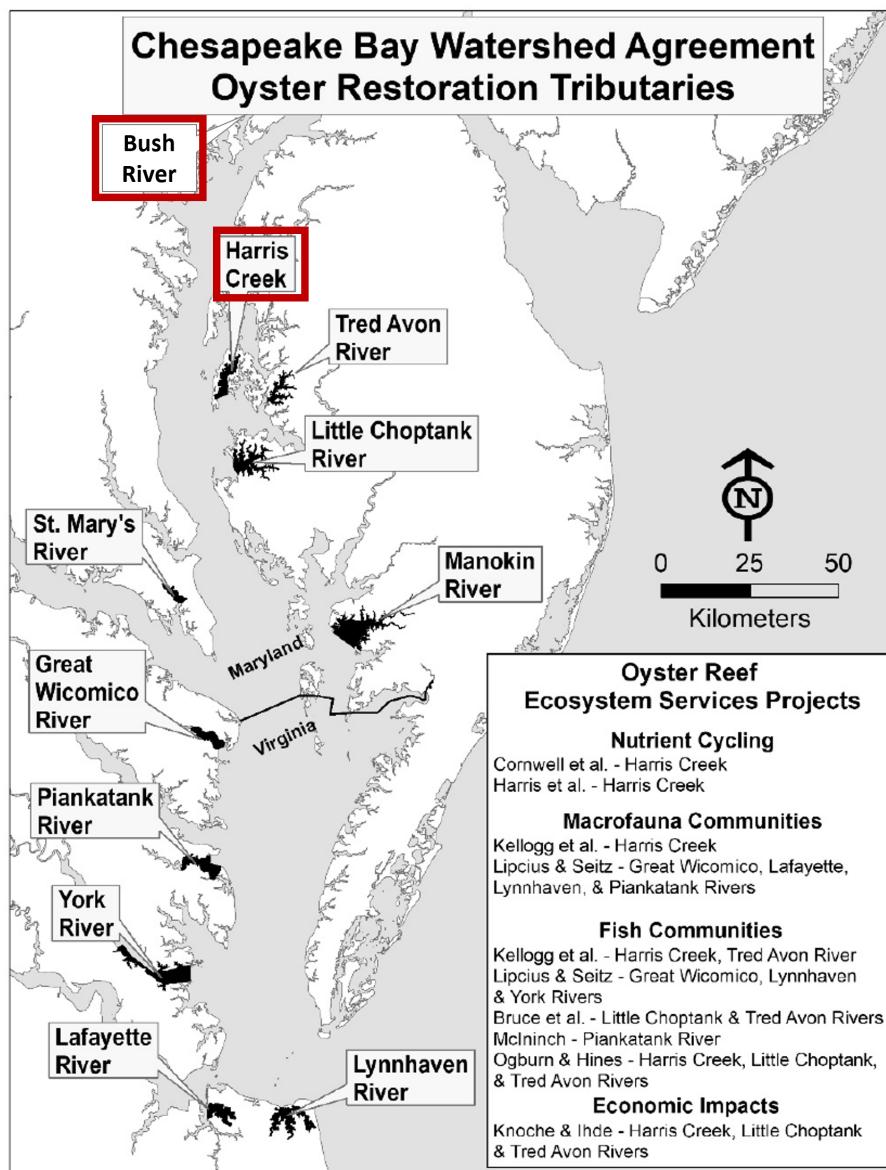
model_coeff <- function(x) {
  model_summary_np <- data_HCCR_BR |>
    dplyr::group_by(model = 1) |>
    dplyr::filter(HUC12 == x) |>
    dplyr::do({
      model <- lm(TDN ~ TDP, data = .)
      dplyr::tibble(
        slope = broom::tidy(model)$estimate[2],
        r_squared = broom::glance(model)$r.squared
      )
    }) |>
    dplyr::ungroup() |>
    dplyr::select(!1)

  return(model_summary_np)
}
```

4) Description of Results: Which nutrient is most limiting in both watersheds? How do you know? (5 points)

5) Interpretation of Results: What could explain the results of N, P, or N and P limitation (refer to information learned from lecture and class readings)? (5 points)

Map (Bruce et al., 2021)



**Figure 3.** Location of Chesapeake Bay tributary-scale oyster restoration activities and ecosystem services research projects.

## Literature

Bruce, D. G., Cornwell, J. C., Harris, L., Ihde, T. F., Lisa, M., Kellogg, S. K., ... & Vogt, B. (2021). A synopsis of research on the ecosystem services provided by Large-scale oyster restoration in the Chesapeake bay. NOAA Tech. Memo. NMFS-OHC, 8, 52.

Humphries, A. T., Ayvazian, S. G., Carey, J. C., Hancock, B. T., Grabbert, S., Cobb, D., ... & Fulweiler, R. W. (2016). Directly measured denitrification reveals oyster aquaculture and restored oyster reefs remove nitrogen at comparable high rates. *Frontiers in Marine Science*, 3, 74.

## Script Part I

```
#data manipulation

if(!require(tidyverse)) install.packages("tidyverse", dependencies = TRUE)
if(!require(ggpubr)) install.packages("ggpubr", dependencies = TRUE)

data <- readr::read_csv("00_data/WaterQualityWaterQualityHUC12_HCCR.csv")

glimpse(data)

data_fltd <- data |>
  dplyr::select(c(SampleDate, Parameter, MeasureValue)) |>
  tidyr::drop_na("MeasureValue") |>
  dplyr::mutate(SampleDates = as.Date(SampleDate, format = "%m/%d/%Y"),
                SampleDate = as.factor(SampleDate))

data_fltd_n_p <- data_fltd |>
  dplyr::mutate(Parameter = "TDNP")

data_fltd_ttl <- rbind(data_fltd, data_fltd_n_p)
data_fltd_ttl

model_coeff <- function(x) {
  model_summary_np <- data_fltd_ttl |>
    dplyr::group_by(model = 1) |>
    dplyr::filter(Parameter == x) |>
    dplyr::do({
      model <- lm(MeasureValue ~ SampleDates, data = .)
      dplyr::tibble(
        slope = broom::tidy(model)$estimate[2],
        r_squared = broom::glance(model)$r.squared
      )
    }) |>
    dplyr::ungroup() |>
    dplyr::select(!1)

  return(model_summary_np)
}

nt.model.coeff <- bind_rows(model_coeff("TDNP"),
                            model_coeff("TDN"),
                            model_coeff("TDP")) |>
  dplyr::mutate(Parameter = c("TDNP", "TDN", "TDP"),
                date_label = "2015-05-14") |>
  dplyr::select(c(3, dplyr::everything()))
nt.model.coeff
```

```

# graphs

lm_model_1pt <- ggplot(data_fltd_ttl, aes(y=MeasureValue, x=SampleDates)) +
  geom_point(shape = 21, color = "black", fill = "grey20", size = 3, alpha = 0.5) +
  geom_smooth(method="lm", col="darkred", linewidth = 2) +
  theme_bw() +
  aes(ymax=Inf) +
  labs(x = "Dates", y = "Total Dissolved Nutrient(Mg/L)") +
  scale_x_date(date_breaks = "year", date_labels = "%Y") +
  theme(axis.text.x= element_text(angle=-75, size = 14, vjust = 0.2),
        axis.text.y= element_text(size = 14),
        axis.title.x = element_text(size = 18),
        axis.title.y = element_text(size = 18),
        strip.text = element_text(size = 12)) +
  facet_wrap(~factor(Parameter, c("TDNP", "TDN", "TDP")),
             labels = c("Nitrogen + Phosphorus", "Nitrogen", "Phosphorus")),
  scales = "free") +
  geom_text(data = nt.model.coeff, aes(label = paste("Slope:", round(slope, 9),
                                              "\nR²:", round(r_squared, 8)),
                                         x = as.Date(date_label), y = -Inf),
            hjust = 0.1, vjust = -10, size = 4, color = "black")

loess_model_1pt <- ggplot(data_fltd_ttl, aes(y=MeasureValue, x=SampleDates)) +
  geom_point(shape = 21, color = "black", fill = "grey20", size = 3, alpha = 0.5) +
  geom_smooth(method="loess", col="darkred", linewidth = 2) +
  theme_bw() +
  aes(ymax=Inf) +
  labs(x = "Dates", y = "Total Dissolved Nutrient(Mg/L)") +
  scale_x_date(date_breaks = "year", date_labels = "%Y") +
  theme(axis.text.x= element_text(angle=-75, size = 14, vjust = 0.2),
        axis.text.y= element_text(size = 14),
        axis.title.x = element_text(size = 18),
        axis.title.y = element_text(size = 18),
        strip.text = element_text(size = 12)) +
  facet_wrap(~factor(Parameter, c("TDNP", "TDN", "TDP")),
             labels = c("Nitrogen + Phosphorus", "Nitrogen", "Phosphorus")),
  scales = "free") +
  geom_text(data = nt.model.coeff, aes(label = paste("Slope:", round(slope, 9),
                                              "\nR²:", round(r_squared, 8)),
                                         x = as.Date(date_label), y = -Inf),
            hjust = 0.1, vjust = -10, size = 4, color = "black")

my_plot_list <- list(lm_model_1pt,loess_model_1pt)

ggpubr::ggarrange(plotlist = my_plot_list, labels = c("A", "B"), nrow = 2) |>
  annotate_figure(top = text_grob("Harris Creek-Choptank River",
                                 color = "black", face = "bold", size = 14))

```

## Script Part II

```
#data manipulation

if(!require(tidyverse)) install.packages("tidyverse", dependencies = TRUE)
if(!require(ggpubr)) install.packages("ggpubr", dependencies = TRUE)

data_HCCR <- readr::read_csv("00_data/WaterQualityWaterQualityHUC12_HCCR.csv") |>
  dplyr::select(c(1,18,20)) |>
  tidyr::drop_na("MeasureValue")
data_HCCR

data_BR <- readr::read_csv("00_data/WaterQualityWaterQualityHUC12_BR.csv") |>
  dplyr::select(c(1,18,20)) |>
  tidyr::drop_na("MeasureValue")
data_BR

data_HCCR_BR <- rbind(data_HCCR, data_BR) |>
  dplyr::group_by(Parameter) %>%
  dplyr::mutate(Row = row_number()) %>%
  tidyr::pivot_wider(names_from = Parameter, values_from = MeasureValue) |>
  dplyr::select(-Row) |>
  dplyr::ungroup() |>
  dplyr::mutate(HUC12 = stringr::str_replace(HUC12, "20600050506", "HCCR"),
               HUC12 = stringr::str_replace(HUC12, "020600030106", "BR"))
data_HCCR_BR

model_coeff_HCCR_BR <- function(x) {
  model_summary_np_HCCR_BR <- data_HCCR_BR |>
    dplyr::group_by(model = 1) |>
    dplyr::filter(HUC12 == x) |>
    dplyr::do({
      model <- lm(TDN ~ TDP, data = .)
      dplyr::tibble(
        slope = broom::tidy(model)$estimate[2],
        r_squared = broom::glance(model)$r.squared
      )
    }) |>
    dplyr::ungroup() |>
    dplyr::select(!1)

  return(model_summary_np_HCCR_BR)
}

model_coeff_HCCR_BR("HCCR")
model_coeff_HCCR_BR("BR")

nt.model.coeff_HCCR_BR <- bind_rows(model_coeff_HCCR_BR("HCCR"),
                                      model_coeff_HCCR_BR("BR")) |>
  dplyr::mutate(HUC12 = c("HCCR", "BR"))
nt.model.coeff_HCCR_BR
```

```

# graphs

lm_model_2pt <- ggplot(data_HCCR_BR, aes(y=TDN, x=TDP)) +
  geom_point(shape = 21, color = "black", fill = "grey20", size = 3, alpha = 0.2) +
  geom_smooth(method="lm", col="darkred", linewidth = 2) +
  theme_bw() +
  #aes(ymax=Inf) +
  labs(x = "Total Dissolved P (Mg/L)", y = "Total Dissolved N (Mg/L)") +
  theme(axis.text.x= element_text(size = 14),
        axis.text.y= element_text(size = 14),
        axis.title.x = element_text(size = 18),
        axis.title.y = element_text(size = 18),
        strip.text = element_text(size = 12)) +
  facet_wrap(~factor(HUC12, c("HCCR", "BR")),
             labels = c("Harris Creek-Choptank River", "Bush River")),
  scales = "free") +
  geom_text(data = nt.model.coeff_HCCR_BR,
            aes(label = paste("Slope:", round(slope, 6), "\nR²:", round(r_squared, 5)),
                x = -Inf, y = -Inf),
            hjust = -0.1, vjust = -7.5, size = 4, color = "black")
lm_model_2pt

loess_model_2pt <- ggplot(data_HCCR_BR, aes(y=TDN, x=TDP)) +
  geom_point(shape = 21, color = "black", fill = "grey20", size = 3, alpha = 0.2) +
  geom_smooth(method="loess", col="darkred", linewidth = 2) +
  theme_bw() +
  aes(ymax=Inf) +
  labs(x = "Total Dissolved P (Mg/L)", y = "Total Dissolved N (Mg/L)") +
  theme(axis.text.x= element_text(size = 14),
        axis.text.y= element_text(size = 14),
        axis.title.x = element_text(size = 18),
        axis.title.y = element_text(size = 18),
        strip.text = element_text(size = 12)) +
  facet_wrap(~factor(HUC12, c("HCCR", "BR")),
             labels = c("Harris Creek-Choptank River", "Bush River")),
  scales = "free") +
  geom_text(data = nt.model.coeff_HCCR_BR,
            aes(label = paste("Slope:", round(slope, 6), "\nR²:", round(r_squared, 5)),
                x = -Inf, y = -Inf),
            hjust = -0.1, vjust = -7.5, size = 4, color = "black")

my_plot_list_pt2 <- list(lm_model_2pt, loess_model_2pt)

ggpubr::ggarrange(plotlist = my_plot_list_pt2, labels = c("A", "B"), nrow = 2) |>
  annotate_figure(top = text_grob("Figure 2. Harris Creek-Choptank River & Bush River",
                                 color = "black", face = "bold", size = 16))

```