

Assignment 7: High Frequency Data

Yixin Wen

OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis on high frequency data

Directions

1. Change “Student Name” on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single pdf file.
5. After Knitting, submit the completed exercise (pdf file) to the dropbox in Sakai. Add your last name into the file name (e.g., “A07_Chamberlin.pdf”) prior to submission.

The completed exercise is due on 16 October 2019 at 9:00 am.

Setup

1. Verify your working directory is set to the R project file,
2. Load the StreamPULSE, streamMetabolizer and tidyverse packages.
3. Set your ggplot theme (can be theme_classic or something else)

```
getwd()
```

```
## [1] "/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Assignment7"
```

```
packages <- c(
  "dataRetrieval",
  "tidyverse",
  "xts",
  "dygraphs",
  "StreamPULSE",
  "streamMetabolizer"
)
invisible(
  suppressPackageStartupMessages(
    lapply(packages, library, character.only = TRUE)
  )
)
theme_set(theme_classic())
```

4. Download data from the Stream Pulse portal using `request_data()` for the Kansas River, (“KS_KANSASR”). Download the discharge (`Discharge_m3s`), dissolved oxygen (`DO_mgL`) and nitrate data (`Nitrate_mgL`) for the entire period of record
5. Reformat the data into one dataframe with columns `DateTime_UTC`, `DateTime_Solar` (using `convert_UTC_to_solartime()`), `SiteName`, `DO_mgL`, `Discharge_m3s`, and `Nitrate_mgL`.

```
Kandat <- request_data(
  sitecode = "KS_KANSASR",
  variables = c('Discharge_m3s', 'DO_mgL', 'Nitrate_mgL'))
```

```
## You may omit the "variables" parameter to automatically retrieve
```

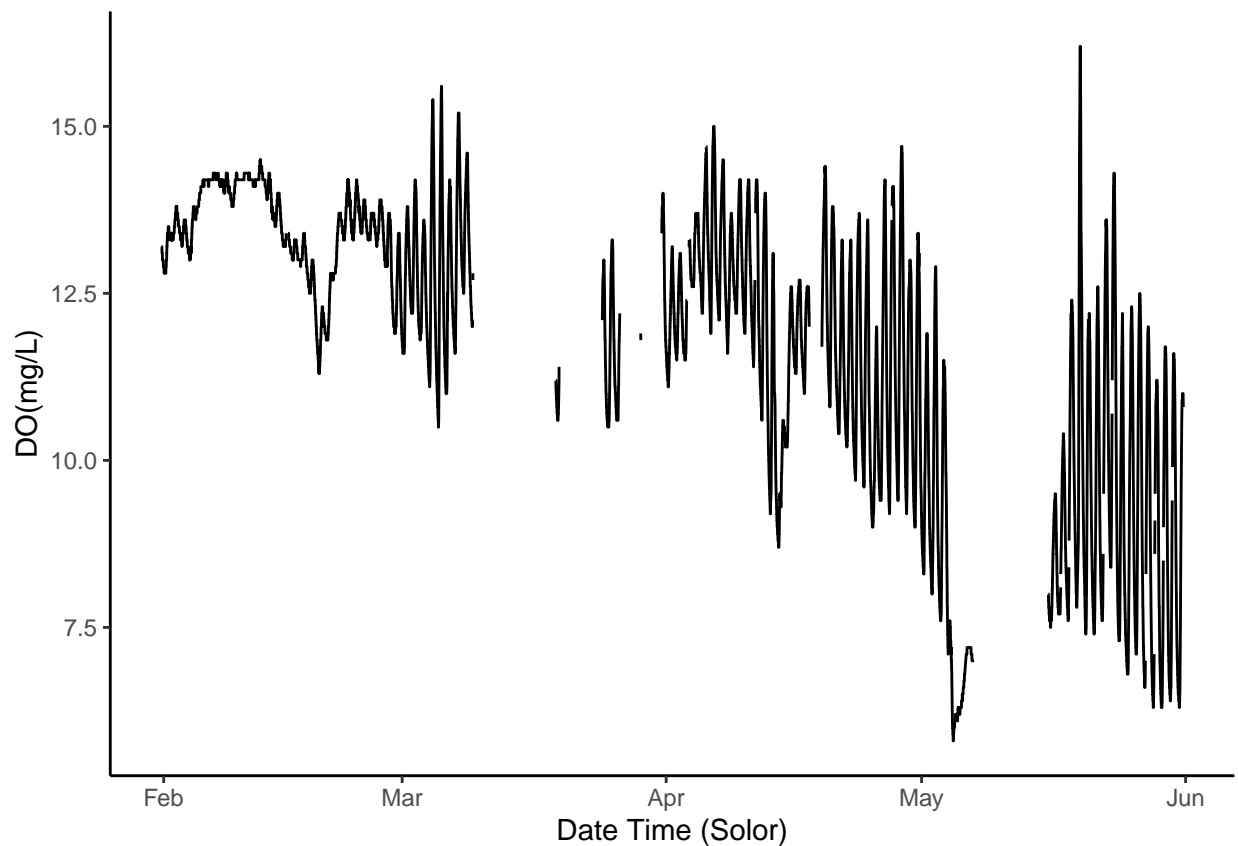
```
## all variables necessary for metabolism modeling.
##
## API call: https://data.streampulse.org/api?sitecode=KS_KANSASR&variables=Discharge_m3s,DO_mgL,Nitrat
##
## Retrieved the following variables:
## DO_mgL, Discharge_m3s, Nitrate_mgL
Kan.lon <- Kandat[[2]]$lon

Kan.para <-
  Kandat[[1]] %>%
  spread(value = value, key = variable)%>%
  mutate(DateTime_Solar = convert_utc_to_solartime(DateTime_UTC, Kan.lon))

Kan.para$flagtype <- NULL
Kan.para$flagcomment <- NULL
```

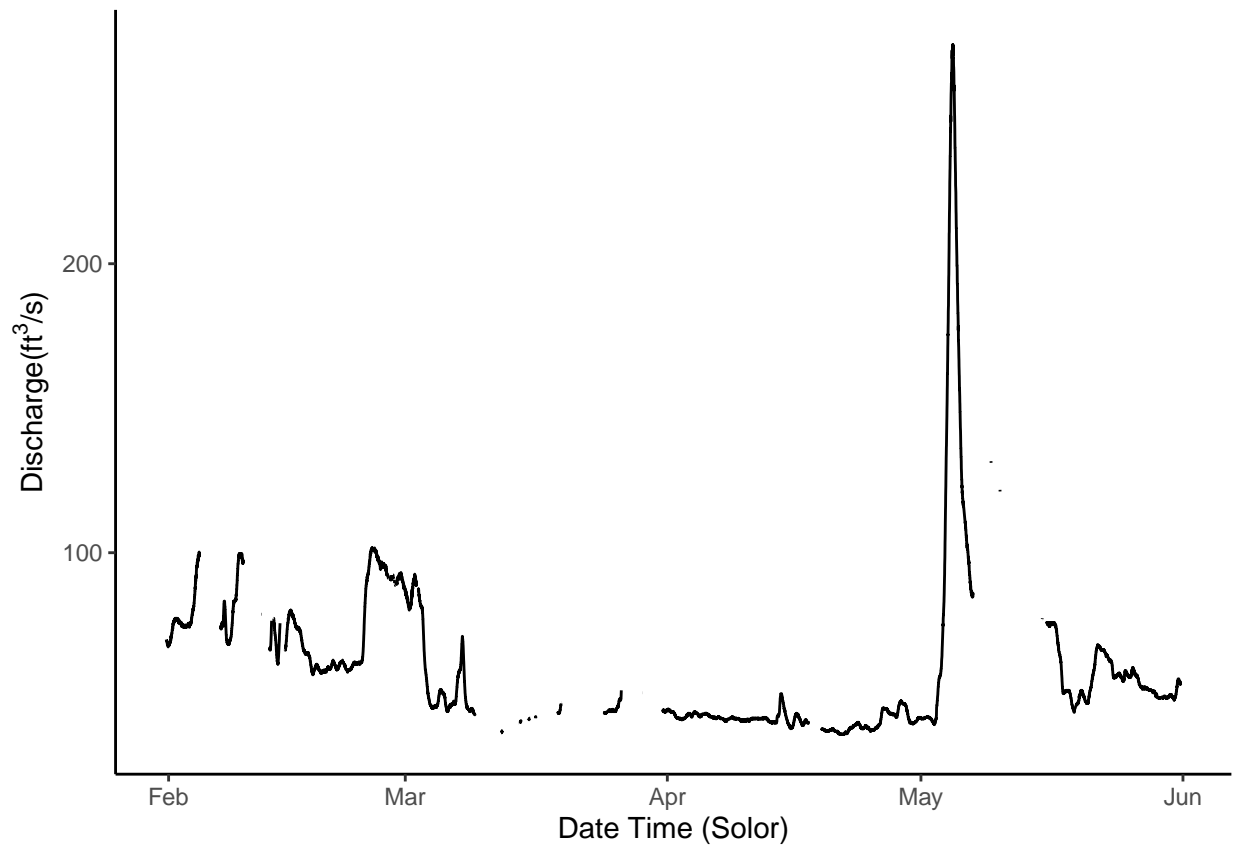
6. Plot each of the 3 variables against solar time for the period of record

```
Kan_DO.plot<-
  ggplot(Kan.para, aes(x = DateTime_Solar, y = DO_mgL))+
  geom_line()+
  labs(x = "Date Time (Solar)", y = "DO(mg/L)")
print(Kan_DO.plot)
```

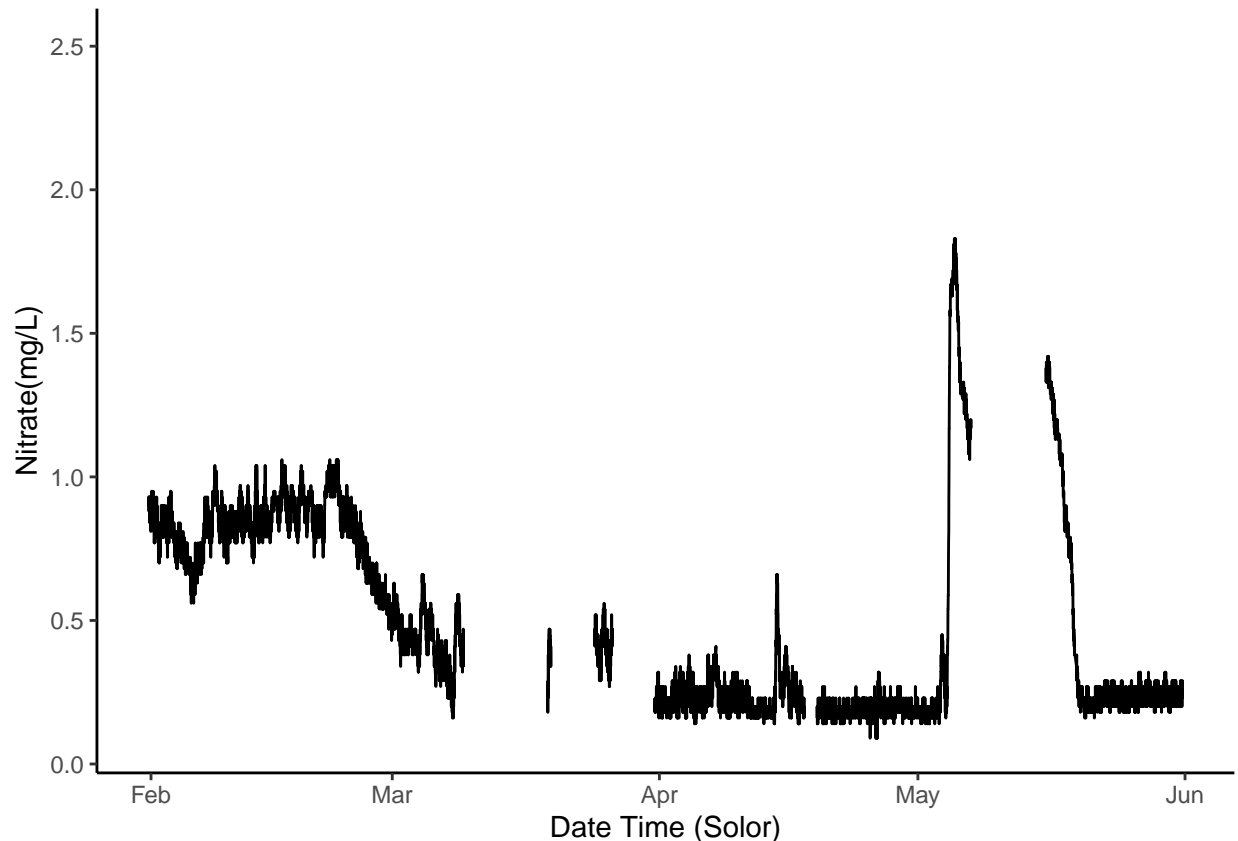


```
Kan_Discharge.plot<-
  ggplot(Kan.para, aes(x = DateTime_Solar, y = Discharge_m3s))+
```

```
geom_line()+
  labs(x = "Date Time (Solor)", y = expression("Discharge(ft\"^3\"/s)"))
print(Kan_Discharge.plot)
```



```
Kan_Nitrate.plot<-
  ggplot(Kan.para, aes(x = DateTime_Solar, y = Nitrate_mgL))+
  geom_line()+
  labs(x = "Date Time (Solor)", y = "Nitrate(mg/L)")
print(Kan_Nitrate.plot)
```



7. How will you address gaps in these dataserries?

I will use linear interplotation or K nearest neighbors to fill the gaps

8. How does the daily amplitude of oxygen concentration swings change over the season? What might cause this?

The range of oxygen concentration is smaller in winter and spring, while it is much wider in summer. The reason for it is the temperature is higher in summer, which may be beneficial to growth of microbes like algae, the respiration process can consume the oxygen in water, and thus decreases the DO value especially at night. At day time, since there's more light in summer, the photosynthesis can make up the loss of oxygen caused by respiration. Thus, the range of daily oxygen concentration is larger than that in winter.

Baseflow separation

9. Use the `EcoHydRology::BaseflowSeparation()` function to partition discharge into baseflow and quickflow, and calculate how much water was exported as baseflow and quickflow for this time period. Use the `DateTime_UTC` column as your timestamps in this analysis.

The `package::function()` notation being asked here is a way to call a function without loading the library. Sometimes the `EcoHydRology` package can mask tidyverse functions like pipes, which will cause problems for knitting. In your script, instead of just typing `BaseflowSeparation()`, you will need to include the package and two colons as well.

10. Create a ggplot showing total flow, baseflow, and quickflow together.

```
Kan.Discharge <- Kan.para%>%
  select("DateTime_UTC", "Discharge_m3s")%>%
```

```

drop_na()

Kan_baseflow <- EcoHydRology::BaseflowSeparation(Kan.Discharge$Discharge_m3s,
  filter_parameter = 0.925,
  passes = 3)

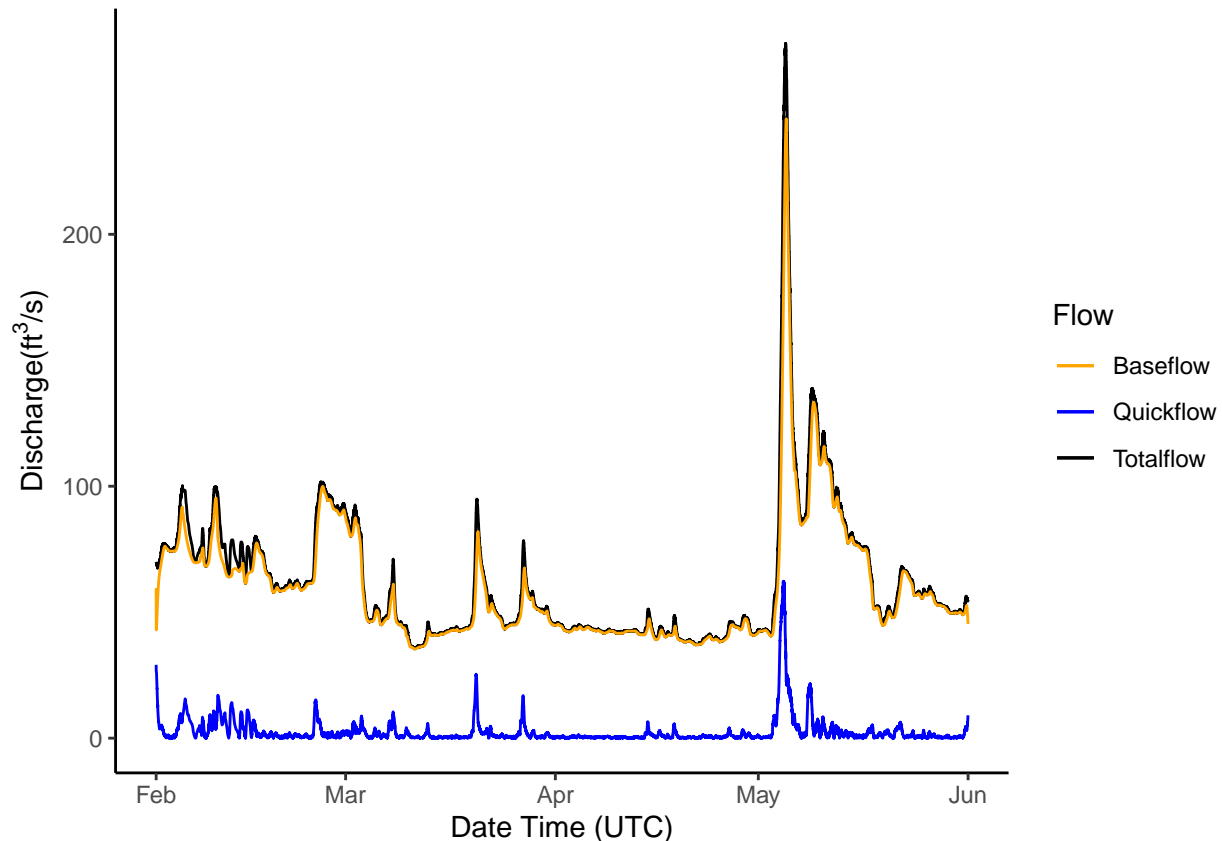
Kan_BQT <- cbind(Kan.Discharge, Kan_baseflow)
names(Kan_BQT)[2:4] <- c("Totalflow", "Baseflow", "Quickflow")

Kan_Export <- Kan_BQT %>%
  mutate(timestep = c(diff(as.numeric(DateTime.UTC))), NA_real_),
  baseflowexport = Baseflow * timestep,
  quickflowexport = Quickflow * timestep) %>%
  summarize(BaseflowExport_cf = sum(baseflowexport, na.rm = T),
    QuickflowExport_cf = sum(quickflowexport, na.rm = T),
    TotalExport_cf = BaseflowExport_cf + QuickflowExport_cf)

Baseflow_percentage = Kan_Export$BaseflowExport_cf / Kan_Export$TotalExport_cf
Quickflow_percentage = Kan_Export$QuickflowExport_cf / Kan_Export$TotalExport_cf

Kan_flow.plot <-
  ggplot(Kan_BQT)+
  geom_line(mapping = aes(x = DateTime.UTC, y = Totalflow, color = "Totalflow"))+
  geom_line(mapping = aes(x = DateTime.UTC, y = Baseflow, color = "Baseflow"))+
  geom_line(mapping = aes(x = DateTime.UTC, y = Quickflow, color = "Quickflow"))+
  scale_color_manual(values = c("Totalflow" = "black",
    "Baseflow" = "orange",
    "Quickflow" = "blue"))+
  labs( x = "Date Time (UTC)", y = expression("Discharge(ft"^3"/s)"), color = "Flow")
print(Kan_flow.plot)

```



11. What percentage of total water exported left as baseflow and quickflow from the Kansas River over this time period?

The baseflow percentage is around 95.7%, the quickflow percentage is around 4.3%.

12. This is a much larger river and watershed than the 2 we investigated in class. How does the size of the watershed impact how flow is partitioned into quickflow and baseflow?

Compared to the 2 rivers which were investigated in class, the larger the watershed is, the bigger the baseflow fraction is and the smaller the quickflow fraction is.

13. The site we are looking at is also further down in its river network (i.e. instead of being a headwater stream, this river has multiple tributaries that flow into it). How does this impact your interpretation of your results?

Tributaries may increase the baseflow contribution in total flow since tributaries are mainly baseflow.

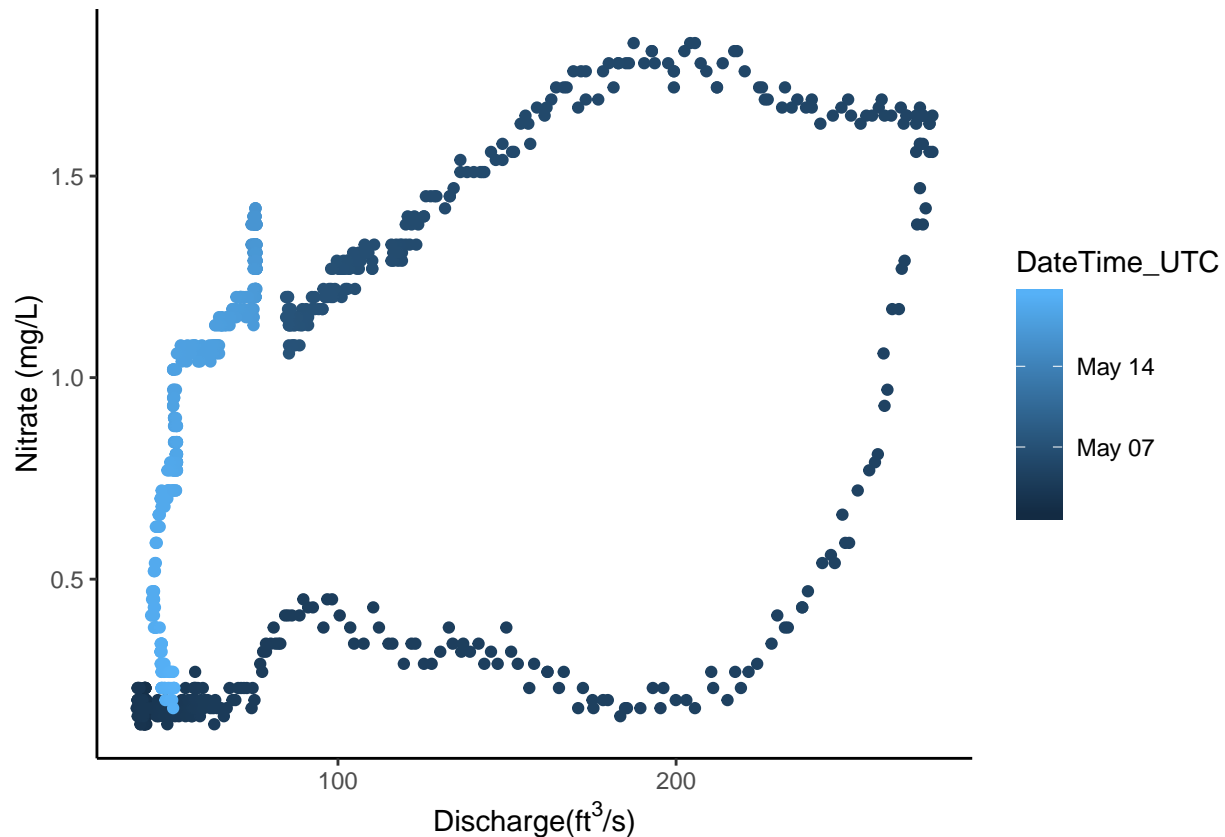
Chemical Hysteresis

14. Create a ggplot of flow vs. nitrate for the large storm in May (~May 1 - May 20). Use color to represent Date and Time.

```
KanStorm <- Kan.para %>%
  select("DateTime_UTC", "Discharge_m3s", "Nitrate_mgL") %>%
  filter(DateTime_UTC > "2018-05-01" & DateTime_UTC < "2018-05-20") %>%
  drop_na()

ggplot(KanStorm, aes(x = Discharge_m3s, y = Nitrate_mgL, color = DateTime_UTC)) +
```

```
geom_point()+
labs(x = expression("Discharge(ft\"^3\"/s)"), y = "Nitrate (mg/L)")
```



15. Does this storm show clockwise or counterclockwise hysteresis? Was this storm a flushing or diluting storm?

This storm shows counterclockwise hysteresis, it's a flushing storm.

16. What does this mean for how nitrate gets into the river from the watershed?

Nitrate gets into the river from quickflow, like pipes and overland flow.

Reflection

17. What are 2-3 conclusions or summary points about high frequency data you learned through your analysis?

1. high frequency data can be used to do analysis on temperature, DO, nutrients and storms.
2. Baseflow indicates flow from groundwater or soil. Quickflow indicates flow from overland and pipes.
3. Chemostatic storm: No slope, concentration does not change with increase of flow; Flushing storm: concentration increases with the increase of flow; Diluting flow: concentration decreases with the increase of flow.

18. What data, visualizations, and/or models supported your conclusions from 17?

The DO plot, nitrate plot, discharge plot, the plot of flow vs. nitrate for the large storm

19. Did hands-on data analysis impact your learning about high frequency data relative to a theory-based lesson? If so, how?

Hands-on data analysis can let me explore the theory on my own, and it can help me understand it better from examples.

20. How did the real-world data compare with your expectations from theory?

The real-world data sometimes can have missing data and gaps, before doing analysis, we need to address this problem first, like we can use linear regression or K Nearest Neighbor method to fill the gaps or clean up the outlier data.