Assignment 7: High Frequency Data

Walker Grimshaw

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)

```r
knitr::opts_chunk$set(warning = FALSE)

getwd()
```

```
## [1] "C:/Users/walke/OneDrive/Documents/Duke/Courses/Fall_2019/Hydrologic_Data_Analysis/Assignments"
```

```r
packages <- c("tidyverse",
              "StreamPULSE",
              "streamMetabolizer",
              "lubridate")
invisible(lapply(packages, library, character.only = TRUE))

theme_set(theme_bw())
```

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.

```r
# 4. download kansas river data
KansasRaw <- 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,Nitrate
##
```

```
## Retrieved the following variables:
##   DO_mgL, Discharge_m3s, Nitrate_mgL
# 5. Spread data
# create kansas longitude variable
KansasLon <- KansasRaw[[2]]$lon
# Start with just the data dataframe
KansasData <- KansasRaw[[1]] %>%
  # spread the variable column into multiple variable columns
  spread(value = value, key = variable) %>%
  # create solar time column using UTC time and the site longitude
  mutate(DateTime_Solar = convert_UTC_to_solartime(DateTime_UTC, KansasLon)) %>%
  # select only columns we care about
  select("DateTime_UTC", "DateTime_Solar", "site",
         "DO_mgL", "Discharge_m3s", "Nitrate_mgL")
# Rename site to SiteName
names(KansasData)[3] <- "SiteName"
```
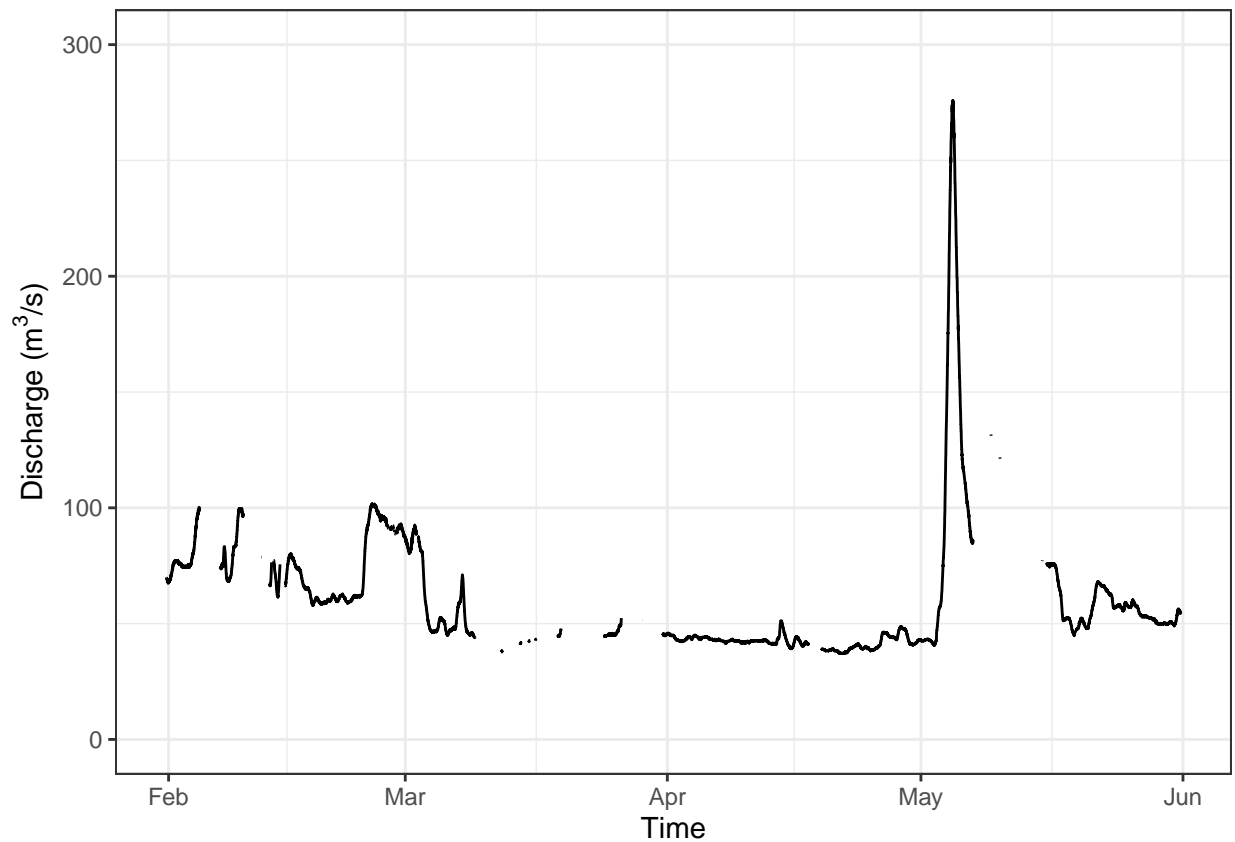
6. Plot each of the 3 variables against solar time for the period of record

```
# DO plot
ggplot(KansasData, aes(x = DateTime_Solar, y = DO_mgL)) +
  geom_line(color = "blue") +
  labs(x = "Time", y = "Dissolved Oxygen (mg/L)") +
  coord_cartesian(ylim = c(0,16))
```
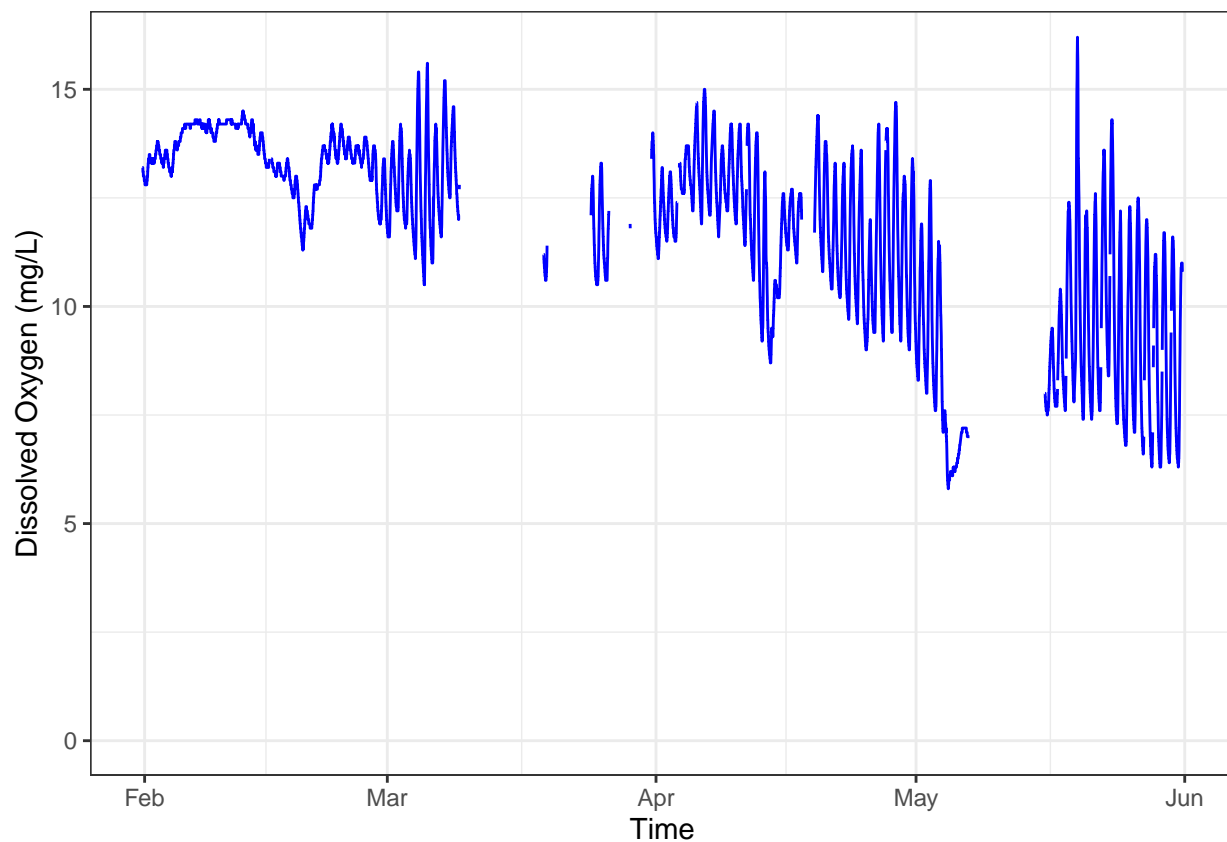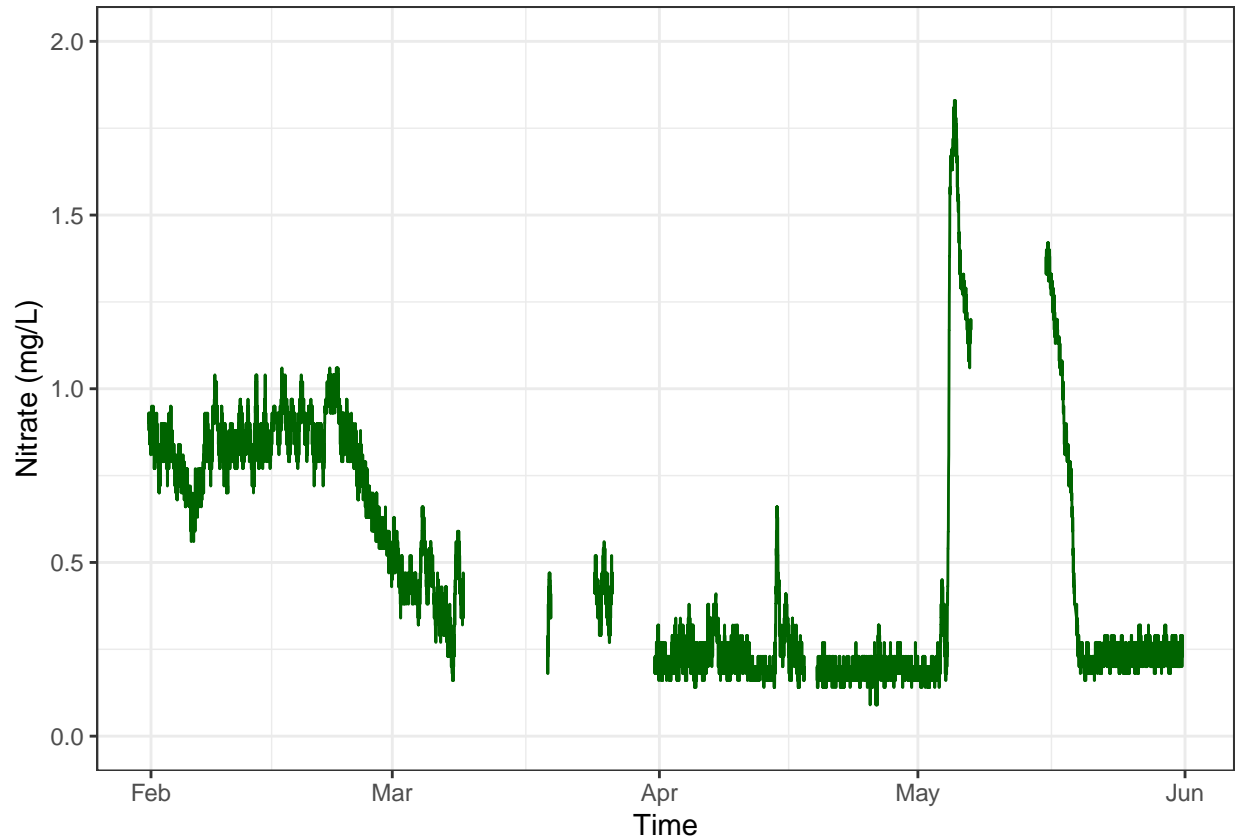


```
# Discharge plot
ggplot(KansasData, aes(x = DateTime_Solar, y = Discharge_m3s)) +
```

```
geom_line() +
labs(x = "Time", y = expression("Discharge (m"^3*"/s)")) +
coord_cartesian(ylim = c(0,300))
```



```
# Nitrate Plot
ggplot(KansasData, aes(x = DateTime_Solar, y = Nitrate_mgL)) +
  geom_line(color = "dark green") +
  labs(x = "Time", y = "Nitrate (mg/L)") +
  coord_cartesian(ylim = c(0,2))
```

7. How will you address gaps in these dataseries?

   Aggregation and subsampling the data do not make sense for these data, as we are interested in the high frequency data points and we are investigating only one season. Thus, I will interpolate the missing values.

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

   The amplitude of the daily DO change increases substantially through the spring, from February to June. As the river receives warms, biological activity within the river increases. With more sunlight, photosynthesis within the river increases and produces more oxygen during the day. At night, organisms continue to respire and consume oxygen in the river more quickly than it can redissolve without photosynthesis.

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.

```
## interpolate Discharge Values
## First, create two column data frame with UTC time and discharge
KansasDischarge <- select(KansasData, DateTime_UTC, Discharge_m3s)
## How many time steps? days * 24 hours * 4 15-minute intervals
timestep_number <- (28+31+30+31) * 24 * 4
## There are 14187 observations, but there are only 11520 15-minute time intervals,
## indicating many observations taken at the same time, in addition to the gaps
table(diff(KansasDischarge$DateTime_UTC))
```

```
##
##      0    900   1800   4500  15300
##   2788  11388      1      2      7
```

```
# this table does show 2,788 observations with no time between measurements

## Interpolate Discharge values
DischargeInterp <- as.data.frame(approx(KansasDischarge,
                                        n = timestep_number,
                                        method = "linear"))
# make the date column date-time again and rename the columns
DischargeInterp$x <- as_datetime(DischargeInterp$x, origin = lubridate::origin)
names(DischargeInterp) <- c("DateTime_UTC", "Discharge")

## Partition Discharge into baseflow and quickflow
FlowPaths <- EcoHydRology::BaseflowSeparation(
  DischargeInterp$Discharge,
  filter_parameter = 0.925, # default parameter
  passes = 3 # default parameter
  )
## recombine with datetime and total flow
KansasFlowPaths <- cbind(DischargeInterp, FlowPaths)

## Plot flow paths and total flow
ggplot(KansasFlowPaths, aes(x = DateTime_UTC)) +
  geom_line(aes(y = Discharge, color = "Total Flow"), size = 2) +
  geom_line(aes(y = bt, color = "Baseflow"), size = 1) +
  geom_line(aes(y = qft, color = "Quickflow")) +
  labs(x = "Time", y = expression("Discharge (m"^3*"/s)")) +
  theme(legend.title = element_blank()) +
  scale_color_brewer(palette = "Dark2")
```
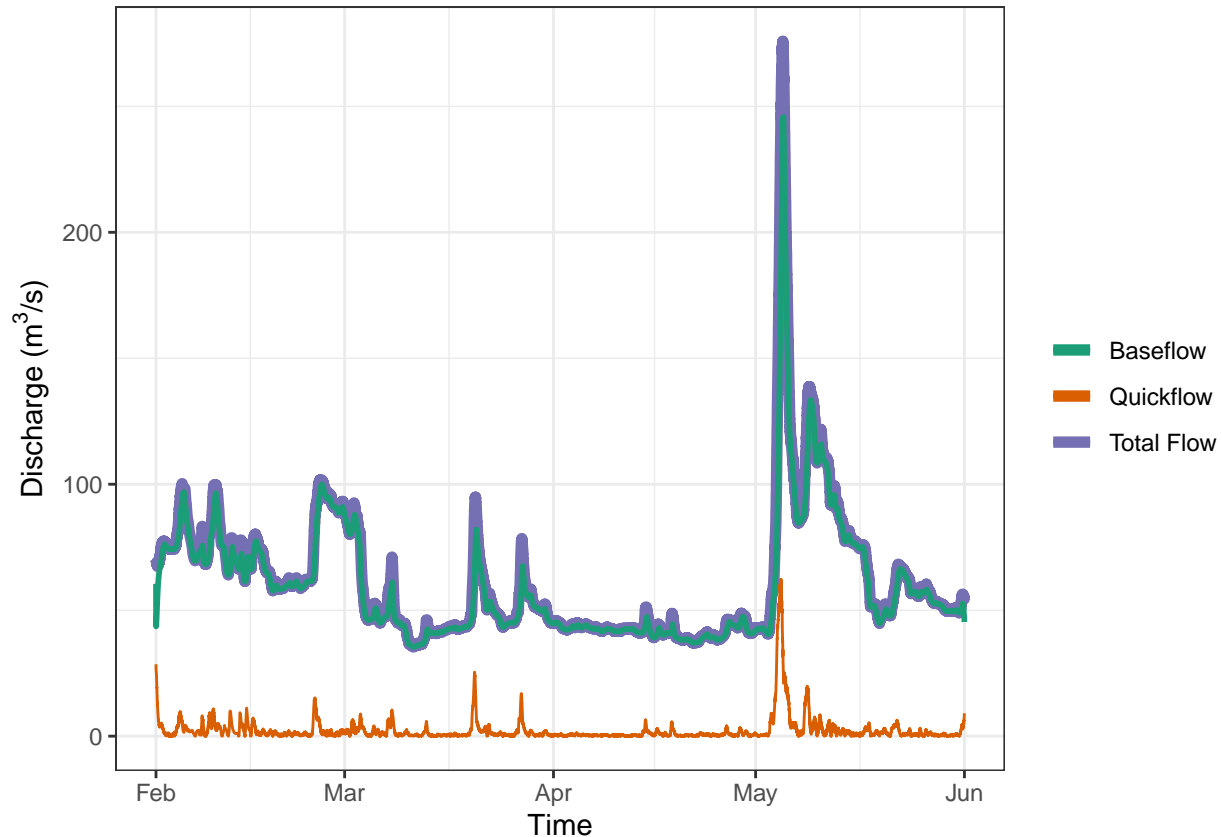
```
## calculate total percentages
## time step is unimportant because they all have the same time step
basepercent <- sum(KansasFlowPaths$bt)/sum(KansasFlowPaths$Discharge)
quickpercent <- sum(KansasFlowPaths$qft)/sum(KansasFlowPaths$Discharge)
```

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

    Approximately 96 percent of the flow was baseflow and 4% of the flow was quickflow from February through May.

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?

    Based on this very small sample size, watershed size is directly related to the percent of baseflow in the river. This makes sense, as larger rivers that are always flowing must have a substantial amount of baseflow. These same large rivers require very large storms that saturate the soil to cause larger flood events.

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?

    In this case, baseflow is more than just groundwater flow. Baseflow would also be from smaller precipitation events at headwater streams, increasing the baseflow of the larger river.

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.

```r
## First, create nitrate dataframe for interpolation
KansasNitrate <- select(KansasData, DateTime_UTC, Nitrate_mgL)

## Interpolate Nitrate values
NitrateInterp <- as.data.frame(approx(KansasNitrate,
                                      n = timestep_number,
                                      method = "linear"))
# make the date column date-time again and rename the columns
NitrateInterp$x <- as_datetime(NitrateInterp$x, origin = lubridate::origin)
names(NitrateInterp) <- c("DateTime_UTC", "Nitrate")

## join NitrateInterp with DischargeInterp by DateTime_UTC
KansasInterp <- inner_join(NitrateInterp, DischargeInterp, by = "DateTime_UTC")

## Create Storm dataframe
MayStorm <- filter(KansasInterp, DateTime_UTC > "2018-05-01" &
                    DateTime_UTC < "2018-05-21")

## hysteresis plot
ggplot(MayStorm, aes(x = Discharge, y = Nitrate, color = DateTime_UTC)) +
  geom_point() +
  labs(x = expression("Discharge (m"^3*"/s)"), y = "Nitrate (mg/L)",
       color = "Time")
```
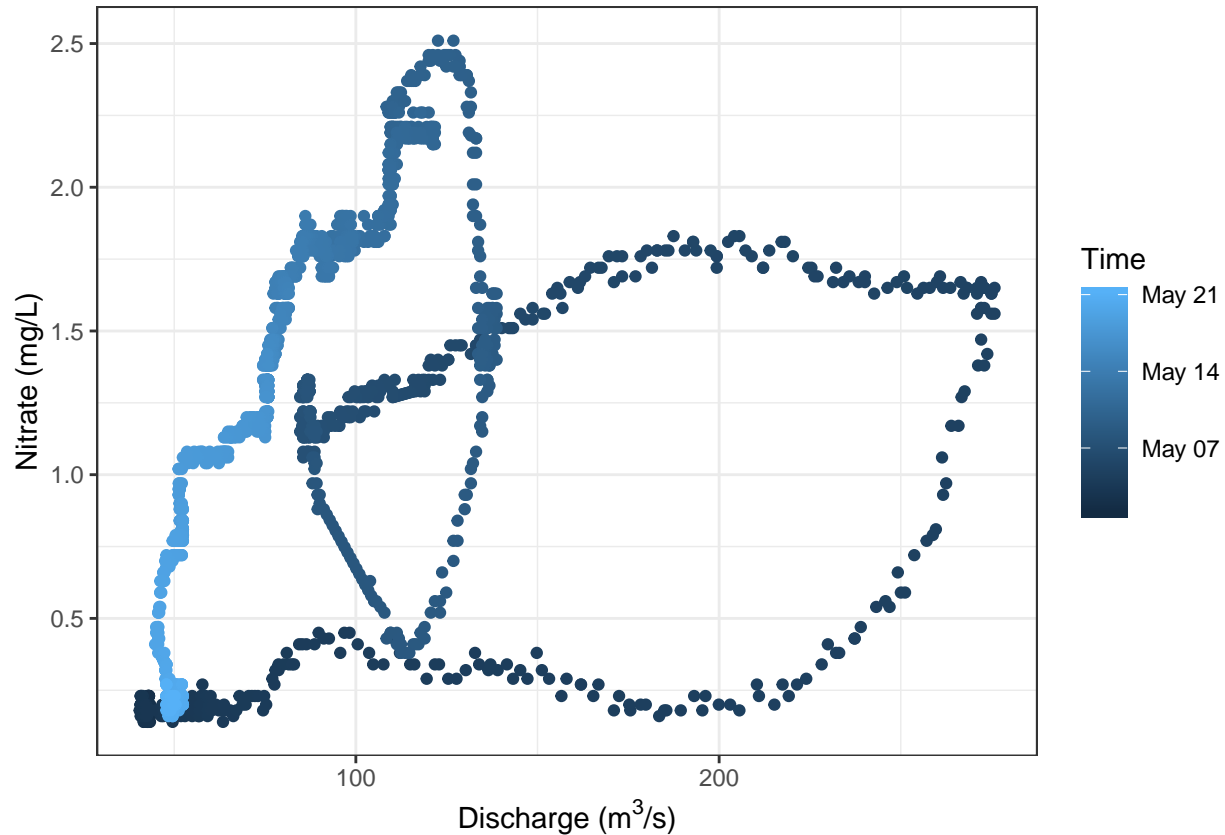
15. Does this storm show clockwise or counterclockwise hysteresis? Was this storm a flushing or diluting storm?

    This was a flushing storm with counterclockwise hysteresis.

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

    Nitrate primarily enters the river through groundwater.

Reflection

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

    Based on a limited number of observations, the size and characteristics of a watershed have a significant impact on the partitioning of precipitation into quickflow and baseflow to a river. Dissolved oxygen also shows high frequency and seasonal trends. The use of different interpolation methods also appears to have significant impact on data analysis.

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

    The partition figure demonstrates much greater amounts of base flow in larger watersheds, while the dissolved oxygen figure shows intradaily changes and a seasonal trend, but only with high frequency data.

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

    Absolutely. I took Jim Heffernan's Stream and River Ecology course last semester and did not fully understand the hydrologic concepts we discussed in class, such as hysteresis. After actually

creating a Q-C plot in class for a storm, I have a much better understanding of this concept.

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

The partitioning of baseflow and quickflow were actually more in line with my theoretical expectations than I thought would occur. The hysteresis loops were also much easier to connect to the theory than I had expected.