Assignment 7: High Frequency Data

Keith Bollt

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
getwd()
```

```
## [1] "Z:/Hydrologic_Data_Analysis2/Assignments"
```

```r
#install.packages("devtools")
library(devtools)
#install_github("streampulse/StreamPULSE")

#install.packages("streamMetabolizer", dependencies = TRUE, repos = c("http://owi.usgs.gov/R", "http://

#install.packages(EcoHydRology)

#install.packages("tidyverse")

library(streamMetabolizer)
library(StreamPULSE)
library(tidyverse)

theme_set(theme_classic(base_size = 12.261996))
```

4. Download data from the Stream Pulse portal using `request_data()` for the Kansas River, ("KS_KANSASR"). Download the discharge (`Discharge_m3s`), disolved 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
Kansasdata <- request_data(
  sitecode = "KS_KANSASR",
  variables = c('Discharge_m3s', 'DO_mgL', 'Nitrate_mgL')
  )
```

```
##
## 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
```

```r
Kansas.lon <- Kansasdata[[2]]$lon

Kansas.DO <- Kansasdata[[1]] %>%
  spread(value = value, key = variable) %>%
  mutate(DateTime_Solar = convert_UTC_to_solartime(DateTime_UTC, Kansas.lon))

colnames(Kansas.DO)
```

```
## [1] "DateTime_UTC"   "region"        "site"           "flagtype"
## [5] "flagcomment"    "Discharge_m3s" "DO_mgL"         "Nitrate_mgL"
## [9] "DateTime_Solar"
```

```r
Kansas.DO.processed <- Kansas.DO %>%
  select("DateTime_UTC", "DateTime_Solar", SiteName = "site", DO_mgL,
         Discharge_m3s, Nitrate_mgL)
```

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

```r
Do_plot <- ggplot(Kansas.DO.processed, aes(x = DateTime_Solar, y = DO_mgL))+
  geom_point()+
  labs(x = "Solar Time", y = "Dissolved Oxygen (mg/L)")
print(Do_plot)
```

```
## Warning: Removed 3047 rows containing missing values (geom_point).
```
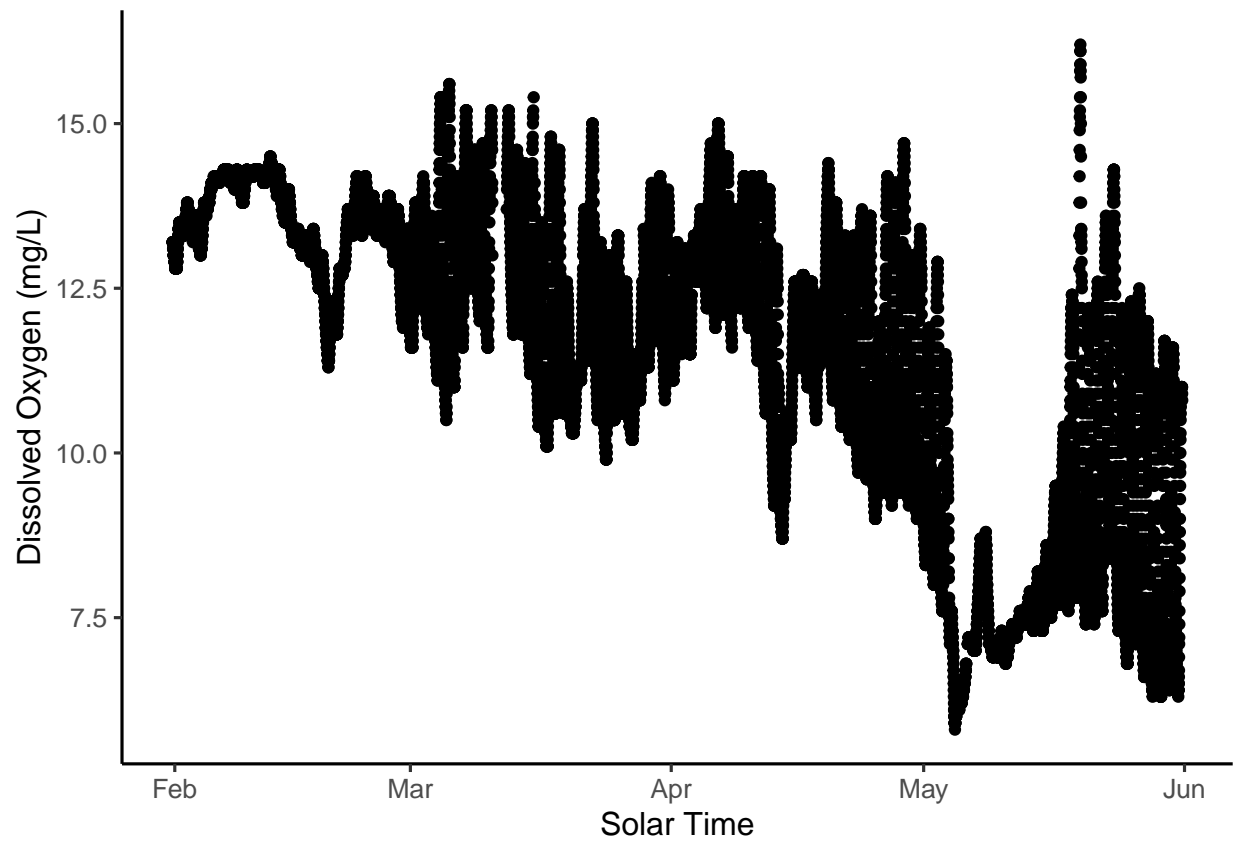
```
Discharge_plot <- ggplot(Kansas.DO.processed, aes(x = DateTime_Solar, y = Discharge_m3s))+
  geom_point()+
  labs(x = "Solar Time", y = "Discharge (m³/s)")
print(Discharge_plot)
```

```
## Warning: Removed 3428 rows containing missing values (geom_point).
```
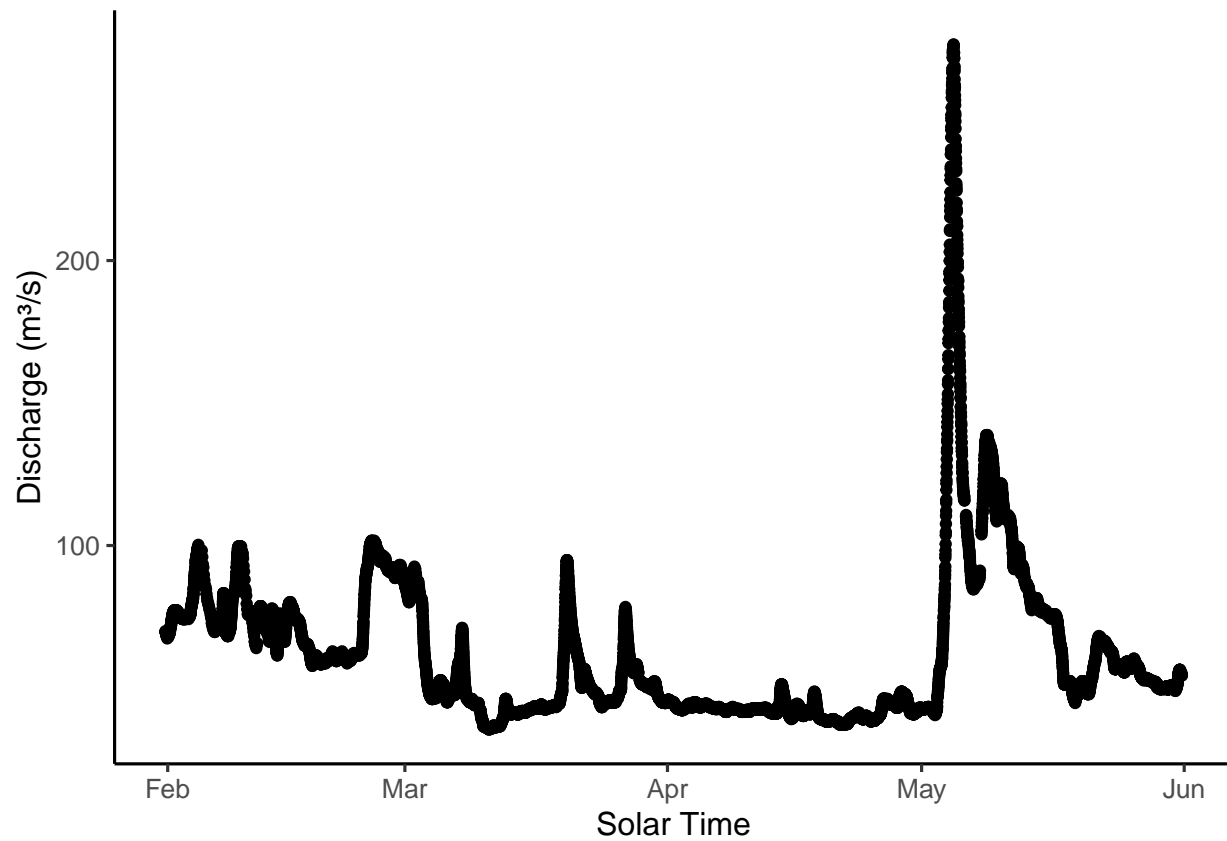
```
Nitrate_plot <- ggplot(Kansas.DO.processed, aes(x = DateTime_Solar, y = Nitrate_mgL))+
  geom_point()+
  labs(x = "Solar Time", y = "Nitrate (mg/L)")
print(Nitrate_plot)
```

```
## Warning: Removed 2908 rows containing missing values (geom_point).
```
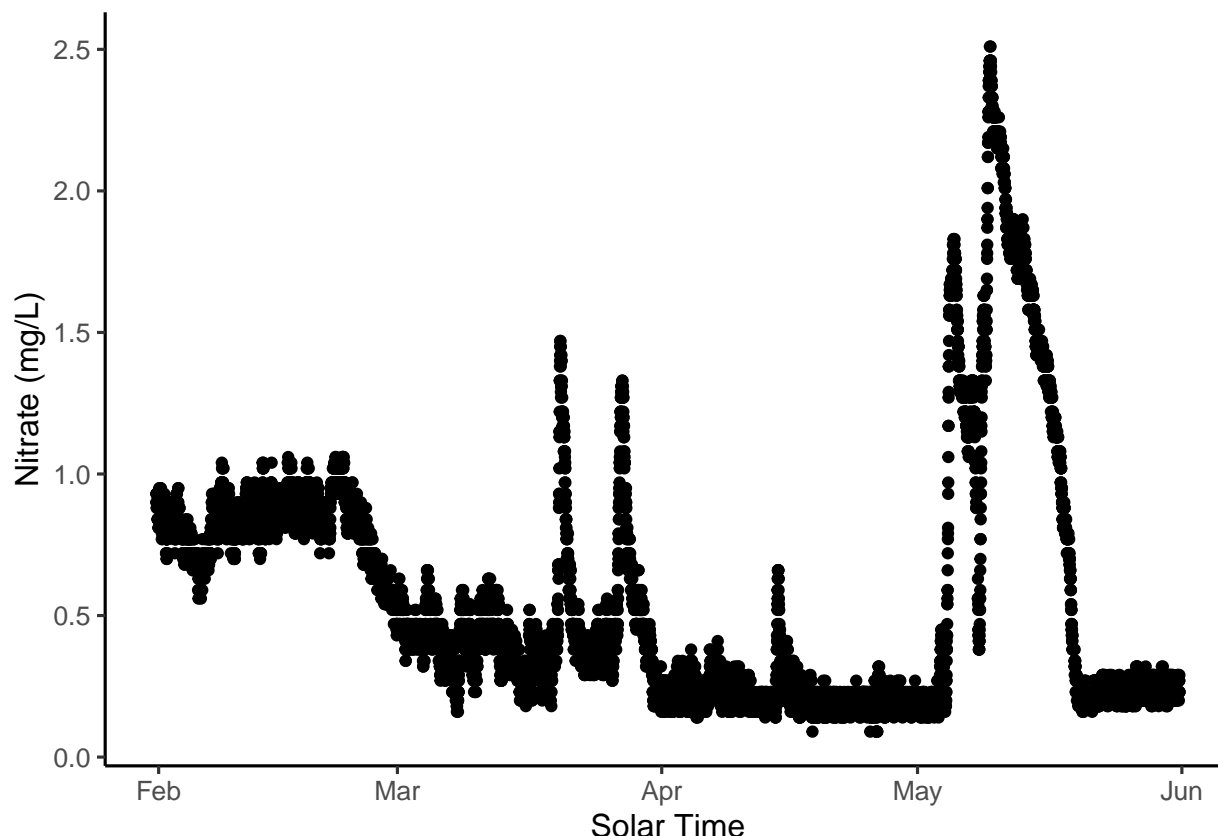
7. How will you address gaps in these dataseries?

For the hysteresis loops, time is not one of the two axes, because we are graphing discharge versus concentration; in an ideal world we would not have a gap, but it is ok that we do for the purpose of our analysis. For baseflow separation analysis, it seems appropriate to omit NA's at the outset. While this is an imperfect solution, it allows us to perform internal analysis of the data we already have, and seems like an appropriate way to separate quickflow and baseflow.

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

The amplitude of these swings increases from winter into spring and early summer. This is likely due to higher levels of photosynthesis and respiration as plants grow in the spring and organisms' metabolisms increase with warmer water.

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.

```r
Kansas.DO.DoublyProcessed <-
  Kansas.DO.processed %>%
  na.omit(Kansas.DO.processed, cols = "Discharge_m3s")

Kansasbaseflow <- EcoHydRology::BaseflowSeparation(
  Kansas.DO.DoublyProcessed$Discharge_m3s,
  filter_parameter = 0.925,
  passes = 3
  )

Kansas2018 <- cbind(Kansas.DO.DoublyProcessed, Kansasbaseflow)

ExportKansas <- Kansas2018 %>%
  mutate(timestep = c(diff(as.numeric(DateTime_UTC)), NA_real_),
         baseflowexport = bt * timestep,
         quickflowexport = qft * timestep) %>%
  summarize(BaseflowExport_cf = sum(baseflowexport, na.rm = T),
            QuickflowExport_cf = sum(quickflowexport, na.rm = T),
            TotalExport_cf = BaseflowExport_cf + QuickflowExport_cf)

DischargePlot2 <- ggplot(Kansas2018, aes(x = DateTime_UTC))+
  geom_point(aes(y = Discharge_m3s), color = "Blue")+
  labs(x = "Date", y = "Total Discharge (m³/s)")

BaseflowPlot <- ggplot(Kansas2018, aes(x = DateTime_UTC))+
  geom_point(aes(y = bt), color = "Light Blue")+
  labs(x = "Date", y = "Baseflow (m³/s)")

QuickflowPlot <- ggplot(Kansas2018, aes(x = DateTime_UTC))+
  geom_point(aes(y = qft), color = "navy")+
  labs(x = "Date", y = "Quickflow (m³/s)")

#install.packages("cowplot")
library(cowplot)

Flowplot <-
  plot_grid(DischargePlot2, BaseflowPlot, QuickflowPlot, ncol = 1, labels = c('', '', ''))
print(Flowplot)
```
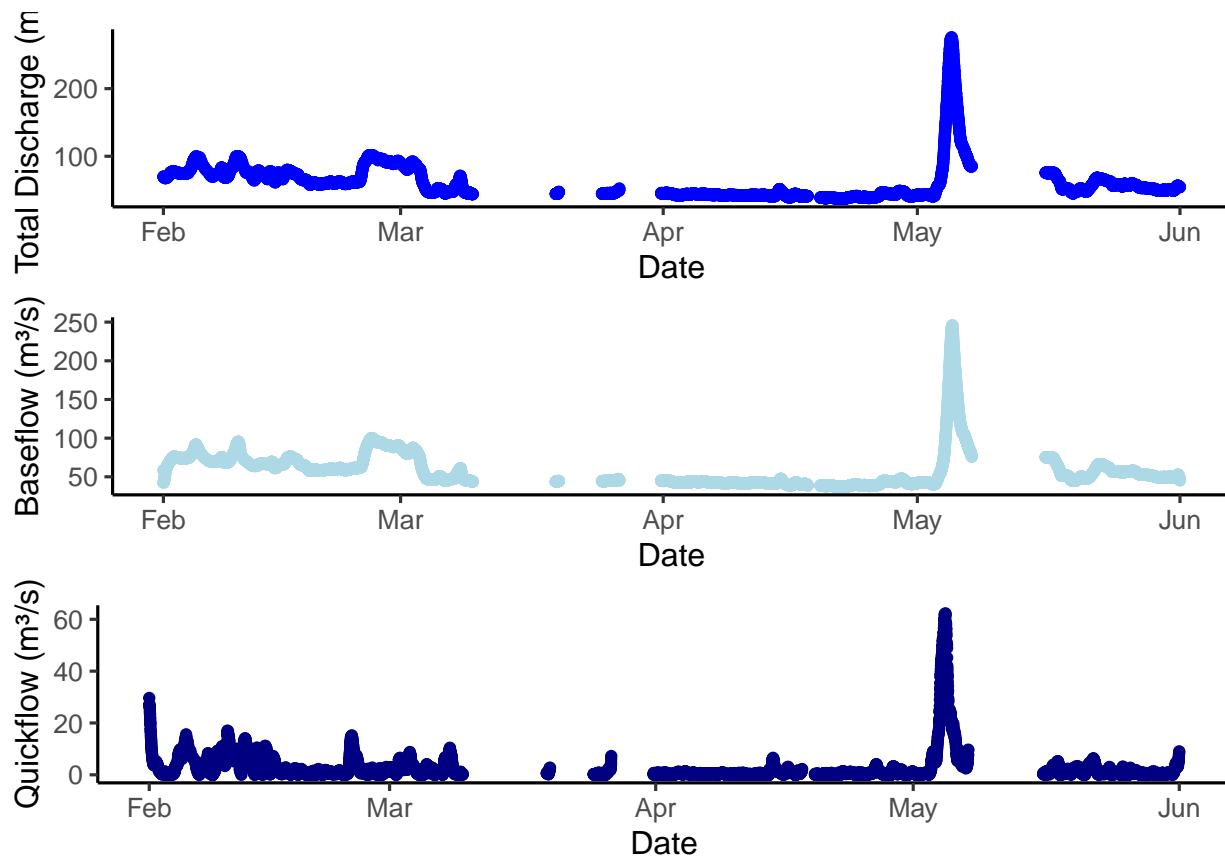
11. What percentage of total water exported left as baseflow and quickflow from the Kansas River over this time period?

    94.6% baseflow; 5.4% quickflow

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?

    Baseflow makes up a higher percentage of discharge in larger rivers and in larger watersheds, all else equal.

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?

    The discharge at a point further down in the water network doesn't have to rely on groundwater alone to maintain its baseflow. Rather, odds are its entire watershed is not in drought, and that some individual tributaries are contributing a fair amount of water to discharge at any given point in time. This results in higher baseflow as a percentage of total discharge.

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.

```
MaystormDF <- Kansas.DO.DoublyProcessed %>%
  filter(DateTime_UTC < "2018-05-20 23:45:00", DateTime_UTC > "2018-05-01 00:00:00")

Hysteresisplot <- ggplot(MaystormDF, aes(x = Discharge_m3s, y = Nitrate_mgL, color = DateTime_UTC))+
```

```
  geom_point()+
  labs(color = "Date", x = "Discharge (m³/s)", y = "Nitrate (mg/L)")
print(Hysteresisplot)
```
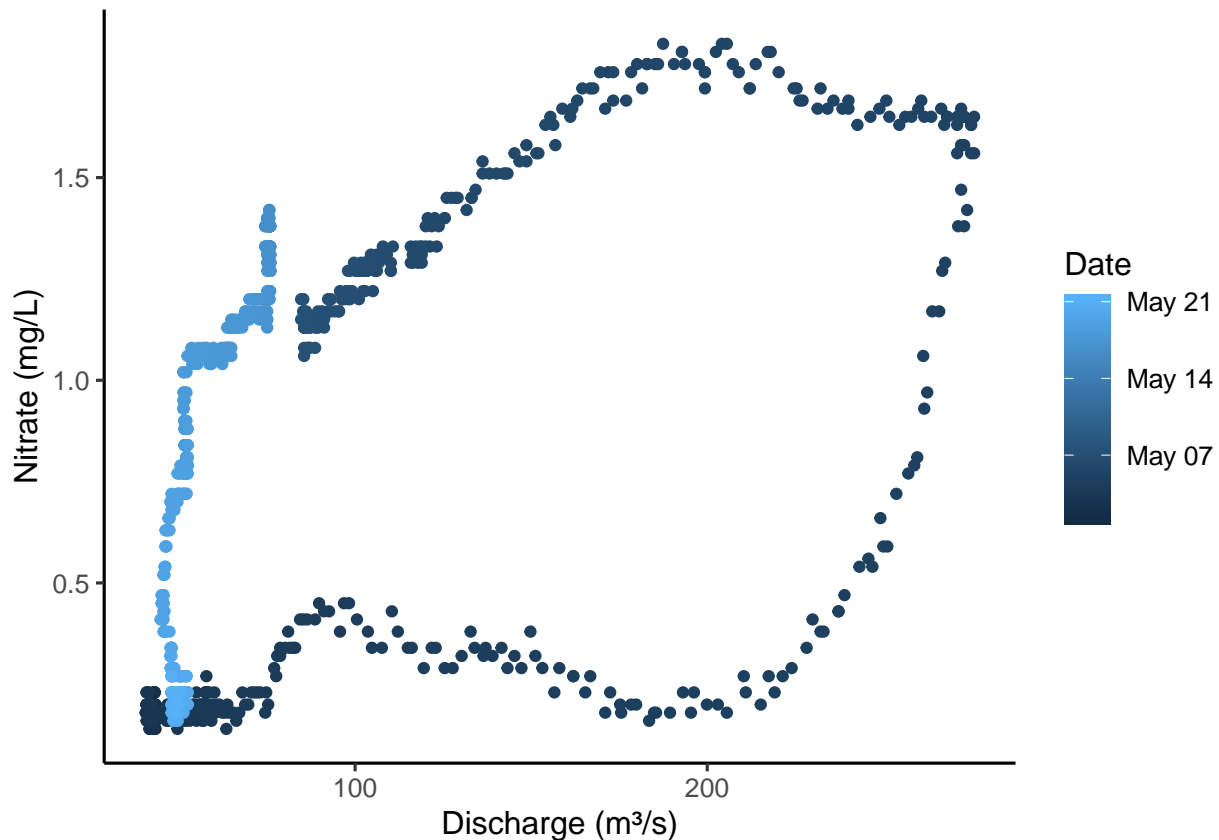

15. Does this storm show clockwise or counterclockwise hysteresis? Was this storm a flushing or diluting storm?

    Counterclockwise hysteresis. Concentration increases as flow increases, so this is a flushing storm.

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

    It enters the watershed as a result of rain events, likely as a result of overland flow. There is likely a lot of corn being grown in the watershed, and corn is treated with high levels of nitrogen.

Reflection

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

    Different analyses of high frequency data deal with data gaps in different ways. Flushing storms show increasing concentration with increasing flow, whereas diluting storms show decreasing concentration with increasing flow. Larger rivers have more baseflow as a proportion of total discharge when compared to smaller rivers.

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

    1) I omitted the data gaps for the flow separation analysis, but kept them in for the hysteresis loop analysis.

2) Looking at the plot I made for question 14, and comparing it to the definitions for flushing and diluting storms from the lesson.

3) I calculated a much higher baseflow percentage for the Kansas River than for the two rivers from class.

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

Yes. I was able to visualize hysteresis loops, which helped me understand the difference between flushing and diluting storms better than just it being explained to me.

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

I am not surprised that a river in the corn belt of the US increases its nitrogen concentration with increased flow. I am surprised how much higher baseflow is relative to quickflow in the Kansas River, when compared to the stream we looked at for class.