Assignment 7: High Frequency Data

Felipe Raby Amadori

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
# 1. Verify your working directory is set to the R project file
getwd()
```

```
## [1] "C:/Users/Felipe/OneDrive - Duke University/1. DUKE/Ramos 3 Semestre/Hydrologic_Data_Analysis/As
```

```r
# 2. Load the StreamPULSE, streamMetabolizer and tidyverse packages.

packages <- c(
  "ggplot2",
  "tidyverse",
  "StreamPULSE",
  "streamMetabolizer"
  )
invisible(
  suppressPackageStartupMessages(
    lapply(packages, library, character.only = TRUE)
    )
  )

# 3. Set your ggplot theme (can be theme_classic or something else)
felipe_theme <- theme_light(base_size = 12) +
  theme(axis.text = element_text(color = "grey8"),
        legend.position = "right", plot.title = element_text(hjust = 0.5))
theme_set(felipe_theme)
```

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
# Kansas River at Desoto, KS

KansasRiver.Desoto.dat <- 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
```

```r
KansasRiver.lon <- KansasRiver.Desoto.dat[[2]]$lon

KansasRiver.Desoto.all <- KansasRiver.Desoto.dat[[1]] %>%
  spread(value = value, key = variable) %>%
  mutate(DateTime_Solar = convert_UTC_to_solartime(DateTime_UTC, KansasRiver.lon)) %>%
  select("DateTime_UTC", "DateTime_Solar", "site", "DO_mgL", "Discharge_m3s", "Nitrate_mgL")

names(KansasRiver.Desoto.all)[3] <- "SiteName"
```

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

```r
KansasRiver.Desoto.all.gathered <- gather(KansasRiver.Desoto.all, variable, value,
                                          "DO_mgL":"Nitrate_mgL")

ggplot(KansasRiver.Desoto.all.gathered, aes(x = DateTime_Solar)) +
  geom_line(aes(y = value, color = variable)) +
  xlab(expression("Solar Time")) +
  ylab(expression("Variable Value")) +
  labs(color = 'Variables') +
  scale_color_brewer(palette = "Set1", labels =
                       c("Discharge (m3/s)", "DO (mg/l)", "Nitrate (mg/l)")) +
  facet_wrap(vars(variable), nrow = 3, scales = "free") +
  theme(plot.title = element_text(hjust = 0.5)) +
  ggtitle("Discharge, DO, and Nitrate vs Solar Time")
```

Discharge, DO, and Nitrate vs Solar Time
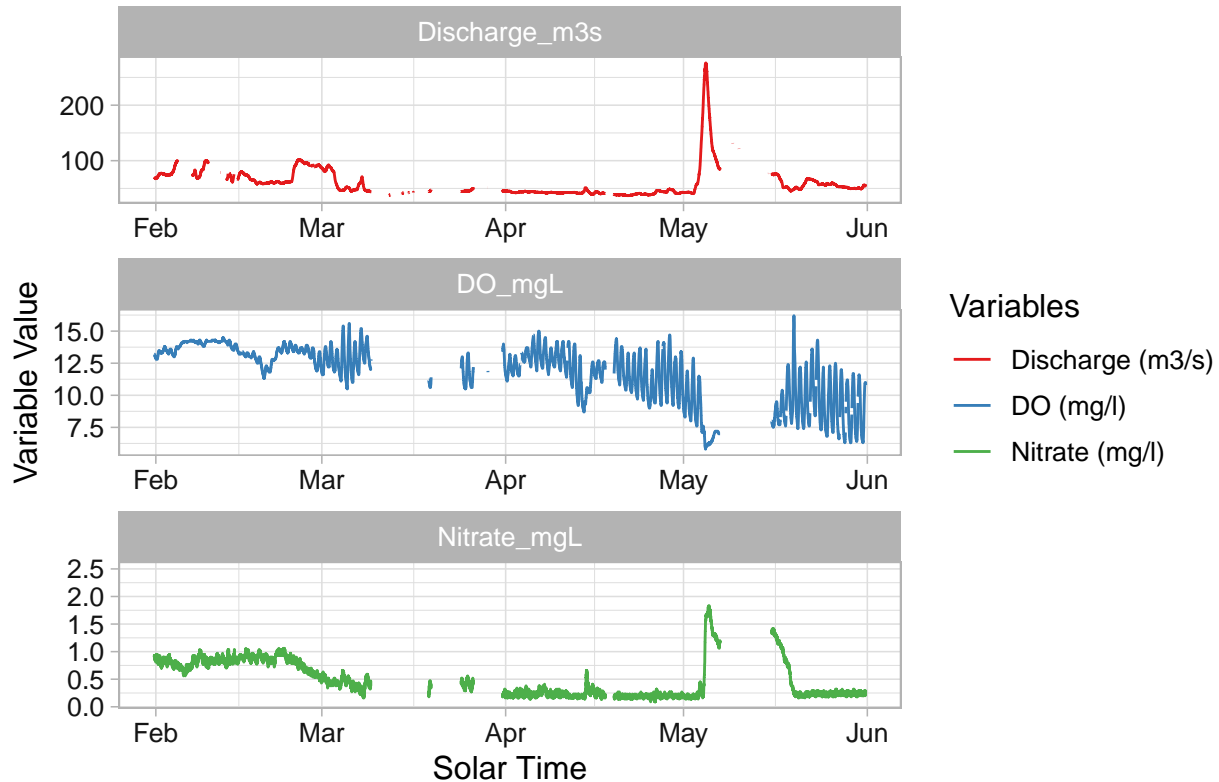
7. How will you address gaps in these dataseries?

We can see in the "Discharge, DO, and Nitrate vs Solar Time" plot that all the variables have gaps in the data; nevertheless, in this assignment we are going to use only Discharge and Nitrate data so we are only going to address the gap in those two variables. In the analysis of baseflow and quickflow discharge it could be relevant to interpolate the data to have a better feel of how discharge changes over time, so we are going to perform a linear interpolation every 15 minutes. For Nitrate, we are analyzing a concentration versus flow plot; therefore, with the data that we already have we will be able to see the patterns that we are looking for. Interpolation is not needed.

```r
#Discharge Linear Interpolation

KansasRiver.Discharge <- select(KansasRiver.Desoto.all, "DateTime_Solar", "Discharge_m3s")


# 120 days every 15 min, n = 120*24*4 = 11,520 + 1 = 11,521
linearinterpolation <- as.data.frame(approx(KansasRiver.Discharge, n = 11521, method = "linear"))

seq1 <- as.data.frame(seq(as.POSIXlt(min(KansasRiver.Discharge$DateTime_Solar)),
                          as.POSIXlt(max(KansasRiver.Discharge$DateTime_Solar)), by=900))

linearinterpolation <- cbind(seq1, linearinterpolation)
linearinterpolation <-
select(
linearinterpolation,
"seq(as.POSIXlt(min(KansasRiver.Discharge$DateTime_Solar)), as.POSIXlt(max(KansasRiver.Discharge$DateTim
```

```
,y)

names(linearinterpolation) <- c("Date", "Discharge_m3s")


#Inspect interpolated values
ggplot(KansasRiver.Discharge, aes(x = DateTime_Solar, y = Discharge_m3s)) +
  geom_point() +
  geom_line() +
  geom_point(data = linearinterpolation, aes(x = Date, y = Discharge_m3s), color = "#c13d75ff") +
  xlab(expression("Solar Time")) +
  ylab(expression("Discharge (m3/s)")) +
  ggtitle("Kansas River at Desoto Interpolated Discharge")
```

```
## Warning: Removed 3428 rows containing missing values (geom_point).
```


Kansas River at Desoto Interpolated Discharge
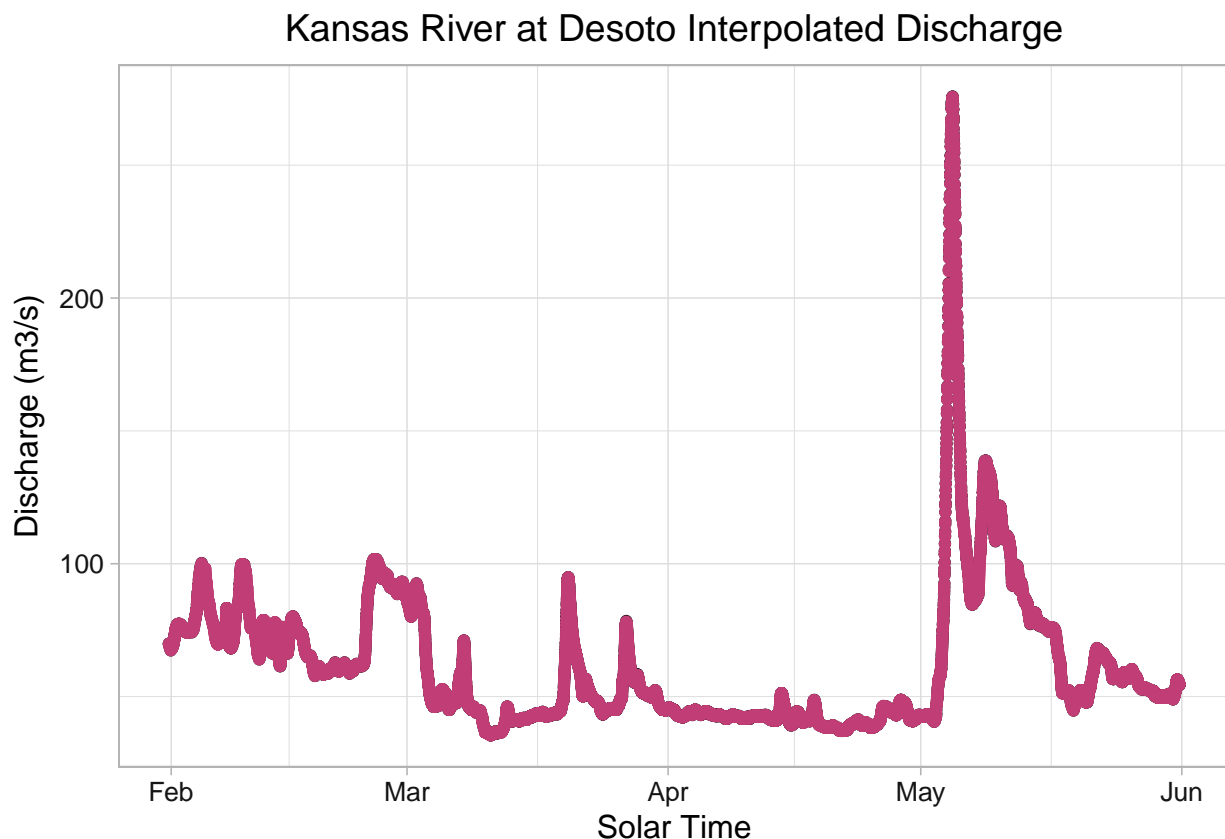
8. How does the daily amplitude of oxygen concentration swings change over the season? What might cause this?

The daily amplitude of oxygen concentration swings change considerably over the season. In February (winter) the DO concentration is relatively stable at values around 14 to 12.5 mg/l. In March the daily amplitude starts to increase with swings of approx. 4 mg/l. The biggest amplitude can be seen in May (summer) with a daily variation of 10 mg/l. The mean concentrations seems to fall from winter to summer. The amount of dissolved oxygen depends on temperature; therefore, the observed patterns could be caused by greater temperatures in summer (less solubility of oxygen) and greater temperature differential between day and night.

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
KansasRiver.Desoto.baseflow <- EcoHydRology::BaseflowSeparation(
  na.omit(linearinterpolation$Discharge_m3s),
  filter_parameter = 0.925,
  passes = 3
  )

KansasRiver.Desoto.Discharge <-linearinterpolation %>%
  cbind(KansasRiver.Desoto.baseflow)

KansasRiver.Desoto.Discharge.gathered <- gather(KansasRiver.Desoto.Discharge,
                                                Discharge_type, "value_m3s",
                                                "Discharge_m3s":"qft")


ggplot(KansasRiver.Desoto.Discharge.gathered) +
  geom_line(aes(x = Date, y = value_m3s, color = Discharge_type)) +
  labs(x = "", y = expression("Discharge (m"^3*"/s)"),
       title = "Kansas River at Desoto Total Flow, Baseflow, and Quickflow Over Time") +
  labs(color = 'Type of Flow') +
  scale_color_brewer(palette = "Set1", labels = c("Baseflow", "TotalFlow", "Quickflow")) +
  theme(plot.title = element_text(hjust = 0.5))
```

ansas River at Desoto Total Flow, Baseflow, and Quickflow Over Time
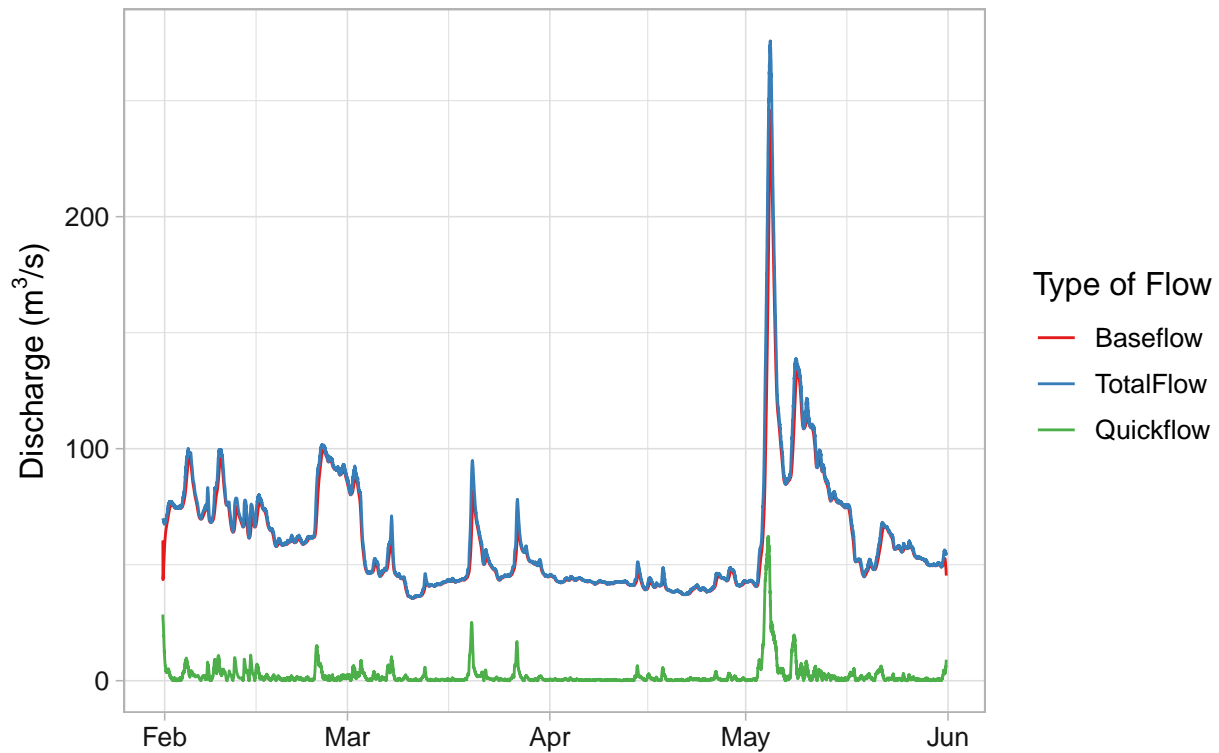
11. What percentage of total water exported left as baseflow and quickflow from the Kansas River over this time period?

```
Export.KansasRiver.Desoto <- KansasRiver.Desoto.Discharge %>%
  mutate(timestep = c(diff(as.numeric(Date)), 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)

PErcentBaseflow <- Export.KansasRiver.Desoto[1]/Export.KansasRiver.Desoto[3]*100
PErcentBaseflow
```

```
##   BaseflowExport_cf
## 1          96.19967
```

```
PErcentQuickflow <- 100-PErcentBaseflow
PErcentQuickflow
```

```
##   BaseflowExport_cf
## 1          3.800333
```

The percentage of total water exported left as baseflow from the Kansas River over this time period was 96.2%. The percentage of total water exported left as quickflow from the Kansas River over this time period was 3.8%.

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?

In a bigger watershed, surface flow takes more time to concentrate in streams; therefore, quickflow tends to have an smaller proportion of the total flow than in smaller watersheds. Moreover, surface water has more time to infiltrate in the soil and a greater volume of soil is "giving" water to streams during a longer period of time, so baseflow tends to have a larger proportion of the total flow.

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?

   If the site is further down in its river network in a river that has multiple tributaries that flow in to it, the effect of the size of the watershed (described in question #12) is more important. Basically, the important parameter to consider are the area of the watershed upstream the gauge and the length that the drops of water have to travel through the watershed to get to the gauge.

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.

```
KansasRiver.Desoto.Storm <- KansasRiver.Desoto.all %>%
  filter(DateTime_Solar > "2018-05-01" & DateTime_Solar < "2018-05-20")

ggplot(KansasRiver.Desoto.Storm, aes(x = Discharge_m3s, y = Nitrate_mgL, color = DateTime_Solar)) +
  geom_point() +
  labs(x = expression("Discharge (m"^3*"/s)"), y = expression("Nitrate (mg/s)"),
       title = "Flow vs. Nitrate May 1 to May 20") +
  labs(color = 'Date Time')
```

```
## Warning: Removed 1638 rows containing missing values (geom_point).
```

Flow vs. Nitrate May 1 to May 20
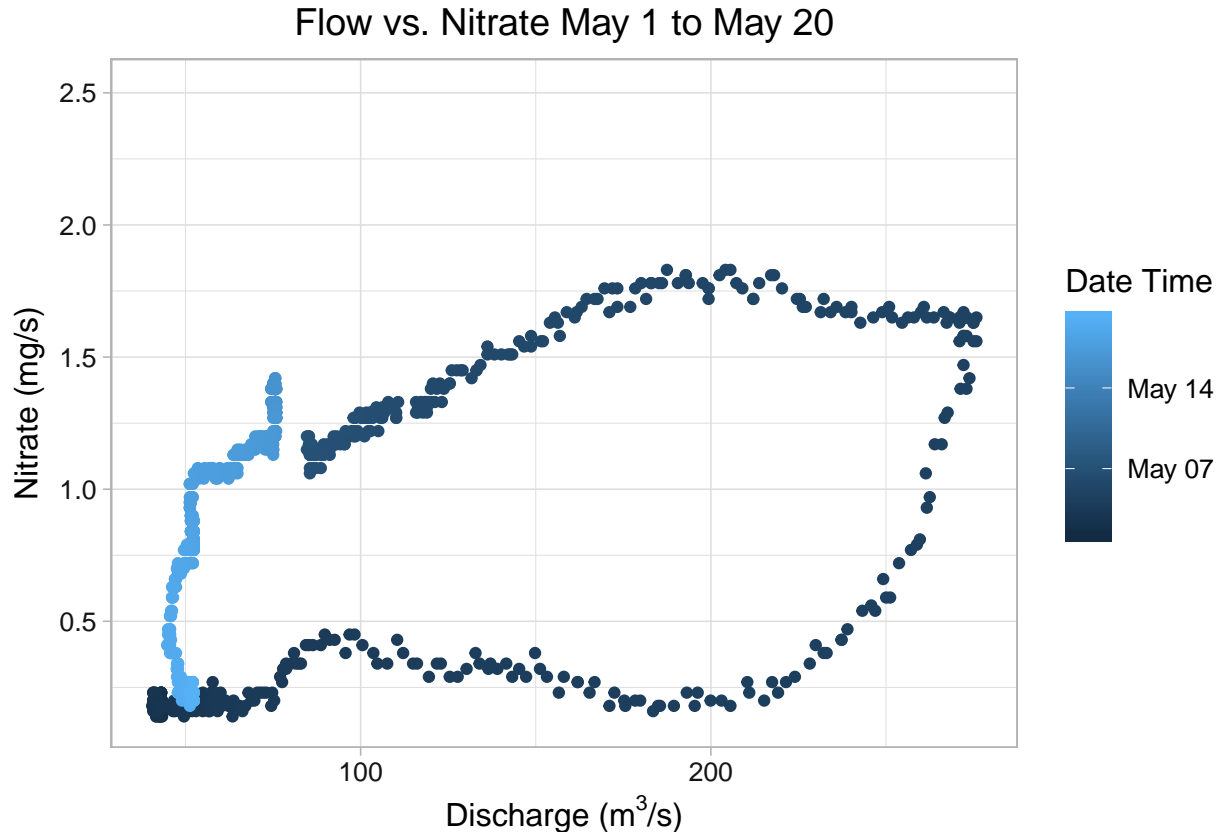
15. Does this storm show clockwise or counterclockwise hysteresis? Was this storm a flushing or diluting storm?

    We can see in the CQ plot that the plot has a positive plot; therefore, is not chemostatic and the storm was a flushing one. Nitrate was flushed into the stream during the storm. The storm also shows a counterclockwise hysteresis.

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

    It means that the concentration of Nitrate is higher in the quick pathways of the watershed. The nitrate is in the surface of the watershed (probably urban or agricultural). Storms are a source of Nitrate to the stream.

Reflection

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

    1. I learned that is harder to work with POSIXct than with regular dates.
    2. I learned that high frequency data points are usually so dense that more scattered points can sometimes not be seen in a plot.
    3. The diff. behavior that quickflow and baseflow can have in different types of rivers.
    4. How to identify a diluting river from a flushing river.

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

    1. Trying to do a linear interpolation with POSIXct data
    2. Looking a the result of the linear interpolation of discharge.

3. Comparing the quickflow baseflow of the river in class and this one.
4. The CQ plot.

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

   Hands-on analysis can be a lot of times harder that theory based lesson because you have to figure out things by yourself and think much more about every step that you are taking. All of these is definitely relevant to learn the lessons and to be able to use them in the future.

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

   I didn't expected the data to have that many gaps, but is understandable considering the high frequency of it. Any repair or even removing sensors to collect the data can interfere with the frequency of the data collection.