Assignment 6: Time Series Analysis

Gabi Richichi

OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis on time series analysis

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., "A06_Salk.html") prior to submission.

The completed exercise is due on 11 October 2019 at 9:00 am.

Setup

1. Verify your working directory is set to the R project file,
2. Load the tidyverse, lubridate, trend, and dataRetrieval packages.
3. Set your ggplot theme (can be theme_classic or something else)
4. Load the ClearCreekDischarge.Monthly.csv file from the processed data folder. Call this data frame ClearCreekDischarge.Monthly.

```
getwd()
```

```
## [1] "/Users/gabriellerichichi/Documents/5th year @ Duke/Stats/Hydrologic_Data_Analysis/Assignments"
```

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------------------ tidyver

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ---------------------------------------------------------------------------- tidyverse_co
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

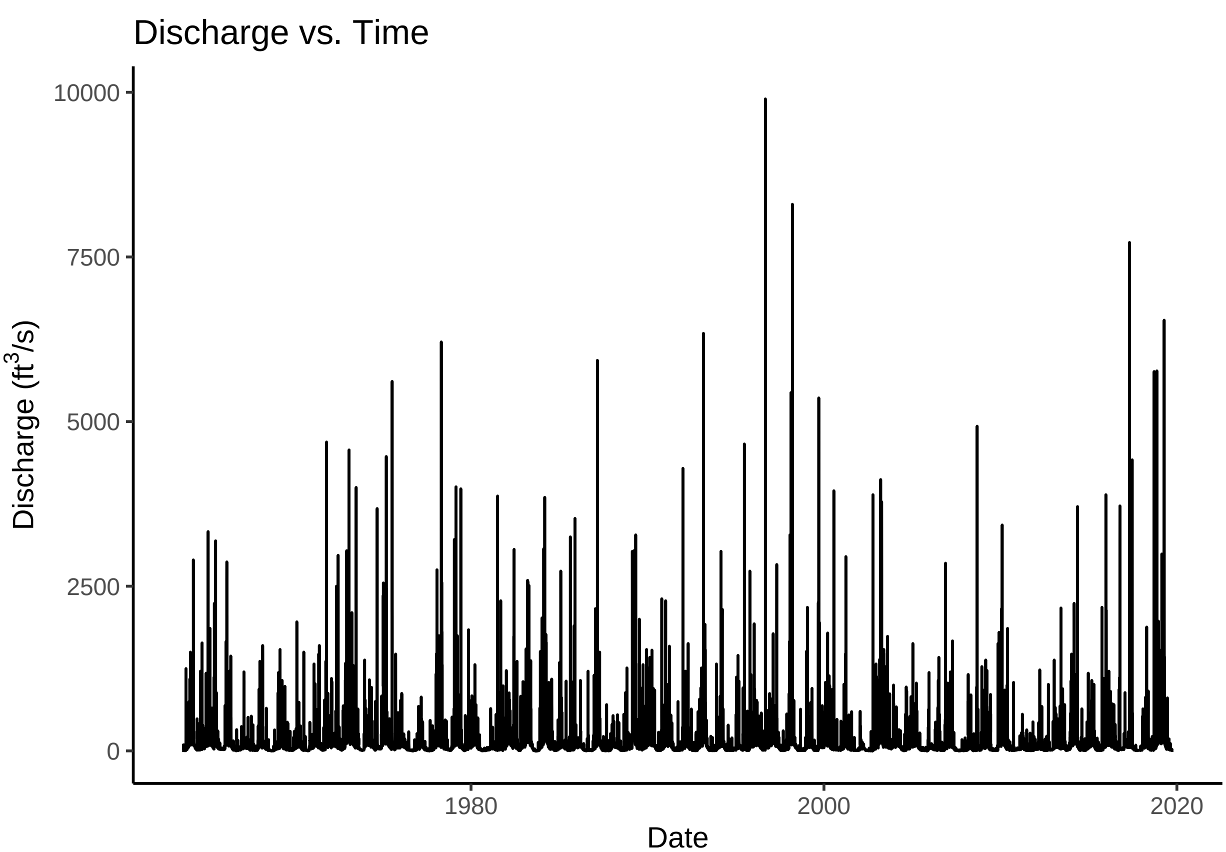
```
library(trend)
library(dataRetrieval)

theme_set(theme_classic())
```

```r
#ClearCreekDischarge.Monthly <- read.csv(file = "./Data/Processed/ClearCreekDischarge.Monthly.csv")
```

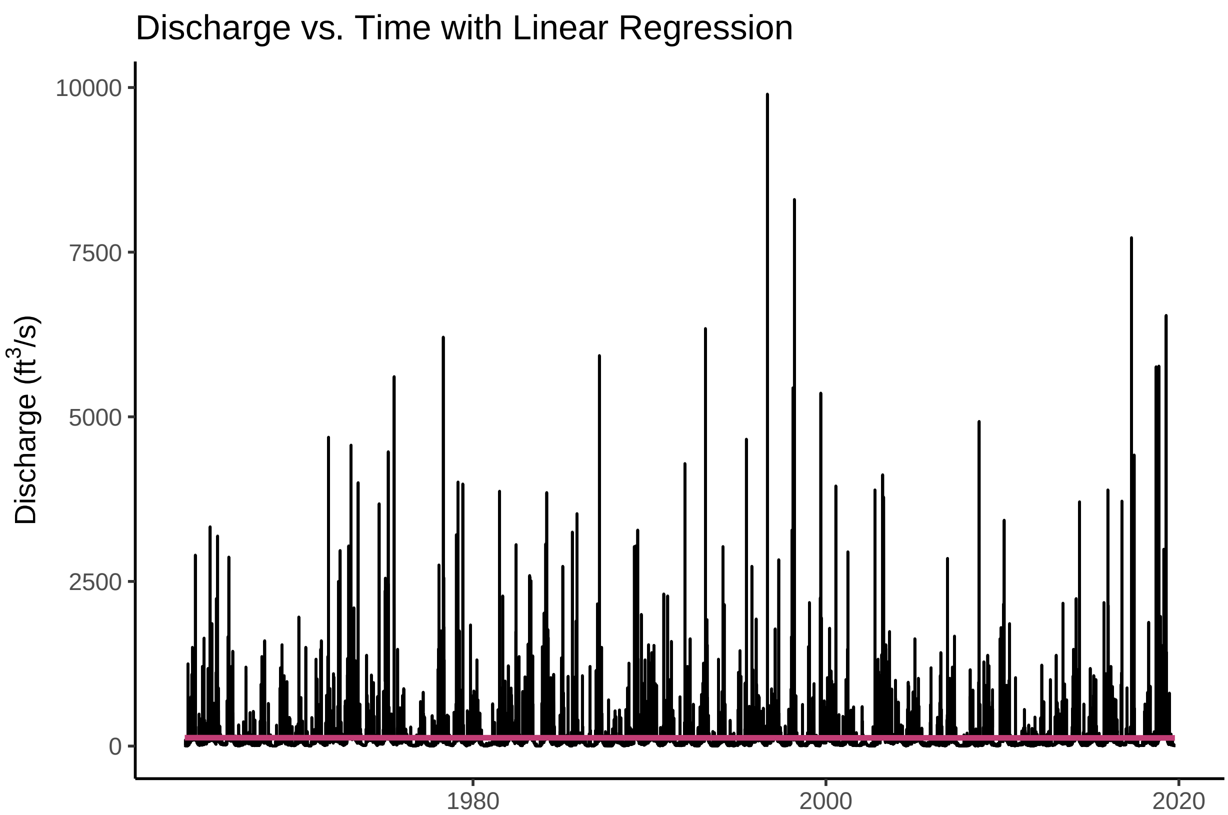
Time Series Decomposition

5. Create a new data frame that includes daily mean discharge at the Eno River for all available dates (`siteNumbers = "02085070"`). Rename the columns accordingly.
6. Plot discharge over time with geom_line. Make sure axis labels are formatted appropriately.
7. Create a time series of discharge
8. Decompose the time series using the `stl` function.
9. Visualize the decomposed time series.

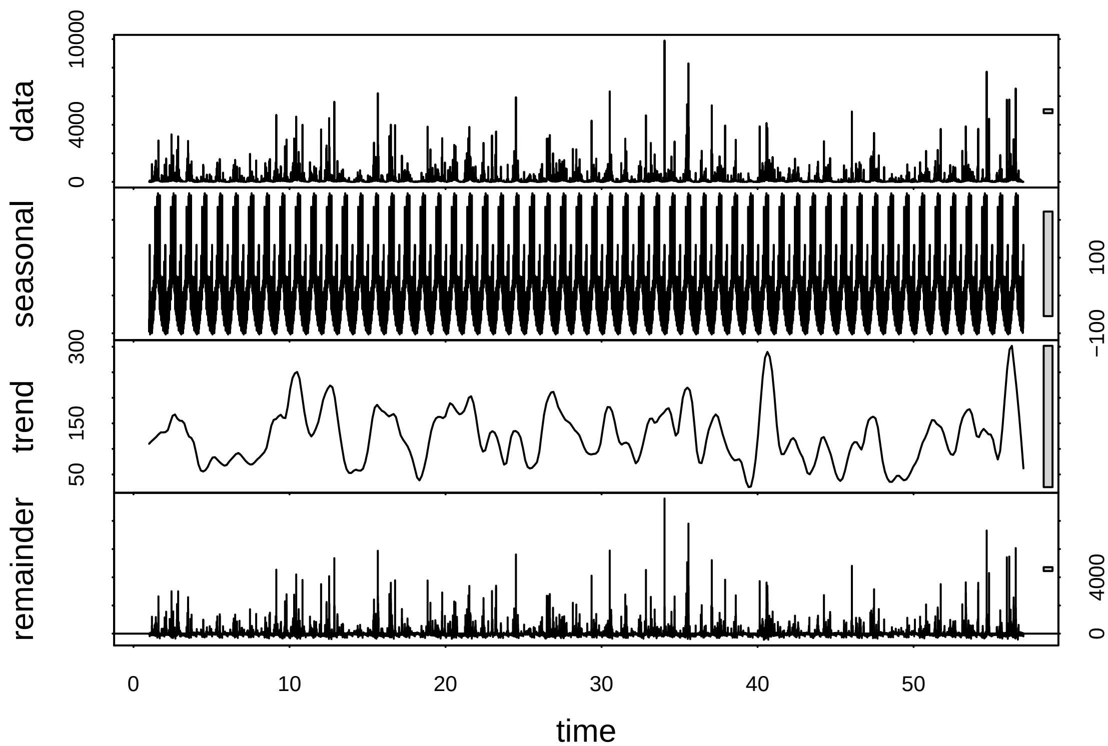
```r
EnoDailyMeanDischarge <- readNWISdv(siteNumbers = "02085070",
                                   parameterCd = "00060",
                                   startDate = "",
                                   endDate = "")
names(EnoDailyMeanDischarge)[4:5] <- c("Discharge", "Approval.Code")

EnoDischargeTimePlot <-
  ggplot(EnoDailyMeanDischarge, aes(x = Date, y = Discharge)) +
  xlab("Date") +
  ylab(expression("Discharge (ft"^3*"/s)")) +
  geom_line() +
  ggtitle("Discharge vs. Time")
print(EnoDischargeTimePlot)
```



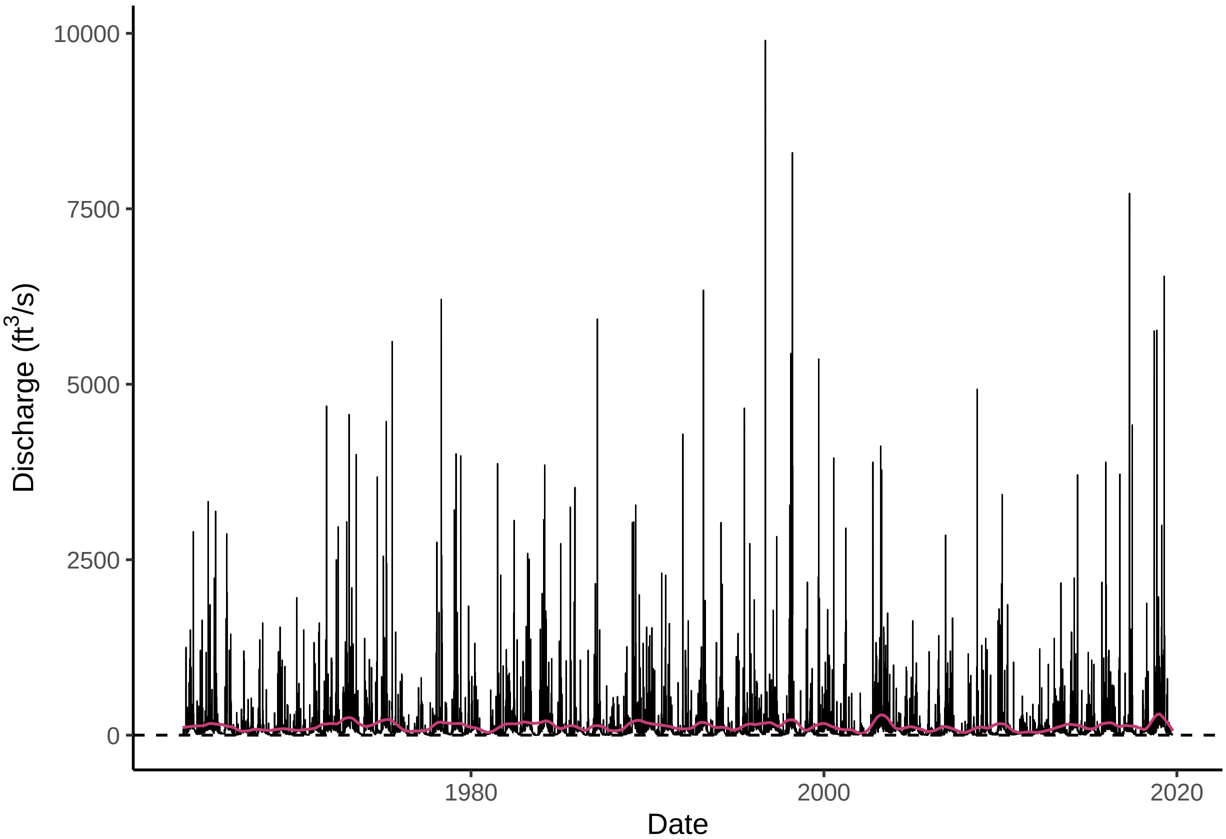
```r
EnoDischargeRegressionPlot <-
  ggplot(EnoDailyMeanDischarge, aes(x = Date, y = Discharge)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE, color = "#c13d75ff") +
  labs(x = "", y = expression("Discharge (ft"^3*"/s)")) +
  ggtitle("Discharge vs. Time with Linear Regression")
print(EnoDischargeRegressionPlot)
```



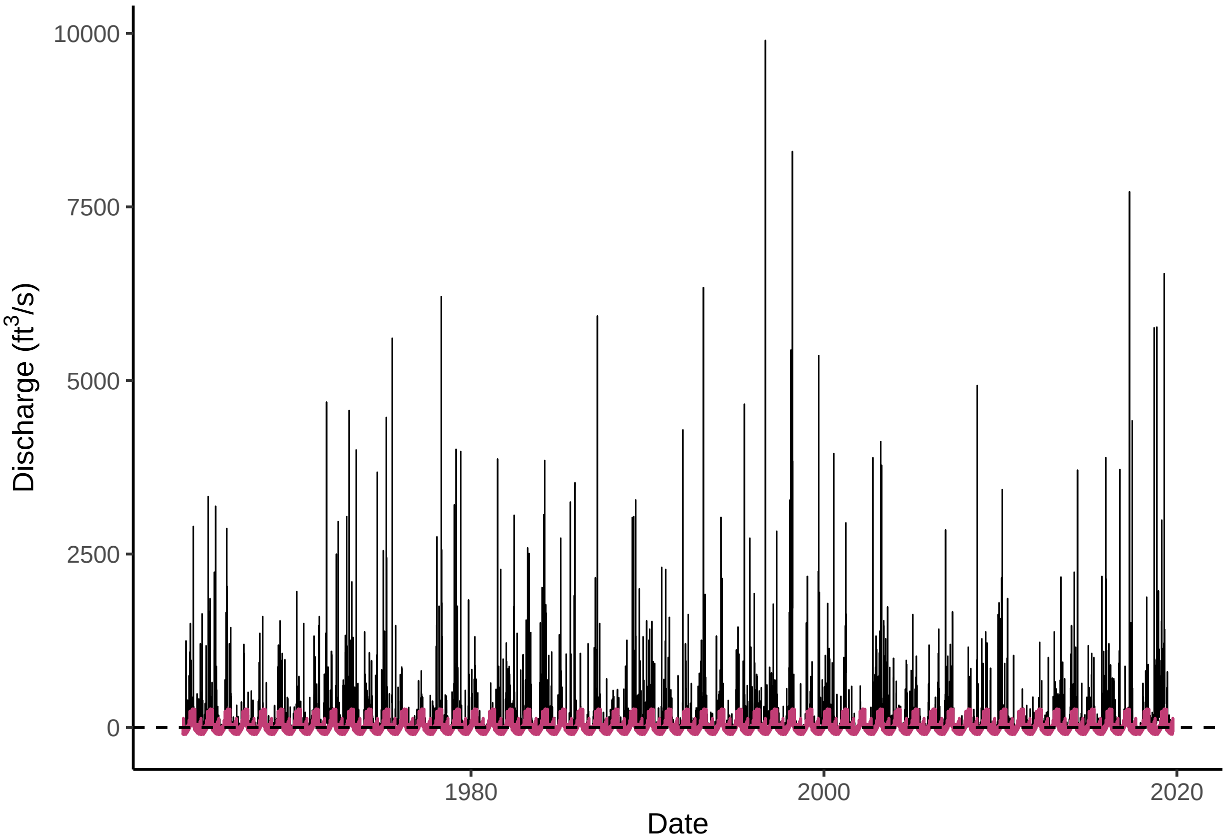

```r
Eno_ts <- ts(EnoDailyMeanDischarge[[4]], frequency = 365)

Eno_Decomposed <- stl(Eno_ts, s.window = "periodic")

plot(Eno_Decomposed)
```




```r
Eno_Components <- as.data.frame(Eno_Decomposed$time.series[,1:3])
Eno_Components <- mutate(Eno_Components,
                         Observed = EnoDailyMeanDischarge$Discharge,
                         Date = EnoDailyMeanDischarge$Date)

ggplot(Eno_Components) +
  geom_line(aes(y = Observed, x = Date),  size = 0.25) +
  geom_line(aes(y = trend, x = Date), color = "#c13d75ff") +
  geom_hline(yintercept = 0, lty = 2) +
  ylab(expression("Discharge (ft"^3*"/s)"))
```




```r
ggplot(Eno_Components) +
  geom_line(aes(y = Observed, x = Date),  size = 0.25) +
  geom_line(aes(y = seasonal, x = Date), color = "#c13d75ff") +
  geom_hline(yintercept = 0, lty = 2) +
  ylab(expression("Discharge (ft"^3*"/s)"))
```



10. How do the seasonal and trend components of the decomposition compare to the Clear Creek discharge dataset? Are they similar in magnitude?

    Seasonal: The range in seasonal discharge values for the Eno dataset is smaller than the range in seasonal discharge values for the Clear Creek dataset. The Eno values span from -100 ft^3/s to +100 ft^3/s, whereas the Clear Creek values range from 0 to 400.

    Trend: The trend components of both the Eno decomposition and the Clear Creek decomposition seem to be similar in magnitude. The Eno data spans from 50 to 300, and the Clear Creek data spans from 50 to 350.

    It is noteworthy that the remainder component of the Eno decomposition is larger than the remainder component of the Clear Creek decomposition. This could explain why the seasonal and trend components are smaller for the Eno decomposition– perhaps a greater proportion of the variability in the Eno data is accounted for by the remainder component rather than the seasonal or trend components.

Trend Analysis

Research question: Has there been a monotonic trend in discharge in Clear Creek over the period of study?

11. Generate a time series of monthly discharge in Clear Creek from the ClearCreekDischarge.Monthly data frame. This time series should include just one column (discharge).
12. Run a Seasonal Mann-Kendall test on the monthly discharge data. Inspect the overall trend and the monthly trends.

```
ClearCreekDischarge <- readNWISdv(siteNumbers = "06719505",
                         parameterCd = "00060", # discharge (ft3/s)
                         startDate = "",
```

```r
                     endDate = "")
names(ClearCreekDischarge)[4:5] <- c("Discharge", "Approval.Code")

ClearCreekDischarge.Monthly <- ClearCreekDischarge %>%
  mutate(Year = year(Date),
         Month = month(Date)) %>%
  group_by(Year, Month) %>%
  summarise(Discharge = mean(Discharge))

ClearCreek_ts <- ts(ClearCreekDischarge.Monthly[[3]], frequency = 12)

CCtrend <- smk.test(ClearCreek_ts)

CCtrend
```

```
##
##  Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data:  ClearCreek_ts
## z = 1.653, p-value = 0.09833
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##      S   varS
##    588 126102
```

```r
summary(CCtrend)
```

```
##
##  Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data: ClearCreek_ts
## alternative hypothesis: two.sided
##
## Statistics for individual seasons
##
## H0
##                      S  varS    tau      z Pr(>|z|)
## Season 1:   S = 0   62 11154  0.060  0.578 0.563546
## Season 2:   S = 0   24 10450  0.024  0.225 0.821984
## Season 3:   S = 0   30 10450  0.030  0.284 0.776650
## Season 4:   S = 0   35 10449  0.035  0.333 0.739425
## Season 5:   S = 0    4 10450  0.004  0.029 0.976588
## Season 6:   S = 0  204 10450  0.206  1.986 0.047054 *
## Season 7:   S = 0  230 10450  0.232  2.240 0.025081 *
## Season 8:   S = 0  148 10450  0.149  1.438 0.150434
## Season 9:   S = 0   94 10450  0.095  0.910 0.362951
## Season 10:   S = 0 -54 10450 -0.055 -0.518 0.604135
## Season 11:   S = 0 -99 10449 -0.100 -0.959 0.337703
## Season 12:   S = 0 -90 10450 -0.091 -0.871 0.383958
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

13. Is there an overall monotonic trend in discharge over time? If so, is it positive or negative?

    There is no significant trend in the data over time. While the z-score is slightly positive, the p-value is greater than 0.05. Therefore, the trend is not significant.

14. Are there any monthly monotonic trends in discharge over time? If so, during which months do they occur and are they positive or negative?

    There are statistically significant trends in discharge over time for the months of June and July. These monthly trends are positive.

Reflection

15. What are 2-3 conclusions or summary points about time series you learned through your analysis?

    1) There are four components to a time series: seasonal variation, a trend over time, an error or remainder component, and sometimes a cyclical component representing an overarching trend on a greater timescale than seasonal variation.
    2) A Mann-Kendall test can be used to determine whether or not there is a monotonic trend in the data over time.
    3) If the data were sampled at erratic and inconsistent rates, some type of interpolation (e.g. pieceweise, linear, spline) can be performed on the data before performing the Mann-Kendall test.

16. What data, visualizations, and/or models supported your conclusions from 12?

    The trend decomposition helped me to understand the different components of the time series. Going through the Mann_Kendall process helped me to understand the interpretation of the trend in the data over time and whether or not it is monotonic.

17. Did hands-on data analysis impact your learning about time series relative to a theory-based lesson? If so, how?

    I believe that having to work through the code encouraged me to understand the components of each of these statistical tools more deeply than if I had just read about them in a textbook.

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

    I think it makes sense that the Eno discharge data had less of a seasonal and overarching trend pattern and more of a random remainder component to its decomposition when compared to the Clear Creek discharge data. This is because the Eno River is located in North Carolina, a state which experiences a wide variety of precipitation events, such as hurricanes and snowfall, that could contribute to variations in discharge. The Clear Creek on the other hand, is located in Colorado, a state which does not experience hurricanes and, thus, would have more consistent precipitation patterns.

    I was surprised that there were not more pronounced monotonic trends in the Clear Creek data, since climate change is occurring and altering natural processes everywhere.