Assignment 6: Time Series Analysis

Gaby Garcia

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.

```
setwd("~/Desktop/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Assignments")
library(tidyverse)
library(lubridate)
library(trend)
library(dataRetrieval)
library(scales)
```

Set GGPlot Theme

```
gabytheme <- theme_bw(base_size = 22) +
  theme(plot.title=element_text(face="bold", size="29", color="IndianRed3", hjust=0.5),
        axis.title=element_text(size=22, color="black"),
axis.text = element_text(face="bold", size=18, color = "black"),
panel.background=element_rect(fill="white", color="darkblue"),
panel.border = element_rect(color = "black", size = 2),
legend.position = "top", legend.background = element_rect(fill="white", color="black"),
            legend.key = element_rect(fill="transparent", color="NA"))

theme_set(gabytheme)
```

Read in CSV

```
setwd("~/Desktop/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Data/Processed")
ClearCreekDischarge.Monthly<-read.csv("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.

Eno River Summary

```
EnoRiverSummary <- whatNWISdata(siteNumbers = "02085070")

EnoRiverDischarge <- readNWISdv(siteNumbers = "02085070",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "", statCd="00003")
```

Rename columns

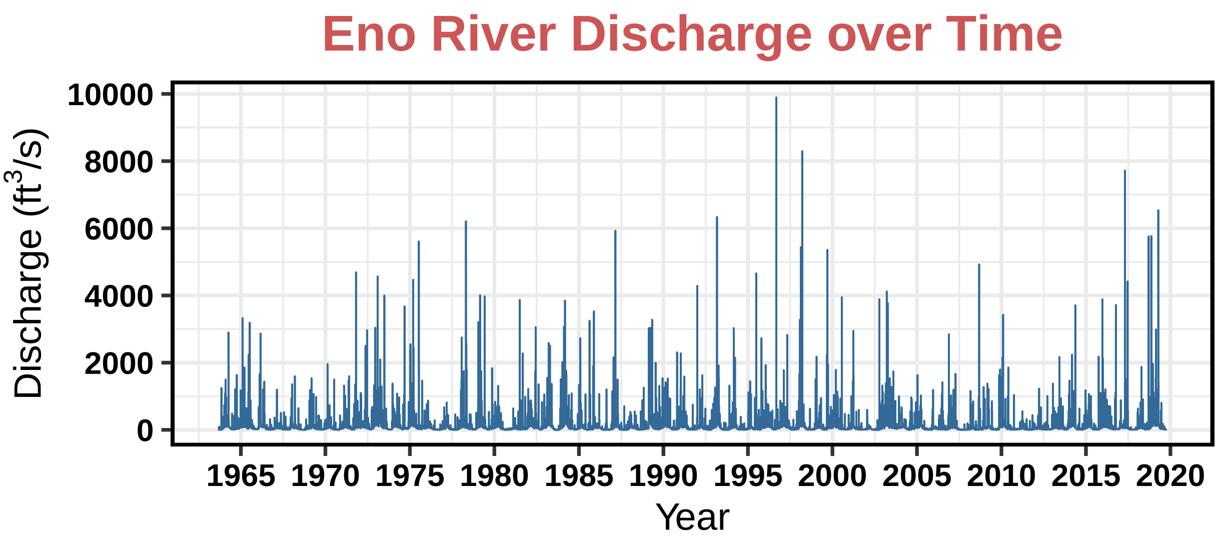
```
names(EnoRiverDischarge)[4:5] <- c("Discharge", "Approval.Code")
class(EnoRiverDischarge$Date)
```

```
## [1] "Date"
```

Eno River Discharge over Time Plot

```
library(ggplot2)
library(scales)
EnoRiverPlot <-
  ggplot(EnoRiverDischarge, aes(x = Date)) +
  geom_line(aes(y=Discharge, color=000000), show.legend=FALSE) +
  labs(title="Eno River Discharge over Time", x = "Year", y = expression("Discharge (ft"^3*"/s)")) +
  theme(plot.title = element_text(margin = margin(b = -10), size = 12),
       axis.title.x = element_blank())+
  scale_y_continuous(breaks=seq(0,10000, by = 2000))+
   scale_x_date(labels = date_format("%Y"), breaks = date_breaks("5 years"))+
  gabytheme

print(EnoRiverPlot)
```



Check for data gaps

```
table(diff(EnoRiverDischarge$Date))
```

```
##
##     1    39
## 20454     1
```

There is one gap with 39 measurements.

Determine Range of Data Gap

```
EnoDataGap<-seq(min(EnoRiverDischarge$Date), max(EnoRiverDischarge$Date), by=1)
EnoDataGap[!EnoDataGap %in% EnoRiverDischarge$Date]
```

```
##  [1] "2017-10-21" "2017-10-22" "2017-10-23" "2017-10-24" "2017-10-25"
##  [6] "2017-10-26" "2017-10-27" "2017-10-28" "2017-10-29" "2017-10-30"
## [11] "2017-10-31" "2017-11-01" "2017-11-02" "2017-11-03" "2017-11-04"
## [16] "2017-11-05" "2017-11-06" "2017-11-07" "2017-11-08" "2017-11-09"
## [21] "2017-11-10" "2017-11-11" "2017-11-12" "2017-11-13" "2017-11-14"
## [26] "2017-11-15" "2017-11-16" "2017-11-17" "2017-11-18" "2017-11-19"
## [31] "2017-11-20" "2017-11-21" "2017-11-22" "2017-11-23" "2017-11-24"
## [36] "2017-11-25" "2017-11-26" "2017-11-27"
```

Data gap between 10/21/2017 and 11/27/2017. Time series objects requires equispaced data, but I will ignore the gap because there is only one.

Create Discharge Time Series

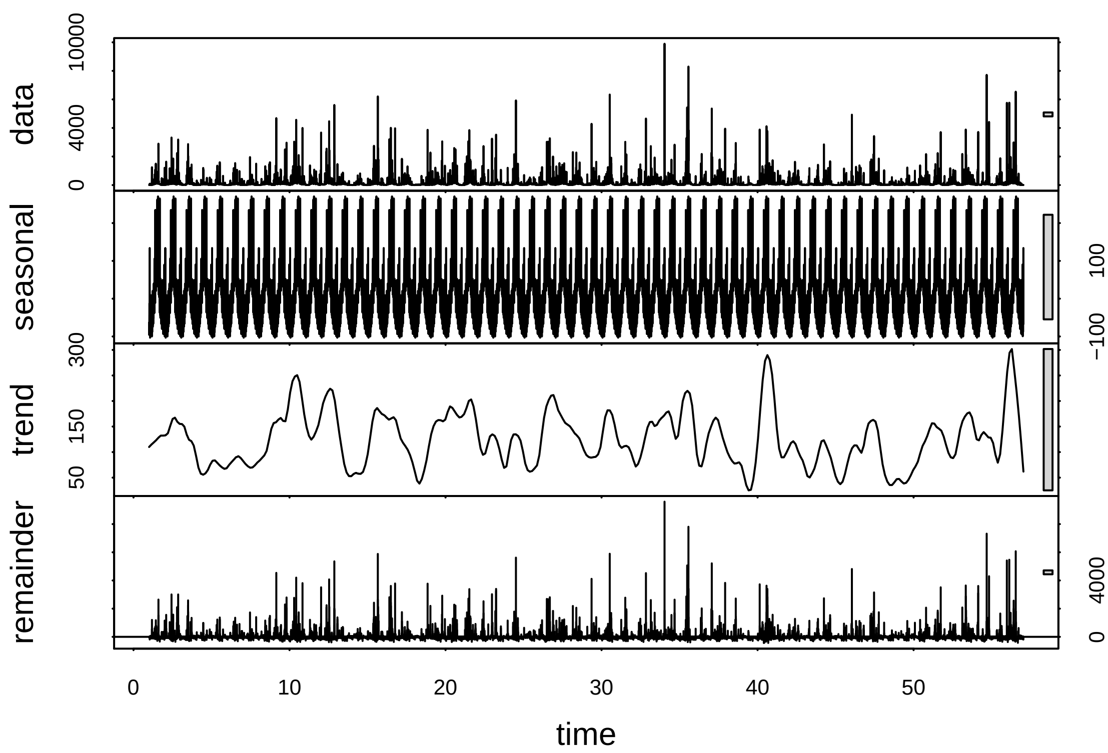
```
EnoRiver_ts <- ts(EnoRiverDischarge[[4]], frequency = 365) #select 4th column-discharge
```

Decompose the Discharge Time Series

```
EnoRiver_Decomposed <- stl(EnoRiver_ts, s.window = "periodic")
```

Visualize the decomposed Discharge Time Series

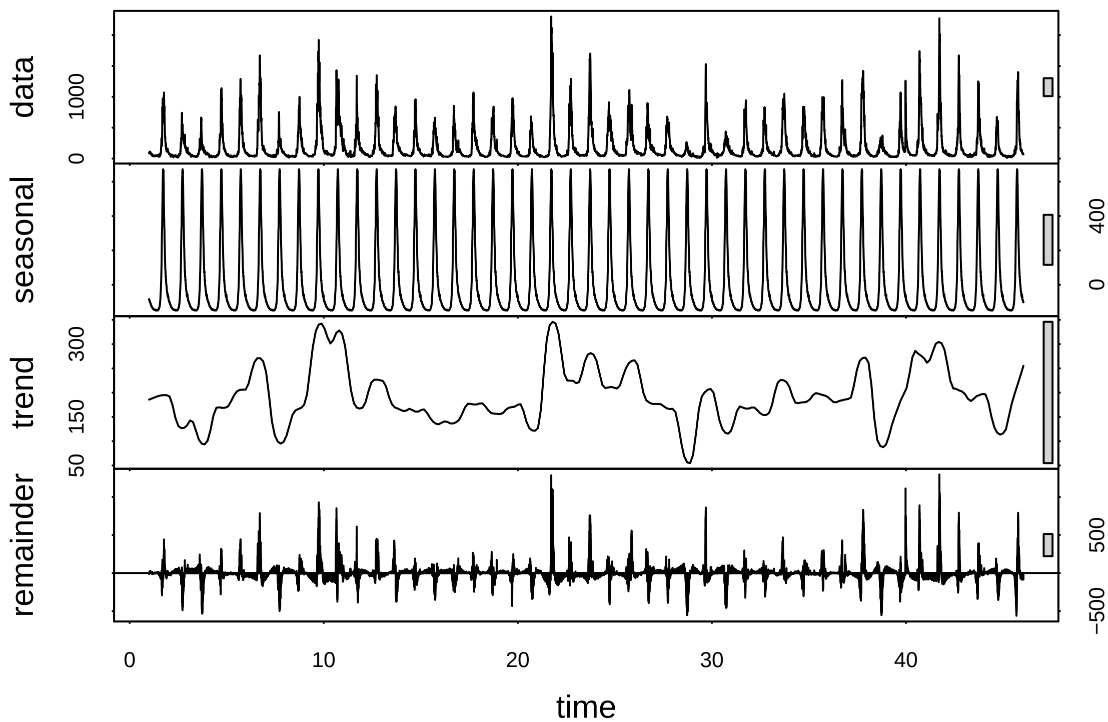
```r
plot(EnoRiver_Decomposed)
```



Revisit Clear Creek Data

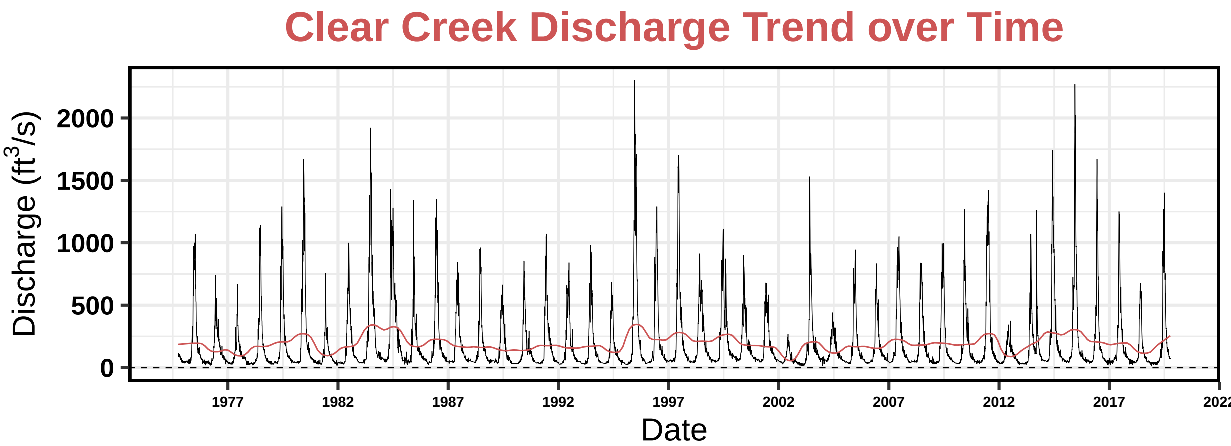
```r
ClearCreekDischarge <- readNWISdv(siteNumbers = "06719505",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")
names(ClearCreekDischarge)[4:5] <- c("Discharge", "Approval.Code")
ClearCreek_ts <- ts(ClearCreekDischarge[[4]], frequency = 365)
ClearCreek_Decomposed <- stl(ClearCreek_ts, s.window = "periodic")
ClearCreek_Components <- as.data.frame(ClearCreek_Decomposed$time.series[,1:3])

ClearCreek_Components <- mutate(ClearCreek_Components,
                      Observed = ClearCreekDischarge$Discharge,
                      Date = ClearCreekDischarge$Date)
plot(ClearCreek_Decomposed)
```



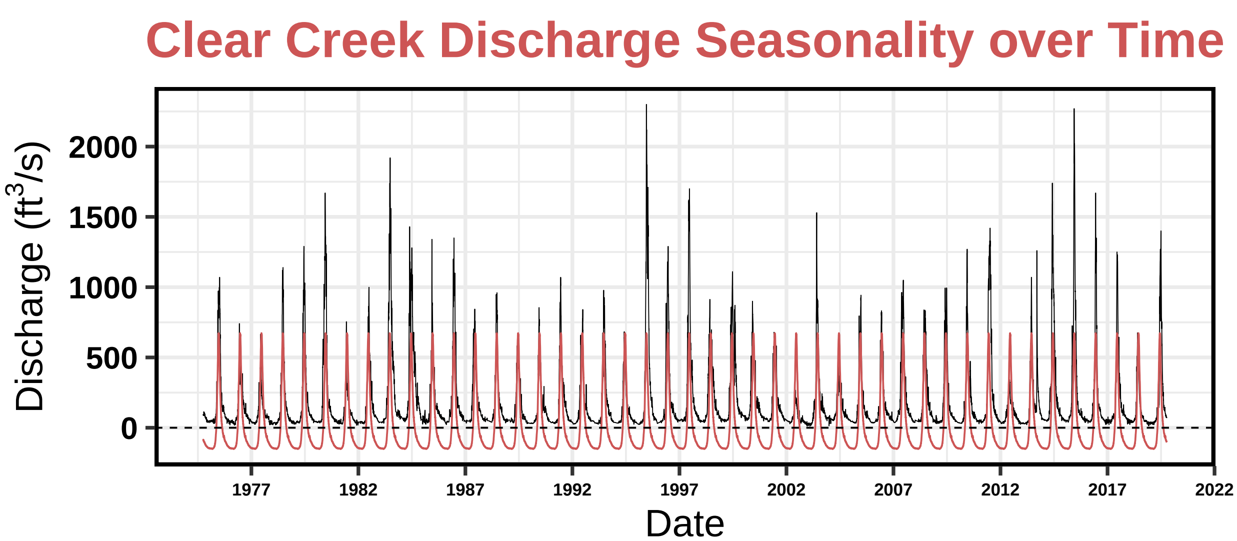
Clear Creek Trend Plot

```
ClearCreek_ComponentsTrendPlot<-ggplot(ClearCreek_Components) +
  geom_line(aes(y = Observed, x = Date),  size = 0.25) +
  geom_line(aes(y = trend, x = Date), color = "#cd5555") +
  geom_hline(yintercept = 0, lty = 2) +
  ylab(expression("Discharge (ft"^3*"/s)"))+
  labs(title="Clear Creek Discharge Trend over Time")+
  scale_x_date(date_breaks="5 years", date_labels="%Y")+
  theme(axis.text.x=element_text(size=10))
print(ClearCreek_ComponentsTrendPlot)
```



Clear Creek Seasonal Plot

```
ClearCreek_ComponentsSeasonalPlot<-ggplot(ClearCreek_Components) +
  geom_line(aes(y = Observed, x = Date),  size = 0.25) +
  geom_line(aes(y = seasonal, x = Date), color = "#cd5555") +
  geom_hline(yintercept = 0, lty = 2) +
  ylab(expression("Discharge (ft"^3*"/s)"))+
  labs(title="Clear Creek Discharge Seasonality over Time")+
  scale_x_date(date_breaks="5 years", date_labels="%Y")+
    theme(axis.text.x=element_text(size=10))
print(ClearCreek_ComponentsSeasonalPlot)
```



Extract Eno River Components

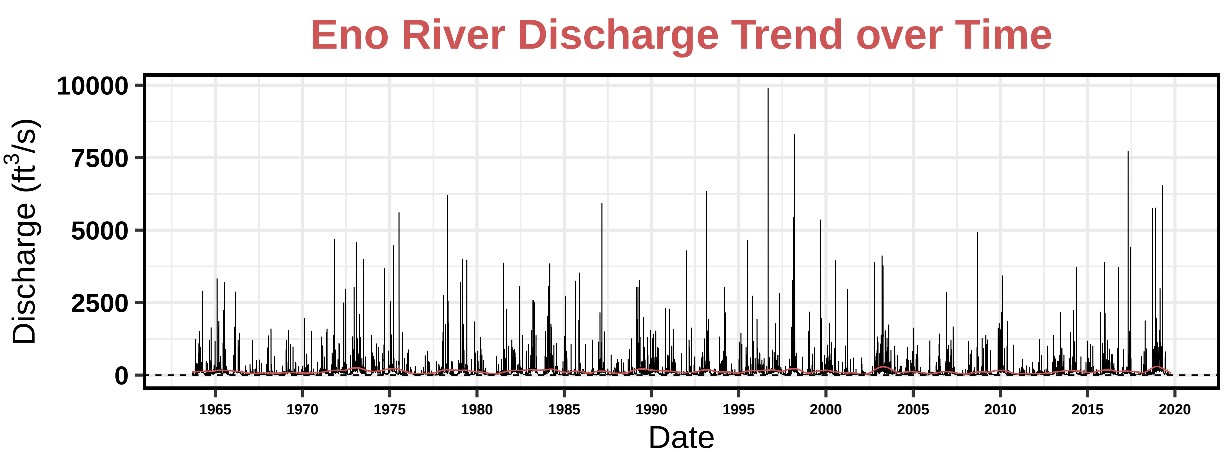
```
EnoRiver_Components <- as.data.frame(EnoRiver_Decomposed$time.series[,1:3])
```

```
EnoRiver_Components <- mutate(EnoRiver_Components,
                              Observed = EnoRiverDischarge$Discharge,
                              Date = EnoRiverDischarge$Date)
```

Observed=discharge

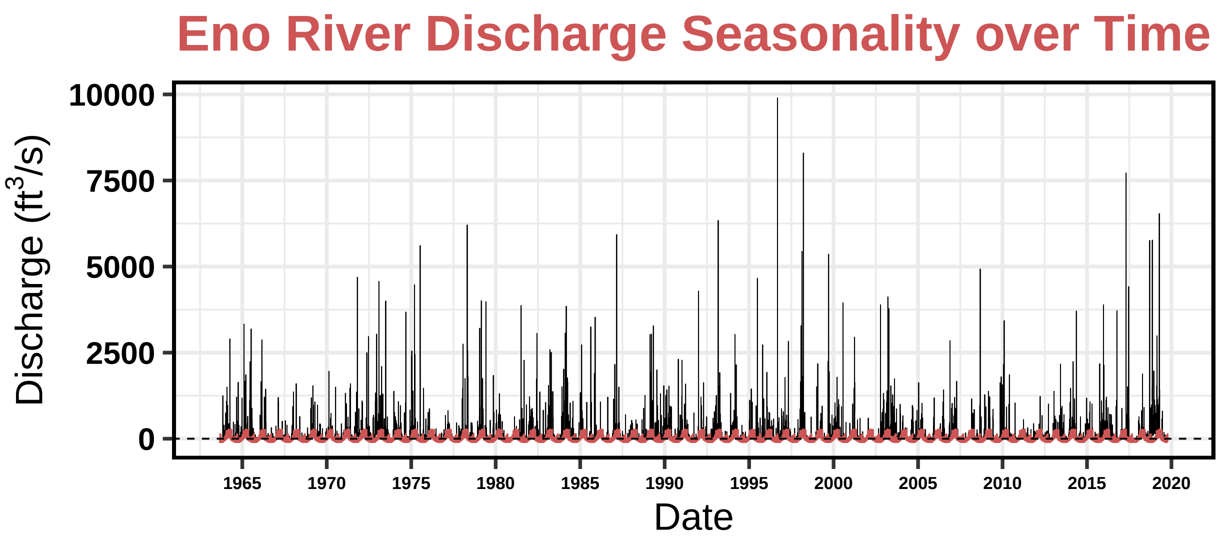
Eno River Trend Plot

```
EnoRiver_ComponentsTrendPlot<-ggplot(EnoRiver_Components) +
  geom_line(aes(y = Observed, x = Date),  size = 0.25) +
  geom_line(aes(y = trend, x = Date), color = "#cd5555") +
  geom_hline(yintercept = 0, lty = 2) +
  ylab(expression("Discharge (ft"^3*"/s)"))+
  labs(title="Eno River Discharge Trend over Time")+
  scale_x_date(date_breaks="5 years", date_labels="%Y")+
   theme(axis.text.x=element_text(size=10))
print(EnoRiver_ComponentsTrendPlot)
```



Eno River Seasonality Plot

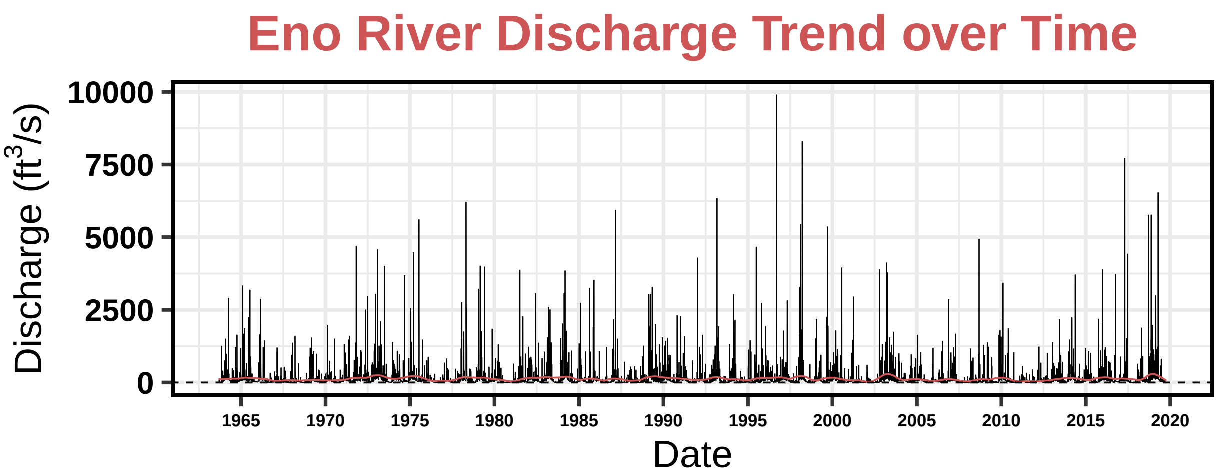
```
EnoRiver_ComponentsSeasonalPlot<-ggplot(EnoRiver_Components) +
  geom_line(aes(y = Observed, x = Date),  size = 0.25) +
  geom_line(aes(y = seasonal, x = Date), color = "#cd5555") +
  geom_hline(yintercept = 0, lty = 2) +
  ylab(expression("Discharge (ft"^3*"/s)"))+
  labs(title="Eno River Discharge Seasonality over Time")+
     theme(axis.text.x=element_text(size=10))+
  scale_x_date(date_breaks="5 years", date_labels="%Y")
print(EnoRiver_ComponentsSeasonalPlot)
```

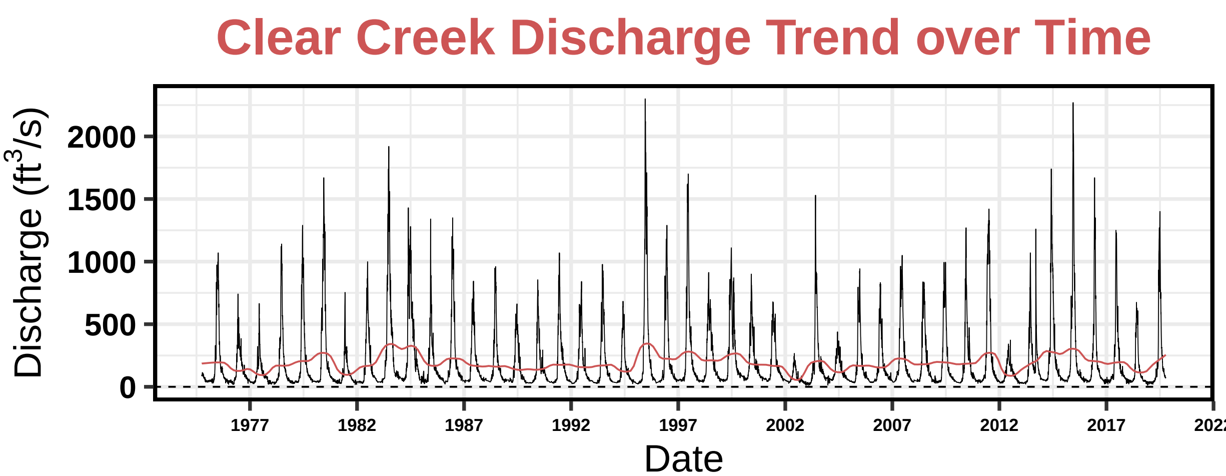


10. How do the seasonal and trend components of the decomposition compare to the Clear Creek discharge dataset? Are they similar in magnitude?

Compare Trend Plots of Eno River and Clear Creek Using Plot_Grid function

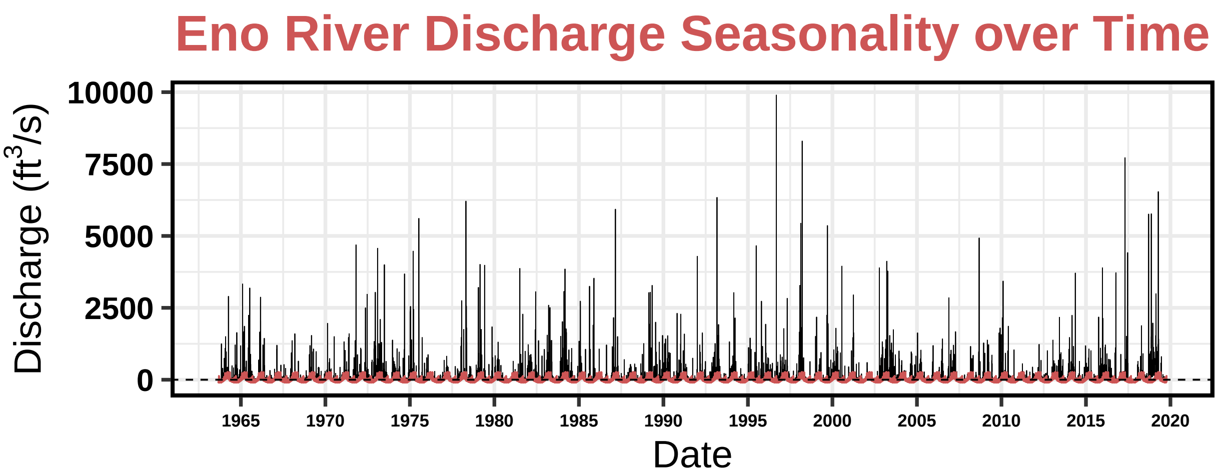
```
library(cowplot)
TrendPlotGrid<-plot_grid(EnoRiver_ComponentsTrendPlot, ClearCreek_ComponentsTrendPlot,
            nrow = 2)
print(TrendPlotGrid)
```

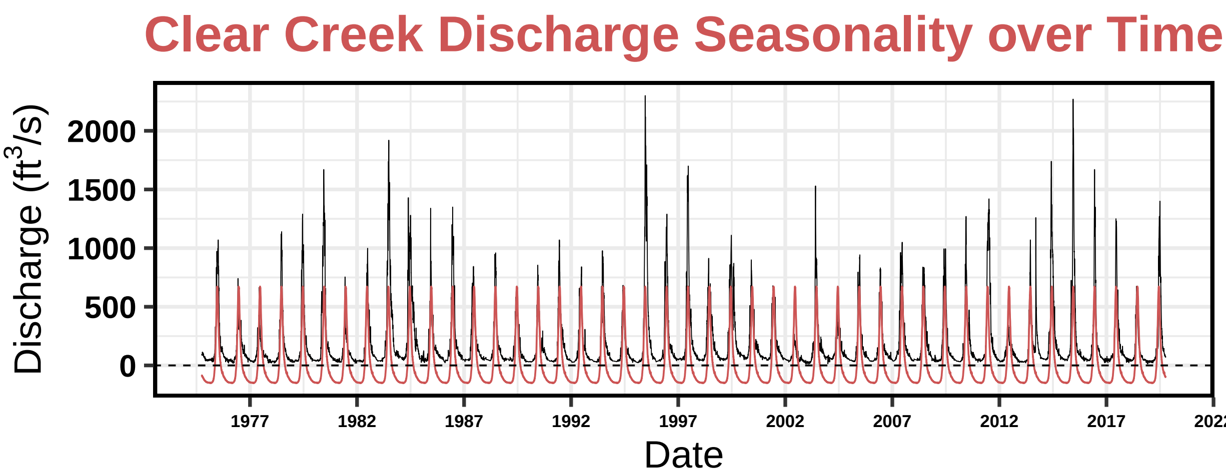




Compare Seasonality Plots of Eno River and Clear Creek using Plot_Grid Function

```
library(cowplot)
SeasonalityPlotGrid<-plot_grid(EnoRiver_ComponentsSeasonalPlot, ClearCreek_ComponentsSeasonalPlot,
            nrow = 2)
print(SeasonalityPlotGrid)
```





Seasonal: There is visible seasonality in both the Eno River and Clear Creek datasets. While the Clear Creek data features one large peak around the springtime of each year (presumably the effect of snowmelt entering the river during the late springtime and increasing river discharge), the Eno River data actually has two seasonal peaks. There is a peak of larger magnitude also around the first few months of each year, but there is a peak of smaller magnitude in between each larger peak, presumably around the spring/summertime. The seasonal component of the Eno River dataset is also more variable, likely because Eno River depends on precipitation in the form of rain, which can be more variable than Clear Creek (which is presumably fed by precipitation in the form of snow).

Trend: For both the Eno River and Clear Creek datasets, the trend peaks correspond to high river discharge events. It is worth noting that Eno River has a larger range in magnitude of river discharge (0.08-9900 cfs) than Clear Creek (12.0-2300 cfs), yet Eno River's trend line has peaks of smaller magnitude than the Clear Creek trend line. The Eno River has more data points than the Clear Creek dataset. Both trend lines for Eno River and Clear Creek are always positive, although Eno River's trend line remains close to 0.

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).

Create Clear Creek Monthly Discharge Time Series

```
ClearCreekMonthly_ts <- ts(ClearCreekDischarge.Monthly[[3]], start=c(1974, 10), frequency = 12)
```

The index of three is to select the third column–>discharge The first part of the start argument is the first year, the second part is the month)

12. Run a Seasonal Mann-Kendall test on the monthly discharge data. Inspect the overall trend and the monthly trends. >A Mann-Kendall test will analyze whether there is a monotonic trend (consistent increase or decrease but not necessarily linear) in the response variable (discharge) over time.

Run Seasonal Mann Kendall Test

```
library(trend)
ClearCreekMKtrend<-smk.test(ClearCreekMonthly_ts)
```

Inspect results of Mann Kendall Test

```
summary(ClearCreekMKtrend)
```

```
##
##  Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data: ClearCreekMonthly_ts
## alternative hypothesis: two.sided
##
## Statistics for individual seasons
##
## H0
##                       S  varS    tau      z Pr(>|z|)
## Season 1:   S = 0   35 10449  0.035  0.333 0.739425
## Season 2:   S = 0    4 10450  0.004  0.029 0.976588
## Season 3:   S = 0  204 10450  0.206  1.986 0.047054 *
## Season 4:   S = 0  230 10450  0.232  2.240 0.025081 *
## Season 5:   S = 0  148 10450  0.149  1.438 0.150434
## Season 6:   S = 0   94 10450  0.095  0.910 0.362951
## Season 7:   S = 0  -54 10450 -0.055 -0.518 0.604135
## Season 8:   S = 0  -99 10449 -0.100 -0.959 0.337703
## Season 9:   S = 0  -90 10450 -0.091 -0.871 0.383958
## Season 10:   S = 0  64 11154  0.062  0.597 0.550828
## Season 11:   S = 0  24 10450  0.024  0.225 0.821984
## Season 12:   S = 0  30 10450  0.030  0.284 0.776650
## ---
## 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?

According to the Mann-Kendall test, we fail to reject the null hypothesis that there is no monotonic trend in the Clear Creek discharge over time (Mann-Kendall test, Z=4.39, p=0.09719). Because the resultant p-value is >0.05, we can't conclude that there is a monotonic trend in Clear Creek discharge over time.

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?

    According to my summary of the seasonal Mann-Kendall test, we reject the null hypothesis that there is no monotonic trend in the Clear Creek discharge over time, and therefore there is a monotonic trend for both March (Mann-Kendall Test, z=1.986, $p=0.047<0.05$) and April (Mann-Kendall Test, z=2.240, $p=0.025<0.05$). The fact that these months are the most significant correlates to seasonal snowpack melting in Colorado and the resulting discharge into the watershed.

Reflection

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

    Time series graphs are valuable because the variances in data, when collected and compared from year to year, can reveal seasonal fluctuation patterns that can serve as the basis for future forecasting. I also learned that there is unexplained variability (the remainder) that wasn't explained by the seasonal and trend components.

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

    The decomposed plots allowed me to visualize the different components of a time series.

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

    Yes, I was able to use real-world data to compare seasonality patterns of different rivers in different parts of the country, and substantiate the theoretical lessons that we covered in class.

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

    The real-world data is reflective of my theoretical expectations because as expected, different rivers in different geographic locations in the United States exhibit seasonality differently. I also expected to see a greater significance in discharge in the months of March and April due to my knowledge of Colorado and its regular snowmelt season.