Assignment 6: Time Series Analysis

Rachel Bash

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] "C:/Users/19524/Documents/DUKE/Hydrologic Data Analytics/Hydrologic_Data_Analysis/Assignments"
```

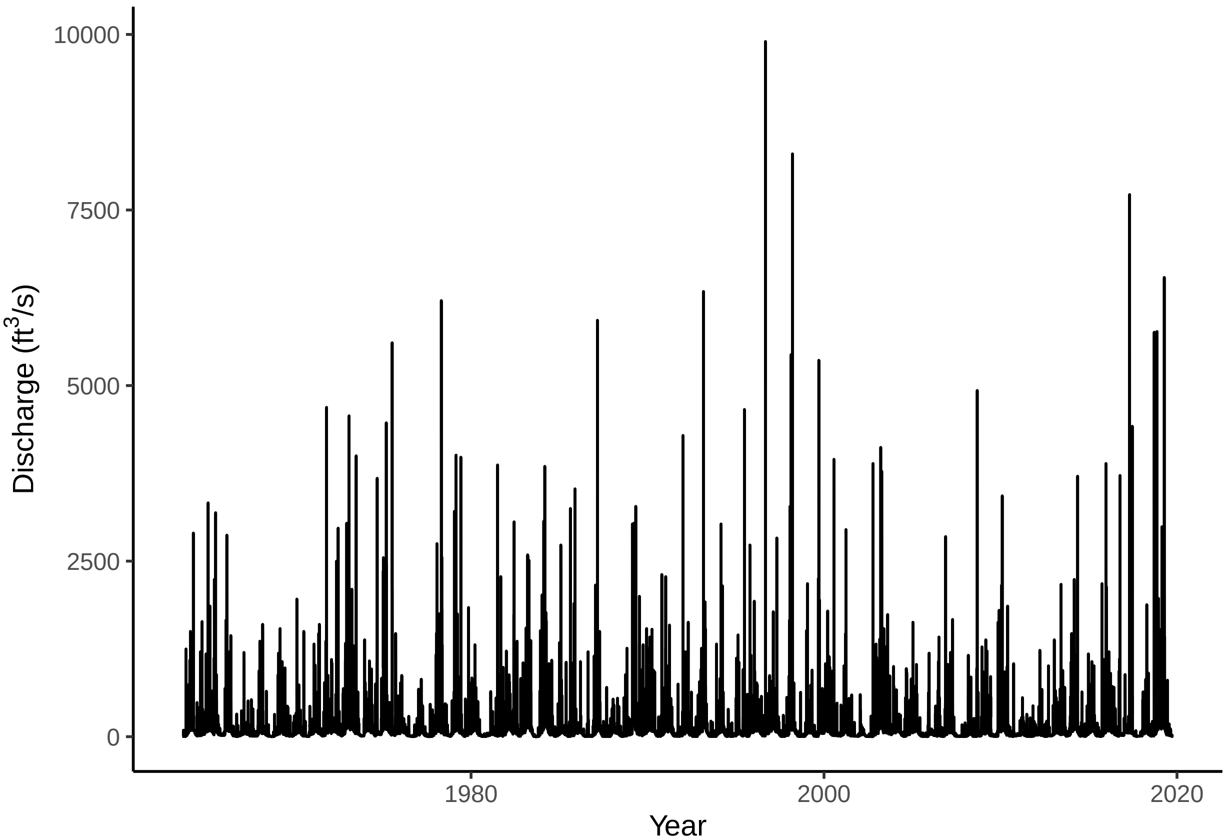
```
library(tidyverse)
library(lubridate)
library(dataRetrieval)
library(trend)

theme_set(theme_classic())

ClearCreekDischarge.Monthly <-
  read.csv("../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.

```
Enodischarge <- readNWISdv(siteNumbers = "02085070",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")
names(Enodischarge)[4:5] <- c("Discharge", "Approval.Code")
```

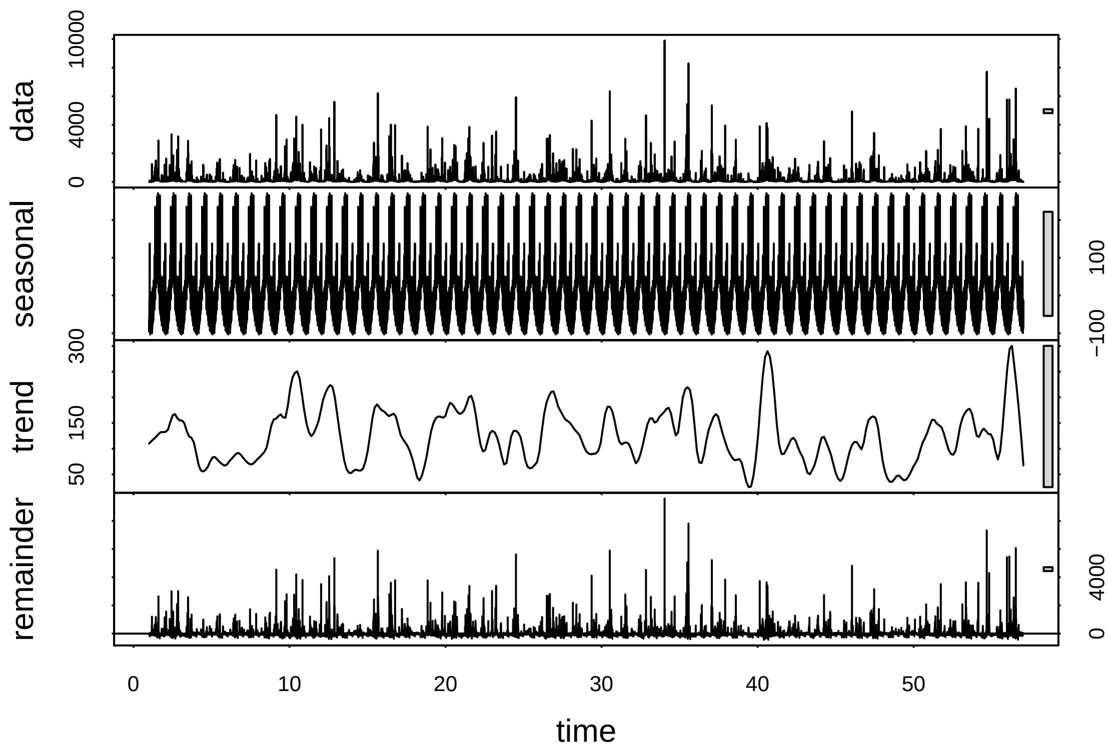
```r
Enodischarge.plot <- ggplot(data=Enodischarge, aes(x=Date, y= Discharge)) +
  geom_line(aes(x=Date, y= Discharge)) +
  labs(x = "Year", y = expression("Discharge (ft"^3*"/s)"))
print(Enodischarge.plot)
```




```r
table(diff(Enodischarge$Date)) #there is technically a gap in the daily data,
```

```
##
##     1    39
## 20452     1
```

```r
#it looks like there is only one, so I will ignore it

Eno.ts <- ts(Enodischarge[[4]], frequency=365)
Eno.decomposed <- stl(Eno.ts, s.window = "periodic")
plot(Eno.decomposed)
```



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 seasonal component of the Eno discharge dataset plots a lot "thicker" meaning that there is more variability in how the seasonality of the river exhibits itself. This is likely because this system is rain-dominated, which is more variable than Clear Creek, which is snow-dominated, and has a much clearer seasonal pattern where snow melts in late spring.

    Trend: Clear Creek and Eno's trend components look pretty similar, with varied upward and downward trends throughout the dataset. Clear Creek has a bigger y-axis range than the Eno trend y-axis, perhaps meaning that there are slightly bigger trends throughout the date range.

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.

```
ClearCreek.ts <- ts(ClearCreekDischarge.Monthly$Discharge, frequency = 12,
                        start = c(1975,10,1), end = c(2019,10,1))

ClearCreektrend <- smk.test(ClearCreek.ts)

ClearCreektrend
```

```
##
```

```
##  Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data:  ClearCreek.ts
## z = 2.2418, p-value = 0.02498
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##        S     varS
##    771.0 117975.7
```

```r
summary(ClearCreektrend)
```

```
##
##  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    63  9774.3  0.067  0.627 0.530583
## Season 2:   S = 0    28  9775.3  0.030  0.273 0.784788
## Season 3:   S = 0   224  9775.3  0.237  2.255 0.024103 *
## Season 4:   S = 0   242  9775.3  0.256  2.438 0.014788 *
## Season 5:   S = 0   190  9775.3  0.201  1.912 0.055928 .
## Season 6:   S = 0    80  9775.3  0.085  0.799 0.424275
## Season 7:   S = 0   -90  9775.3 -0.095 -0.900 0.368030
## Season 8:   S = 0  -115  9774.3 -0.122 -1.153 0.248876
## Season 9:   S = 0   -74  9775.3 -0.078 -0.738 0.460307
## Season 10:   S = 0   85 10449.0  0.086  0.822 0.411217
## Season 11:   S = 0   66  9775.3  0.070  0.657 0.510906
## Season 12:   S = 0   72  9775.3  0.076  0.718 0.472688
## ---
## 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?

    Yes, there is a significant overall monotonic trend in discharge over time. It is a positive trend, meaning discharge is significantly increasing over time. We can tell by the z-score, which quantifies the direction and magnitude of the trend.

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?

    Yes, there are two months that have significant monotonic trends – March and April. Both are positive trends. This makes sense that these are the two significant months because these are the months when the most discharge is likely occuring due to the melting snowpack.

Reflection

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

    There is a difference between seasonal and trend components over time, and there is also inherent unexplained variability that cannot be explained by the seasonal or trend components. I also learned that you cannot only determine trends with linear regression, especially with datasets with a lot of years.

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

    The plot that decomposed the time series helped me make sense of time series and the different components of it.

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

    Yes, I understand timeseries better since I've done one myself, rather than reading from a book about it.

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

    I liked seeing the real-world data and understanding that all systems are different. Many rivers may have seasonality, but they exhibit seasonality differently from each other.