Assignment 6: Time Series Analysis

Keqi He

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/keqi/Desktop/Hydrologic_Data_Analysis"
```

```
#install.packages("tidyverse")
#install.packages("lubridate")
#install.packages("trend")
#install.packages("dataRetrieval")

packages <- c("tidyverse",
              "lubridate",
              "trend",
              "dataRetrieval")
invisible(lapply(packages, library, character.only = TRUE))
```

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

## 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 ------------------------------------------------------------------------- tid
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

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

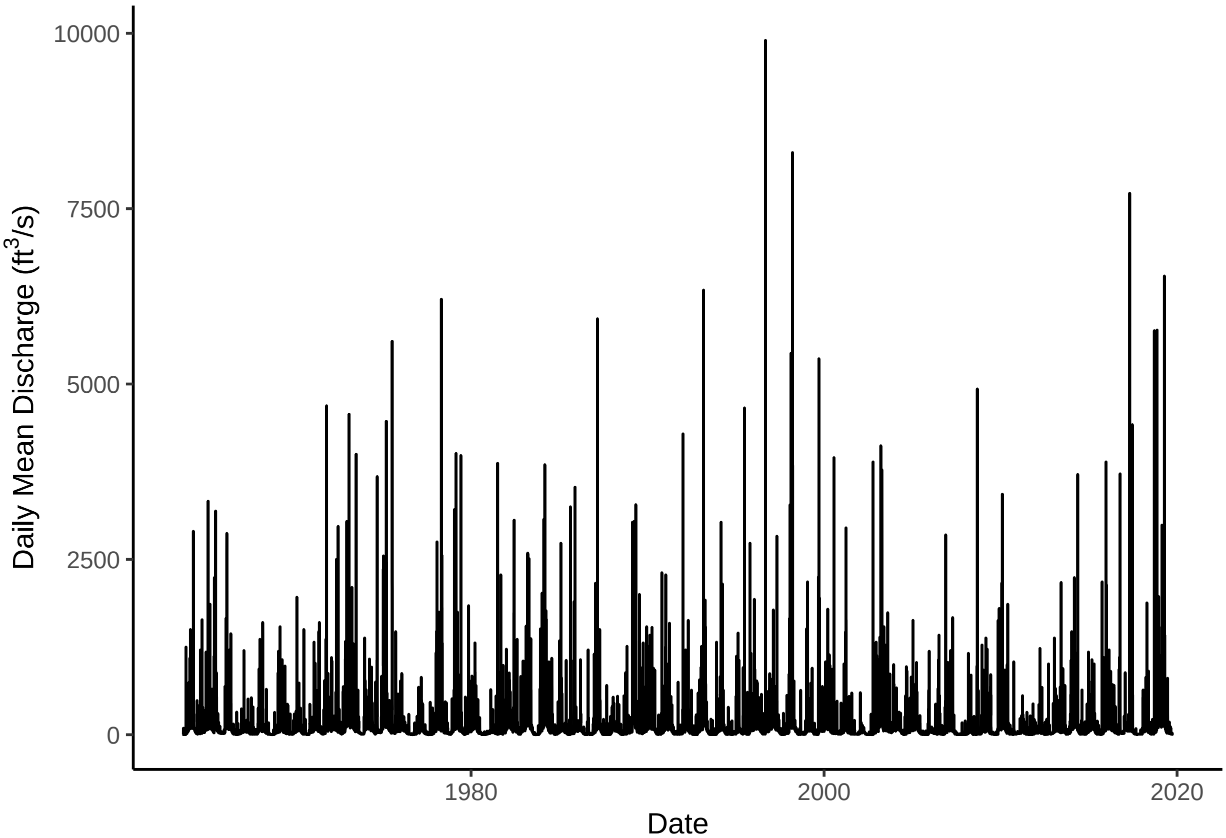
```
## The following object is masked from 'package:base':
##
##     date
```

```r
theme_set(theme_classic())

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.

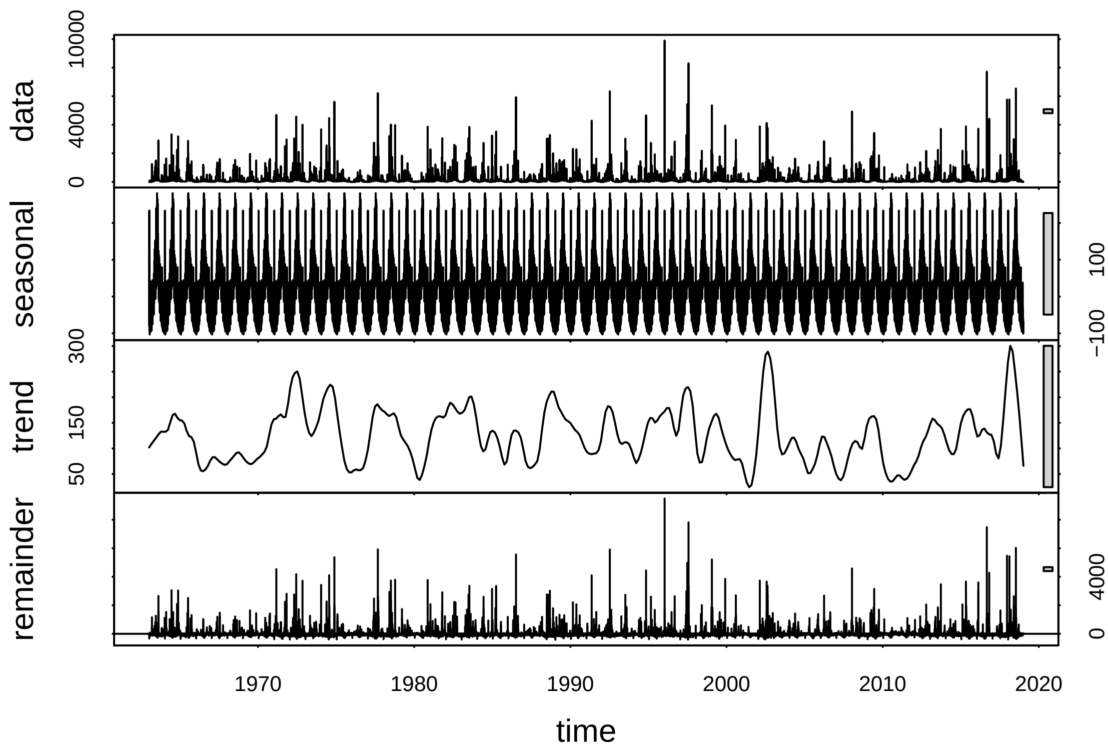
```r
EnoDischarge <- readNWISdv(siteNumbers = "02085070",
                     parameterCd = "00060", # discharge (ft3/s)
                     startDate = "",
                     endDate = "")
names(EnoDischarge)[4:5] <- c("Daily.Mean.Discharge", "Approval.Code")

EnoDischargePlot <-
  ggplot(EnoDischarge,aes(x = Date, y = Daily.Mean.Discharge)) +
  geom_line() +
  labs(y = expression("Daily Mean Discharge (ft"^3*"/s)"))
print(EnoDischargePlot)
```




```r
#install.packages("zoo")
library(zoo)
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
z <- zooreg(EnoDischarge, start = as.Date("1963-09-01"))
# remove Feb 29's
zz <- z[format(time(z), "%m %d") != "02 29"]
Enotimeseries <- ts(zz$Daily.Mean.Discharge, frequency = 365,
                        start = c(1963, 9, 1))

Enotimeseriesstl <- stl(Enotimeseries, s.window='periodic')

plot(Enotimeseriesstl)
```



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 decomposition of the Eno River discharge dataset is not as smooth as that of the Clear Creek discharge. But they are both periodic. The peaks of discharge both occur in summer. However they are not similar in magnitude. The peak of discharge of the Clear Creek is much greater than that of the Eno River.

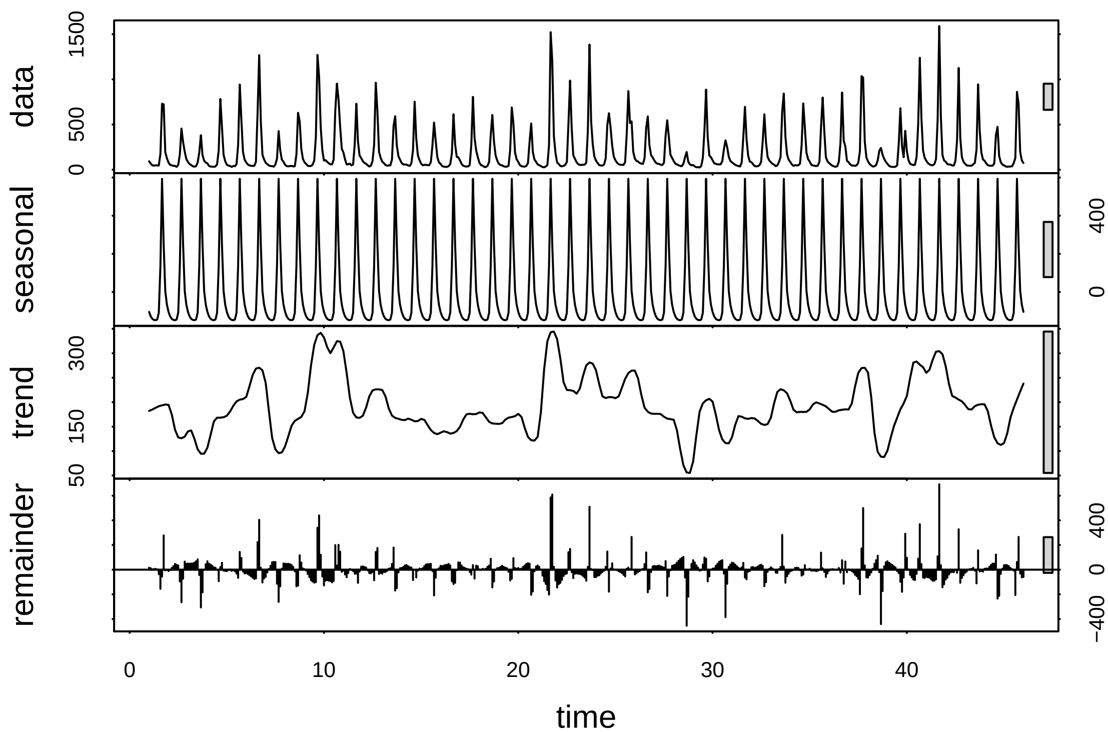
    Trend: The trend component of the decomposition of the Eno River discharge dataset is more volatile than that of the Clear Creek discharge. However, the magnitude of the trend of the Eno River discharge is smaller than that of the Clear Creek discharge.

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[[3]], frequency = 12)
ClearCreektimeseriesstl <- stl(ClearCreek_ts, s.window='periodic')
plot(ClearCreektimeseriesstl)
```




```r
# Run SMK test
ClearCreektrend <- smk.test(ClearCreek_ts)

# Inspect results
ClearCreektrend
```

```
##
##  Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data:  ClearCreek_ts
## z = 1.6586, p-value = 0.09719
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##      S   varS
##    590 126102
```

```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   64 11154  0.062  0.597 0.550828
## 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?

    No, there is not an overall monotonic trend in 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?

    Yes, positive trend occurs during June and July.

Reflection

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

    Time series analysis is very useful in analyzing the trend of variables of rivers or lakes. In addition, we can use the time series model to forecast the conditions in the future. However, the dataset is often discontinuous. Therefore, we should use different types of interpolation to make it continuous.

16. What data, visualizations, and/or models supported your conclusions from 15?

    Data used includes discharge data from Clear Creek in Colorado, USA, total phosphorus data from Lake Mendota. Ggplot, acf, pacf, stl function, ARMA models are used to draw those conclusions.

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

    Yes. Hands-on data analysis convinced me that time series analysis is very useful in the research of rivers or lakes. Also it can help me understand the theory and master some useful tools of time series analysis better.

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

    There are quite big differences. Real-world data is not perfect. It can have some gaps and not continuous, which is a challenge for time series analysis. But in general, real-world data has a theoretical law.