

STAT 443: Lab 1

Caden Hewlett (Student #40404246)

12 January, 2024

Question 1

```
# read in data to an object called dat  
dat <- read.csv("LakeLevels.csv")  
head(dat, 10)
```

```
##      Date LakeLevel  
## 1  1/1/2007   3732.65  
## 2  1/2/2007   3732.65  
## 3  1/3/2007   3732.65  
## 4  1/4/2007   3732.64  
## 5  1/5/2007   3732.64  
## 6  1/6/2007   3732.64  
## 7  1/7/2007   3732.64  
## 8  1/8/2007   3732.64  
## 9  1/9/2007   3732.64  
## 10 1/10/2007  3732.64
```

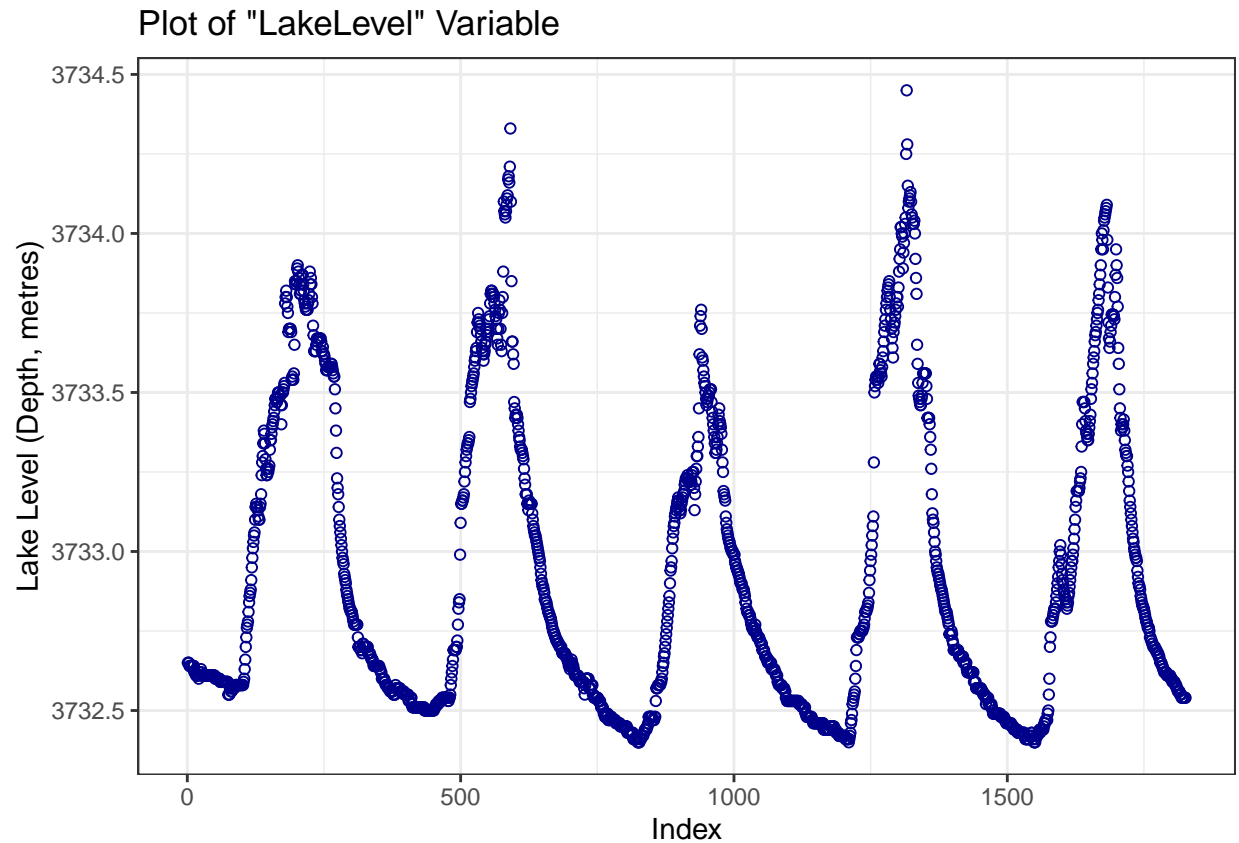
Then, we can check the `names` of this data frame.

```
names(dat)
```

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

Now, we create a plot of the `LakeLevel` variable.

```
ggplot(dat, aes(x = 1:nrow(dat), y = LakeLevel)) +  
  geom_point(shape = 1, col = "darkblue") +  
  xlab("Index") +  
  ylab("Lake Level (Depth, metres)") +  
  ggtitle("Plot of \"LakeLevel\" Variable") +  
  theme_bw()
```



This plot is different from what we want for a time series because the value on the x -axis is the Index of the Lake Level data points in the `dat` dataframe rather than time. Hence, this is not a time series (yet!)

Question 2

Now, we can test if this is a time series object.

```
is.ts(dat)
```

```
## [1] FALSE
```

Since it isn't, we'll convert it.

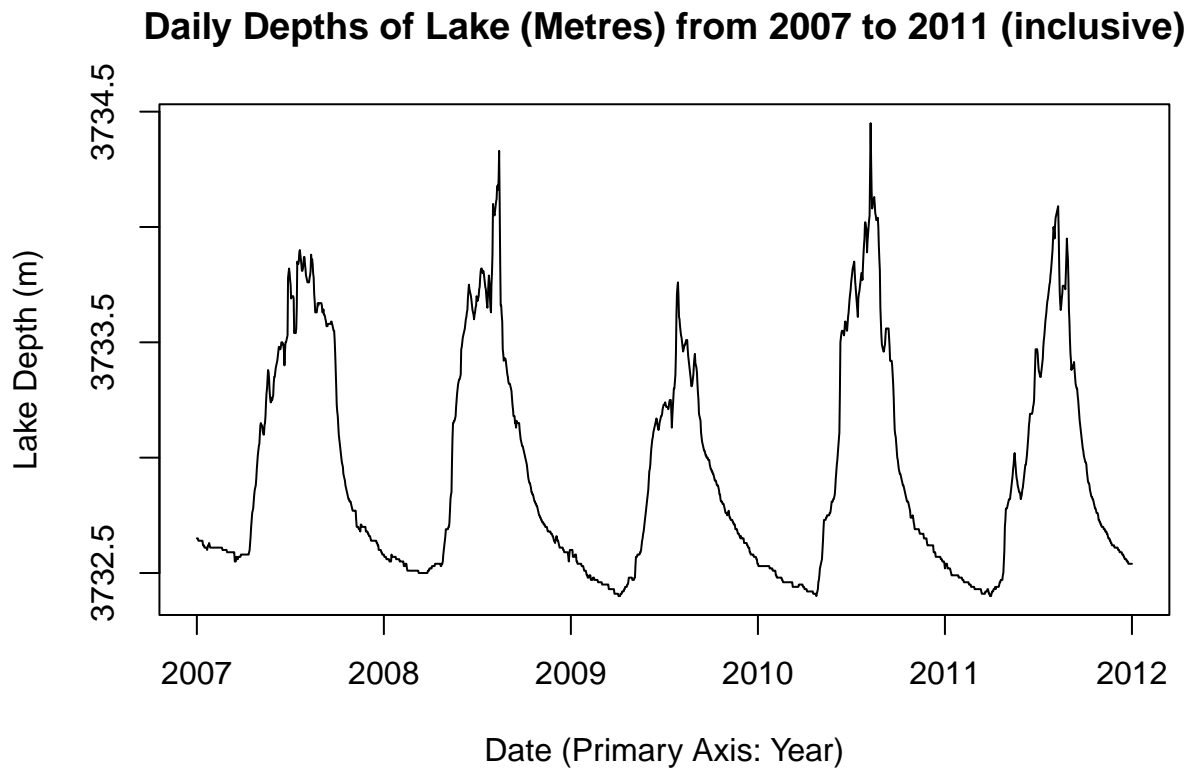
```
# convert to ts
x <- ts(dat$LakeLevel,
        start = c(2007, 1),
        end = c(2012, 1), # note: (2011, 12) omitted December
        frequency = 365)

# check that we haven't lost any data
length(x) == nrow(dat)
```

```
## [1] TRUE
```

Question 3

```
plot(x,  
     main = "Daily Depths of Lake (Metres) from 2007 to 2011 (inclusive)",  
     ylab = "Lake Depth (m)",  
     xlab = "Date (Primary Axis: Year)"  
)
```



We see a few key differences in this plot compared to the one in Question 1. Firstly, it automatically forms a continuous line graph rather than a discrete scatter plot, which is better for interpretation in the context of time series. Secondly, we see that the date (with year as the primary axis label) is now on the x -axis, rather than simply the `Index` we saw before.

Question 4

Comment on the main features of this time series and provide physical justification(s) for what you observe.

- Firstly, we see that there is a noticeable seasonal component to this time series. Simply from observation, it seems as if the depths follow a yearly period, peaking at about mid-year. This makes sense based on the longer-term cycle of Earth's seasons.
- As a physical justification, if we're in the Northern Hemisphere, we can expect more snow melt in the Spring/Summer, which may result in higher water levels. This is reflected in the changing amplitude of the lake depth over the course of the year - specifically, the "peaks" at around mid-year.

- Further, we see in the colder months that the lake depth tends to decrease. This could be due to the formation of ice on the surface of the water, decreasing the amount of water in the lake (and hence the depth overall.)
- Secondly, there doesn't appear to be a significant trend component to this time series. If there were a trend component, we would see the data increasing or decreasing overall over time. From rudimentary observation, such a trend doesn't seem present. However, it would be interesting to observe more data on this particular lake beyond the time frame we've been provided, to inspect if there is a climate-change based larger trend to the water levels. Unfortunately, we don't have a large enough time frame to observe such a gradual effect.