STAT 443: Lab 1

Wenxuan Zan (61336194)

13 January, 2022

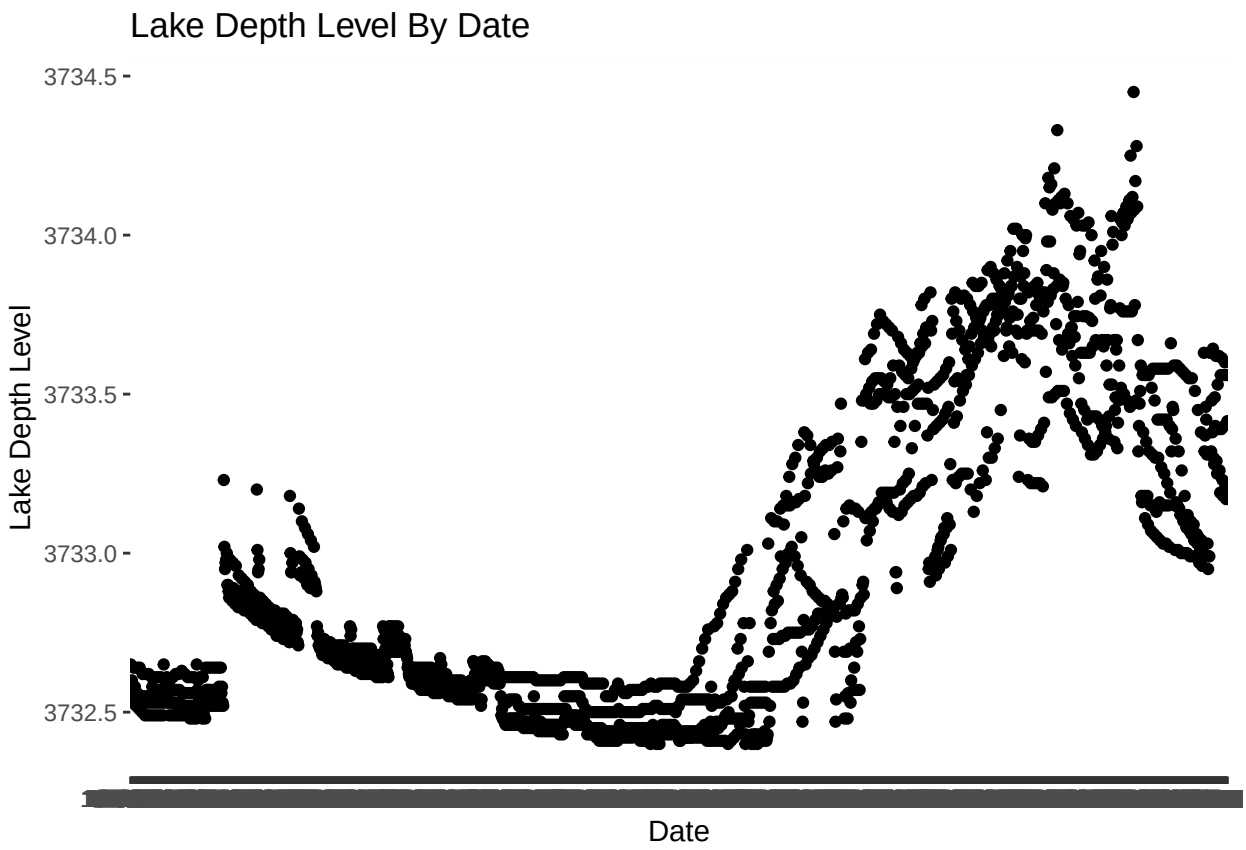
Question 1

(a)

```
data <- read.csv("LakeLevels.csv", header = TRUE)
head(data,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
```

```
data %>%
    ggplot(aes(x = Date, y = LakeLevel)) +
    geom_point() +
    labs(x = "Date",
         y = "Lake Depth Level",
         title = "Lake Depth Level By Date")
```

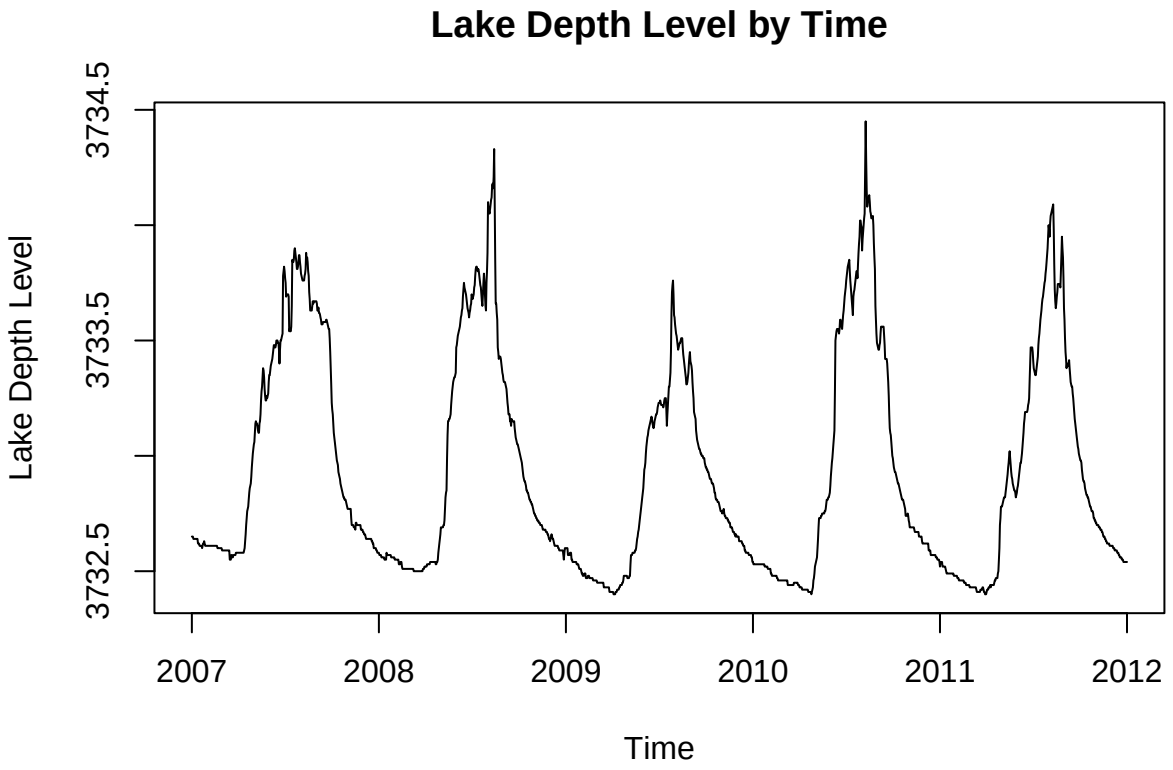


(b)

```
x <- ts(data = data[,-1], start = c(2007,1), frequency = 365)
```

(c)

```
plot(x,
     xlab = "Time",
     ylab = "Lake Depth Level",
     main = "Lake Depth Level by Time")
```



i) In this new plot, we can clearly see the change in lake's depth level according according to time, whereas in the first plot this is not easy to see.

ii) We observe seasonality from the above plot where the lake's depth is high in the middle of the year and becomes lower at the end of a year.

iii) We do not observe a trend that there is no overall increase or decrease in lake's depth level year over year.

iv) The observations are serially dependent.

v) There is cyclical variation in the above plot, where we see the same pattern of increasing and decreasing depth year over year.

Question 2

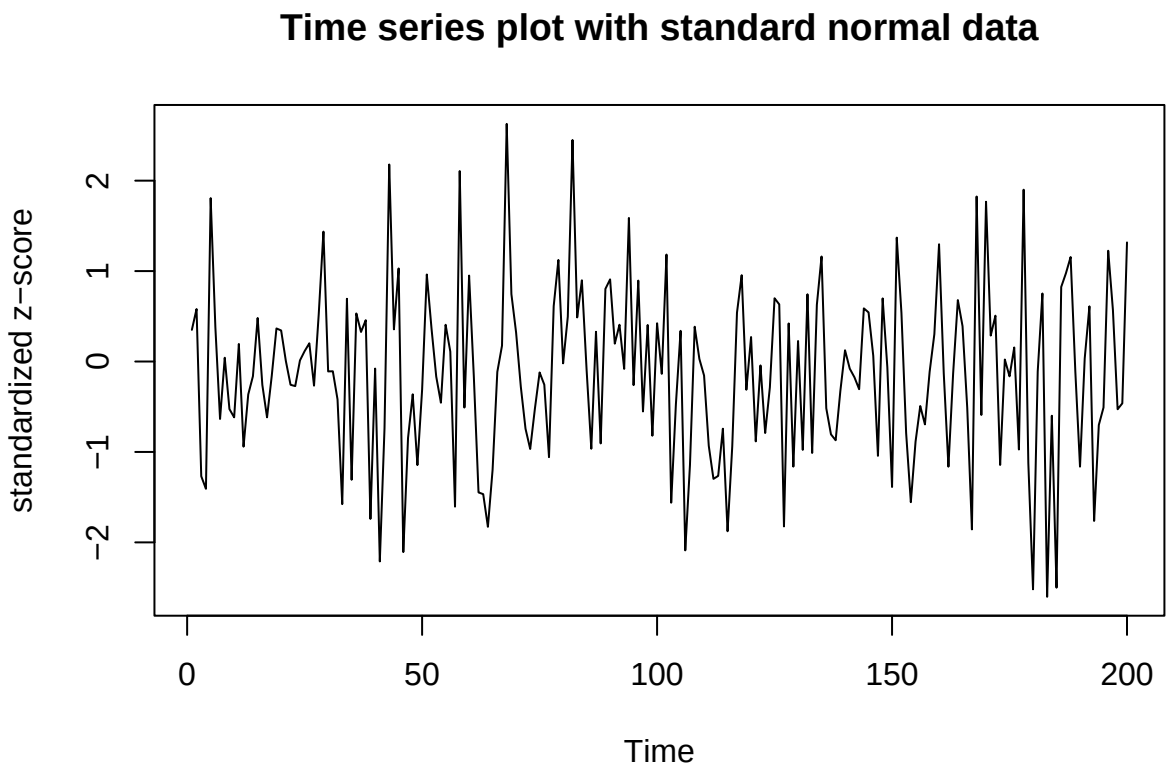
(a)

```
set.seed(443)
std_data <- rnorm(200)
std_ts <- ts(std_data)
max(std_data)
```

```
## [1] 2.627609
```

(b)

```
plot(std_ts,
     ylab = "standardized z-score",
     main = "Time series plot with standard normal data")
```




```
# calculate percentage of observation fall outside of +/- 2
mean(std_data < -2.0 | std_data > 2.0)
```

```
## [1] 0.05
```

```
# expected percentage of observations fall outside of +/- 2
round(2*pnorm(2.0, lower.tail = FALSE),4)
```

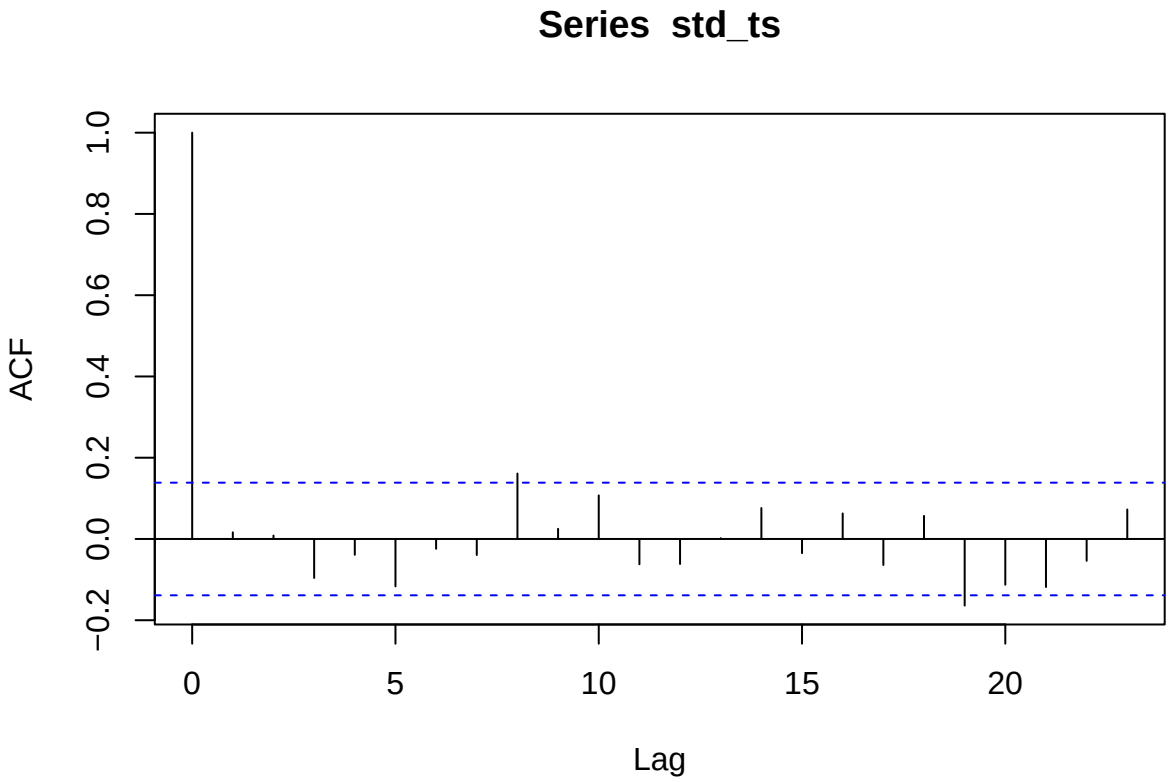
```
## [1] 0.0455
```

i) About 5% of the observations fall outside of the range ± 2

ii) Since the observations are draw from a standard normal distribution, we would expect around 4.55% of the observations to fall outside of the range ± 2.

(c)

```
acf(std_ts, type = "correlation")
```



Looking at the above correlogram, most of the sample acf are within the $\pm 2/\sqrt{200}$ bound. We do not see the acf alternates, nor there is any oscillation. All observations in the sample are equally uncorrelated, such that observations close in time are equally less correlated as observations that are farther in apart in time.

More information on R Markdown

This is an R Markdown document, which can be used as a template for STAT 443 labs and assignments. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Using the function *kable*, it produces a nicer table

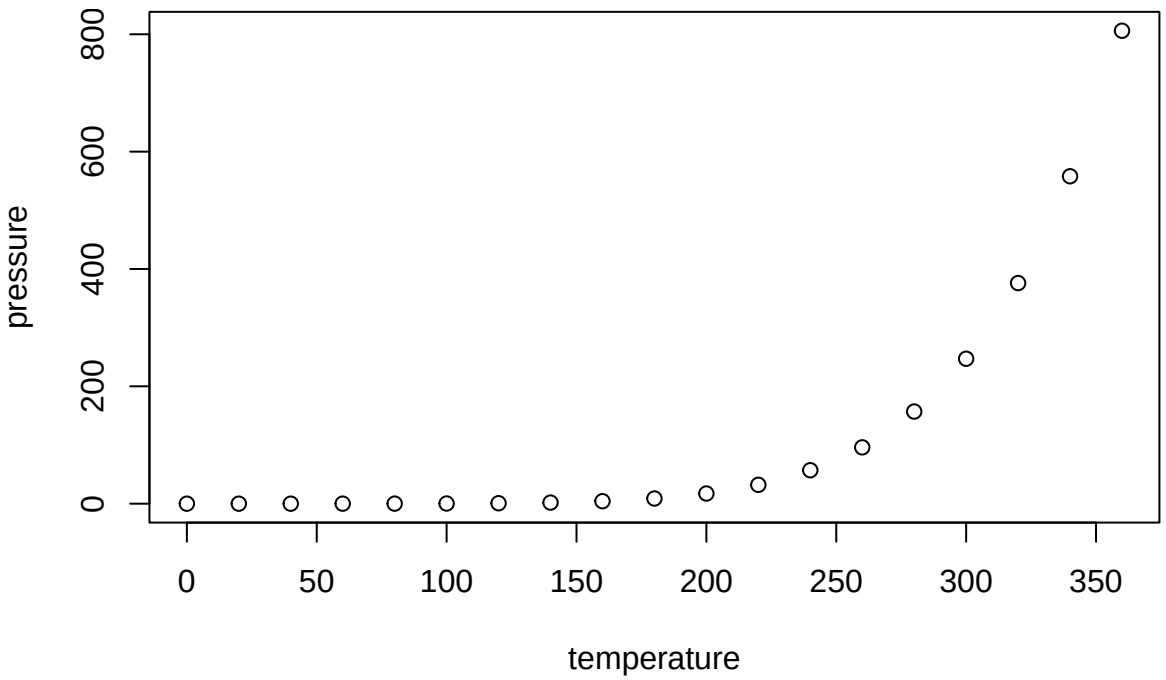
```
kable(summary(cars))
```

speed	dist
Min. : 4.0	Min. : 2.00
1st Qu.:12.0	1st Qu.: 26.00
Median :15.0	Median : 36.00
Mean :15.4	Mean : 42.98
3rd Qu.:19.0	3rd Qu.: 56.00
Max. :25.0	Max. :120.00

Including Plots

You can also embed plots, for example:

```
plot(pressure)
```



Note that specifying `echo = FALSE` parameter would prevent printing of the R code that generated the plot. This is something you may want to do for larger reports that would not require display of the R code.

You can also modify the size and alignment of the figure.

```
plot(pressure)
```

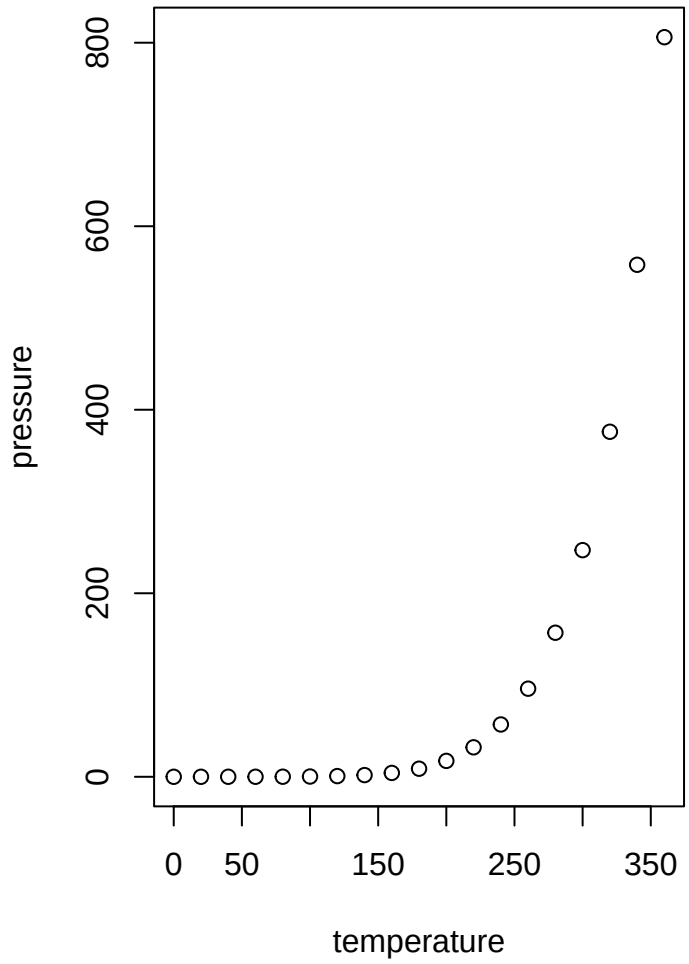


Figure 1: title