DSC 425 Time Series Analysis Forecasting

Project: Sea Ice Extend

Exploratory Data Analysis

```
> seaice <- read_csv("Desktop/seaice.csv")
Rows: 26354 Columns: 7
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (2): Source Data, hemisphere
dbl (5): Year, Month, Day, Extent, Missing

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
> head(seaice)
# A tibble: 6 × 7
   Year Month   Day Extent Missing `Source Data`                                                                hemisphere
  <dbl> <dbl> <dbl>  <dbl>   <dbl> <chr>                                                                        <chr>
1  1978    10    26   10.2       0 ['ftp://sidads.colorado.edu/pub/DATASETS/nsidc0051_gsfc_nasateam_seaice/final-… north
2  1978    10    28   10.4       0 ['ftp://sidads.colorado.edu/pub/DATASETS/nsidc0051_gsfc_nasateam_seaice/final-… north
3  1978    10    30   10.6       0 ['ftp://sidads.colorado.edu/pub/DATASETS/nsidc0051_gsfc_nasateam_seaice/final-… north
4  1978    11     1   10.7       0 ['ftp://sidads.colorado.edu/pub/DATASETS/nsidc0051_gsfc_nasateam_seaice/final-… north
5  1978    11     3   10.8       0 ['ftp://sidads.colorado.edu/pub/DATASETS/nsidc0051_gsfc_nasateam_seaice/final-… north
6  1978    11     5   11.0       0 ['ftp://sidads.colorado.edu/pub/DATASETS/nsidc0051_gsfc_nasateam_seaice/final-… north
>
```

We removed the missing column and source data from our data, and after that combine the year, month, day into a single column and gave that column new name called date.

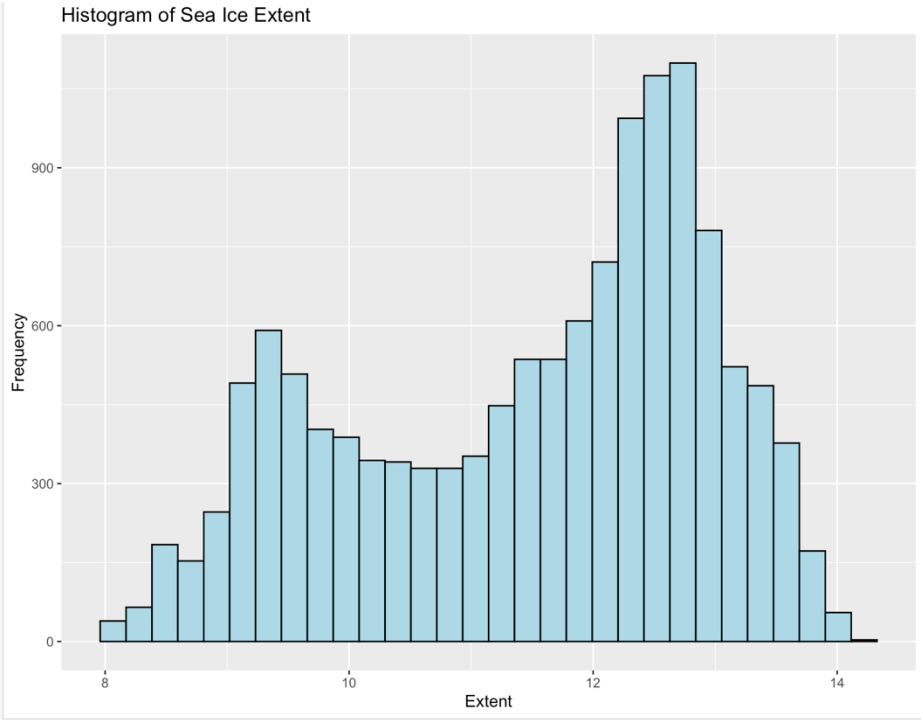
```
> seaice <- seaice[c(-5, -6)]
> seaice$Date <- as.Date(with(seaice, paste(Year, Month, Day, sep = '-')), "%Y-%m-%d")
> seaice <- seaice %>% select(-Year, -Month, -Day)
> head(seaice)
# A tibble: 6 × 3
  Extent hemisphere Date
   <dbl> <chr>      <date>
1   10.2 north      1978-10-26
2   10.4 north      1978-10-28
3   10.6 north      1978-10-30
4   10.7 north      1978-11-01
5   10.8 north      1978-11-03
6   11.0 north      1978-11-05
>
```

Moreover, we separate out the hemisphere into two parts i.e., North, and South. After that combine their extend value in one by taking average.

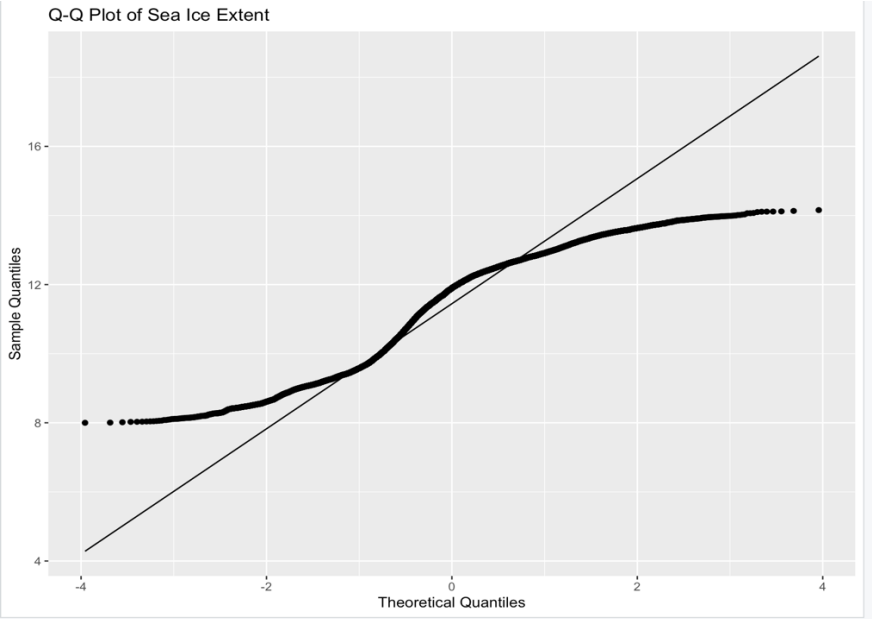
```
> north_seaice <- filter(seaice, hemisphere == "north")
> head(north_seaice)
# A tibble: 6 × 3
  Extent hemisphere Date
   <dbl> <chr>      <date>
1   10.2 north      1978-10-26
2   10.4 north      1978-10-28
3   10.6 north      1978-10-30
4   10.7 north      1978-11-01
5   10.8 north      1978-11-03
6   11.0 north      1978-11-05
> south_seaice <- filter(seaice, hemisphere == "south")
> head(south_seaice)
# A tibble: 6 × 3
  Extent hemisphere Date
   <dbl> <chr>      <date>
1   17.6 south      1978-10-26
2   17.8 south      1978-10-28
3   17.7 south      1978-10-30
4   17.5 south      1978-11-01
5   17.5 south      1978-11-03
6   17.3 south      1978-11-05
>
> combined_seaice <- seaice %>%
+   group_by(Date) %>%
+   summarize(newExtent = mean(Extent))
>
> head(combined_seaice)
# A tibble: 6 × 2
  Date       newExtent
  <date>         <dbl>
1 1978-10-26      13.9
2 1978-10-28      14.1
3 1978-10-30      14.1
4 1978-11-01      14.1
5 1978-11-03      14.1
6 1978-11-05      14.2
>
```

Plotting Histogram, QQ plot and Jarque Bera Test to check whether data is normally distributed.

```
> # Histogram
> ggplot(combined_seaice, aes(x = newExtent)) +
+   geom_histogram( fill = "lightblue", color = "black") +
+   labs(title = "Histogram of Sea Ice Extent", x = "Extent", y = "Frequency")
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
>
```




```
> # Q-Q plot
> ggplot(combined_seaice, aes(sample = newExtent)) +
+   geom_qq() +
+   geom_qq_line() +
+   labs(title = "Q-Q Plot of Sea Ice Extent", x = "Theoretical Quantiles", y = "Sample Quantiles")
>
```



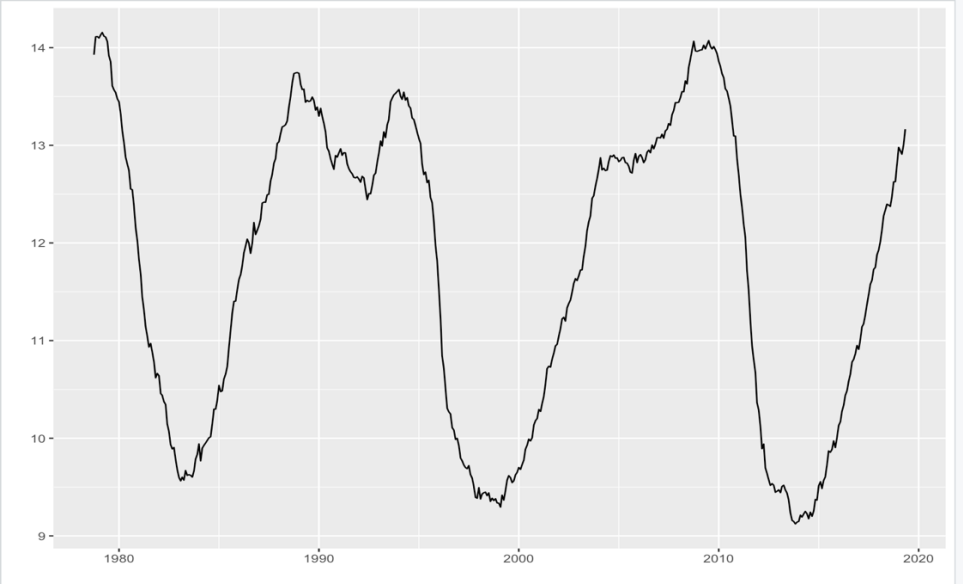
From the above plots, it seems that data is not normally distributed.

```
>
> # Perform Jarque-Bera test
> jb_test <- jarque.bera.test(combined_seaice$newExtent)
>
> # Print the test results
> print(jb_test)

        Jarque Bera Test

data:  combined_seaice$newExtent
X-squared = 989.8, df = 2, p-value < 2.2e-16
```

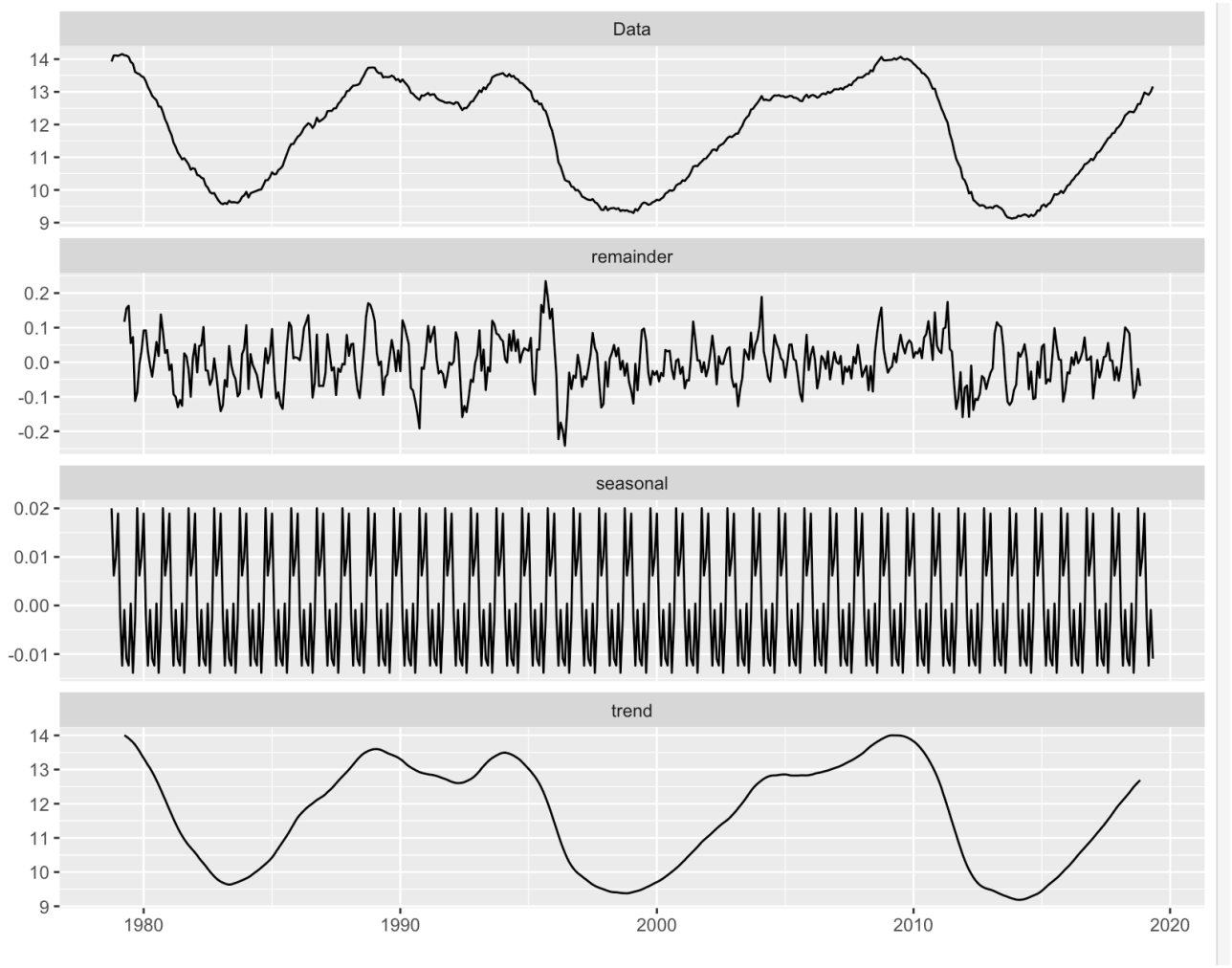
Based on these results, we can conclude that the data in combined_seaice$newExtent significantly deviates from a normal distribution. The extremely small p-value suggests strong evidence against the null hypothesis of normality.
Therefore, it is not appropriate to assume that the sea ice extent data follows a normal distribution.

Creating Time series:

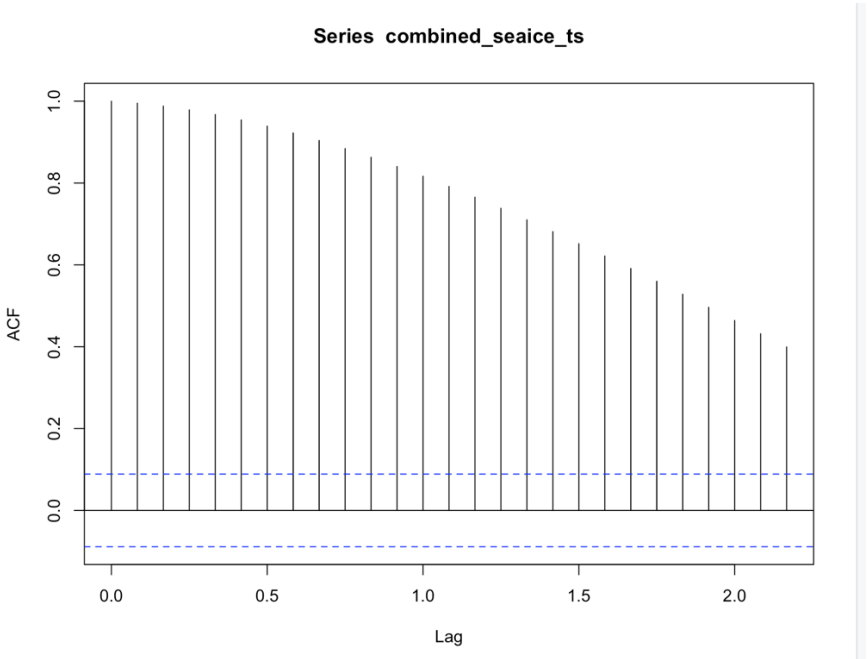
```
> combined_seaice_ts = ts(combined_seaice $newExtent, start=c(1978, 10),end  = c(2019, 5), frequency=12)
> autoplot(combined_seaice_ts)
```



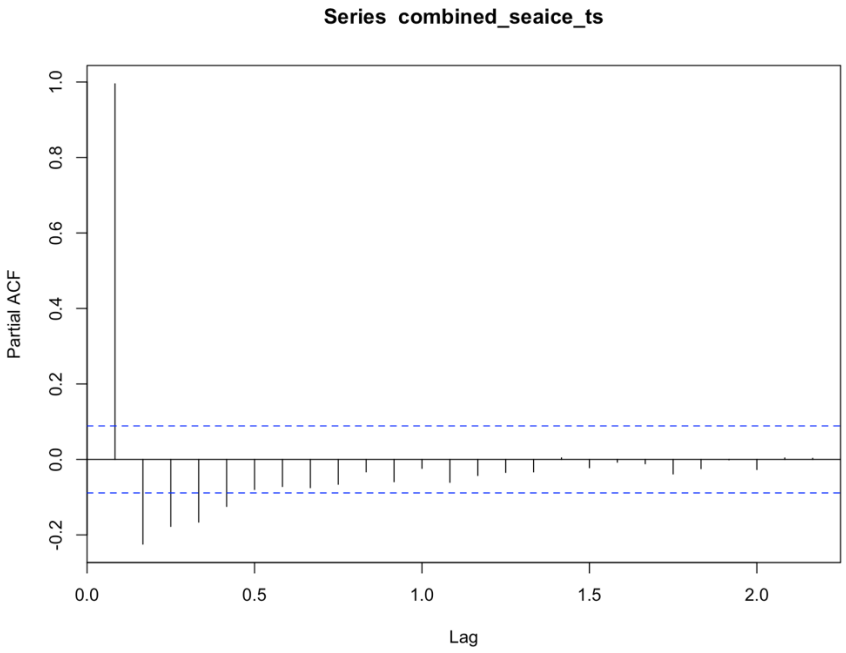
autoplot(decompose(combined_seaice_ts))



acf(combined_seaice_ts)



From, the above graphs series appear to be non- stationary.



The pacf plot have a significant correlation at lag 0 and then negative correlation at lag 1.

Further Analysis:
Now, we are considering making models based on interpretation of graphs, we are planning to make 2-3 models and check which fits best.

Code:

```
library(readr)
 library(dplyr)
 library(ggplot2)   # For qplot
 library(fBasics)
 library(lubridate) # for mdy date conversion
 library(ggfortify) # for ts autoplot
 library(zoo)
 library(tseries)
seaice <- read_csv("Desktop/seaice.csv")
head(seaice)
seaice <- seaice[c(-5, -6)]
seaice$Date <- as.Date(with(seaice, paste(Year, Month, Day, sep = '-')), "%Y-%m-%d")
seaice <- seaice %>% select(-Year, -Month, -Day)
head(seaice)
tail(seaice)
north_seaice <- filter(seaice, hemisphere == "north")
head(north_seaice)
south_seaice <- filter(seaice, hemisphere == "south")
head(south_seaice)

combined_seaice <- seaice %>%
  group_by(Date) %>%
  summarize(newExtent = mean(Extent))

head(combined_seaice)


# Histogram
ggplot(combined_seaice, aes(x = newExtent)) +
  geom_histogram( fill = "lightblue", color = "black") +
  labs(title = "Histogram of Sea Ice Extent", x = "Extent", y = "Frequency")

# Q-Q plot
```

```
ggplot(combined_seaice, aes(sample = newExtent)) +
  geom_qq() +
  geom_qq_line() +
  labs(title = "Q-Q Plot of Sea Ice Extent", x = "Theoretical Quantiles", y = "Sample Quantiles")

# Perform Jarque-Bera test
jb_test <- jarque.bera.test(combined_seaice$newExtent)

# Print the test results
print(jb_test)


combined_seaice_ts = ts(combined_seaice $newExtent, start=c(1978, 10),end  = c(2019, 5),
frequency=12)
autoplot(combined_seaice_ts)
autoplot(decompose(combined_seaice_ts))

acf(combined_seaice_ts)
pacf(combined_seaice_ts)


decomp = decompose(combined_seaice_ts)

ifelse(all(decomp$random == 0), "Additive", "Multiplicative")
```