Final project

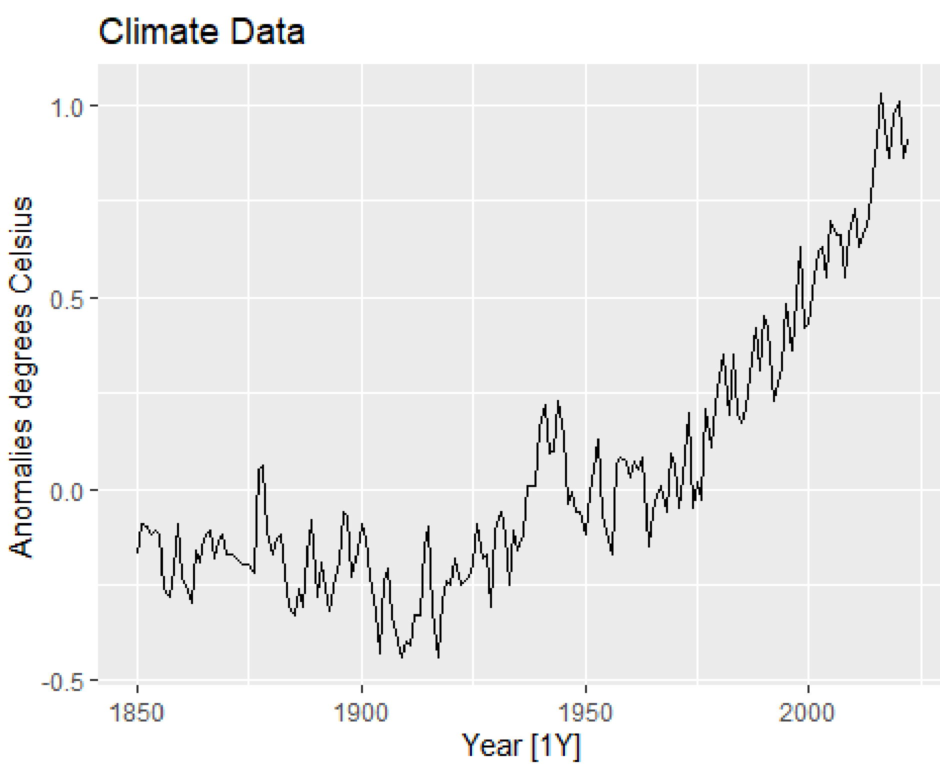
2023-04-24

```
library(fpp3)

## ── Attaching packages ──────────────────────────────────────────── fpp3
0.4.0 ──

## ✓ tibble      3.1.8      ✓ tsibble     1.1.3
## ✓ dplyr       1.0.10     ✓ tsibbledata 0.4.1
## ✓ tidyr       1.3.0      ✓ feasts      0.3.0
## ✓ lubridate   1.9.1      ✓ fable       0.3.2
## ✓ ggplot2     3.4.0

## ── Conflicts ─────────────────────────────────────────────────
fpp3_conflicts ──
## ✖ lubridate::date()    masks base::date()
## ✖ dplyr::filter()      masks stats::filter()
## ✖ tsibble::intersect() masks base::intersect()
## ✖ tsibble::interval()  masks lubridate::interval()
## ✖ dplyr::lag()         masks stats::lag()
## ✖ tsibble::setdiff()   masks base::setdiff()
## ✖ tsibble::union()     masks base::union()

library("readxl")

data <- read_excel("C:\\Users\\jlgan\\OneDrive\\Desktop\\Time Series
Data\\Climate Data.xlsx")
attach(data)
```

Plot data

```
head(data)

## # A tibble: 6 × 2
##    Year Value
##   <dbl> <dbl>
## 1  1850 -0.17
## 2  1851 -0.09
## 3  1852 -0.1
## 4  1853 -0.12
## 5  1854 -0.11
## 6  1855 -0.12

data.ts <- data|>
  as_tsibble(index= Year)
data.ts |>
```

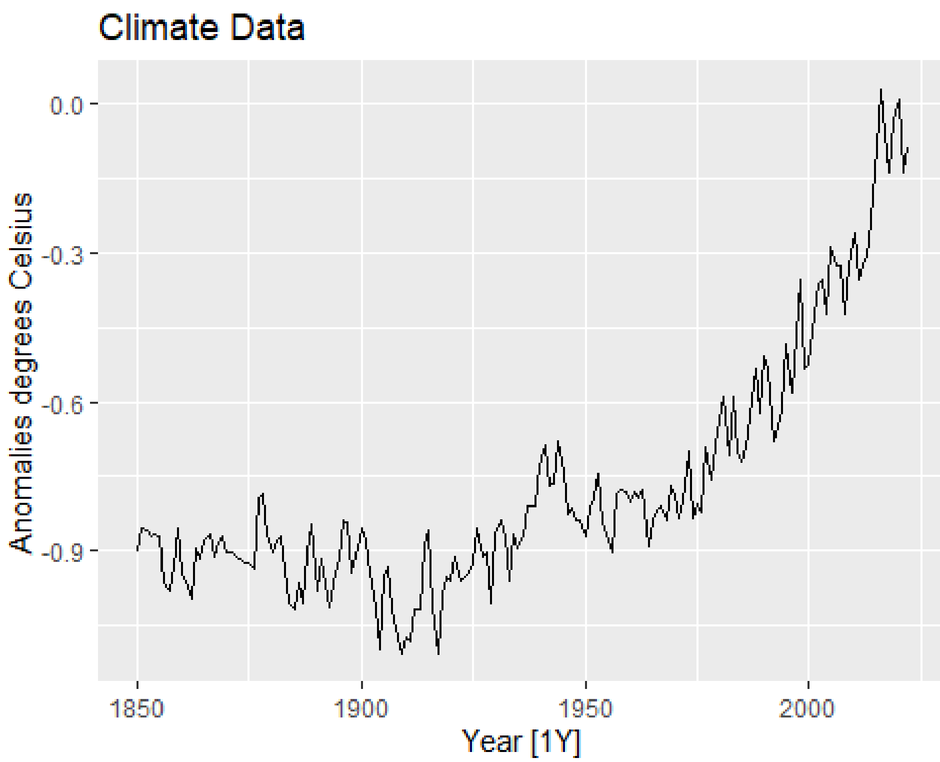
```
  autoplot() +
  labs(title = "Climate Data",
       y="Anomalies degrees Celsius")

## Plot variable not specified, automatically selected `.vars = Value`
```



##The data above is a time series plot of our data. As we can see, the variance tende to increase, so we will now transform and difference the data. The trend towards the last 30 years has increased dramatically, so it gives proof for the ever pressing issue of climate change.

##Transformation

```
lambda <- data.ts |>
  features(Value, features = guerrero) |>
  pull(lambda_guerrero)
lambda

## [1] 1.232127

data.ts |>
  autoplot(box_cox(Value, lambda)) +
  labs(title = "Climate Data",
       y="Anomalies degrees Celsius")
```




```
data.tsBC <- data.ts|>
  mutate(ValueBC = box_cox(Value,lambda))
```

##Above is the transformation of our data. As is clearly presented in the graph, the Box Cox transformation had little to no impact, thus we will not be transforming our data.

Descriptive statistics

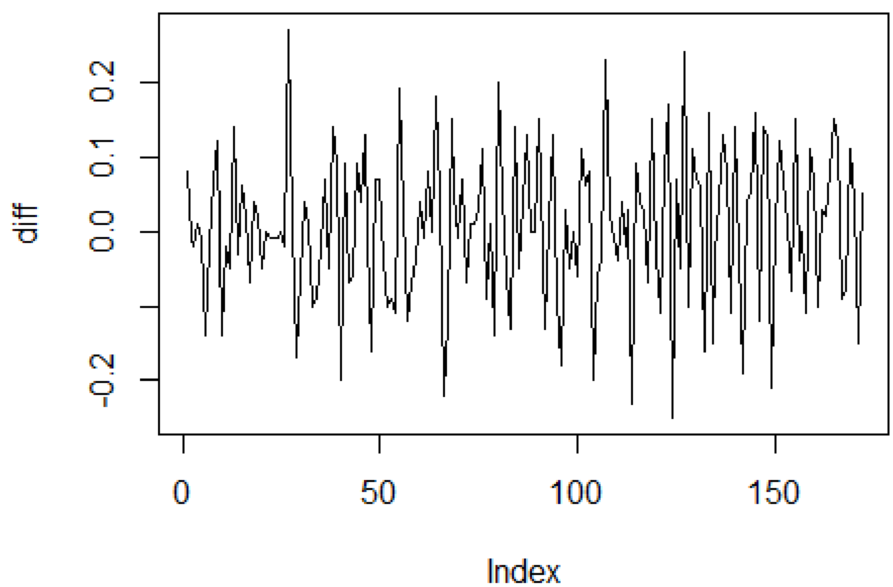
```
summary(data)

##       Year          Value
##  Min.   :1850   Min.   :-0.44000
##  1st Qu.:1893   1st Qu.:-0.19000
##  Median :1936   Median :-0.07000
##  Mean   :1936   Mean   : 0.04439
##  3rd Qu.:1979   3rd Qu.: 0.21000
##  Max.   :2022   Max.   : 1.03000

data.ts |>
  features(Value, unitroot_ndiffs)

## # A tibble: 1 × 1
##   ndiffs
##    <int>
## 1      1

data.tsBC |>
  features(ValueBC, unitroot_ndiffs)
```

```
## # A tibble: 1 × 1
##   ndiffs
##    <int>
## 1      1
```

##This is a representation of the data through basic statistics. As is clearly presented, the max data was in 2022 at 1.08 degrees celsius. The unit root test informs the necessity for differencing to make the data stationary, which will be applied in the next graph.

Make data stationary

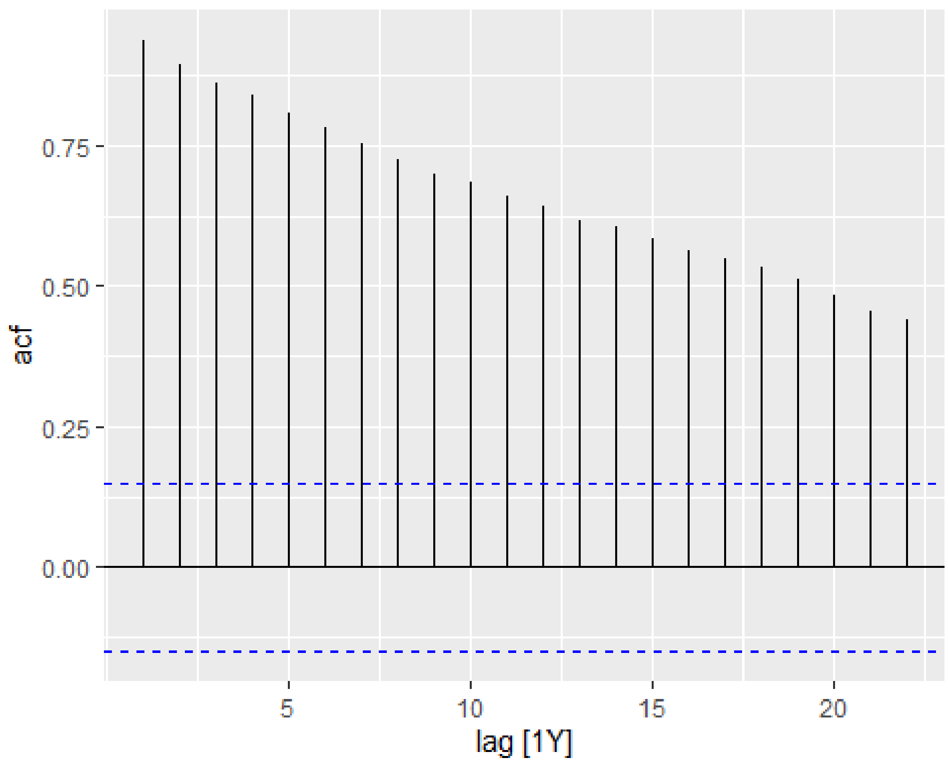
```
diff <- diff(Value)
plot(diff, type = "l")
```



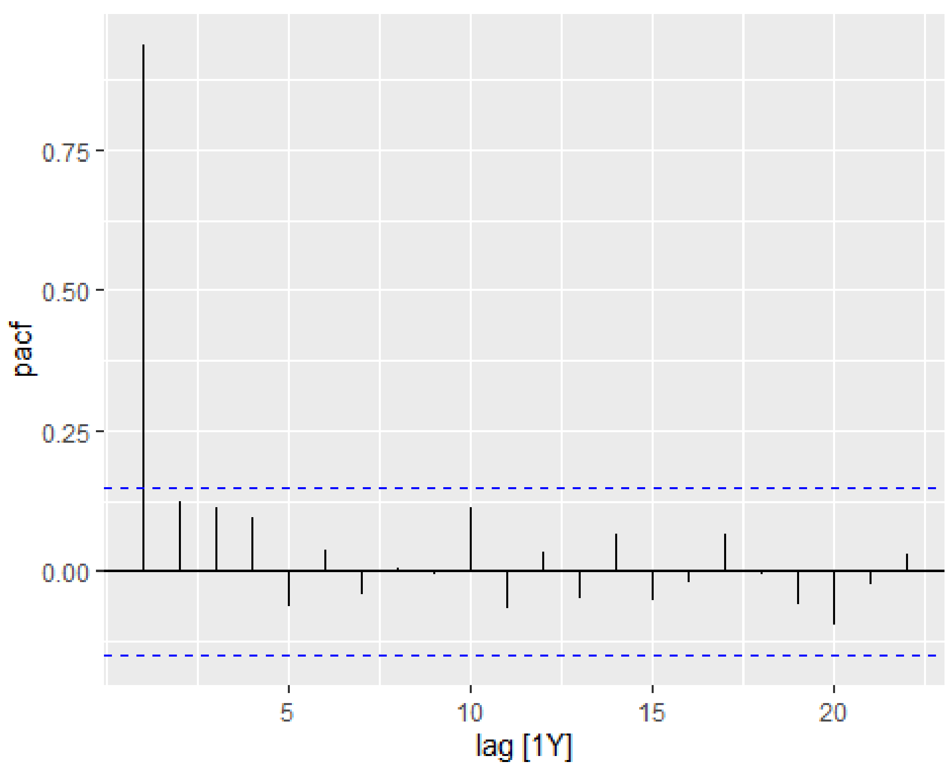
##Here is a presentation of the differenced data, these values are sinusoidal with no trend, thus allowing us to create a more accurate forecast.

ACF

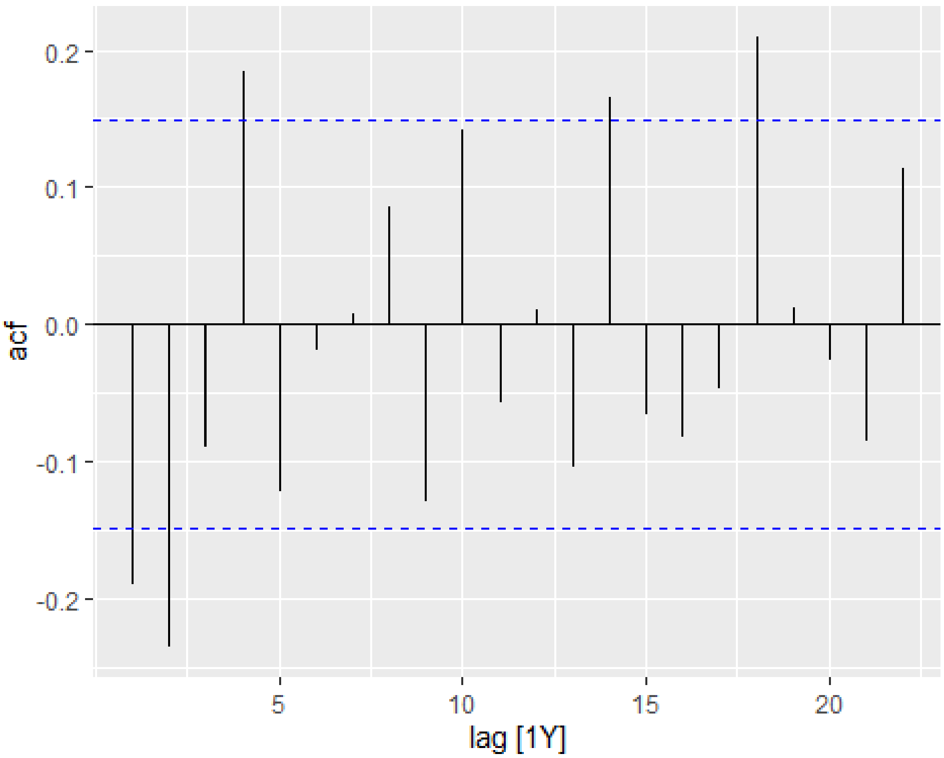
```
data.ts |> ACF(Value) |>
  autoplot()
```



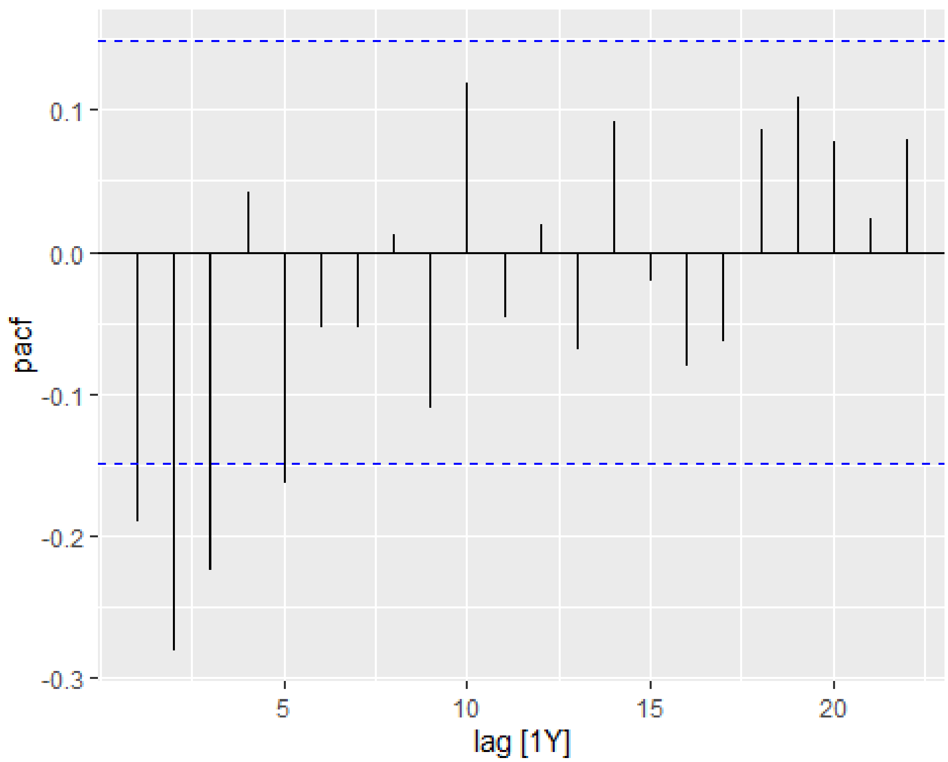

```
## Not seasonal they are annual, so they are highly autocorrelated.
data.ts |> PACF(Value) |>
  autoplot()
```



```r
data.ts |> ACF(difference(Value)) |>
  autoplot()
```




```r
data.ts |> PACF(difference(Value)) |>
  autoplot()
```



##The ACF of the differenced data varies greatly with the ACF of the non-differenced data. Thus, the overall differencing of the data has allowed for it to be stationary.

ARIMA model

```
fit <- data.ts |>
  model(arima1c = ARIMA(Value ~ 1 + pdq(1,1,3)),
        arima510c = ARIMA(Value ~ 1 + pdq(5,1,0)),
        arima0118c = ARIMA(Value ~ 1 +  pdq(0,1,18)),
        arima1 = ARIMA(Value ~ 0 + pdq(1,1,3)),
        arima510 = ARIMA(Value ~ 0+ pdq(5,1,0)),
        arima0118 = ARIMA(Value ~ 0 + pdq(0,1,18)),
        stepwise = ARIMA(Value),
        search = ARIMA(Value, stepwise = FALSE)
        )
fit |> report()

## Warning in report.mdl_df(fit): Model reporting is only supported for
individual
## models, so a glance will be shown. To see the report for a specific model,
use
## `select()` and `filter()` to identify a single model.

## # A tibble: 8 × 8
##   .model      sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
##   <chr>        <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
## 1 arima1c    0.00851    168. -324. -324. -305. <cpl [1]> <cpl [3]>
## 2 arima510c  0.00867    167. -320. -319. -298. <cpl [5]> <cpl [0]>
```

```
## 3 arima0118c 0.00815    176. -312. -306. -249. <cpl [0]> <cpl [18]>
## 4 arima1     0.00867    166. -322. -322. -306. <cpl [1]> <cpl [3]>
## 5 arima510   0.00878    165. -319. -318. -300. <cpl [5]> <cpl [0]>
## 6 arima0118  0.00816    175. -313. -308. -253. <cpl [0]> <cpl [18]>
## 7 stepwise   0.00851    168. -324. -324. -305. <cpl [1]> <cpl [3]>
## 8 search     0.00851    168. -324. -324. -305. <cpl [1]> <cpl [3]>
```

```
fit |>
  select(arima1c) |>
  report()
```

```
## Series: Value
## Model: ARIMA(1,1,3) w/ drift
##
## Coefficients:
##           ar1     ma1      ma2      ma3  constant
##       -0.9561  0.6883  -0.5473  -0.3463    0.0123
## s.e.   0.0370  0.0800   0.0743   0.0688    0.0056
##
## sigma^2 estimated as 0.008511:  log likelihood=168.06
## AIC=-324.12   AICc=-323.61   BIC=-305.23
```

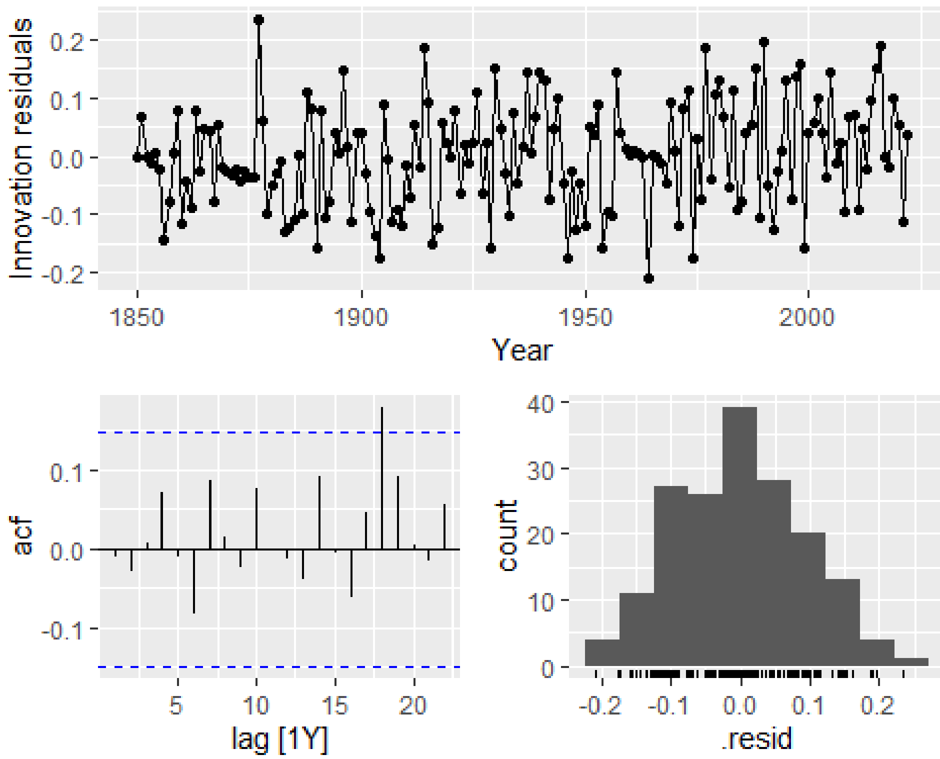
```
fit |>
  select(stepwise) |>
  report()
```

```
## Series: Value
## Model: ARIMA(1,1,3) w/ drift
##
## Coefficients:
##           ar1     ma1      ma2      ma3  constant
##       -0.9561  0.6883  -0.5473  -0.3463    0.0123
## s.e.   0.0370  0.0800   0.0743   0.0688    0.0056
##
## sigma^2 estimated as 0.008511:  log likelihood=168.06
## AIC=-324.12   AICc=-323.61   BIC=-305.23
```

```
fit |>
  select(search) |>
  report()
```

```
## Series: Value
## Model: ARIMA(1,1,3) w/ drift
##
## Coefficients:
##           ar1     ma1      ma2      ma3  constant
##       -0.9561  0.6883  -0.5473  -0.3463    0.0123
## s.e.   0.0370  0.0800   0.0743   0.0688    0.0056
##
## sigma^2 estimated as 0.008511:  log likelihood=168.06
## AIC=-324.12   AICc=-323.61   BIC=-305.23
```

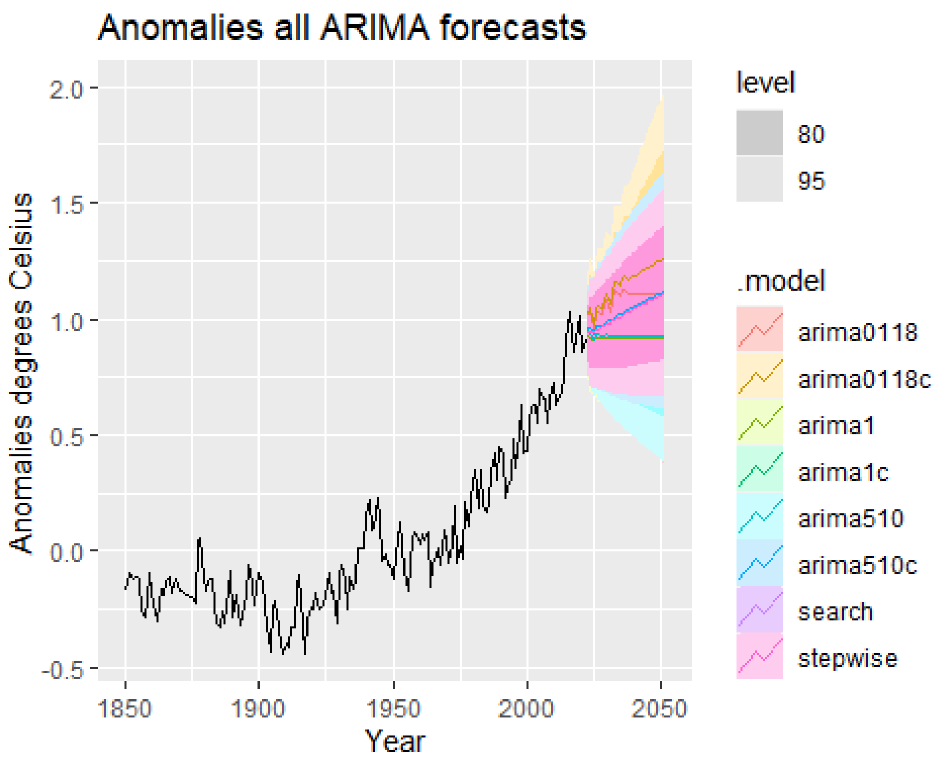
```
fit2 <- data.ts |>
  model(stepwise1 = ARIMA(Value),
         search1 = ARIMA(Value, stepwise = FALSE))

fit2 |>
  select(search1) |>
  gg_tsresiduals()
```



##This presents the search for the "Best" ARIMA model to forecast our data. As is clear, the AICc is lowest for the stepwise, search, and arima1c models. It appears these three models are equivalent.

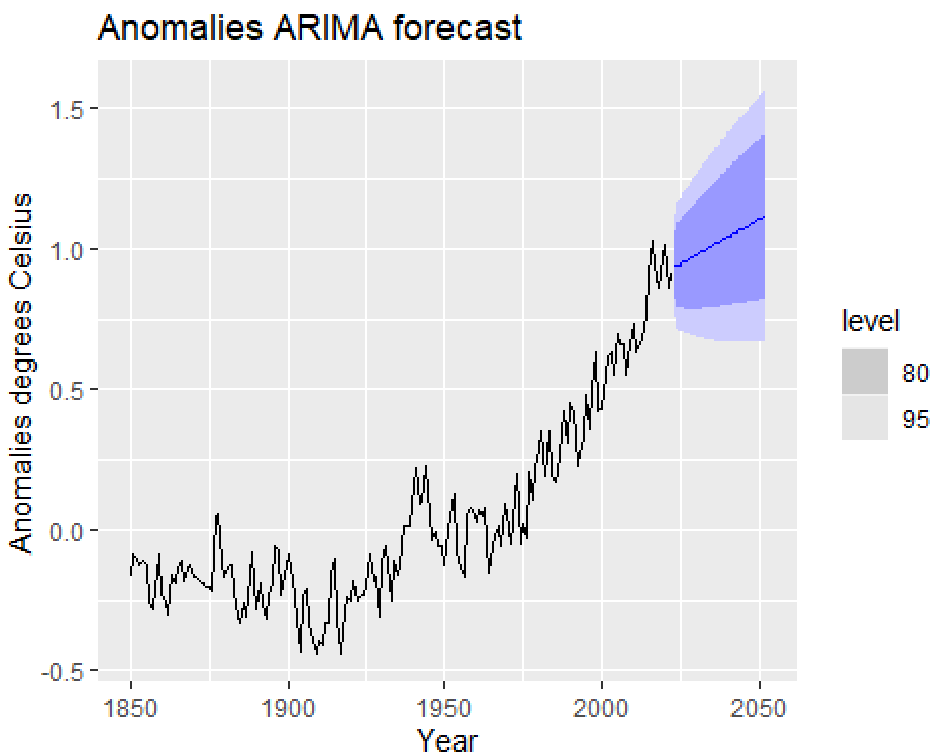
Forecasting

```
all_forecast <- fit %>% forecast(h=30) %>% autoplot(data.ts) +
  labs(title = "Anomalies all ARIMA forecasts",
       y="Anomalies degrees Celsius")
all_forecast
```



##This visually digests the entirety of the forecasts used above. It allows us to see how the separate forecasts vary from one another. The most accurate forecast is the arima1c/search/stepwise model.

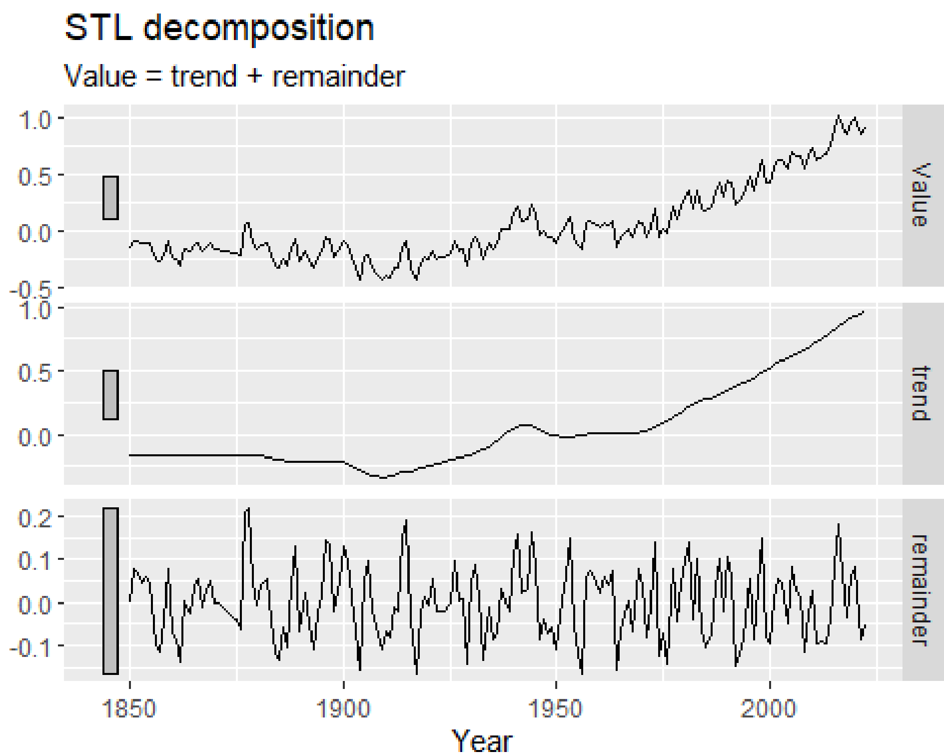
```
stepwiseforecast <- fit |>
  forecast(h=30) |>
  filter(.model=='stepwise') |>
  autoplot(data.ts) +
  labs(title = "Anomalies ARIMA forecast",
       y="Anomalies degrees Celsius")
stepwiseforecast
```



##This is a visual representation of only the arima1c model. As we can see the trend of the data gradually increases, once again providing insight on Climate Change data.

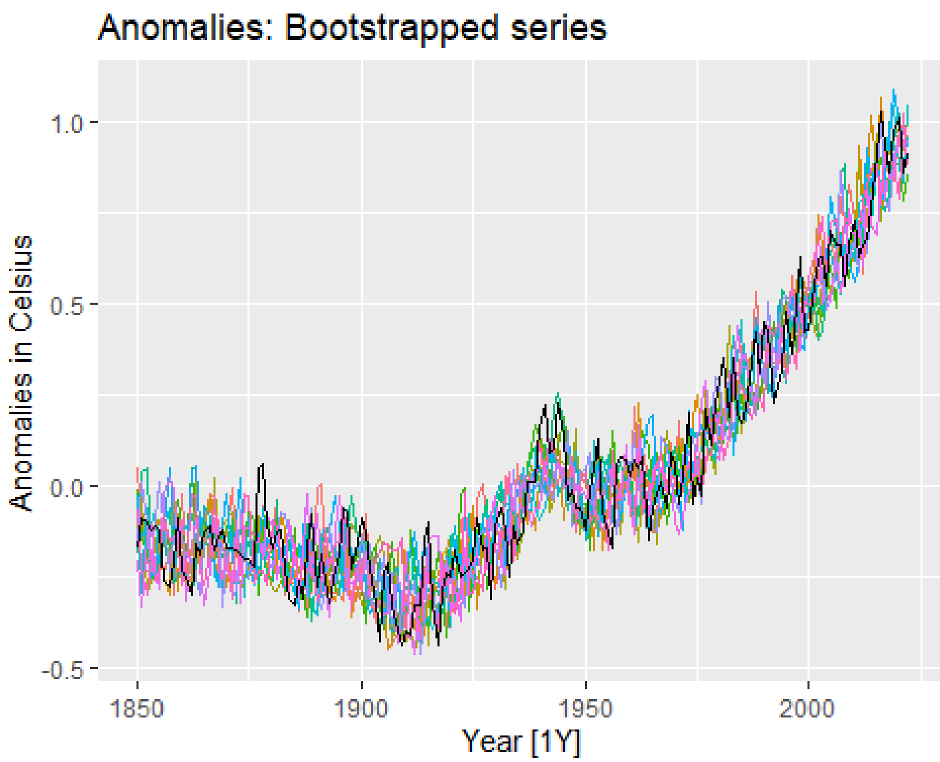
To receive a secondary more accurate model we will be using the Bootstrapping method. The Bootstrapping method requires the following procedures: the decomposion into trend, seasonal and remainder components using STL.Then it shuffles the remainder component to resample the dataset and get bootstrapped remainder series. Our data used blocked bootstrap (due to autocorrelation)- where adjacent sections of data are selected at random and joined together. The Bootstrapped remainder series is then added to trend and seasonal components. Finally, transformation is reversed to create variations on original time series data

```
data.ts_stl <- data.ts |>
  model(stl = STL(Value))
data.ts_stl |>
  components() |>
  autoplot()
```



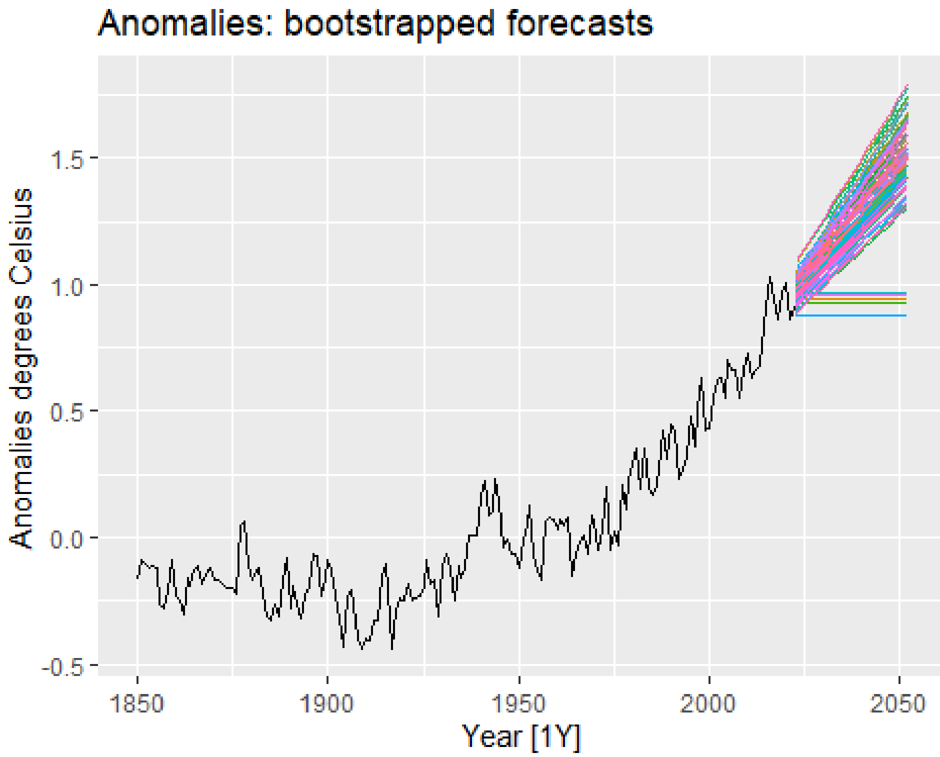
##This presents the STL decomposition or the Seasonal Decomposition of Time Series by Loess.

```
data.ts_stl |>
  generate(new_data = data.ts, times = 10,
           bootstrap_block_size = 8) |>
  autoplot(.sim) +
  autolayer(data.ts, Value) +
  guides(colour = "none") +
  labs(title = "Anomalies: Bootstrapped series",
       y="Anomalies in Celsius")

## Warning in max(vapply(x$seasons, `[[`, double(1L), "period")): no non-
missing
## arguments to max; returning -Inf
```



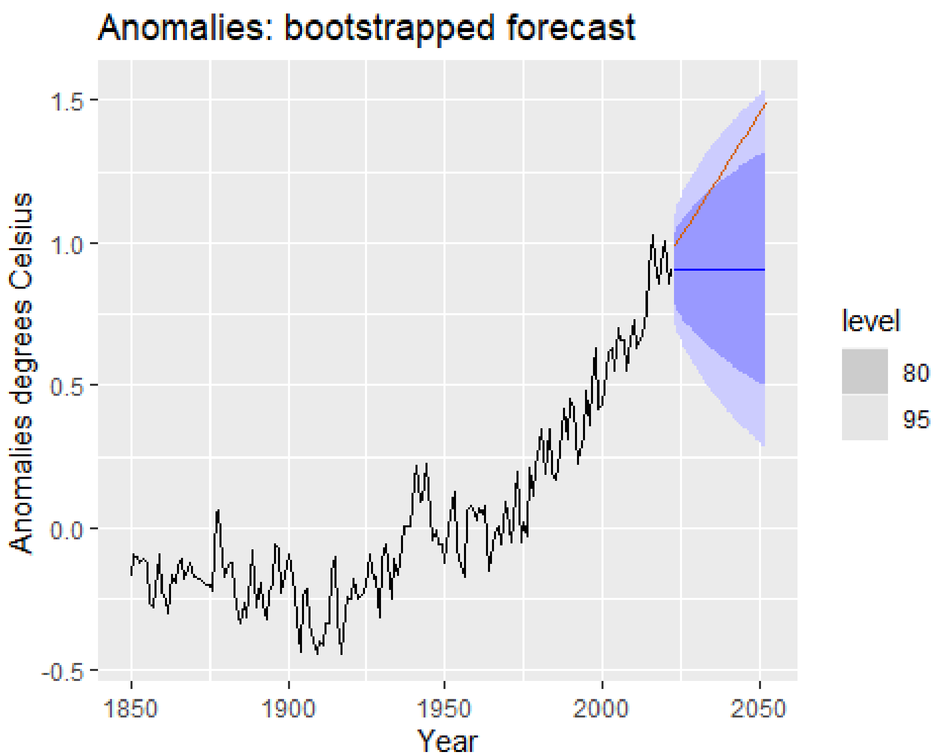
##Here is the actual bootstrapping of our data with the data being generated 10 times.

```
sim <- data.ts_stl |>
  generate(new_data = data.ts, times = 100,
           bootstrap_block_size = 8) |>
  select(-.model, -Value)

## Warning in max(vapply(x$seasons, `[[`, double(1L), "period")): no non-
missing
## arguments to max; returning -Inf

ets_forecasts <- sim |>
  model(ets = ETS(.sim)) |>
  forecast(h = 30)
ets_forecasts |>
  update_tsibble(key = .rep) |>
  autoplot(.mean) +
  autolayer(data.ts, Value) +
  guides(colour = "none") +
  labs(title = "Anomalies: bootstrapped forecasts",
       y="Anomalies degrees Celsius")
```



##This is the final presentation of all of the Bootstrap forecasts for the next 30 years, in the next step the most accurate model is presented.

```
bagged <- ets_forecasts |>
  summarise(bagged_mean = mean(.mean))
data.ts |>
  model(ets = ETS(Value)) |>
  forecast(h = 30) |>
  autoplot(data.ts) +
  autolayer(bagged, bagged_mean, col = "#D55E00") +
  labs(title = "Anomalies: bootstrapped forecast",
       y="Anomalies degrees Celsius")
```



##This is the final forecast. It is the most accurate model since it was bootstrapped; however, the forecast goes towards the 80% confidence interval, proving the forecast may become more unreliable as time increases. With this said, environmental impact may also affect the projection of the data with the implementation of more gas emitting industries. The data projects the average increase in temperature in 2050 to be 1.5 degrees celsius.

##In accordance with a report by climate scientists convened by the UN, if the climate were to warm by 1.5 degrees celsius it would have devastating effects on the environment. In 2018, they showed that if this were to occur sea levels would rise between 10 and 30 inches, which would cause 10 million people to be at risk of flooding and coastal storms. It would also have detrimental impacts on the environment, where 90% of coral reefs would die and 7% of Earth's land would shift into a new Biome; for example, grasslands would turn to deserts, and tundra's into forests. Also, earth's population would be subject to extreme heat waves every 5 years.