BirthTimeSeries

Part A decomposition

```
library(forecast)
```

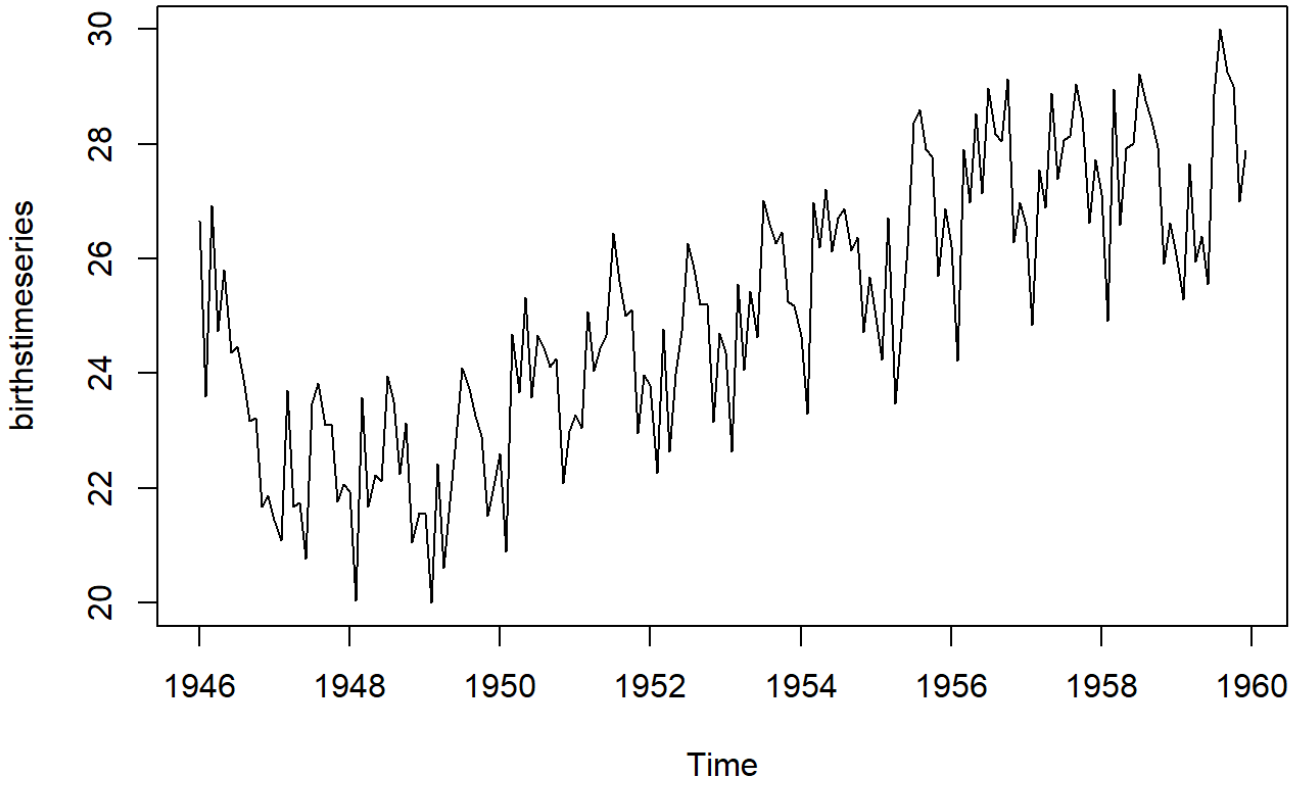
```
## Warning: package 'forecast' was built under R version 3.4.4
```

```
births <- scan("http://robjhyndman.com/tsdldata/data/nybirths.dat")
birthstimeseries <- ts(births, frequency=12, start=c(1946,1))
birthstimeseries
```

```
##         Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct
## 1946 26.663 23.598 26.931 24.740 25.806 24.364 24.477 23.901 23.175 23.227
## 1947 21.439 21.089 23.709 21.669 21.752 20.761 23.479 23.824 23.105 23.110
## 1948 21.937 20.035 23.590 21.672 22.222 22.123 23.950 23.504 22.238 23.142
## 1949 21.548 20.000 22.424 20.615 21.761 22.874 24.104 23.748 23.262 22.907
## 1950 22.604 20.894 24.677 23.673 25.320 23.583 24.671 24.454 24.122 24.252
## 1951 23.287 23.049 25.076 24.037 24.430 24.667 26.451 25.618 25.014 25.110
## 1952 23.798 22.270 24.775 22.646 23.988 24.737 26.276 25.816 25.210 25.199
## 1953 24.364 22.644 25.565 24.062 25.431 24.635 27.009 26.606 26.268 26.462
## 1954 24.657 23.304 26.982 26.199 27.210 26.122 26.706 26.878 26.152 26.379
## 1955 24.990 24.239 26.721 23.475 24.767 26.219 28.361 28.599 27.914 27.784
## 1956 26.217 24.218 27.914 26.975 28.527 27.139 28.982 28.169 28.056 29.136
## 1957 26.589 24.848 27.543 26.896 28.878 27.390 28.065 28.141 29.048 28.484
## 1958 27.132 24.924 28.963 26.589 27.931 28.009 29.229 28.759 28.405 27.945
## 1959 26.076 25.286 27.660 25.951 26.398 25.565 28.865 30.000 29.261 29.012
##         Nov    Dec
## 1946 21.672 21.870
## 1947 21.759 22.073
## 1948 21.059 21.573
## 1949 21.519 22.025
## 1950 22.084 22.991
## 1951 22.964 23.981
## 1952 23.162 24.707
## 1953 25.246 25.180
## 1954 24.712 25.688
## 1955 25.693 26.881
## 1956 26.291 26.987
## 1957 26.634 27.735
## 1958 25.912 26.619
## 1959 26.992 27.897
```

ts is a timeseries function that will convert data to time series.
Frequency 12 means data will be represented on a monthly basis and start defines the starting date (in this case from january 1946)

```
plot(birthstimeseries)
```



Ploting values on Y axis.Time is represented on X axis.

```
birthstimeseriescomponents <- decompose(birthstimeseries)
```

To estimate the trend, seasonal and irregular components of this time series.
The estimated values of the seasonal, trend and irregular components are now stored in variables.

```
birthstimeseriescomponents$seasonal
```

```
##              Jan        Feb        Mar        Apr        May        Jun
## 1946 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1947 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1948 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1949 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1950 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1951 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1952 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1953 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1954 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1955 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1956 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1957 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1958 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
## 1959 -0.6771947 -2.0829607  0.8625232 -0.8016787  0.2516514 -0.1532556
##              Jul        Aug        Sep        Oct        Nov        Dec
## 1946  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1947  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1948  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1949  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1950  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1951  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1952  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1953  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1954  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1955  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1956  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1957  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1958  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
## 1959  1.4560457  1.1645938  0.6916162  0.7752444 -1.1097652 -0.3768197
```

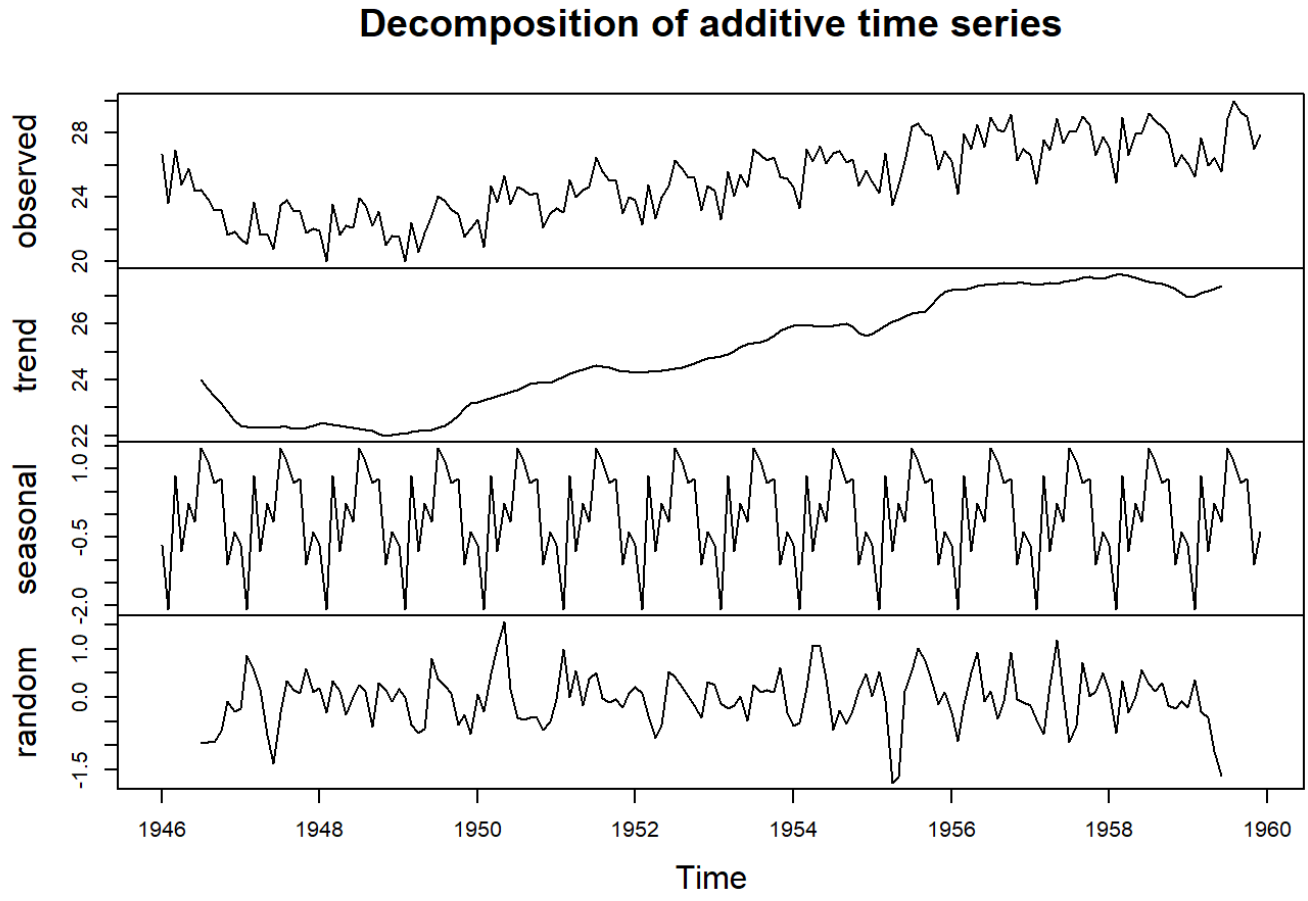
get the estimated values of the seasonal component

```
birthstimeseriescomponents$trend
```

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 1946       NA       NA       NA       NA       NA       NA 23.98433
## 1947 22.35350 22.30871 22.30258 22.29479 22.29354 22.30562 22.33483
## 1948 22.43038 22.43667 22.38721 22.35242 22.32458 22.27458 22.23754
## 1949 22.06375 22.08033 22.13317 22.16604 22.17542 22.21342 22.27625
## 1950 23.21663 23.26967 23.33492 23.42679 23.50638 23.57017 23.63888
## 1951 24.00083 24.12350 24.20917 24.28208 24.35450 24.43242 24.49496
## 1952 24.27204 24.27300 24.28942 24.30129 24.31325 24.35175 24.40558
## 1953 24.78646 24.84992 24.92692 25.02362 25.16308 25.26963 25.30154
## 1954 25.92446 25.92317 25.92967 25.92137 25.89567 25.89458 25.92963
## 1955 25.64612 25.78679 25.93192 26.06388 26.16329 26.25388 26.35471
## 1956 27.21104 27.21900 27.20700 27.26925 27.35050 27.37983 27.39975
## 1957 27.44221 27.40283 27.44300 27.45717 27.44429 27.48975 27.54354
## 1958 27.68642 27.76067 27.75963 27.71037 27.65783 27.58125 27.49075
## 1959 26.96858 27.00512 27.09250 27.17263 27.26208 27.36033       NA
##           Aug      Sep      Oct      Nov      Dec
## 1946 23.66213 23.42333 23.16112 22.86425 22.54521
## 1947 22.31167 22.26279 22.25796 22.27767 22.35400
## 1948 22.21988 22.16983 22.07721 22.01396 22.02604
## 1949 22.35750 22.48862 22.70992 22.98563 23.16346
## 1950 23.75713 23.86354 23.89533 23.87342 23.88150
## 1951 24.48379 24.43879 24.36829 24.29192 24.27642
## 1952 24.44475 24.49325 24.58517 24.70429 24.76017
## 1953 25.34125 25.42779 25.57588 25.73904 25.87513
## 1954 25.98246 26.01054 25.88617 25.67087 25.57312
## 1955 26.40496 26.45379 26.64933 26.95183 27.14683
## 1956 27.44150 27.45229 27.43354 27.44488 27.46996
## 1957 27.56933 27.63167 27.67804 27.62579 27.61212
## 1958 27.46183 27.42262 27.34175 27.25129 27.08558
## 1959       NA       NA       NA       NA       NA
```

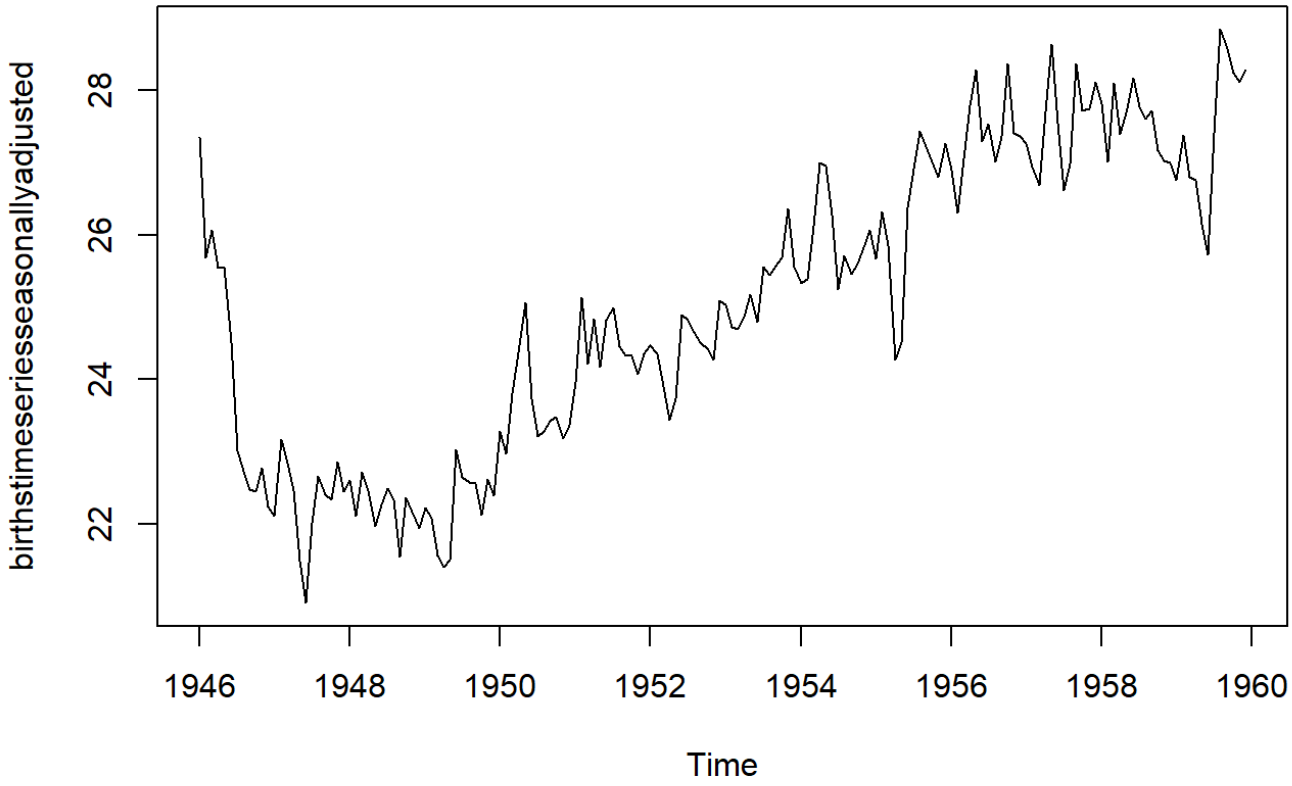
get the estimated values of the trend component

```
plot(birthstimeseriescomponents)
```



First graph is between Observed value and Time
Second graph is between Trend and Time.Trend defines the kind of trend the value is forming over a time period
Third graph is between Seasonal and time.Seasonal componenet
Fourth graph is irregular component

```
birthstimeseriesseasonallyadjusted <- birthstimeseries - birthstimeseriescomponents
$seasonal
plot(birthstimeseriesseasonallyadjusted)
```



Seasonal time series that can be described using an additive model
Seasonal variation has been removed from the seasonally adjusted time series.
The seasonally adjusted time series now just contains the trend component and an irregular component.

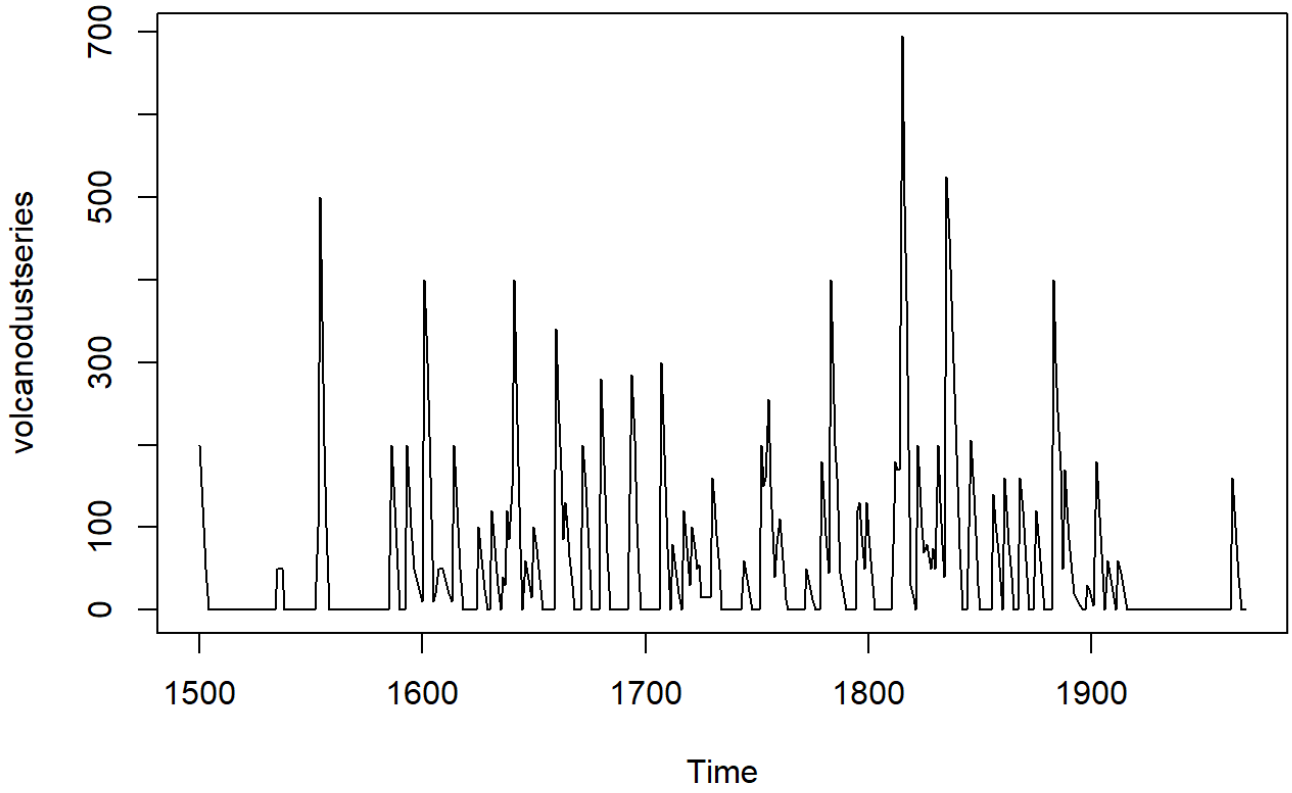
Part B Using Arima

```
volcanodust <- scan("http://robjhyndman.com/tsdldata/annual/dvi.dat", skip=1)

volcanodustseries <- ts(volcanodust,start=c(1500))
class(volcanodustseries)
```

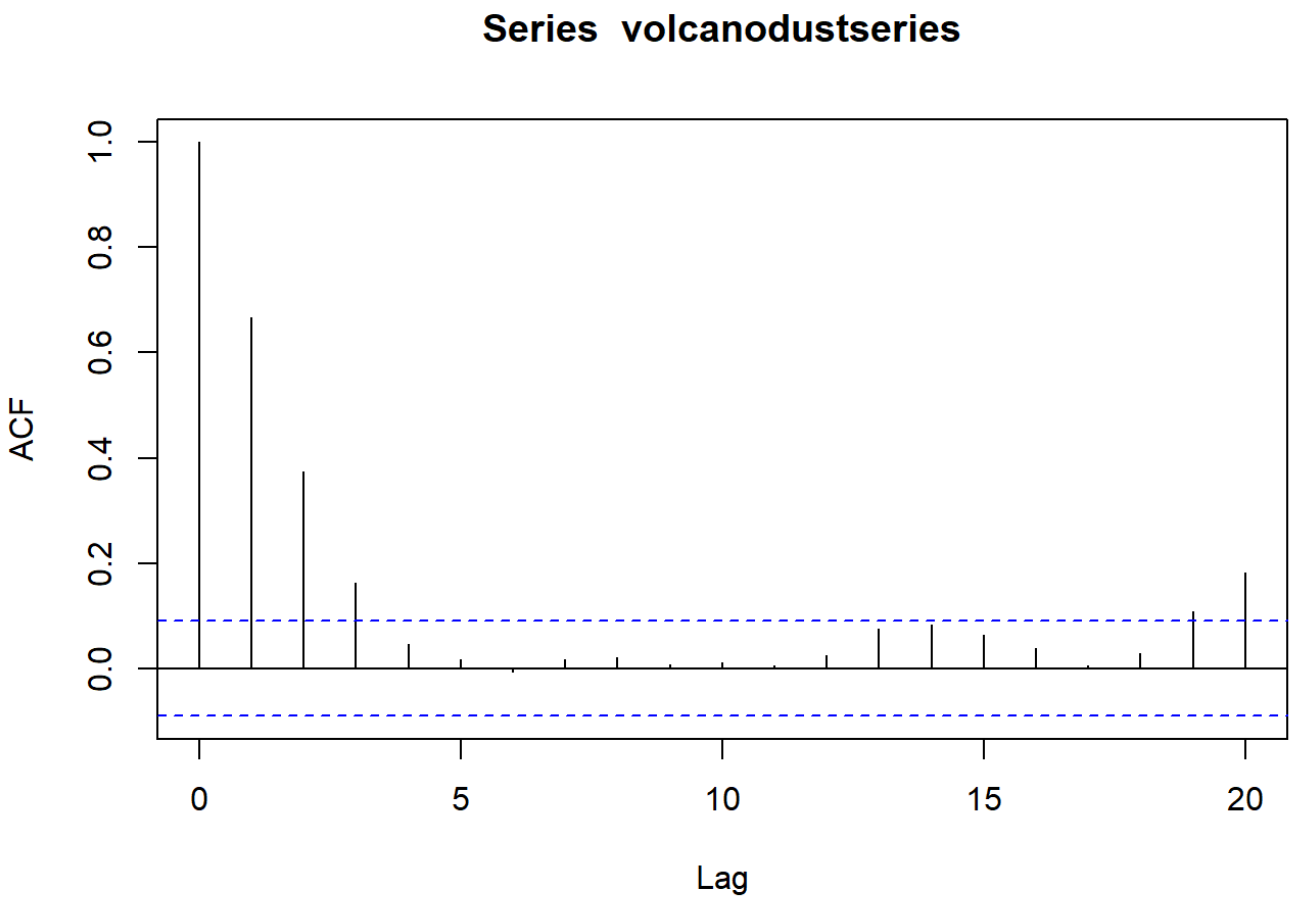
```
## [1] "ts"
```

```
plot.ts(volcanodustseries)
```



From the time plot, it appears that the random fluctuations in the time series are roughly constant in size over time, so an additive model is probably appropriate for describing this time series.

```
acf(volcanodustseries, lag.max=20)
```

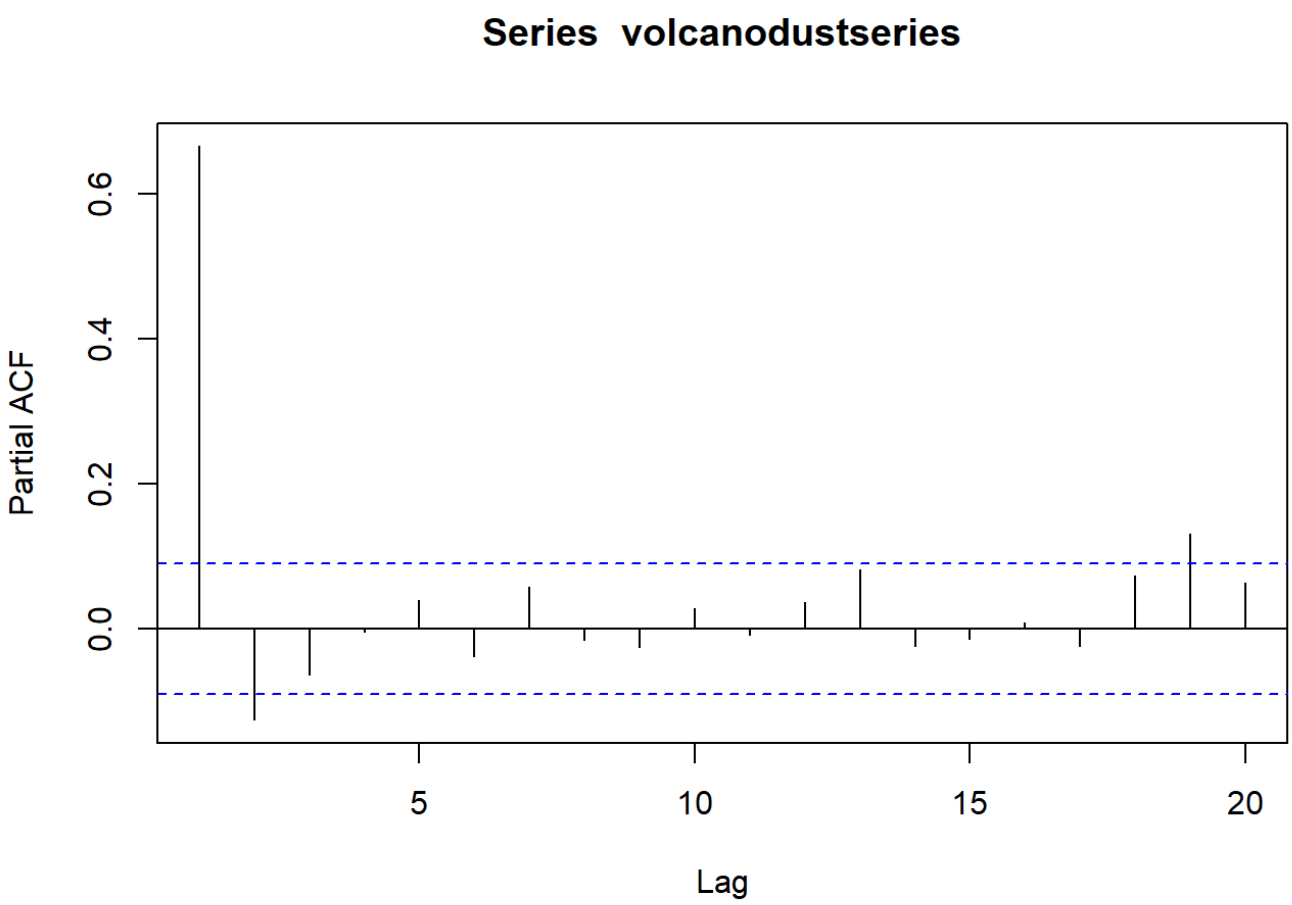



```
acf(volcanodustseries, lag.max=20, plot=FALSE)
```

```
##
## Autocorrelations of series 'volcanodustseries', by lag
##
##      0      1      2      3      4      5      6      7      8      9
##  1.000  0.666  0.374  0.162  0.046  0.017 -0.007  0.016  0.021  0.006
##     10     11     12     13     14     15     16     17     18     19
##  0.010  0.004  0.024  0.075  0.082  0.064  0.039  0.005  0.028  0.108
##     20
##  0.182
```

Autocorrelations for lags 1, 2 and 3 exceed the significance bounds, and that the autocorrelations tail off to zero after lag 3.
The autocorrelations for lags 1, 2, 3 are positive, and decrease in magnitude with increasing lag.

```
pacf(volcanodustseries, lag.max=20)
```




```
pacf(volcanodustseries, lag.max=20, plot=FALSE)
```

```
##
## Partial autocorrelations of series 'volcanodustseries', by lag
##
##      1      2      3      4      5      6      7      8      9     10
##  0.666 -0.126 -0.064 -0.005  0.040 -0.039  0.058 -0.016 -0.025  0.028
##     11     12     13     14     15     16     17     18     19     20
## -0.008  0.036  0.082 -0.025 -0.014  0.008 -0.025  0.073  0.131  0.063
```

Partial autocorrelation at lag 1 is positive and exceeds the significance bounds (0.666), while the partial autocorrelation at lag 2 is negative and also exceeds the significance bounds (-0.126).
The partial autocorrelations tail off to zero after lag 2.

```
volcanodustseriesarima <- arima(volcanodustseries, order=c(2,0,0))
volcanodustseriesarima
```

```
##
## Call:
## arima(x = volcanodustseries, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       0.7533  -0.1268    57.5274
## s.e.  0.0457   0.0458     8.5958
##
## sigma^2 estimated as 4870:  log likelihood = -2662.54,  aic = 5333.09
```

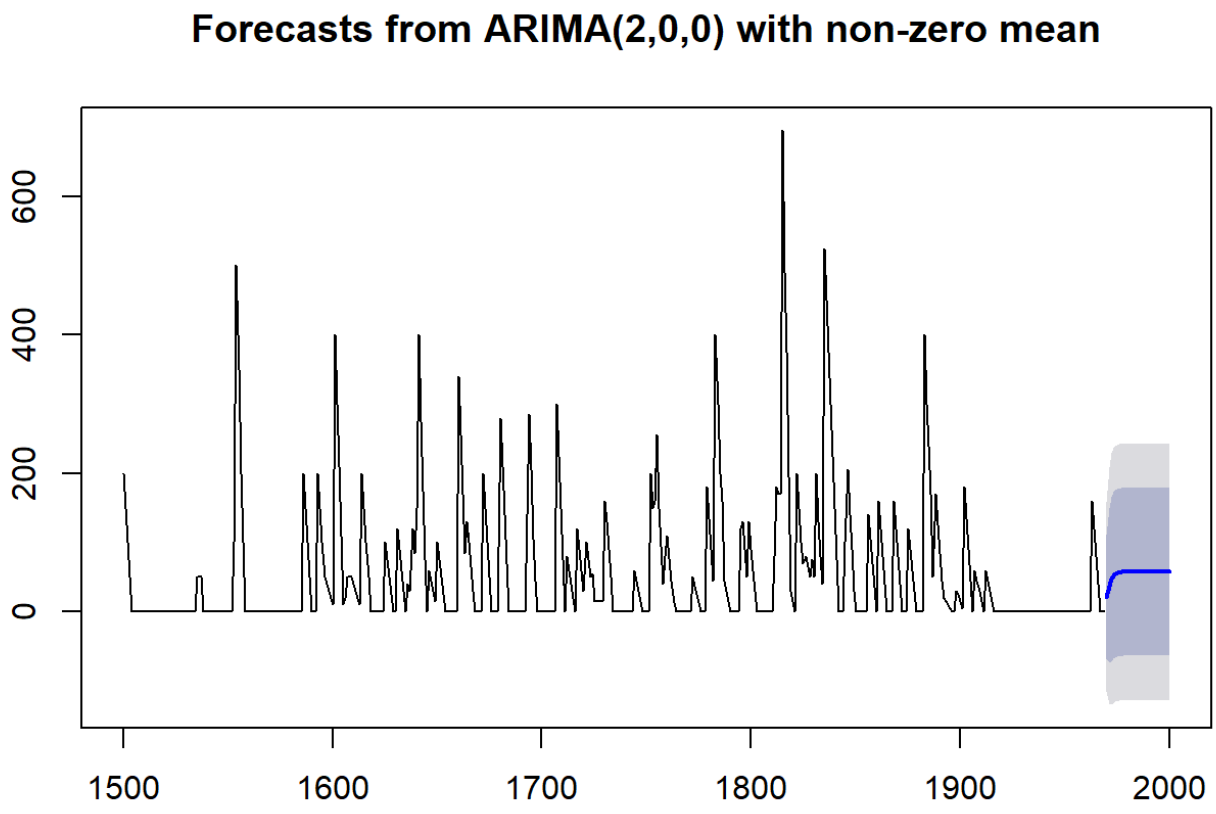
ARIMA model for the time series of volcanic dust veil index may be an ARIMA(2,0,0) model. To fit an ARIMA(2,0,0) model to this time series

```
volcanodustseriesforecasts <- forecast(volcanodustseriesarima, h=31)
volcanodustseriesforecasts
```

```
##      Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## 1970       21.48131 -67.94860 110.9112 -115.2899 158.2526
## 1971       37.66419 -74.30305 149.6314 -133.5749 208.9033
## 1972       47.13261 -71.57070 165.8359 -134.4084 228.6737
## 1973       52.21432 -68.35951 172.7881 -132.1874 236.6161
## 1974       54.84241 -66.22681 175.9116 -130.3170 240.0018
## 1975       56.17814 -65.01872 177.3750 -129.1765 241.5327
## 1976       56.85128 -64.37798 178.0805 -128.5529 242.2554
## 1977       57.18907 -64.04834 178.4265 -128.2276 242.6057
## 1978       57.35822 -63.88124 178.5977 -128.0615 242.7780
## 1979       57.44283 -63.79714 178.6828 -127.9777 242.8634
## 1980       57.48513 -63.75497 178.7252 -127.9356 242.9059
## 1981       57.50627 -63.73386 178.7464 -127.9145 242.9271
## 1982       57.51684 -63.72330 178.7570 -127.9040 242.9376
## 1983       57.52212 -63.71802 178.7623 -127.8987 242.9429
## 1984       57.52476 -63.71538 178.7649 -127.8960 242.9456
## 1985       57.52607 -63.71407 178.7662 -127.8947 242.9469
## 1986       57.52673 -63.71341 178.7669 -127.8941 242.9475
## 1987       57.52706 -63.71308 178.7672 -127.8937 242.9479
## 1988       57.52723 -63.71291 178.7674 -127.8936 242.9480
## 1989       57.52731 -63.71283 178.7674 -127.8935 242.9481
## 1990       57.52735 -63.71279 178.7675 -127.8934 242.9481
## 1991       57.52737 -63.71277 178.7675 -127.8934 242.9482
## 1992       57.52738 -63.71276 178.7675 -127.8934 242.9482
## 1993       57.52739 -63.71275 178.7675 -127.8934 242.9482
## 1994       57.52739 -63.71275 178.7675 -127.8934 242.9482
## 1995       57.52739 -63.71275 178.7675 -127.8934 242.9482
## 1996       57.52739 -63.71275 178.7675 -127.8934 242.9482
## 1997       57.52739 -63.71275 178.7675 -127.8934 242.9482
## 1998       57.52739 -63.71275 178.7675 -127.8934 242.9482
## 1999       57.52739 -63.71275 178.7675 -127.8934 242.9482
## 2000       57.52739 -63.71275 178.7675 -127.8934 242.9482
```

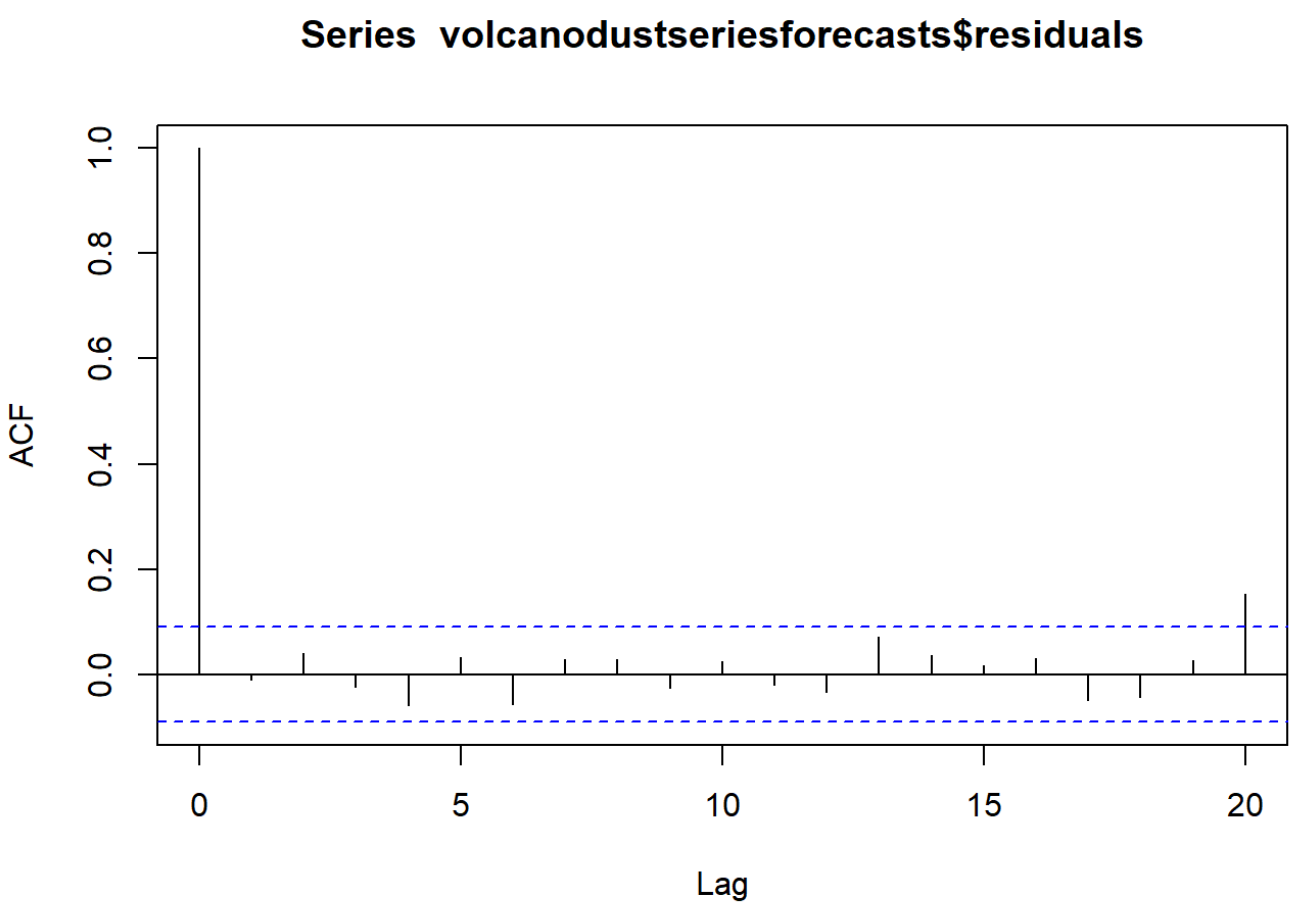
forecast() model to predict future values of the volcanic dust veil index.
The original data includes the years 1500-1969.
To make predictions for the years 1970-2000 (31 more years)

```
plot(forecast(volcanodustseriesforecasts))
```



The model has predicted negative values for the volcanic dust veil index, but this variable can only have positive values

```
acf(volcanodustseriesforecasts$residuals, lag.max=20)
```

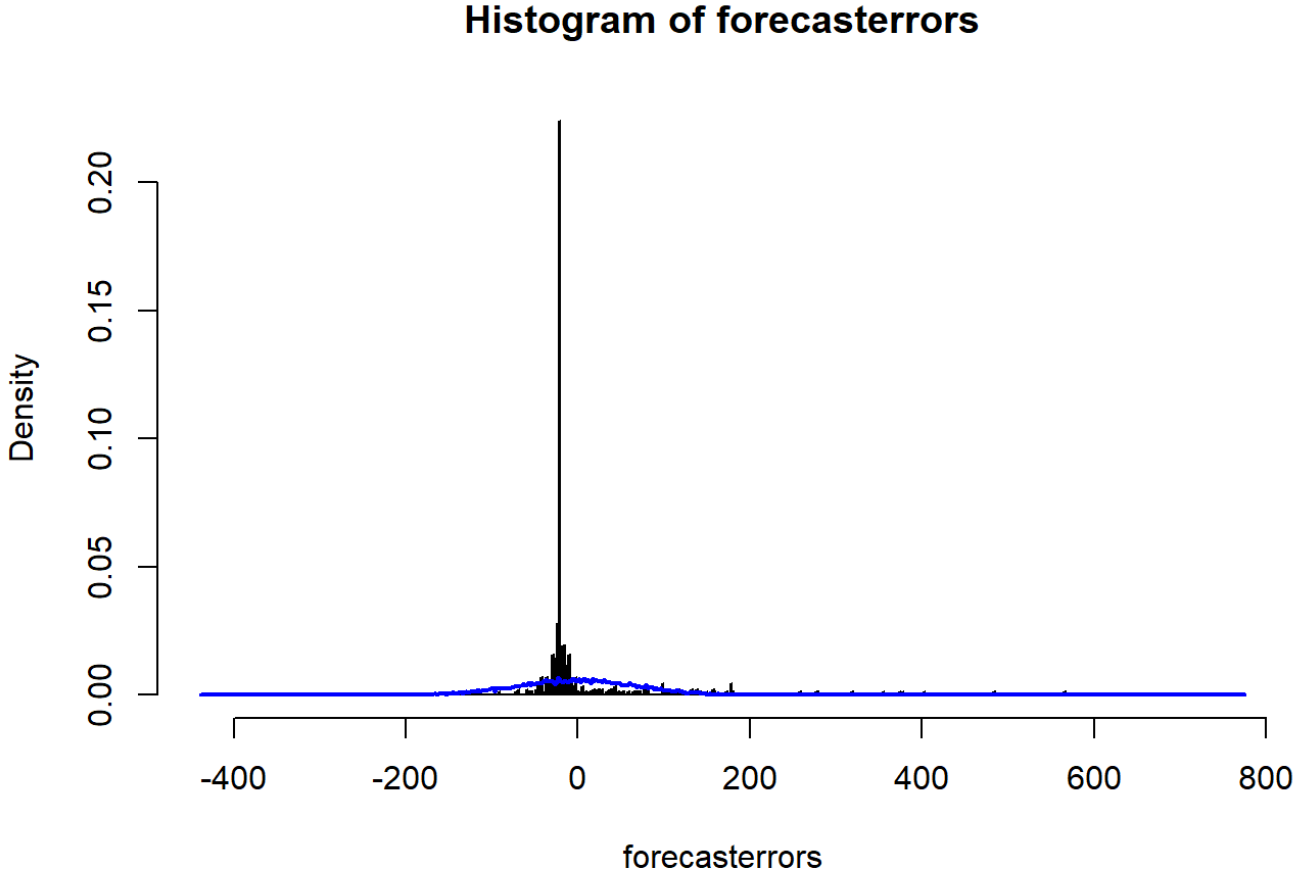



```
Box.test(volcanodustseriesforecasts$residuals, lag=20, type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  volcanodustseriesforecasts$residuals
## X-squared = 24.364, df = 20, p-value = 0.2268
```

Correlogram shows that the sample autocorrelation at lag 20 exceeds the significance bounds.

```
plotForecastErrors(volcanodustseriesforecasts$residuals)
```




```
mean(volcanodustseriesforecasts$residuals)
```

```
## [1] -0.2205417
```

Time plot of forecast errors shows that the forecast errors seem to have roughly constant variance over time. However, the time series of forecast errors seems to have a negative mean, rather than a zero mean. We can confirm this by calculating the mean forecast error, which turns out to be about -0.22: