# Time Series Analysis in R

Jazmyn Singh

2020-01-10

## Introduction

This analysis demonstrates how to handle, explore, and model a time series dataset in R.
We will use the built-in **AirPassengers** dataset as an example, which contains monthly airline passenger totals from 1949 to 1960.

---

## Load and Explore Data

```
# Load dataset
data("AirPassengers")
ts_data <- AirPassengers

# Basic overview
summary(ts_data)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   104.0   180.0   265.5   280.3   360.5   622.0
```

```
start(ts_data); end(ts_data); frequency(ts_data)
```

```
## [1] 1949    1
```

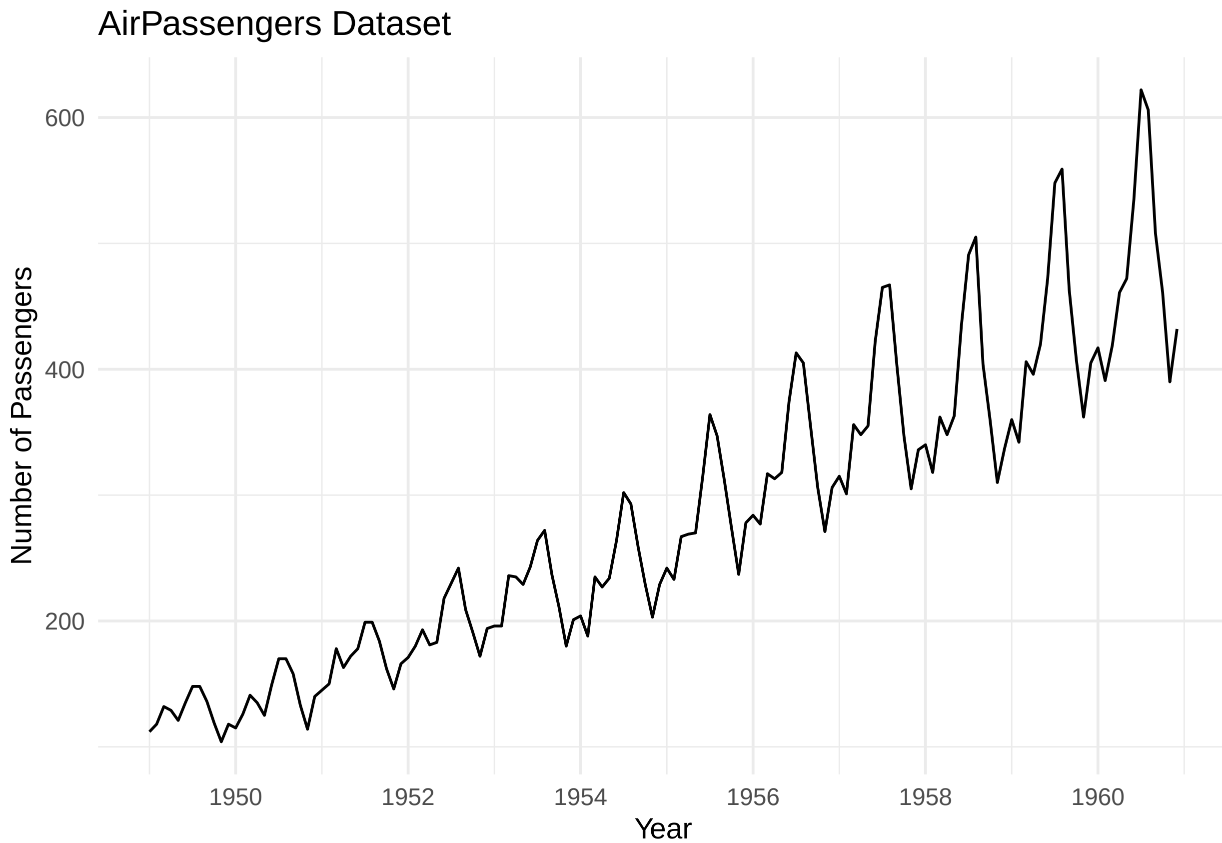
```
## [1] 1960   12
```

```
## [1] 12
```

---

## Plot the Time Series

```
autoplot(ts_data) +
  labs(title="AirPassengers Dataset",
       x="Year", y="Number of Passengers") +
  theme_minimal()
```

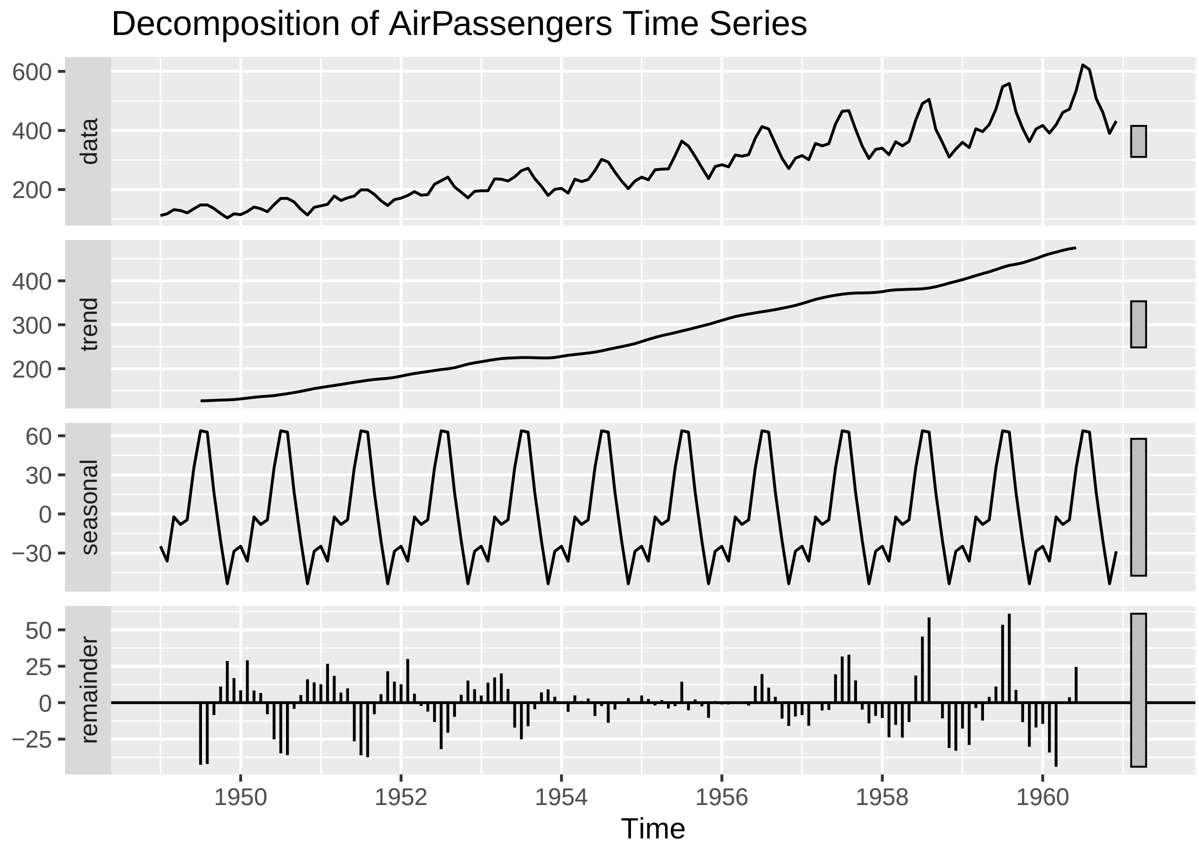



*Comment:* We can see a strong upward trend and seasonality pattern.

---

### Decomposition

```
decomp <- decompose(ts_data)
autoplot(decomp) + labs(title="Decomposition of AirPassengers Time Series")
```




*Comment:* This shows **trend**, **seasonality**, and **residuals** separately.

---

## Stationarity Check

```
adf_test <- adf.test(ts_data)
adf_test
```

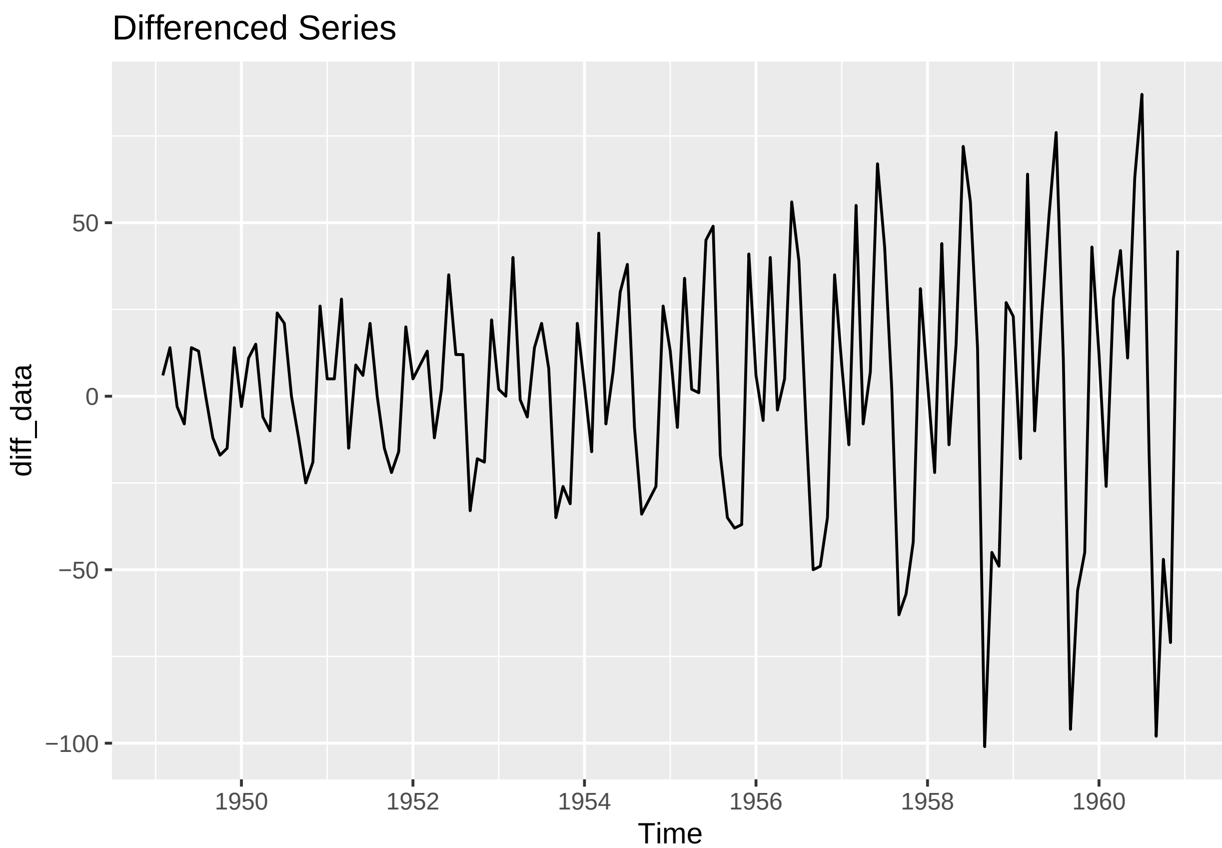
```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts_data
## Dickey-Fuller = -7.3186, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

*Comment:* If the p-value $> 0.05$, the series is **not stationary** and needs differencing.

---

## Differencing to Achieve Stationarity

```r
diff_data <- diff(ts_data)
autoplot(diff_data) + labs(title="Differenced Series")
```




```r
adf.test(diff_data)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff_data
## Dickey-Fuller = -7.0177, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

*Comment:* After differencing, the series should become stationary.

---

## ARIMA Model

```r
fit <- auto.arima(ts_data)
summary(fit)
```

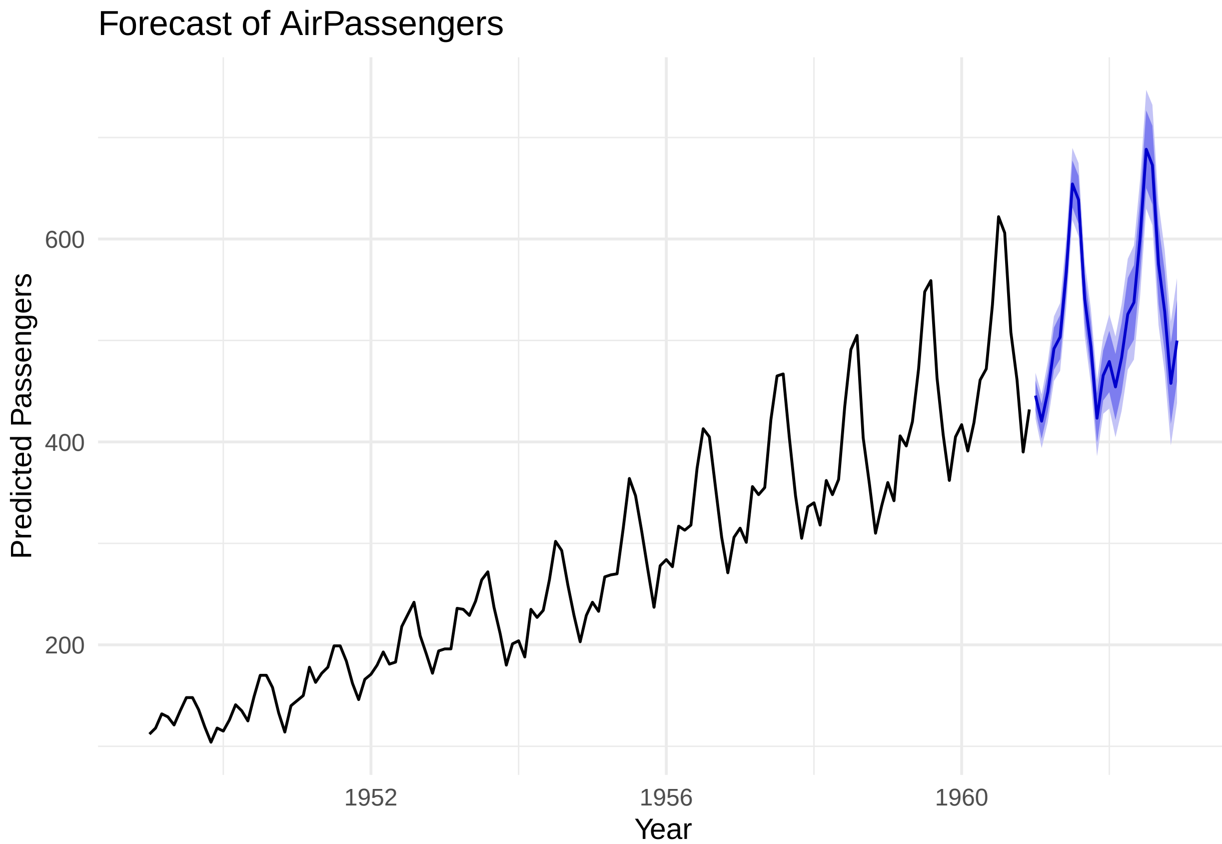
```
## Series: ts_data
## ARIMA(2,1,1)(0,1,0)[12]
##
## Coefficients:
##          ar1     ar2      ma1
##       0.5960  0.2142  -0.9819
## s.e.  0.0888  0.0880   0.0292
##
## sigma^2 = 132.3:  log likelihood = -504.92
## AIC=1017.85   AICc=1018.17   BIC=1029.35
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE    MAPE     MASE          ACF1
## Training set 1.342189 10.84619 7.867506 0.4206662 2.80045 0.245627 -0.001286643
```

*Comment:* `auto.arima()` automatically selects the best model based on AIC/BIC.

---

## Forecasting

```r
fc <- forecast(fit, h=24)  # forecast next 24 months
autoplot(fc) +
  labs(title="Forecast of AirPassengers",
       x="Year", y="Predicted Passengers") +
  theme_minimal()
```

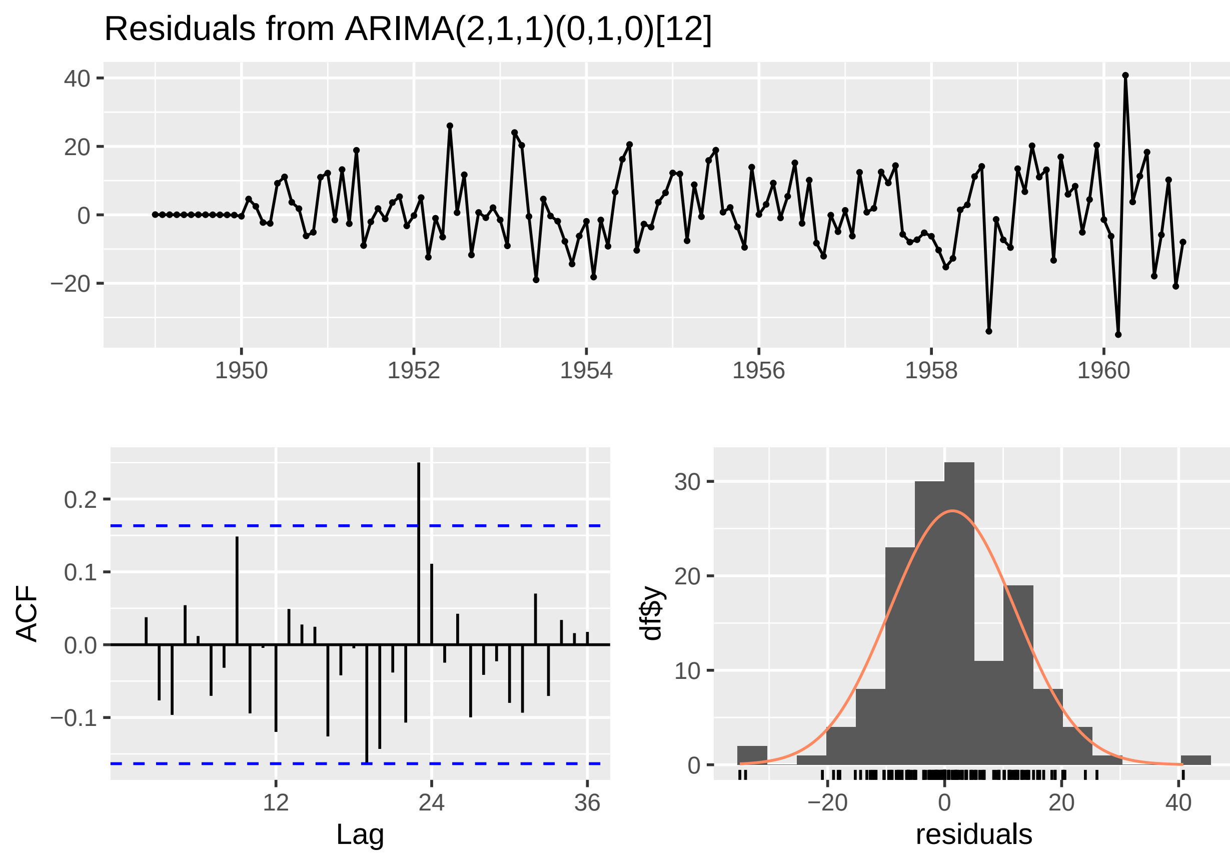



*Comment:* The forecast shows continued growth with seasonality.

---

### Residual Diagnostics

```
checkresiduals(fit)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,1)(0,1,0)[12]
## Q* = 37.784, df = 21, p-value = 0.01366
##
## Model df: 3.   Total lags used: 24
```

*Comment:* This helps ensure residuals are white noise (no autocorrelation left).