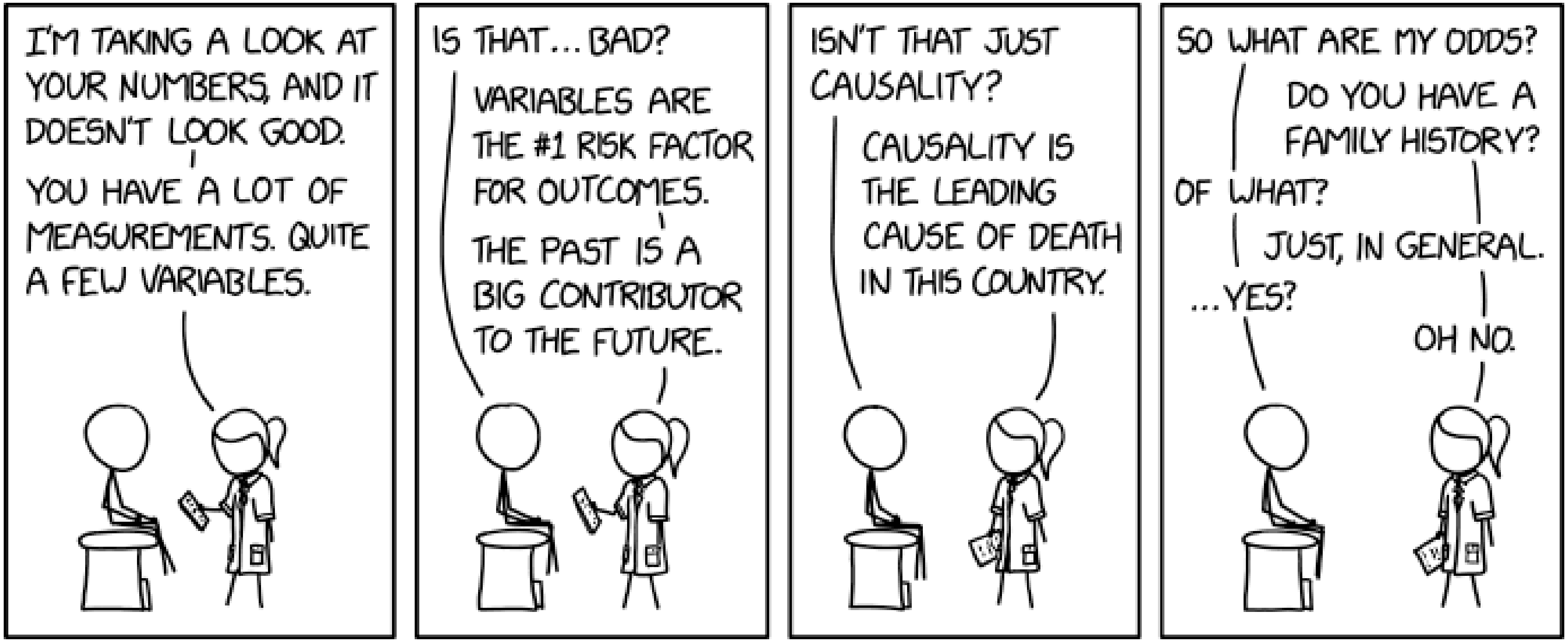
# INTRODUCTION TO FORECASTING & TIME SERIES STRUCTURE

Dr. Susan Simmons

Institute for Advanced Analytics




Source: xkcd.com/2620

# TIME SERIES DATA

# Time Series Data

- A time series is an ordered sequence of observations.
  - Ordering is typically through ***equally spaced*** time intervals.
  - Possibly through space as well.
- Used in a variety of fields:
  - Agriculture: Crop Production
  - Economics: Stock Prices
  - Engineering: Electric Signals
  - Meteorology: Wind Speeds
  - Social Sciences: Crime Rates

# Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

# Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

| Date | Y |
|---|---|
| January 2000 | 23 |
| February 2000 | 18 |
| March 2000 | 20 |
| April 2000 | 25 |
| May 2000 | 21 |

# Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

| Date | Y |
|---|---|
| January 2000 | 23 |
| February 2000 | 18 |
| March 2000 | 20 |
| April 2000 | 25 |
| May 2000 | 21 |

$Y_1$

# Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

| Date | Y |
|---|---|
| January 2000 | 23 |
| February 2000 | 18 |
| March 2000 | 20 |
| April 2000 | 25 |
| May 2000 | 21 |

$Y_2$

# Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

| Date | Y |
|---|---|
| January 2000 | 23 |
| February 2000 | 18 |
| March 2000 | 20 |
| April 2000 | 25 |
| May 2000 | 21 |

$Y_3$

# Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

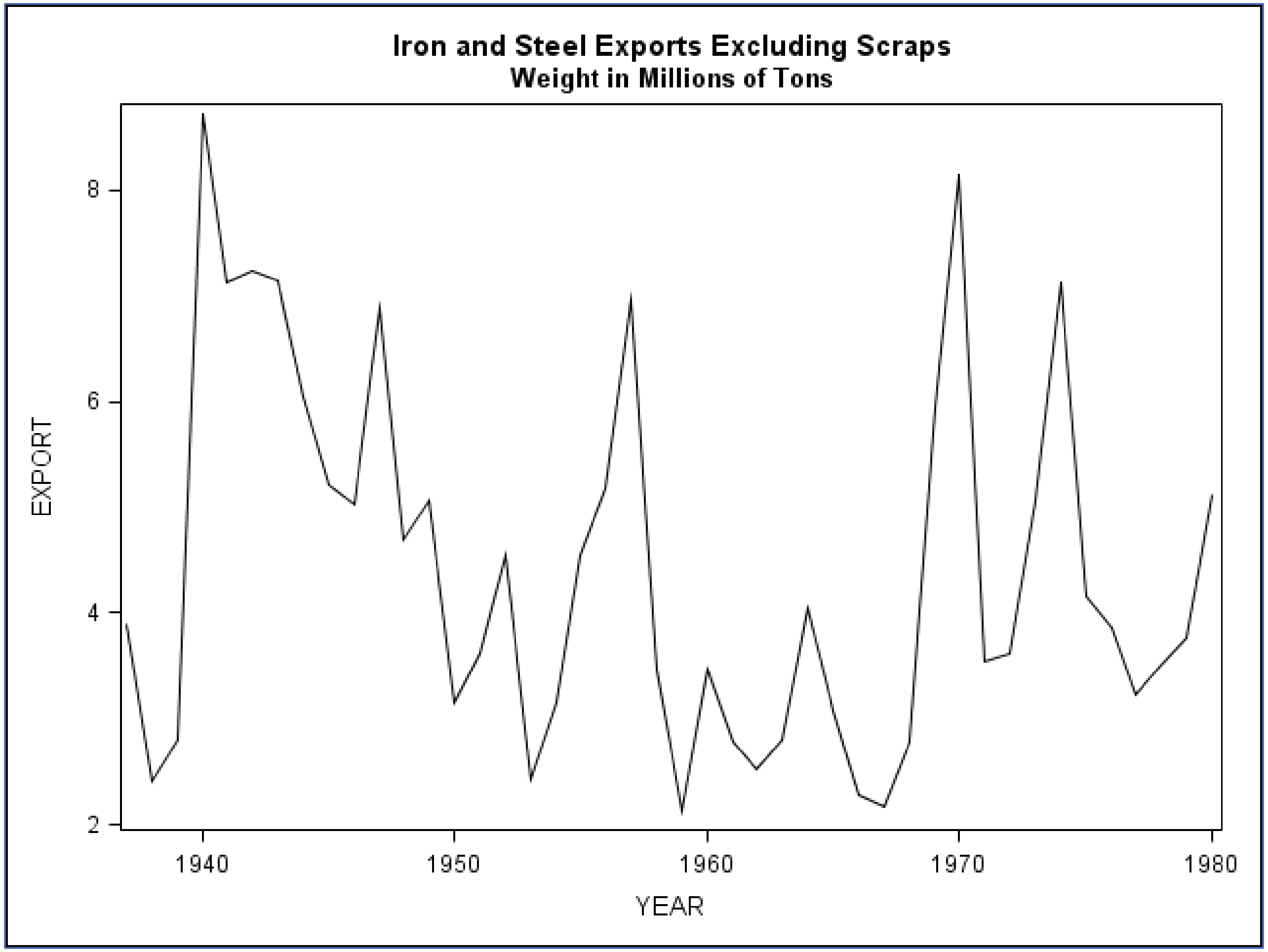
| Date | Y |
|---|---|
| January 2000 | 23 |
| February 2000 | 18 |
| March 2000 | 20 |
| April 2000 | 25 |
| May 2000 | 21 |

$Y_3$

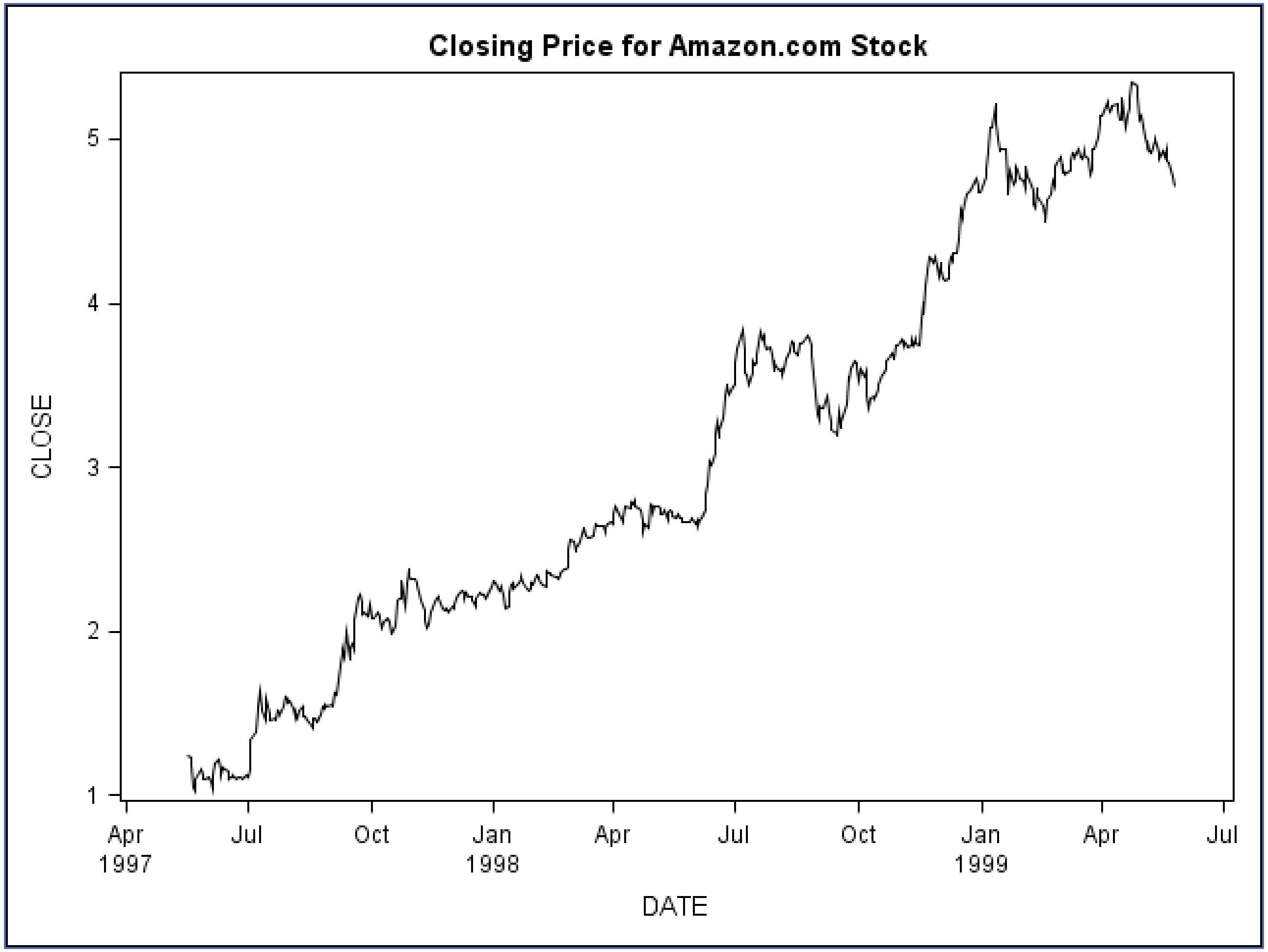
$Y_t$

***CAREFUL: Since we are assuming equally spaced, you will need to take care of missing values !!***

# Example 1: Iron and Steel Exports

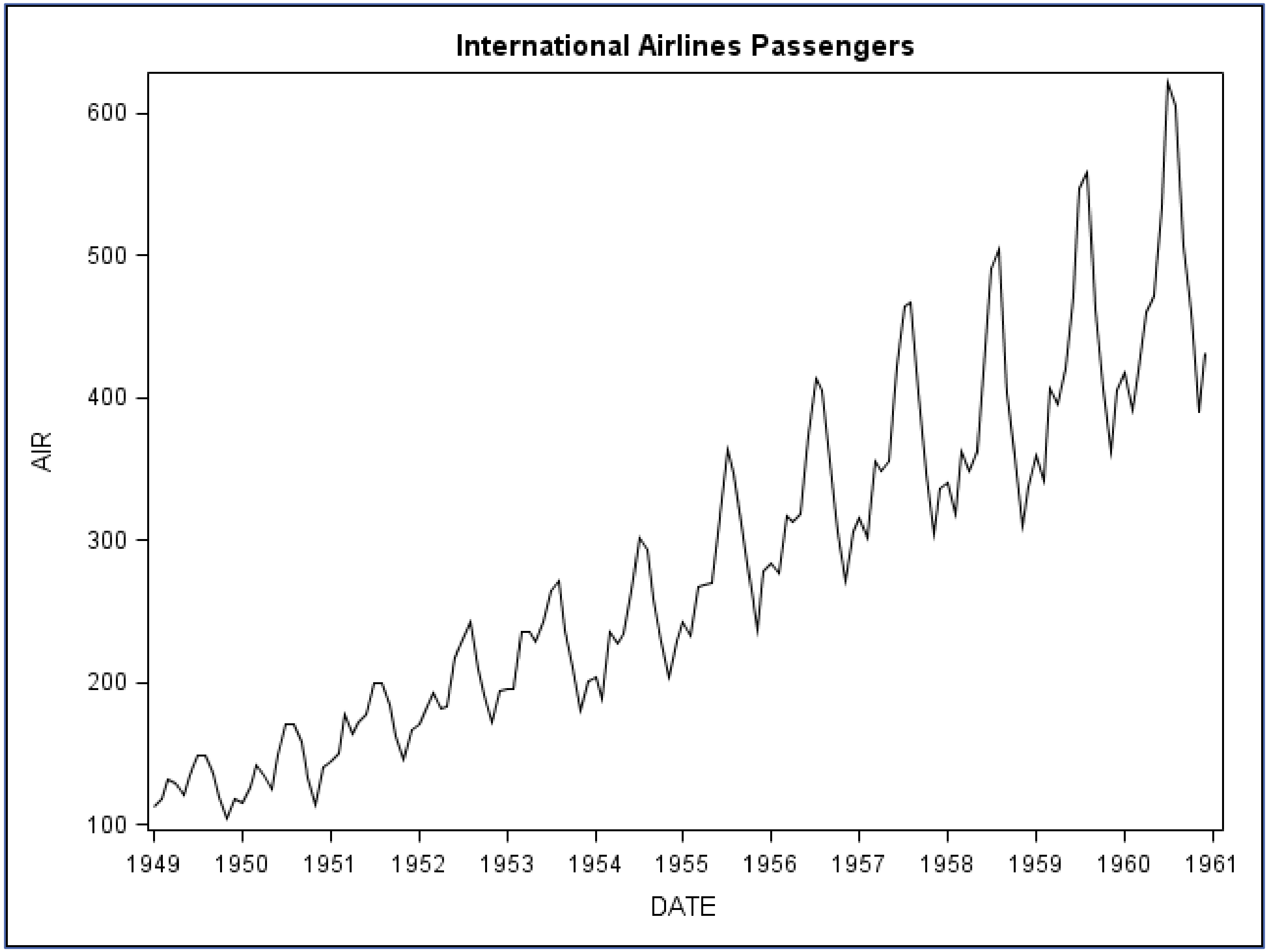


# Example 2: Amazon.com Stock

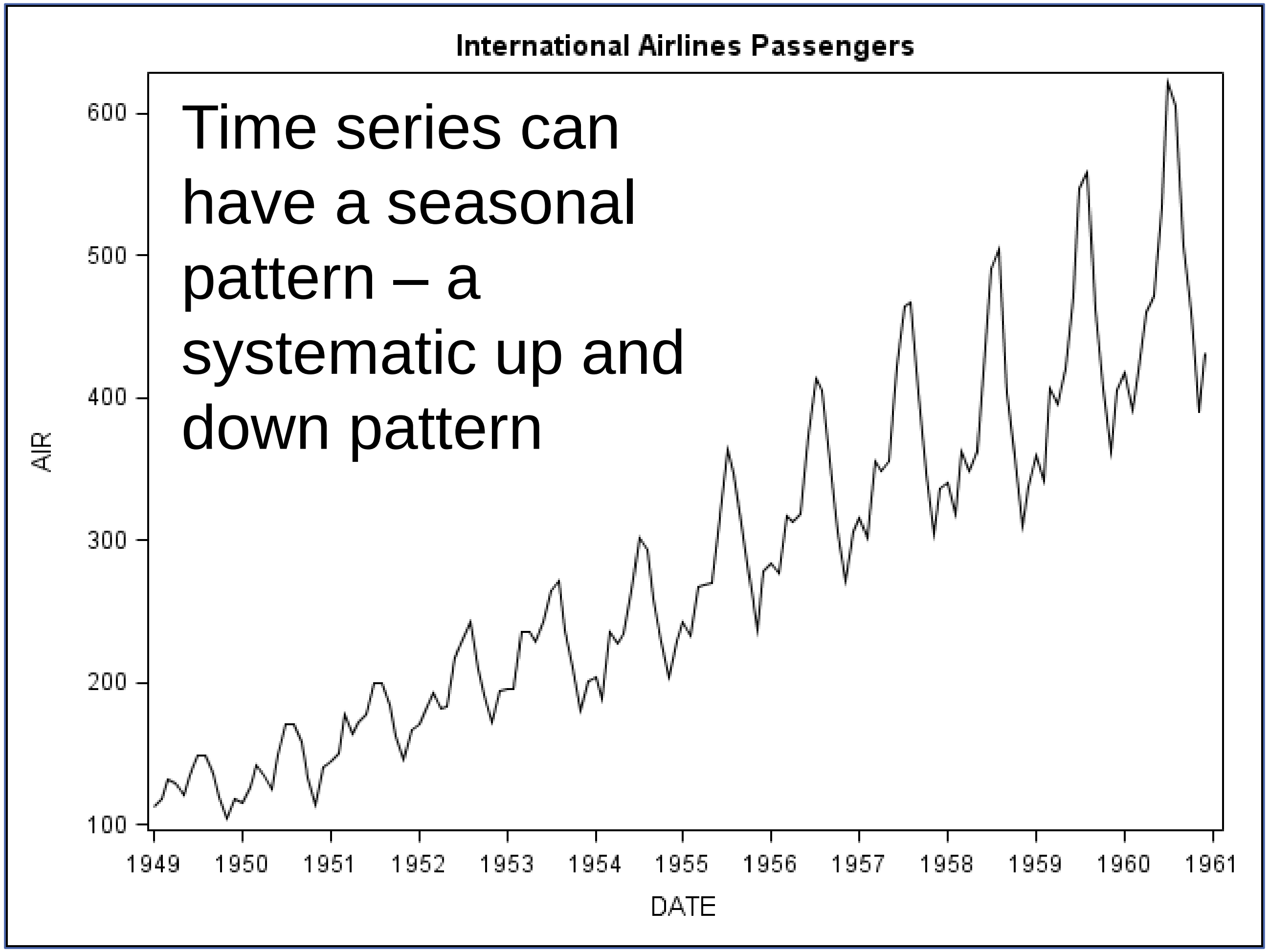


# Example 2: Amazon.com Stock

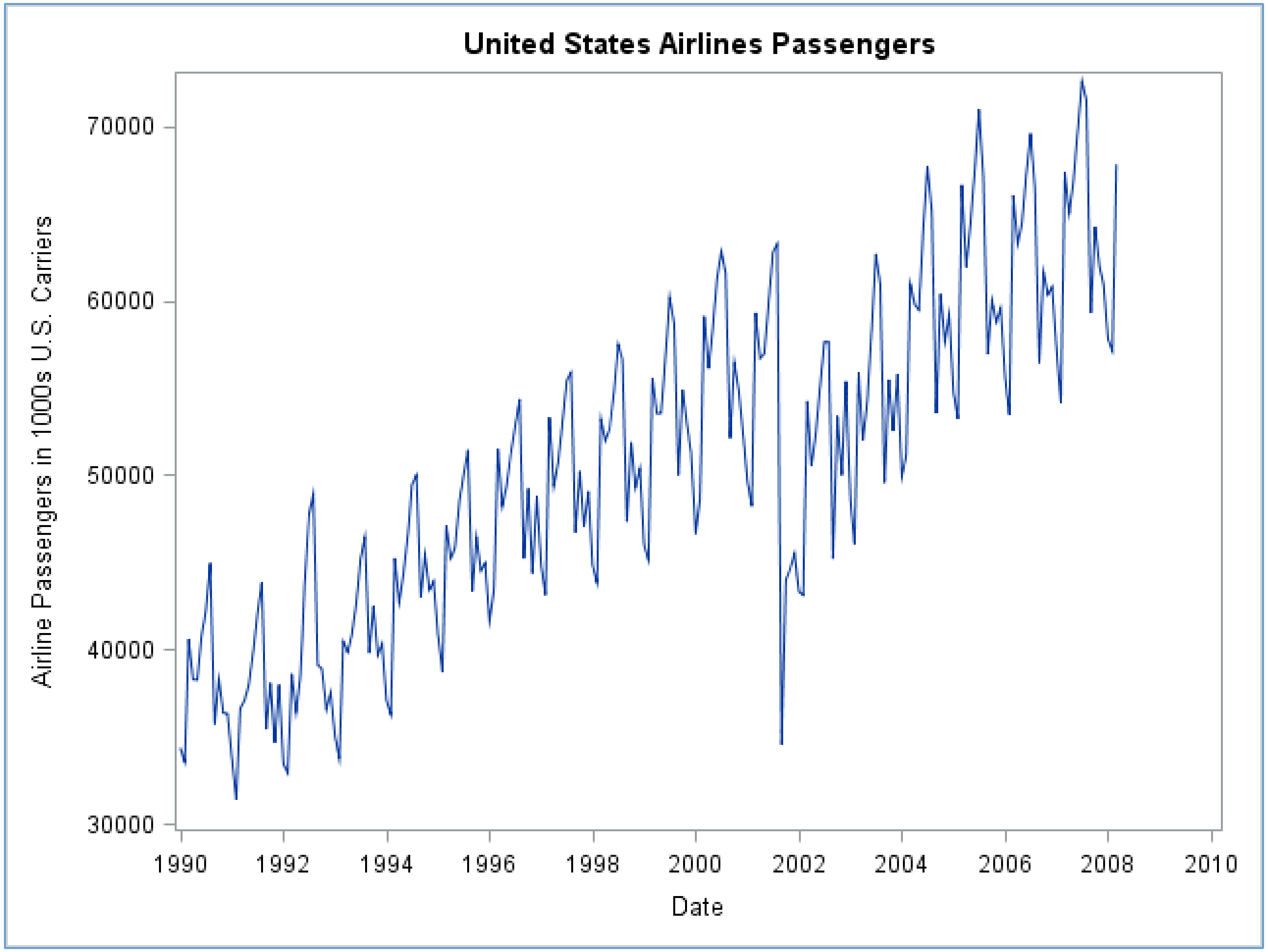
# Example 3: Airlines Passengers



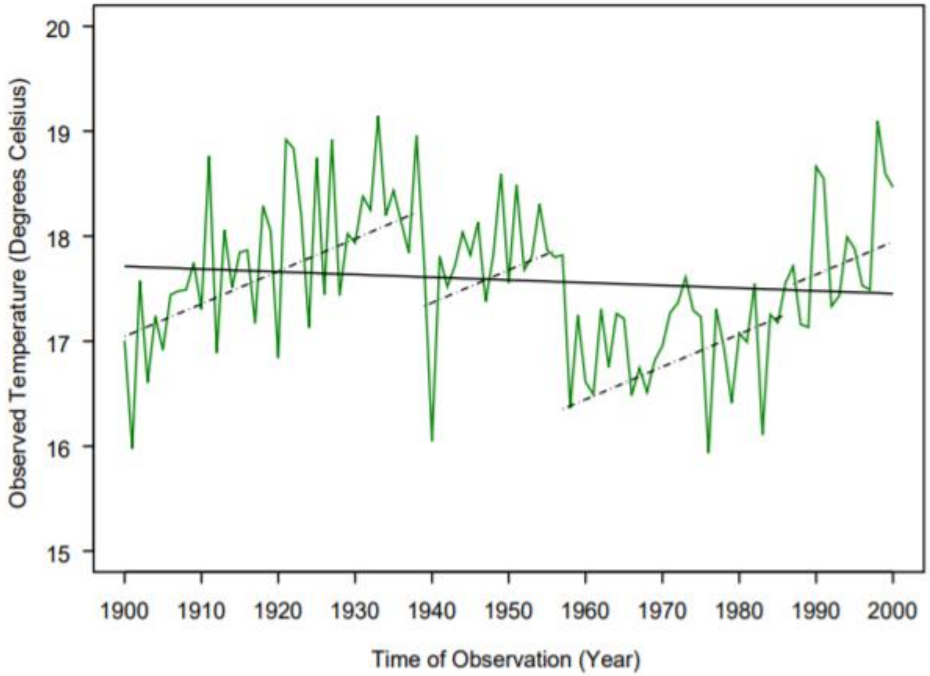
# Example 3: Airlines Passengers



# Example 5: Airline Passengers Again



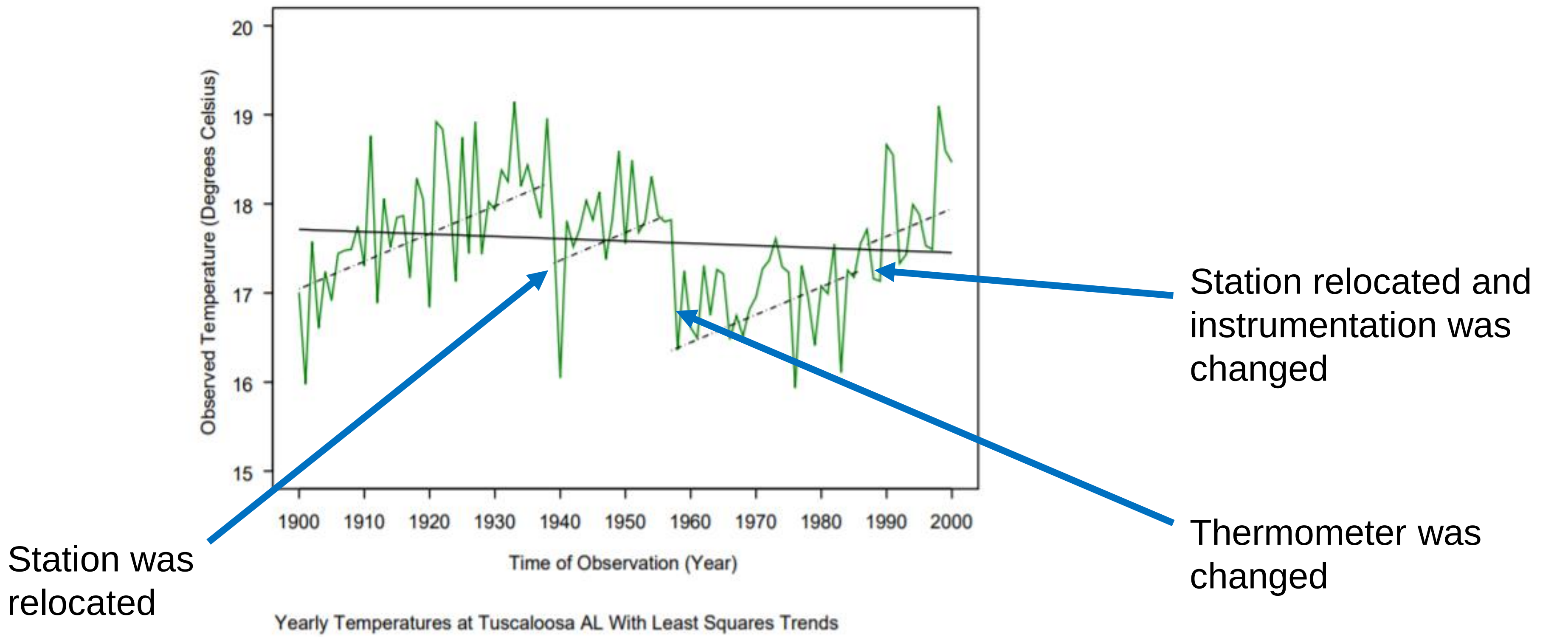
# Temperature over the past century for Tuscaloosa, Alabama




Yearly Temperatures at Tuscaloosa AL With Least Squares Trends

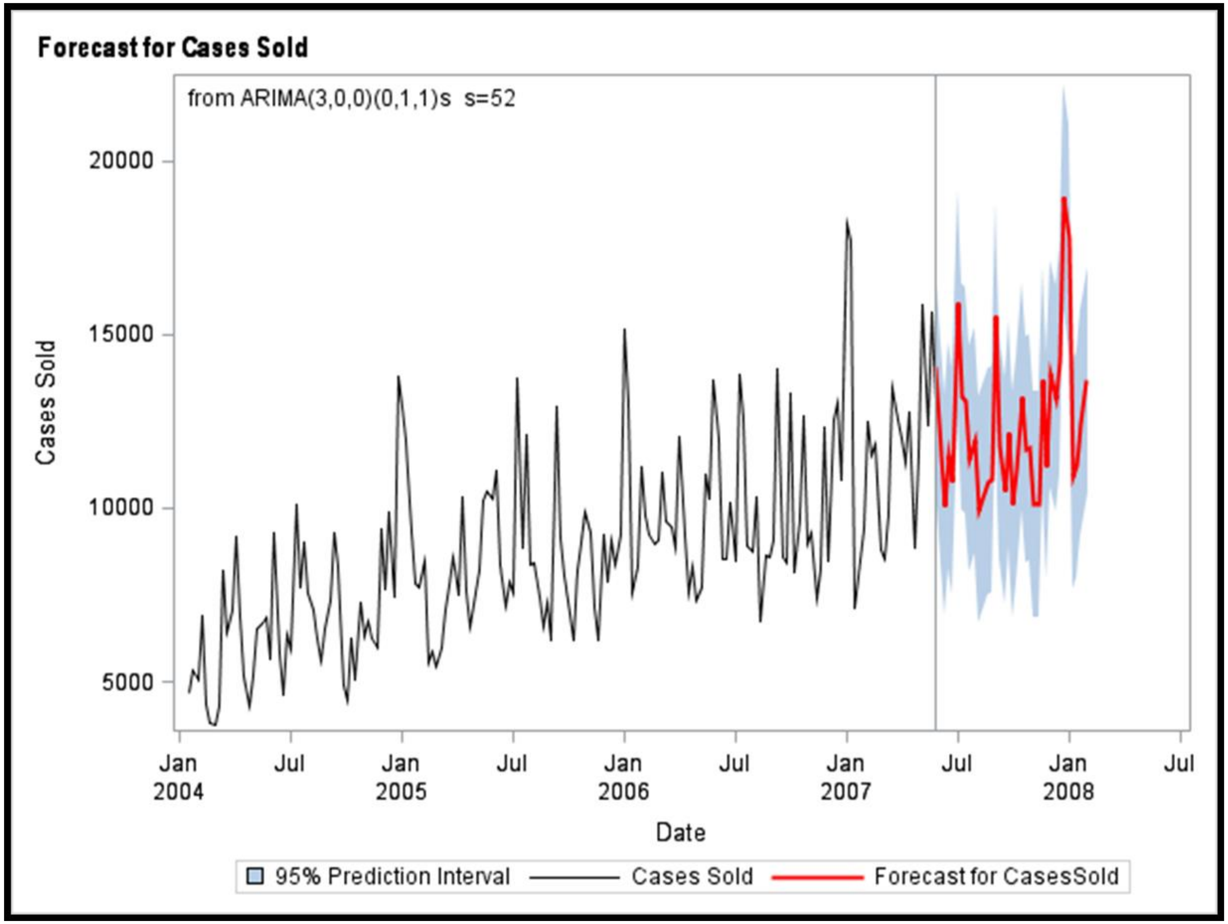
Source: Dr. Robert Lund

# Temperature over the past century for Tuscaloosa, Alabama

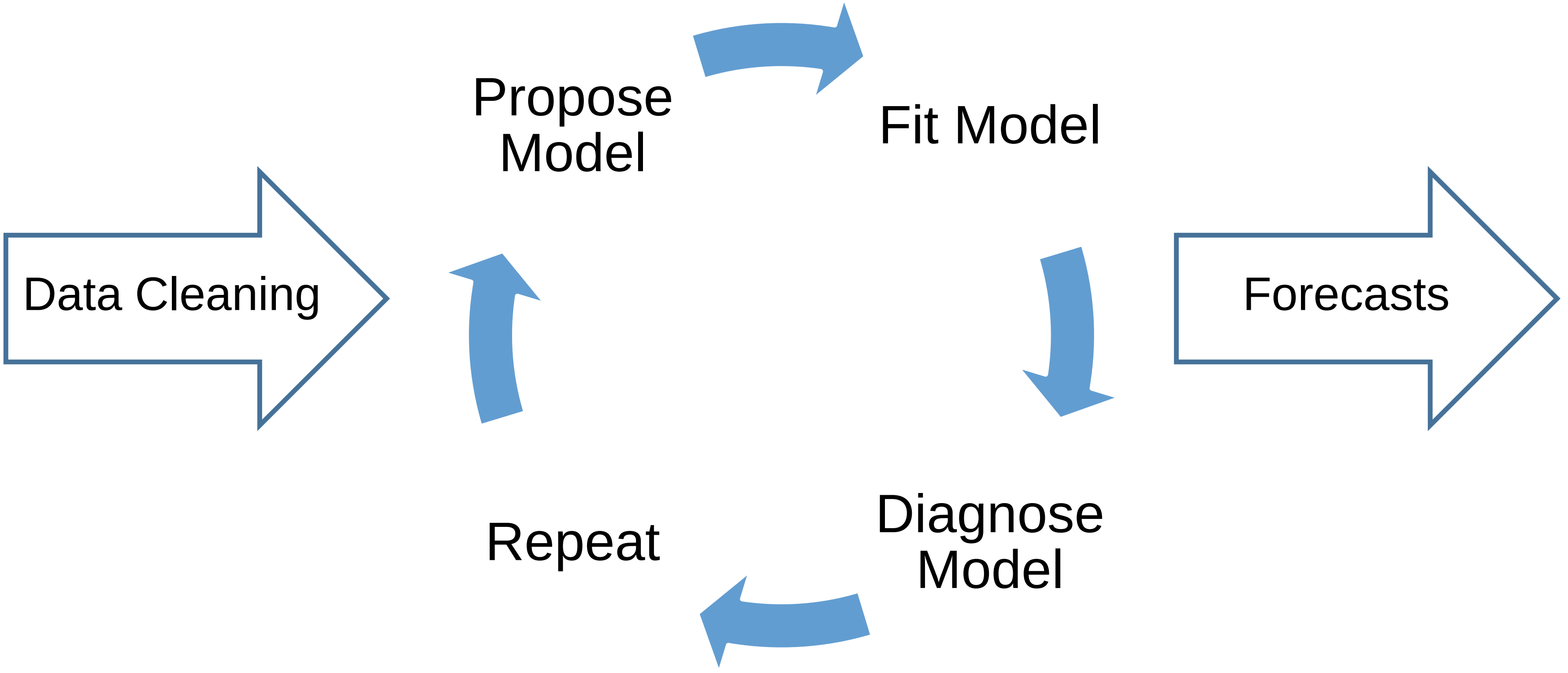



Source: Dr. Robert Lund

# Time Series to Forecast



# Forecasting Process

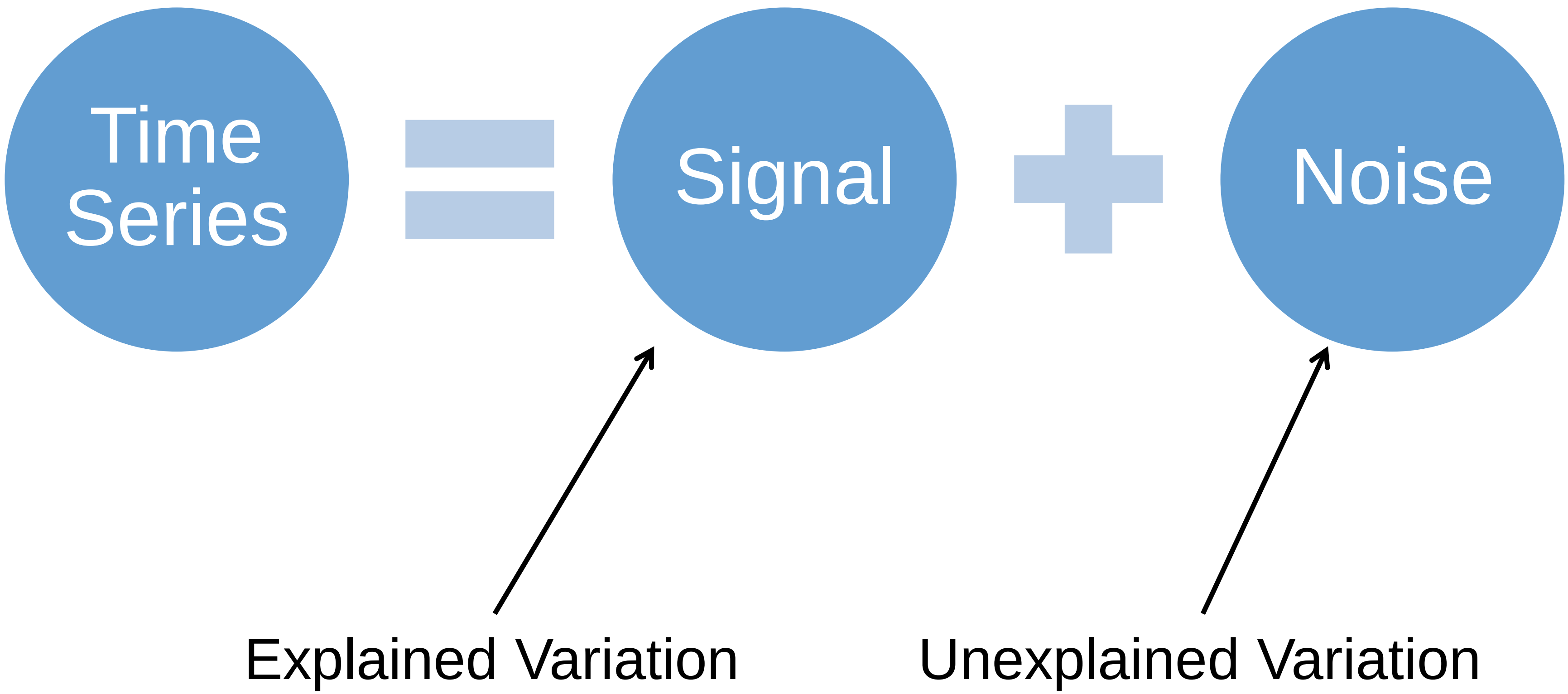


# SIGNAL AND NOISE

# Statistical Forecasting

Time Series = Signal + Noise

# Statistical Forecasting



# Statistical Forecasting

Time Series = Signal + Noise

Explained variation:
Trend/Cycle
Seasonality

Error

# Statistical Forecasting

Time Series = Signal + Noise

Explained variation:
Trend/Cycle
Seasonality
Dependence Structure

Error

This will be discussed in ARIMA

# Statistical Forecasting

Time Series = Signal + Noise

Forecasts extrapolate signal portion of model.

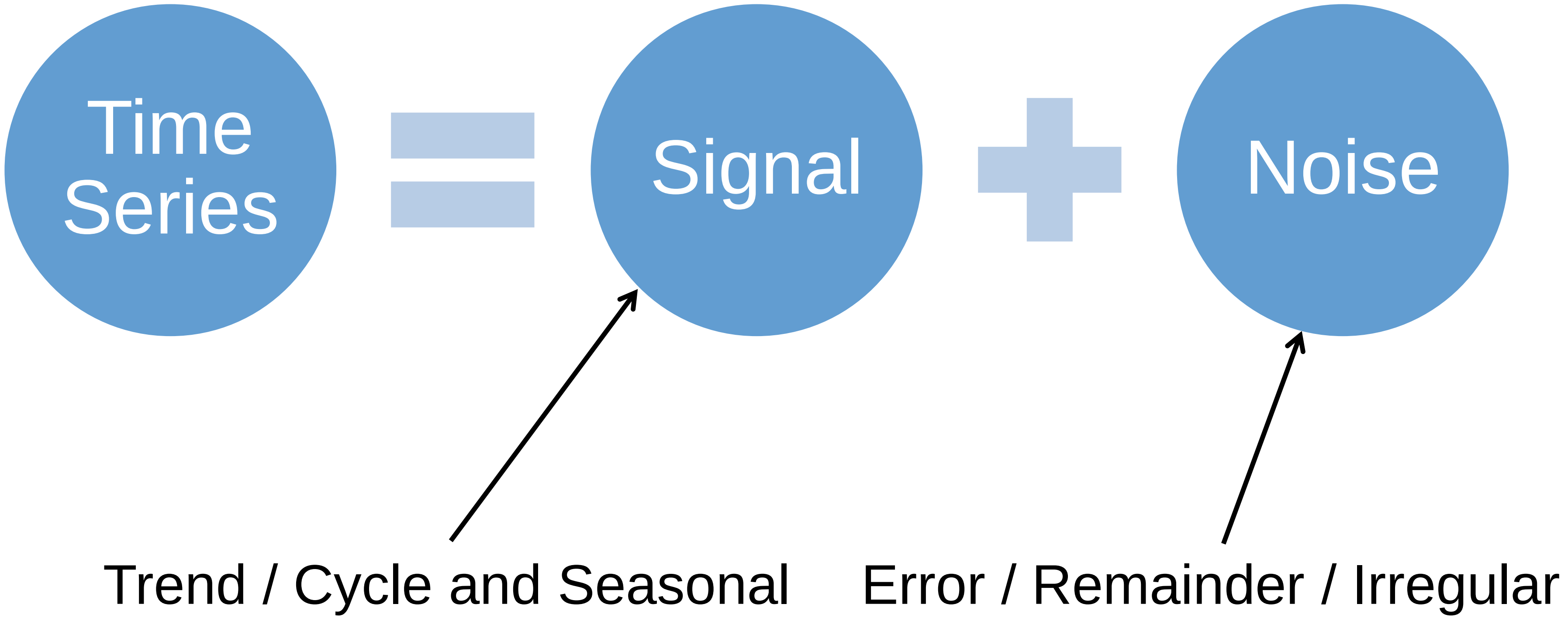
Confidence intervals account for uncertainty.

# Time Series Decomposition

- If a time series only has trend/cycle patterns, there is no need to decompose
- If a time series has both trend/cycle patterns AND seasonal variation, we can decompose series into these individual parts:
  - Trend/Cycle patterns
  - Seasonal variation
  - Error

# Time Series Decomposition

- The signal part of the time series can typically be broken down into two components:



# Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

# Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

Trend / Cycle

  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

# Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

Seasonal

  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

# Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

Error

  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

# Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

  - Multiplicative:

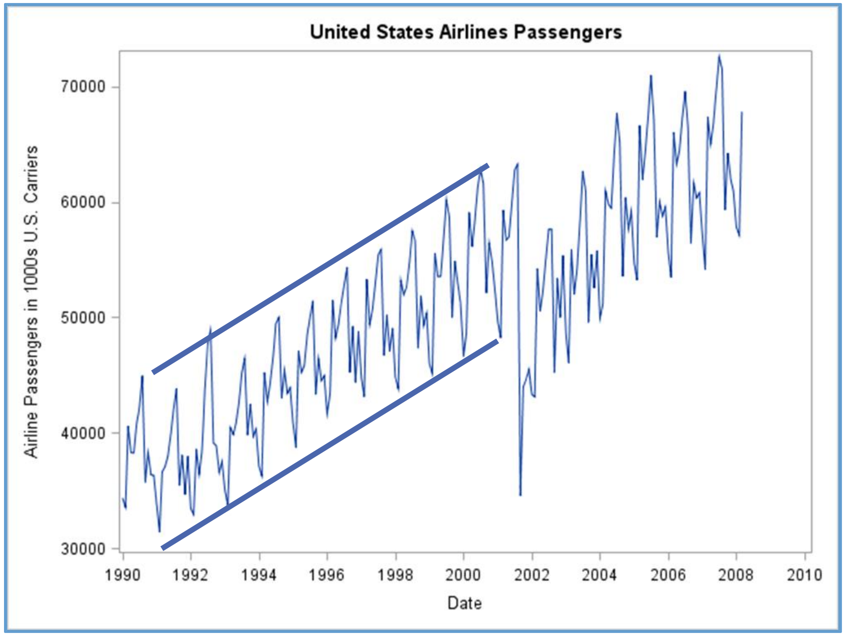
$$Y_t = T_t \times S_t \times E_t$$

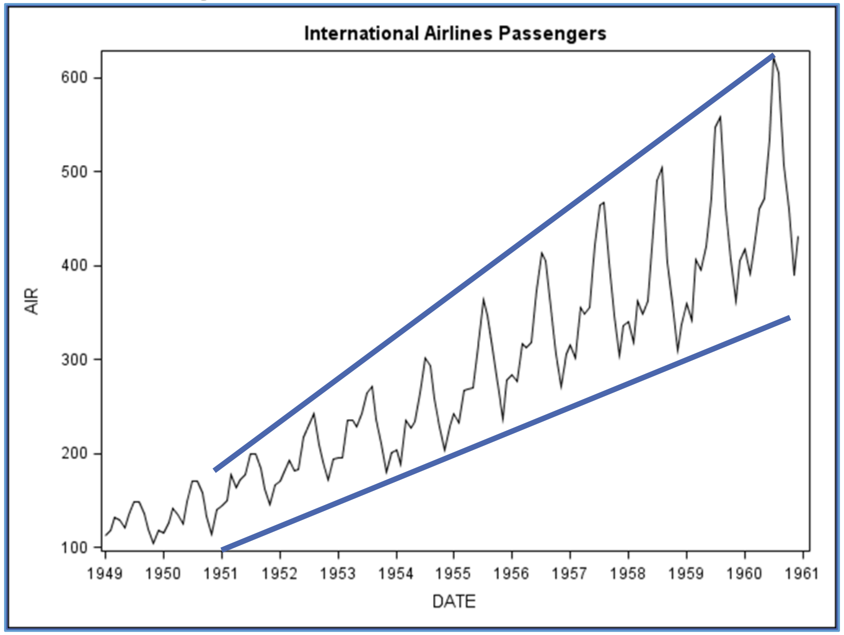
OR

$$\log(Y_t) = \log(T_t) + \log(S_t) + \log(E_t)$$

# Additive vs. Multiplicative

- Additive – magnitude of variation around trend / cycle remains constant.
- Multiplicative – magnitude of the variation around trend / cycle proportionally changes.



# Seasonally Adjusted Data

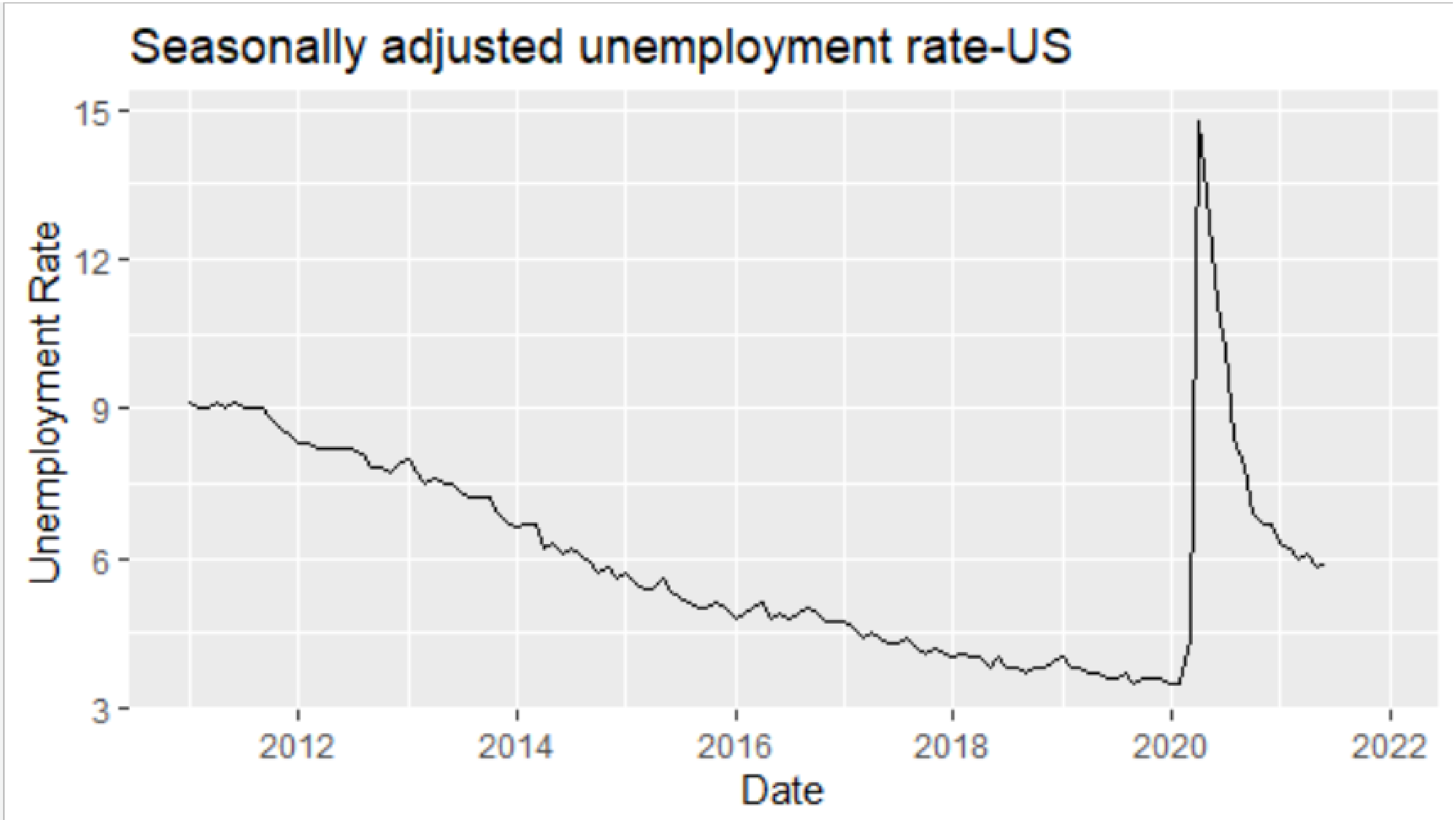
One advantage of time series decomposition is that we are able to create seasonally adjusted data (i.e. remove the “effect of Seasonality”)

This allows analysts to understand the trend of the series

$$Y_t = T_t + S_t + E_t$$

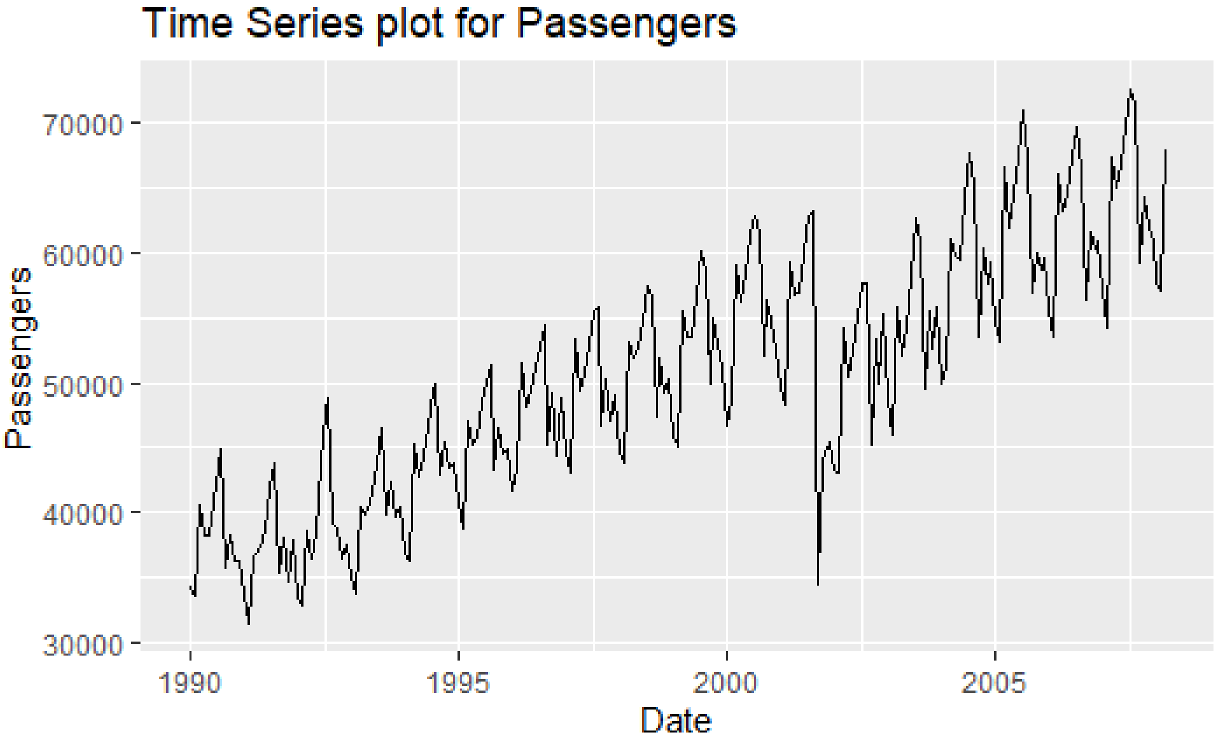
$$Y_t - S_t \qquad\qquad (T_t + E_t)$$

The seasonal length of the time series is the length of one season (how long til the series repeats the “pattern”)


Seasonally adjusted unemployment rate-US
15
12
9
6
3
Unemployment Rate
2012
2014
2016
2018
2020
2022
Date

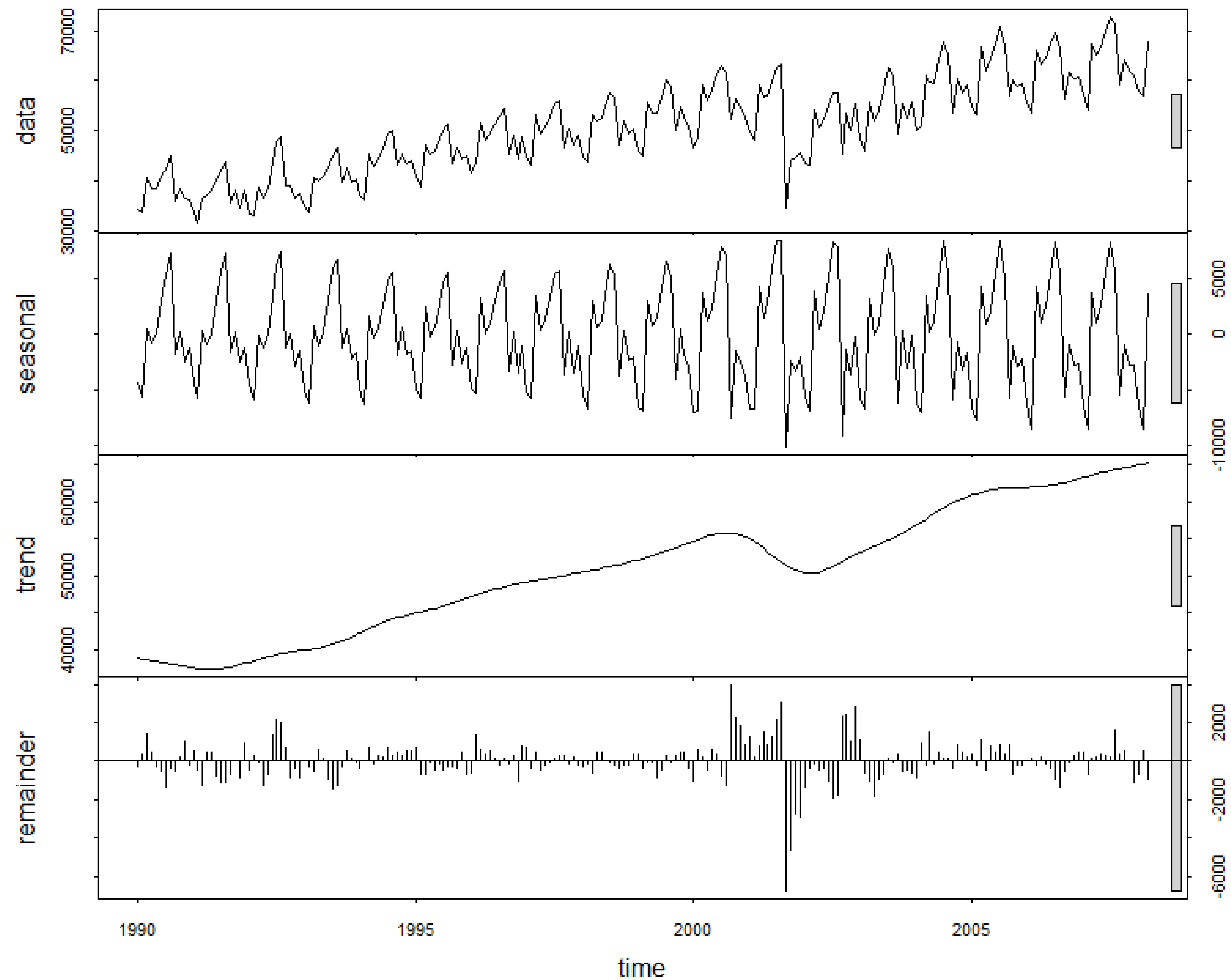
# Airline data set

- Data contains number of US airline passengers from January 1990 – March 2008
- Data is monthly (length of season is 12…repeats pattern every 12 observations)



# Time Series Decomposition

$$Y_t = S_t + T_t + E_t$$



# Time Series Decomposition

$Y_t$

$=$

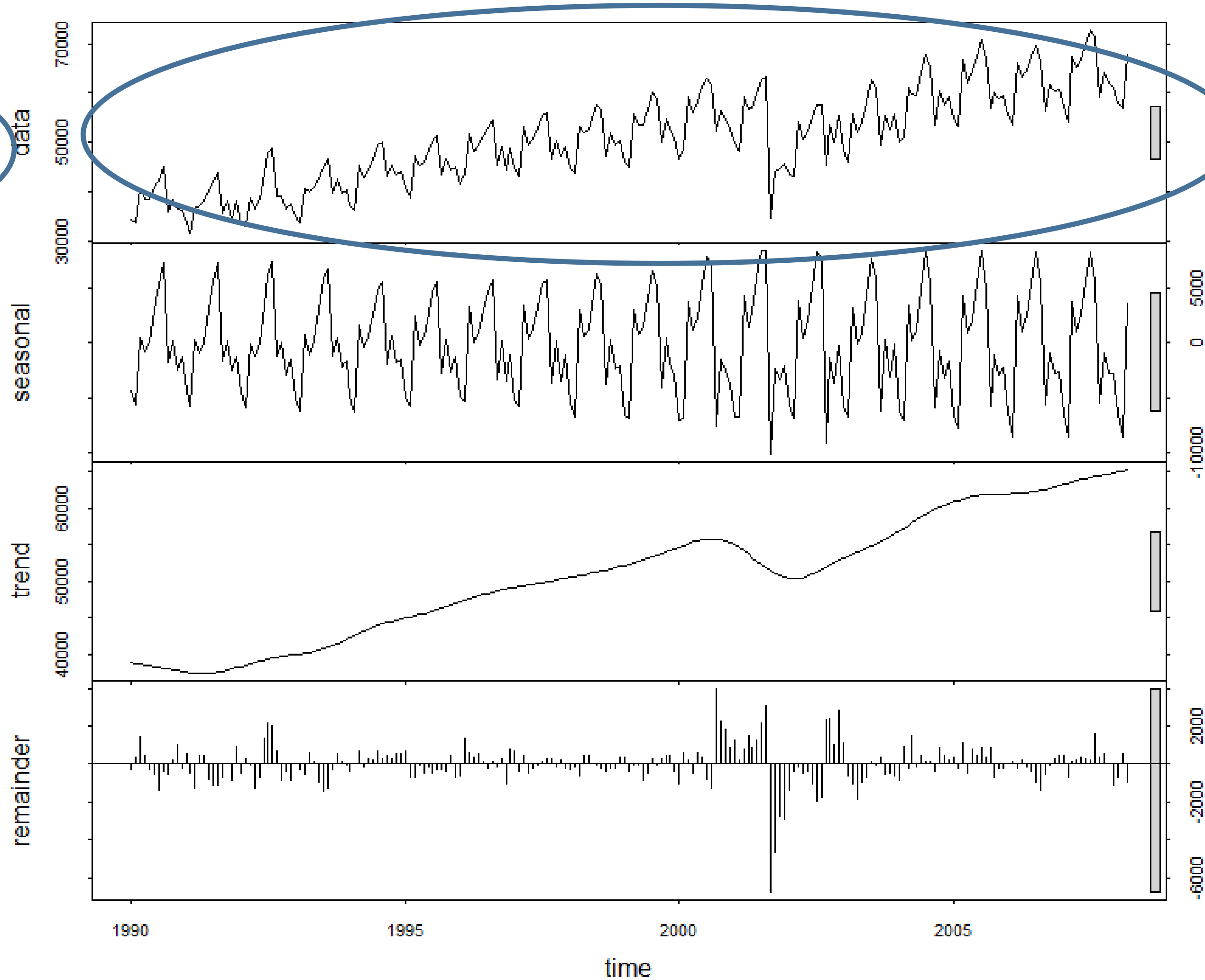
$S_t$

$+$

$T_t$

$+$

$E_t$




Original data

# Time Series Decomposition

$Y_t$

$=$

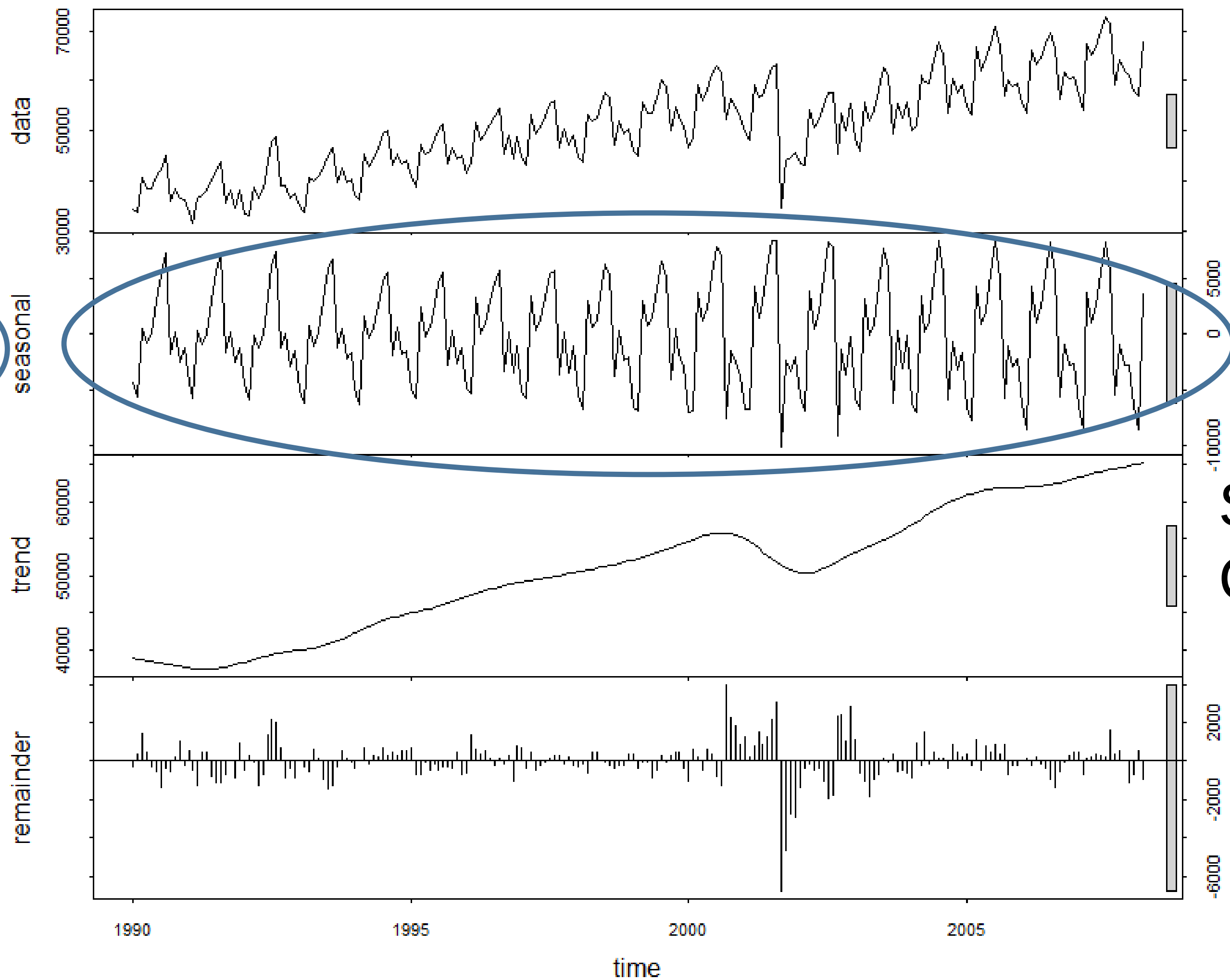
$S_t$

$+$

$T_t$

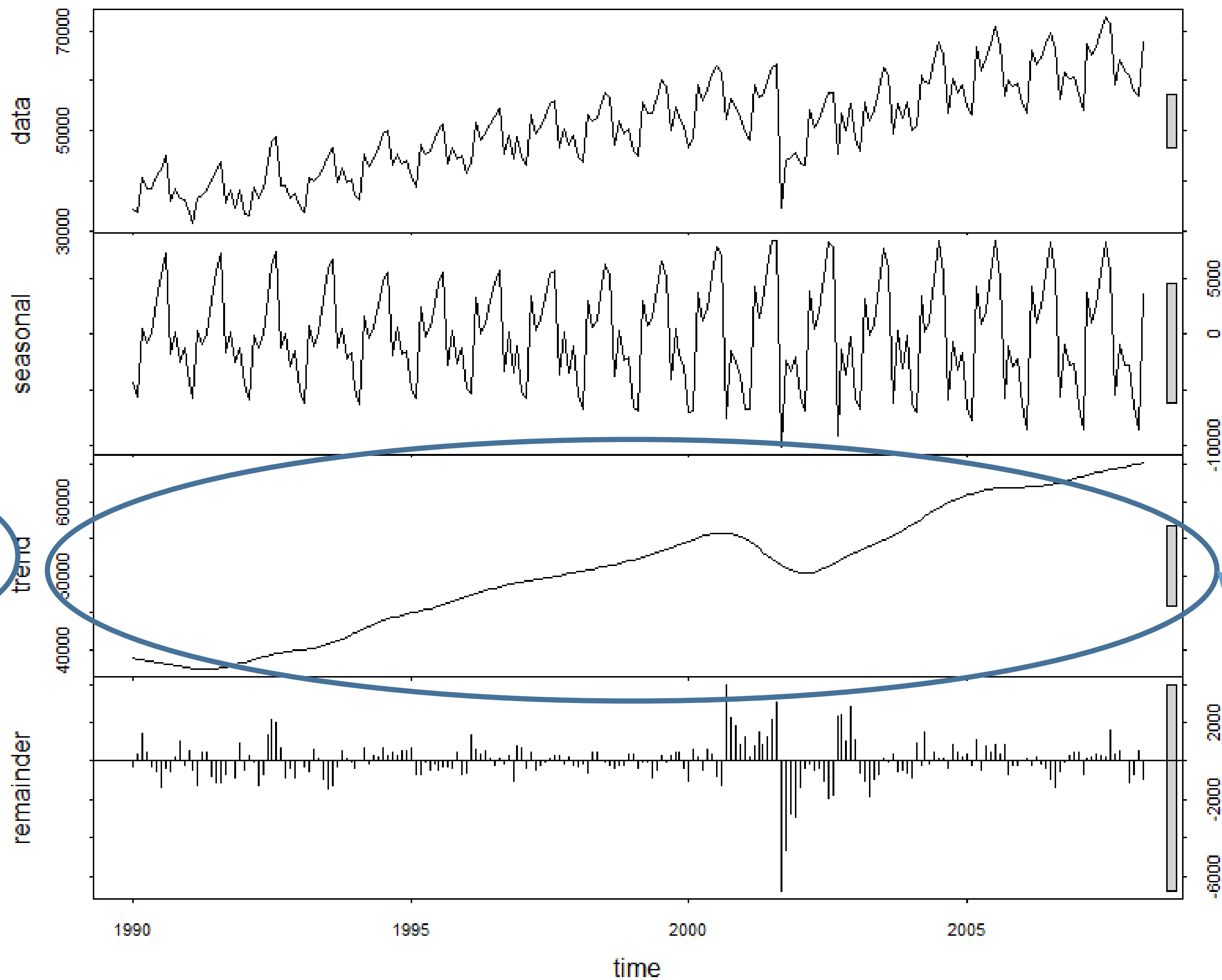
$+$

$E_t$




Seasonal Component

# Time Series Decomposition




$Y_t$

$=$

$S_t$

$+$

$T_t$

$+$

$E_t$

Trend Component

# Time Series Decomposition

$Y_t$

$=$

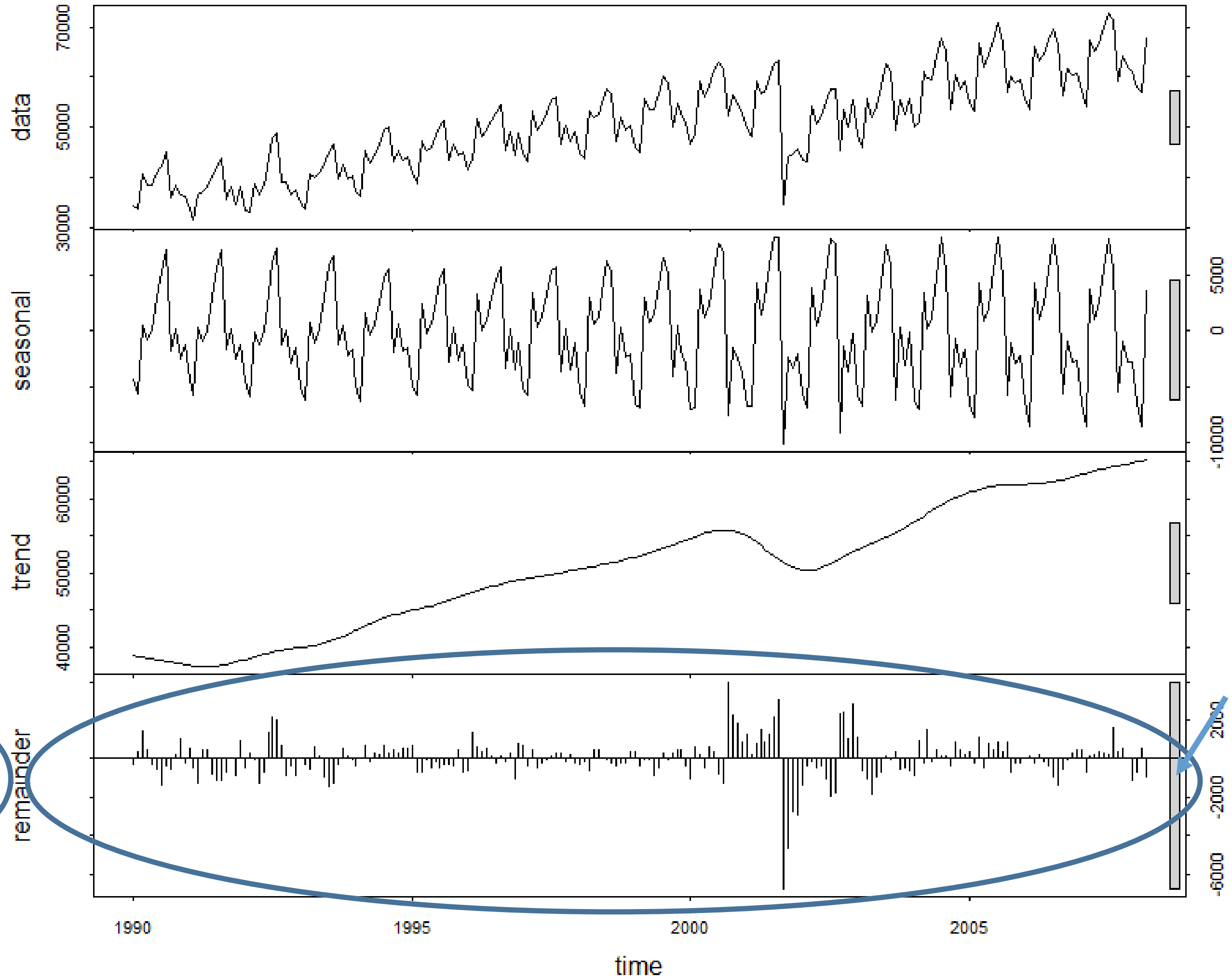
$S_t$

$+$

$T_t$

$+$

$E_t$



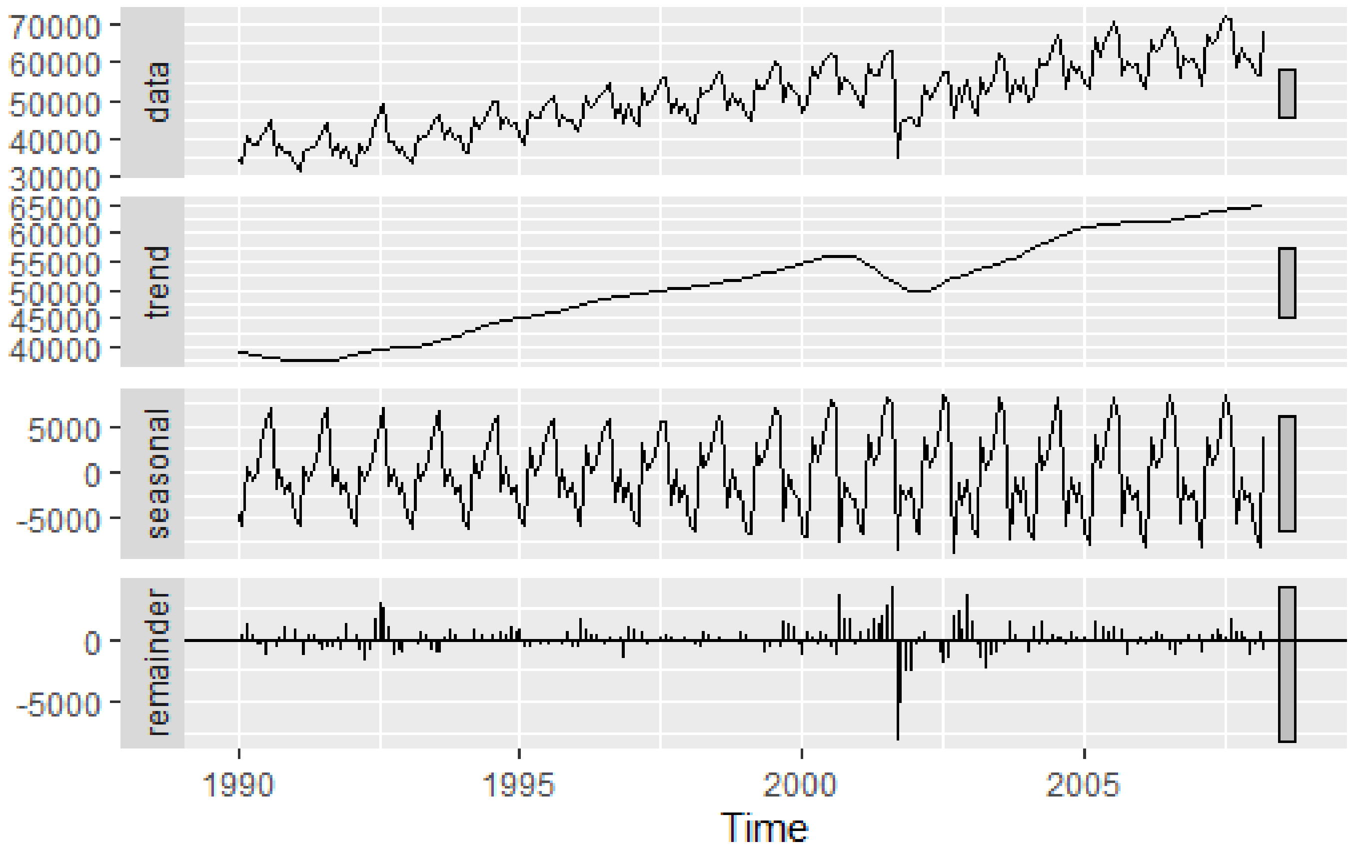

Remainder (What's left over)

# Time Series Decomposition

```
# Time Series Decomposition ...STL#
Passenger <- ts(USAirlines$Passengers, start = 1990, frequency =12)
decomp_stl <- stl(Passenger, s.window = 7)

# Plot the individual components of the time series
plot(decomp_stl)

autoplot(decomp_stl)
```

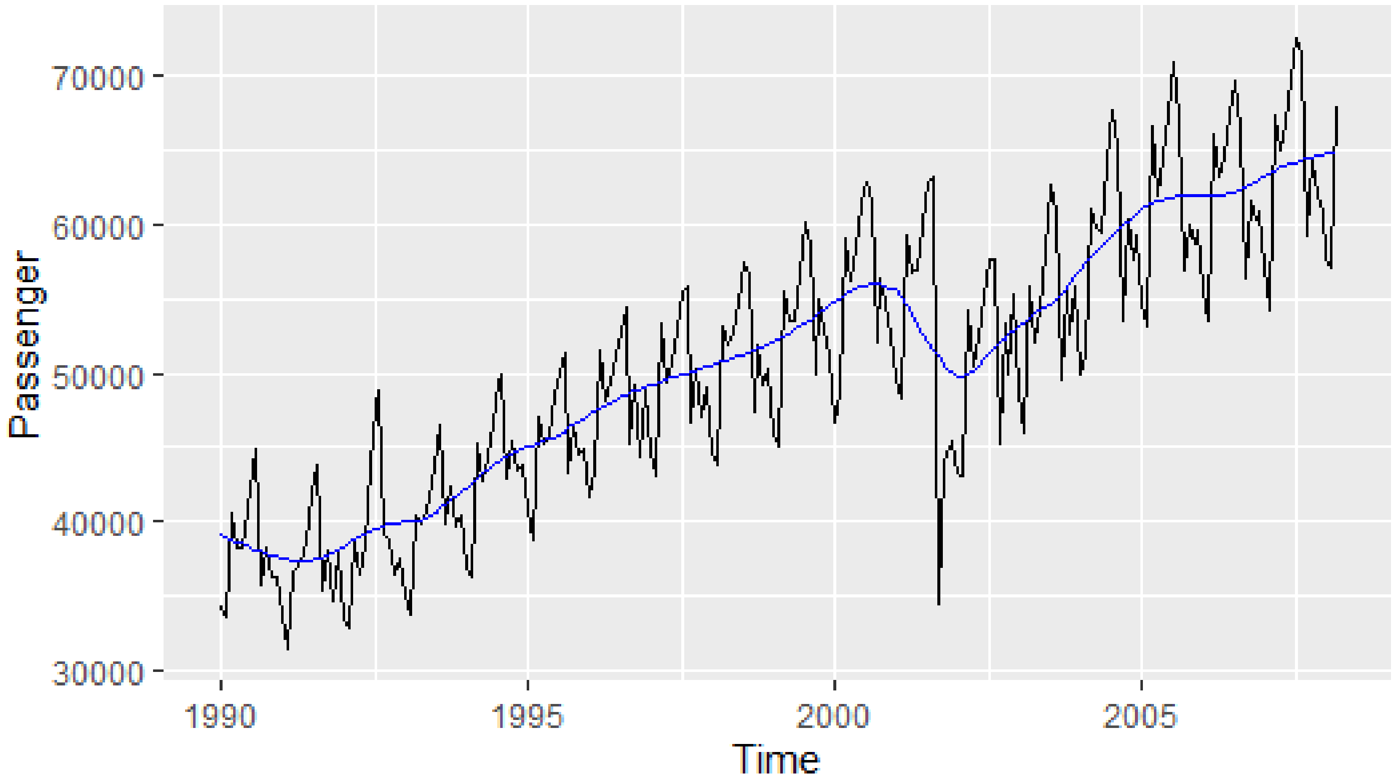


70000
60000
50000
40000
30000
data
65000
60000
55000
50000
45000
40000
trend
5000
0
-5000
seasonal
0
-5000
remainder
1990
1995
2000
2005
Time

# Time Series Decomposition

```
autoplot(Passenger)+
geom_line(aes(y=decomp_stl$time.series[,2]),
color="blue")
```

Overlay the trend component


70000
60000
50000
40000
30000
Passenger
1990
1995
2000
2005
Time

# Time Series Decomposition

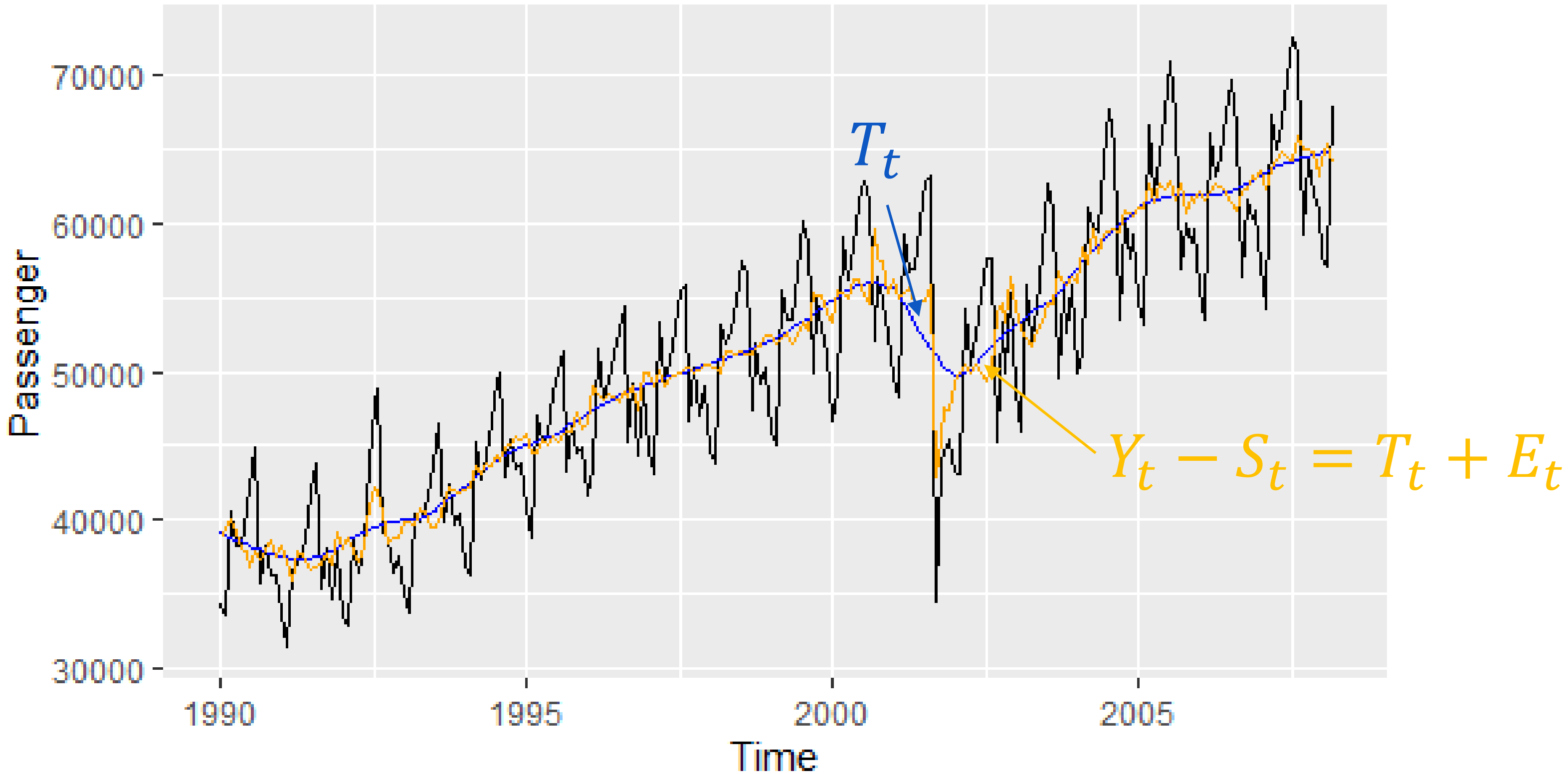
```
seas_adj=Passenger-decomp_stl$time.series[,1]

autoplot(Passenger) +
  geom_line(aes(y=decomp_stl$time.series[,2]),color="blue")
+ geom_line(aes(y=seas_adj),color="orange")
```

Overlay the trend component

Overlay seasonally adjusted

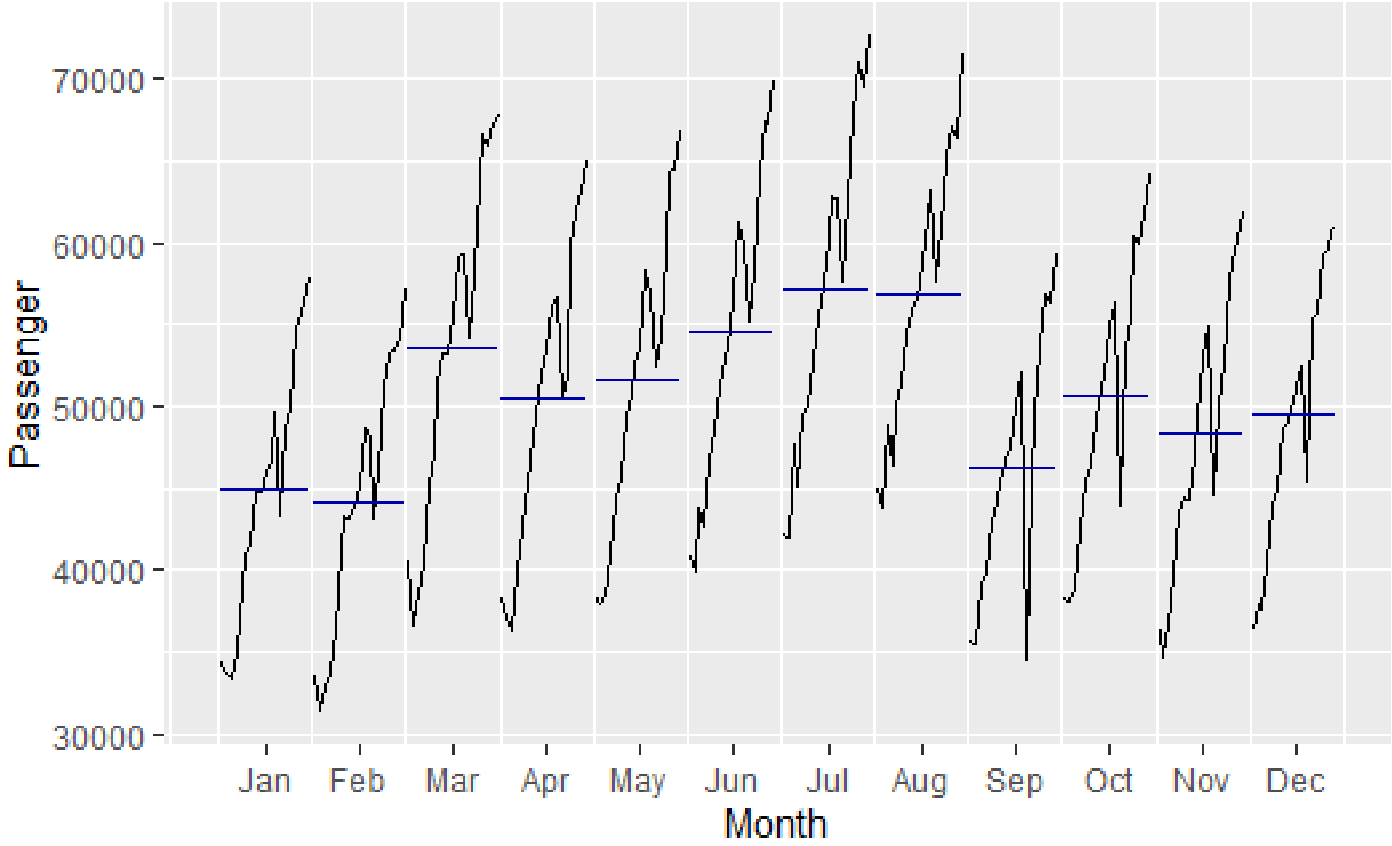
# Time Series Decomposition



# Time Series Decomposition – R

```
ggsubseriesplot(Passenger)
```

# Time Series Decomposition



# Decomposition Techniques

- There are many different ways to calculate the trend/cycle and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition

# Decomposition Techniques

- There are many different ways to calculate the trend/cycle and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition
     a. Default in SAS (Can be done in R)
     b. Trend – Uses Moving / Rolling Average Smoothing
     c. Seasonal – Average De-trended Values Across Seasons

# Decomposition Techniques

- There are many different ways to calculate the trend/cycle and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition
  2. X-13 ARIMA Decomposition (self study)
     a. Trend – Uses Moving / Rolling Average Smoothing
     b. Seasonal – Uses Moving / Rolling Average Smoothing
     c. Iteratively Repeats Above Methods and ARIMA Modeling
     d. Can handle outliers

# Decomposition Techniques

- There are many different ways to calculate the trend/cycle, and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition
  2. X-12 ARIMA Decomposition
  3. STL (Seasonal and Trend using LOESS estimation) Decomposition
     a. Default of stl Function in R (Not available in SAS)
     b. Uses **LO**cal regr**ESS**ion Techniques to Estimate Trend and Seasonality
     c. Allows Changing Effects for Trend and Season
     d. Adapted to Handle Outliers

# Comparison of Classical versus STL seasonal decomposition