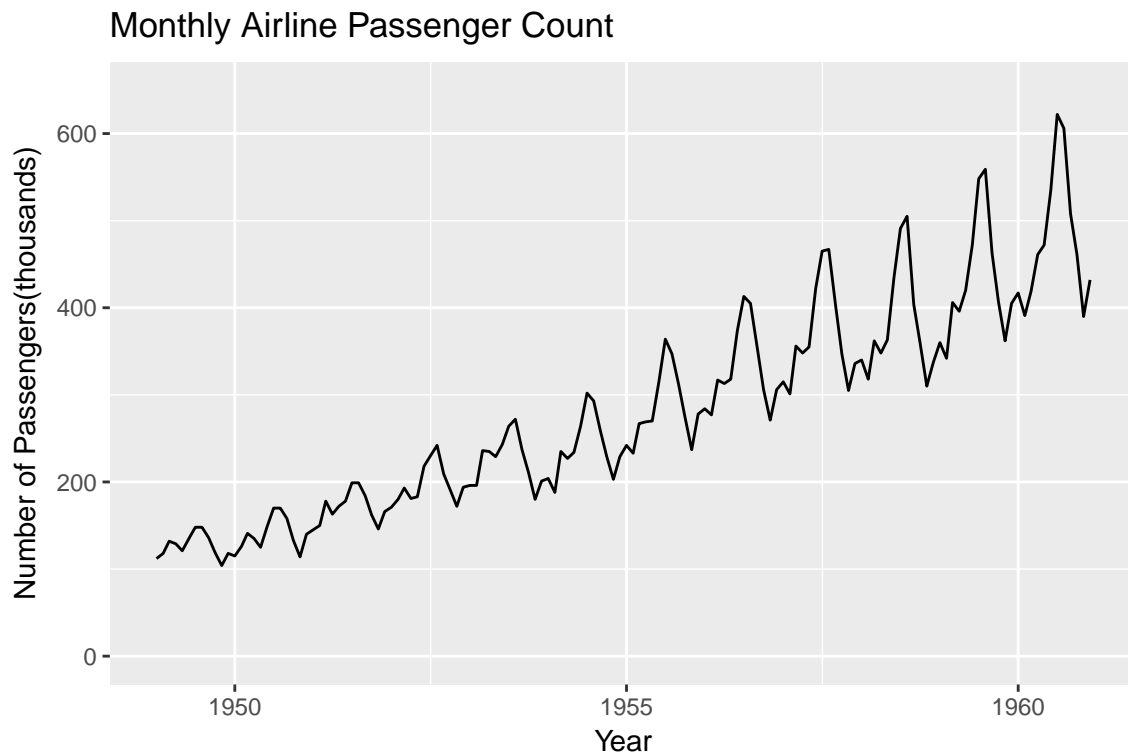


# STAT 436 / 536 - Lecture 3

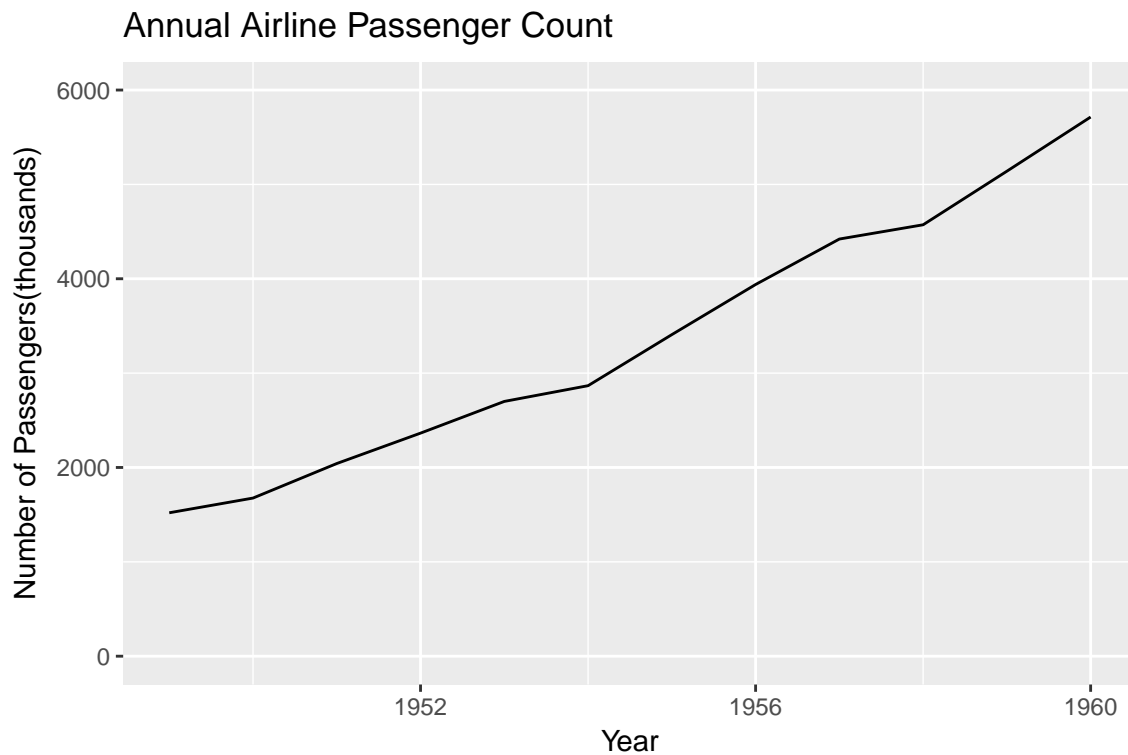
September 5, 2018

## Time Series Decomposition - Visually

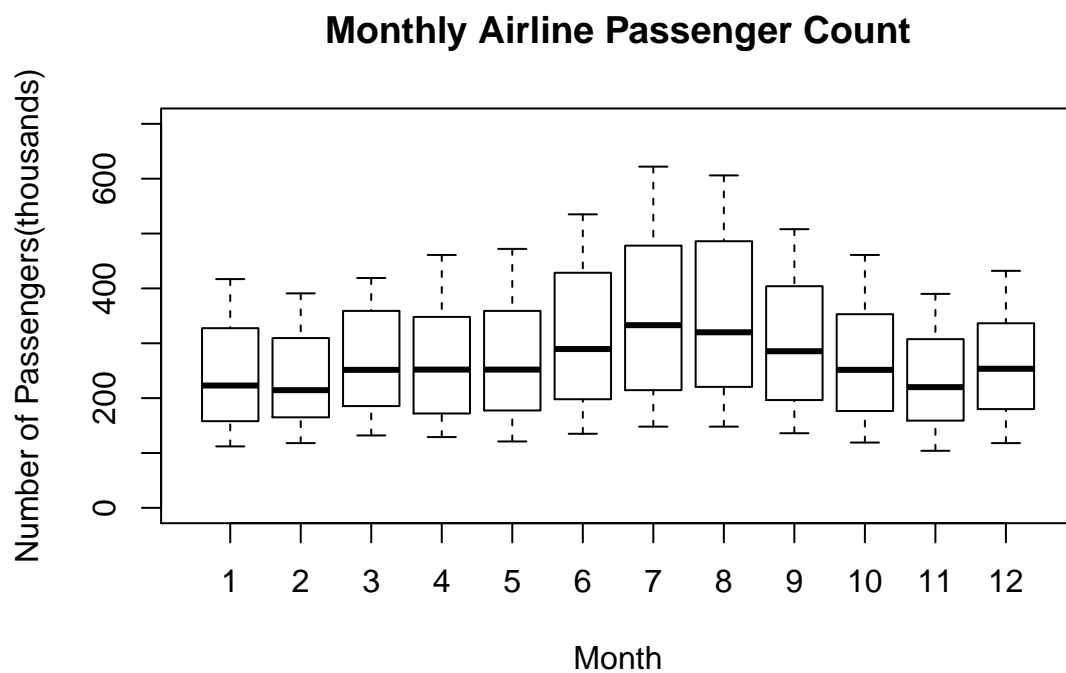
```
library(datasets)
library(ggplot2)
library(ggfortify)
data(AirPassengers)
autoplot(AirPassengers) + ylim(0,650) +
  labs(title="Monthly Airline Passenger Count", y="Number of Passengers(thousands)", x= 'Year')
```



```
autoplot(aggregate(AirPassengers)) + ylim(0,6000) +
  labs(title="Annual Airline Passenger Count", y="Number of Passengers(thousands)", x= 'Year')
```



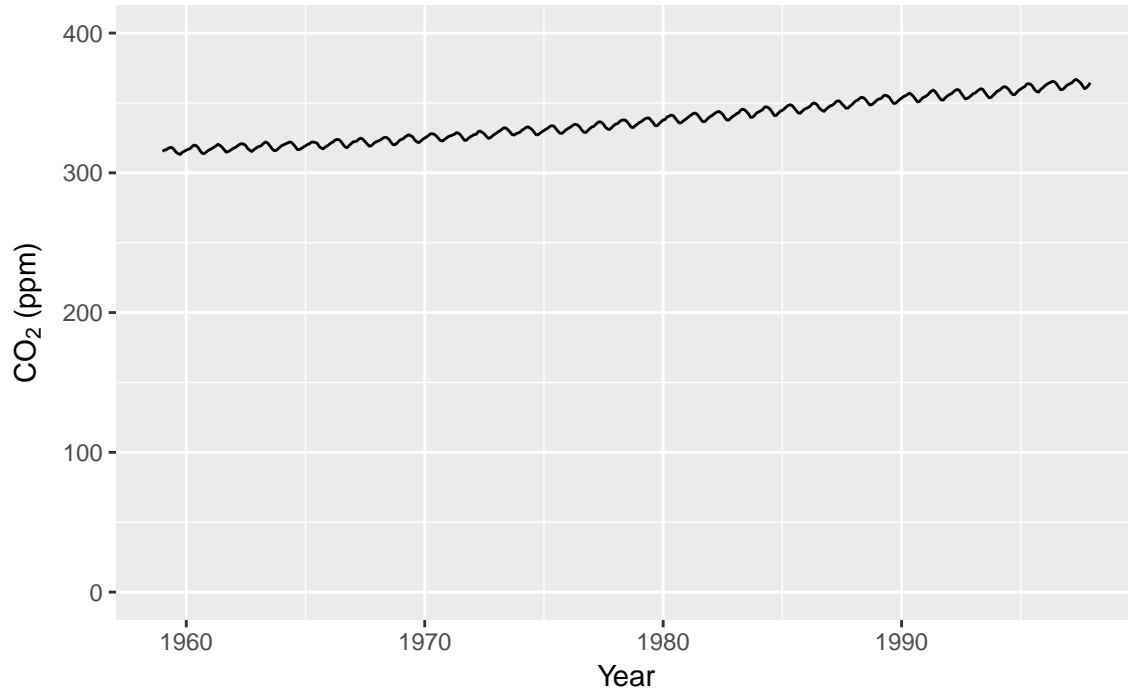
```
boxplot(AirPassengers ~ cycle(AirPassengers), ylim = c(0,700), xlab = 'Month',  
        ylab = "Number of Passengers(thousands)", main = "Monthly Airline Passenger Count")
```



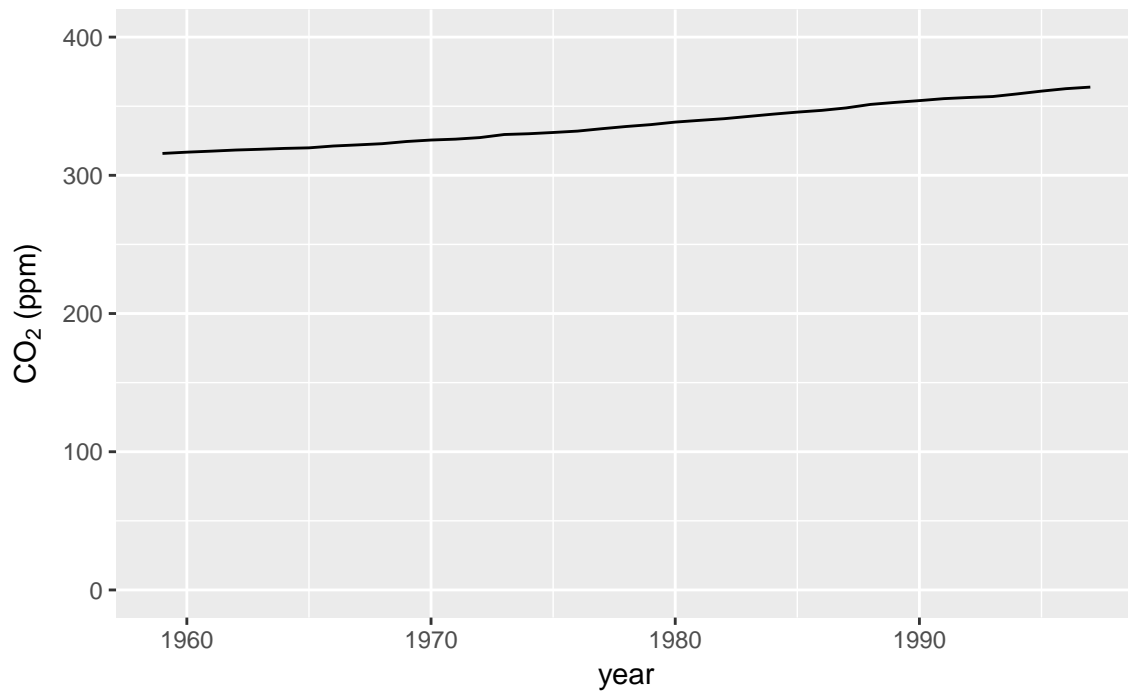
### Exercise: Mauna Loa $CO_2$ Observations

Using the Mauna Loa dataset, create three figures: 1.) the complete time series, 2.) annual decomposition, and 3.) monthly series. Think carefully about the appropriate scale for the y-axis of these plots.

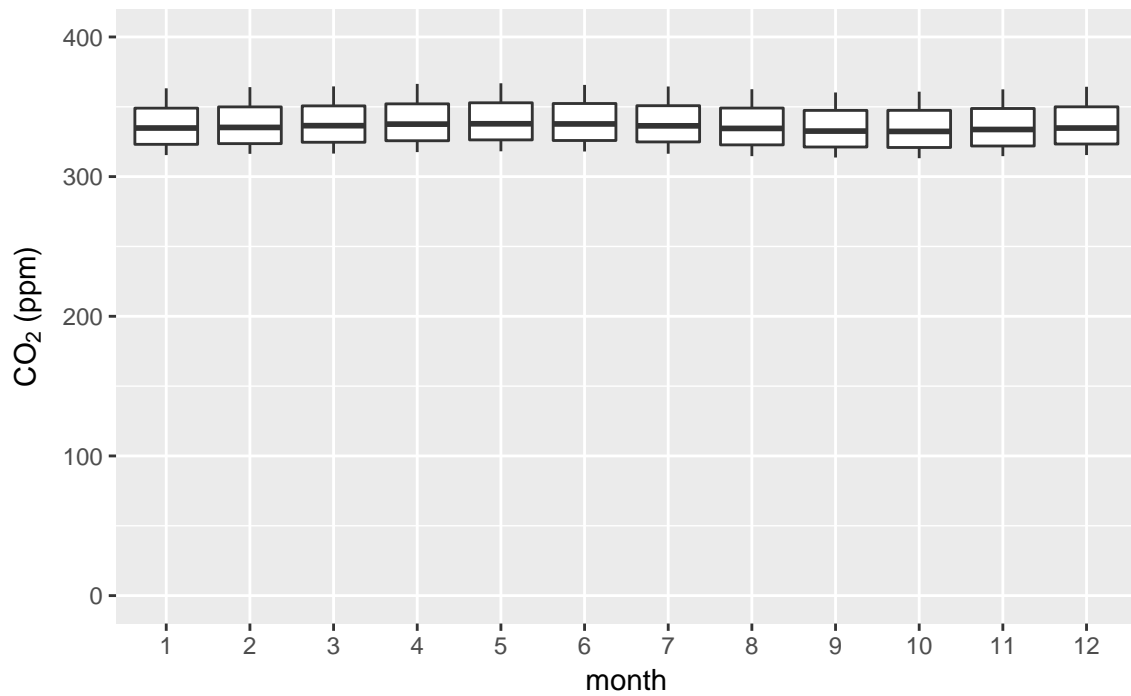
#### Monthly $CO_2$ concentration at Maua Loa



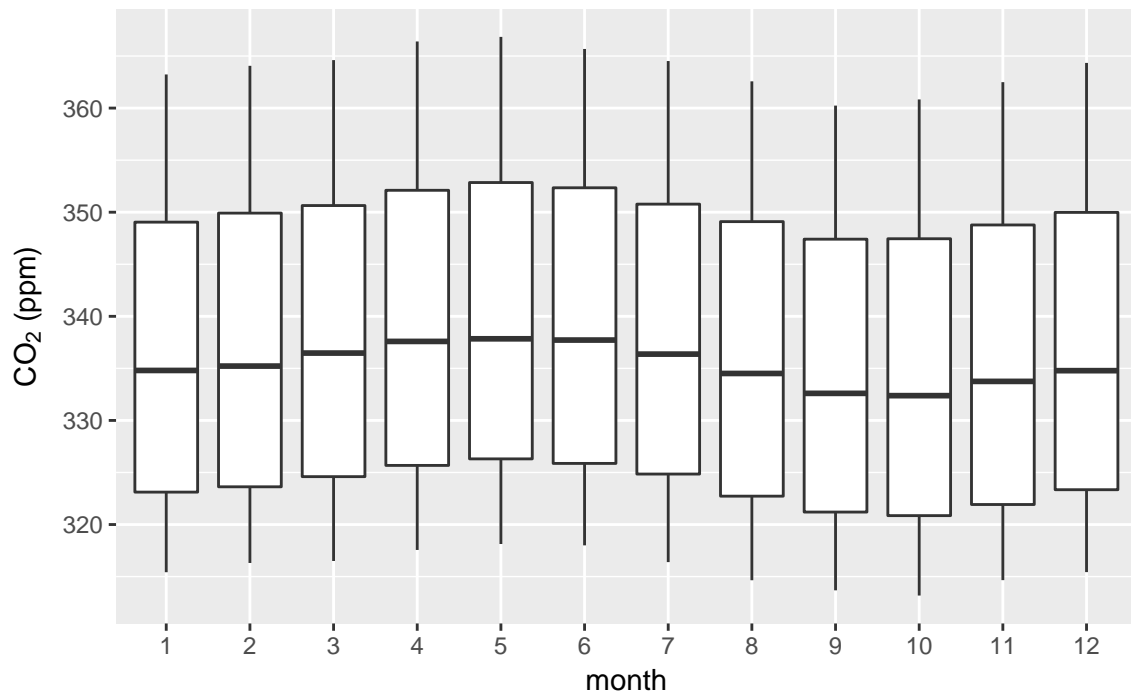
#### Average Annual $CO_2$ Concentration at Maua Loa



Monthly Average CO<sub>2</sub> Concentration at Maua Loa



Monthly Average CO<sub>2</sub> Concentration at Maua Loa



## Time Series Decomposition - Modeling and Notation

### Time Series Mathematical Notation

- Let  $\{x_t\} = \{x_1, \dots, x_T\}$  be a time series of length  $T$ .
- A time series is a sequence of random variables and the observed time series is one realization of this process. Unlike some other settings, the same notation is typically used for the random variables and observed values.
- The hat notation will often be used for prediction,  $\hat{x}_{t+k|1:t}$ , meaning the observation at time  $t+k$ , given observations up to time  $t$ .
- For those of you familiar with Bayesian notation, we will use  $p(x_{t+k}|x_{1:t})$  to represent the distribution at time  $t+k$  given observations from time 1 to time  $t$ .

### Time Series Models

- We have seen how to visually decompose a time series into an annual trend and a seasonal pattern. So this additive model could be written as:

$$x_t = m_t + s_t$$

where  $m_t$  is the trend and  $s_t$  is the seasonal pattern.

- What is missing from the previous model? *the error term:  $z_t$ .*

$$x_t = m_t + s_t + z_t$$

- **Question:** how should we think about estimating a trend?
  - One option is a moving average centered at  $x_t$ , but the question is how many observations are necessary? Make sure that the moving average is sufficiently long to smooth out the seasonal effects.

- With monthly data this can be written as:

$$\hat{m}_t = \frac{1/2x_{t-6} + x_{t-5} + \dots x_t + \dots x_{t+5} + 1/2x_{t+6}}{12}$$

- **Question:** Now how about the seasonal effect?
- The seasonal effect for time  $t$  can be estimated by subtracting the trend line.

$$s_t = x_t - m_t$$

- Then the overall seasonal effect  $\bar{s}$  is the average of the seasonal components in that seasonal cycle.

- This procedure results in a *smoothed* estimate of the time series.

- A “seasonally adjusted” time series can also be produced that removes the seasonal patterns.

$$x_t - \bar{s}$$

- With an `ts` object in R, the `decompose()` function will automatically do this procedure.

```
decompose(AirPassengers)
```