

# BIG DATA ANALYTICS

WEEK-12 | Time-series analysis

Yonsei University  
Jungwon Seo

# 시계열 분석

Samsung Electronics Co., Ltd. (005930.KS) ☆

KSE - KSE Delayed Price. Currency in KRW

50,000.00 -100.00 (-0.20%)

At close: April 29 3:30PM KST

⊕ Indicators ⊕ Comparison ⚡ Events | 📅 Mar 01, 2010 - Apr 29, 2020 1D 5D 1M 3M 6M YTD 1Y 2Y 5Y Max 📄 Interval 1M ⚡ Line 📄 Draw



# 시계열 분석

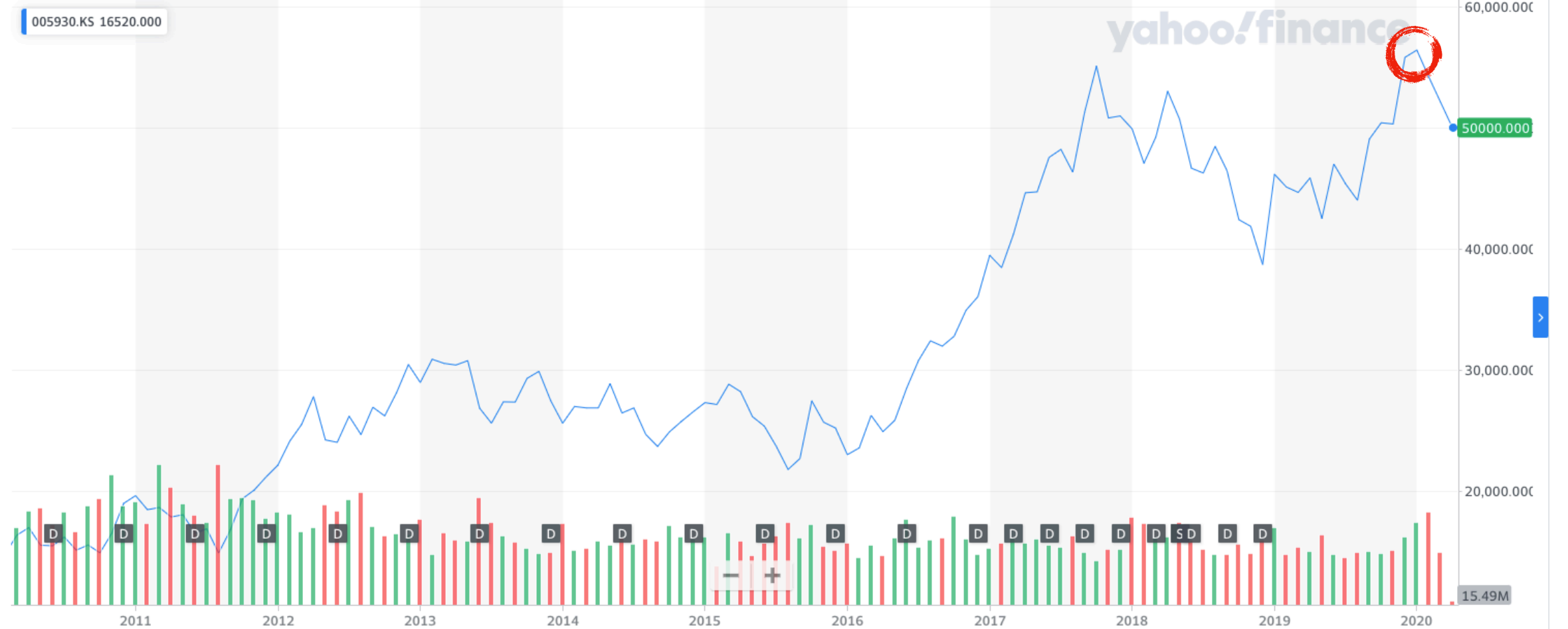
Samsung Electronics Co., Ltd. (005930.KS) ☆

KSE - KSE Delayed Price. Currency in KRW

50,000.00 -100.00 (-0.20%)

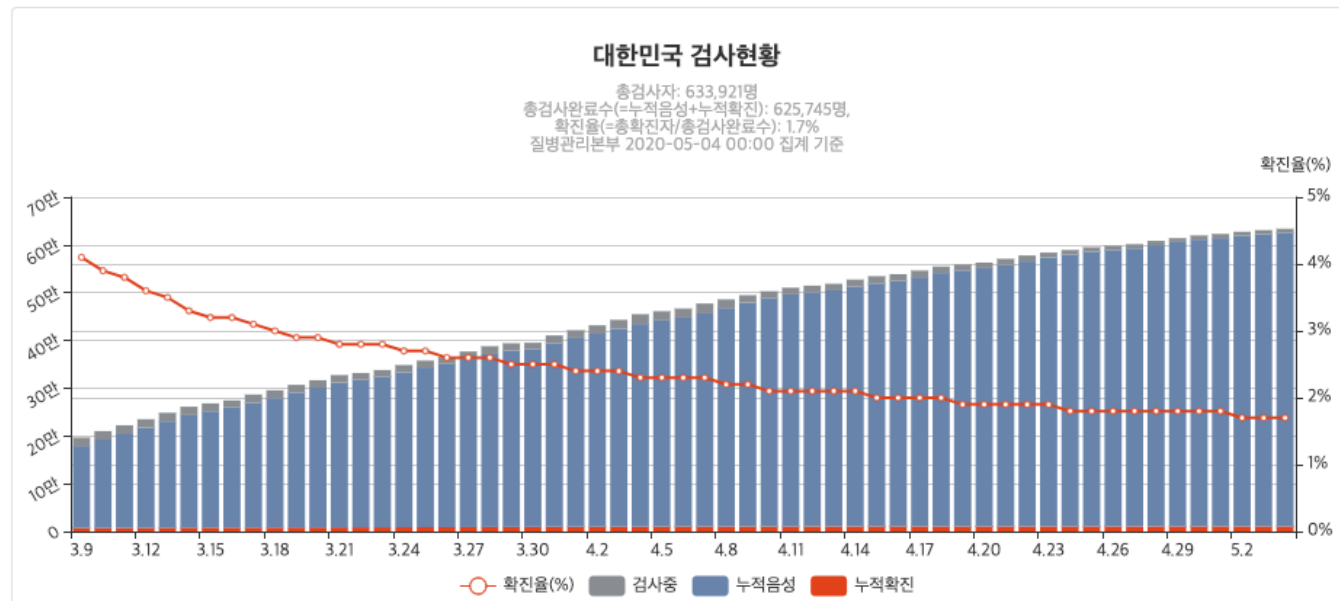
At close: April 29 3:30PM KST

⊕ Indicators ⊕ Comparison ⚡ Events | 📅 Mar 01, 2010 - Apr 29, 2020 1D 5D 1M 3M 6M YTD 1Y 2Y 5Y Max | 📄 Interval 1M ⚡ Line 📄 Draw ⚙️ Settings

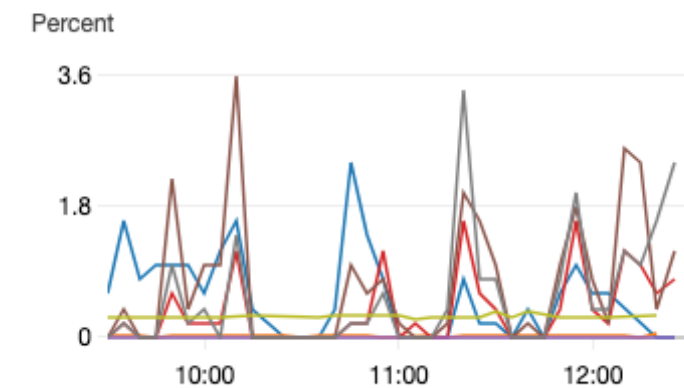


# 시계열 데이터

- 관측치가 시간적 순서를 포함
  - 예) 주가, 월별 매출액, 월별 실업률, 연도별 인구수, 환율, 공장 로그, 웹사이트 로그 등



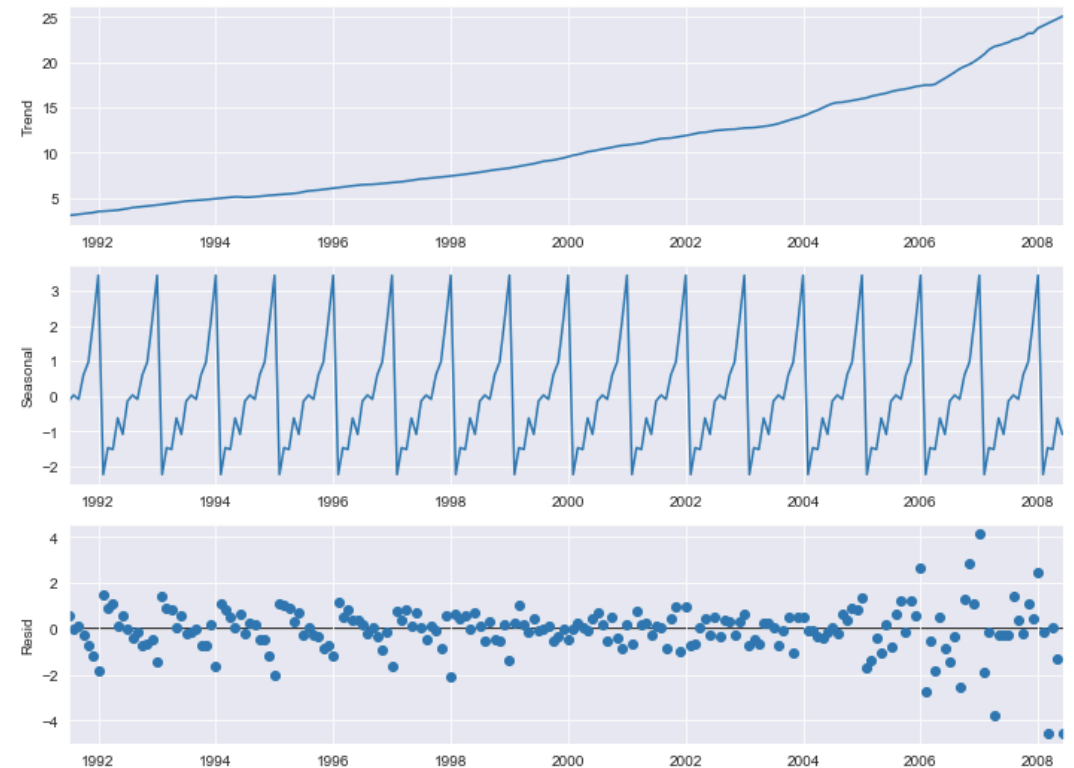
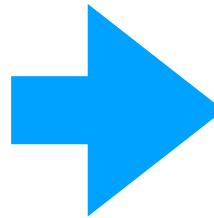
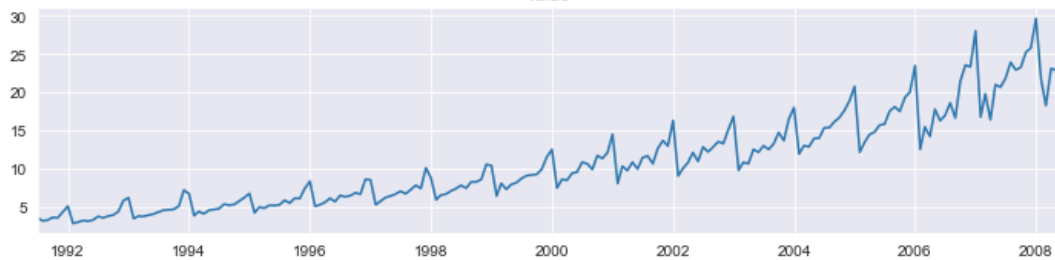
CPU Utilization Average



# 시계열 데이터

- 시계열 데이터의 성분

- 계절성(Seasonal): 고정된 주기로 특정 패턴이 반복되는 성질
- 추세(Trend): 전반적인 경향
- 랜덤(Random): 추세 및 계절성으로 설명되지 않는 데이터

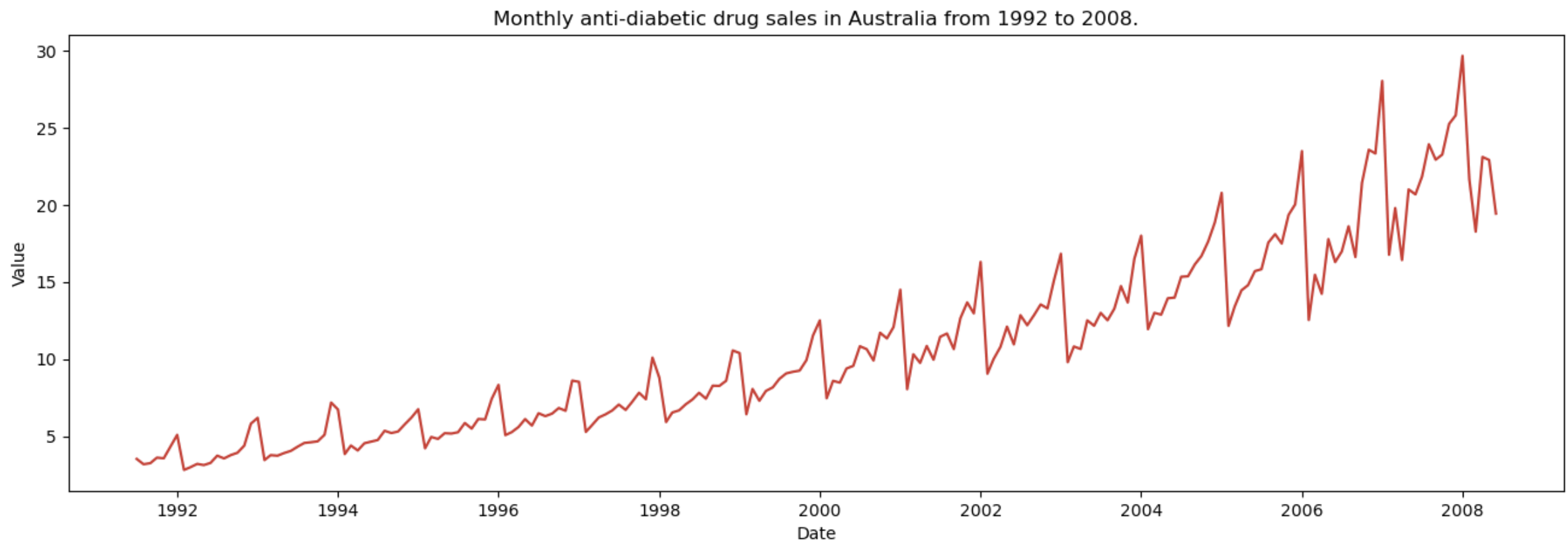


# 시계열 분석의 목적

- 미래 값의 값을 예측
  - 일주일간 주가 예측
  - 2050년 대한민국 인구
  - 다음달 매출액 예측
- 시계열 데이터 특성 파악
  - 경향(trend)
  - 주기(cycle)
  - 계절성(seasonality)
  - 불규칙성( irregular)

# 시계열 데이터 시각화

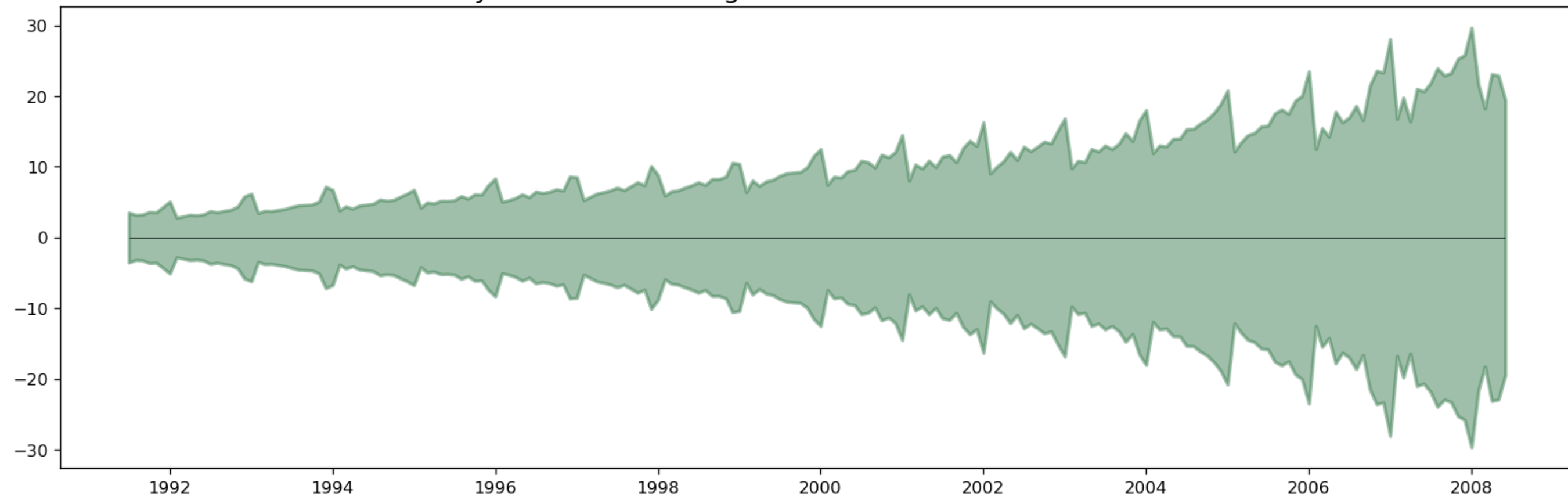
- 기본적으로 line chart로 표현가능



# 시계열 데이터 시각화

- 만약 모든데이터가 양수이고,  $y$ 값의 증가를 강조하고 싶다면,
- Tw-side view로 표현

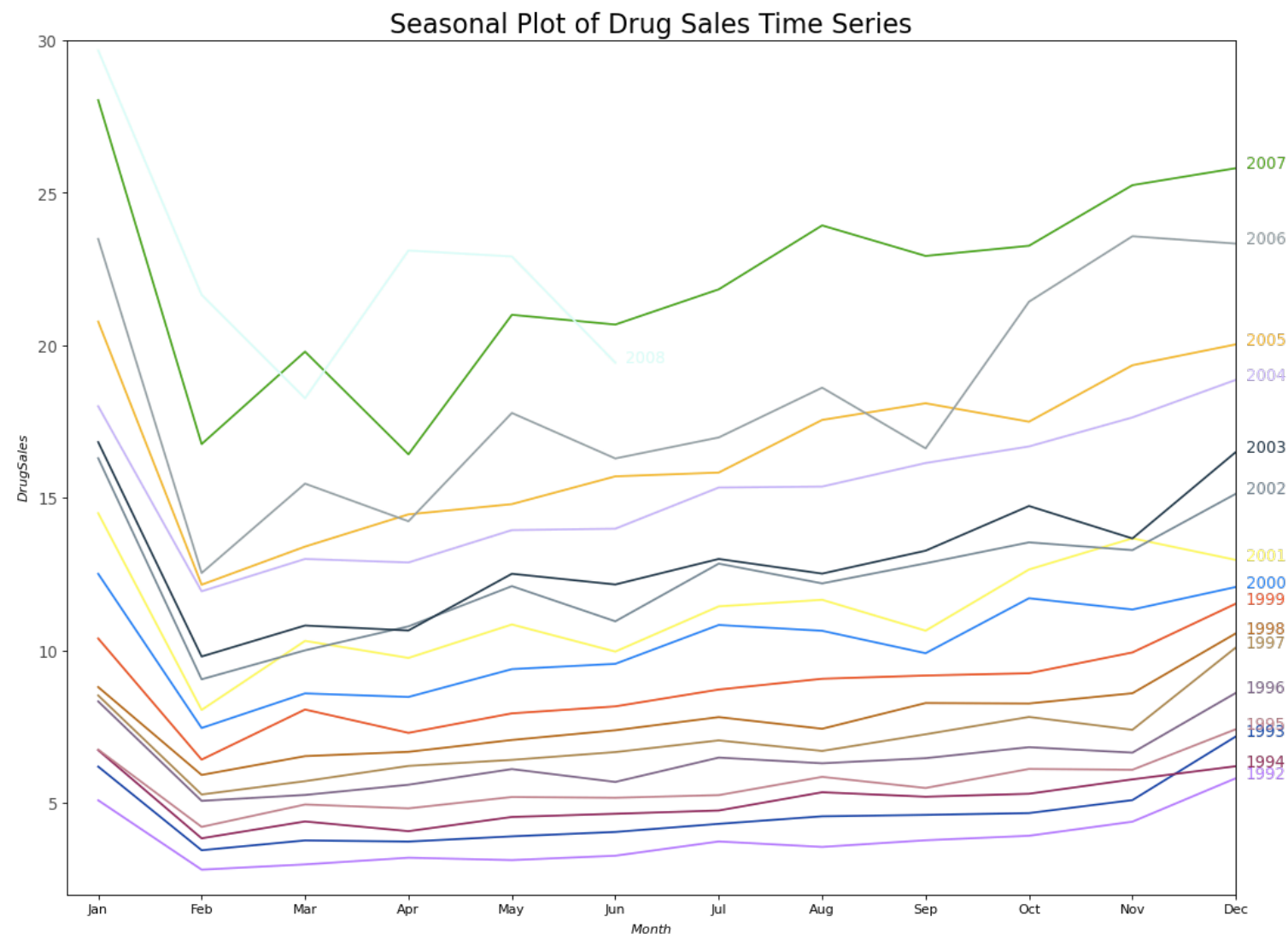
Monthly anti-diabetic drug sales in Australia from 1992 to 2008.





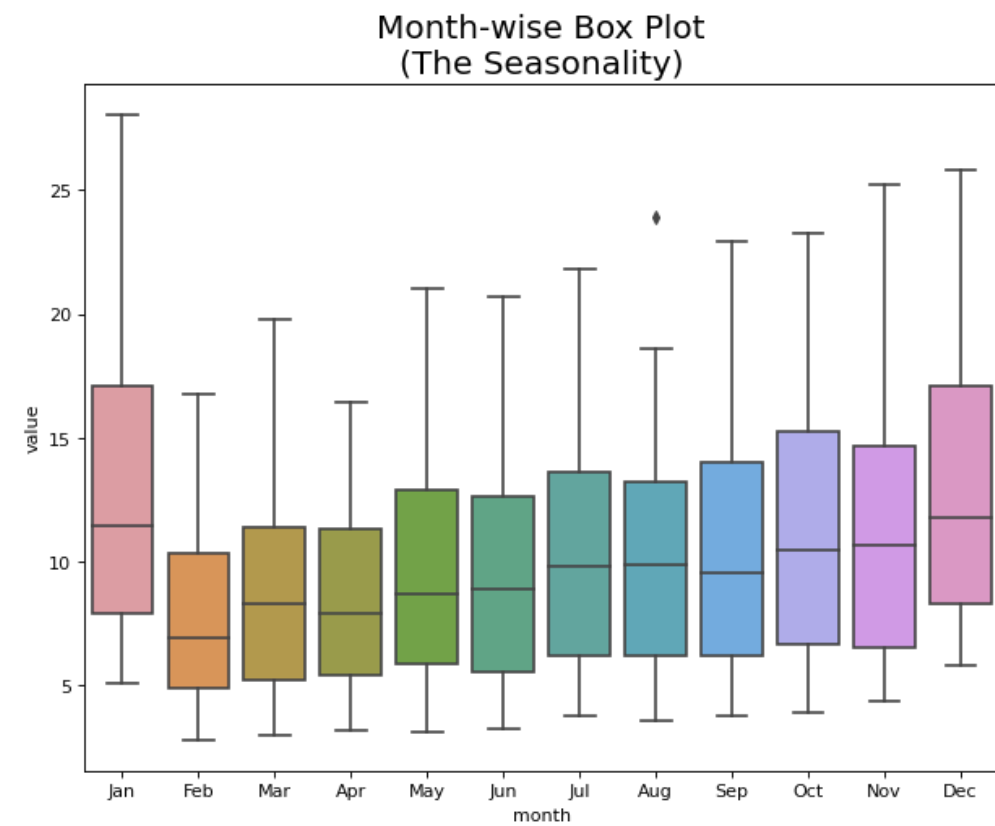
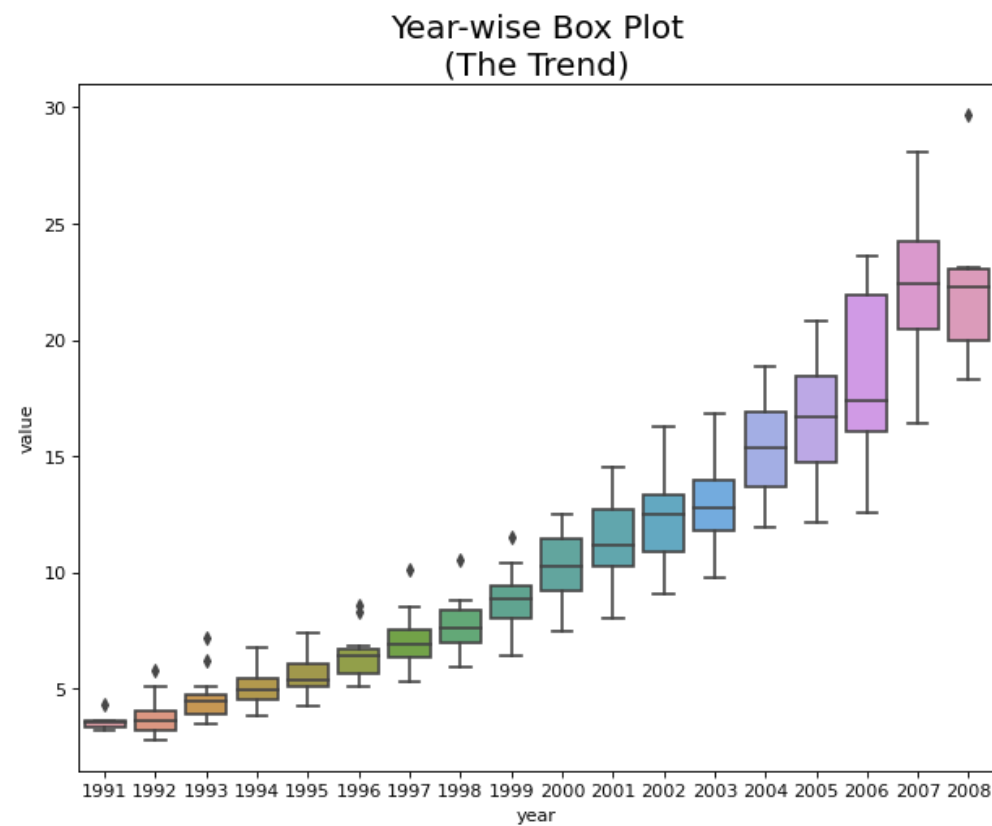
# 시계열 데이터 시각화

- 만약 주기성(매주,매달)이 있고 이를 연도별로 비교하여 표현하고 싶다면,



# 시계열 데이터 시각화

- 통계적인 데이터를 (중간값, 이상치) 포함하여 표현하고 싶다면,

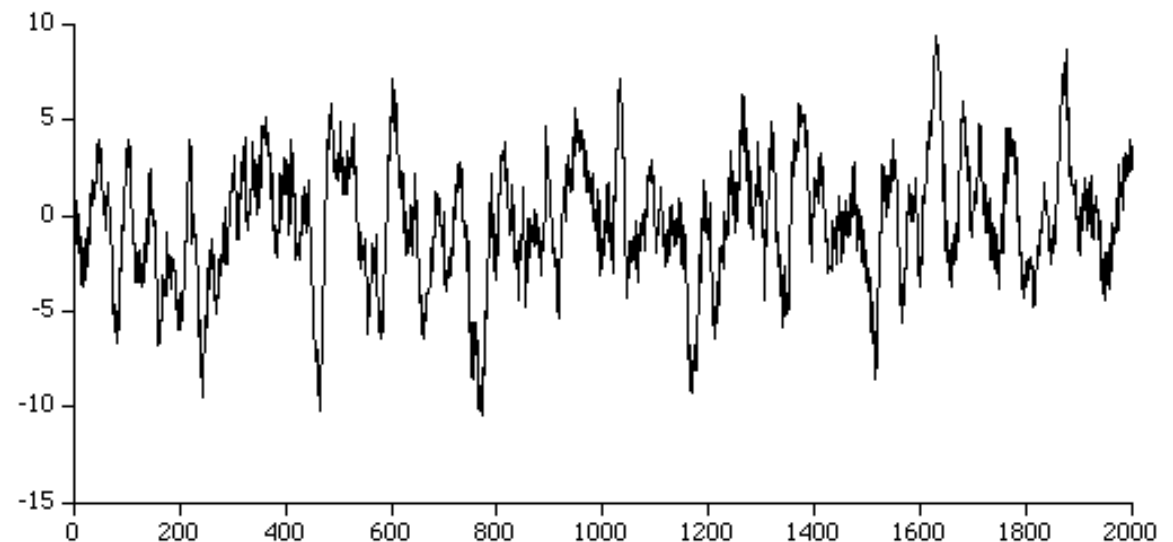


# 일변량 시계열 데이터 분석

- 변수 하나에만 의존한 분석
- 일변량(univariate) 데이터의 예측치는 자신들의 과거의 값에만 의존
  - $S_t = \alpha + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \cdots + \beta_p S_{t-p}$
- ARMA, 지수평활법, 시계열 분해법

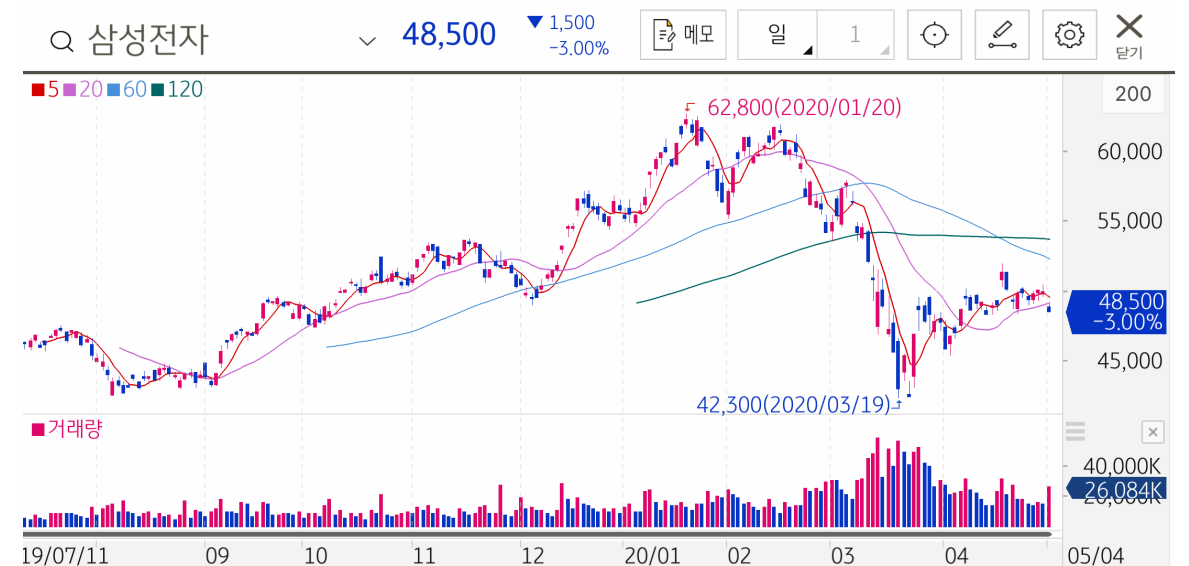
# 자기회귀 (AR) 모형

- 자기회귀(AR : Auto Regressive)
  - 주식에서 흔히, U자 반등, V자 반등과 같이 이전 값들이 현재 값에 영향을 미치는 자기 상관 특성을 갖고 있음
- 직전 데이터 (t-1)이 현재 데이터 (t)의 값에 영향을 미침
  - AR(1):  $Y_t = \phi Y_{t-1} + \epsilon_t$
  - $\epsilon_t$ 는 백색잡음 (white noise: 평균 0, 표준편차 1인 정규분포에서 도출된 랜덤 값)
- 더 이전 시점을 포함 할 시
  - AR(P):  $Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t$



# 이동평균 (MA) 모형

- 시간이 지날 수록 어떠한 변수의 평균값이 지속적으로 감소하거나 증가 할 수 있음
- 이동평균의 종류
  - 단순이동평균
  - 가중이동평균
  - 누적이동평균
  - 지수이동평균

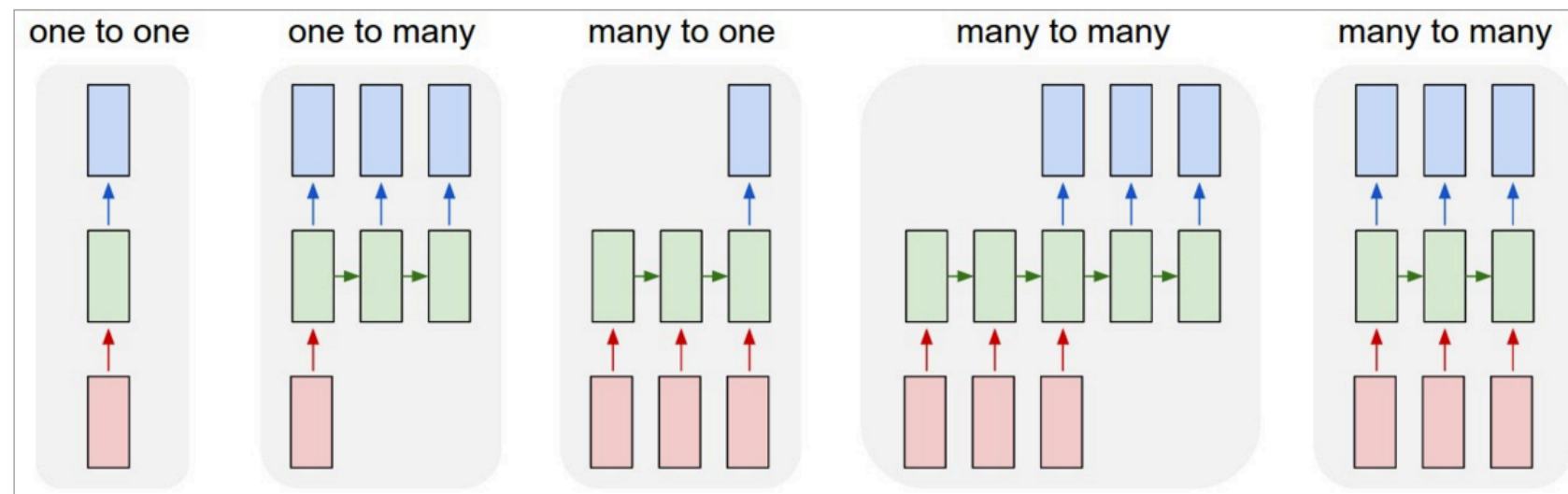


# 자기회귀 이동평균 (ARMA)

- 선형회귀 (Linear Regression)
  - $Y_i = \beta X_i + \epsilon_i$
- 자기회귀 (AR : Auto Regressive)
  - $Y_t = \phi Y_{t-1} + \epsilon_t$
- 이동평균 (MA : Moving Average)
  - $\epsilon_t = W_t + \theta W_{t-1}$
- 자기회귀 이동평균
  - $X_t = \phi X_{t-1} + W_t + \theta W_{t-1}$

# 인공신경망 기반 시계열 분석

- 데이터 특성을 파악함에 있어서는, 전통적인 방식이 작용을 하지만, 예측에 있어서는 명확한 한계가 존재
- 많은 테스트들이 딥러닝 기반의 방법론을 쓰는 이유처럼, 만약 패턴을 인공신경망이 알아서 결정하게 한다면?
- Recurrent Neural Networks



**E.O.D**