

2-기본 차트의 사용

2. Line Plot 사용하기

안수빈

서울대학교 컴퓨터공학부

Contents

1. 기본 Line plot

- 1.1 Line Plot이란?
- 1.2 Line의 요소
- 1.3 Line plot을 위한 전처리

2. 정확한 Line plot

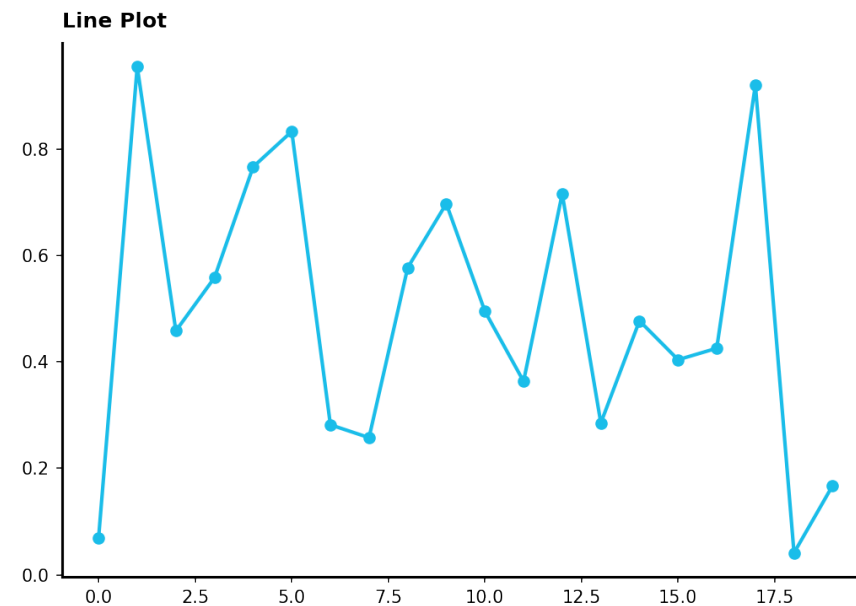
- 2.1 추세에 집중
- 2.2 간격
- 2.3 보간
- 2.4 이중 축 사용
- 2.5 ETC

1. 기본 Line Plot

Line plot의 여러가지 요소와 테크닉을 알아보자

1.1 Line plot이란?

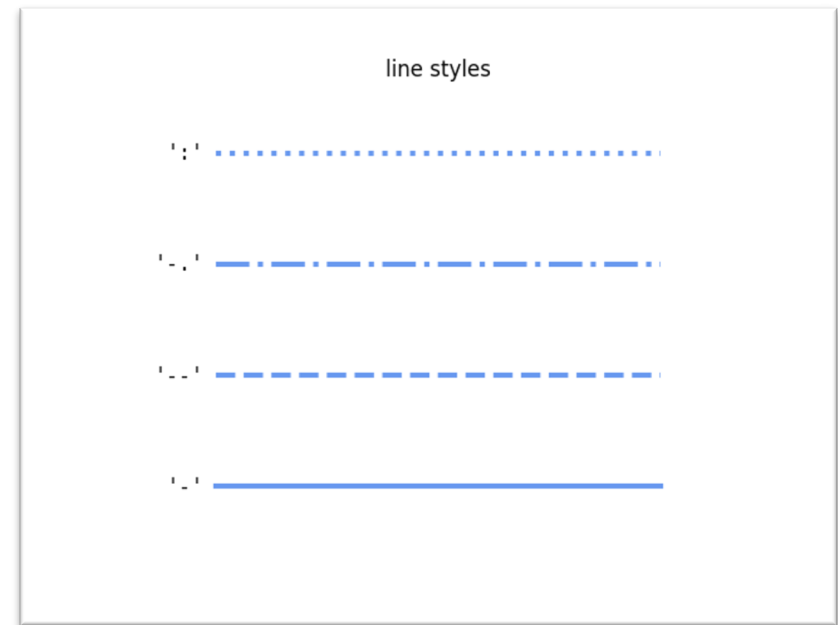
- **Line Plot**은 연속적으로 변화하는 값을 순서대로 점으로 나타내고, 이를 선으로 연결한 그래프
- 꺾은선 그래프, 선 그래프, line chart, line graph 등의 이름으로 사용됨
- 시간/순서에 대한 변화에 적합하여 추세를 살피기 위해 사용
 - 시계열 분석에 특화!!
- 신기하게 .line이 아니라 **.plot()**



1.2 Line plot의 요소

1. 기본 Line Plot

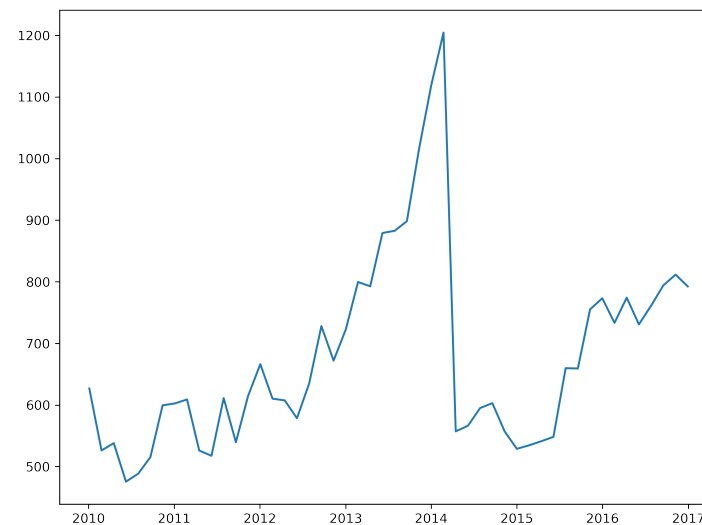
- 5개 이하의 선을 사용하는 것을 추천
 - 더 많은 선은 중첩으로 인한 가독성 하락
- 그렇다면 이를 구별하는 요소는 어떤 것이 있을까?
 1. 색상 (color)
 2. 마커 (marker, markersize)
 3. 선의 종류 (linestyle, linewidth)



1.3 Line plot을 위한 전처리

1. 기본 Line Plot

- 시시각각 변동하는 데이터는 Noise로 인해 패턴 및 추세 파악이 어려움
- Noise의 인지적인 방해를 줄이기 위해 smoothing을 사용



2.

정확한 Line Plot

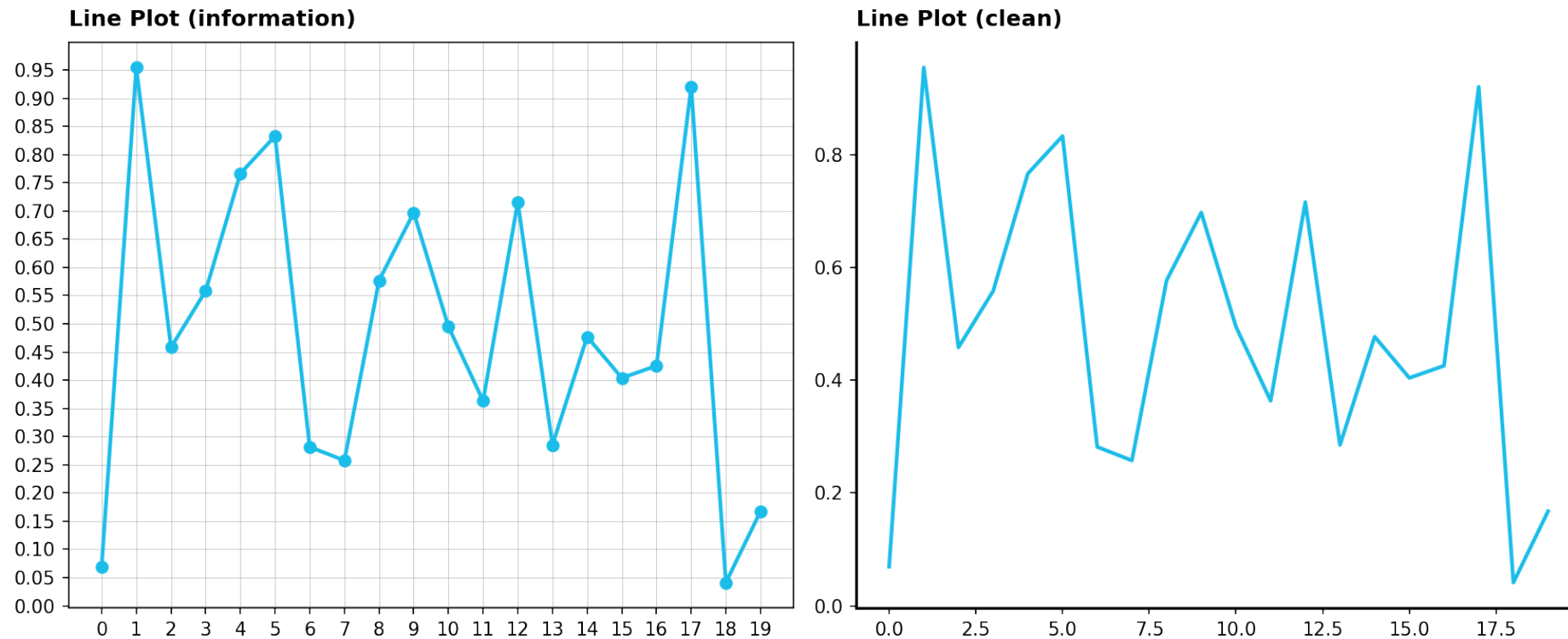
Line plot의 오용을 막아봅시다.

2.1 추세에 집중

- Bar plot과 다르게 꼭 축을 0에 초점을 둘 필요는 없음
 - 추세를 보기 위한 목적이므로
- 너무 구체적인 line plot보다는 **생략된 line plot**이 더 나을 수 있다.
 - Grid, Annotate 등 모두 제거
 - 디테일한 정보는 표로 제공하는 것을 추천
- 생략되지 않는 선에서 범위를 조정하여 변화율 관찰 (**.set_ylim()**)

2.1 추세에 집중

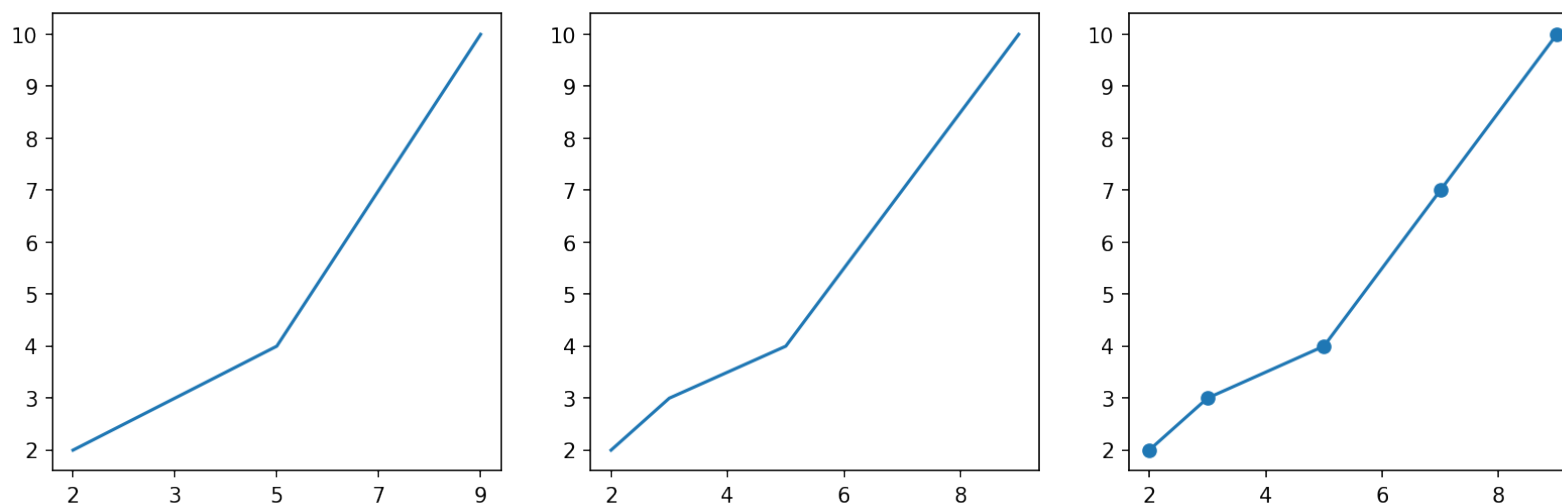
2. 정확한 Line Plot



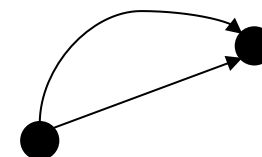
2.2 간격

2. 정확한 Line Plot

- 규칙적인 간격이 아니라면 오해를 줄 수 있다!
 - 그래프 상에서 규칙적일 때 : 기울기 정보의 오해
 - 그래프 상에서 간격이 다를 때 : 없는 데이터에 대해 있다고 오해
- 규칙적인 간격의 데이터가 아니라면 각 관측 값에 점으로 표시하여 오해를 줄이자



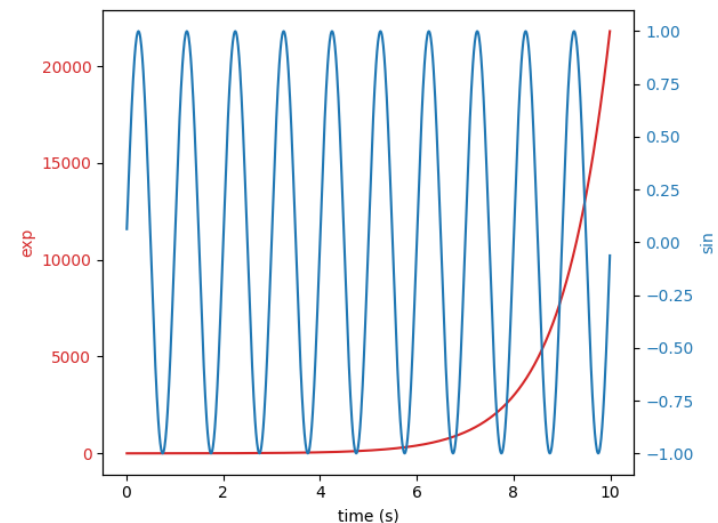
- Line은 점을 이어 만드는 요소. -> 점과 점 사이에 데이터가 없기에 이를 잇는 방법(보간)
- 데이터의 error나 noise가 포함되어 있는 경우, 데이터의 이해를 돕는 방법
 - Moving Average
 - Smooth Curve with Scipy
 - `scipy.interpolate.make_interp_spline()`
 - `scipy.interpolate.interp1d()`
 - `scipy.ndimage.gaussian_filter1d()`
- Presentation에는 좋은 방법일 수 있으나
 - 없는 데이터를 있다고 생각하게 할 수 있으며
 - 작은 차이를 없앨 수 있음
 - **일반적인 분석에서는 지양할 것!**



2.4 이중 축 사용

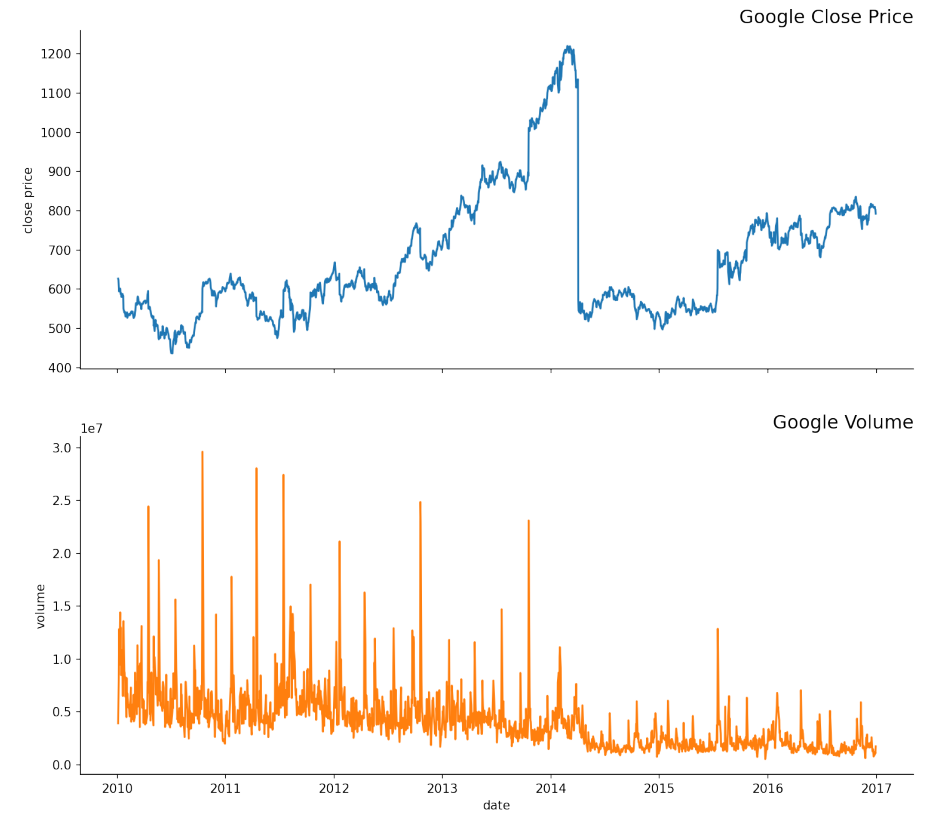
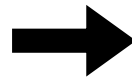
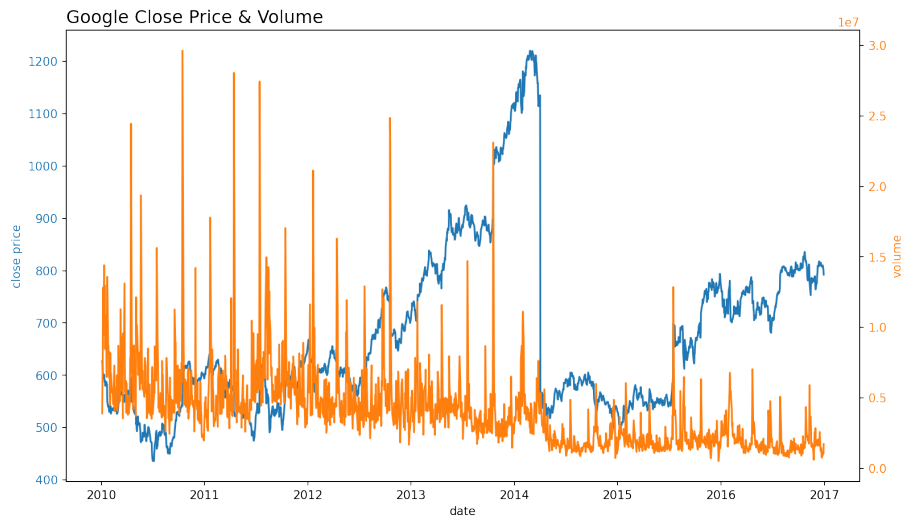
2. 정확한 Line Plot

- 한 plot에 대해 2개의 축을 **이중 축(dual axis)**라고 함
- 같은 시간 축에 대해 서로 다른 종류의 데이터를 표현하기 위해서는 축이 2개가 필요
 - **.twinx()**를 사용
- 한 데이터에 대해 다른 단위 (ex. radian과 degree)
 - **.secondary_xaxis()**, **.secondary_yaxis()** 사용
- 2개의 plot을 그리는 것 >>>> 이중 축 사용
 - **이중 축은 지양**할 것.

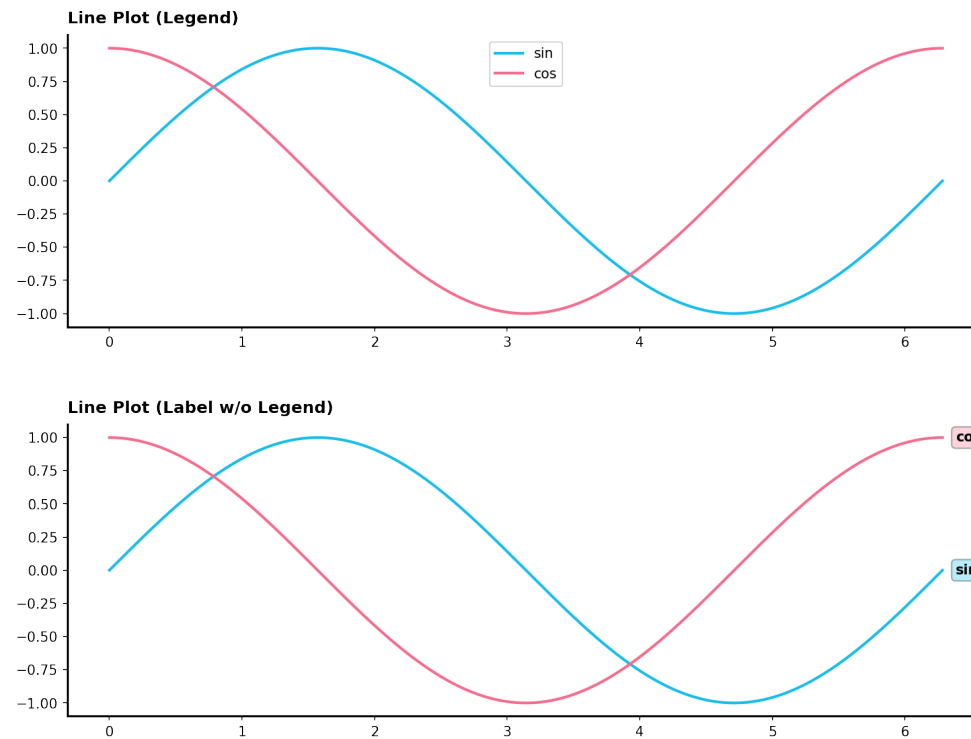


2.4 이중 축 사용

2. 정확한 Line Plot



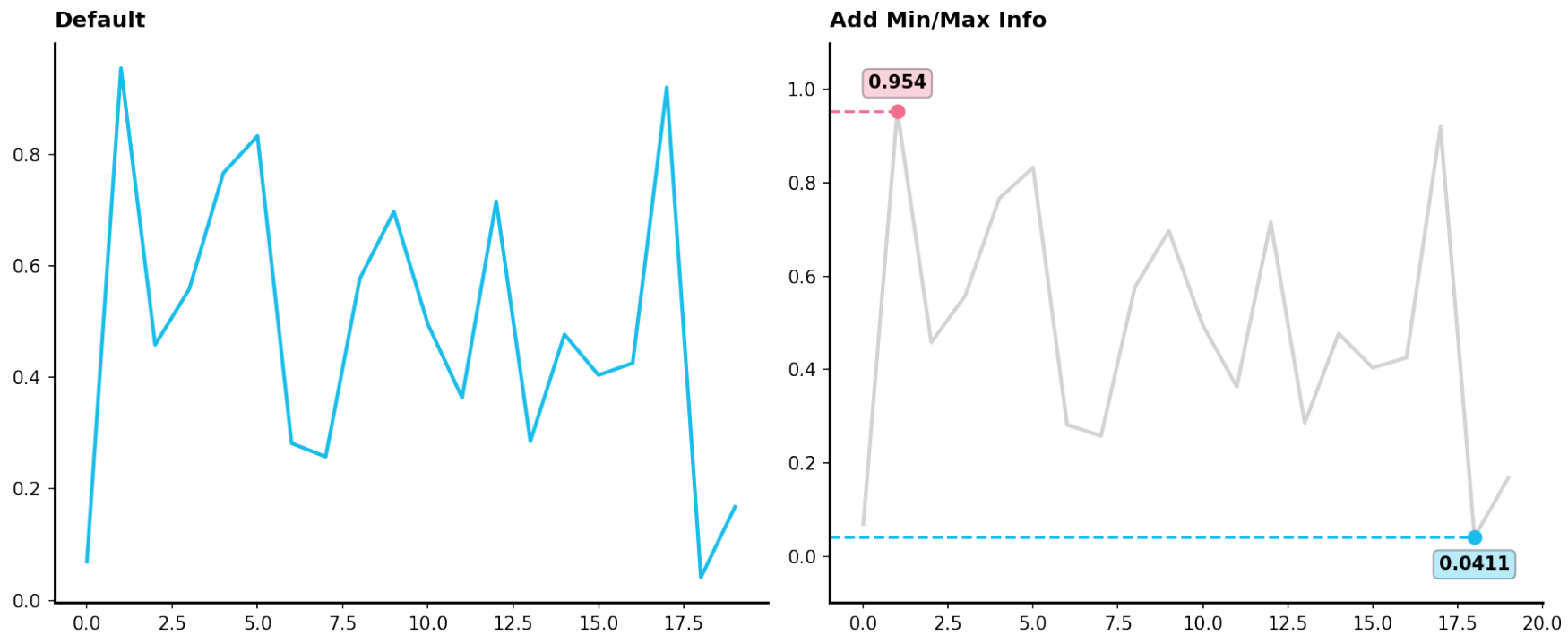
- 라인 끝 단에 레이블을 추가하면 식별에 도움 (범례 대신)



2.5 ETC

2. 정확한 Line Plot

- Min/Max 정보(또는 원하는 포인트)는 추가해주면 도움이 될 수 있음 (annatation,)



- 보다 연한 색을 사용하여 uncertainty 표현 가능 (신뢰구간, 분산 등)

