



2강 데이터 시각화 기초 다지기

강사: 김진용

목차



1. 시각화 : 독자를 설득하는 디자인

- a. 데이터 시각화
- b. 시각 속성 체계
- c. 색 사용법
- d. 인지 부하 낮추는 대표 원칙

2. 차트의 요소

- a. 색
- b. 텍스트

3. 어떤 차트를 써야할까?

- a. 막대 그래프
- b. 선 그래프
- c. 산포도

시각화의 효과



운전해본 경험이 있으신가요?

갈림길에서 빠져야 하는지?

갈림길로 빠졌더니 갈림길이 또 2개나 나온 경우를 경험해보셨나요?

고속도로에서 길을 잘못 판단할 경우 먼 거리를 다시 우회할 수도 있고 자칫 사고가 날 수도 있습니다.

시각화의 효과



출처 : <https://blog.naver.com/ladynmom/222996972924>

시각화의 효과

분홍색 선 : 갈라지는 길이 하나일 때 사용

초록색 선 : 갈라지는 길이 2개일 때 분홍색과 같이 사용



시각화의 효과





1. 시각화 : 독자를 설득하는 디자인

1. 시각화 : 독자를 설득하는 디자인



1.1 데이터 시각화

데이터 시각화란?

→ 데이터에 숨어있는 중요한 정보를 사람의 인지 시스템이 쉽게 찾아낼 수 있게 데이터의 형태를 표현한 것

1. **시각적 이해:** 시각화는 복잡한 데이터와 통계적인 패턴을 시각적으로 이해하기 쉽게 만들어줍니다.
2. **인사이트 도출:** 데이터 시각화는 데이터로부터 인사이트를 도출하는 데 도움을 줍니다.
3. **효과적인 커뮤니케이션:** 데이터 시각화는 효과적인 커뮤니케이션 도구로 사용됩니다.

1. 시각화 : 독자를 설득하는 디자인

1.1 데이터 시각화

퀴즈:

숫자 7이 몇개인지 10초 안에 찾아보세요.

123948192581928341892635917283479182
759812734897129835671982374891263498
172389561892374981273489162356198237
498126539812374981237894791285698123



1. 시각화 : 독자를 설득하는 디자인



1.1 데이터 시각화

7은 몇개일까요?



1. 시각화 : 독자를 설득하는 디자인

1.1 데이터 시각화



대답하신 개수가 정말 확실한가요?



1. 시각화 : 독자를 설득하는 디자인

1.1 데이터 시각화

퀴즈:

숫자 7이 몇개인지 10초 안에 찾아보세요.

힌트 - 숫자 7의 색상은 빨간색입니다.

123948192587192834189263591728347918
275981234891298356719823748912634981
723895618923749812734891623561982349
812653981237498123789479128569812383



1. 시각화 : 독자를 설득하는 디자인



1.1 데이터 시각화

7은 몇개일까요?



1. 시각화 : 독자를 설득하는 디자인



1.1 데이터 시각화

데이터 시각화의 목적은 숫자 7을 빠르게 셀 수 있는 사람이 답을 하기 위해 하는 것이 아닙니다.

즉, 데이터를 해석할 수 있는 일부의 사람 뿐 아니라 적절하게 시각화해서 모두가 같은 사실을 인지하도록 하는 것을 목표로 합니다.

123948192587192834189263591728347918
275981234891298356719823748912634981
723895618923749812734891623561982349
812653981237498123789479128569812383

1. 시각화 : 독자를 설득하는 디자인

1.1 데이터 시각화

앤스컴 콰르텟

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.5
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

- ❑ 데이터셋을 만든 사람 : 프랭크 앤스컴
- ❑ 데이터 표현 형태 : 표

어떻게 분석할 것인가?

객관적으로 데이터를 파악하는 방법 →
통계!

출처 :

https://ko.wikipedia.org/wiki/%EC%95%A4%EC%8A%A4%EC%BB%B4_%EC%BD%B0%EB%A5%B4%ED%85%9F

1. 시각화 : 독자를 설득하는 디자인

1.1 데이터 시각화

항목	값	정확도
x 평균	9	정확
x 표본분산	11	정확
y 평균	7.50	소수점 2자리
y 표본분산	4.125	정확
x 와 y 의 상관	0.816	소수점 3자리
선형회귀선	$y = 3.00 + 0.500x$	각 소수점 2자리, 소수점 3자리
선형회귀 결정계수	0.67	소수점 2자리

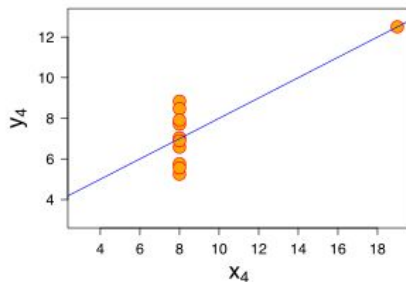
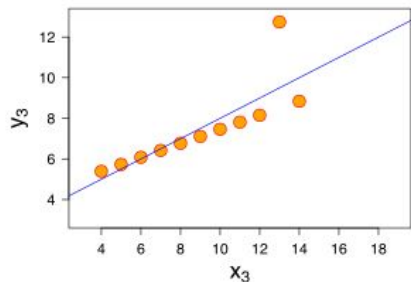
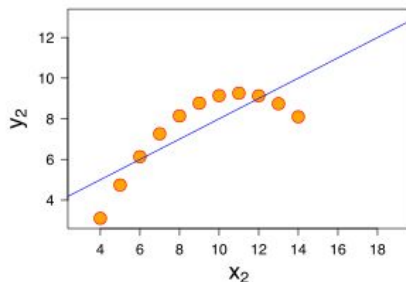
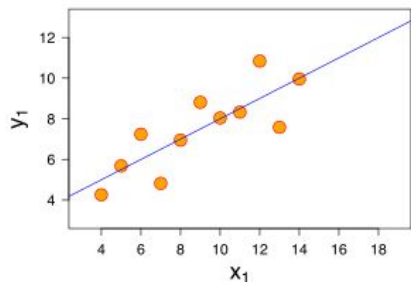
통계치(평균, 분산 등)가 동일

→ 객관적으로 볼 때 4가지
데이터가 같은 경향을 가짐.

정말 그럴까?

1. 시각화 : 독자를 설득하는 디자인

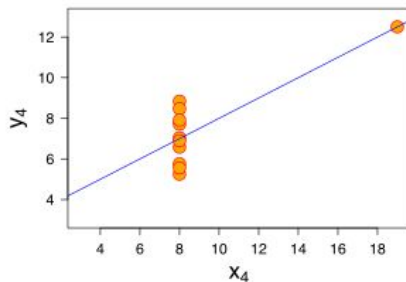
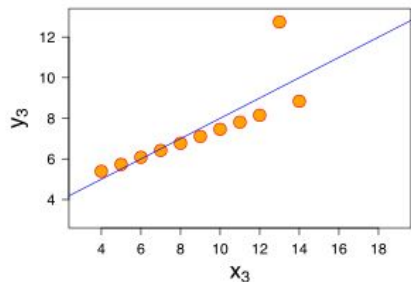
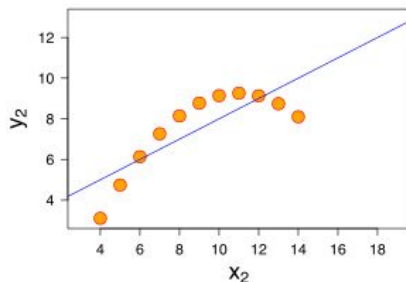
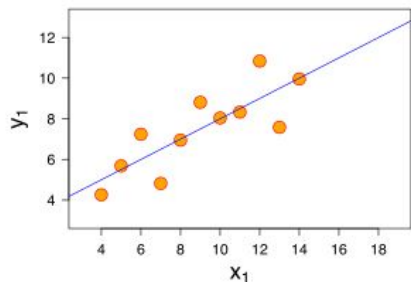
1.1 데이터 시각화



1. 다소 퍼져 있지만 x 와 y 가 같이 증가하는 경향
2. x 는 증가하고, y 는 9을 기준으로 감소하는 경향
3. x 와 y 가 선형으로 증가하는 경향, y 하나가 벗어남.
4. x 는 동일한 값이고 y 는 증가하는 경향, 값 하나가 벗어남

1. 시각화 : 독자를 설득하는 디자인

1.1 데이터 시각화



시각화를 해야 하는 이유

1. 표나 통계만으로 파악하기 어려운 데이터의 추세를 알아볼 수 있음.
2. 데이터가 가진 경향이 다르면 비즈니스 관점에서 취할 행동이나 아이디어가 크게 달라질 수 있음.

1. 시각화 : 독자를 설득하는 디자인



1.2 시각 속성

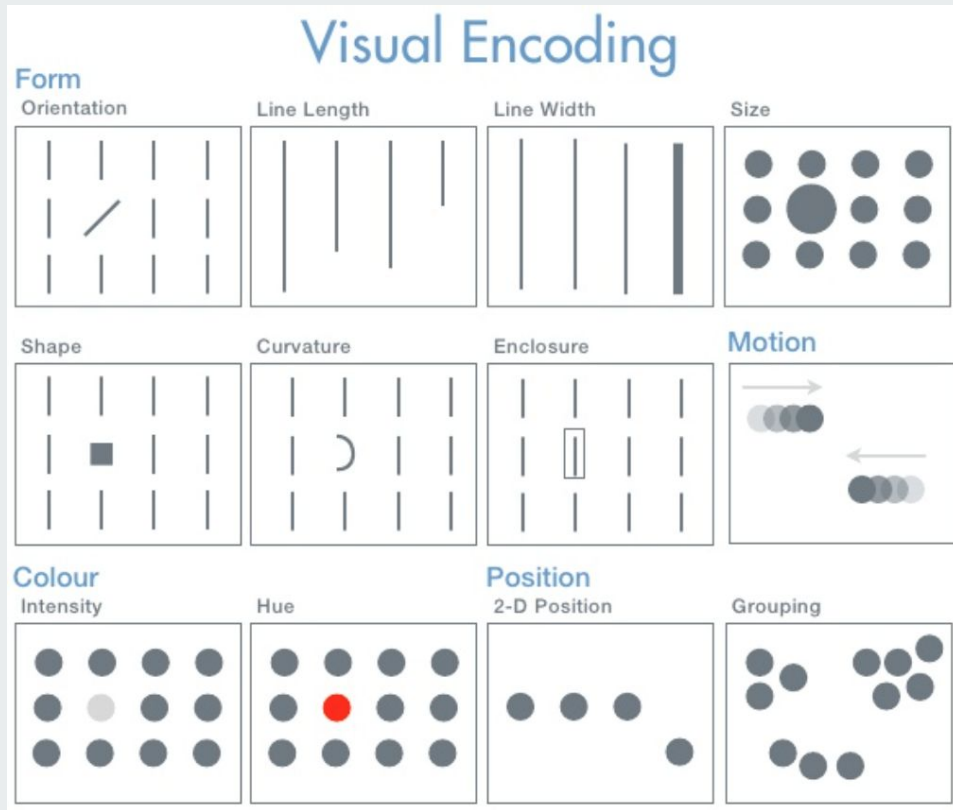
7을 빨리 셀 수 있었던 이유

“색” = 밀리초 단위로 차이에 반응할 수 있는 **시각 속성**

1. 시각화 : 독자를 설득하는 디자인

1.2 시각 속성

시각 속성을 적절히 활용해
데이터 시각화를 하면
상대의 기억에 잘 남게 할 수
있다.



출처

<https://www.pinterest.co.kr/pin/preattentive-attributes--396316835946556714/>

1. 시각화 : 독자를 설득하는 디자인

1.2 시각 속성

시각 속성을 부적절하게 사용하는 경우

모든 숫자에 색깔을 넣는 것은 7이 몇개인지 세는 것에 도움이 되지 않습니다.

123948192587192834189263591728347918
275981234891298356719823748912634981
723895618923749812734891623561982349
812653981237498123789479128569812383



1. 시각화 : 독자를 설득하는 디자인



1.3 색 사용법

색을 사용하는 목적

- ❑ 순차적 색상
- ❑ 확산적 색상
- ❑ 카테고리 색상
- ❑ 하이라이트 색상

1. 시각화 : 독자를 설득하는 디자인

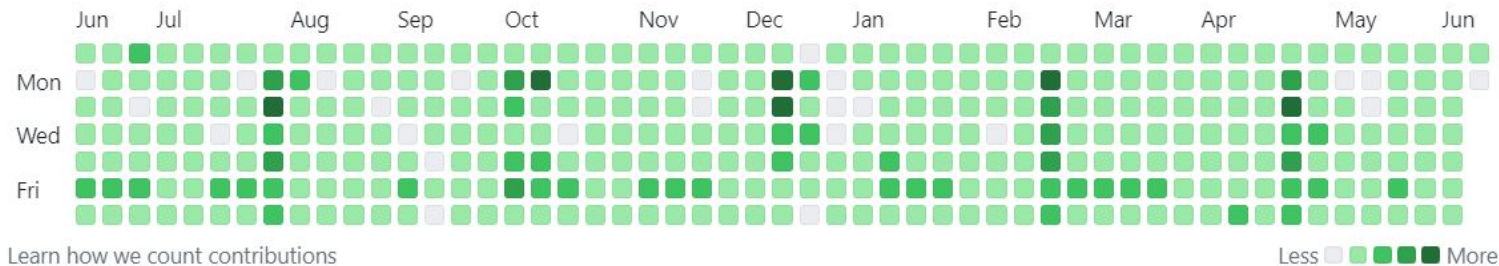
1.3 색 사용법

순차적 색상

색의 명도(밝기)로 데이터를 표현



2,486 contributions in the last year

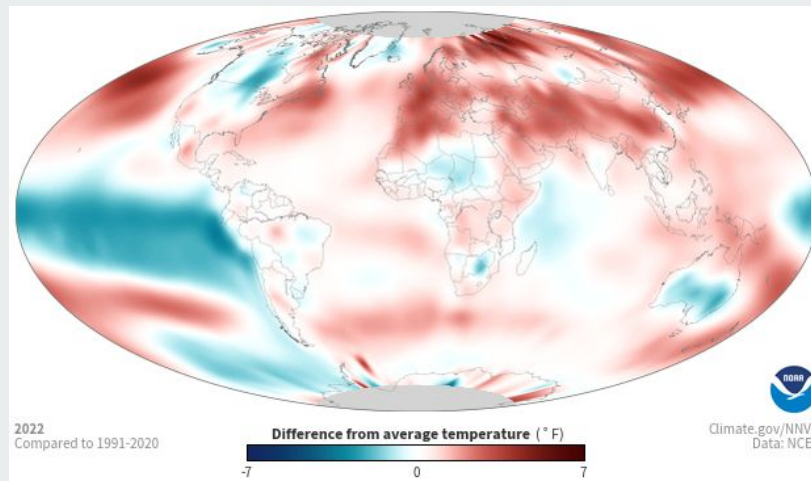


1. 시각화 : 독자를 설득하는 디자인

1.3 색 사용법

확산적 색상

중앙점을 기점으로 2가지 색을 순차적 색상으로 표현



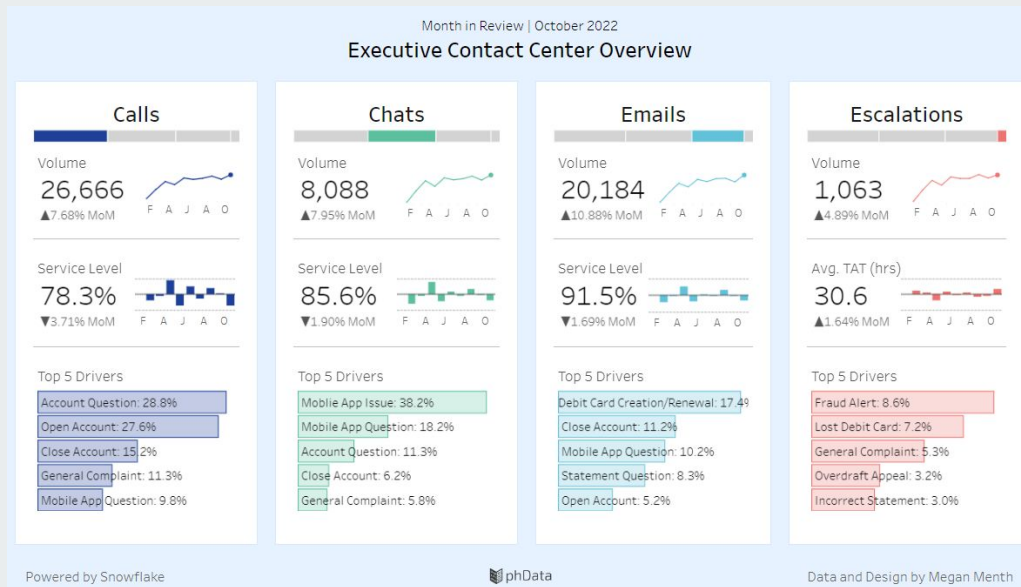
출처 :
<https://www.climate.gov/maps-data/data-snapshots/data-source/temperature-global-yearly-difference-average>

1. 시각화 : 독자를 설득하는 디자인

1.3 색 사용법

카테고리 색상

서로 다른 분류나 구분을 색상으로 표현
4가지 색 이하로 구분하는 것을 추천



출처 :

<https://public.tableau.com/app/profile/megan.menth/viz/CallCenterExecutiveView/CallCenterExecView>

1. 시각화 : 독자를 설득하는 디자인

1.3 색 사용법

하이라이트 색상

독자에게 강조하고자 하는 정보를
하이라이트 표시

Student Sleep Deficit

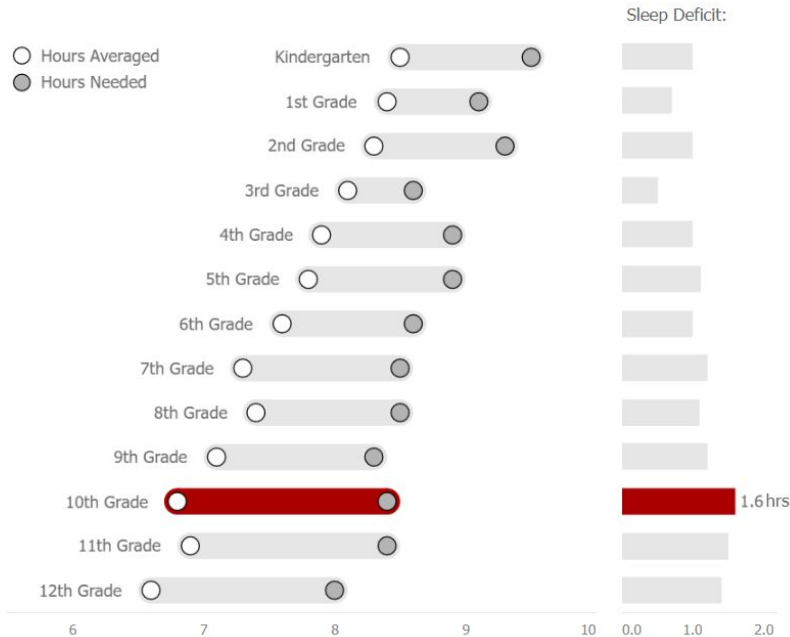
Sleep hours needed vs averaged on school night deficits by grade

Survey of 1000 parents was conducted about their children's sleep health to know if they were getting enough sleep.

CDC* recommends 9-10 hours of sleep for 6-12 years olds and 8 to 10 hours of sleep for 13-18 year olds. On average none of the age groups meet these recommendations.

The sleep deficit gets worse as students grow older. **As per the results, 10th graders have the highest sleep deficit of 1.6 hours.**

Getting enough sleep is as critical as getting enough food and nutrients. Rest and a good night's sleep is important for a robust immune system, focus and positive mood.



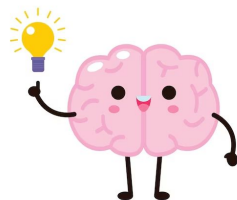
Design: Zunaira (@ZeeVizGirl) | Project: #MakeoverMonday | Data: savvysleeper.org

1. 시각화 : 독자를 설득하는 디자인

1.4 인지 부하 낮추는 대표 원칙

데이터 시각화에서 인지부하란?

- ❑ 독자는 데이터 시각화를 이해하기 위해서 뇌의 에너지를 사용합니다.
- ❑ 뇌의 에너지를 많이 사용해야 할수록 부담이 가고 눈길이 가지 않게 됩니다.
- ❑ 여러분이 혼신을 다한 시각화 작품이 누군가에게는 인지적인 부담으로 다가와 그들의 관심에서 벗어날 가능성이 있습니다.
- ❑ 그렇기에 인지 부하를 최소화하는 노력이 필요합니다.



1. 시각화 : 독자를 설득하는 디자인



1.4 인지 부하 낮추는 대표 원칙

- ❑ 데이터 잉크 레이션
- ❑ 게슈탈트 법칙

1. 시각화 : 독자를 설득하는 디자인

1.4 인지 부하 낮추는 대표 원칙

▶ 데이터 잉크 레이션 - 에드워드 터프티(Edward Tufte)

그래프에 존재하는 모든 요소가 데이터를 나타내거나 시사하는 바를 전달하는데 필수적이어야 한다.

$$\frac{\text{data-ink}}{\text{total-ink}} = \frac{\text{Elements conveying data information}}{\text{All elements in the chart}}$$

$$= \frac{\text{데이터를 나타내거나 정보를 전달하기 위해 사용된 잉크의 양}}{\text{그래프에 있는 모든 요소에 사용된 잉크의 양}}$$

1. 시각화 : 독자를 설득하는 디자인



1.4 인지 부하 낮추는 대표 원칙

데이터 잉크 레이션스를 왜 지켜야 하는가?

- **시사점 명확히 전달** - 그래프에서 필수적인 요소만을 남김으로써 여러분의 독자는 여러분이 전달하고자 하는 바를 빠르고 명확하게 이해할 수 있습니다.
- **시간 절약** - 독자가 그래프를 잘 이해하지 못하여 재질문을 하거나 헷갈려할 가능성을 낮춰 전달하는 흐름을 끊지 않을 수 있습니다.
- **공간 절약** - 여러 개의 차트가 한꺼번에 나타낼 때 필요없는 요소를 제거함으로써 동일한 공간에 정보를 함축할 수 있습니다.

1. 시각화 : 독자를 설득하는 디자인



1.4 인지 부하 낮추는 대표 원칙

▶ 게슈탈트 법칙

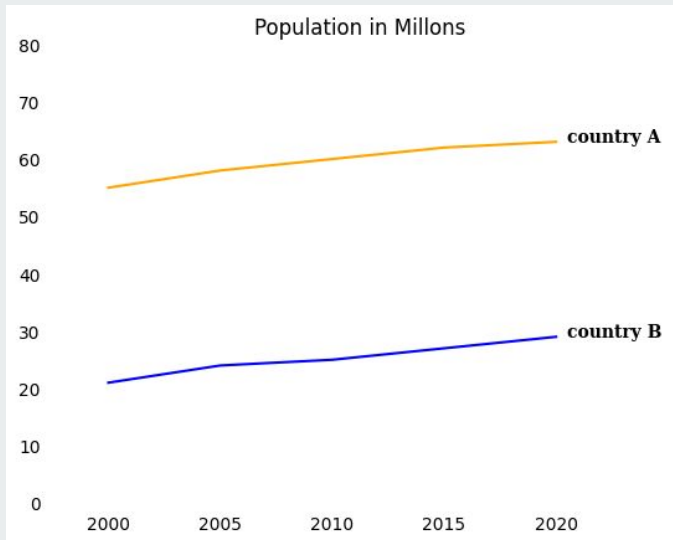
- ❑ 근접성 법칙(Law of Proximity)
- ❑ 유사성 법칙(Law of Similarity)
- ❑ 인클로저 법칙(Law of Enclosure)
- ❑ 폐쇄성 법칙(Law of Closure)
- ❑ 연속성 법칙(Law of Continuity)
- ❑ 연결성 법칙(Law of Connection)

1. 시각화 : 독자를 설득하는 디자인

1.4 인지 부하 낮추는 대표 원칙

근접성 법칙(Law of Proximity)

→ 물리적으로 같은 거리에 있는 대상을 같은 그룹으로 간주하는 법칙



누군가 알려주지 않아도

오렌지 선이 country A

파란색 선이 country B

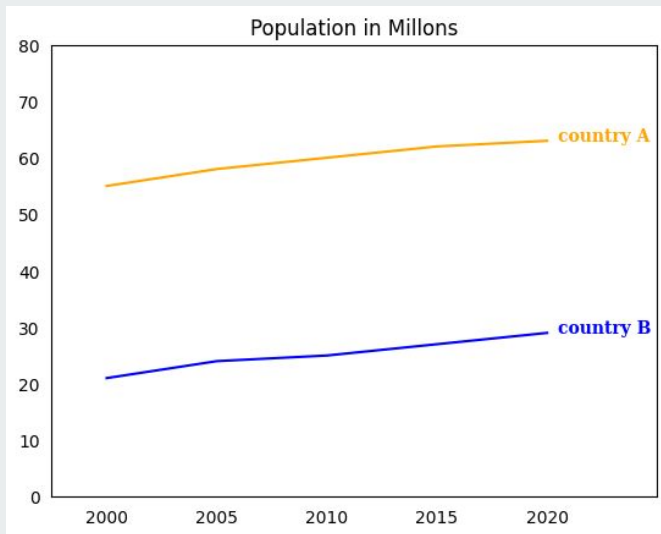
라는 것을 알 수 있습니다.

1. 시각화 : 독자를 설득하는 디자인

1.4 인지 부하 낮추는 대표 원칙

유사성 법칙(Law of Similarity)

→ 같은 색, 형태, 방향을 가진 것을 같은 그룹으로 인지하는 법칙



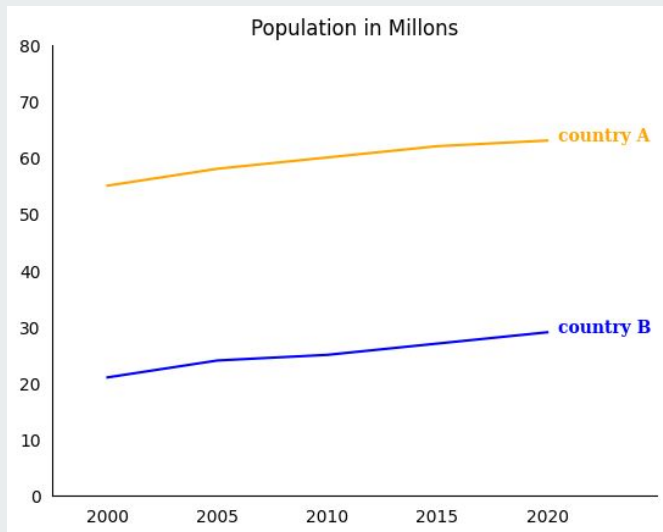
country A, country B line과 같은 색깔로 라벨링하여 더 눈에 쉽게 눈에 띄게 할 수 있습니다.

1. 시각화 : 독자를 설득하는 디자인

1.4 인지 부하 낮추는 대표 원칙

폐쇄성 법칙(Law of Enclosure)

→ 빠진 부분이 있더라도 알고 있던 이미지로 대상을 인지하려고 하는 법칙



그래프 주위를 감싸고 있는 선들을 제거하더라도 그래프 자체를 인지할 수 있습니다.

→ 더 깔끔한 이미지!



2. 차트의 요소

2. 차트의 요소

2.1 색

색 속성 효과적으로 사용하는 방법

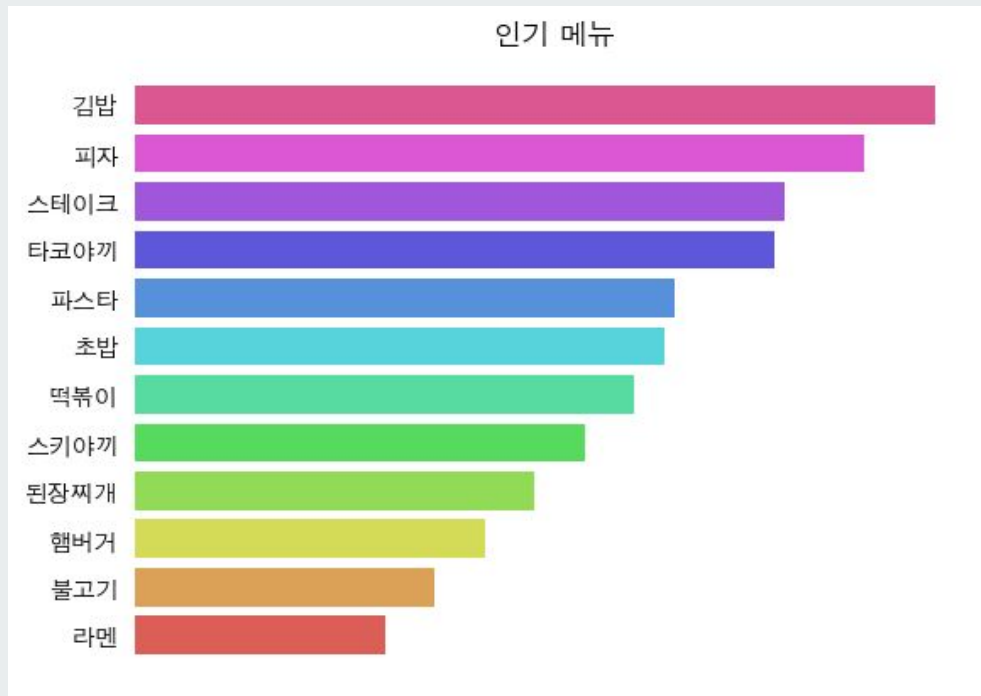
사용하는 색의 수를 줄여라!

→ 사람이 동시에 인지 가능한 색은 **최대 8가지**



2. 차트의 요소

2.1 색

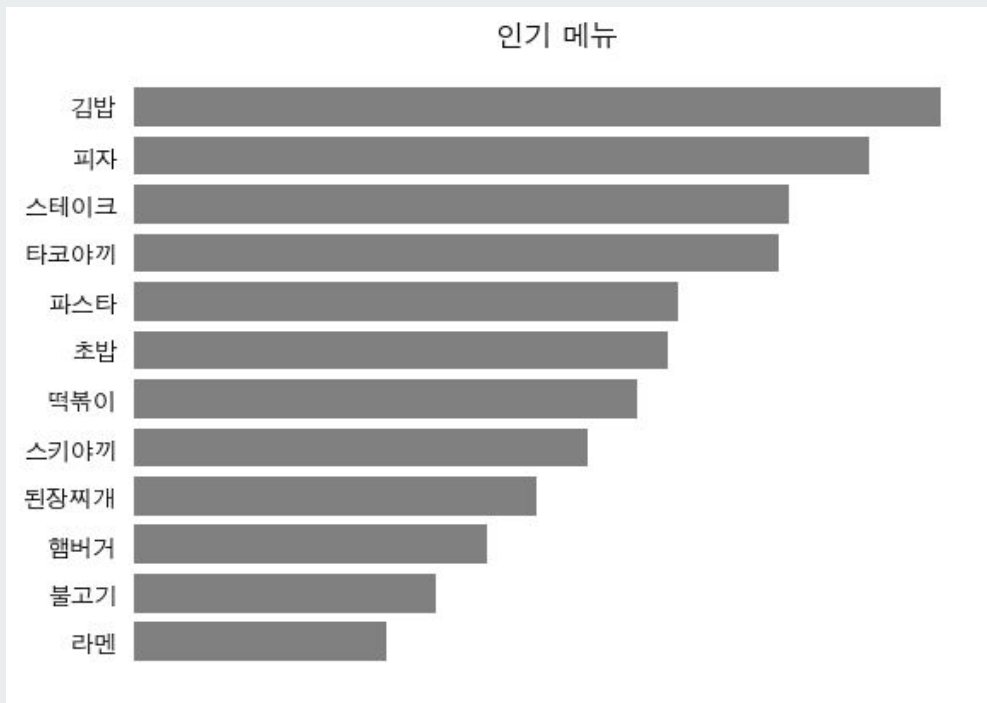


데이터의 개수만큼 많은 색
→ 데이터 시각화의 목적은
화려하게 보여주는 것이
아님



2. 차트의 요소

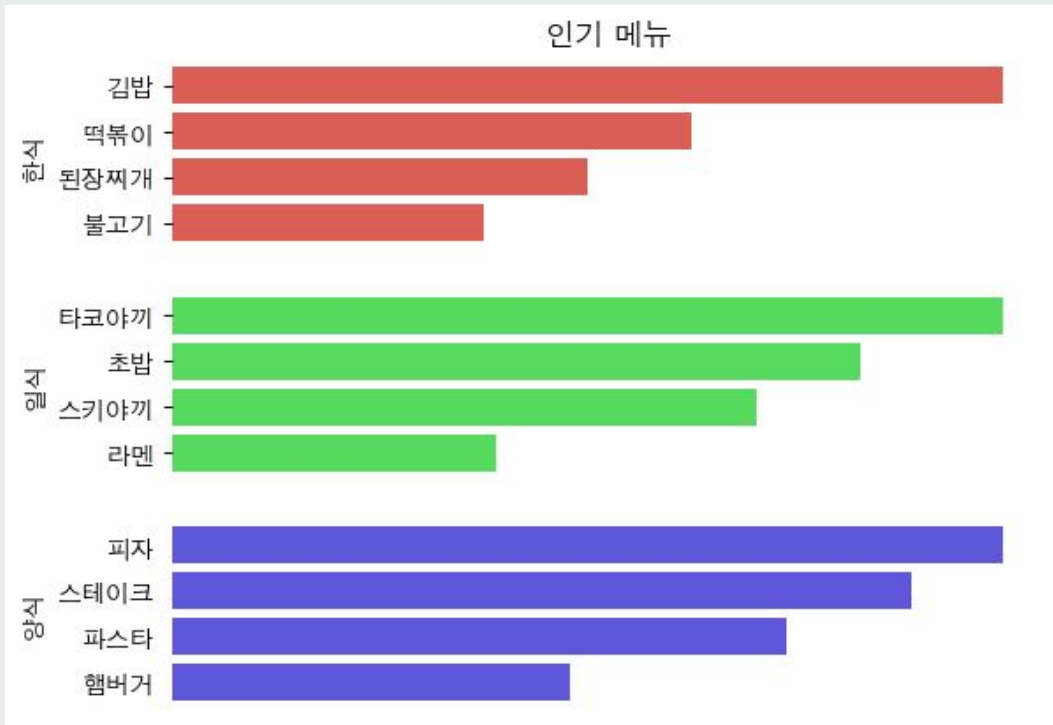
2.1 색



Simple is the best!

2. 차트의 요소

2.1 색



카테고리별로 색깔 분류

→ 색깔의 개수 : 3

→ 깔끔한 느낌

2. 차트의 요소



2.1 색

어떤 색을 사용해야 좋을까요?

[How to pick more beautiful colors for your data visualizations](#)

2. 차트의 요소



2.2 텍스트

데이터 시각화에서 텍스트의 역할

- ❑ 주제를 설명합니다.
- ❑ 데이터의 상세한 정보를 라벨링합니다.
- ❑ 그래프만으로 알 수 없는 세부적인 내용을 보충합니다.

2.2 텍스트

- 피가 흐르는 듯한 이미지
- 전쟁의 참혹함이 강조됨
- 아래 방향으로 막대 그래프를 그림

[illegible]

2. 차트의 요소

2.2 텍스트

제목 : 이라크 전쟁 : 사상자 감소

→ 희생자가 감소하고 있다는 희망적인 메시지

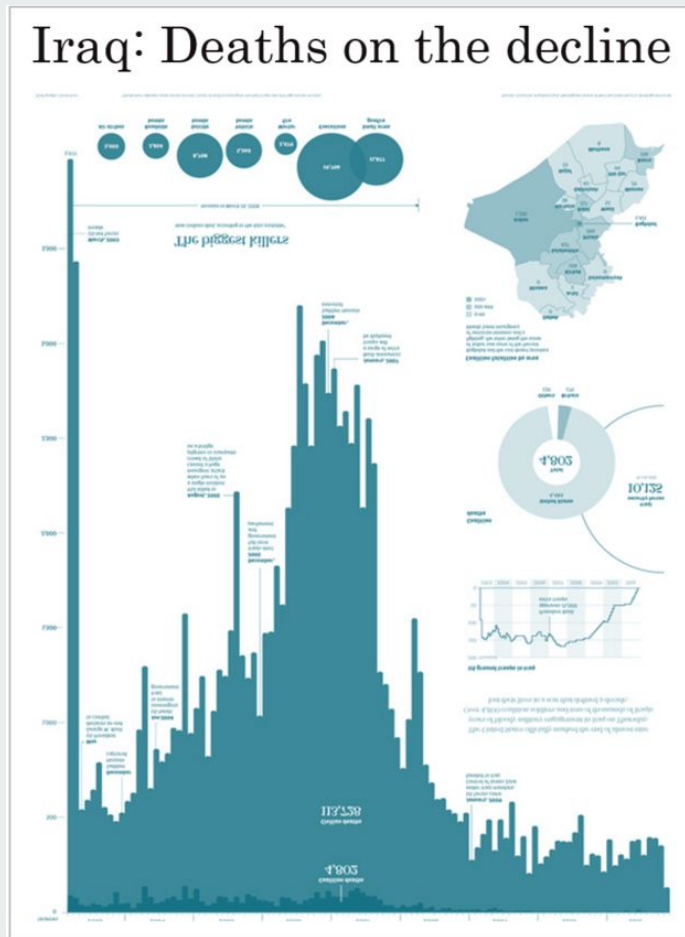
→ 막대 그래프를 위쪽 방향으로 그리고 색을 변경

같은 데이터

→ 서로 상반된 주장이 가능하기 때문에
독자에게 어떤 메시지를 전달할 것이냐를 명확히
하는 것이 중요합니다.

출처 :

<https://www.infoworld.com/article/3088166/why-how-to-lie-with-statistics-did-us-a-disservice.html>

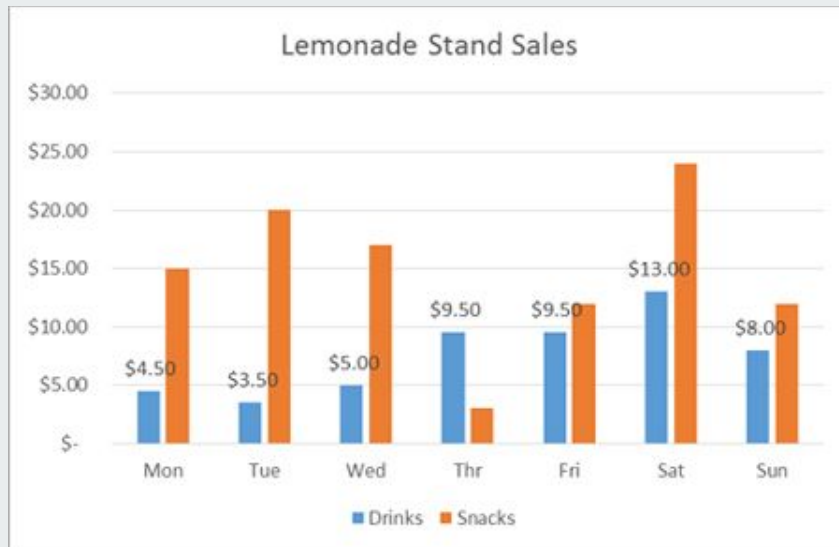


2. 차트의 요소

2.2 텍스트

라벨 : 데이터 참조 역할

* 구체적인 숫자를 표시





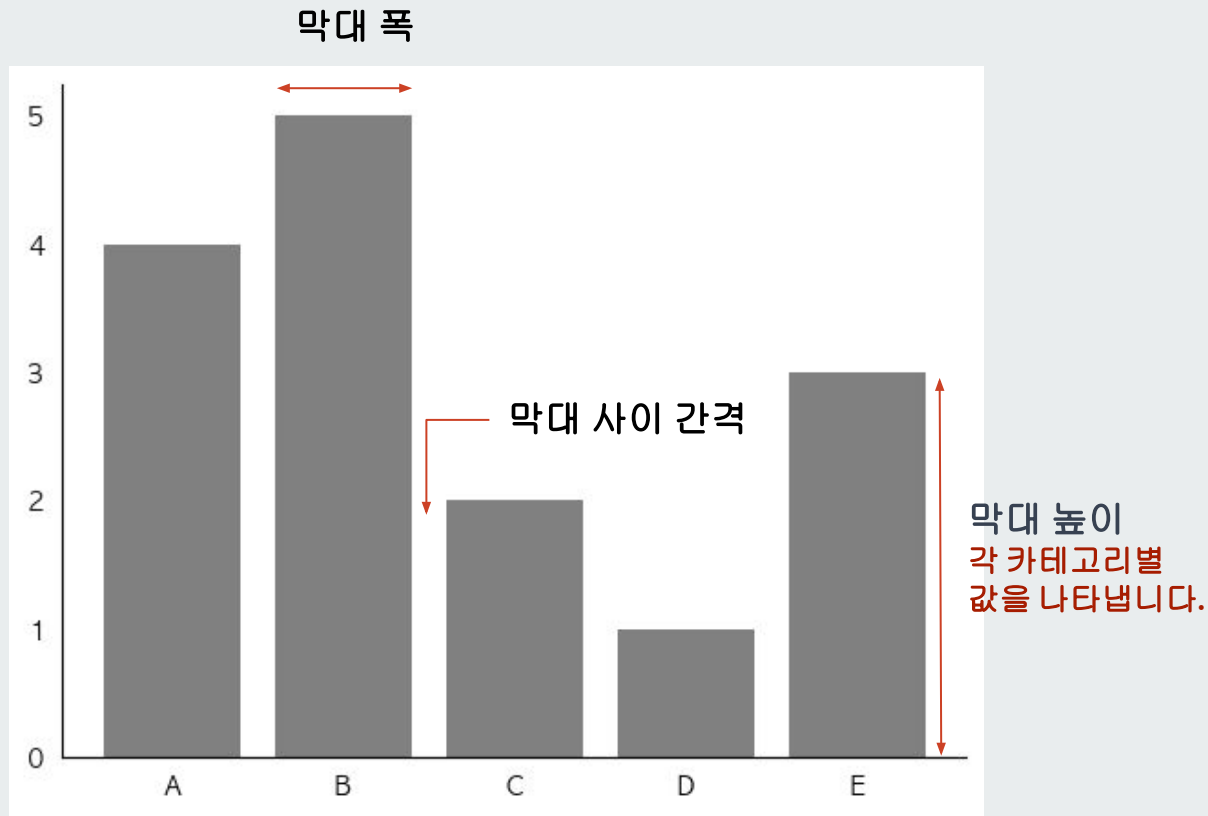
3. 어떤 차트를 써야할까?

3. 어떤 차트를 써야할까?

3.1 막대 그래프

막대 그래프란?

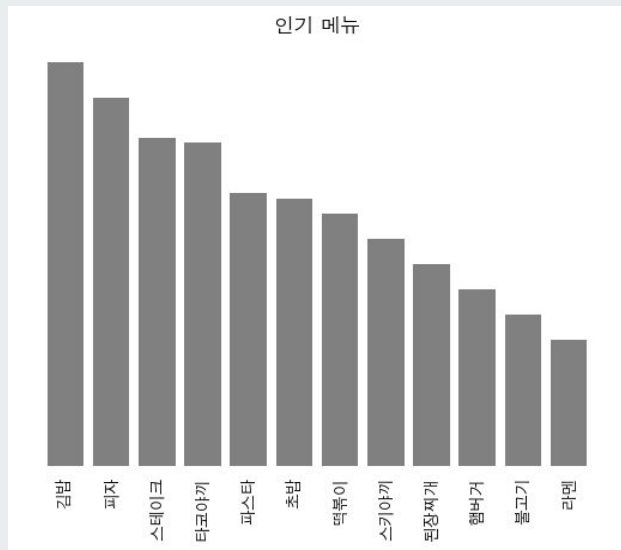
- 양을 비교할 때 사용
- 시각 속성 중 높이와 위치를 활용하며 높이가 높을수록 값이 크고 모든 막대가 0이라는 위치에서 시작함



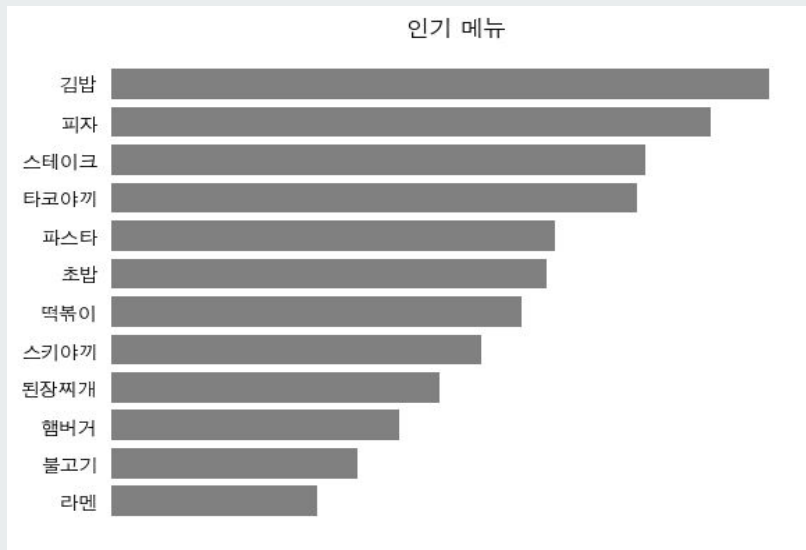
3. 어떤 차트를 써야할까?

3.1 막대 그래프

막대 그래프 사용 시 주의할 점



세로 막대 그래프



가로 막대 그래프

1. 카테고리가 많다면 가로 방향으로 그리기

- > 세로 막대 그래프는 카테고리가 많아질수록 지면이 부족하여 뻥뻥해지고 가독성이 떨어짐
- > 라벨이 길어져 90도를 회전하게 되면 독자가 고개를 틀어서 봐야 하므로 인지부하가 커짐

3. 어떤 차트를 써야할까?

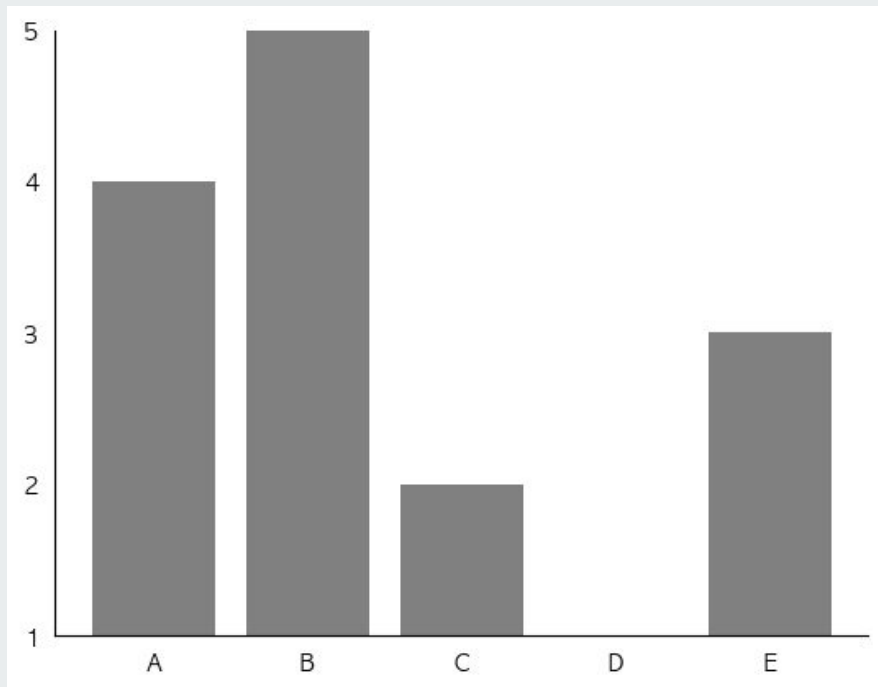
3.1 막대 그래프

막대 그래프 사용 시 주의할 점

2. 막대그래프는 0부터 시작

→ 1부터 시작시 D의 데이터는 사라짐.

→ B 막대 높이는 C 막대 높이의 4배로 보임.

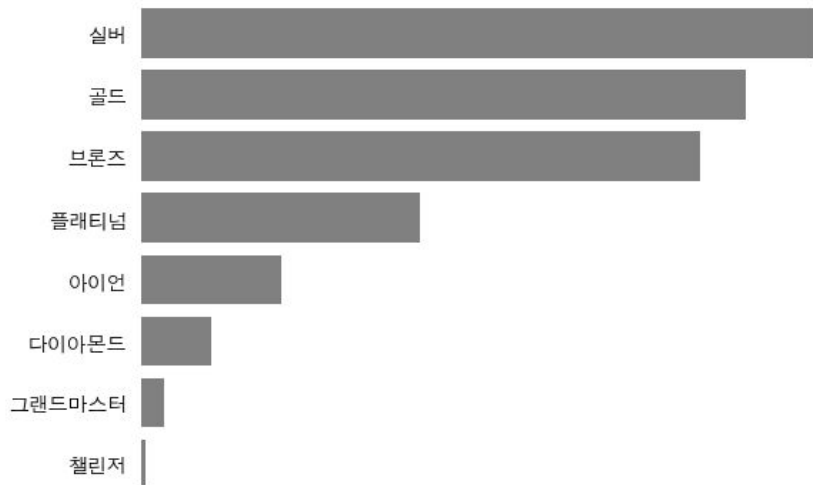


3. 어떤 차트를 써야할까?

3.1 막대 그래프

막대 그래프 사용 시 주의할 점

랭크 분포도

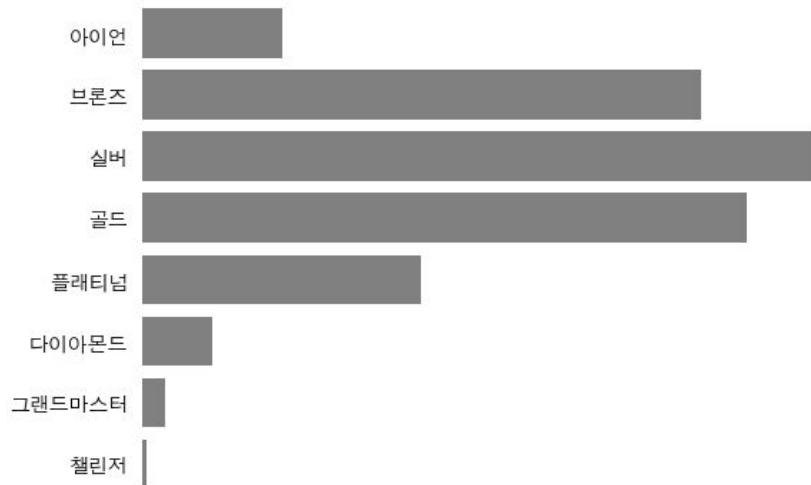


3. 순서를 가진 카테고리 무조건 내림차순 정렬하지 않기

-> 혼란을 야기할 수 있음

-> 순서 그대로 정렬하는 것이 좋을 수도 있음

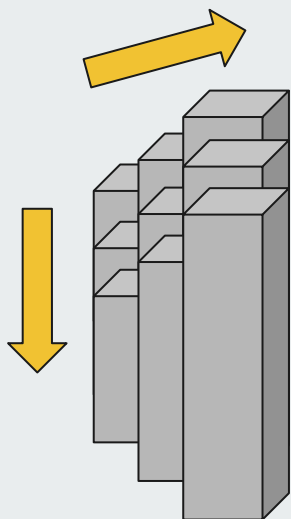
랭크 분포도



3. 어떤 차트를 써야할까?

3.1 막대 그래프

막대 그래프 사용 시 주의할 점

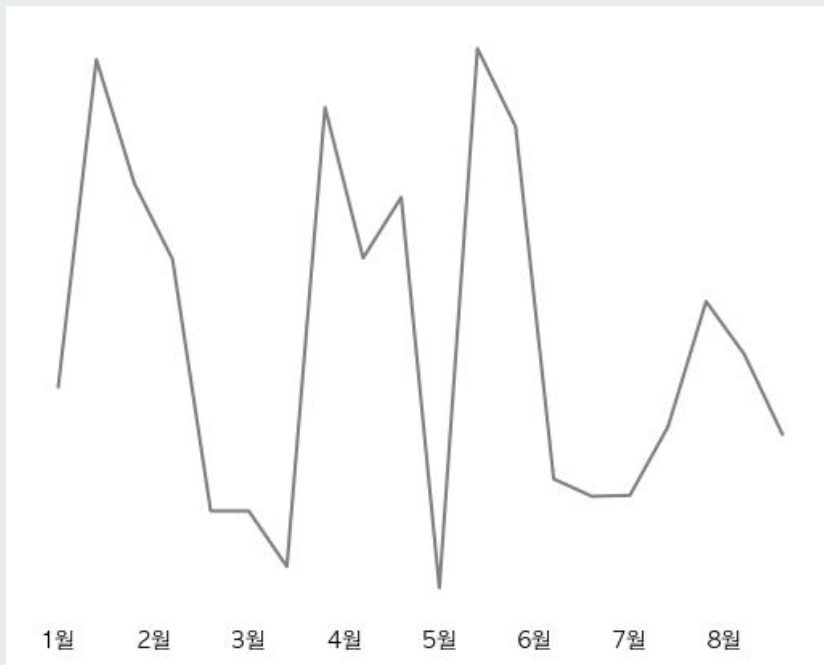


4. 3D 막대 사용하지 않기

- > 막대 그래프의 시작점이 0이 되지 않아 막대 높이 비교가 어려움
- > 뒤의 막대 그래프를 가릴 경우 보이지 않을 수 있음
- > 시각적으로 화려해보일 수는 있지만 전달하고자 하는 메시지를 잘 나타내지 못할 수 있음

3. 어떤 차트를 써야할까?

3.2 선 그래프



선 그래프란?

□ 추이를 표시할 때 사용

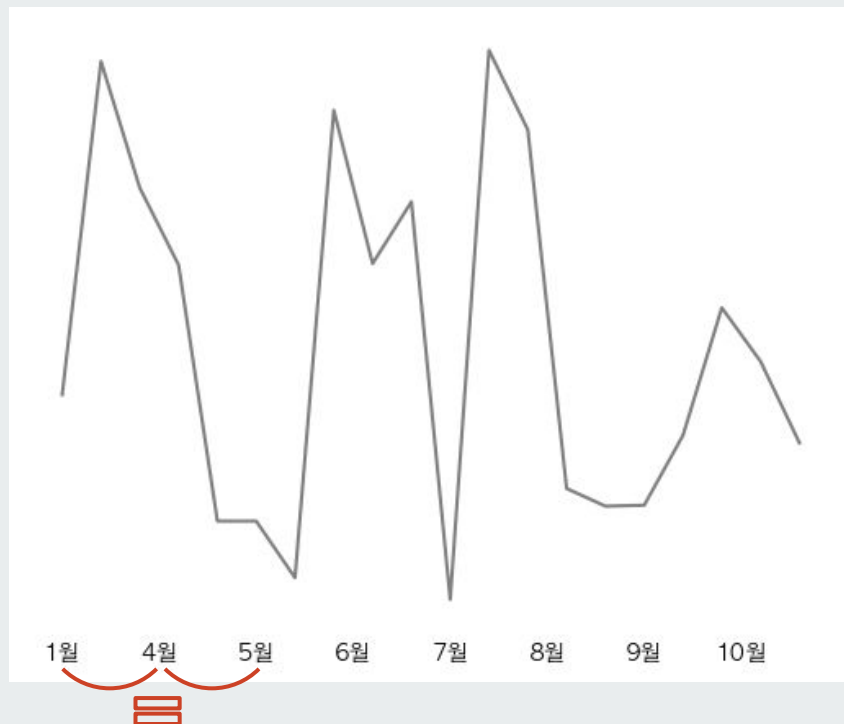
3. 어떤 차트를 써야할까?

3.2 선 그래프

선 그래프 사용 시 주의할 점

1. 가로축 간격의 왜곡

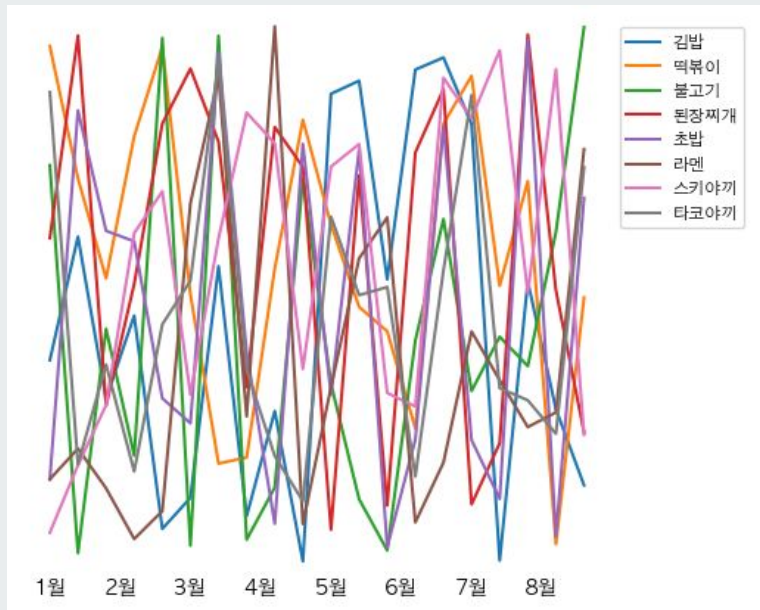
- > 점과 점 사이의 선의 기울기를 통해 추이를 파악하기 때문에 간격이 왜곡되면 사실을 나타내지 못함
- > 1월과 4월 사이의 간격에 2월, 3월만큼의 추가적인 공간을 확보해야함



3. 어떤 차트를 써야할까?

3.2 선 그래프

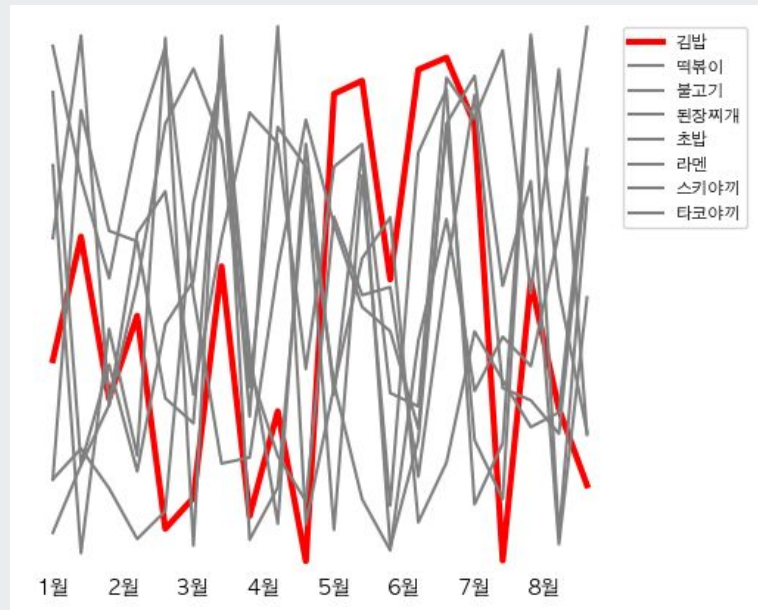
선 그래프 사용 시 주의할 점



2. 주목하고 싶은 부분만 강조하기

-> 색이 너무 많을 경우 인지하기 어려움

-> 선의 색이나 굵기를 바꾸기

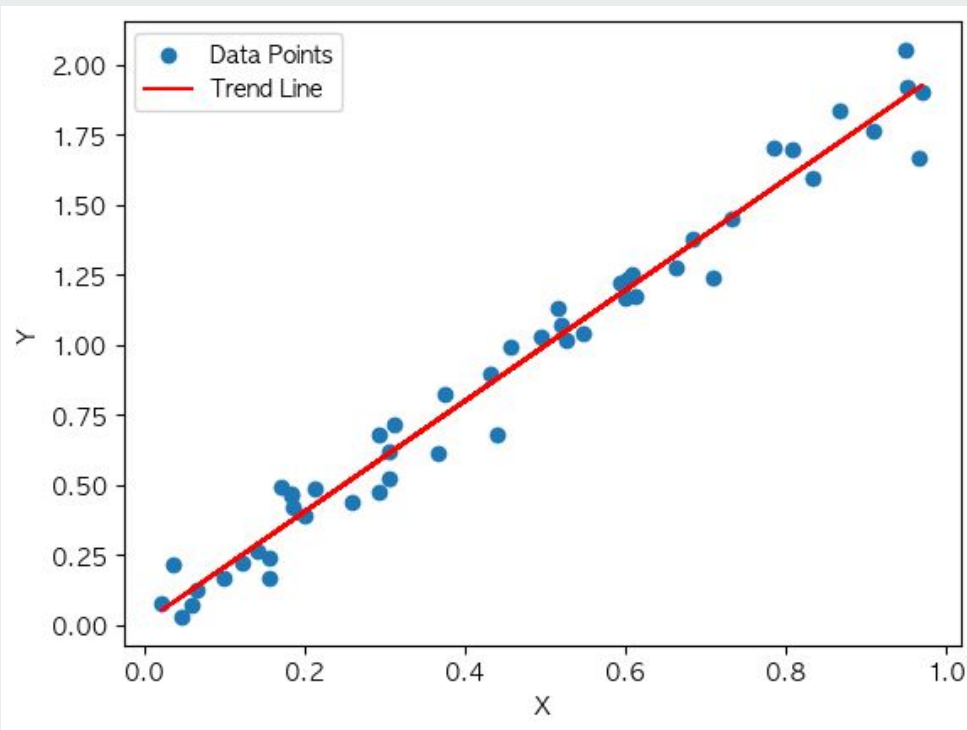


3. 어떤 차트를 써야할까?

3.3 산포도

산포도란?

- ❑ 변수 2개의 관계성을 표시할 때 사용
- ❑ 데이터가 오른쪽 그래프와 같은 분포를 나타낼 때 X, Y는 양의 상관관계가 있음



정리하기



1. 시각화 : 독자를 설득하는 디자인

- a. 데이터 시각화
- b. 시각 속성 체계
- c. 색 사용법
- d. 인지 부하 낮추는 대표 원칙

2. 차트의 요소

- a. 색
- b. 텍스트

3. 어떤 차트를 써야할까?

- a. 막대 그래프
- b. 선 그래프
- c. 산포도

1. 시각화 : 독자를 설득하는 디자인

참고

- 
- <https://www.holistics.io/blog/data-ink-ratio/>
 - <https://daydreamingnumbers.com/blog/gestalt-laws-data-visualization/>