

데이터 시각화

13. 그래프 디자인의 기본원칙(3)

최대영 교수



고려사이버대학교
THE CYBER UNIVERSITY OF KOREA



Data Visualization

데이터 시각화

13주차

그래프 디자인의 기본 원칙(3)

최대영 교수



1

● 학습리뷰

1 도표 설명 요소와 표

☑ 중복 코딩(Redundant coding)

- 정보의 전달이 아닌 도표를 보기 좋게 꾸미기 위한 용도로 시각적 요소를 사용
- 데이터에 따라 중복 코딩을 이용하면 더 명확한 시각화가 가능

☑ 색과 모양을 이용하는 산점도

- 범주의 구별이 명확하도록 색과 모양을 선정하여 중복 코딩

☑ 레이블 달기

- 범례 대신 선 옆에 레이블을 표시하여 간결하게 표현

2

1 도표 설명 요소와 표

✍ 도표 제목의 위치

- 매체에 따라 적절한 제목의 위치를 선정(캡션 내, 도표 안)

✍ 축 레이블과 범례 제목

- 명시적으로 적지 않더라도 맥락 정보로 파악할 수 있는 경우에는 축 레이블과 범례 제목을 생략하여 간명하게 표현

✍ 표 레이아웃의 핵심원칙

- 세로 선은 넣지 않고 가로 선은 제한적으로 사용
- 데이터 유형에 따라 정렬 방법을 다르게 적용

2 면분할과 복합형 도표

✍ 면분할(Faceting)

- 면분할은 논리적으로 타당하게 배열
- 축의 레이블과 범위는 통일되게 설정

✍ 복합형 도표

- 도표를 조합하여 연관되어 있는 정보를 전달
- 도표의 색상과 위치를 일관성 있게 적용하고 정렬

● 학습목표

● 데이터에 맞는 맥락 요소의 사용에 대해 설명할 수 있다.

● 피해야 할 시각화 방법에 대해 설명할 수 있다.

● 데이터의 맥락 정보 관련 matplotlib 라이브러리를 이해하고 활용할 수 있다.

5

● 학습내용

1 데이터와 맥락 요소

2 피해야 할 시각화 방법

3 실습

6



데이터와 맥락 요소

7

1. 적절한 수준의 맥락 요소 활용

≡ 데이터-잉크 비율(Data-Ink Ratio)

🔗 가능한한 적은 양의 픽셀로 가능한한 많은 양의 데이터를 표현

$$\text{데이터 잉크 비율} = \frac{\text{시각화한 데이터}}{\text{사용된 픽셀수}} \rightarrow \text{최대화}$$

🔗 데이터-잉크 비율은 타당한 범위 안에서 최대화 되어야 함

- 비 데이터 잉크를 무조건 줄이는 방식이 아님
- 지나치게 비 데이터 잉크를 줄이면 조악해짐

8

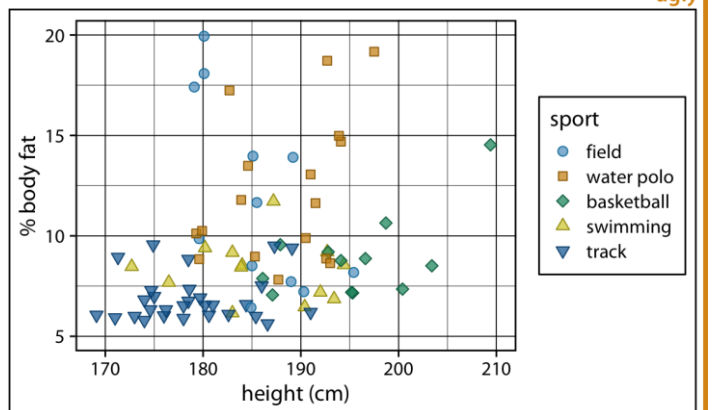
1. 적절한 수준의 맥락 요소 활용

산점도의 맥락 요소

불필요한 테두리와 격자를 과도하게 사용

→ 맥락을 나타내는 주변 요소로 인해 실제 데이터 포인트로 전하려는 메시지가 약해짐

호주 출신 남성 운동선수들의 체지방률과 신장



[출처] Fundamentals of Data Visualization

9

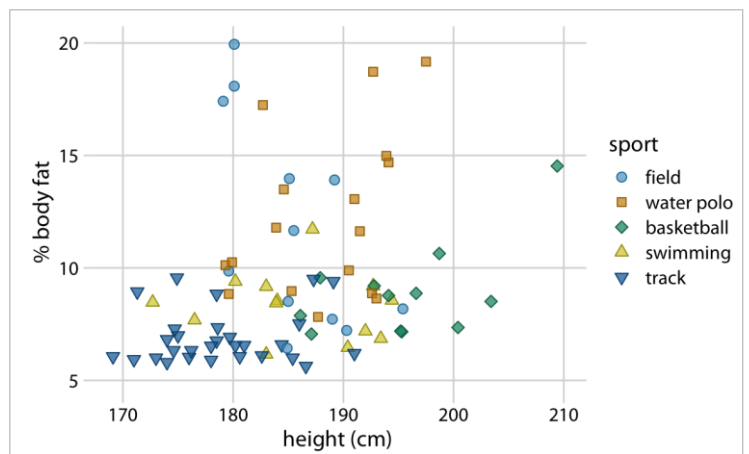
1. 적절한 수준의 맥락 요소 활용

산점도의 맥락 요소

산점도의 테두리와 격자를 연회색으로 표현

→ 실제 데이터 포인트를 훨씬 명료하게 드러냄

호주 출신 남성 운동선수들의 체지방률과 신장



[출처] Fundamentals of Data Visualization

10

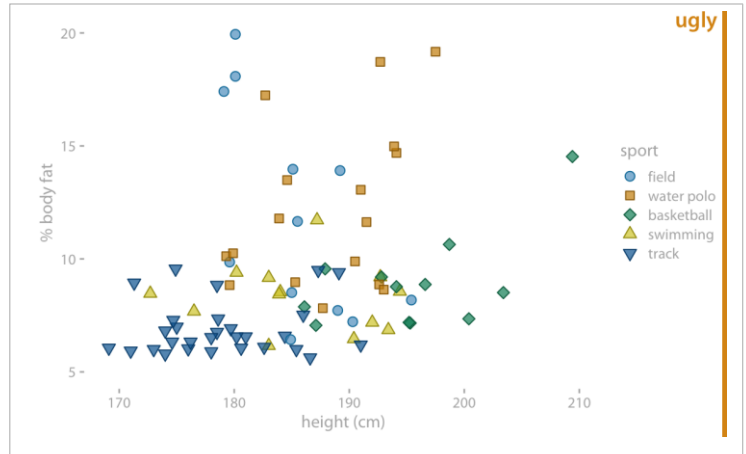
1. 적절한 수준의 맥락 요소 활용

산점도의 맥락 요소

✎ 비 데이터 잉크를 줄이기 위해
맥락 요소를 최소화

→ 레이블과 제목이 흐려서
알아보기 어렵고 범례와
도표의 경계가 없음

호주 출신 남성 운동선수들의 체지방률과 신장



[출처] Fundamentals of Data Visualization

11

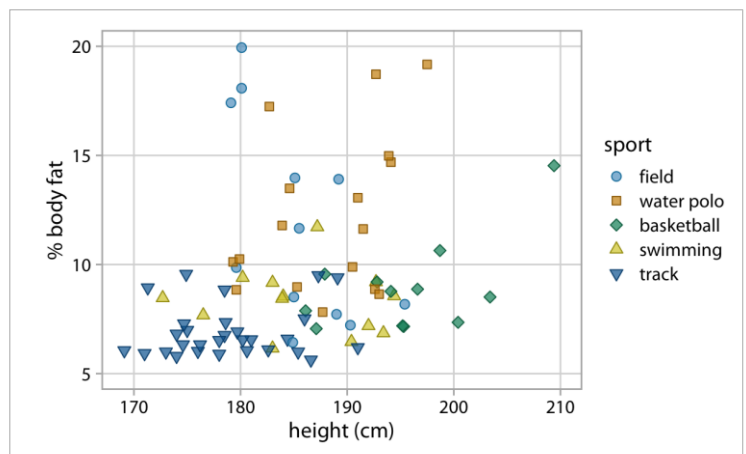
1. 적절한 수준의 맥락 요소 활용

산점도의 맥락 요소

✎ 그래프 패널에 테두리를 넣어
범례와 데이터를 구분

→ 범례와 데이터를 명확하게
구분할 수 있음, 패널 테두리의
사용여부는 선호의 차이

호주 출신 남성 운동선수들의 체지방률과 신장



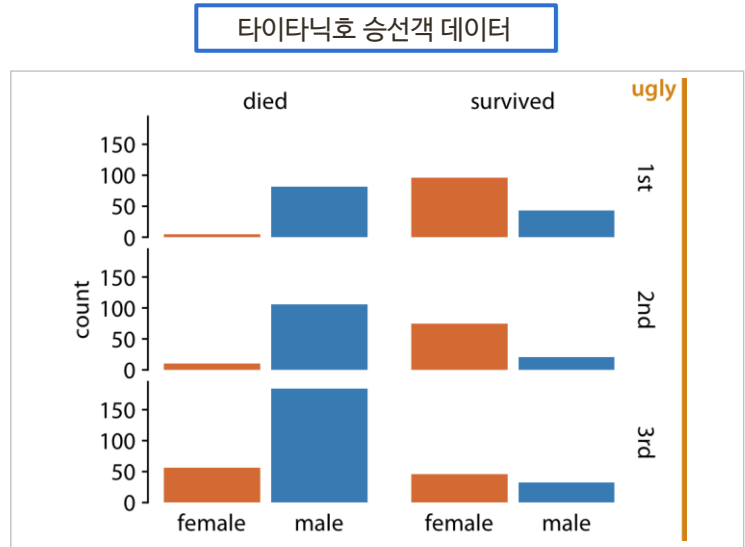
[출처] Fundamentals of Data Visualization

12

1. 적절한 수준의 맥락 요소 활용

다중패널 도표의 맥락 요소

- 비 데이터 잉크를 최소화
→ 기준선이 없어 패널의 영역을 구분하기 어려움



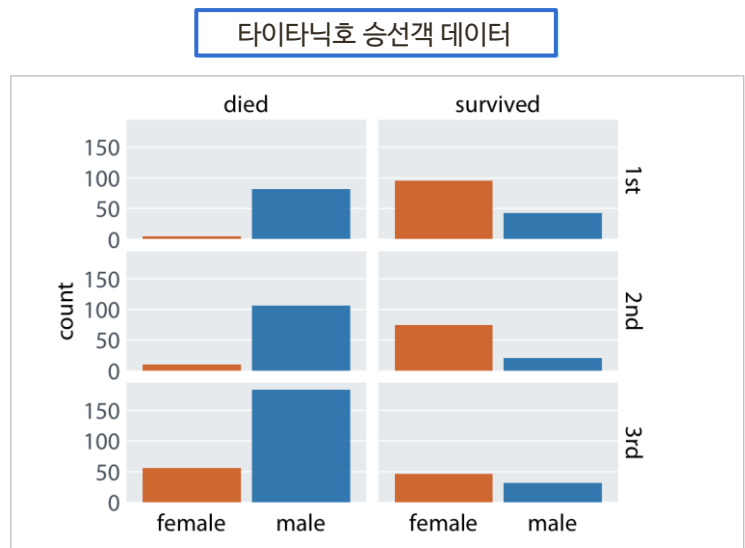
[출처] Fundamentals of Data Visualization

13

1. 적절한 수준의 맥락 요소 활용

다중패널 도표의 맥락 요소

- 배경을 깔고 가는 선으로 기준선을 표현
→ 패널의 구분이 명확하고 막대의 값을 확인하기 쉬움



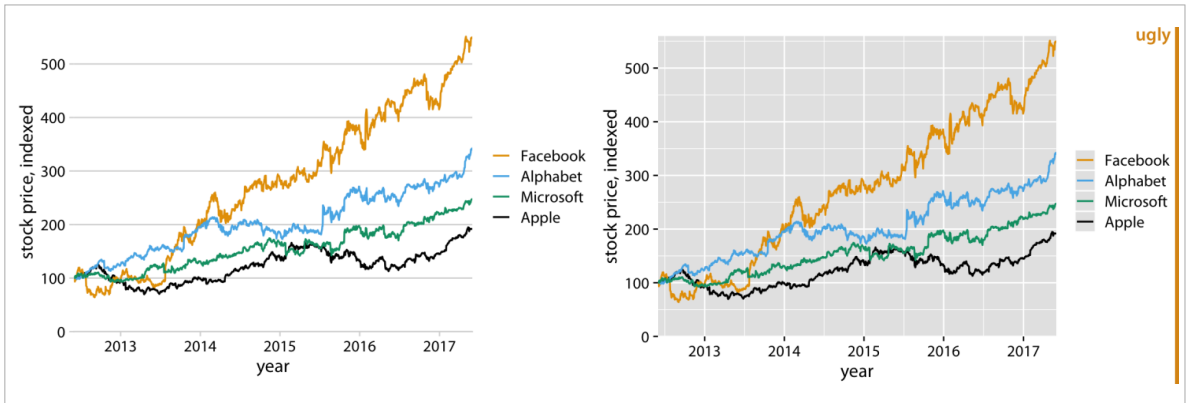
[출처] Fundamentals of Data Visualization

14

2. 배경 격자와 사선

≡ 배경 격자의 역할

- 특정 데이터 값을 파악하거나 여러 데이터 값을 비교하는 용도로 활용
 - 진하거나 촘촘한 격자선으로 인해 오히려 도표가 조잡해 보일 수 있음



[출처] Fundamentals of Data Visualization

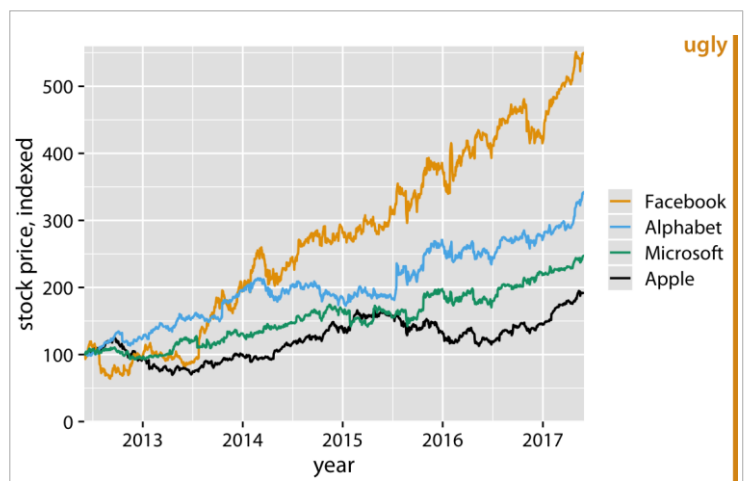
15

2. 배경 격자와 사선

≡ 색이 있는 배경 격자

- 회색 배경에 굵은 격자선과 가는 격자선을 교차
 - 격자선을 데이터 선 만큼 두드러지게 표현해 데이터를 제대로 강조하지 못함

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

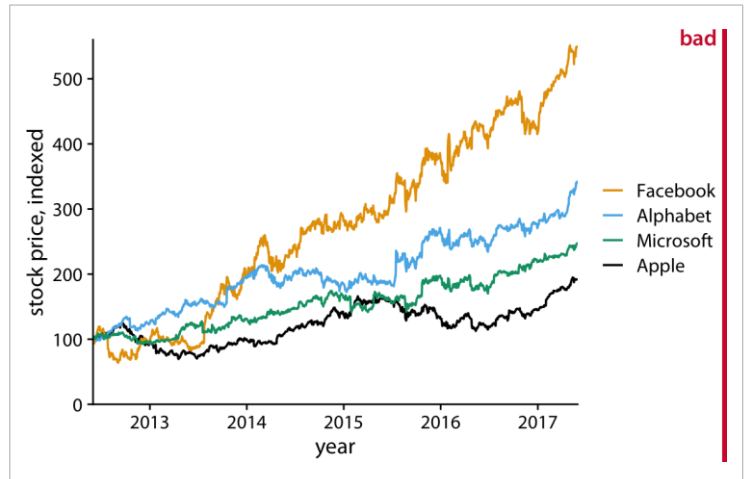
16

2. 배경 격자와 사선

기준선이 없는 도표

- 주가의 출발점을 표시하는 기준선이 없음
- 기준이 되는 지수값 100으로 부터 얼마나 변동되었는지 알기 어려움

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

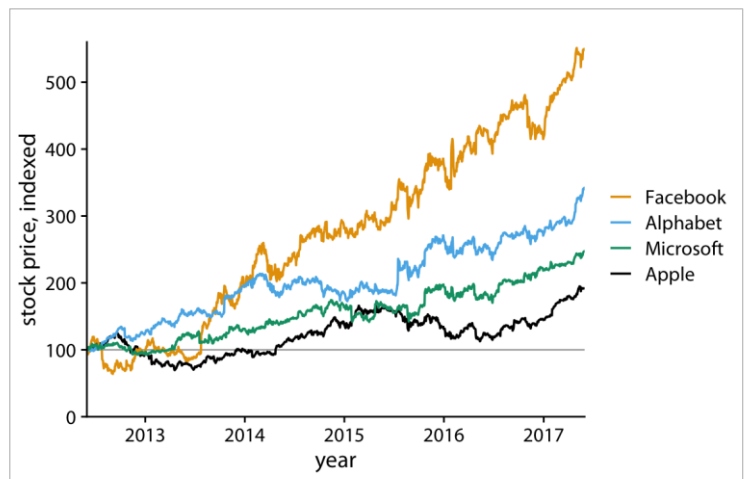
17

2. 배경 격자와 사선

기준선이 있는 도표

- 기준이 되는 수치에 선을 표현
- 데이터 값을 식별하기 쉬움

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

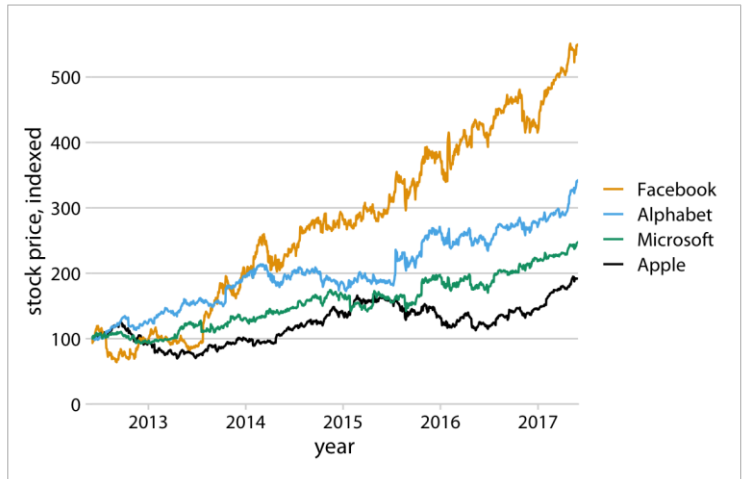
18

2. 배경 격자와 사선

≡ 배경 격자가 적절한 도표

- ☞ 눈금마다 격자선을 표시하고 세로선을 삭제
→ 데이터 값을 식별하기 쉽고 간결함

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

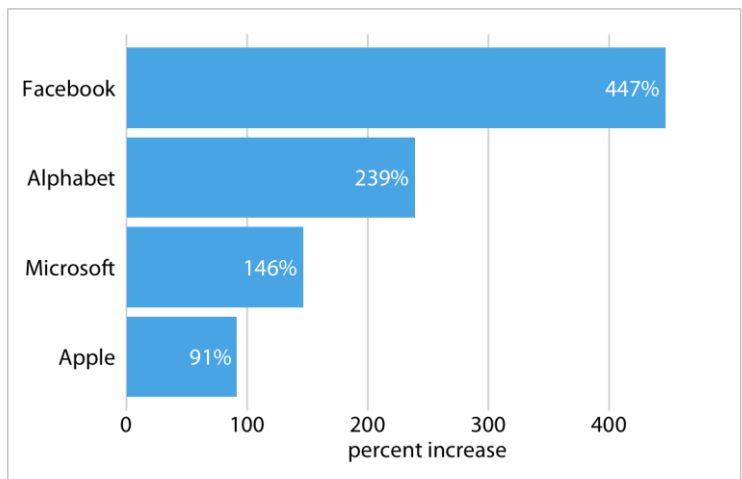
19

2. 배경 격자와 사선

≡ 막대 도표의 격자선

- ☞ 가로 막대의 격자선은 세로로 표시 → 보통 격자선이 막대 위로 지나가도록 표현

4개 기업의 주가 상승률



[출처] Fundamentals of Data Visualization

20

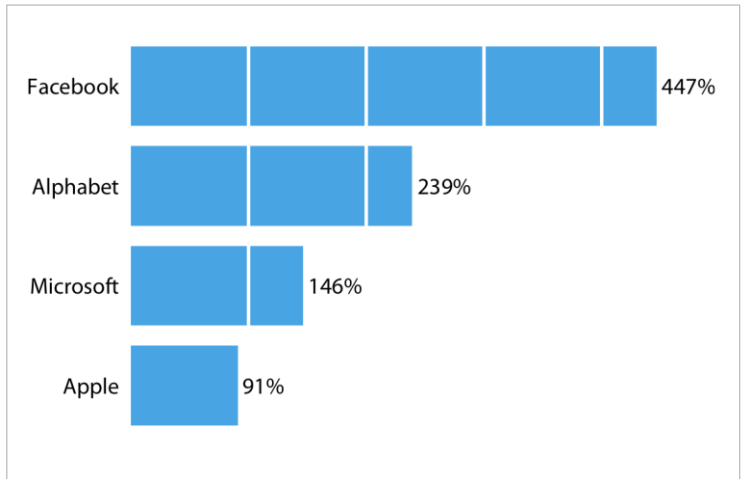
2. 배경 격자와 사선

≡ 막대 도표의 격자선

✎ 배경색과 같은 흰색으로 격자선을 표시

→ 격자선에 의해 조각이 등분되는 효과, 막대 길이의 인지가 쉽지만 시각적으로 덩어리를 이루지 못함

4개 기업의 추가 상승률



[출처] Fundamentals of Data Visualization

21

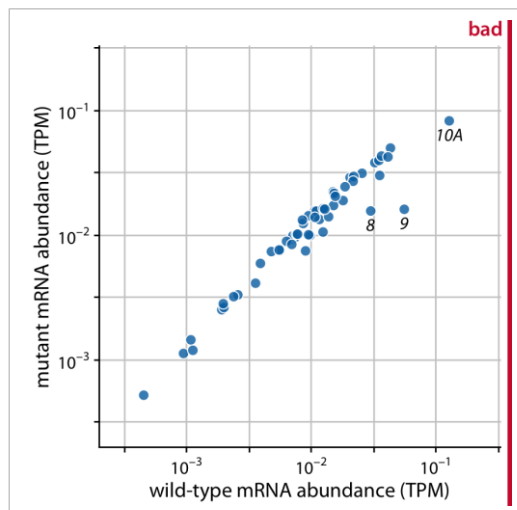
2. 배경 격자와 사선

≡ 격자가 있는 쌍 데이터

✎ 격자를 기준선으로 표현

→ 쌍 데이터의 데이터를 비교하기 어려움

유전자 발현수준 비교



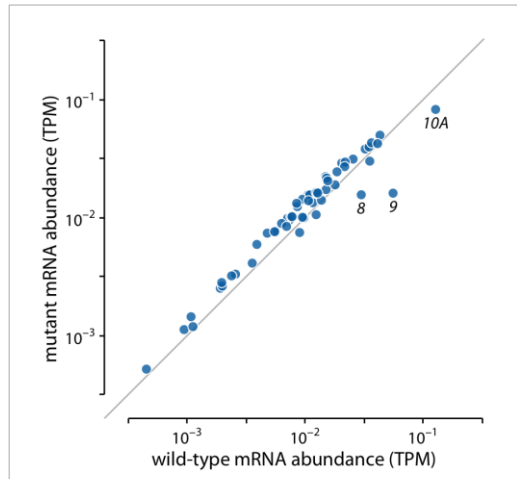
22

2. 배경 격자와 사선

≡ 쌍 데이터의 사선

- ☞ 사선을 기준으로 표현
→ 쌍 데이터의 비교에 적합

유전자 발현수준 비교



[출처] Fundamentals of Data Visualization

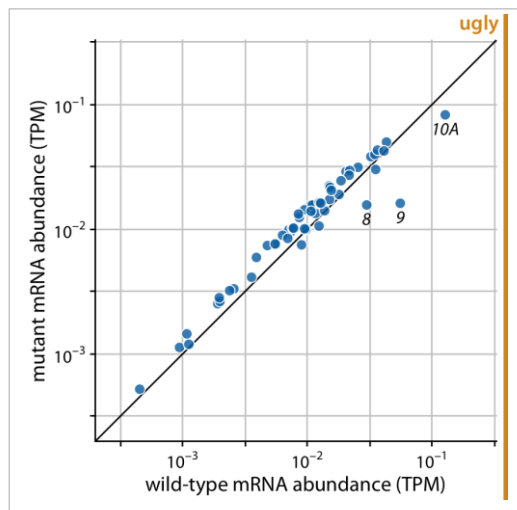
23

2. 배경 격자와 사선

≡ 사선과 격자가 모두 있는 쌍 데이터

- ☞ 사선과 격자를 기준으로 표현
→ 쌍 데이터의 비교는 가능하나
도표가 너무 복잡해 보임

유전자 발현수준 비교



24



피해야 할 시각화 방법

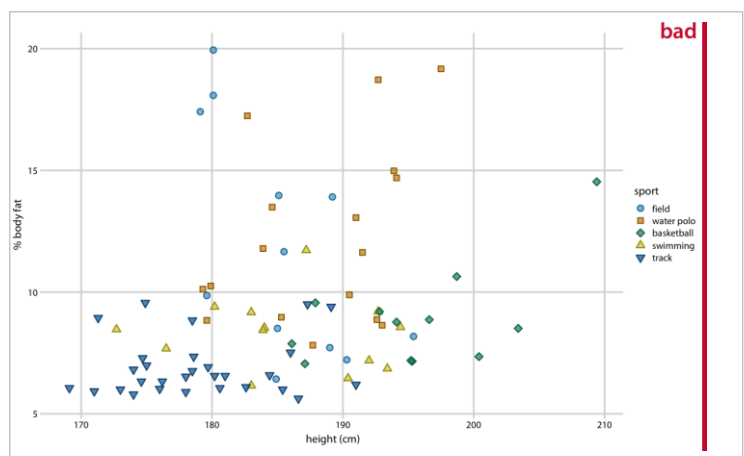
25

1. 도표 글자 크기

너무 작은 글자

- 축과 범례의 글자가 너무 작음
→ 읽기 어려워 독자에게 불필요한 노력을 요구

호주 출신 남성 운동선수들의 체지방률과 신장



[출처] Fundamentals of Data Visualization

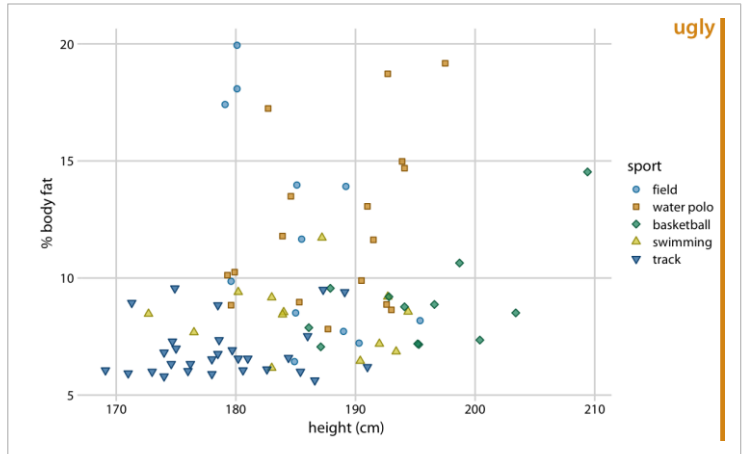
26

1. 도표 글자 크기

시각적 요소들의 크기 불균형

- 축과 범례의 글자가 커졌으나
전체적으로 도표의 균형이
맞지 않음
→ 정확한 시각적 인식을 위한
개선이 필요

호주 출신 남성 운동선수들의 체지방률과 신장



[출처] Fundamentals of Data Visualization

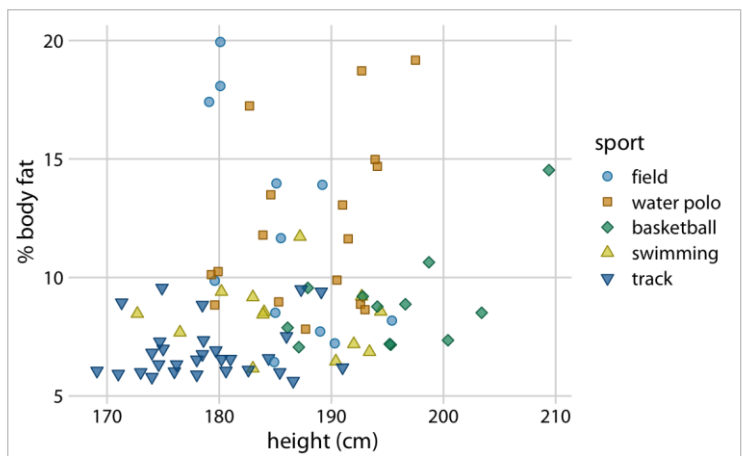
27

1. 도표 글자 크기

균형 있는 도표

- 글자와 함께 점의 크기를 키움
→ 도표와 시각적 요소들의
크기가 전체적으로 조화로움

호주 출신 남성 운동선수들의 체지방률과 신장



[출처] Fundamentals of Data Visualization

28

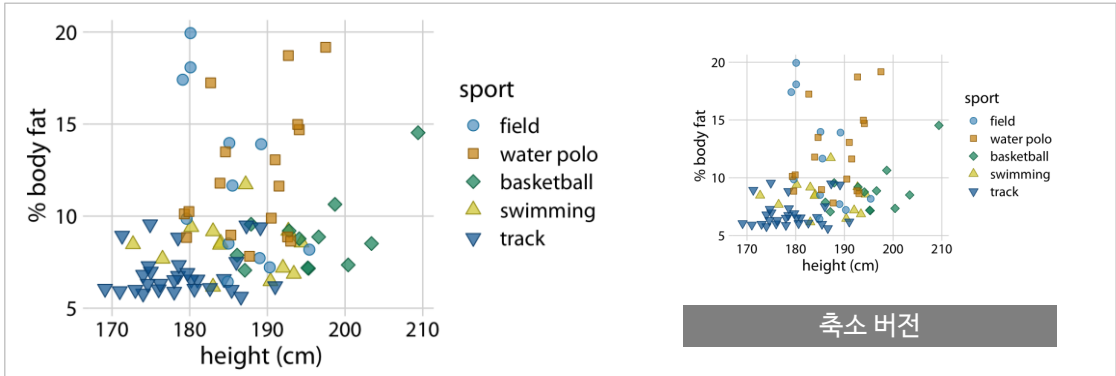
1. 도표 글자 크기

축소해도 문제가 없는 도표

모든 요소들의 크기를 키움

→ 도표를 축소해도 시각적 요소가 충분히 커서 정보의 전달에 쉬움

호주 출신 남성 운동선수들의 체지방률과 신장



[출처] Fundamentals of Data Visualization

29

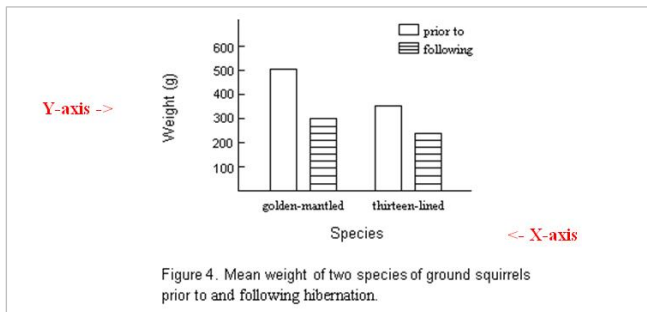
2. 선 그림

선 그림의 역사

선 그림은 20세기 과학 분야에서 손으로 직접 도표를 그려 시각화하고 흑백으로 복사했던 것에서 시작

선 그림 안을 선을 긋거나 점을 찍는 방법으로 면을 채움

초기 도표 제작 SW는 선 그림을 본 따고 확장하는 경우가 많았음



[출처] Acadia University

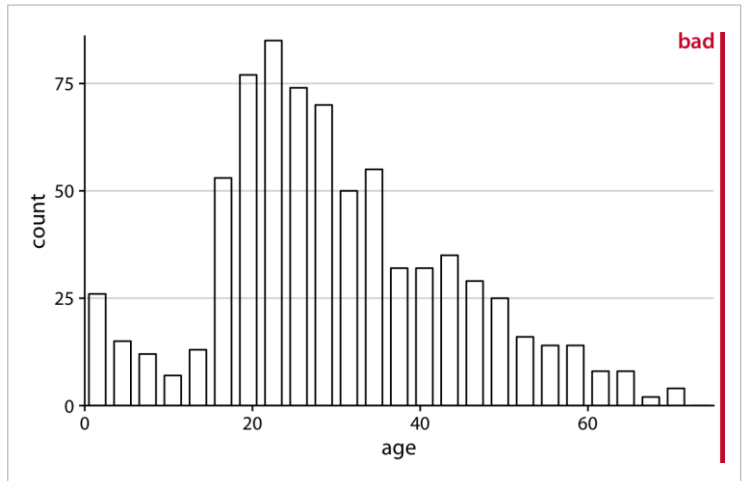
30

2. 선 그림

히스토그램

- ☞ 윤곽선 만으로 막대를 표현
→ 막대 안팎을 한눈에 구분하기 힘들고 막대들 사이의 틈으로 인해 도표의 의미를 해석하는데 오래 걸림

타이타닉호 승선객의 연령



[출처] Fundamentals of Data Visualization

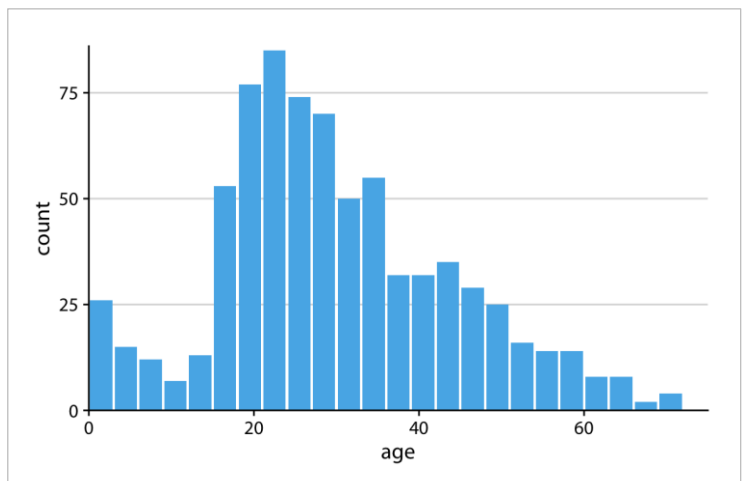
31

2. 선 그림

히스토그램

- ☞ 윤곽선 안을 연한 색으로 채움
→ 배경과 막대의 구분이 명확해 짐

타이타닉호 승선객의 연령



[출처] Fundamentals of Data Visualization

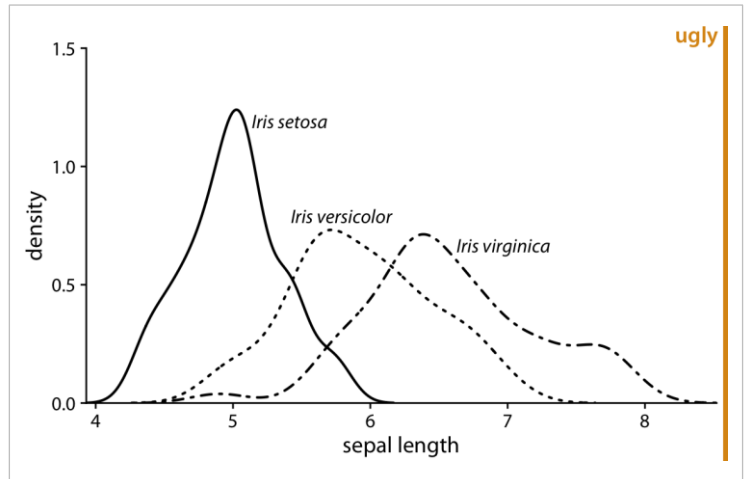
32

2. 선 그림

밀도 도표

- ☞ 품종을 구분하기 위해 선의 스타일을 다르게 적용
→ 점선과 쇄선의 위아래 영역의 구분이 모호함

3가지 품종 붓꽃의 꽃받침 길이



[출처] Fundamentals of Data Visualization

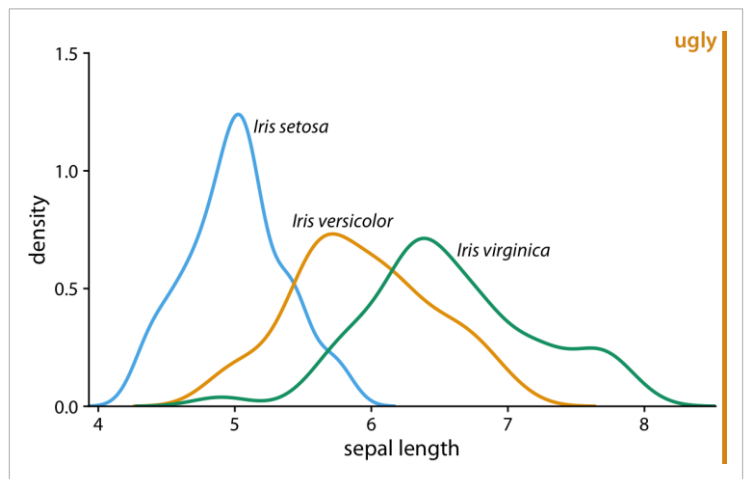
33

2. 선 그림

밀도 도표

- ☞ 품종을 구분하기 위해 선의 색상을 다르게 적용
→ 선 사이가 벌어진 문제는 해결되었지만 면적이 명확하게 구분되지는 않음

3가지 품종 붓꽃의 꽃받침 길이



[출처] Fundamentals of Data Visualization

34

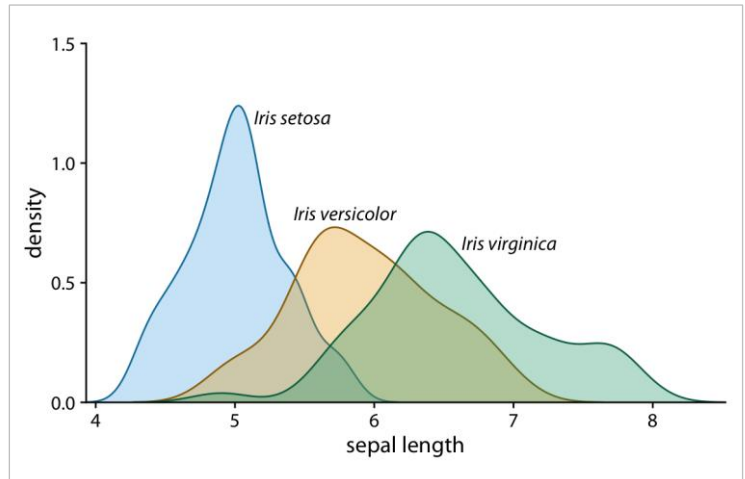
2. 선 그림

밀도 도표

선 아래 영역에 반투명한 색을 채움

→ 영역별로 구분이 명확해져서 각 품종의 분포를 파악하기 쉬움

3가지 품종 붓꽃의 꽃받침 길이



[출처] Fundamentals of Data Visualization

35

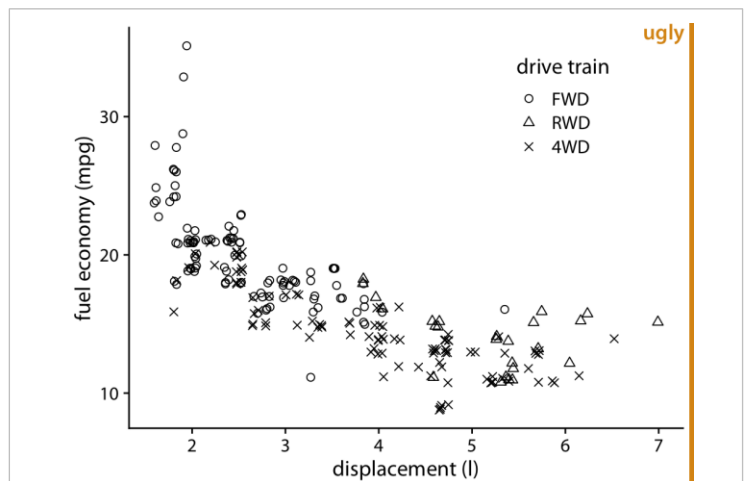
2. 선 그림

산점도

데이터 포인트의 모양으로 범주를 구분

→ 데이터 포인트가 겹쳐져 있는 부분에 대한 해석이 어렵고 도표가 번잡해 보임

인기차종 234종의 도시 주행 연비와 엔진 배기량



[출처] Fundamentals of Data Visualization

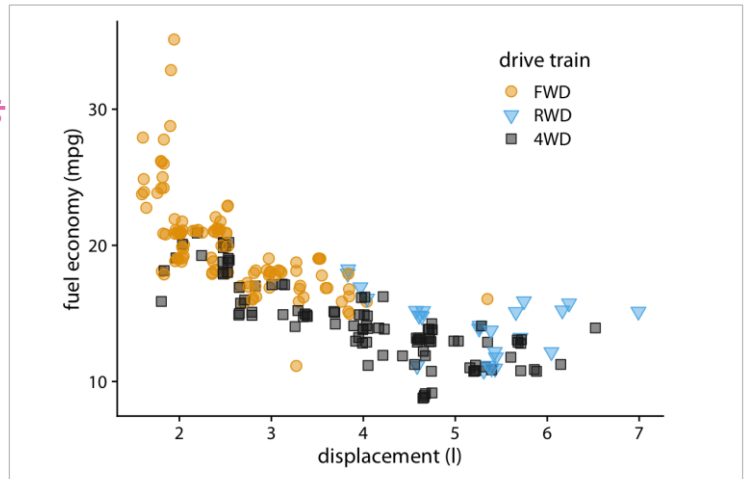
36

2. 선 그림

산점도

- 데이터 포인트 모양을 변경하고 색을 채움
- 범주 구분이 명확하고 흑백으로 출력해도 구분 가능

인기차종 234종의 도시 주행 연비와 엔진 배기량



[출처] Fundamentals of Data Visualization

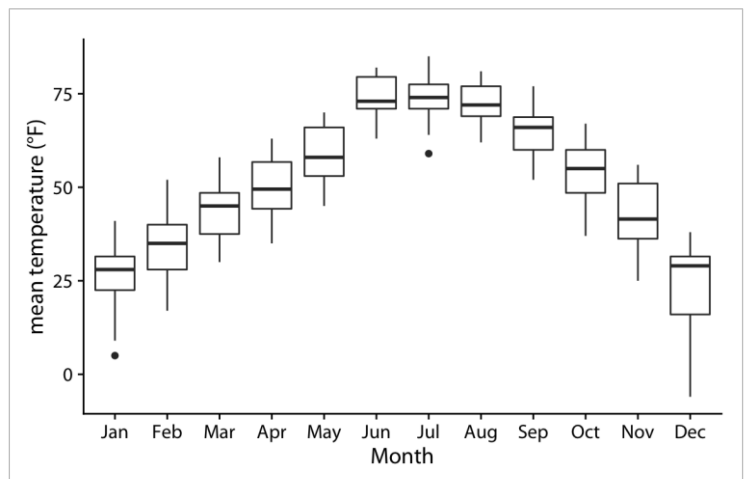
37

2. 선 그림

박스 도표

- 윤곽선만으로 박스를 표현
- 박스와 선이 많아지면 박스 안 밖의 구분이 모호

네브레스카 링컨시의 일평균 기온



[출처] Fundamentals of Data Visualization

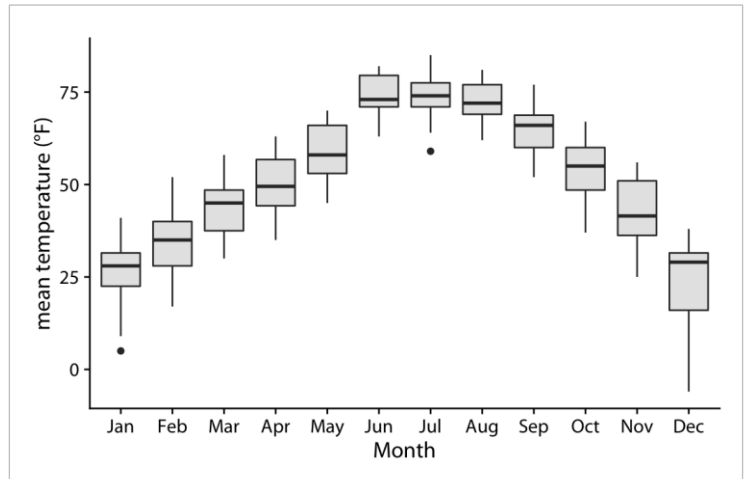
38

2. 선 그림

☰ 박스 도표

- ☞ 박스 안에 연한 색을 채움
→ 박스와 배경이 명확하게 구분됨

네브레스카 링컨시의 일평균 기온



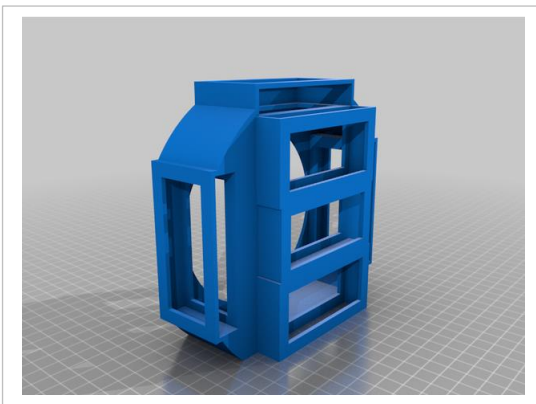
[출처] Fundamentals of Data Visualization

39

3. 3차원 그래픽과 차트

☰ 3차원 차트의 문제점

- ☞ 인간의 시각계는 3차원 이미지를 2차원 평면에 투사한 결과물을 다시 3차원 공간에 투영하여 데이터의 왜곡을 보정



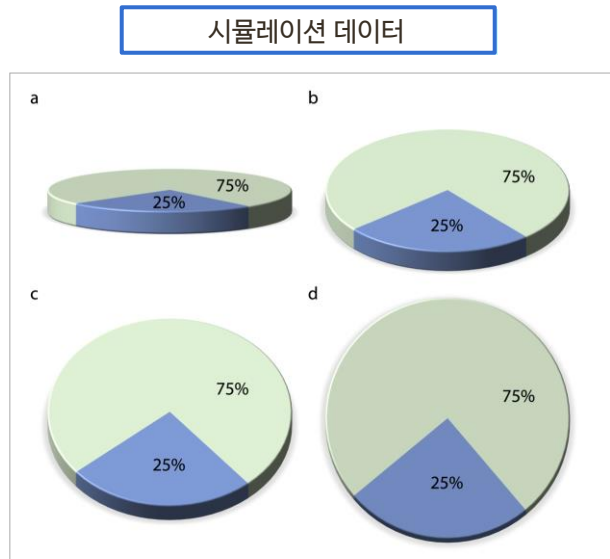
[출처] Thingiverse

40

3. 3차원 그래픽과 차트

3차원 파이 차트

- ✎ 파이 차트를 3차원으로 만들고 보이는 각도를 다르게 하여 표현
→ 실제 데이터의 비율과 조각의 면적이 일치하지 않음



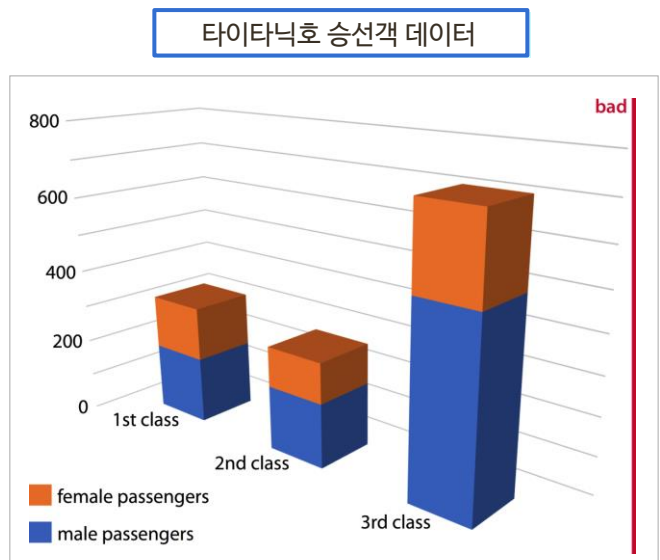
[출처] Fundamentals of Data Visualization

41

3. 3차원 그래픽과 차트

3차원 막대 도표

- ✎ 막대를 3차원으로 만들고 두 벽면으로 부터 떨어지게 위치하여 표현
→ 실제 데이터 보다 막대가 표현하는 값이 더 작게 표현됨



[출처] Fundamentals of Data Visualization

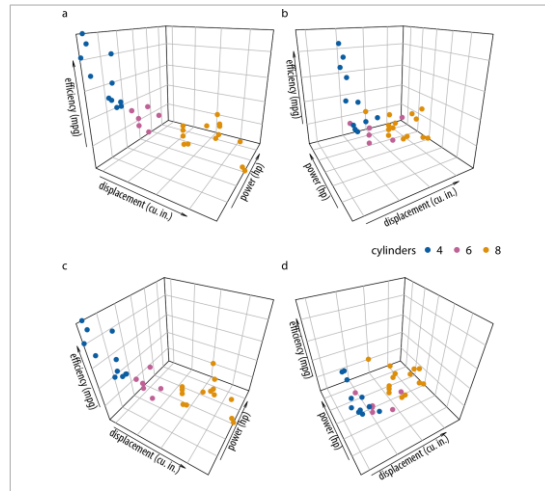
42

3. 3차원 그래픽과 차트

3차원 산점도

- 동일한 데이터를 다른 각도의 3차원 산점도로 표현
→ 점들이 실제로 공간상에 어떻게 분포되어 있는지 상상하기 어려움

32대 차량의 연비, 배기량, 마력 비교



[출처] Fundamentals of Data Visualization

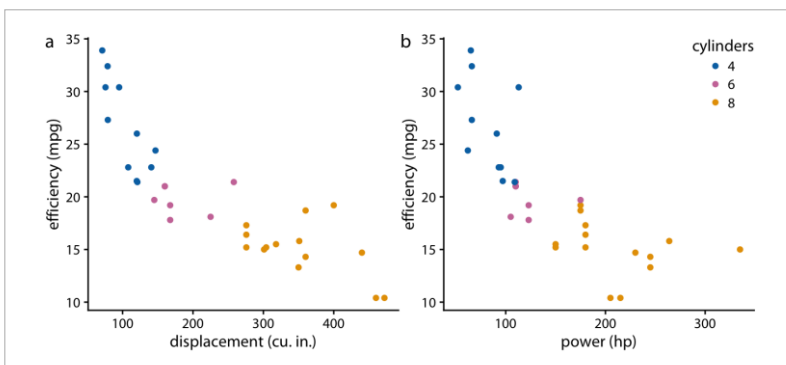
43

3. 3차원 그래픽과 차트

3차원 산점도를 보완하는 방법

- 변수를 나눠 2개의 2차원 산점도로 표현
→ 3차원 산점도로 표현한 3개의 변수 중에 하나가 중요한 경우에 유용

32대 차량의 연비, 배기량, 마력 비교



[출처] Fundamentals of Data Visualization

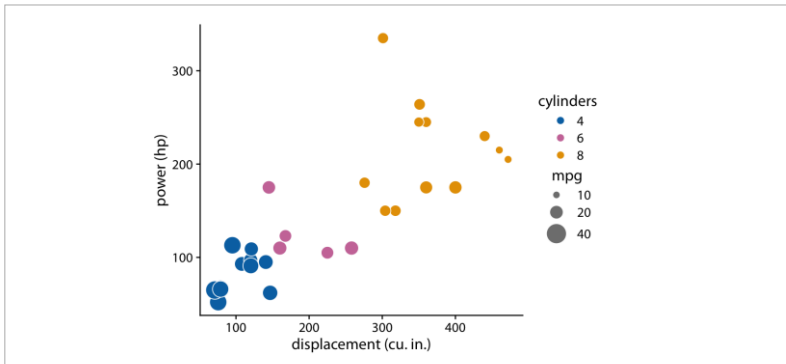
44

3. 3차원 그래픽과 차트

3차원 산점도를 보완하는 방법

- 2차원 산점도로 나타내고 점의 크기를 이용하여 나머지 변수를 표현
→ 3개의 변수 중 하나가 부차적일 때 유용

32대 차량의 연비, 배기량, 마력 비교



[출처] Fundamentals of Data Visualization

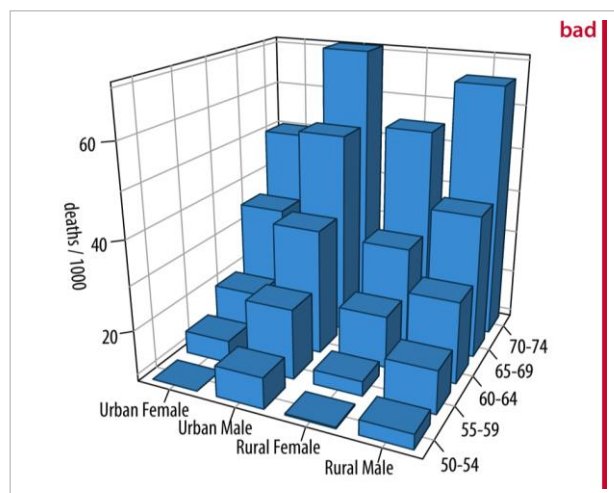
45

3. 3차원 그래픽과 차트

3차원 막대 도표

- 3차원 막대로 사망률을 표현
→ 막대가 가려져 있지 않아
전체적인 해석에는 문제가
없으나 막대들의 직접 비교는
여전히 제한

1940년 버지니아주의 사망률



[출처] Fundamentals of Data Visualization

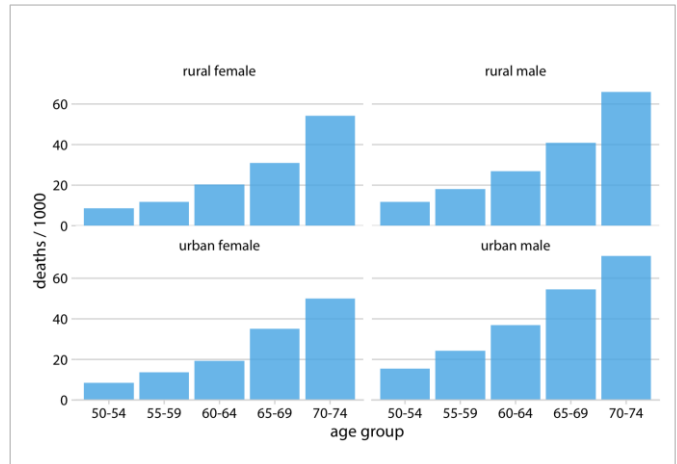
46

3. 3차원 그래픽과 차트

3차원 막대 도표를 보완하는 방법

- ☞ 변수를 나눠 4개의 막대 도표로
표현 → 막대 도표들 사이의
정확한 비교가 가능

1940년 버지니아주의 사망률



[출처] Fundamentals of Data Visualization

47

3. 3차원 그래픽과 차트


3차원 시각화가 적절한 경우

- 01 사용자가 도표를 회전할 수 있는 상호작용이 있는 시각화
- 02 도표가 천천히 회전하여 3차원 공간에서 재구성이 가능한 경우
- 03 실제 3차원 객체나 3차원 공간에 놓인 데이터를 보여주는 경우



[출처] Fundamentals of Data Visualization

48



03

chapter

실습

49

1. 맥락 요소의 활용

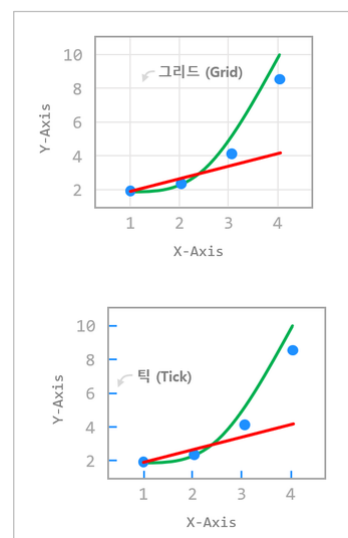
≡ 배경 격자와 눈금

📌 배경 격자

- `grid(color='red', alpha=0.5, linestyle='--')`

📌 눈금

- `xticks([0, 1, 2])`
- `yticks(np.arange(1, 6))`



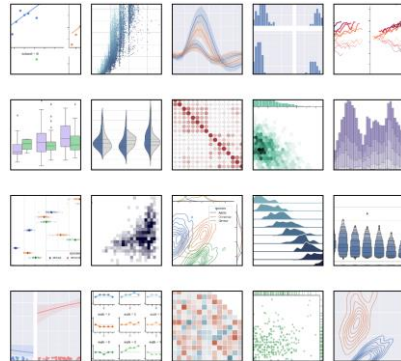
[출처] Matplotlib Tutorial - 파이썬으로 데이터 시각화하기

50

2. 시각화 라이브러리

seaborn

- ✍ matplotlib의 기능과 스타일을 확장한 라이브러리
- ✍ 단순한 코딩스타일로 초보자도 쉽게 사용 가능
- ✍ 기본설정에서 화려한 디자인과 색감 제공



[출처] seaborn

51

● 학습정리

1 데이터와 맥락 요소

✍ 데이터-잉크 비율(Data-Ink Ratio)

- 가능한한 적은 양의 픽셀로 가능한한 많은 양의 데이터를 표현
- 데이터-잉크 비율은 타당한 범위 안에서 최대화 되어야 함

✍ 적절한 수준의 맥락 요소 활용

- 테두리와 배경 격자는 데이터 요소를 가리지 않게 사용
- 쌍 데이터에는 격자 대신 사선을 사용

52

● 학습정리

2 피해야 할 시각화 방법

✍ 도표 글자 크기

- 도표의 글자는 너무 작지 않게 설정
- 도표 글자, 점과 선 등이 전체적으로 조화롭게 크기 설정

✍ 선 그림

- 선 그림은 배경과의 구분이 모호할 수 있어 피하고 색을 채움

✍ 3차원 그래픽과 차트

- 3차원은 데이터 값을 왜곡할 수 있어 가능한한 2차원으로 표현

53

● 참고문헌

📁 「데이터 시각화 교과서」, Claus O. Wilke, 책만, 2020.

📁 「Fundamentals of Data Visualization」, Claus O. Wilke, O'Reilly Media, 2019.

※ 서체 출처 | 넥슨Lv2고딕-(넥슨코리아)www.levelup.nexon.com / 나눔바른고딕(네이버)

54

저작권 안내

이 강의록은 저작권법에 의해 보호받는 저작물로서
저작권자의 허락 없이 저작재산권 일체(복제권,
배포권, 대여권, 공연권, 공중전송권, 전시권, 2차적
저작물 작성권)를 침해 시 저작권법에 의거 처벌받을
수 있습니다.