

데이터 시각화

12. 그래프 디자인의 기본원칙(2)

최대영 교수



고려사이버대학교
THE CYBER UNIVERSITY OF KOREA



Data Visualization

데이터 시각화

12주차

그래프 디자인의 기본 원칙(2)

최대영 교수



1

학습리뷰

1 잉크 양 비례의 원칙

☞ 잉크 양 비례의 원칙 (Principle of proportional ink)

- 음영 영역의 크기는 해당 영역이 나타내는 데이터 값에 비례해야 함

☞ 축이 있는 도표

- 잉크 양 비례의 원칙에 부합하기 위해 막대나 선은 축의 0부터 표현해야 함

☞ 면적을 나타내는 도표

- 파이 조각의 면적은 각도에 비례, 각도는 데이터 값에 비례
- 인간의 인지가 면적보다는 길이를 먼저 판단하기 때문에 막대 도표가 파이 차트와 트리맵 보다 데이터 값의 차이를 더 두드러지게 나타냄

2

2 오버플로팅과 색

📌 오버플로팅(Overplotting)

- 여러 개의 점이 같은 좌표에 겹쳐서 표현되어 있는 현상

📌 반투명 색 사용

- 점마다 반투명 색을 이용하여 데이터 값의 차이를 표현

📌 지터링(Jittering)

- 데이터를 왜곡하지 않는 한에서 점마다 약간씩 위치를 옮겨 표현

2 오버플로팅과 색

📌 2차원 히스토그램

- 2차원 평면을 사각형 등으로 나눠 데이터의 수에 따라 색으로 표현

📌 등고선 그래프(Contour lines)

- 도표 전체의 점 밀도를 추정하고 그에 따라 등고선을 그리는 방법

2 오버플로팅과 색

📌 색상 스케일의 수와 의미

- 3~5개의 범주를 색으로 나타낼 때 효과적
- 색이 의미하는 바를 명확히 하고 채도를 조절하여 피로를 낮춰야 함

📌 색각 이상자(color-vision deficiency)에 대한 고려

- 색각 이상자도 구별할 수 있는 색상 스케일로 인코딩
- Color Brewer 등 도구를 이용해 시각화에 쓰일 색상을 선정

● 학습목표

📌 도표 설명 요소와 표의 적절한 사용에 대해 설명할 수 있다.

📌 면분할과 복합형 도표의 적절한 사용에 대해 설명할 수 있다.

📌 도표 설명 요소 관련 matplotlib 라이브러리를 이해하고 활용할 수 있다.

● 학습내용

1 도표 설명 요소와 표

2 면분할과 복합형 도표

3 실습



7



도표 설명 요소와 표

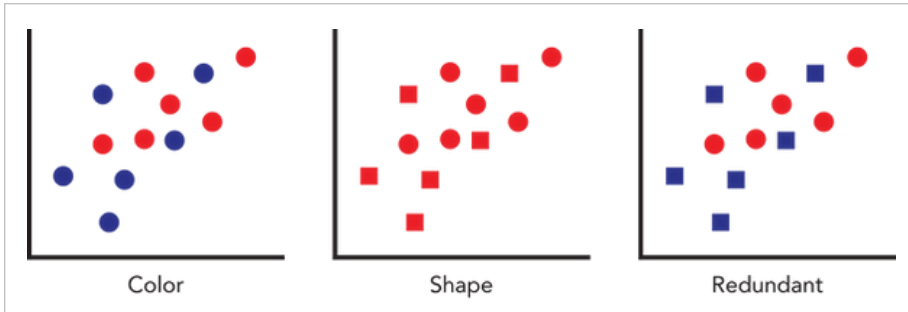
8

1. 중복 코딩과 범례

중복 코딩(Redundant coding)

☞ 정보의 전달이 아닌 도표를 보기 좋게 꾸미기 위한 용도로 시각적 요소를 사용

- 장식 목적의 요소를 많이 넣으면 데이터에 불필요하게 많은 의미가 부여됨



[출처] Nothelfer et al., Redundant Coding Can Speed Up Segmentation in Multiclass Displays, 2016

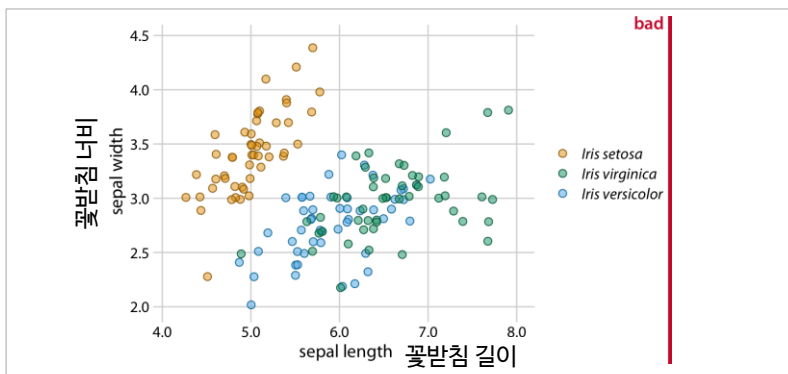
9

1. 중복 코딩과 범례

색을 이용하는 산점도

☞ 3가지 색으로 3종류의 붓꽃을 구별 → virginica와 versicolor 품종의 색이 유사하여 뚜렷하게 대비되지 않음

붓꽃(iris) 데이터



[출처] Fundamentals of Data Visualization

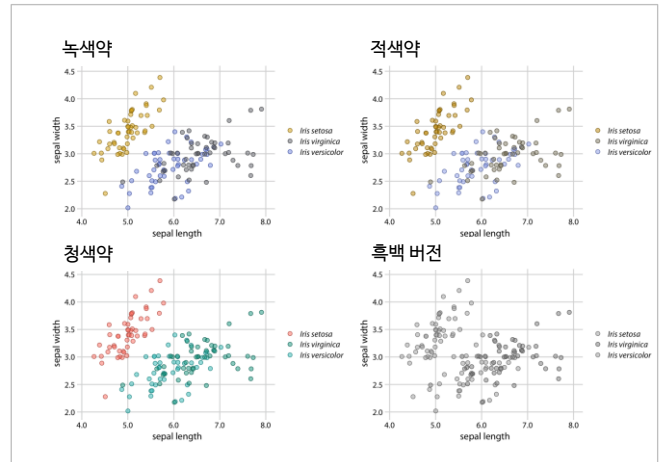
10

1. 중복 코딩과 범례

≡ 색을 이용하는 산점도

- ☞ 3가지 색으로 3종류의 붓꽃을 구별
 - 색깔 이상자도 구분하기 어려움,
 - 채도 값을 제거한 경우에는 3종을 모두 구분하기 어려움

붓꽃(iris) 데이터



[출처] Fundamentals of Data Visualization

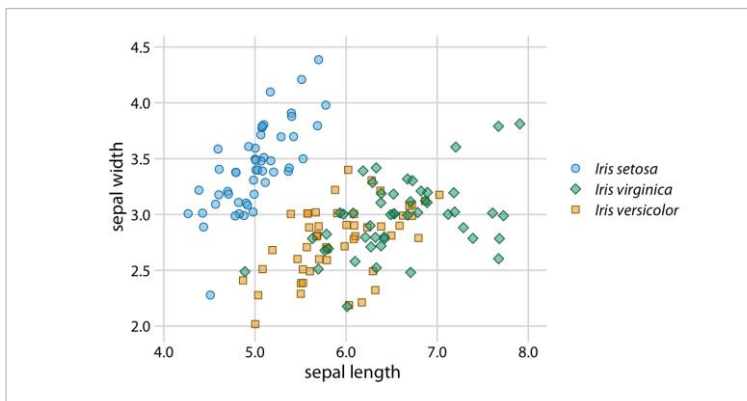
11

1. 중복 코딩과 범례

≡ 모양을 이용하는 산점도

- ☞ 점의 색과 모양으로 3종류의 붓꽃을 구별 → 붓꽃의 품종을 명확하게 구분 가능

붓꽃(iris) 데이터



[출처] Fundamentals of Data Visualization

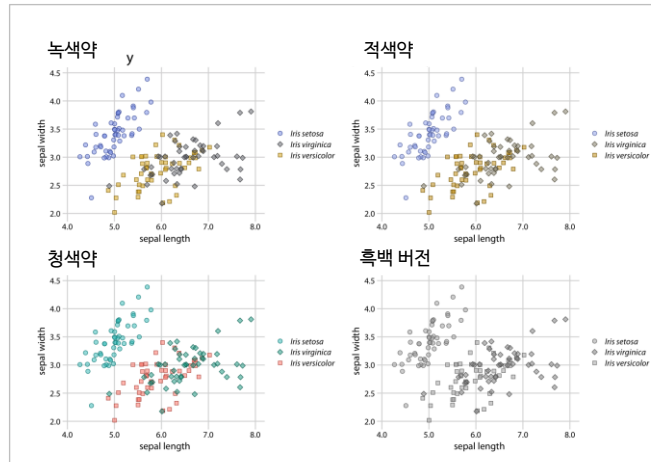
12

1. 중복 코딩과 범례

모양을 이용하는 산점도

- 점의 색과 모양으로 3종류의 붓꽃을 구별
- 색각 이상자도 구분이 가능하고 흑백 버전으로 프린트해도 구분이 가능

붓꽃(iris) 데이터



[출처] Fundamentals of Data Visualization

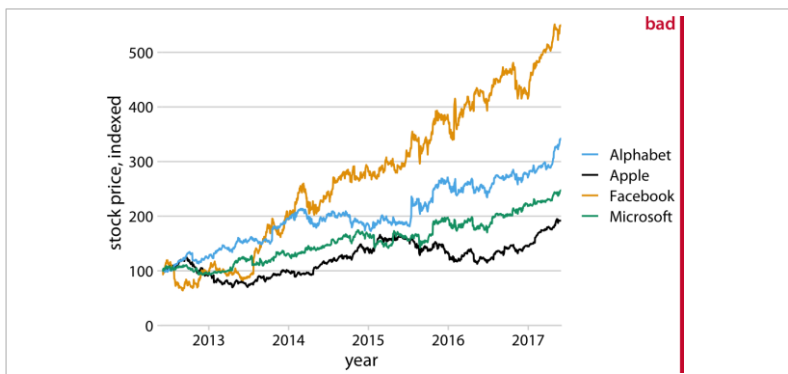
13

1. 중복 코딩과 범례

색을 이용하는 선 그래프

- 4가지 색으로 4개의 기업을 구별 → 선과 범례의 순서가 달라 둘을 비교하는 과정이 불편(시각화 도구들이 자동으로 범례를 지정)

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

14

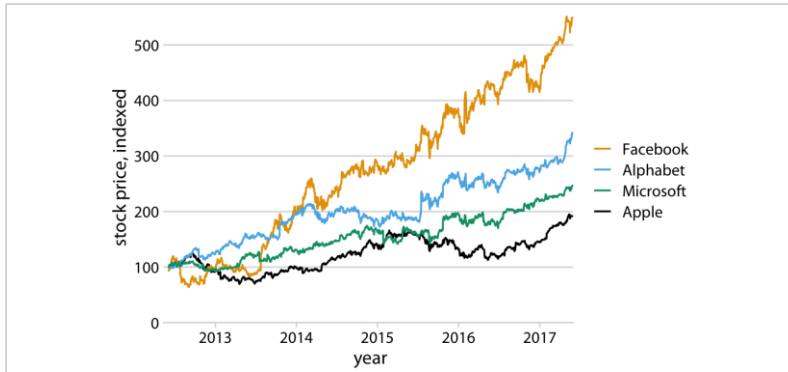
1. 중복 코딩과 범례

≡ 색을 이용하는 선 그래프

📌 4가지 색으로 4개의 기업을 구별

→ 선과 범례의 순서가 일치하고 색약 친화적인 색을 사용하여 구별하기 쉬움

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

15

1. 중복 코딩과 범례

≡ 모양을 이용하는 선 그래프

📌 3가지 선의 모양으로 3개의 범주를 구별

→ 점선이나 색선은 선 그래프를 산만하게 만들 수 있음

시뮬레이션 데이터



[출처] Fundamentals of Data Visualization

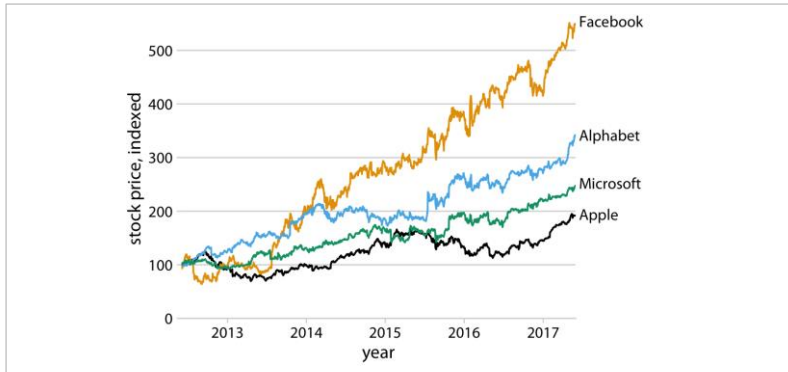
16

1. 중복 코딩과 범례

레이블 달기

☞ 범례 대신 선 옆에 레이블을 표시 → 각 선이 뜻하는 바를 즉각적이고 확실하게 알 수 있음, 가능하면 범례가 없는 도표를 만드는 것이 중요

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

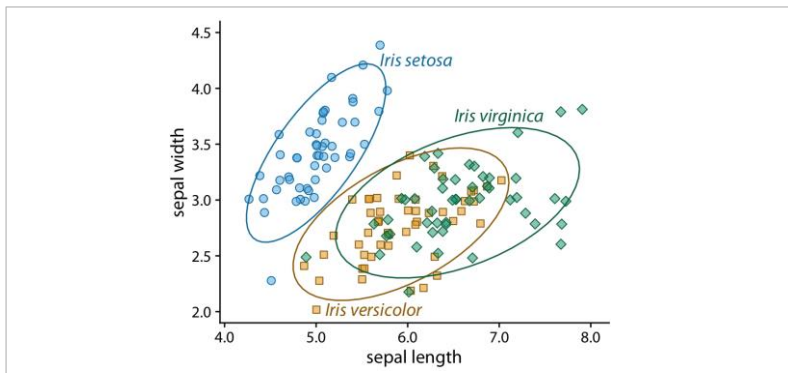
17

1. 중복 코딩과 범례

레이블 달기

☞ 그룹 단위로 타원으로 표현하고 레이블을 표시 → 그룹의 분류는 명확해 지나 도표가 조악해질 수 있음, 타원의 의미에 대한 오해가 생길 수 있음

붓꽃(iris) 데이터



[출처] Fundamentals of Data Visualization

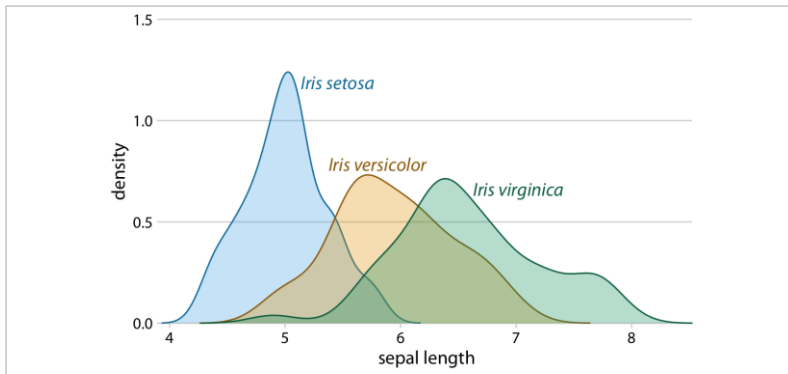
18

1. 중복 코딩과 범례

레이블 달기

- ☞ 밀도 추정 선에 레이블을 표시 → 레이블, 윤곽선, 색칠 영역의 색감을 조화롭게 사용해야 시각화 해석의 혼란과 조악함을 피할 수 있음

붓꽃(iris) 데이터



[출처] Fundamentals of Data Visualization

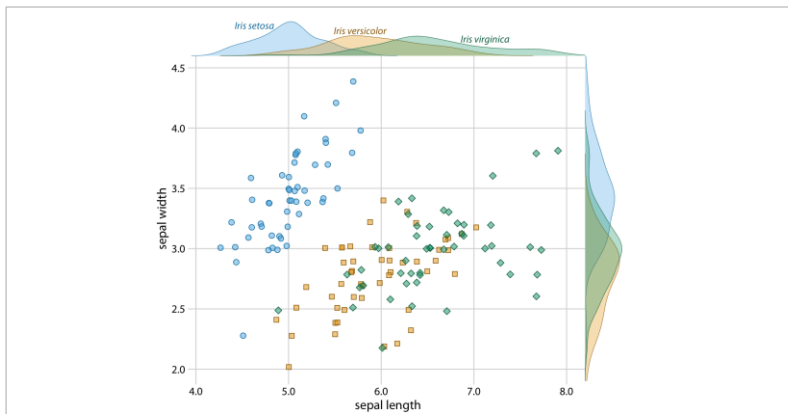
19

1. 중복 코딩과 범례

레이블 달기

- ☞ 산점도와 밀도 도표를 함께 그리고 레이블을 표시

붓꽃(iris) 데이터



[출처] Fundamentals of Data Visualization

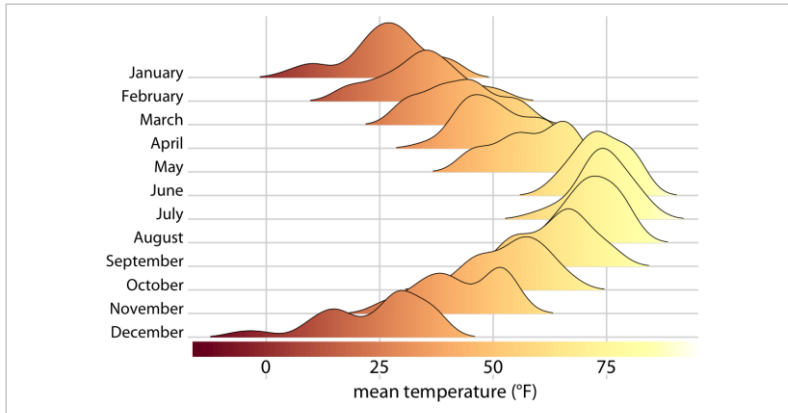
20

1. 중복 코딩과 범례

≡ 축과 범례의 통합

📌 색상 범례를 x축에 통합 → 기온 숫자 값과 색상을 통합하여 간명하게 표현

네브레스카 링컨시의 기온



[출처] Fundamentals of Data Visualization

21

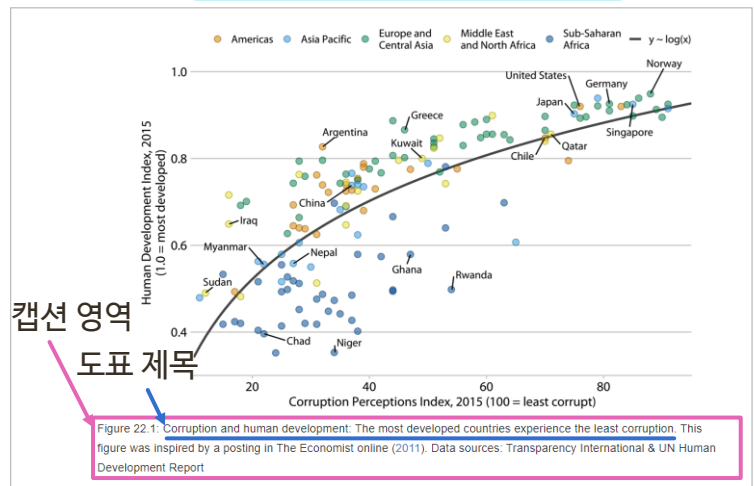
2. 제목과 캡션

≡ 도표 제목의 위치

📌 도표의 제목을 캡션 영역의 처음에 표시

→ 책이나 논문의 레이아웃에 적합, 캡션의 첫 부분은 제목으로 시작해야 함

부패 인지 지수와 인간 개발 지수



[출처] Fundamentals of Data Visualization

22

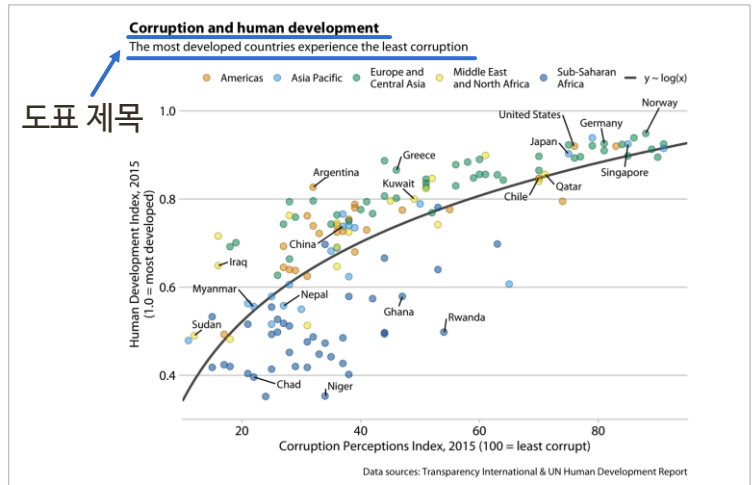
2. 제목과 캡션

도표 제목의 위치

☞ 도표 안에 제목을 표시

→ 도표를 단독으로 싣거나
소셜미디어, 웹페이지에
설명 글없이 게시하는 경우에
적합

부패 인지 지수와 인간 개발 지수



[출처] Fundamentals of Data Visualization

23

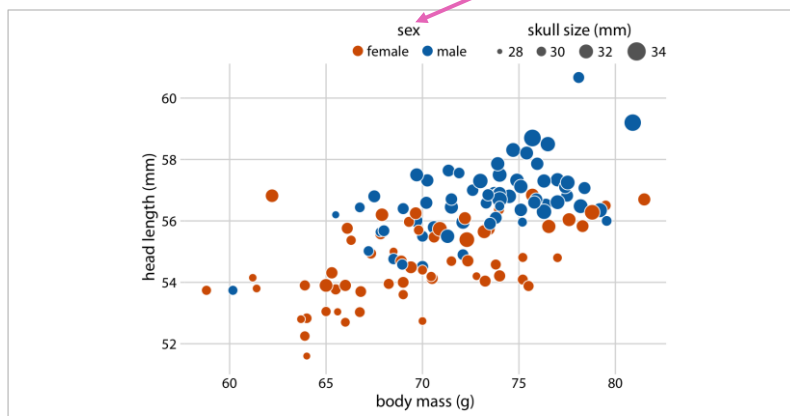
2. 제목과 캡션

축 레이블과 범례 제목

☞ 축 레이블에는 도표에 표시된 변수명과 함께 단위를 표시

큰어치 123마리 데이터

불필요한 범례명



[출처] Fundamentals of Data Visualization

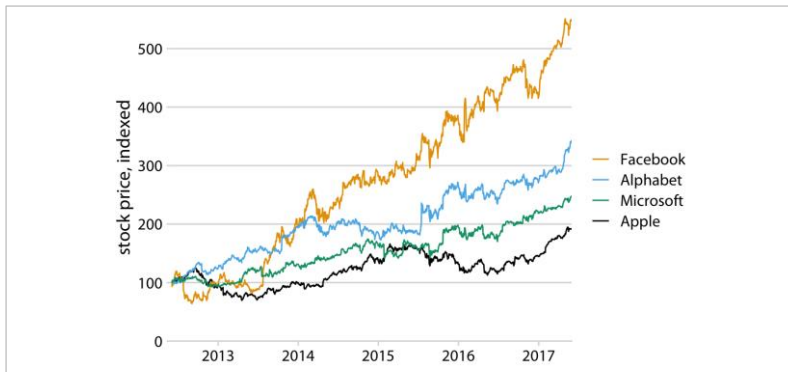
24

2. 제목과 캡션

축 레이블과 범례 제목

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

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

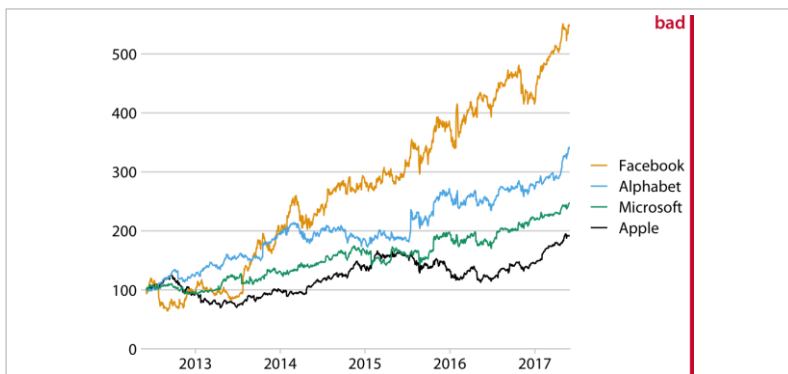
25

2. 제목과 캡션

축 레이블과 범례 제목

☞ y축의 축 레이블을 생략 → 도표 제목과 캡션에 y축을 추측할 수 있는 정보가 있다고 하더라도 독자가 모호하게 받아들일 수 있음

4개 기업의 주식 가격



[출처] Fundamentals of Data Visualization

26

3. 표

표 레이아웃의 핵심원칙

- 01 세로선을 넣지 않는다.
- 02 데이터의 행과 행사이에 가로 선을 넣지 않는다.
(가로 선은 헤더 행과 첫 데이터의 행을 구분하는 용도, 전체 테두리 용도)
- 03 글자 열은 왼쪽을 기준으로 정렬해야 한다.
- 04 숫자 열은 오른쪽을 기준으로 정렬하고,
모든 숫자의 소수점 자릿수를 맞춘다.
- 05 열에 숫자나 문자가 1글자만 들어갈 경우에는 가운데로 정렬한다.
- 06 헤더 칸은 아래쪽 데이터 열과 정렬을 맞춰야 한다.

27

3. 표

표 예시

시각적 어수선했음이 정보 전달을 방해할 수 있으므로 원칙을 준수해야 함

a

Rank	Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

b

Rank	Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

c

Rank	Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

d

Rank	Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

[출처] Fundamentals of Data Visualization

28



면분할과 복합형 도표

29

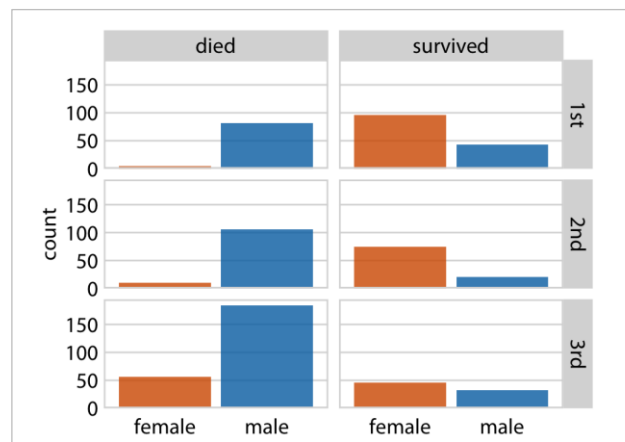
1. 면분할

면분할(Faceting)

성별, 객실 등급, 생존 여부에 따라
다중 패널로 표현

→ 여러 변수가 있는 데이터를
직관적으로 이해하기 쉬움

타이타닉호 승선객 데이터



[출처] Fundamentals of Data Visualization

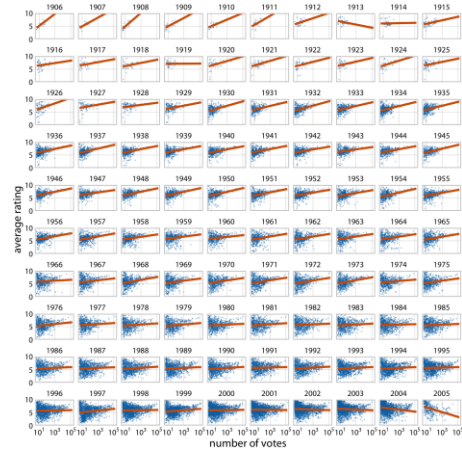
30

1. 면분할

면분할(Faceting)

- 2개의 변수를 연도별로 구분하고 연도 순으로 도표를 배치
→ 연도에 따른 추세를 알 수 있음, 도표의 배치 순서가 중요

영화 평균 순위와 득표수



[출처] Fundamentals of Data Visualization

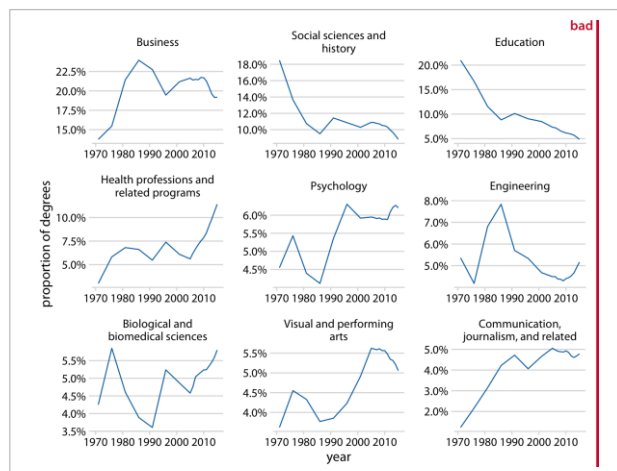
31

1. 면분할

면분할의 축 설정

- y축의 범위가 도표 별로 다름
→ 독자가 y축의 정보를 잘못 이해할 수 있음

미국의 학사학위 취득 추세



[출처] Fundamentals of Data Visualization

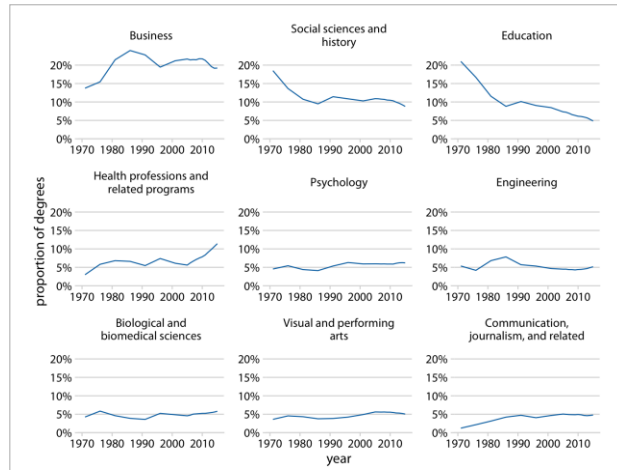
32

1. 면분할

면분할의 축 설정

- ☞ y축의 범위를 통일
 - 전공별 추세를 정확하게 비교할 수 있음

미국의 학사학위 취득 추세



[출처] Fundamentals of Data Visualization

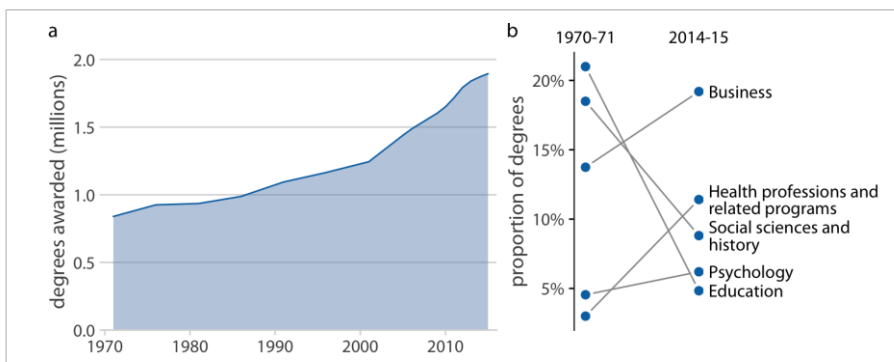
33

2. 복합형 도표

도표를 조합하여 중요한 정보를 전달

- ☞ 전체 추세는 선 그래프로 인기 있는 5개 전공의 추세는 경사 도표로 표현
 - 전체적인 추세와 세부 정보를 적절하게 함께 표현

미국의 학사학위 취득 추세



[출처] Fundamentals of Data Visualization

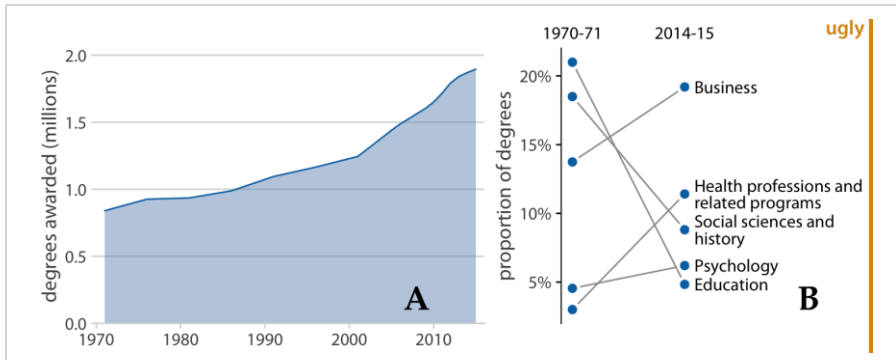
34

2. 복합형 도표

도표의 번호

☞ 도표의 번호를 오른쪽 아래에 크게 표시 → 도표를 읽는데 방해가 되는 위치에
도표 번호가 있고 전체 글꼴과 어울리지 않음

미국의 학사학위 취득 추세



[출처] Fundamentals of Data Visualization

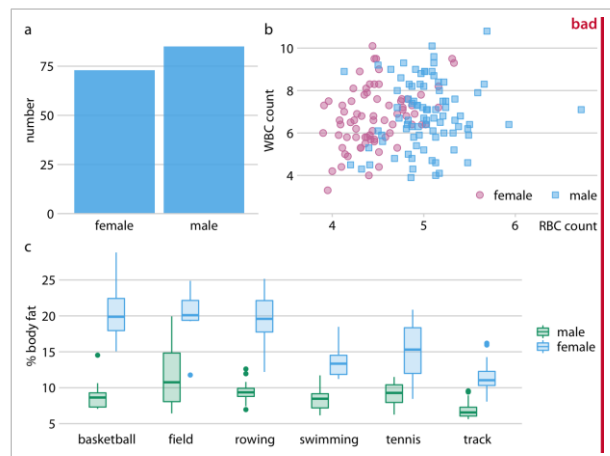
35

2. 복합형 도표

도표의 일관성

☞ 남녀의 위치와 색상이 도표마다
다름 → 3개의 도표를 일관되게
해석할 수 없음, 시각화를
해석하는데 혼란이 있고 시간이
오래 걸림

운동선수의 생리 기능과 체성분



[출처] Fundamentals of Data Visualization

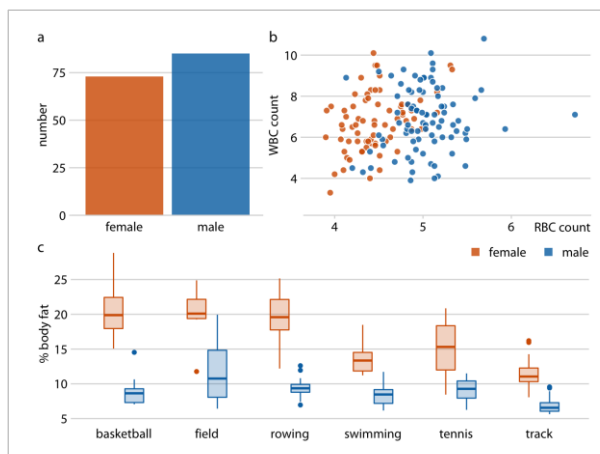
36

2. 복합형 도표

☞ 도표의 일관성

☞ 남녀의 위치와 색상을 일관성 있게 사용

운동선수의 생리 기능과 체성분



[출처] Fundamentals of Data Visualization

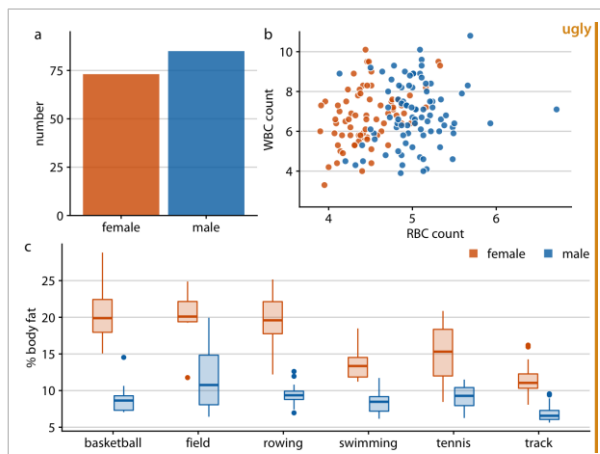
37

2. 복합형 도표

☞ 도표의 정렬

☞ 모든 도표가 조금씩 어긋나 있음
→ 시각화 해석을 크게 방해 하지는
않으나 조악함

운동선수의 생리 기능과 체성분



[출처] Fundamentals of Data Visualization

38



03

chapter

실습

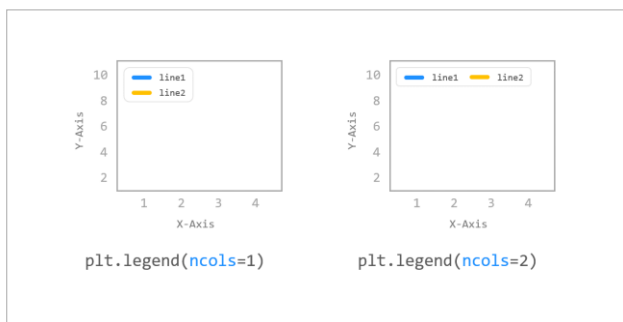
39

1. 도표 설명 요소

범례와 텍스트 삽입

범례

- `legend(loc='upper left', ncols=n)`



[출처] Matplotlib Tutorial - 파이썬으로 데이터 시각화하기

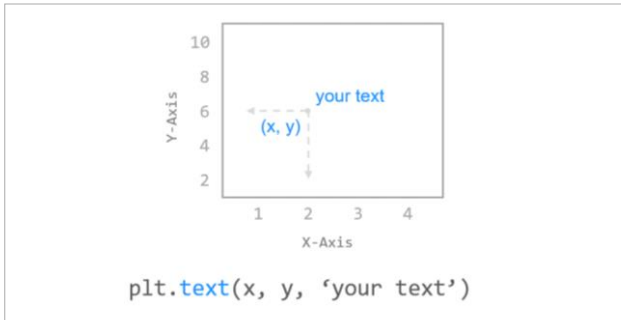
40

1. 도표 설명 요소

≡ 범례와 텍스트 삽입

📌 텍스트 삽입

- `text(x, y, 'your text')`



[출처] Matplotlib Tutorial - 파이썬으로 데이터 시각화하기

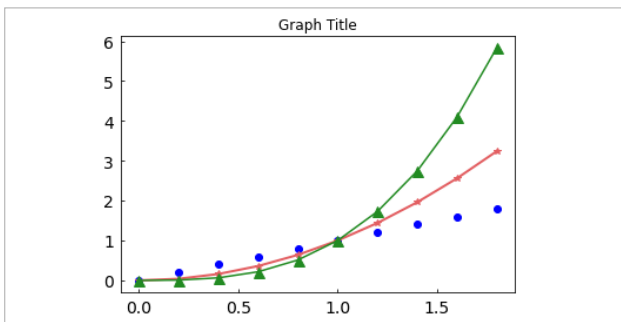
41

1. 도표 설명 요소

≡ 도표 제목과 축 레이블

📌 도표 제목

- `title('Graph Title', loc='center', pad=20))`



[출처] Matplotlib Tutorial - 파이썬으로 데이터 시각화하기

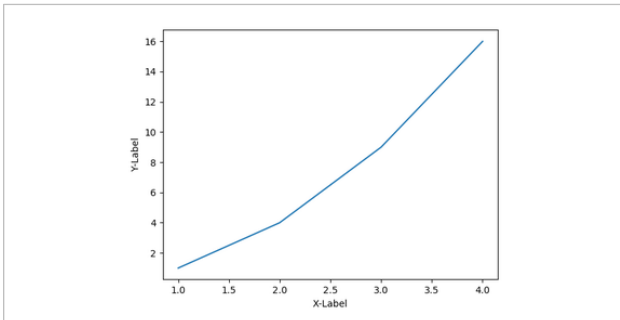
42

1. 도표 설명 요소

도표 제목과 축 레이블

축 레이블

- xlabel('X-Label')
- ylabel('Y-Label')



[출처] Matplotlib Tutorial - 파이썬으로 데이터 시각화하기

43

학습정리

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

44

● 학습정리

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

45

● 학습정리

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

46

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)

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.