

# 데이터 시각화

## 09. 비율 데이터의 시각화

최대영 교수



**고려사이버대학교**  
THE CYBER UNIVERSITY OF KOREA



Data Visualization

데이터 시각화

# 비율 데이터의 시각화

최대영 교수

09주차



1

## ● 학습리뷰

### 1 수량 데이터의 시각화

#### ☞ 막대 도표 (Bar plot)

- 수치 집합의 크기(수량)를 나타내야 하는 경우에 사용

#### ☞ 묶은 막대 도표 (Grouped bar)

- 두 가지 이상의 범주를 표현해야 하는 경우에 사용

#### ☞ 누적 막대 도표 (Stacked bar)

- 막대들을 쌓아서 합을 도출하는 것이 의미가 있는 경우에 사용

2

## 1 수량 데이터의 시각화

### 📌 점 도표(Dot plot)

- 막대 도표가 기준점 0으로 부터 길이로 정량 값을 표현하는 것의 단점을 보완
- 기준점을 조정하여 데이터를 더 간명하게 표현 가능

### 📌 히트맵(Heatmap)

- 막대나 점 대신 색으로 데이터 값을 표현하여 전반적인 추세 확인이 쉬운 방법

## 2 분포 데이터의 시각화

### 📌 히스토그램(Histogram)

- 가로에 범주 구간, 세로에 구간에 포함된 데이터의 수를 표현

### 📌 밀도 도표(Density plot)

- 데이터의 분포를 연속적인 곡선으로 표현, 곡선 아래의 면적은 1

### 📌 박스 도표(Box plot)

- 데이터를 사분위로 나누고 박스안에 50%의 데이터를 표현

## ● 학습목표

● 단일 범주 비율 데이터의 시각화 방법에 대해 설명할 수 있다.

● 내포 비율의 시각화 방법에 대해 설명할 수 있다.

● 비율 데이터 관련 matplotlib 라이브러리를 이해하고 활용할 수 있다.

5

## ● 학습내용

1 단일 범주 비율 데이터의 시각화

2 내포 비율의 시각화

3 실습

6



# 단일 범주 비율 데이터의 시각화

7

## 1. 파이 차트와 막대 도표

### ≡ 비율(proportion) 데이터의 시각화

어떤 집단, 조직체, 수량이 개개의 조각으로 나뉘어 있을 때  
전체에서 차지하는 비율을 시각화로 표현

예 한 집단에서의 남녀 성비, 정당별 선거 득표율, 기업의 시장 점유율



[출처] Fundamentals of Data Visualization

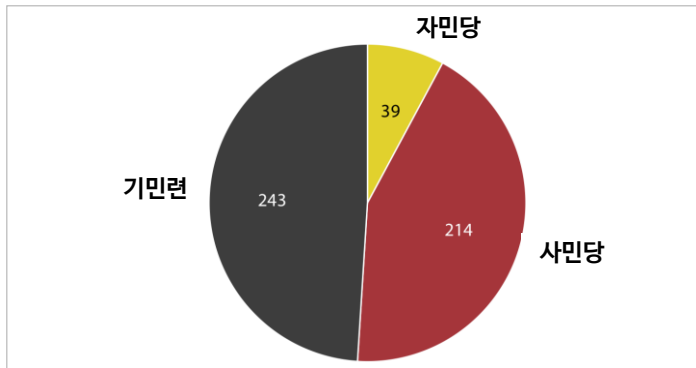
8

## 1. 파이 차트와 막대 도표

### ≡ 파이 차트(Pie chart)

📌 전체 데이터에 해당하는 원을 부분 데이터가 차지하는 비중에 비례하는 크기의 조각으로 분할

독일 8대 국회의 정당 구성



[출처] Fundamentals of Data Visualization

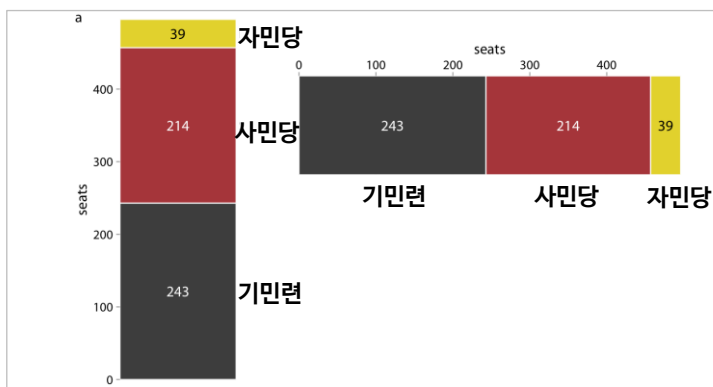
9

## 1. 파이 차트와 막대 도표

### ≡ 누적 막대 도표(Stacked bar chart)

📌 원 대신 직사각형의 조각으로 표현 → 사민당과 자민당의 합이 기민련 보다 크다는 것을 알기 어려움

독일 8대 국회의 정당 구성



[출처] Fundamentals of Data Visualization

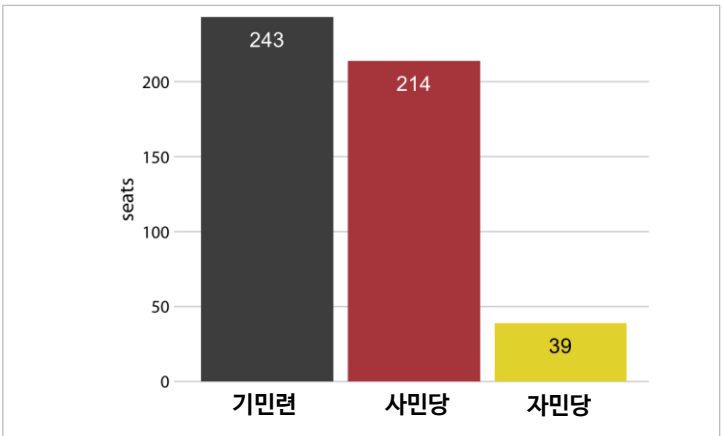
10

# 1. 파이 차트와 막대 도표

## ≡ 막대 도표(Bar plot)

📝 병렬 막대로 표현 → 각 막대가 전체에서 차지하는 비율을 알기 어려움

독일 8대 국회의 정당 구성



[출처] Fundamentals of Data Visualization

11

# 1. 파이 차트와 막대 도표

## ≡ 비율 데이터 시각화 방법의 비교

특 성	파이 차트	누적 막대	병렬 막대
데이터가 전체에서 차지하는 비율을 확실하게 보여줌	●	●	✕
상대적인 비율을 비교하기 쉬움	✕	✕	●
1/2, 1/3, 1/4 등 단순한 분수 비율을 강조하기 쉬움	●	✕	✕
데이터셋이 매우 작은 경우에도 시각적으로 보기 좋음	●	✕	●
전체가 매우 많은 조각으로 나뉘어 있을 때 효과적임	✕	✕	●
시간의 흐름에 따른 비율을 시각화 할 때 효과적임	✕	●	✕

[출처] Fundamentals of Data Visualization

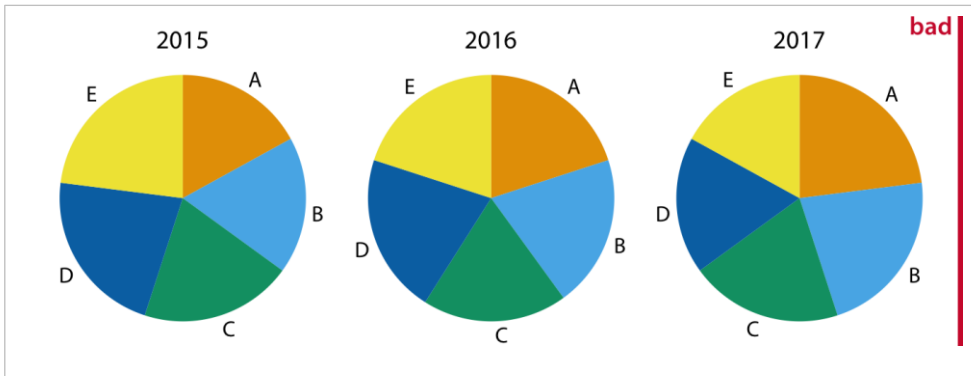
12

## 1. 파이 차트와 막대 도표

### ≡ 파이 차트가 효과적이지 않은 경우

📌 3년 연속 시장 점유율을 표현 → 추이와 점유율의 차이를 확인하기 어려움

가상 회사의 시장 점유율



[출처] Fundamentals of Data Visualization

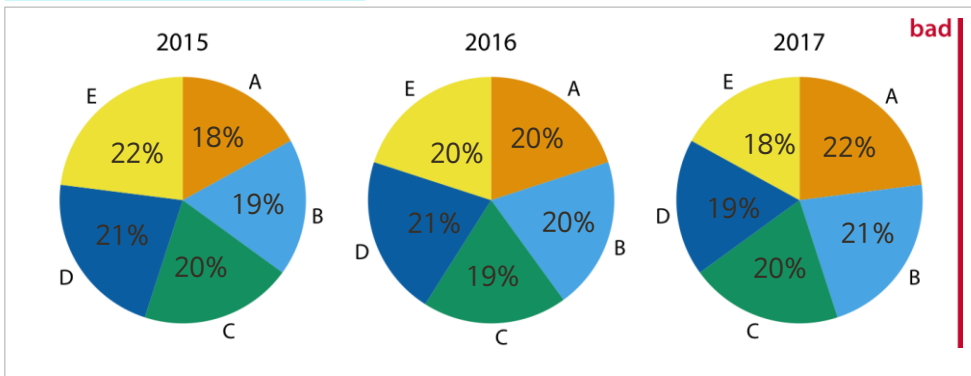
13

## 1. 파이 차트와 막대 도표

### ≡ 파이 차트가 효과적이지 않은 경우

📌 파이 차트안에 비율을 표현 → 여전히 3년 간의 추이가 두드러지지 않음

가상 회사의 시장 점유율



[출처] Fundamentals of Data Visualization

14

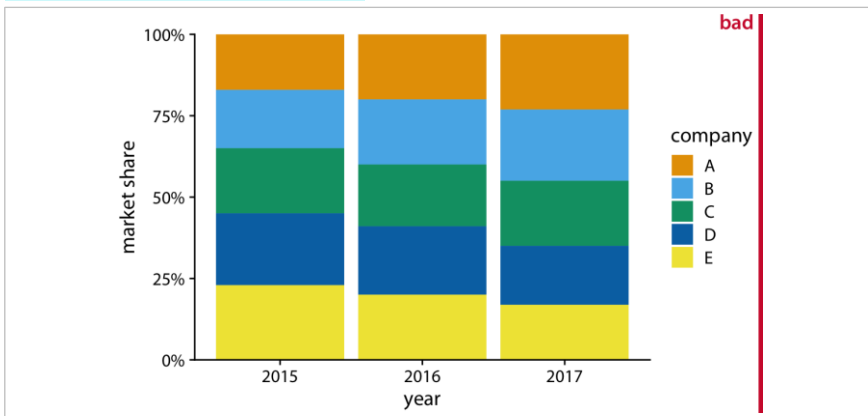


## 1. 파이 차트와 막대 도표

### 누적 막대 도표

☞ A와 E의 변화량을 뚜렷하게 볼 수 있음 → B, C, D의 변화는 보기 어려움

가상 회사의 시장 점유율



[출처] Fundamentals of Data Visualization

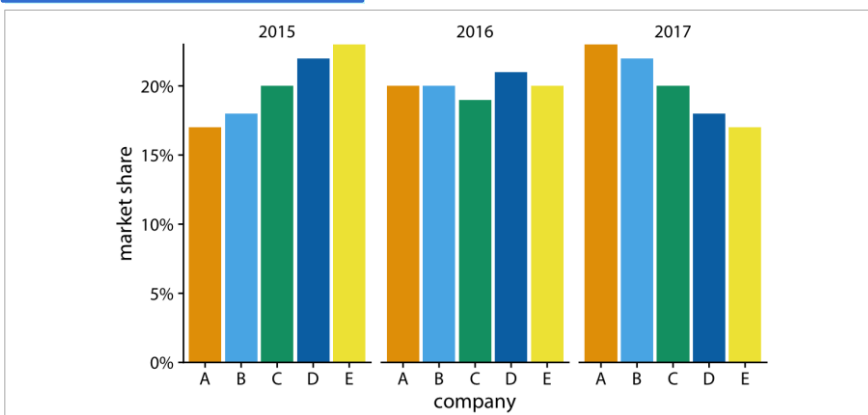
15

## 1. 파이 차트와 막대 도표

### 병렬 막대

☞ 회사별 점유율을 각각 막대로 표현 → 시간에 따른 변화와 연도내 순서를 보여줌

가상 회사의 시장 점유율



[출처] Fundamentals of Data Visualization

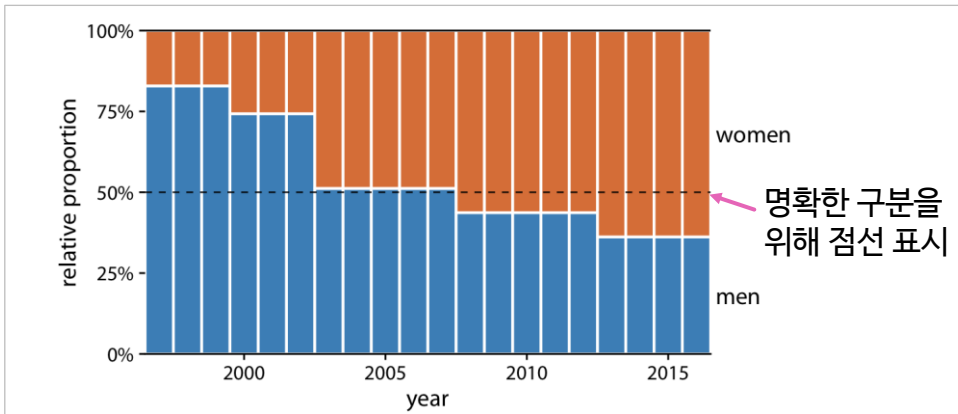
16

## 2. 누적 막대와 누적 밀도

### ≡ 누적 막대가 적절한 경우

📌 2개의 성별을 누적하여 표현 → 막대의 위치가 변하는 문제가 없이 간명함

르완다 국회의 남녀 성비



[출처] Fundamentals of Data Visualization

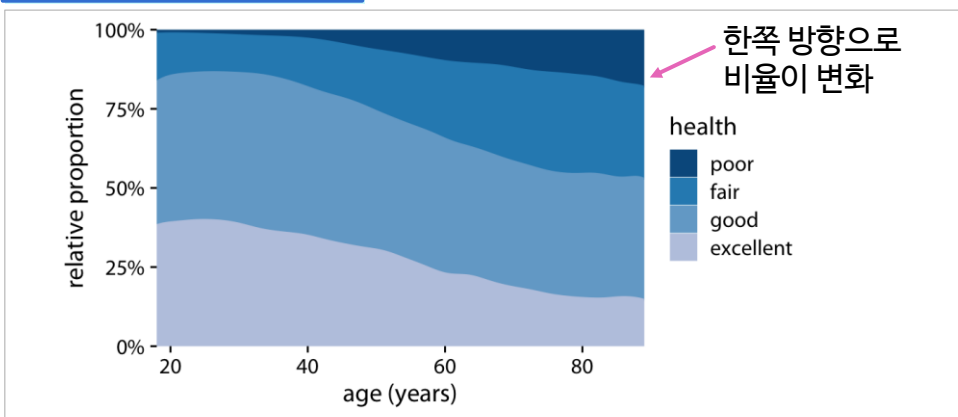
17

## 2. 누적 막대와 누적 밀도

### ≡ 누적 밀도가 적절한 경우

📌 연속형 데이터에 대해 밀도를 누적하여 표현 → 데이터의 특성에 부합

연령별 건강 상태



[출처] Fundamentals of Data Visualization

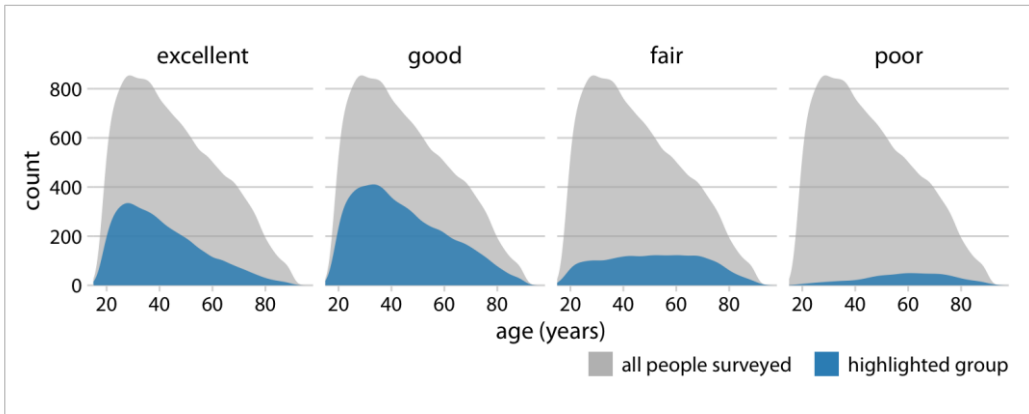
18

## 2. 누적 막대와 누적 밀도

### 전체 대비 부분 비율의 표현

전체 분포를 배경으로 두고 각 범주를 표현 → 연령별 인원수 차이를 반영

연령별 건강 상태



[출처] Fundamentals of Data Visualization

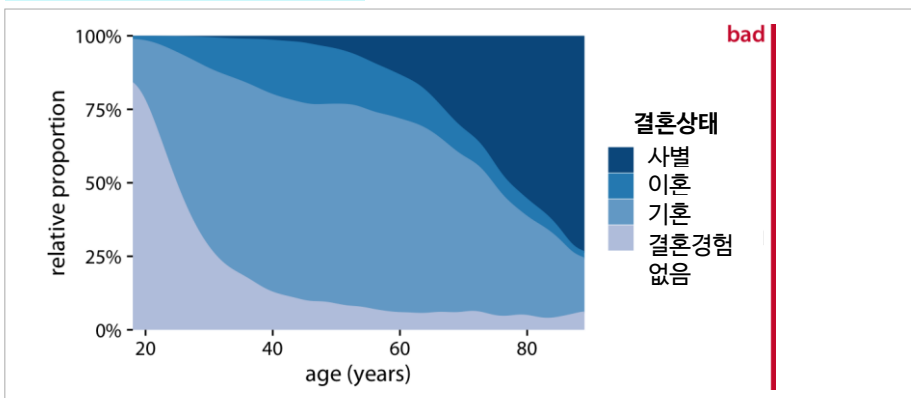
19

## 2. 누적 막대와 누적 밀도

### 누적 밀도가 부적절한 경우

연속형 데이터에 대해 밀도를 누적하여 표현 → 결혼경험 없음과 사별의 연령별 변화가 급격하여 이혼과 기혼의 해석이 어려움

연령별 결혼 상태



[출처] Fundamentals of Data Visualization

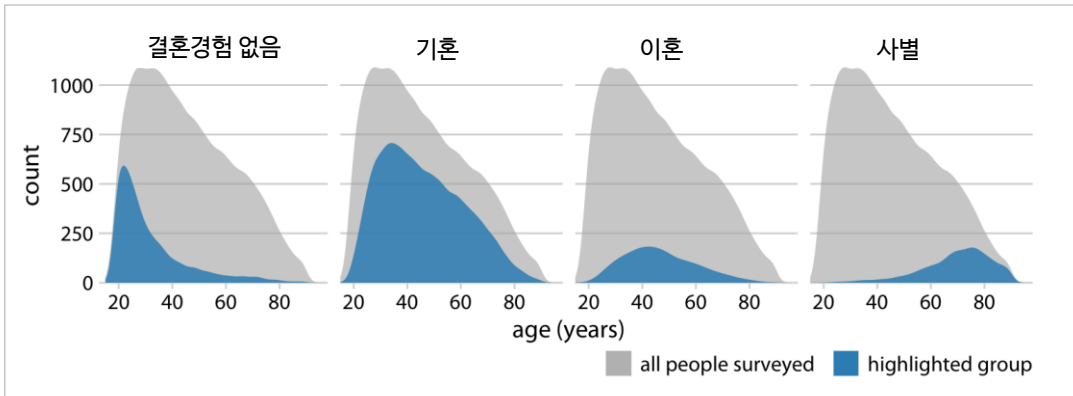
20

## 2. 누적 막대와 누적 밀도

### 전체 대비 부분 비율의 표현

전체 분포를 배경으로 두고 각 범주를 표현 → 결혼 상태별 연령 특성을 표현

연령별 결혼 상태



[출처] Fundamentals of Data Visualization

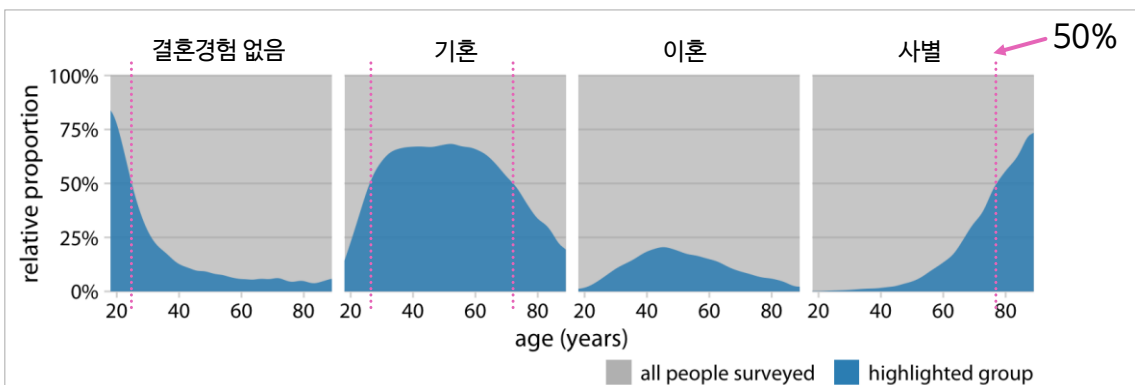
21

## 2. 누적 막대와 누적 밀도

### 전체 대비 부분 비율의 표현

강조하려는 결혼 상태와 이외의 상태의 비율을 표현 → 특정 연령대의 결혼 상태의 비율을 파악하기 쉬움

연령별 결혼 상태



[출처] Fundamentals of Data Visualization

22



# 내포 비율의 시각화

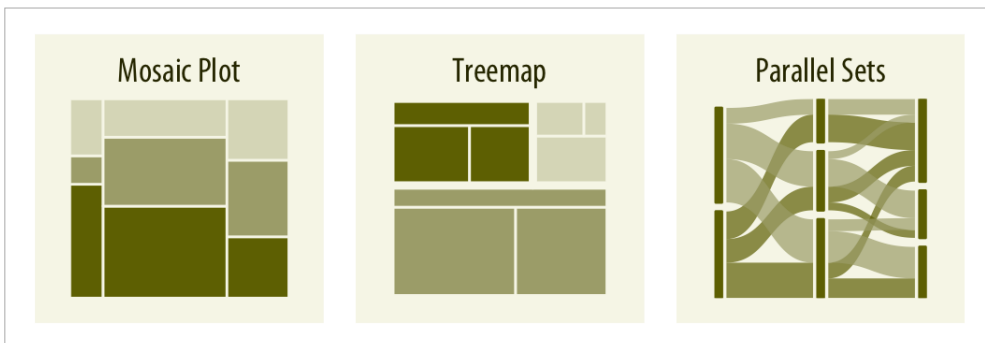
23

## 1. 모자이크 도표와 트리맵

### ≡ 내포 비율(Nested proportion)

📌 하나의 범주 하위에 또 다른 범주가 있는 데이터의 **하위 범주의 비율**

예 정당별 성비, 결혼 상태별 건강상태



[출처] Fundamentals of Data Visualization

24

## 1. 모자이크 도표와 트리맵

### ≡ 피츠버그 다리(bridge) 데이터

다리명	건설 자재	건설 시기
Bridge 1	나무(wood)	전통식(craft)
Bridge 2	철(iron)	초기(emerging)
⋮	⋮	⋮
Bridge 105	강철(steel)	성숙기(mature)
Bridge 106	강철(steel)	근대식(modern)

[출처] Fundamentals of Data Visualization

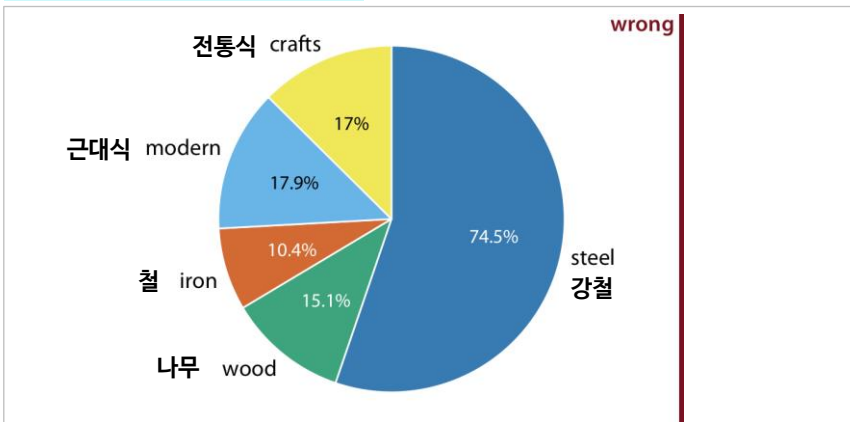
25

## 1. 모자이크 도표와 트리맵

### ≡ 내포 비율 시각화의 잘못된 사례

🔗 파이 차트 조각의 총합이 135% → 교량의 개수를 이중으로 계산

피츠버그 다리 데이터



[출처] Fundamentals of Data Visualization

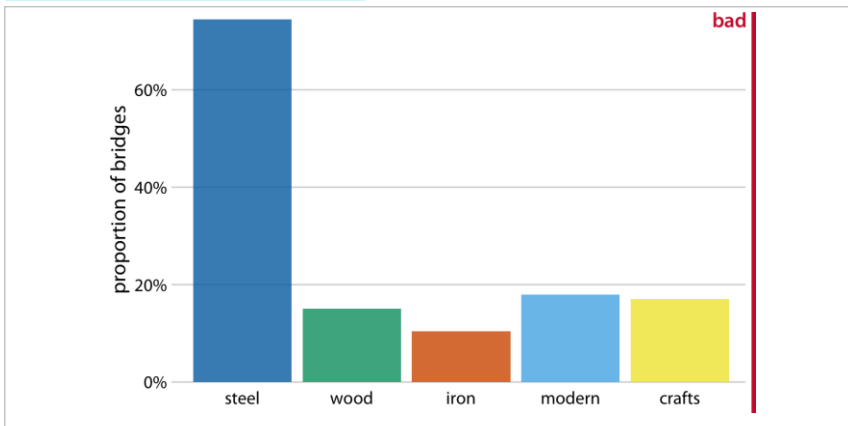
26

## 1. 모자이크 도표와 트리맵

### ≡ 막대 도표로 두 범주를 동시에 표현

📌 두 범주의 상대적인 값의 크기 비교 → 중복 집계가 있다는 것을 표현하지 않음

피츠버그 다리 데이터



[출처] Fundamentals of Data Visualization

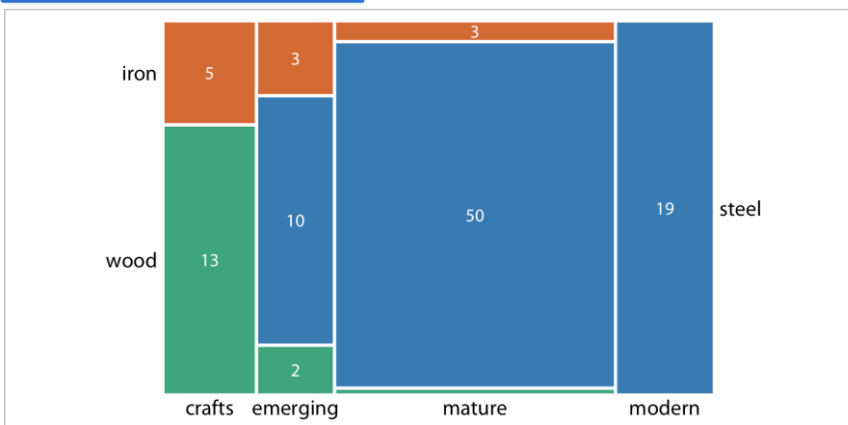
27

## 1. 모자이크 도표와 트리맵

### ≡ 모자이크 도표(Mosaic plot)

📌 길이로 범주 내 비율을 표현 → 가로 범주의 변화(차이)를 강조하는데 효과적

피츠버그 다리 데이터



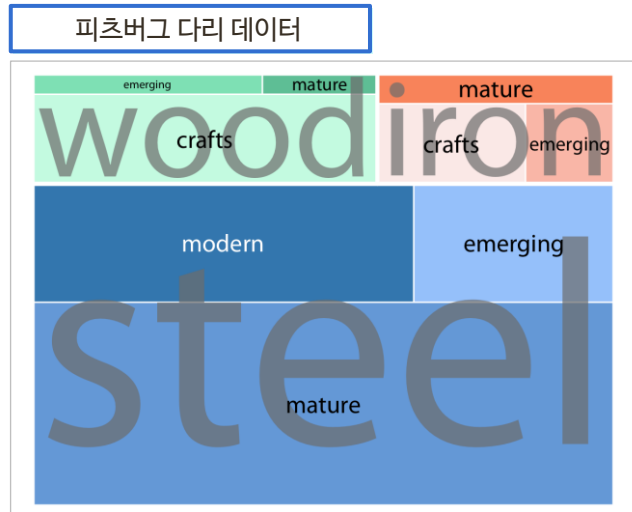
[출처] Fundamentals of Data Visualization

28

## 1. 모자이크 도표와 트리맵

### 트리맵(Treemap)

- 사각형 안에 작은 사각형을 반복적으로 계층을 가지며 쪼갬  
→ 여러 범주형 변수의 조합이 중요한 의미를 가지지 않을 때 효과적임



[출처] Fundamentals of Data Visualization

29

## 1. 모자이크 도표와 트리맵

### 트리맵(Treemap)

- 사각형 안에 작은 사각형을 반복적으로 계층을 가지며 쪼갬  
→ 하나의 범주 안에 속한 데이터는 다른 범주 안에 속한 데이터와 무관함



[출처] Fundamentals of Data Visualization

30

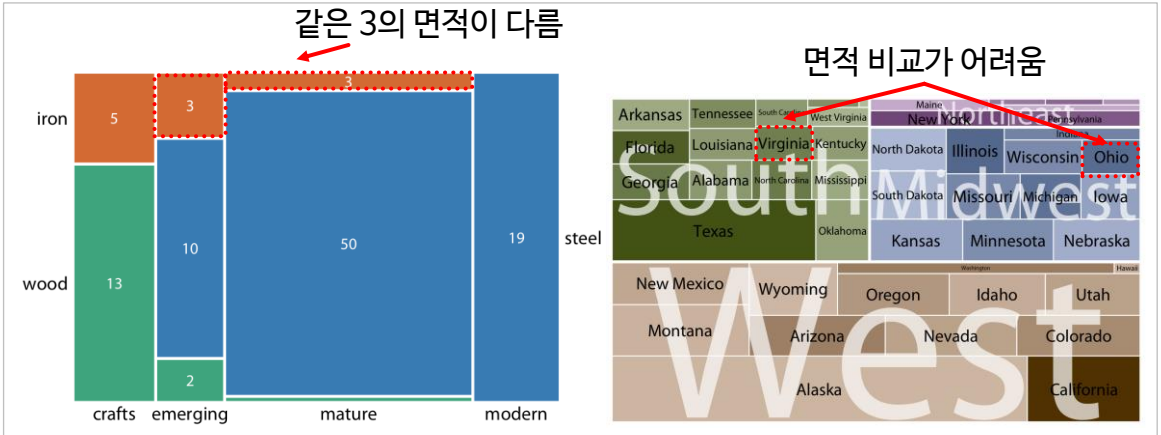


## 1. 모자이크 도표와 트리맵

### 모자이크 도표와 트리맵의 단점

직사각형의 모양이 제각각이라 데이터의 크기 비교가 어려움

→ 숫자나 백분율을 도표 내에 표시



[출처] Fundamentals of Data Visualization

31

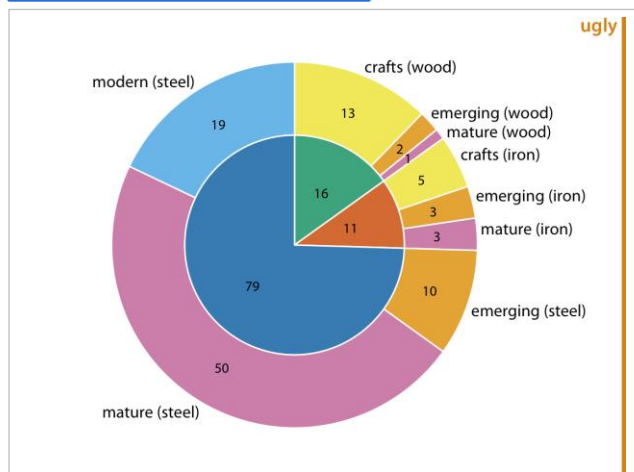
## 2. 내포 파이 차트와 평행 집합 도표

### 내포 파이 차트(Nested pie)

외부원과 내부원으로 구분하여 2개의 범주를 각각 표현

→ 복잡하고 폰트가 작아 간명하게 알아볼 수 없음, 2개의 범주가 있다는 것을 인식하기 어려움

피츠버그 다리 데이터



[출처] Fundamentals of Data Visualization

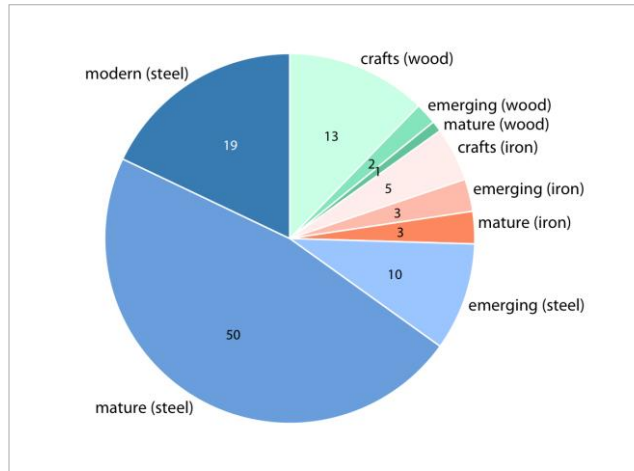
32

## 2. 내포 파이 차트와 평행 집합 도표

### ≡ 내포 파이 차트(Nested pie)

- 범주의 조합을 조각과 색으로 표현(색감: 건설 자재, 밝기: 건설 시기) → 트리맵 대비 공간활용이 비효율적임, 파이 조각이 작아 레이블 표현이 어려움

피츠버그 다리 데이터



[출처] Fundamentals of Data Visualization

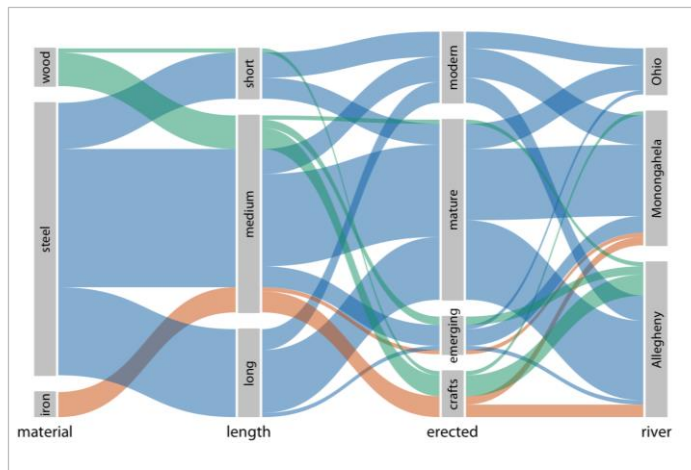
33

## 2. 내포 파이 차트와 평행 집합 도표

### ≡ 평행 집합 도표(Parallel sets plot)

- 2개 이상의 범주가 있는 데이터의 범주별 분류와 분류간의 관계성을 나타냄 → 파이 차트, 모자이크 차트, 트리맵에 비해 여러 범주를 나타내는데 효율적임

피츠버그 다리 데이터



[출처] Fundamentals of Data Visualization

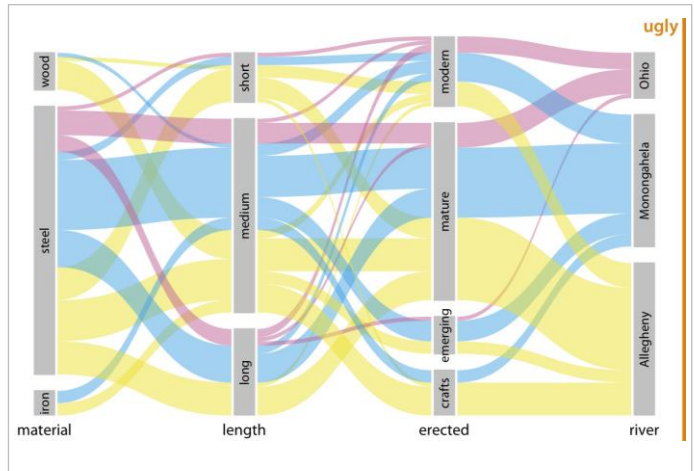
34

## 2. 내포 파이 차트와 평행 집합 도표

### ≡ 평행 집합 도표(Parallel sets plot)

- ✎ 색을 칠하는 기준을 변경하면 전하려는 메시지가 달라짐  
→ 어떤 유형의 다리가 어느 강에 설치되어 있는지를 잘 나타냄, 너무 많은 교차가 발생하여 혼란스러움

피츠버그 다리 데이터



[출처] Fundamentals of Data Visualization

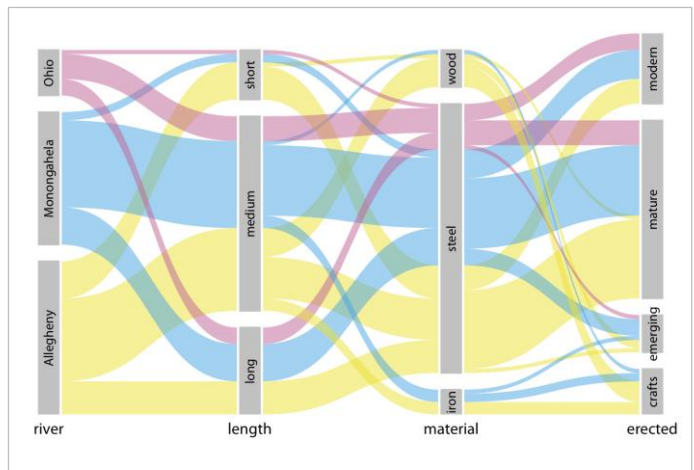
35

## 2. 내포 파이 차트와 평행 집합 도표

### ≡ 평행 집합 도표(Parallel sets plot)

- ✎ 범주의 순서를 변경하면 메시지 전달이 쉬워짐 → 왼쪽에서 오른쪽으로 정보가 나열되어 인지가 쉽고 각 띠의 교차점이 줄어들어 간명해짐

피츠버그 다리 데이터



[출처] Fundamentals of Data Visualization

36



03  
chapter

# 실습

37

## 1. 비율 데이터의 시각화

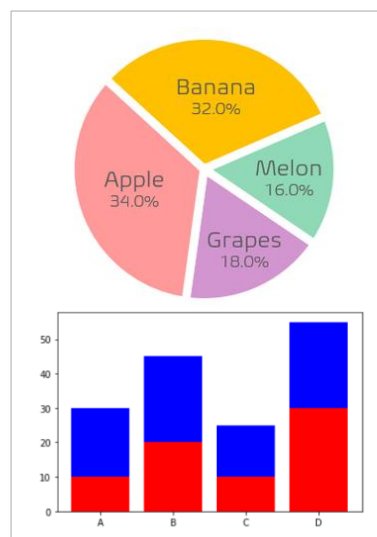
### ≡ 파이 차트와 누적 막대 도표

#### 📌 파이 차트

- `pie(ratio, labels=labels)`

#### 📌 누적 막대 도표

- `bar(x, y1, color='r')`  
`bar(x, y2, bottom=y1, color='b')`



[출처] Matplotlib Tutorial - 파이썬으로 데이터 시각화하기

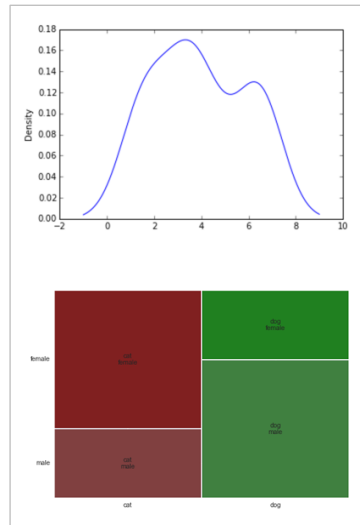
38

## 1. 비율 데이터의 시각화

### 밀도 도표와 모자이크 도표

#### 밀도 도표

- `pd.df.plot(kind='density')`
  - \* pandas dataframe의 `plot` 함수를 이용



#### 모자이크 도표

- `mosaic(data, ['pet', 'gender'])`
  - \* statsmodels 패키지를 이용

[출처] Matplotlib Tutorial - 파이썬으로 데이터 시각화하기

39

## 학습정리

### 1 단일 범주 비율 데이터의 시각화

#### 파이 차트 (Pie chart)

- 전체 데이터에 해당하는 원을 부분 데이터가 차지하는 비율에 비례하는 크기의 조각으로 분할

#### 누적 막대 도표 (Stacked bar chart)

- 파이 차트의 원 대신 직사각형의 조각으로 표현

#### 병렬 막대

- 누적 막대 도표의 직사각형 조각을 나열하여 비율을 표현

40

## ● 학습정리

### 1 단일 범주 비율 데이터의 시각화

#### ☞ 누적 밀도

- 연속형 데이터에 대해 밀도를 누적하여 표현

#### ☞ 전체 대비 부분 비율의 표현

- 전체 분포를 배경으로 두고 각 범주를 표현

41

## ● 학습정리

### 2 내포 비율의 시각화

#### ☞ 모자이크 도표 (Mosaic plot)

- 가로, 세로의 길이로 범주 내 비율을 표현

#### ☞ 트리맵 (Treemap)

- 사각형 안에 작은 사각형을 반복적으로 계층을 가지며 쪼개어 표현

42

## ● 학습정리

### 2 내포 비율의 시각화

#### 📌 내포 파이 차트(Nested pie)

- 외부원과 내부원으로 구분하여 2개의 범주를 각각 표현

#### 📌 평행 집합 도표(Parallel sets plot)

- 2개 이상의 범주가 있는 데이터의 범주별 분류와 분류간의 관계성을 나타냄

43

## ● 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)

44

## 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.