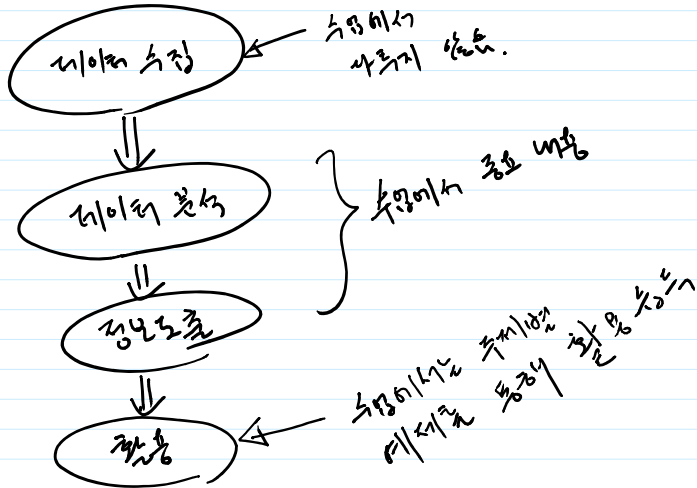


강의 간략 소개

통계란 : 혼란스러운 데이터로부터 의미 찾기

통계는 어디에나 있다.

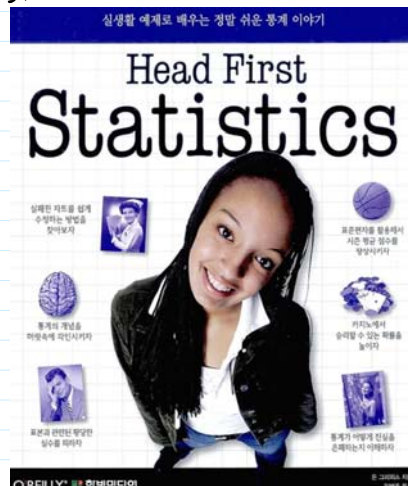


통계를 배워야 하는 이유 : 데이터를 바탕으로 하여 필요한 정보를 얻어낼 수 있음.

하지만 한번 정보를 여러 방식으로 해석할 수 있음.
 심지어 거짓말도 가능.

강의 소개: Head First Statistics

- 한빛미디어
- 저자: 존 그리피스
- 번역: 임 백준



책의 특징: 다양한 예시로 이해하기 용이하며 통계의 기본 개념을 이해 시키려는 노력함

강의 평가: 중간과, 기말과사, 출석, 숙제 등으로
 80% 10% 10%

1장 정보의 시각화

내용: 정보를 시각화하는 세 가지 방법 소개

- ① 파이차트
- ② 막대 그래프
- ③ 히스토그램

p. 43

예제: 아래 도표를 이용하시다. 어떤 회사의 월별 수익을 나타낸.

월	7	8	9	10	11	12
이익	2.0	2.1	2.2	2.1	2.3	2.4

어떻게 보이나요?

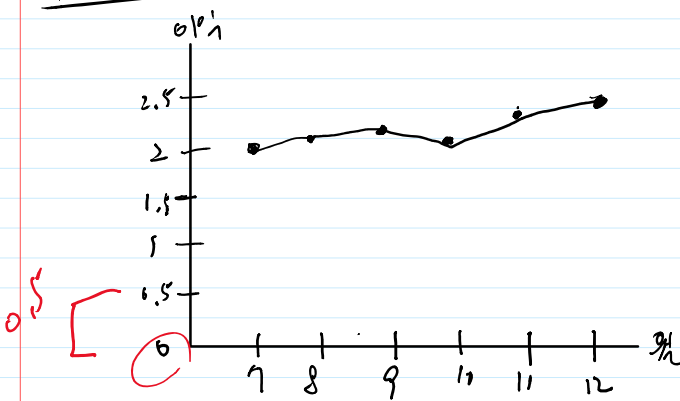
- ① 꾸준히 이익 나고 있지만 특별한 건 없음
- ② 이익의 변화가 매우 심함.

⇒ 어떻게 이런 해석의 차이가 가능한가?

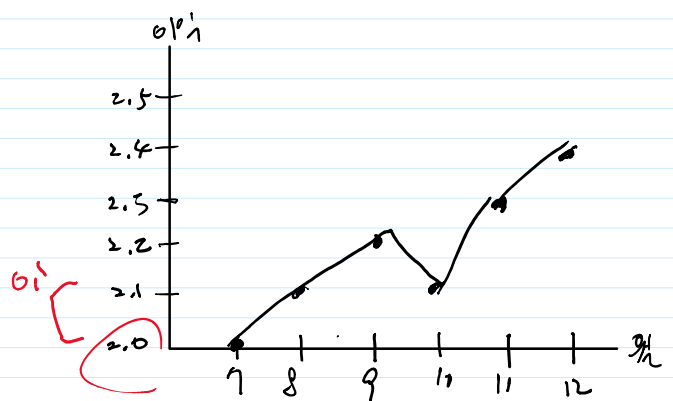
동일한 데이터에 대해

p. 44

방식 1



방식 2



결론: 동일한 데이터를 어떻게 시각화 하는가에 따라 해석이 달라짐.

p.45

Q&A :

데이터	정보
취급된 자료	의미가 부여된 데이터

| 장에서는 정보를 시각화는 다양한 방법 소개.
그래프를 이용하여

p.48

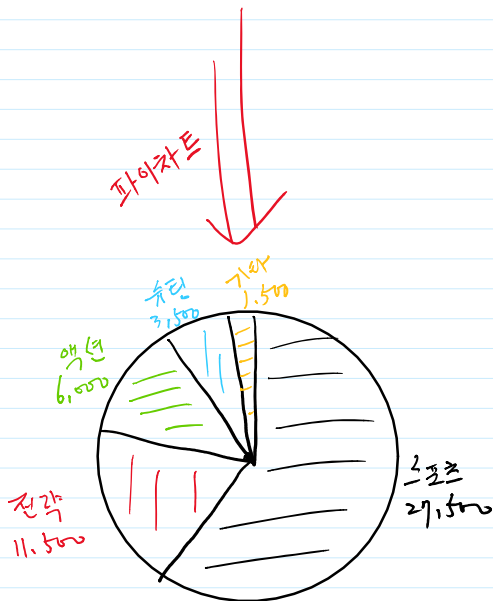
파이차트

예제: 기업회사에서 판매한 기원의 품종별 판매량

반경

품종	판매량
스프링	21,500
전각	11,500
백색	6,000
슈림	3,500
기타	1,500

반경 300

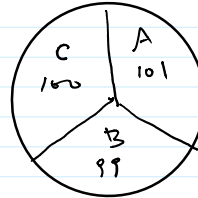


원사 제품
전체 판매량 중에서
각 제품이 차지하는
비율을 한눈에 볼 수
있게 도와 줌.

파이차트의 장점: 변주형 도수의 상대적 비율을 쉽게 보여줌
 파이차트의 단점: 각 변주의 도수의 차이가 미미할 경우 차이를
 보기 어려움
 예)

제품	판매량
A	101
B	99
C	100

=>



각 파이의 크기가
 구분되지 않음

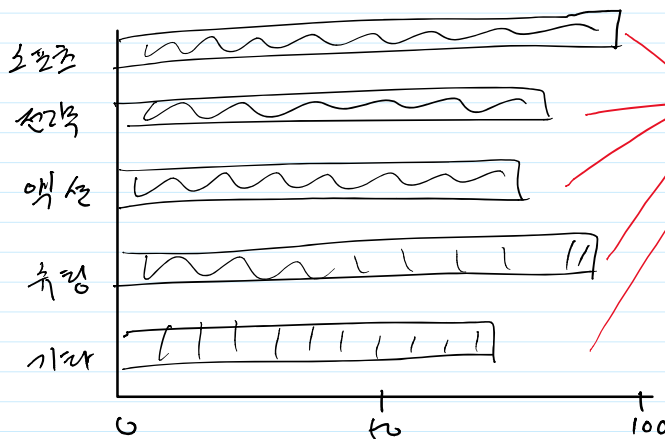
막대그래프

변주형 상대적 비율이 아니라
 변주형 미소한 차이를 명확히 보여주고자 할 때.

예) 게임 장르별 만족도

장르	만족도(%)
소도전	99%
전략	90%
액션	85%
슈팅	95%
기타	80%

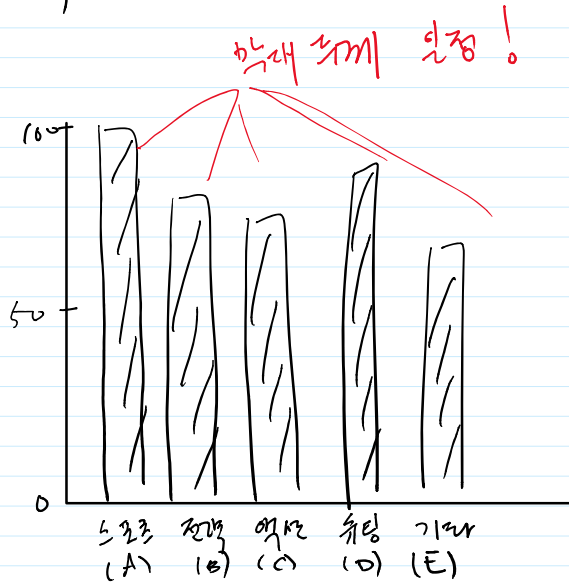
① 수평막대 그래프: 변주형이 길게 표현



만족도의 차이가
 일명해야 함

p.51

② 수직 막대 그래프



* 수직 VS. 수평 선택 방법:

변수들의 이론이 길 경우 수평 막대 그래프 추천.

하지만 반드시 그럴 필요 없다.

필요에 따라 선택 가능!

* 리스도 그래프

(수치별 정보로 유인 데이터 다룰 때 사용)

변적적 데이터
(정성적 데이터)

변적의 성질이나
특성에 따라
나뉘어진 데이터

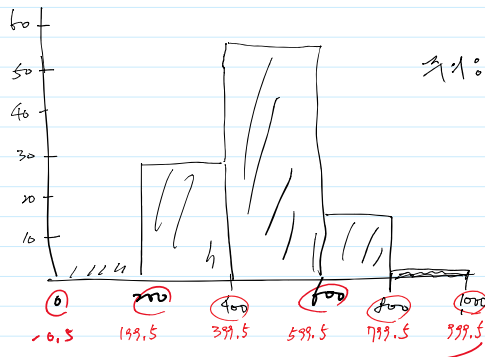
수치적 데이터
(정량적 데이터)

무게, 길이, 시간 등
숫자로 변치기 따라
나뉘어진 데이터

p.62

예1) 세 구간별 수

구간	수
0-199	5
200-399	29
400-599	56
600-799	19
800-999	3



구간의 연속성과
전체를 포함함을
보여주기 위해 필요.

주의: 빈칸 사이에 간격이 없어야 함.
또한 각 영역은 정지시 안동에서
오른 편차 방식을 포함해야 함

각 구간의 크기: 200으로
동반.
동시에 모든 편차를
포함함.

밀집하게 하자면 -0.5, 199.5, 399.5, 599.5, 799.5, 999.5
999.5 등으로 구간을 정해야 하지만
굳이 그럴 필요 없음.

p.64

주요미노

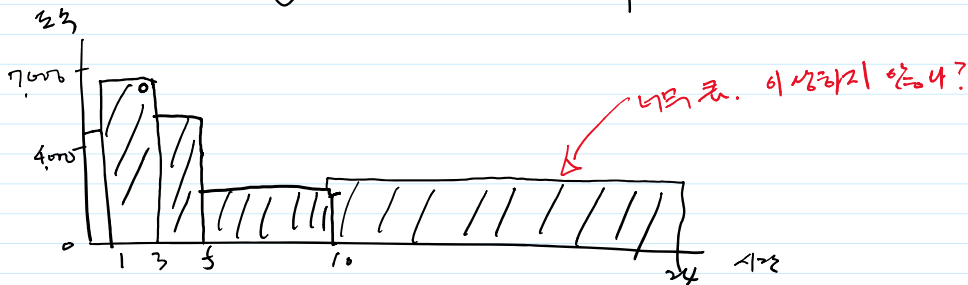
구간의 크기가 다른 경우

주요미노를 계산한 후 리프로그램 작성!

예) 사람들이 게임을 한 시간

시간	소수
0-1	4,300
1-3	6,900
3-5	4,900
5-10	2,000
10-24	2,100

↓
이제부터 동일한 방식으로 그려보자...



문제점: 예들 들어 10~24의 구간이 표시된 막대가 너무 커 보인다.

진실: 위 그래프를 이용하면 예들 들어 13시간 정도 게임하는 사람들의 수를 2,100명으로 추정하게 된다.

하지만 이 추정은 믿을 수 없다.

10~24시간 동안 게임하는 사람들의 수가 2,100명인데 13시간 정도 게임하는 이들의 수가 2,100명 된 가능성이 거의 없다.

7.15

해결책: 막대의 높이를 도수에 비례하게 만들어야 한다

방식: 막대의 변직을 도수와 일치시킨

$$\text{도수} = \text{막대의 폭} \times \text{막대의 높이}$$

p. 67

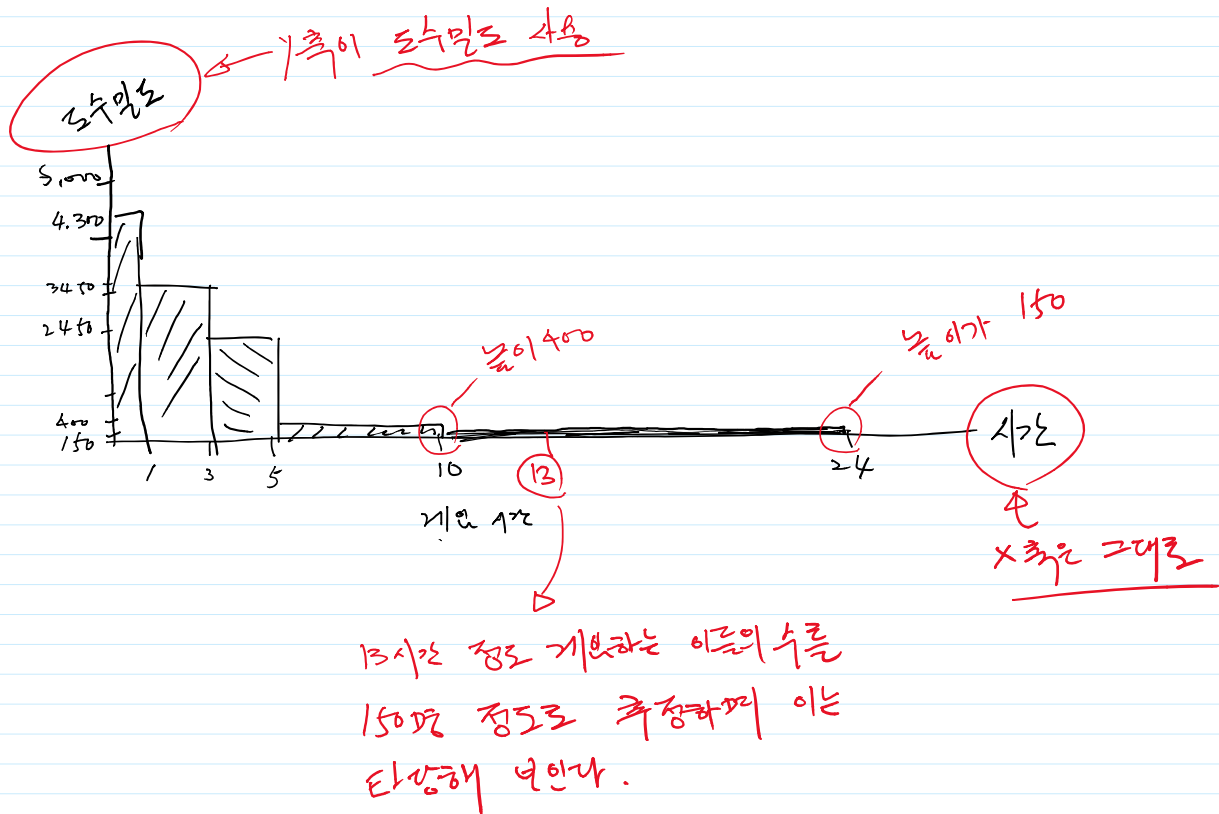
이 예제의 도수밀도 계산:

$$\text{빈도의 높이} = \frac{\text{도수}}{\text{빈도의 폭}}$$

"도수밀도"라 부름

p. 68

시간	도수	폭	높이 (도수밀도)
0-1	4,300	1	4,300
1-3	6,800	2	3,400
3-5	4,900	2	2,450
5-10	2,000	5	400
10-24	2,100	14	150



* 누적소득 그래프

p. 74

누적소득 : 특정한 값에 이르기까지 소득층의 소득 합계

예) 앞서 다룬 계급 사본의 소득 합계

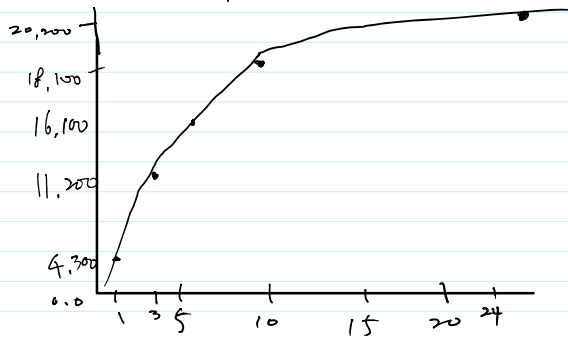
시간	소득
0-1	4,300
1-3	6,900
3-5	4,900
5-10	21,000
10-24	2,100

질문 : 4시간 이하 계급을 통틀어 소득 합계는 얼마?

시간	소득	승환시간	누적소득
0	0	0	0
0-1	4,300	1	4,300
1-3	6,900	3	11,200
3-5	4,900	5	16,100
5-10	21,000	10	37,100
10-24	2,100	24	39,200

p. 75

⇒ 누적소득 그래프 그리기 (선 그래프 활용)



※※ 선택 사항 ※※

이해 부분은 강의에서 다루지 않았지만
지금까지의 내용을 이해하였다면
충족할 수 있다.

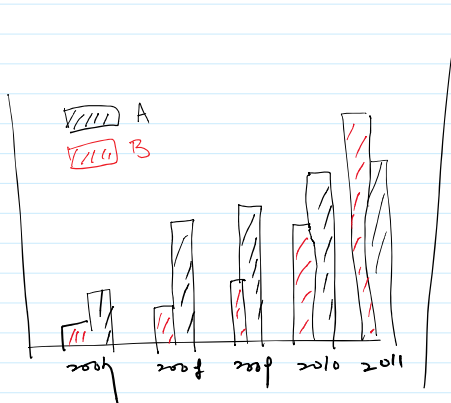
적절한 차트 고르기

P. 79

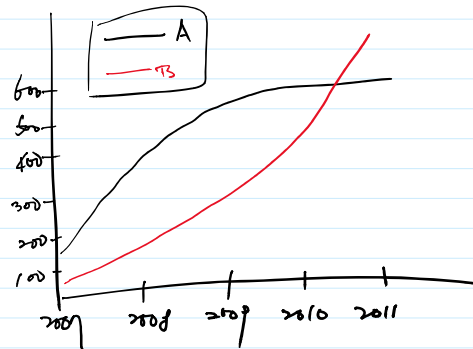
= 어떤 그래프를 선택하는 것이 따라 주어진 데이터를 가장 잘
표현할 수 있는가.

예제 : A, B 두 회사의 연산량 비교

① 막대 그래프로 나타낼 때



② 선 그래프로 나타낼 때



VS.

질문: 어떤 방식 선택? 이유는?