

Chapter05

시각화

I 시각화

M T W T F S S

FASTCAMPUS
ONLINE

금융공학/퀀트 I

강사. 장순용

I 키포인트

- 시각화의 목적.
- 탐색적 시각화.
- 시각화 유형별 용도.
- 착시현상.

I 시각화의 목적

- 가설간의 비교를 목적.
- 인과관계, 상응관계, 구조적 관계를 보여주는 목적.
- 다변량 데이터를 요약해서 보여주는 목적.
- 분석 결과를 요약해서 보여주는 목적.
- 기초 데이터와 결과를 뒷바침하는 증거를 정리하여 보여주는 목적.
- 결국은 콘텐츠 (스토리)의 유/무가 시각화의 효과를 결정한다!

I 탐색적 시각화

- 데이터의 통계적 특성을 알아본다.
- 데이터에 패턴이 있는지 육안으로 확인해 본다.
- 향후 분석 방향을 정하는데 도움이 된다.
- 결과를 보고하는 목적으로도 사용된다.

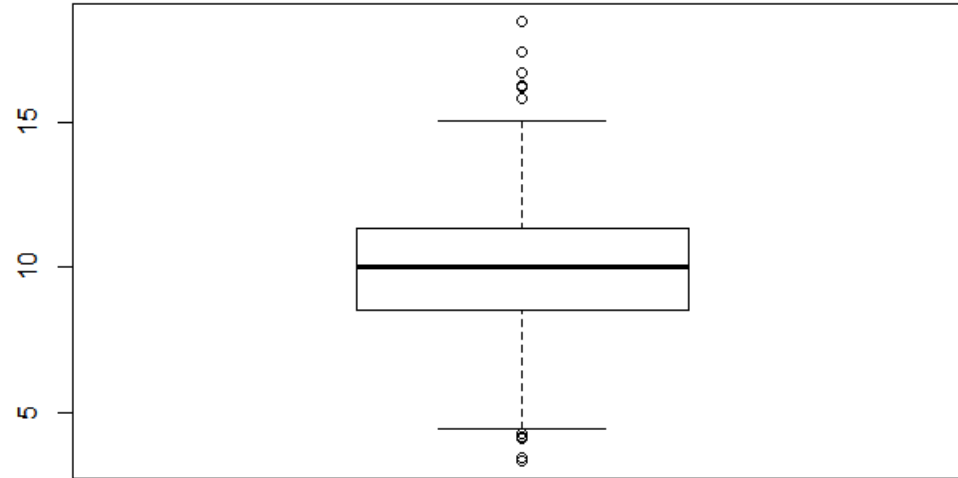
I 탐색적 시각화의 특징

- 많은 노력을 들이지 않고 짧은 시간안에 완성.
- 많은 수의 그래프 생성.
- 분석가 본인의 직관적인 이해를 우선시 한다.
- 타이틀, 레이블 등은 크게 중요하지 않다.
- 컬러, 폰트 등은 꼭 필요할 때만 사용.

I 단변량 기술통계 요약의 용도

- 상자그림 (boxplot):

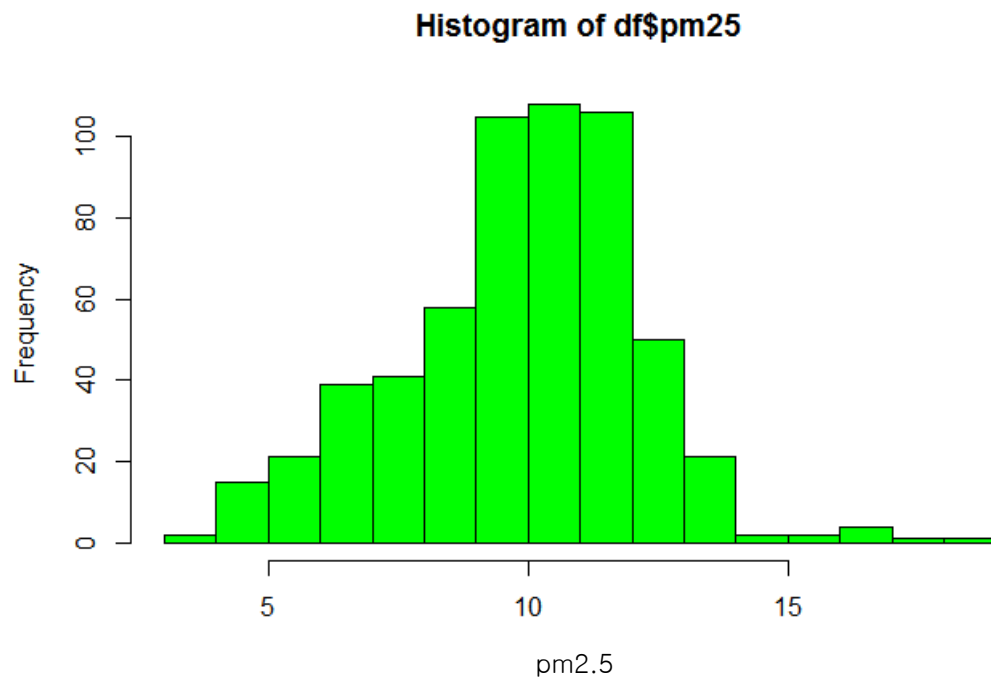
⇒ Box, whiskers, outlier로 구성됨. 연속형 변수 사용.



I 단변량 기술통계 요약의 용도

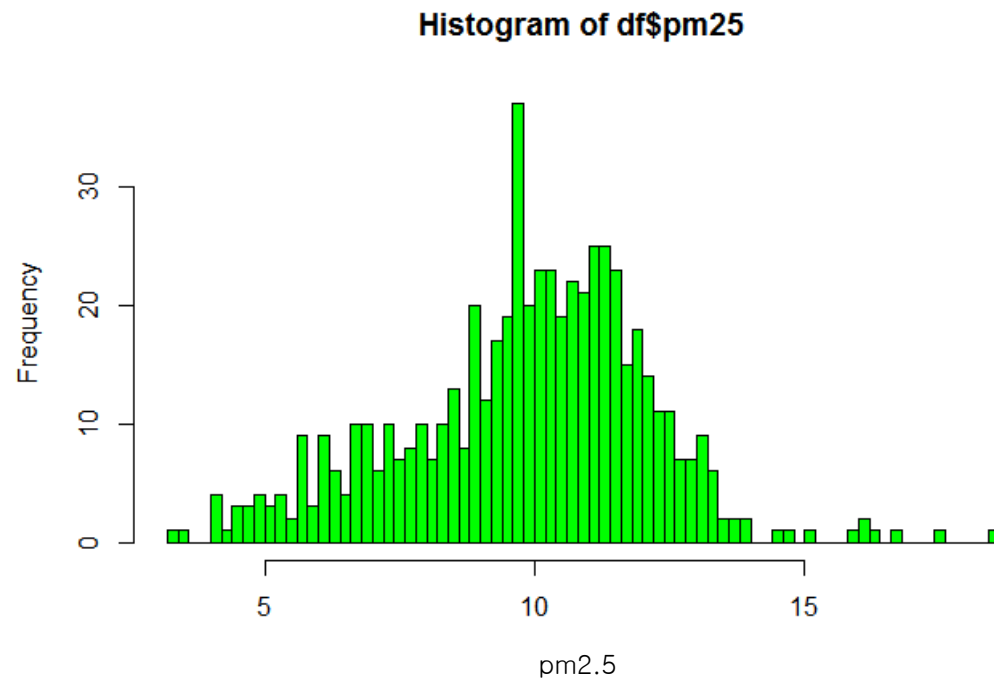
- 히스토그램 (histogram):

⇒ 계급에 해당하는 도수를 보여줌. 연속형 변수 사용.



I 단변량 기술통계 요약의 용도

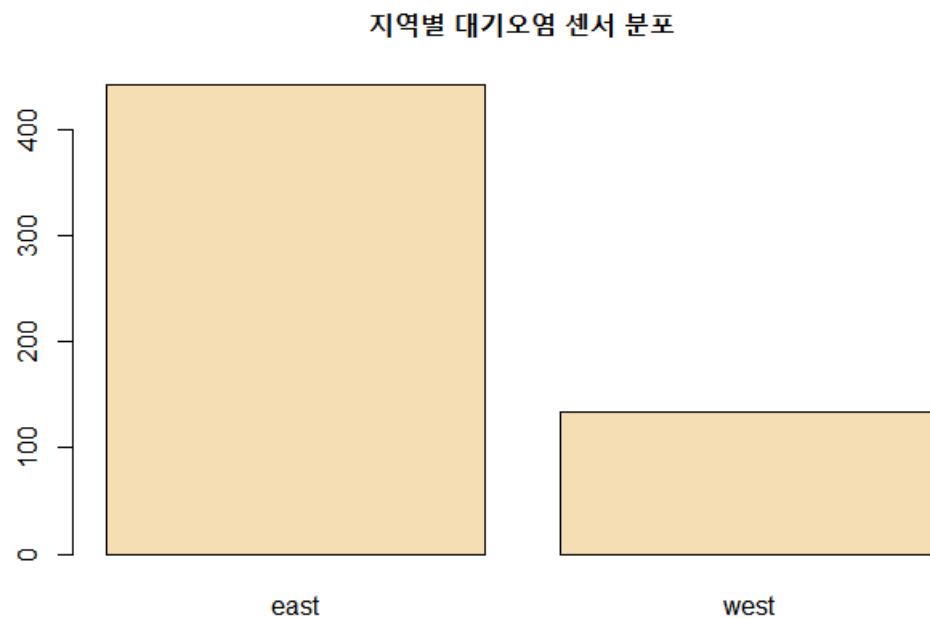
- 히스토그램 (histogram):
⇒ 계급의 크기를 조정할 수 있다.



I 단변량 기술통계 요약의 용도

- 막대그림 (bar chart):

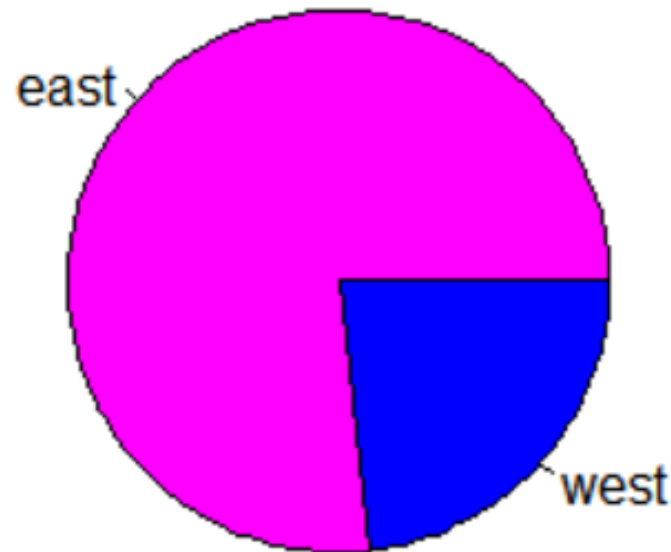
⇒ 명목형 변수의 도수를 보여준다.



I 단변량 기술통계 요약의 용도

- 파이 (pie):

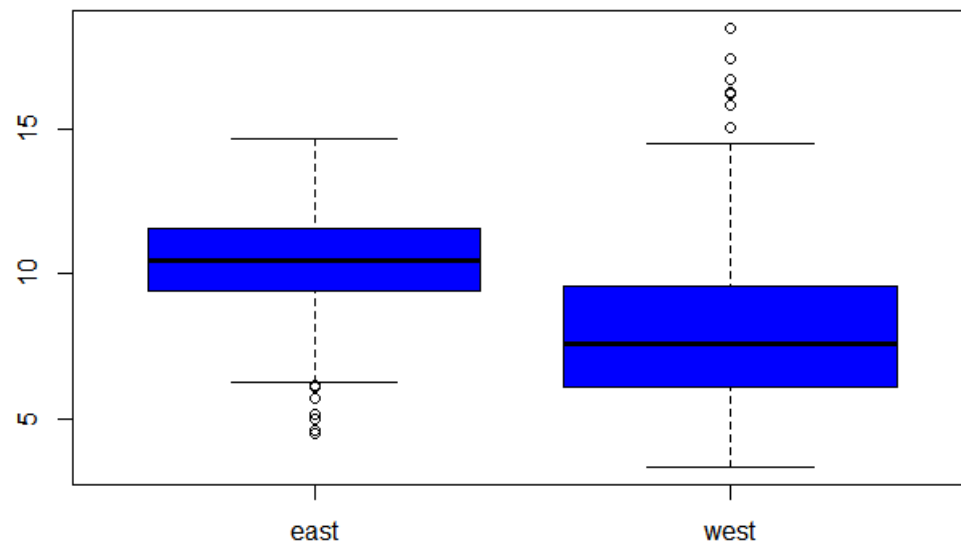
⇒ 명목형 변수의 상대도수를 보여준다.



I 비교, 상관관계 확인의 용도

- 다중 상자그림 (multiple boxplot):

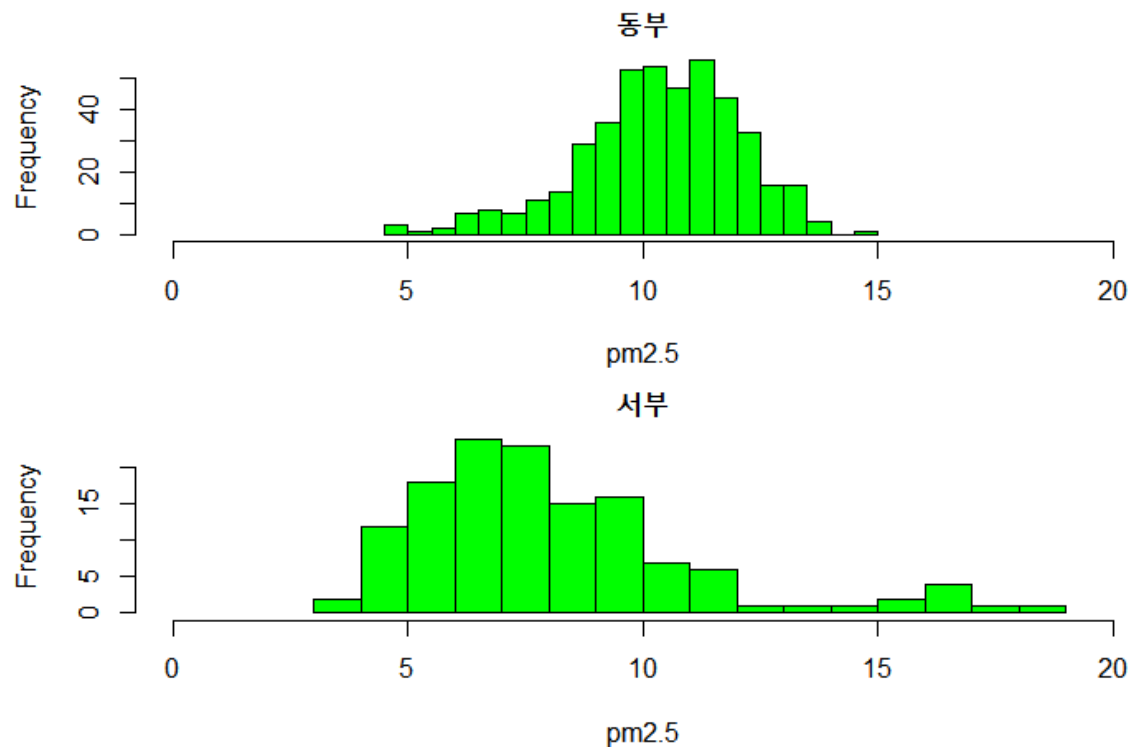
⇒ 제2의 **명목형** 변수에 의해서 여러 개의 상자그림이 생성된다.



I 비교, 상관관계 확인의 용도

- 다중 히스토그램 (multiple histogram):

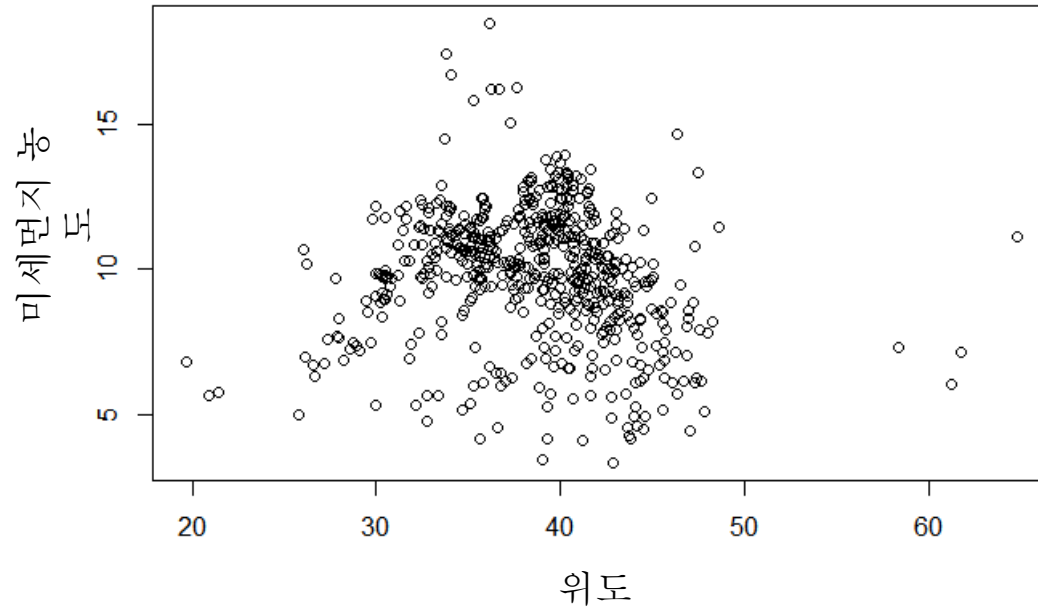
⇒ 제2의 **명목형** 변수에 의해서 여러 개의 히스토그램이 생성된다.



I 비교, 상관관계 확인의 용도

- 산점도 (scatter plot):

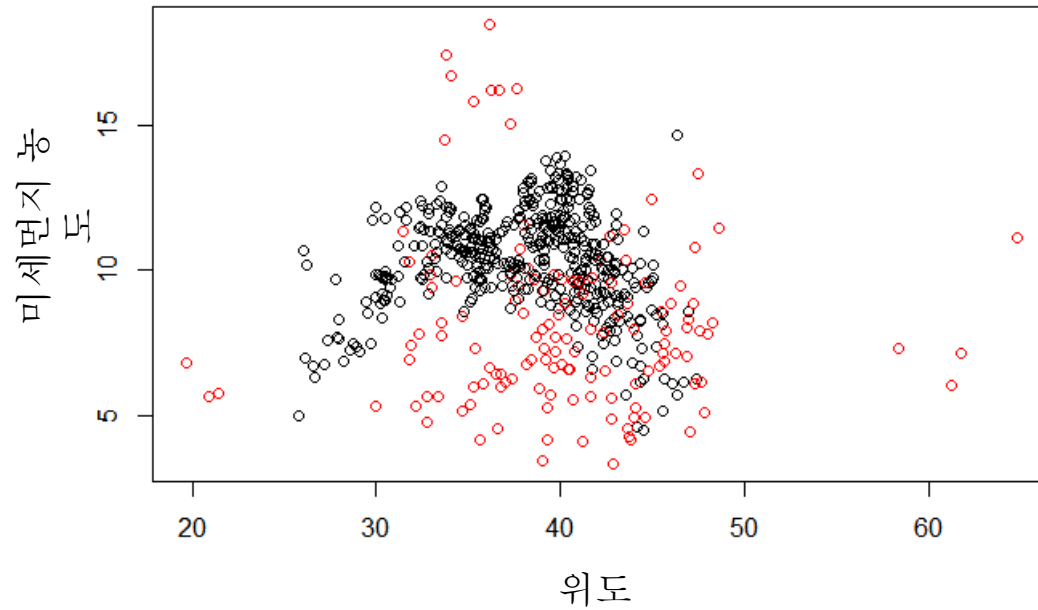
⇒ X 와 Y 두 개의 변수 사이의 관계를 보여준다.



I 비교, 상관관계 확인의 용도

- 산점도 (scatter plot):

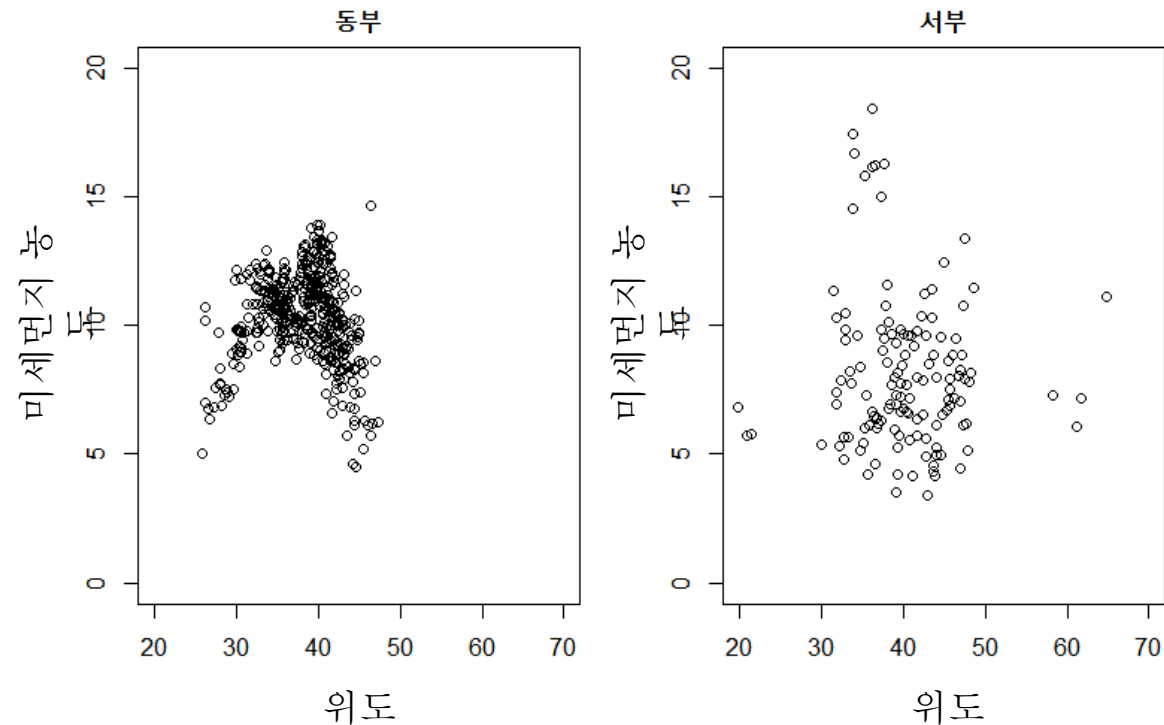
⇒ 컬러로 제3의 명목형 변수의 유형을 보여줄 수도 있다.



I 비교, 상관관계 확인의 용도

- 다중 산점도 (multiple scatter plot):

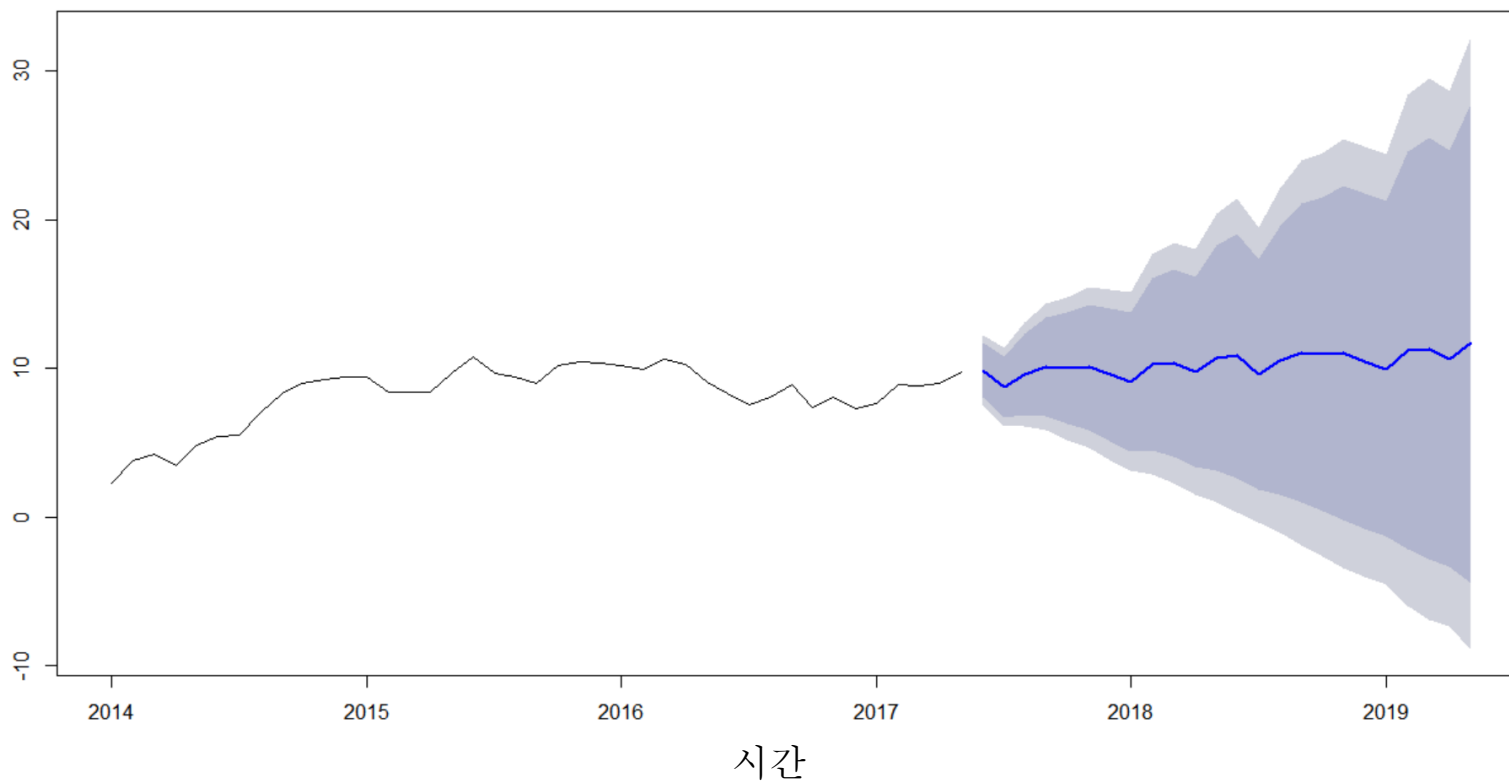
⇒ 제3의 **명목형** 변수에 의해서 여러 개의 산점도가 생성된다.



I 시계열

- 시계열 그래프:

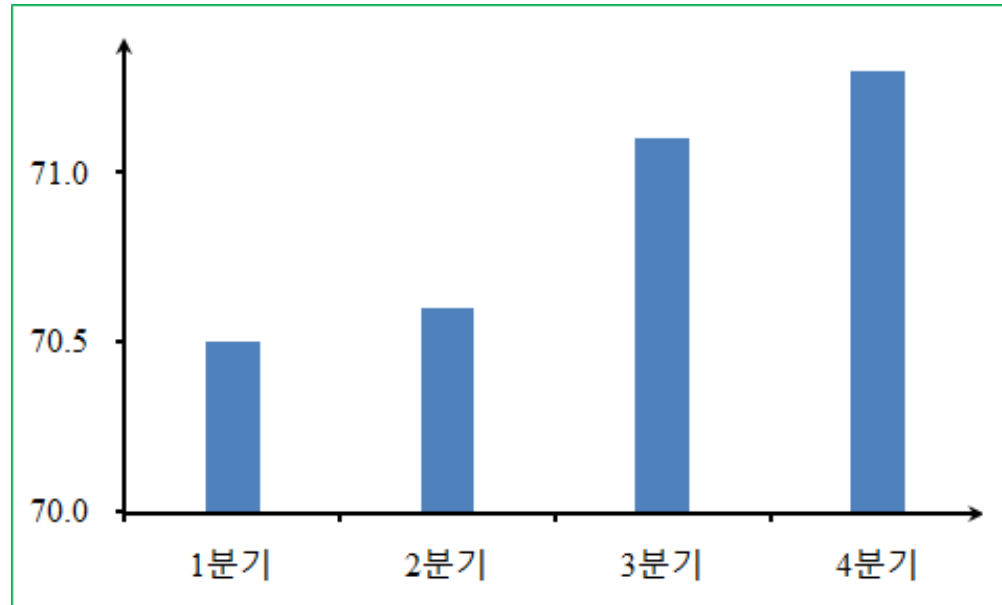
⇒ 시간이 지남에 따라서 변하는 수치를 나타냄. 예측 신뢰구간 표현.



I 착시현상

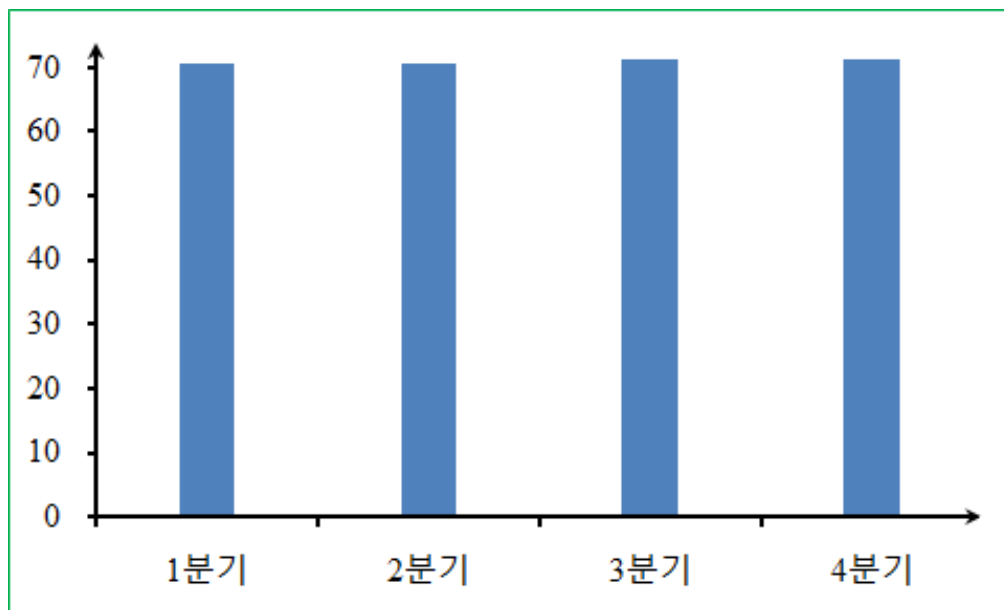
- 다음은 A사의 분기별 매출 현황을 나타낸 막대그림이다.

⇒ 큰 폭의 매출 증가?



I 착시현상

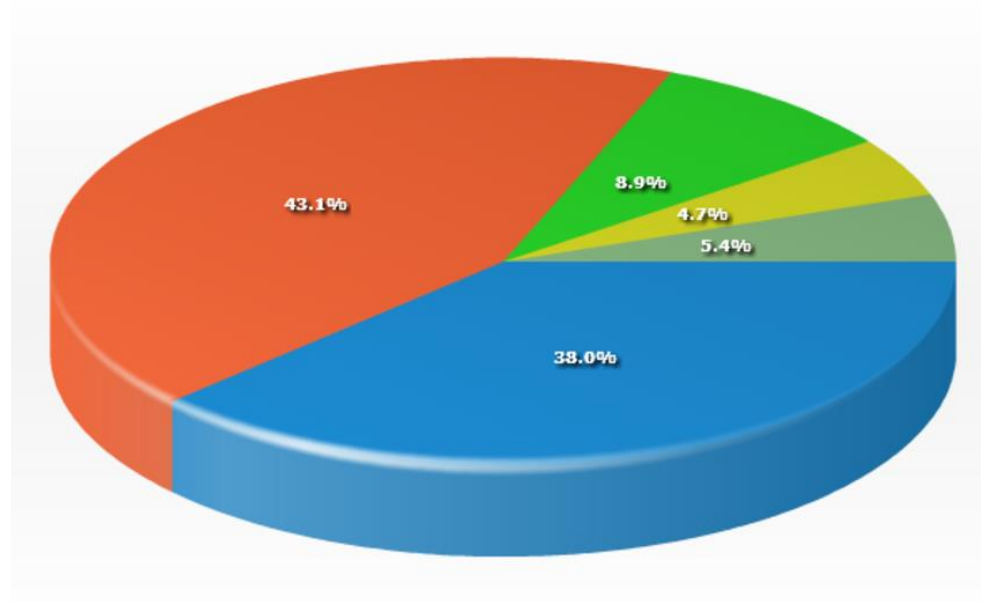
- 다음은 A사의 분기별 매출 현황을 나타낸 막대그림이다.
- ⇒ 세로축 스케일을 조정하여 전체를 본다. 대략 2% 매출 증가.



I 착시현상

- 파이 그래프는 주의하여 사용한다.

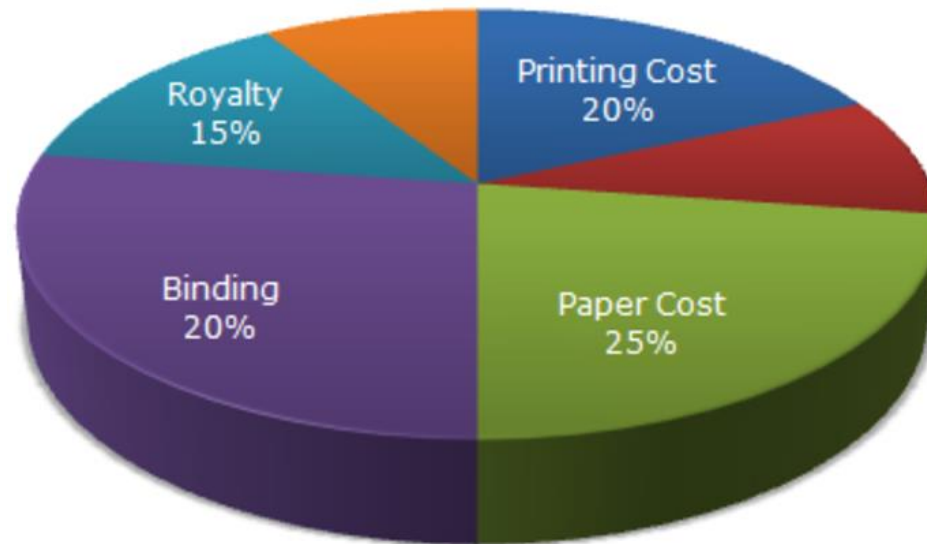
⇒ 특히 3D 파이 그래프는 원근법 때문에 **큰 착시** 현상 발생!



I 착시현상

- 파이 그래프는 주의하여 사용한다.

⇒ 특히 3D 파이 그래프는 원근법 때문에 **큰 착시** 현상 발생!



I 끝.

감사합니다.

