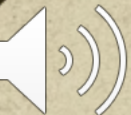


데이터-정보 리터러시

- 책, 자료, 통계에 대한 우리의 생각과 태도

국민대학교 객수범

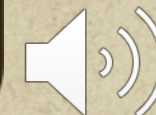
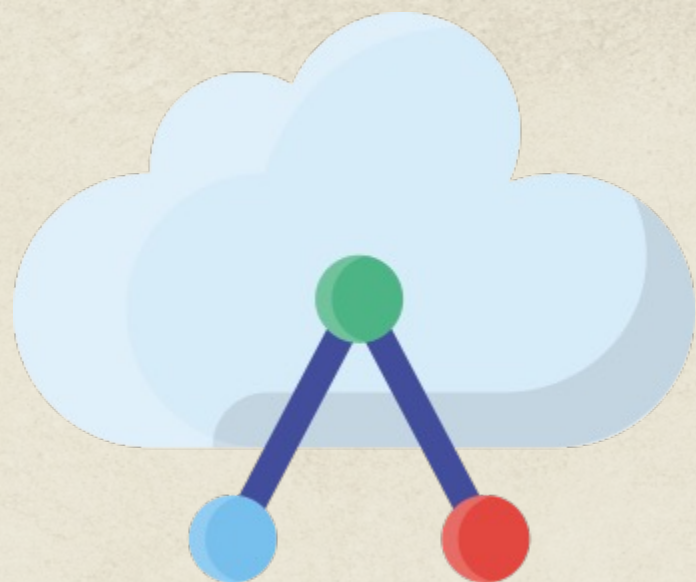


통계

자연과학: 자연현상의 객관적 사실과 법칙을 발견하는 학술 분과

사회과학: 자연과학 방법을 인간 사회의 다양한 주제, 문제에 적용하는 학문

- 사회과학의 대표적인 도구는 바로 데이터와 통계



통계

식당 수익과 옐프Yelp 별점 사이의 상관관계

- 별점 평점 1점이 오르면 매출액이 5%, 크게는 10% 가까이 상승
- 소비자의 후기나 별점이 좋으면 새로운 소비자의 구매 심리를 자극



통계

Likert Scale Representation



Strongly
Disagree



Disagree



Neutral



Agree



Strongly
Agree



통계

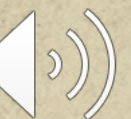


vs



통계

- 설문/평가 참여자는 몇 명인가? (표본 크기)
- 별점 분포는 어떠한가? (평균과 점수 분포)
- 메뉴에 따른 평가 결과 차이가 있는가?
(특정 메뉴에 대한 평가가 유독 좋거나 나쁠 수 있음)
- 아웃라이어는 얼마나 존재하는가?
(습관적으로/다른 이유로 별점을 아주 박하게/후하게 매기는 사람 있음)



[그림] 쿠팡 직원 추정 리뷰어들의 행위 요약






쿠팡 픽업 후점 리뷰어 5명

마스크, 안전장갑, 티타늄 식도, 고양이 모래, 고속충전기 등 완전히 동일한 PB 상품을 유사날짜 구매 후 조직적으로 후기 작성

한달간 고양이모래 210L구매
평균 고양이 1마리 1달 5L 사용
"집에 고양이 가 몇 마리?"
무리 30마리 추정 


"칼이 너무 좋아요!"
"무더지면 재구매할게요!"
일주일 후 동일한 식칼 재구매

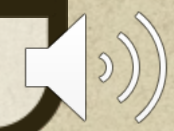
기이한 구매 행태네요...


한달간 10여회에 걸쳐 장갑
630매 구매, 모두 만족 
"사이즈가 꼭 맞아요"
동일제품인데도 구매 사이
S,M,L로 제각각? 

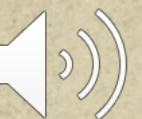
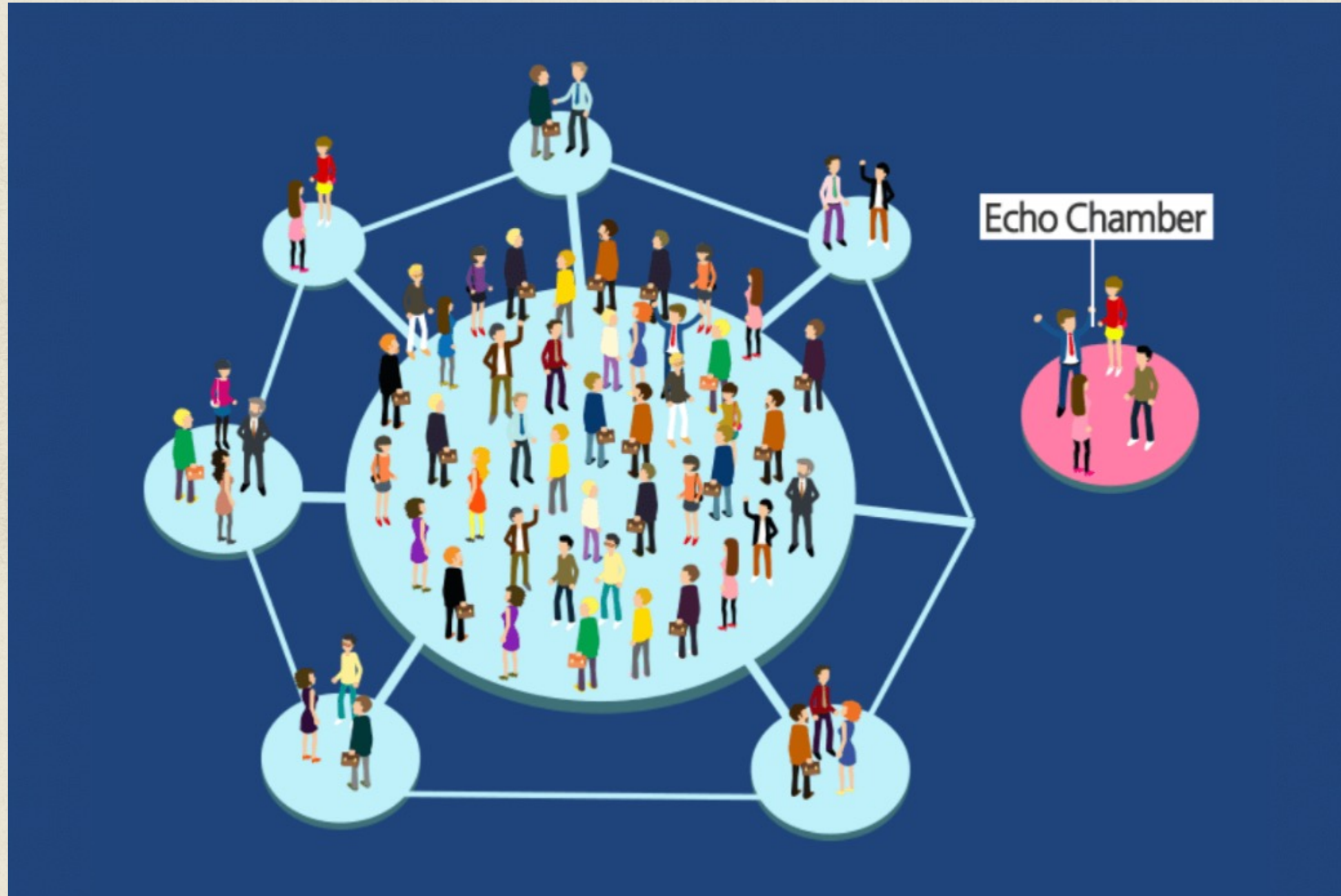
리뷰 아르바이트생의 허위 리뷰 작성 과정

- 인터넷 쇼핑몰에서 리뷰 조작 의뢰 업체의 상품 검색
- ↓
- 해당 상품 무통장입금 방식으로 구매
- ↓
- 가상계좌를 업체 측에 전달
- ↓
- 업체가 대신 상품 구매 비용 입금
- ↓
- 구매자에게 빈 박스 발송
- ↓
- 구매자는 가짜 상품 수령
- ↓
- 업체 측이 제시한 가이드라인에 맞춰 리뷰 작성
- ↓
- 업체는 구매자에게 수수료 지급

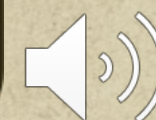
① 중앙일보



Echo Chamber



가짜 뉴스를 찾아라

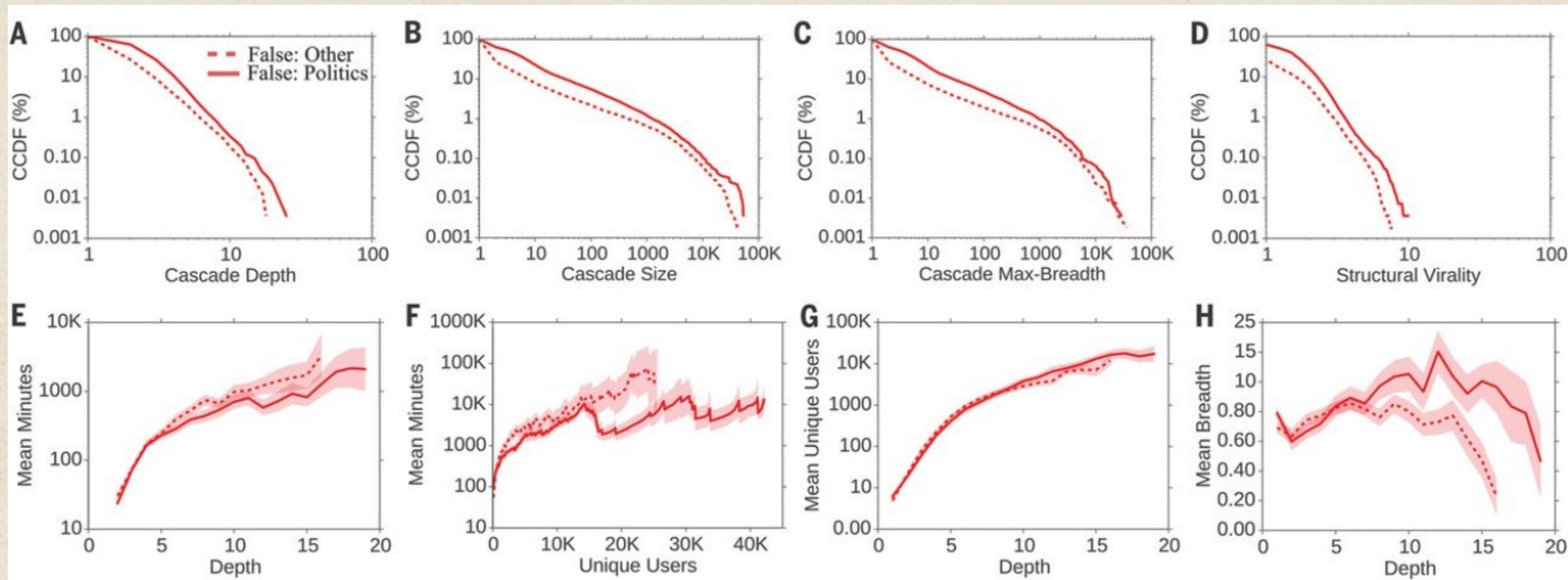


가짜 뉴스를 찾아라

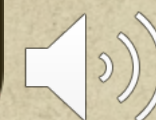
1. 맨땅에 갑자기 솟아오른 화산(Volcano suddenly springs up)
2. Ross라는 이름의 지구와 유사한 행성 발견
(Newly discovered Earth-like planet named "Ross")
3. 이제껏 알려지지 않은 비틀즈의 노래, 도서관 지하에서 발견
(Previously unknown Beatles song found in library basement)
4. 토끼를 돌봐 줄 자원봉사자 모집
(City looking for volunteers to cuddle rabbits)



사람들을 왜 가짜 뉴스를 만들까



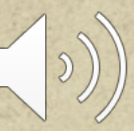
방문자 수, 팔로워 수, 광고비, 여론 조작, 선동,



상관 관계 / 인과 관계



미국 뉴욕주 초등학교 학생을 대상으로 연구한 결과 신발 사이즈가 클수록, 읽기 능력이 뛰어난 것으로 나타났다.

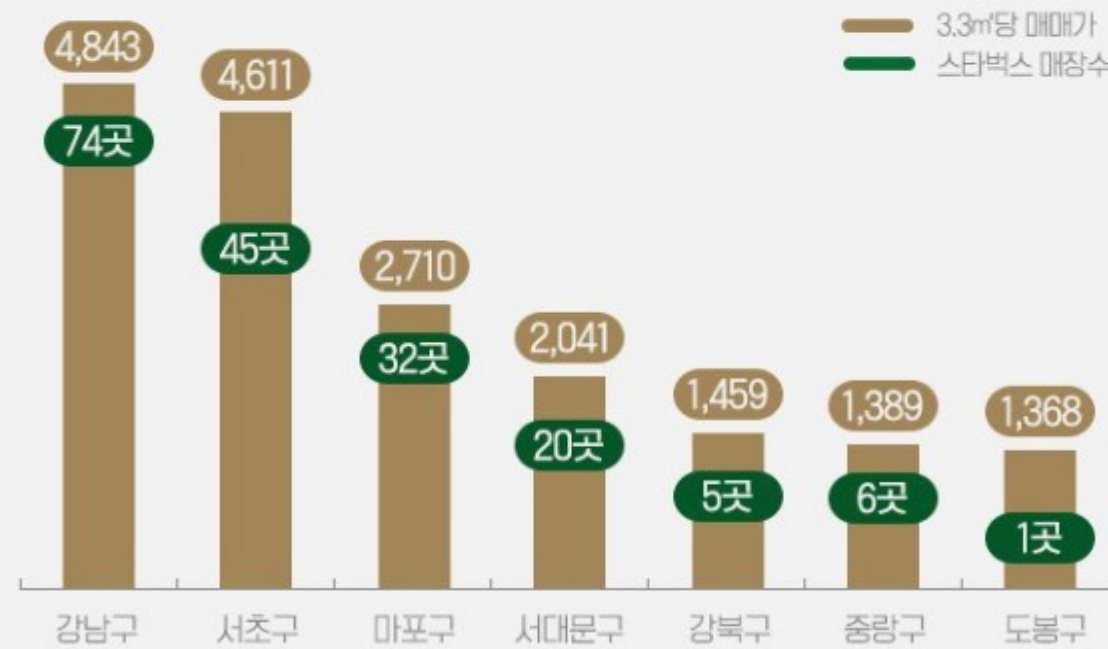


상관 관계 / 인과 관계

kt estate

스타벅스 많은 지역이 집값도 높아~

서울 주요 구 3.3㎡당 아파트 평균 매매가 및 스타벅스 매장수



* 출처: 시가정보통신(14), 매장수(스타벅스 홈페이지)

현재 3월 기준



〈기술 통계〉

Q: 성별에 따라 선호하는
독서 문종이 있는가?

	여학생	남학생
서사	63	67
사회	68	44
과학	29	29

결론: 여학생은 정보글(사회), 남학생은 서사글을 더 좋아한다.



〈기술 통계〉

Q: 성별에 따라 선호하는
독서 문종이 있는가?

	여학생	남학생
서사	63	67
사회	68	44
과학	29	29

결론: 여학생은 정보글(사회),
남학생은 서사글을 더 좋아한다.

〈교차분석(카이제곱 검정)〉

Q: 성별에 따라 선호하는
독서 문종이 있는가?

	여학생	남학생
서사	63	67
사회	68	44
과학	29	29

$$\chi^2 = 3.950, p = .149$$

결론: 여학생은 정보글(사회),
남학생은 서사글을 더 좋아하는 것처럼
보이지만 유의한 차이가 아니다.

