

국내 COVID-19 데이터 분석

11조

1828013 송혜준
1828027 조문선
1922043 이상은
1976076 김우정
1976099 김채원

01

서론

가설설정



02

본론

가설검정



03

결론

결론도출

01

서론

가설설정

02

본론

가설검정

03

결론

결론도출

1. 개인 감염보다 '집단감염' 비율이 더 높을 것이다.
2. '확진자 수'는 유동성이 높은 20대, 그리고 남성일 때 가장 클 것이다.
3. '확진자 대비 사망비율'은 면역력이 약한 노인층(80대 이상),
그리고 남성일 때 가장 클 것이다.
4. 누적 검사자 수와 누적 확진자 수는 상관관계가 있을 것이다.
5. 서울에서 유동인구가 많은 곳에서 집단감염이 더 많을 것이다.
6. 서울에서 기온이 낮을수록 감염 가능성이 더 클 것이다.
7. 교육 정책 변화에 따라 확진자 수에 유의미한 영향이 있을 것이다.
8. '코로나 바이러스' 검색량과 확진자의 수는 유의미한 관계가 있을 것이다.
9. 월별 확진자 수가 많을수록 월별 지하철 이용량이 적을 것이다.

01

서론

가설설정

02

본론

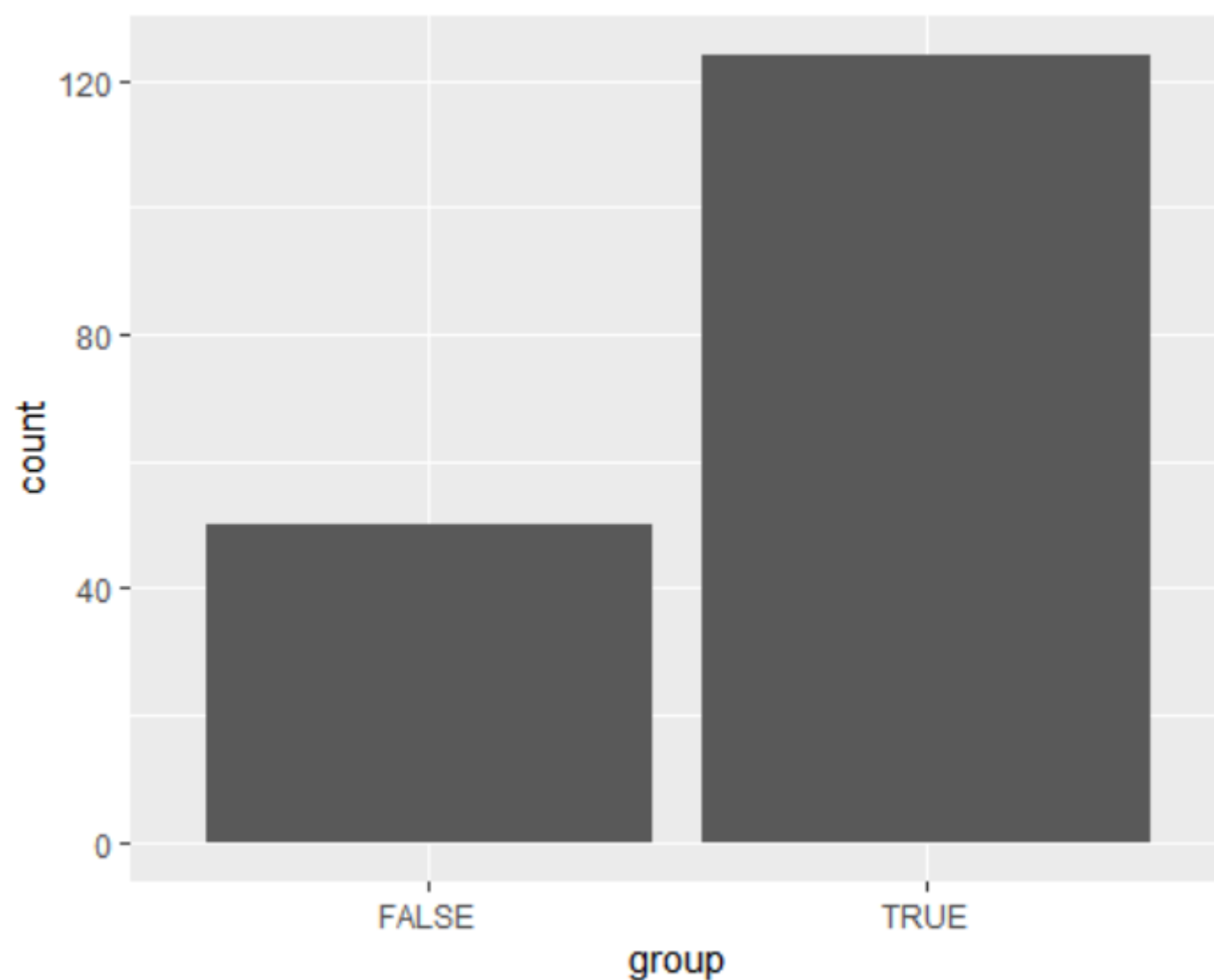
가설검정

03

결론

결론도출

1. 개인 감염보다 '집단감염' 비율이 더 높을 것이다.

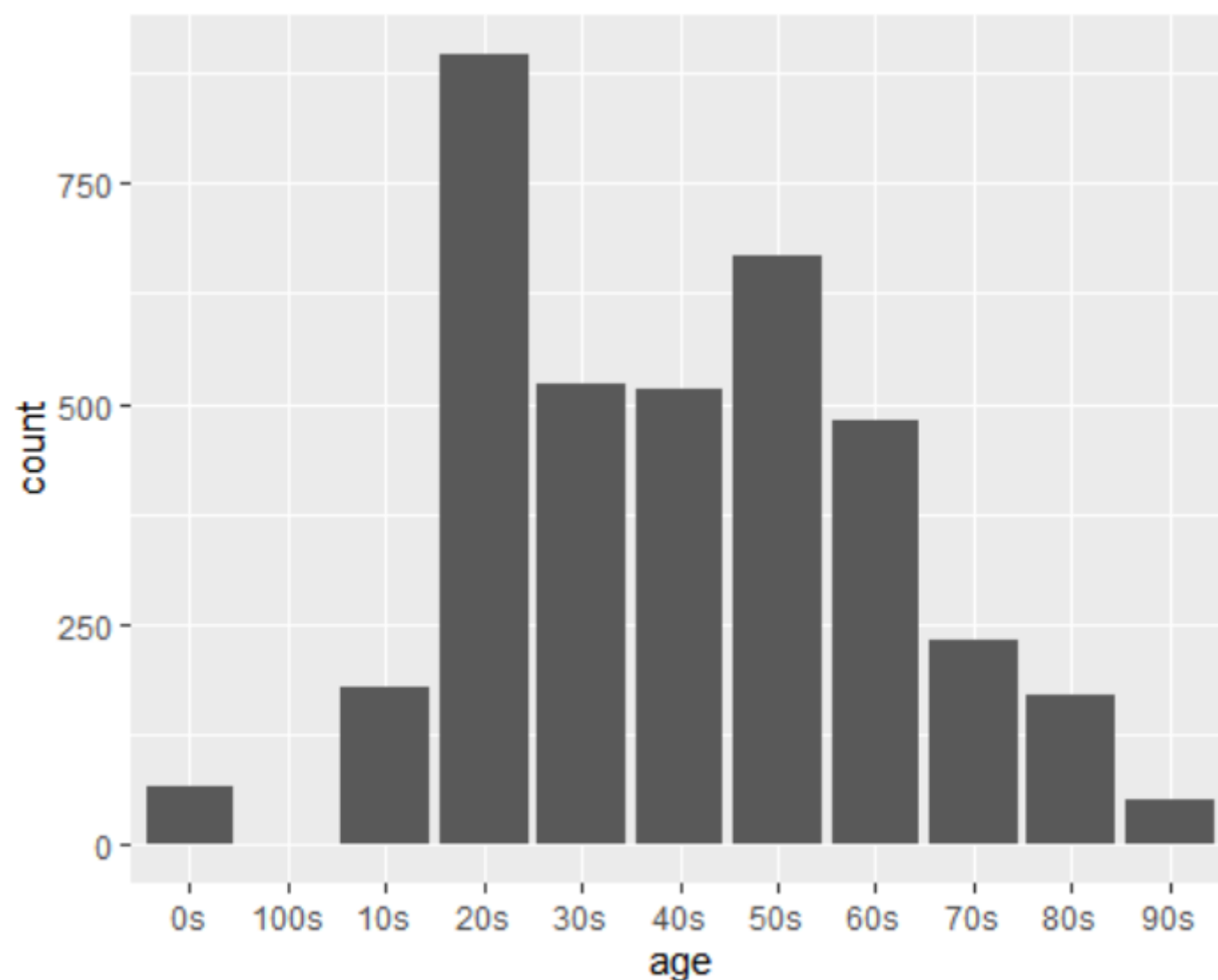


- group 변수가 TRUE일 때가 FALSE보다 더 많을 것이다.
TRUE : 집단감염 / FALSE : 개인감염

**-> 집단감염으로 감염된 사람이
개인감염으로 감염된 사람보다 많다.**

가설 수용

2. '확진자 수'는 유동성이 높은 20대, 그리고 남성일 때 가장 클 것이다.

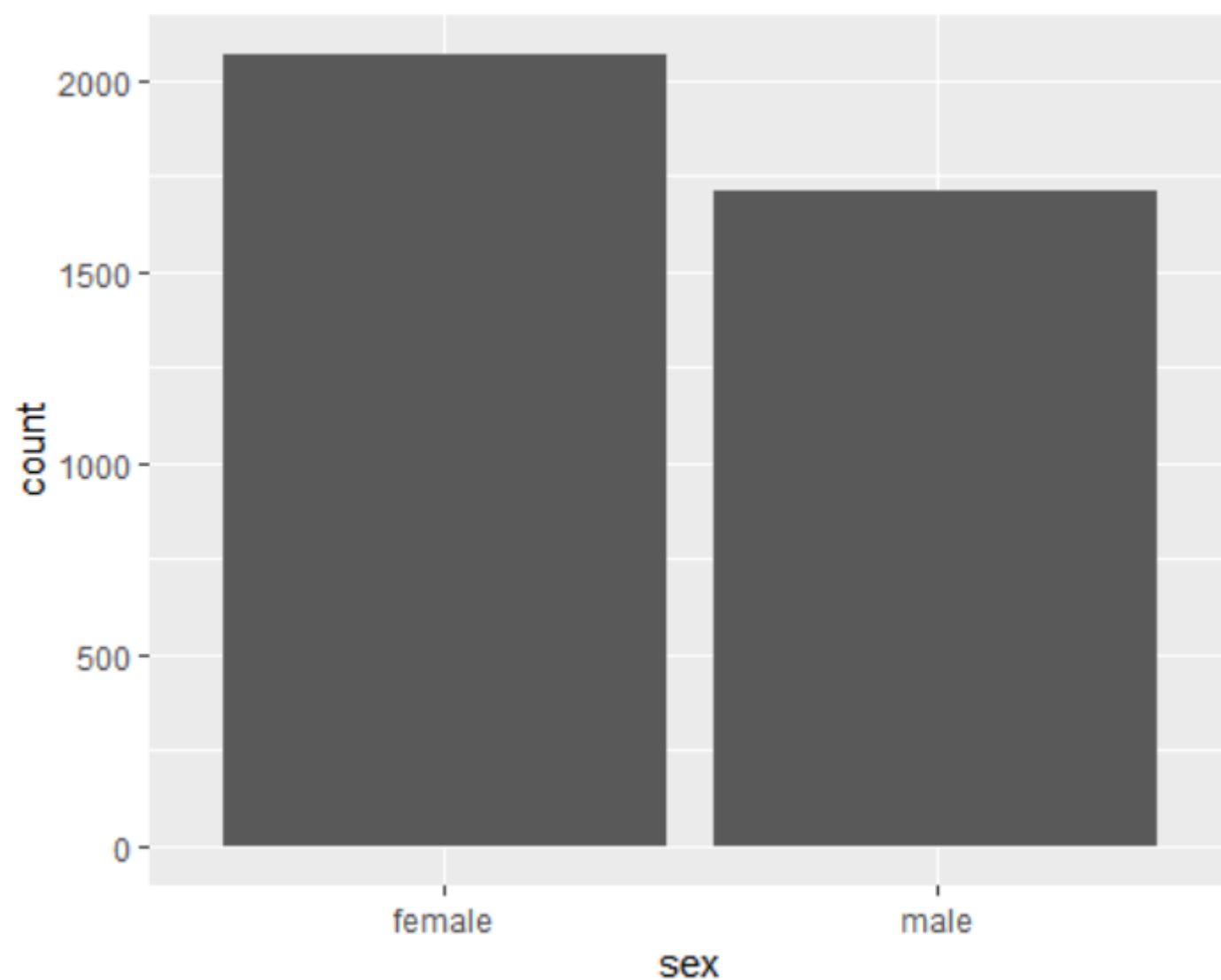


- age 변수가 20s일 때 발생빈도가 가장 높을 것이다.

-> **확진자 중
가장 많은 연령대는 20대이다.**

가설 수용

2. '확진자 수'는 유동성이 높은 20대, 그리고 남성일 때 가장 클 것이다.

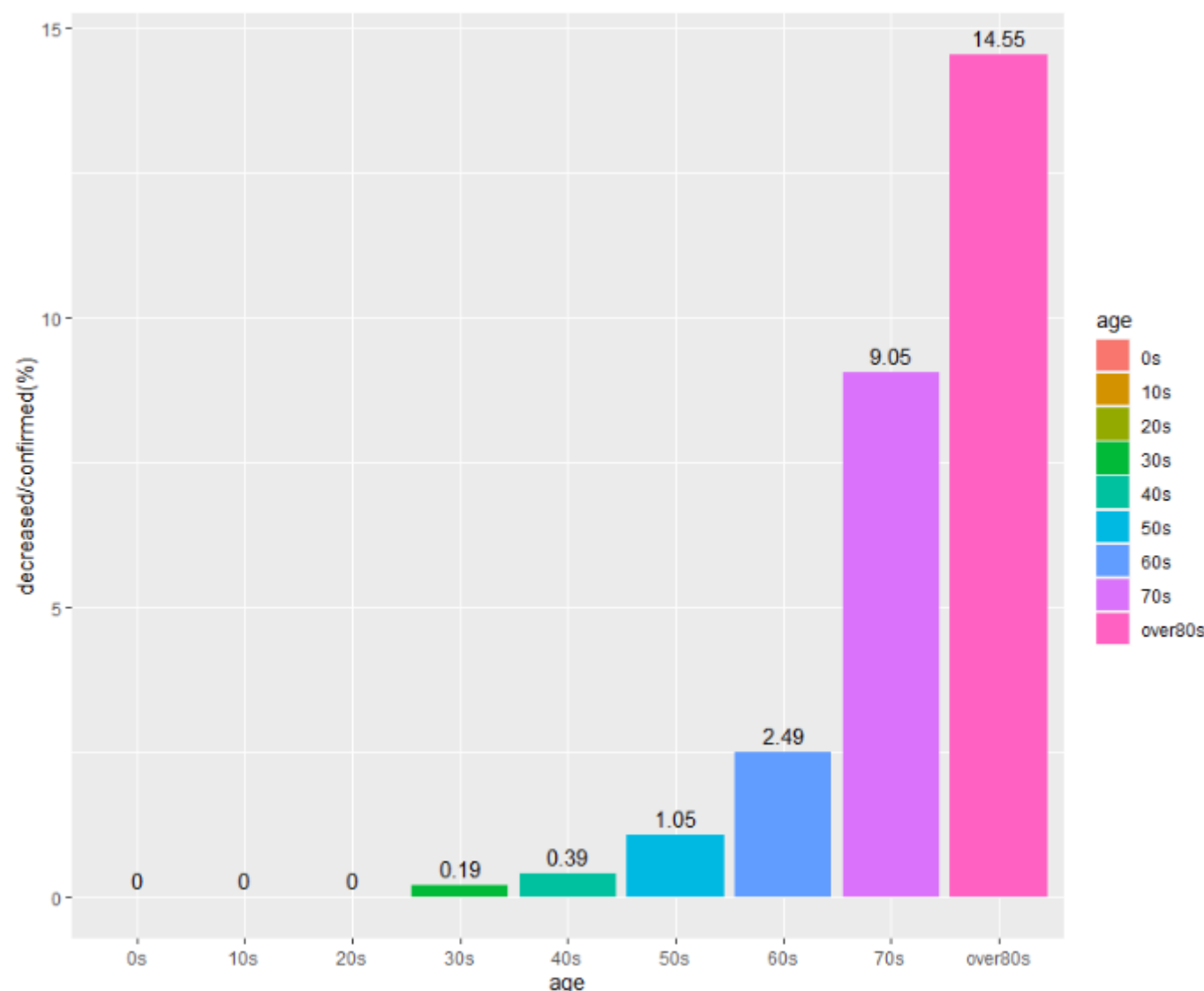


- sex 변수가 male일 때 female보다 발생빈도가 높을 것이다.

-> 확진자 중 여성이 남성보다 많다.

가설 기각

3. '확진자 대비 사망비율'은 면역력이 약한 노인층(80대 이상), 그리고 남성일 때 가장 클 것이다.

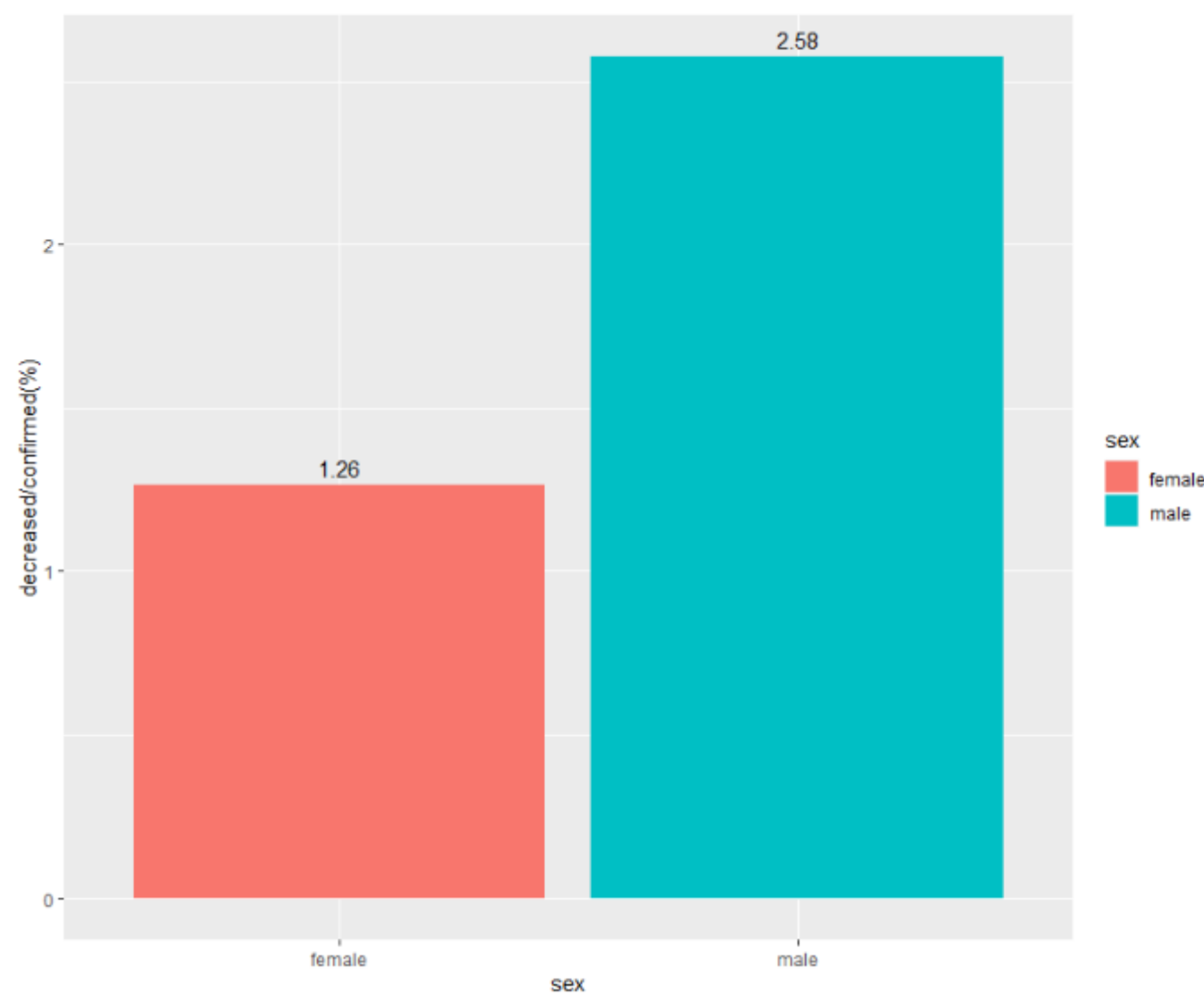


- 사망자 수/확진자 수 비율(%)'은
면역력이 약한 노인층(80대 이상)이 가장 클 것이다.

-> 80대 이상의
'사망자 수/확진자 수 비율(%)'이
가장 크다.

가설 수용

3. '확진자 대비 사망비율'은 면역력이 약한 노인층(80대 이상), 그리고 남성일 때 가장 클 것이다.

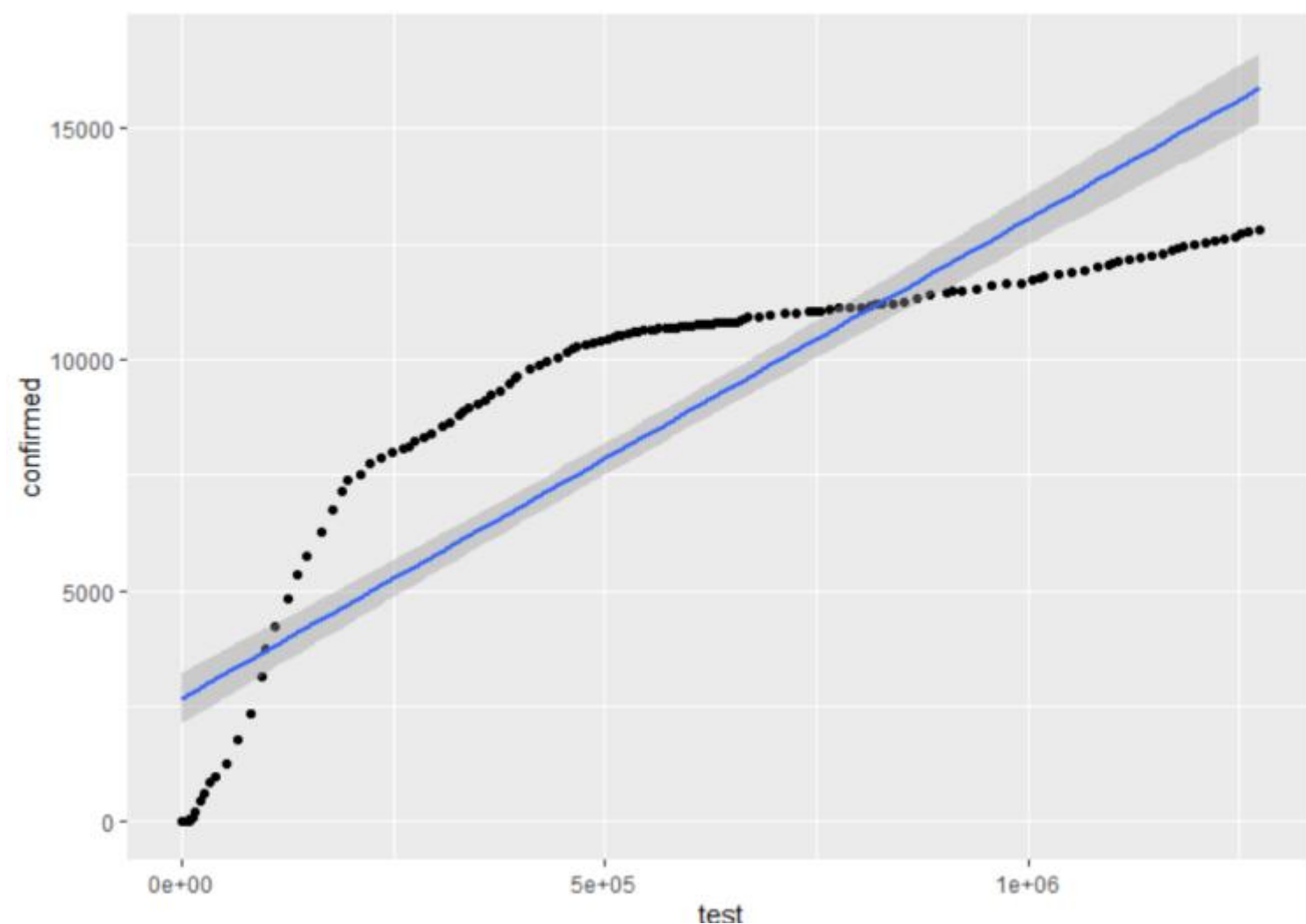


- 사망자 수/확진자 수 비율(%)은
면역력이 약한 노인층(80대 이상)이 가장 클 것이다.

-> 남성이 여성보다
'사망자 수/확진자 수 비율(%)'이
더 크다.

가설 수용

4. 누적 검사자 수와 누적 확진자 수는 상관관계가 있을 것이다.

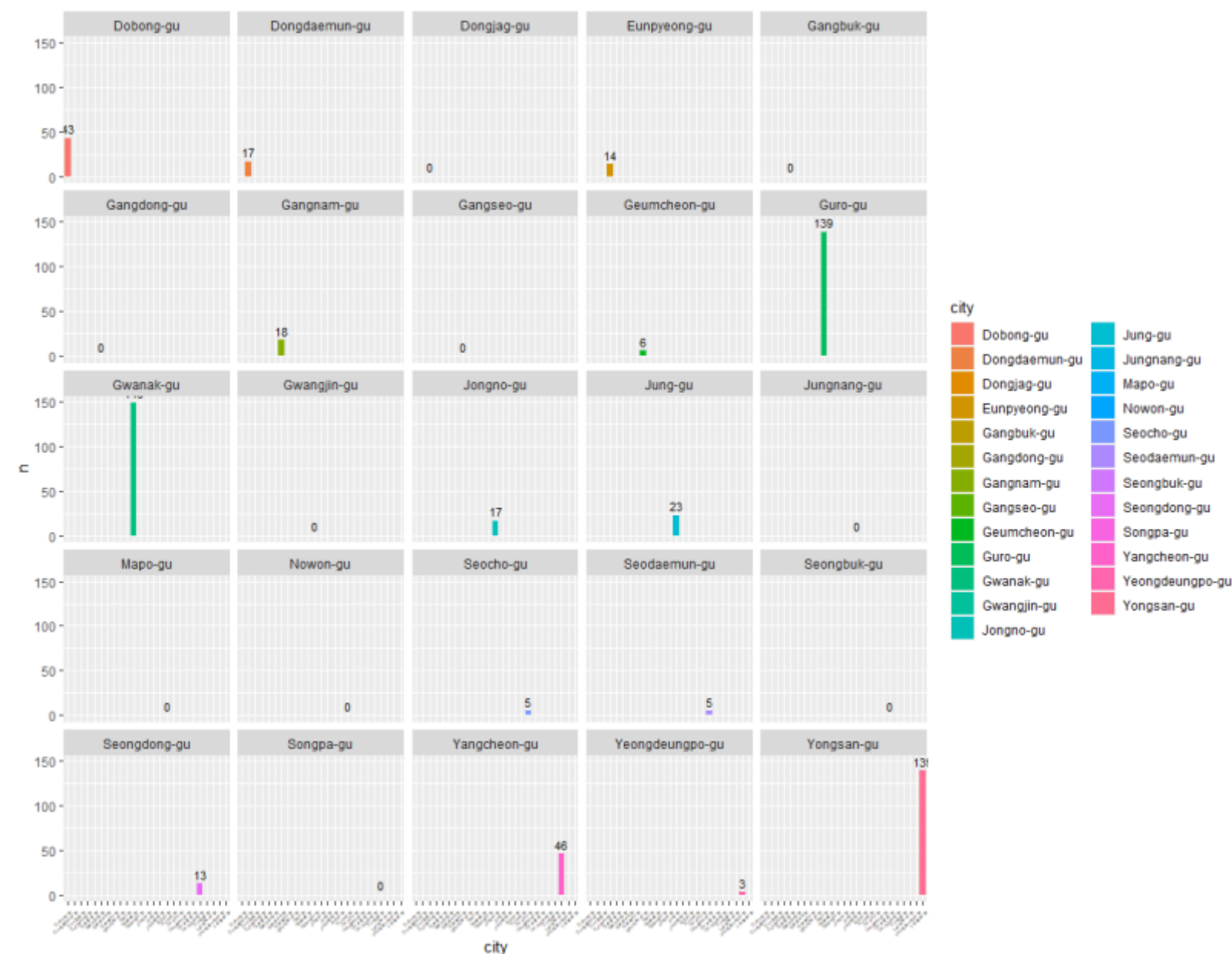
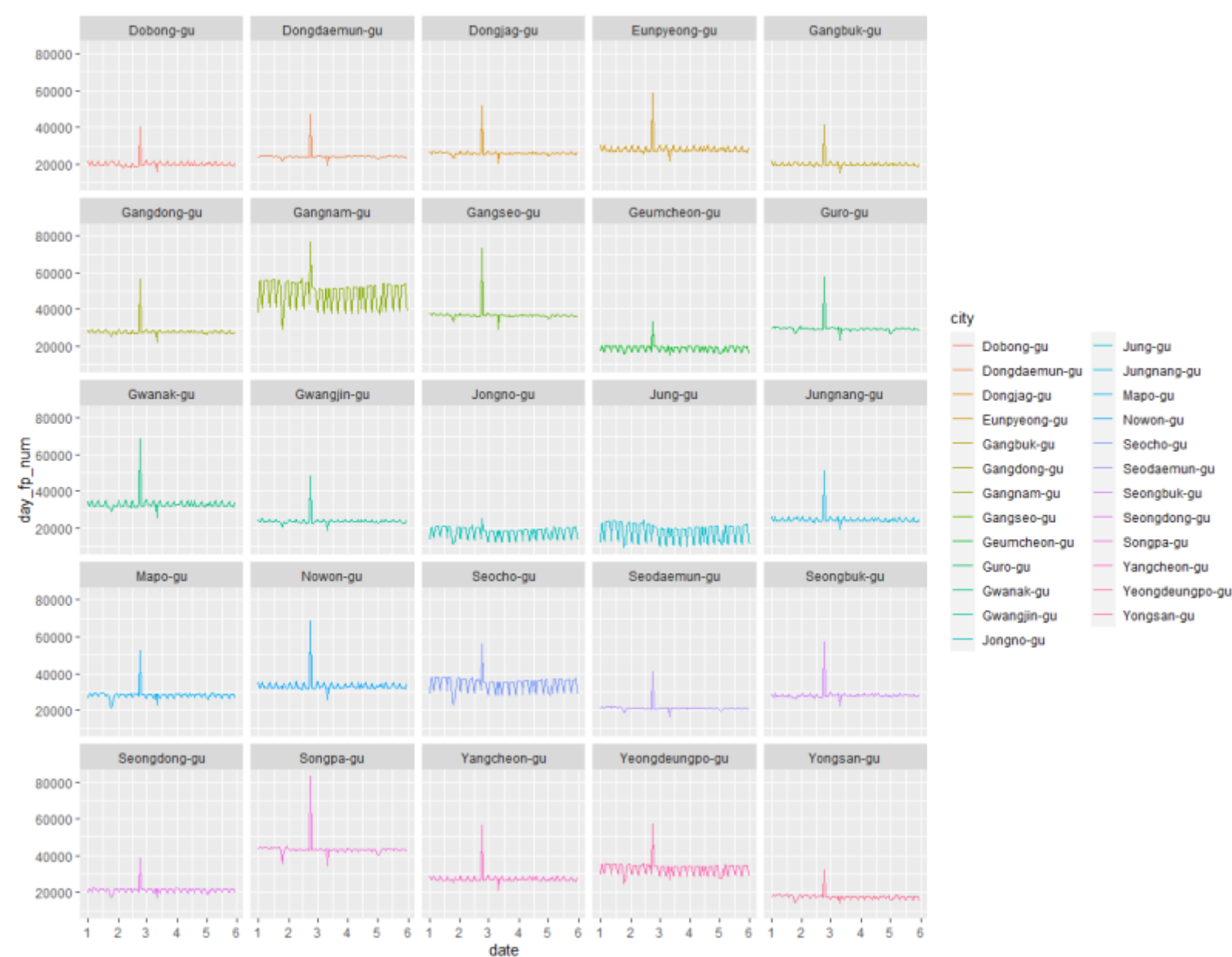


- 누적 검사수에 따른 누적 확진자 수의 산점도 및 회귀직선
누적 검사자 수가 증가할수록 누적 확진자 수가 증가하는 추세이다.

**-> 상관계수는 0.88로,
코로나바이러스 누적 검사자 수와
누적 확진자 수 간에
강한 양의 상관관계가 있음을 알 수 있다.**

가설 수용

5. 서울에서 유동인구가 많은 곳에서 집단감염이 더 많을 것이다.



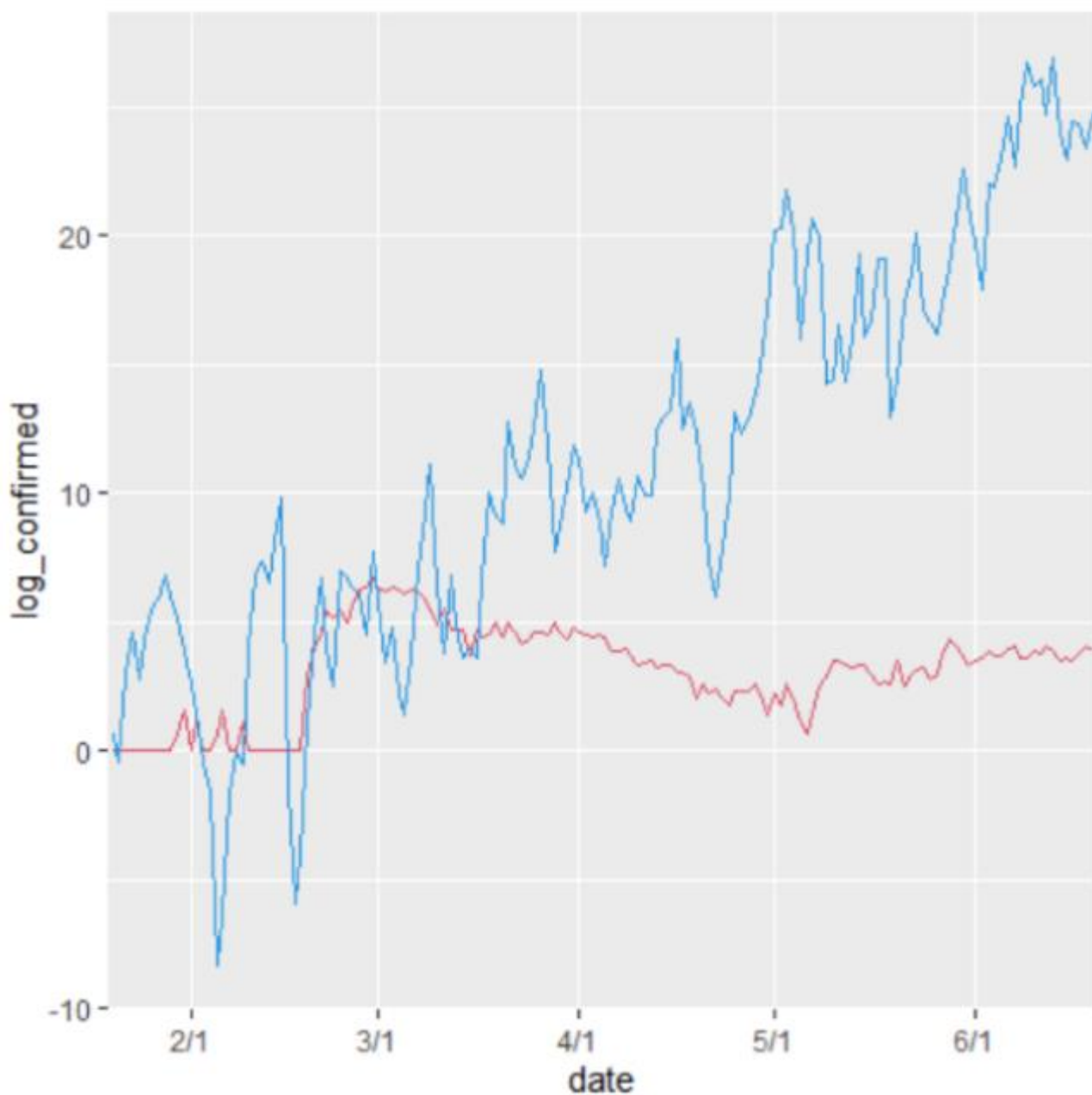
5. 서울에서 유동인구가 많은 곳에서 집단감염이 더 많을 것이다.

- 서울에서 유동인구가 많은 구가 집단감염이 더 많을 것이다.

**-> 강남구의 유동인구 수의 평균이 높지만, 집단감염 확진자는 적다.
반대로 용산구의 유동인구 수의 평균은 낮은 편이지만,
집단감염 확진자는 매우 높다.**

가설 기각

6. 서울에서 기온이 낮을수록 감염 가능성이 더 클 것이다.

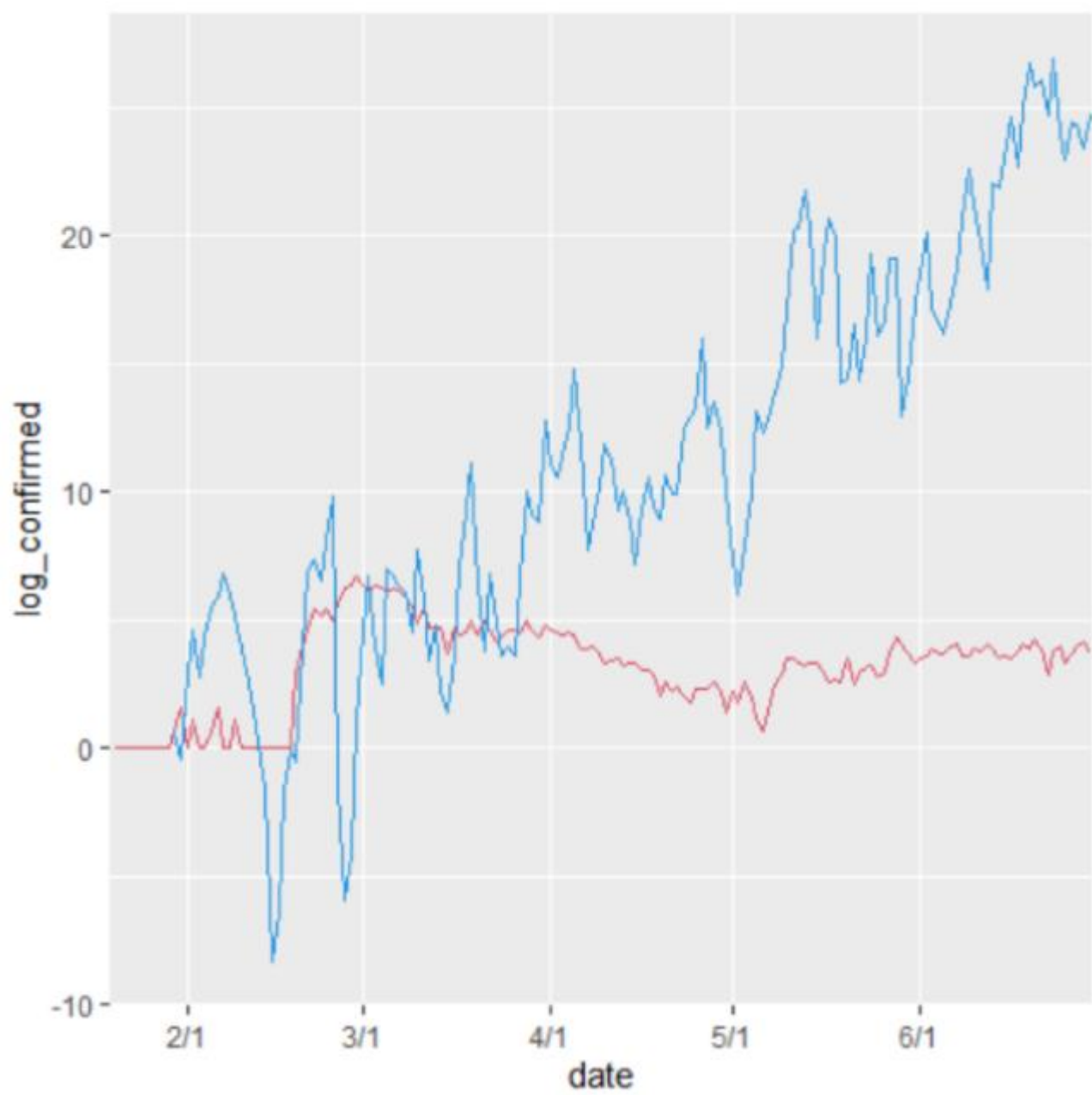


- Weather에서 서울의 avg_temp가 낮은 날일수록 확진자 수가 증가할 것이다.

— 시간에 따른 일일 확진자 수에 로그값을 취한 시계열 그래프
— 시간에 따른 일평균 기온을 나타낸 시계열 그래프

-> 일평균 기온과 일일 확진자 수 사이에는 일정한 패턴이 없다.

6. 서울에서 기온이 낮을수록 감염 가능성이 더 클 것이다.



- Weather에서 서울의 avg_temp가 낮은 날일수록 확진자 수가 증가할 것이다.

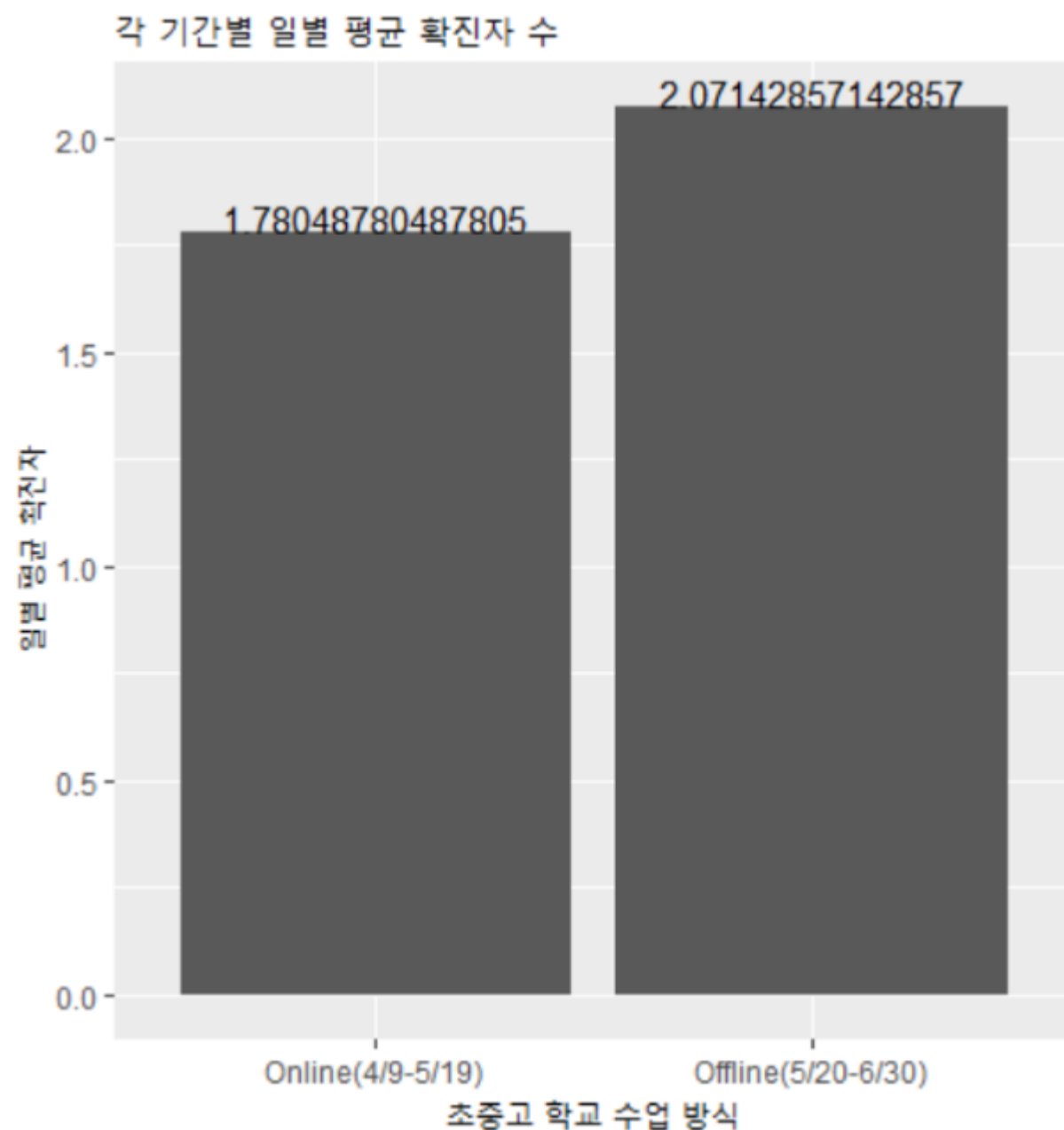
— 시간에 따른 일일 확진자 수에 로그값을 취한 시계열 그래프

— 10일 전 시간에 따른 일평균 기온을 나타낸 시계열 그래프

**-> 10일 전 일평균 기온과
일일 확진자 수 사이에는
큰 상관관계가 없다.**

가설 기각

7. 교육 정책 변화에 따라 확진자 수에 유의미한 영향이 있을 것이다.



- 교육 정책(온라인 등교)에 따라 확진자 수에 유의미한 영향이 있다.

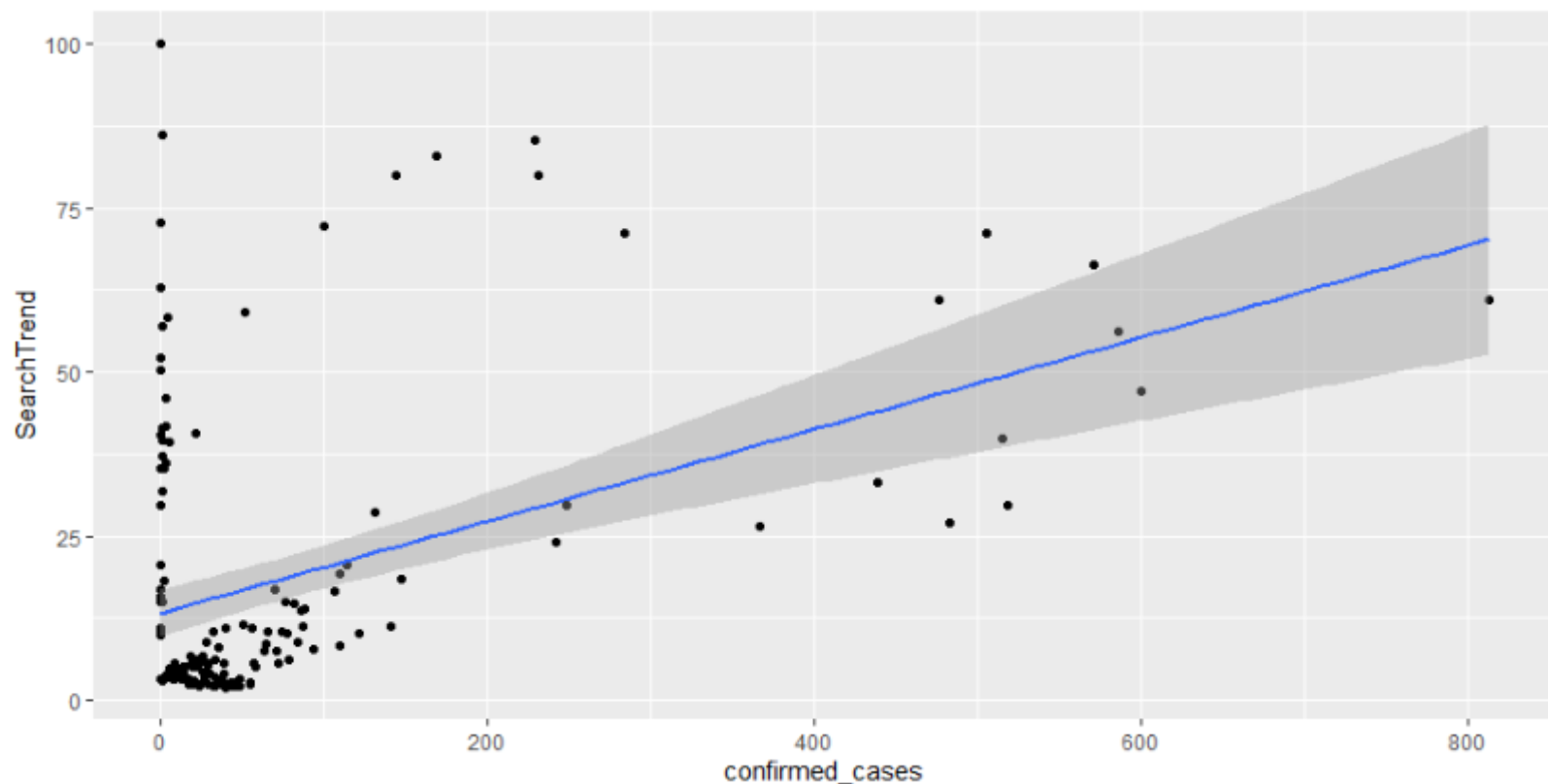
온라인 등교 기간(41일) / 오프라인 등교 기간(42일)

각 기간별 일평균 10대 확진자 수

-> 온라인 등교 정책이 시행되지 않았을 때의 일평균 10대 확진자 수가 더 많다.

가설 수용

8. '코로나 바이러스' 검색량과 확진자의 수는 유의미한 관계가 있을 것이다.

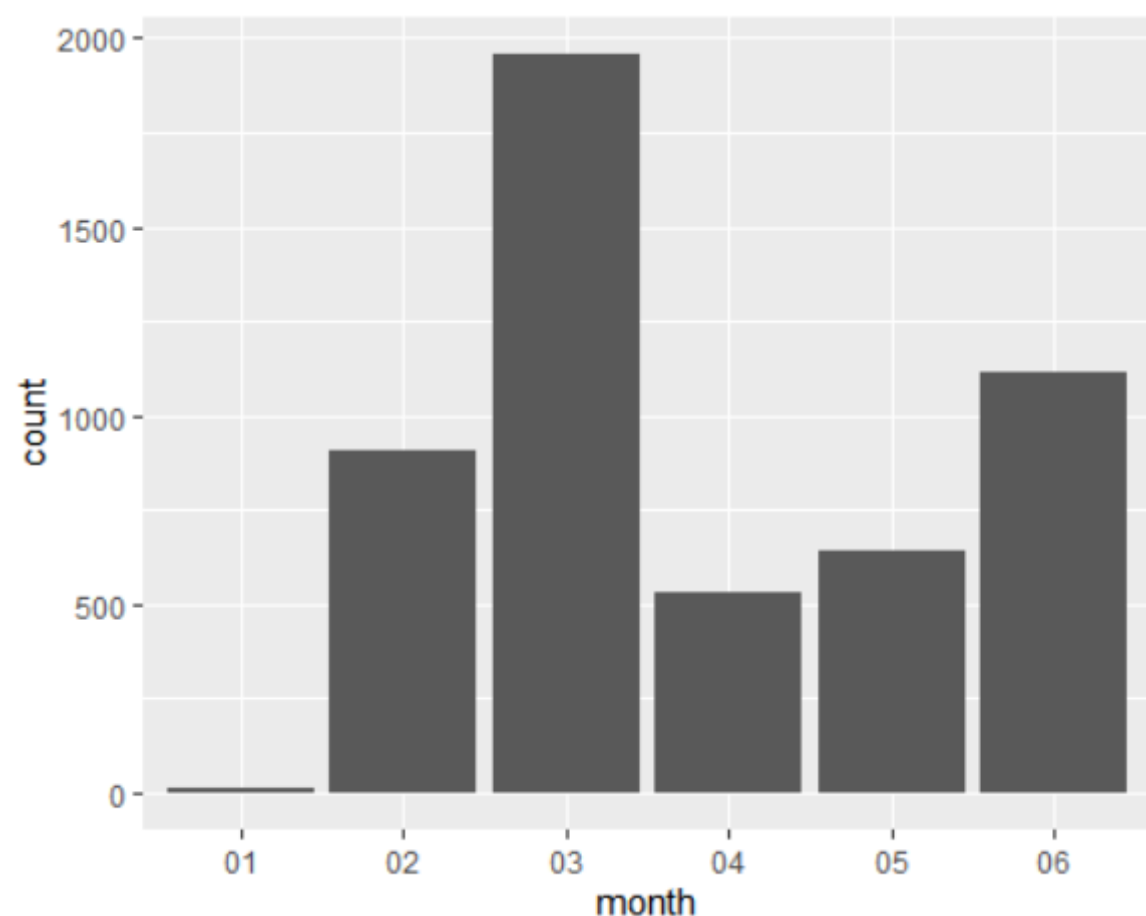


- 확진자 수에 따른 검색량의 산점도 및 회귀선
확진자가 없을 때도 검색량은 많았지만
확진자가 증가할수록 검색량 또한 증가하는 추세이다.

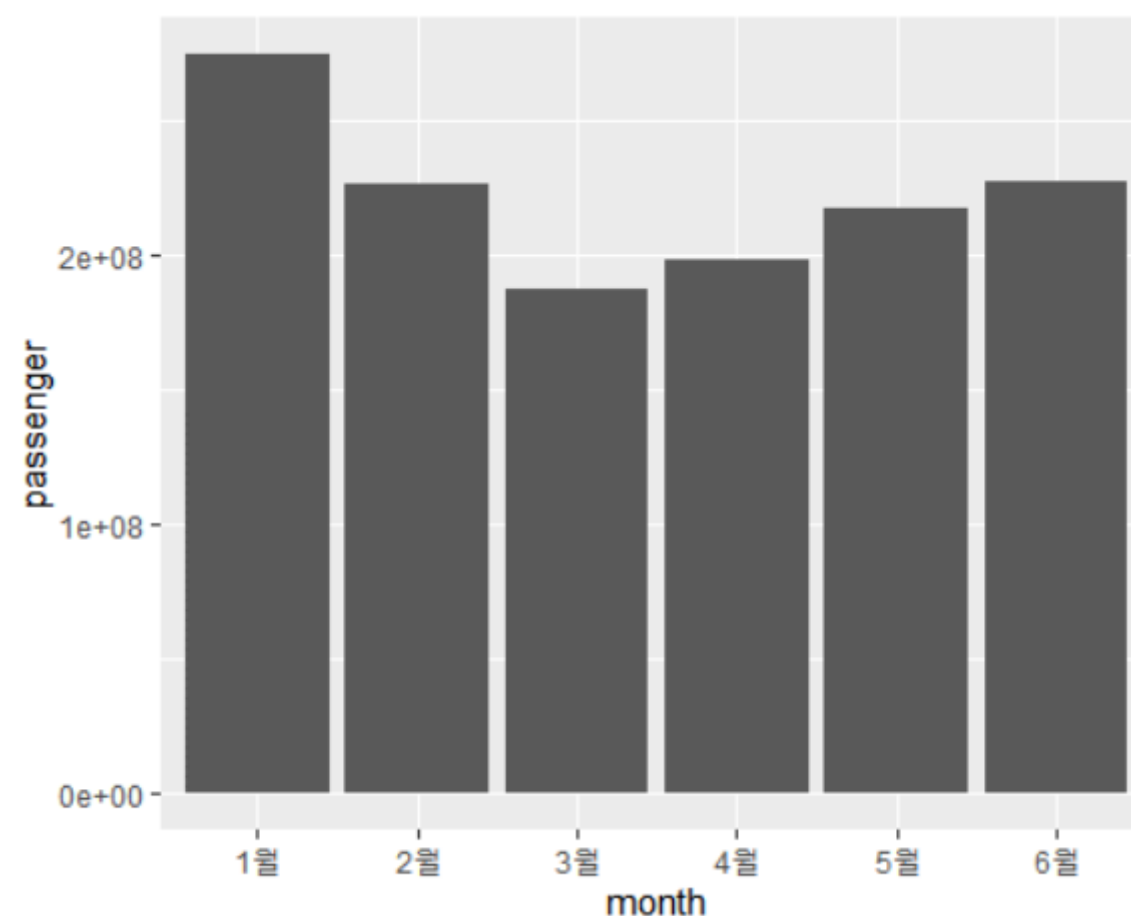
-> 상관계수는 0.42로, 확진자 수와 '코로나 바이러스' 검색량 간에 약한 양의 상관관계가 있음을 알 수 있다.

가설 수용

9. 월별 확진자 수가 많을수록 월별 지하철 이용량이 적을 것이다.



월별 확진자 수



월별 지하철 이용량

- 확진자가 제일 많은 3월에 지하철 이용량이 제일 적다. 그러나 4월은 6월에 비해 확진자가 적지만, 지하철 이용량이 6월에 비해 적다.

**-> 월별 확진자 수에
월별 지하철 이용량이
비례하지는 않는다.**

가설 기각

01

서론

가설설정

02

본론

가설검정

03

결론

결론도출

1. 개인 감염보다 '집단감염' 비율이 더 높다.
2. '확진자 수'는 유동성이 높은 20대, 그리고 여성일 때 가장 크다.
3. '확진자 대비 사망비율'은 면역력이 약한 노인층(80대 이상), 그리고 남성일 때 가장 크다.
4. 누적 검사자 수와 누적 확진자 수는 상관관계가 있다.
5. 서울에서 유동인구가 많은 곳에서 집단감염이 더 많은 것은 아니다.
6. 서울에서 일평균 기온과 일일 확진자 수 사이에는 큰 상관관계가 없다.
7. 교육 정책 변화에 따라 확진자 수에 유의미한 영향이 있다.
8. '코로나 바이러스' 검색량과 확진자의 수는 유의미한 관계가 있다.
9. 월별 확진자 수에 월별 지하철 이용량이 비례하지는 않는다.

빅데이터를 이용한 통계그래픽스

감사합니다!