

# 데이터 시각화 경진대회

우평식 팀

21600402 안수현,  
21700675 정지훈, 21700589 이찬효



# COVID-19 + 유동인구 데이터 시각화 및 분석



주어진 데이터 해석



데이터 시각화



시각화된 데이터를  
바탕으로 분석

# 신종 코로나 바이러스감염증

중국 우한시 등 확산 중

국내 감염자 발생에 대비하고,  
지역사회 전파를 예방하기 위해서  
국민 여러분의 동참이 필요합니다!!!

자세히보기 —



## 코로나 바이러스의 진행 과정

중국 우한  
최초 발생

2020.01.25, 호남교육신문

YTN 주요뉴스

**[속보]**

**국내 신종 코로나 환자  
총 15명으로 늘어**

2020.02.01, YTN 속보

# 코로나 바이러스의 진행 과정

국내 확진자  
발생 및 확산





# 활용한 데이터 목록

유동인구 데이터 (서울 지역, 1 ~ 6월)	데이터 패키지
코로나 확진자 정보	데이터 패키지
검색 트렌드	데이터 패키지
인구밀도	서울시 열린 데이터 광장
서울시 지도 데이터	(주)지오서비스 GIS Developer

# 데이터 해석 과정 - 유동인구 데이터

Column Name	포함하고 있는 데이터	활용
etl_dt	데이터셋에 추가된 시간	일괄 처리되어 있는 값이므로 필요없음
base_ymd	기준 년 월 일	월별, 주별, 일별 분류시 기준
tmzon_se_code	24시간대 구분코드	X
sexdstn_se_code	성별구분코드	성별 유동인구 추세 관측
agrde_se_code	5세단위 연령대 구분코드	연령별 유동인구 추세 관측
adstrd_code	행정동 코드	지역 구분시 행정동 코드를 매핑
popltn_cascnt	유동인구 수	유동인구
etl_date	처리 일자	일괄 처리되어 있는 값이므로 필요없음

# 데이터 해석 과정 - 코로나 확진자 정보

Column Name	포함하고 있는 데이터	활용
patient_id	환자번호	X
sex	성별	X
age	연령대	X
country	국적	X
province	특별시-광역시-도	서울 지역만 추출할 때 활용
city	시-군-구	서울의 자치구 구분시 활용
infection_case	감염 케이스	X
infected_by	감염원	X
contact_number	접촉자수	결측 값이 많아 활용하기 어려움
symptom_onset_date	증상발현날짜	결측 값이 많아 활용하기 어려움
confirmed_date	확진날짜	결측 값이 많아 활용하기 어려움
release_date	완치날짜	결측 값이 많아 활용하기 어려움
deceased_date	사망날짜	결측 값이 많아 활용하기 어려움
state	상태	결측 값이 많아 활용하기 어려움

# 데이터 해석 과정 - 검색 트렌드

Column Name	포함하고 있는 데이터	활용
date	날짜	16년 1월 1일 ~ 20년 6월 29일, 기준 값
cold	감기 검색 비율	사람들의 관심도를 알 수 있음
flu	독감 검색 비율	사람들의 관심도를 알 수 있음
pneumonia	폐렴 검색 비율	사람들의 관심도를 알 수 있음
coronavirus	코로나바이러스 검색 비율	사람들의 관심도를 알 수 있음

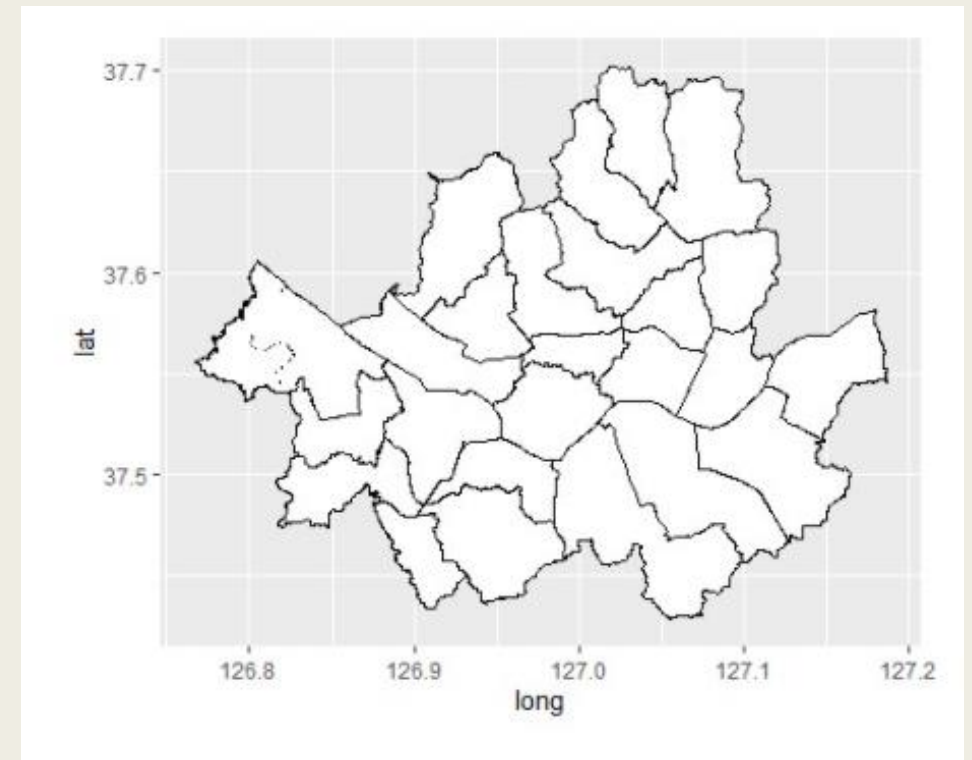


# 데이터 해석 과정 - 인구밀도

Column Name	포함하고 있는 데이터	활용
기간	해당 통계의 년도	X
지역	서울의 각 자치구 이름	다른 데이터의 지역 이름과 매핑
인구	해당 년도 자치구 인구수	X
면적	자치구 면적	X
인구밀도	$km^2$ 당 인구수	밀도와 다른 데이터 간의 상관관계 분석

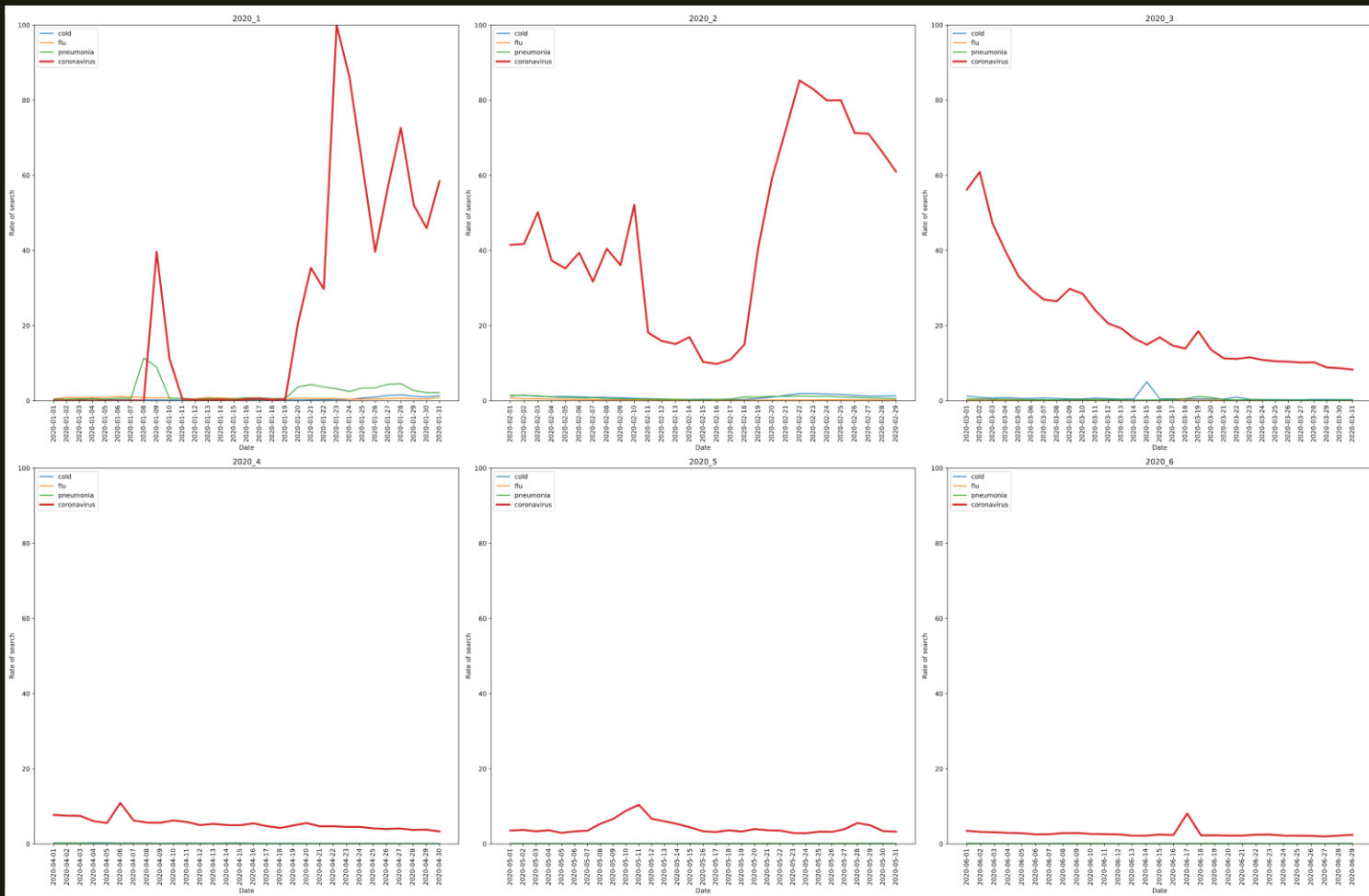
# 데이터 설명 - 서울시 지도 좌표 데이터

- 해당 데이터는 Table 형태가 아닌, .dbf, .prj, shp, .shx 로 구성 되어 있으며 R 언어의 Shapefile과 관련된 패키지들을 사용하여 위도, 경도 좌표계로 변환하여 Plot을 그릴 때 사용하였음
- 우측의 이미지는 해당 데이터를 사용하여 서울시와 각 자치구의 경계선만 그린 그림임



# 데이터 분석 및 시각화

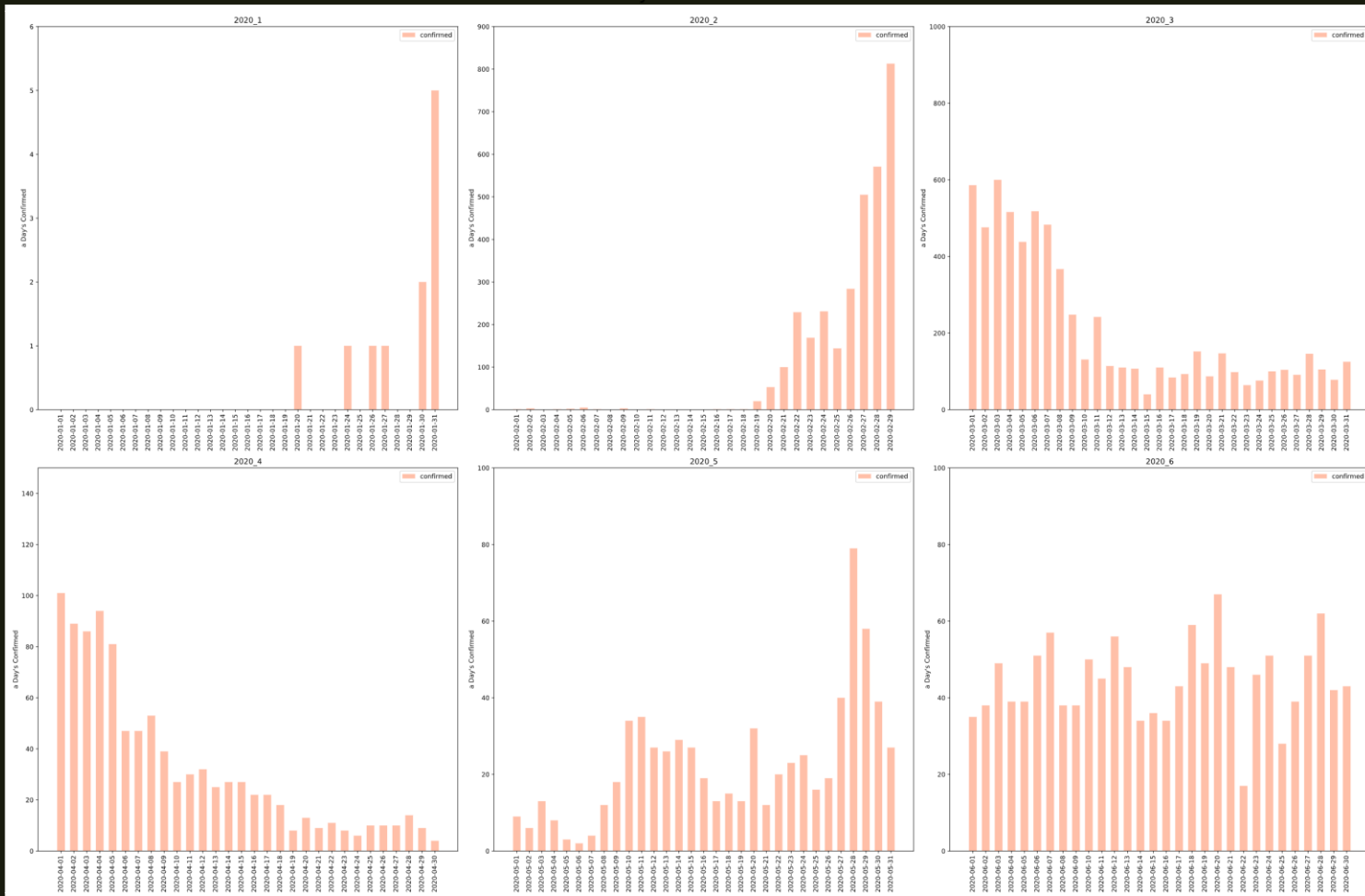
Search Trend



# 코로나 검색 추이

푸른색 - cold  
주황색 - flu  
초록색 - pneumonia  
붉은색 - **coronavirus**

Day's Confirmed



# 일별 확진자수

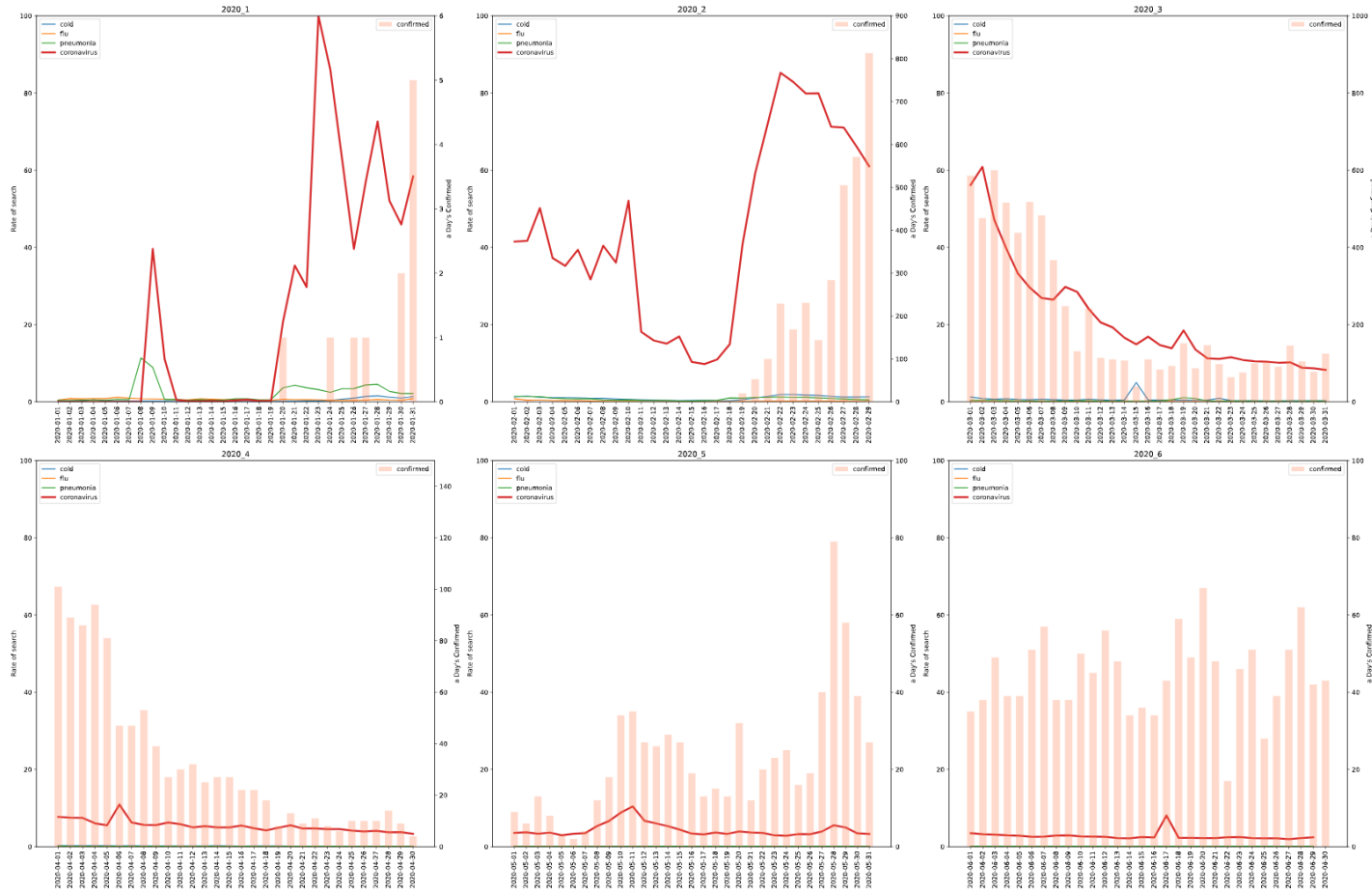
1월 ~ 6월

\*\*세로막대의 Scale은  
월별로 다름



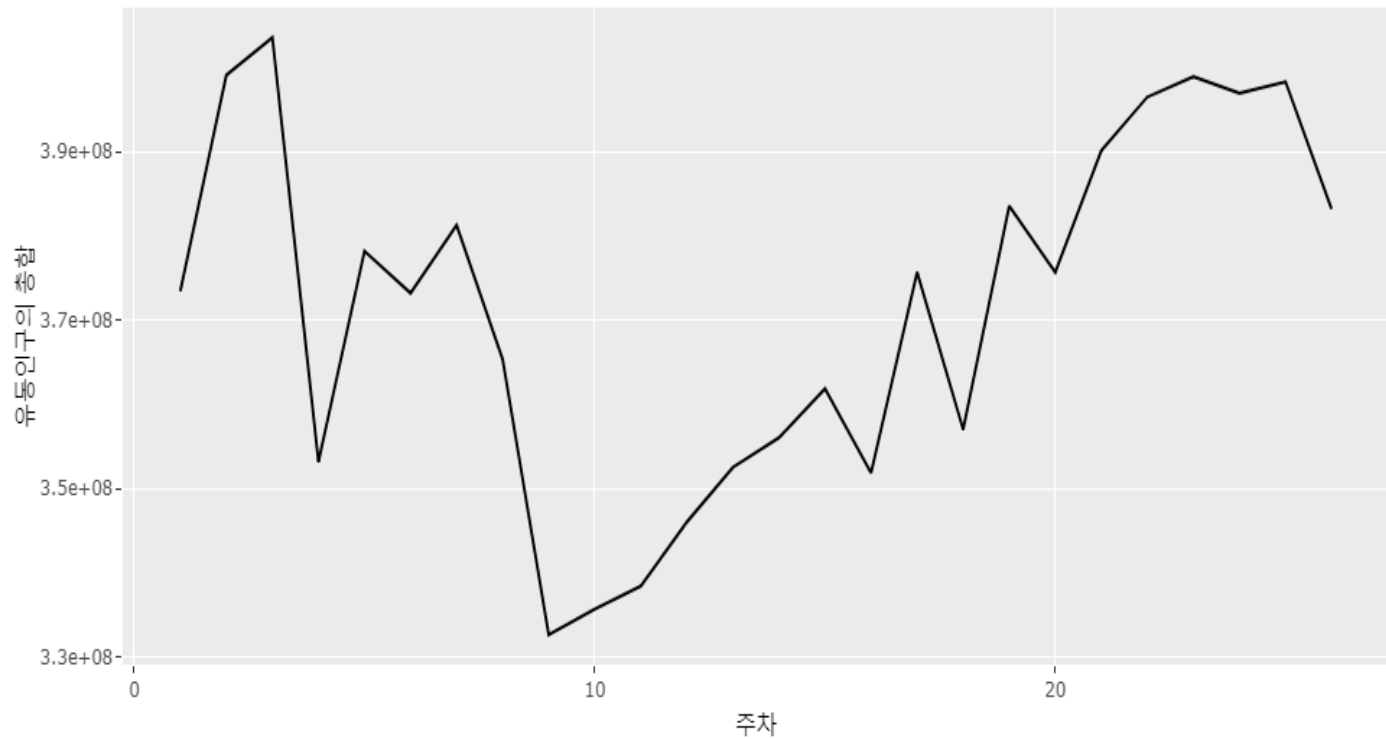
# 일별 확진자 수 & 검색률 비교

Day's Confirmed & Search Trend



실선 - 검색추이  
막대 - 일별 확진자 수

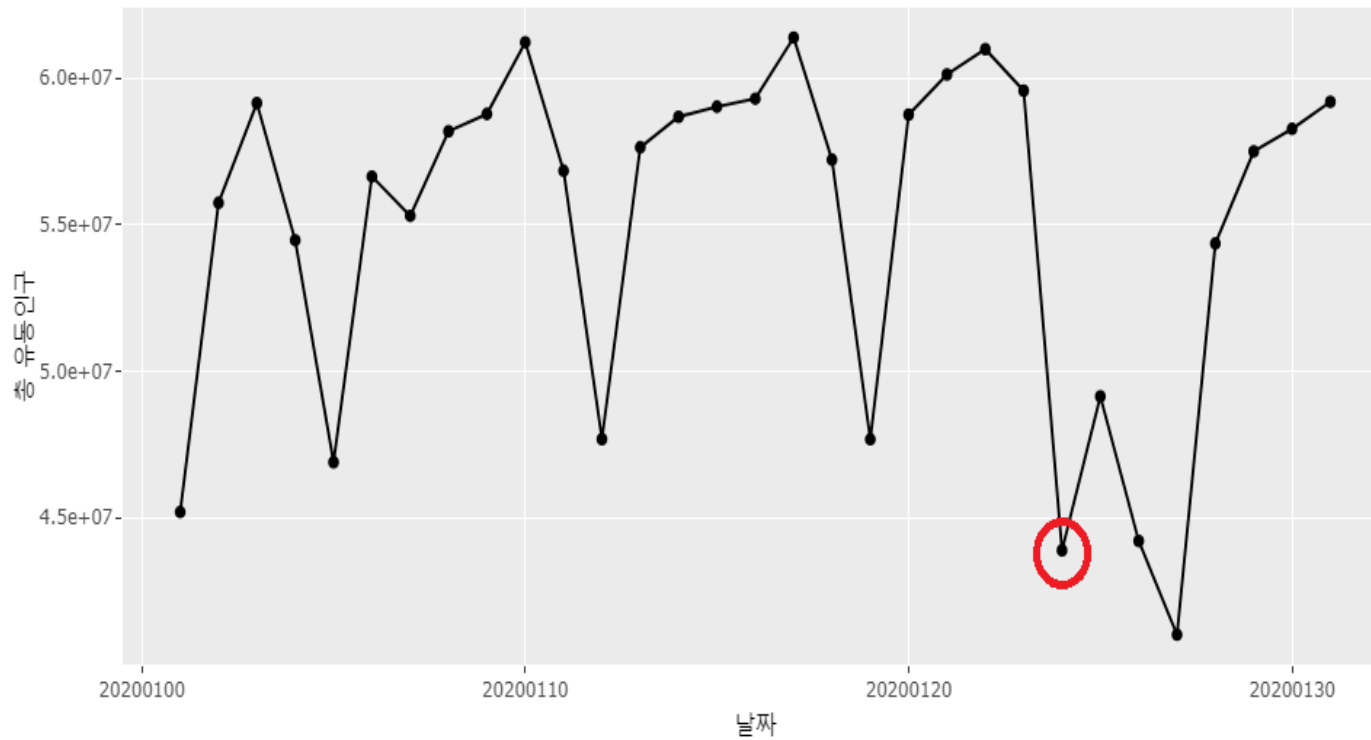
유동인구 주차별 비교



# 유동인구 주차별 비교

코로나 사건 발생이후  
2월부터 3월(5~10주차)  
에서 큰 감소폭을 보이  
다 점차 회복

1월 유동인구 일별 추이



# 1월 유동인구 일별 추이

큰 감소폭을 보이는  
1월 24일의 모습

오늘의 주요뉴스



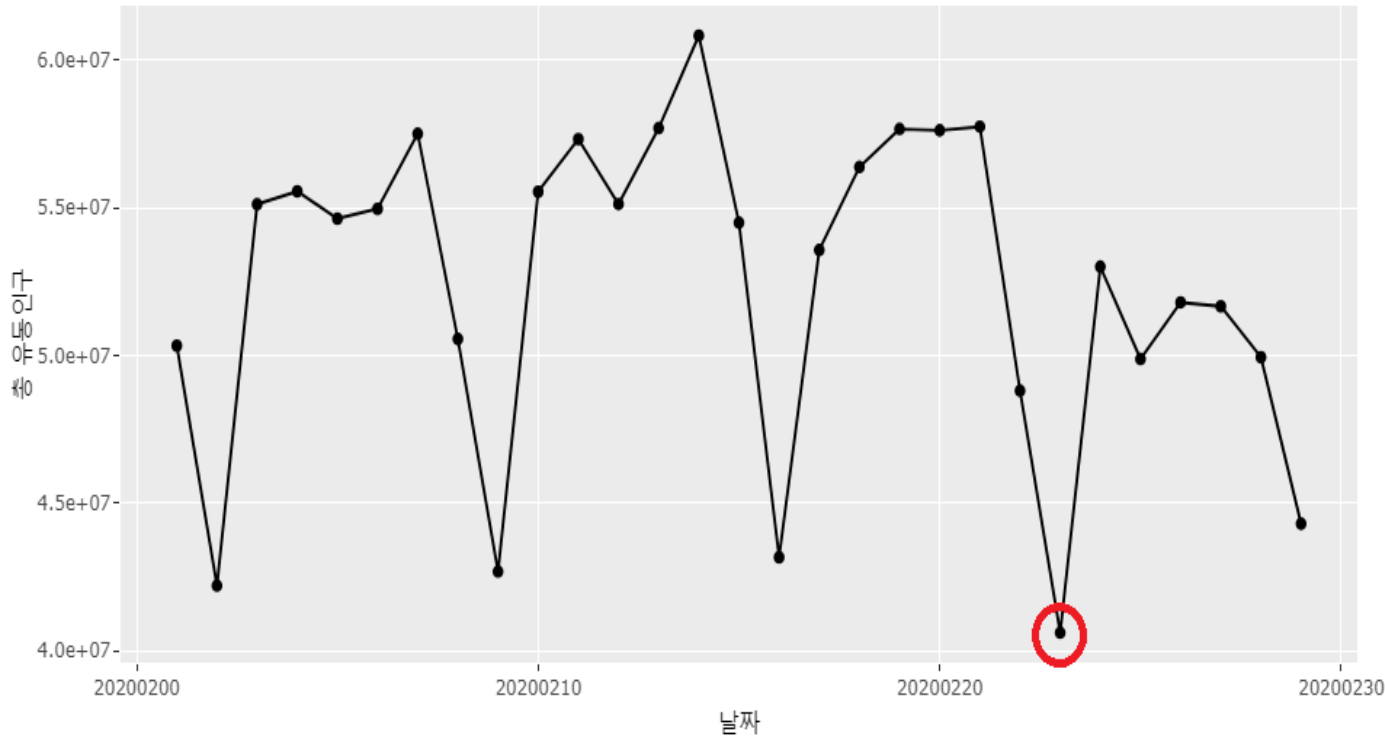
한국인 첫 확진

우한 머물다 귀국한 50대 남성

# 1월 24일의 사건

한국인 첫 확진자 발생

2월 유동인구 일별 추이



# 2월 유동인구 일별 추이

큰 감소폭을 보이는  
2월 23일의 모습



오늘의 주요뉴스

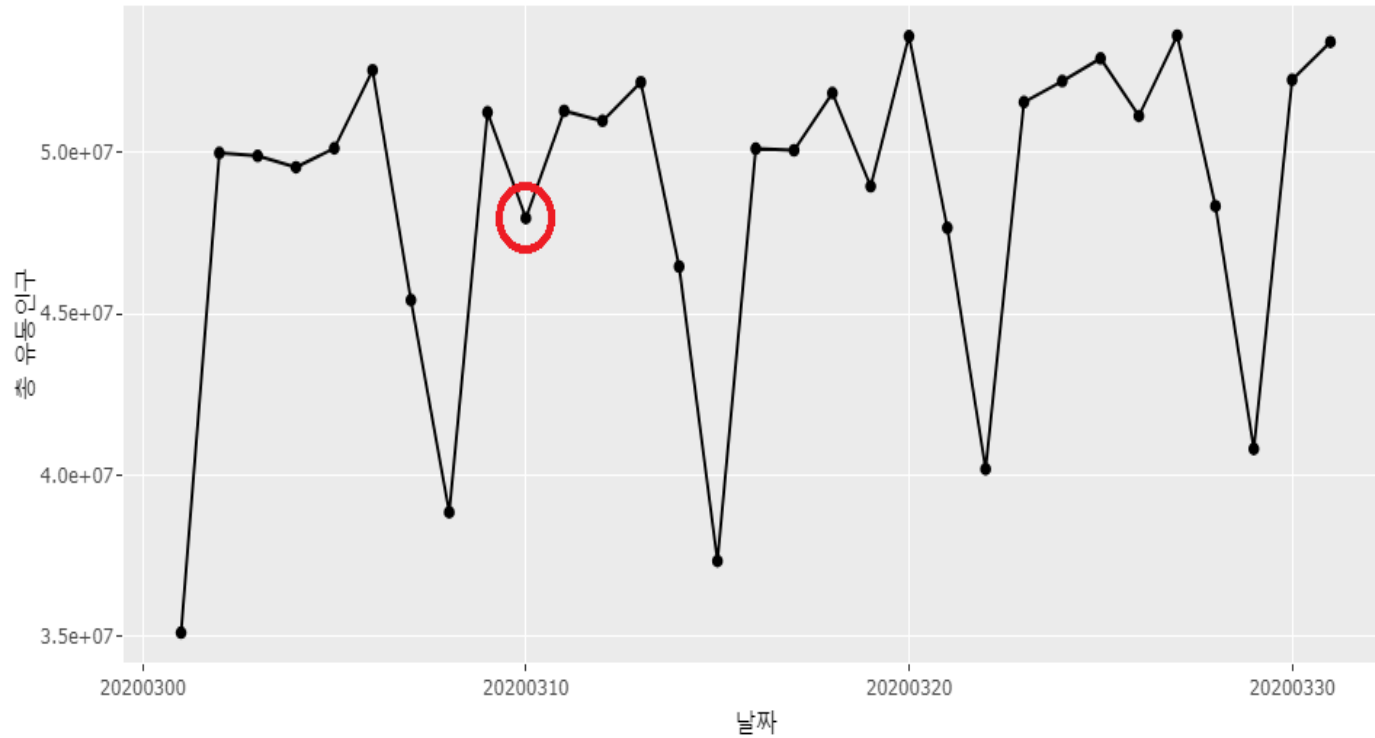
전국이 다 뚫렸다

오늘만 229명 추가...총 확진 433명

## 2월 23일의 사건

전국적인 코로나  
확진자 발생  
대구신천지 사건

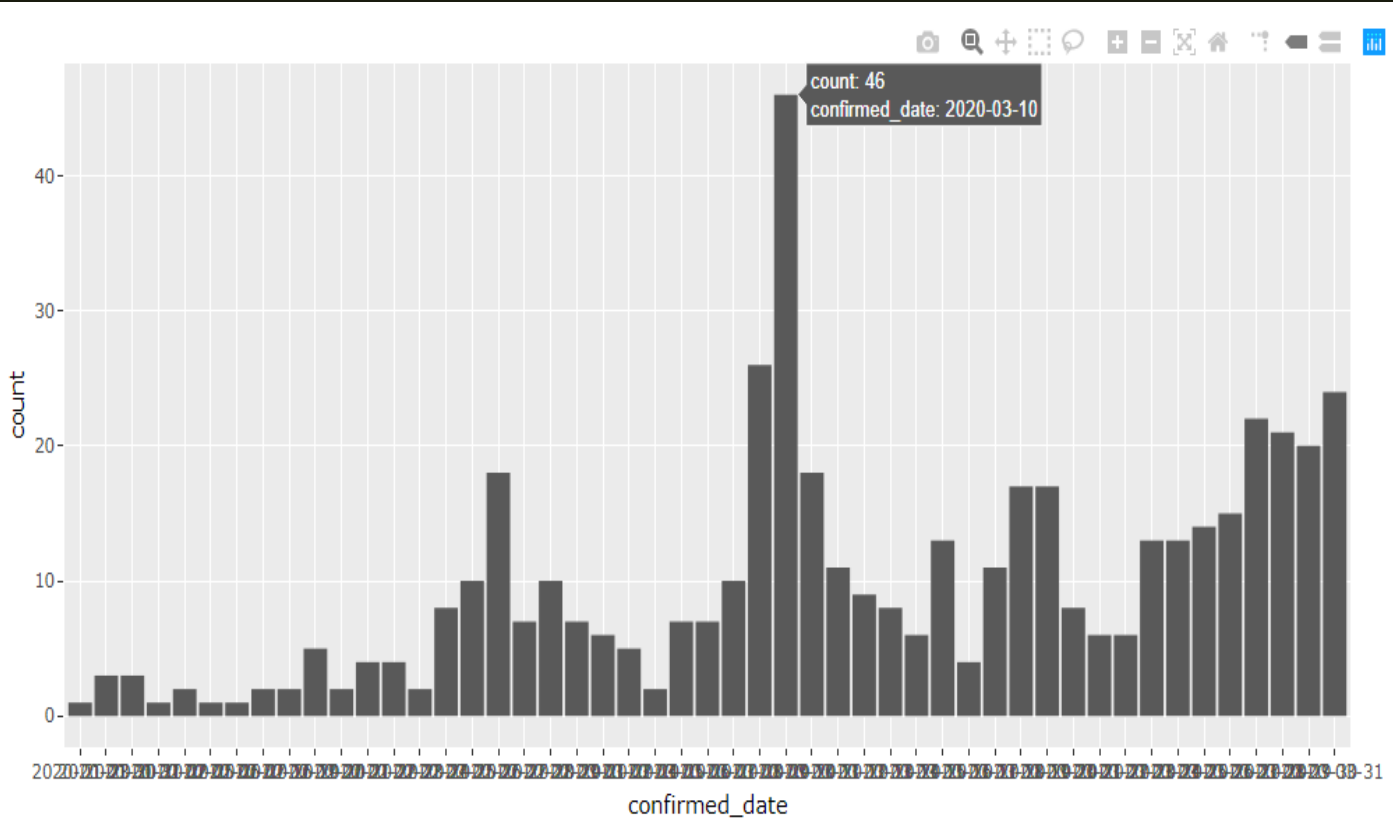
3월 유동인구 일별 추이



# 3월 유동인구 일별 추이

큰 감소폭을 보이는  
**3월 10일**의 모습

# 서울시 일별 확진자 현황



3월 10일에  
46명의 확진자가 발생

오늘의 주요뉴스

### 건물 폐쇄 공고문

이 건물 11층에 입주하고 있는 전화콜센터업체 [메타넷애플렛폼(주)]에서 **코로나19 확진자가 집단 발생**했습니다. 이에 코로나19 확산방지를 위해 아래와 같이 **건물을 폐쇄명령**하니 입주 기업체(사무실)에서는 적극 협조하여 주시기 바랍니다.

가. 폐쇄장소 : 코리아빌딩 1층 ~ 12층 전체

나. 폐쇄기간 : ' 20. 3. 9 ~ 별도 안내시까지(입주민 검사추이확인)

2020. 3. 9.

서울 최대 집단감염

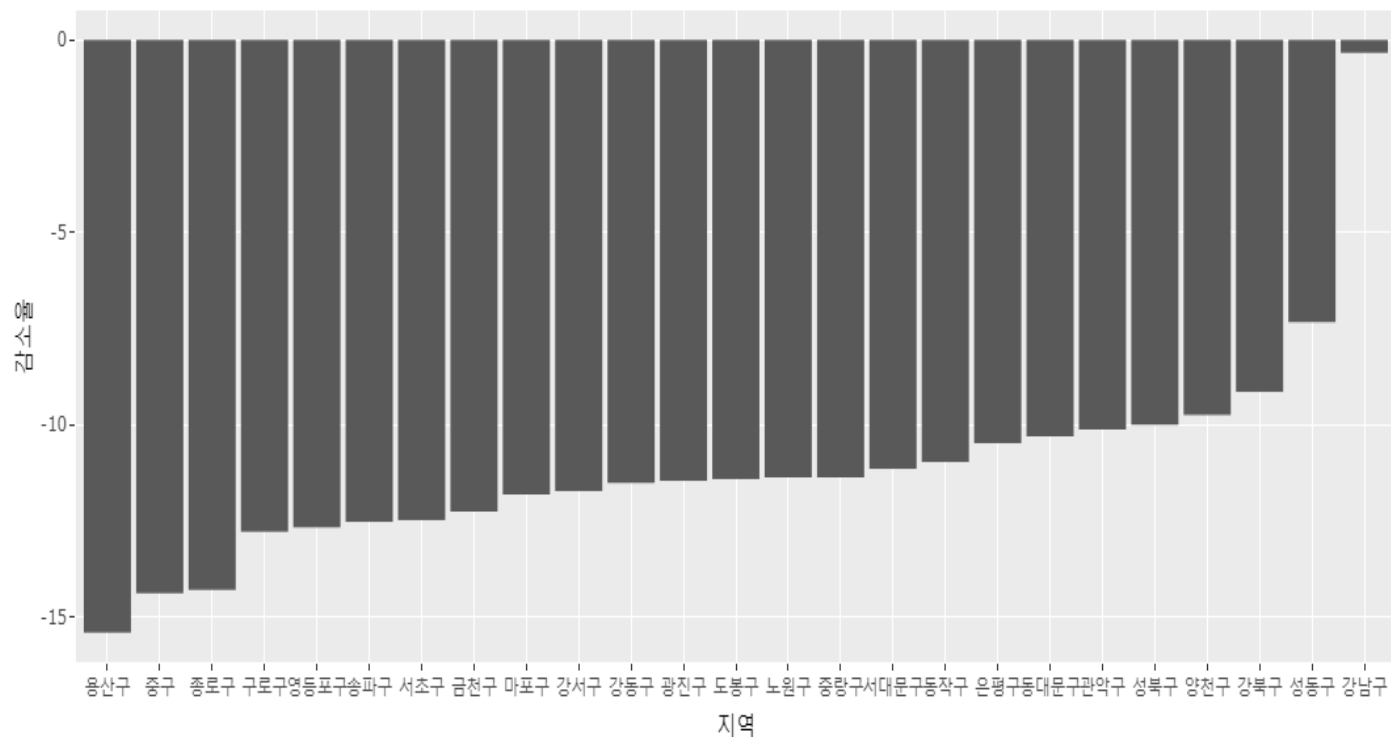
구로구 코로나19 대책본부

**구로 보험사 콜센터서 64명 무더기 확진**

## 3월 10일의 사건

서울 최대 집단감염  
구로 콜센터 확진 사건

1월 대비 2월 유동인구 지역별 감소 추이

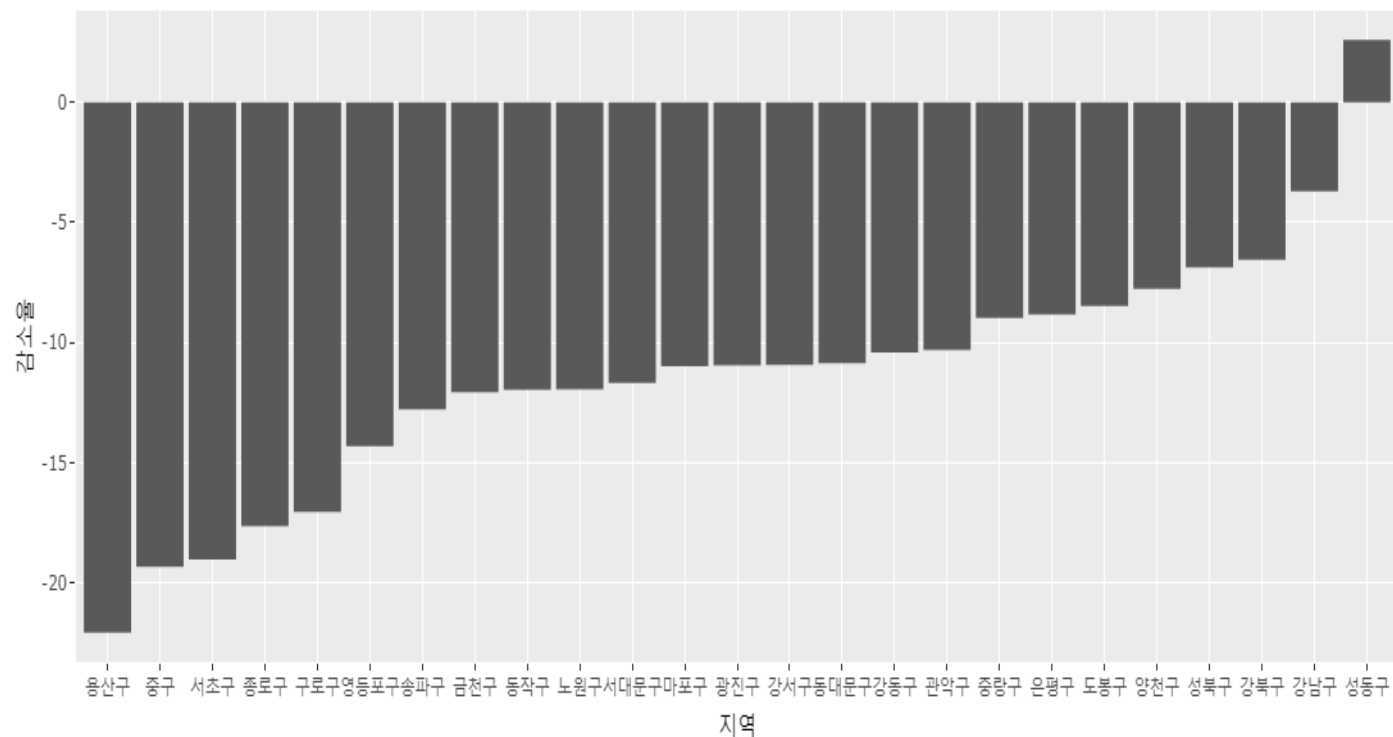


1월 대비  
2월 유동인구  
지역별 감소 추이

지역에 따라 서로 다른  
감소폭을 보임을 확인



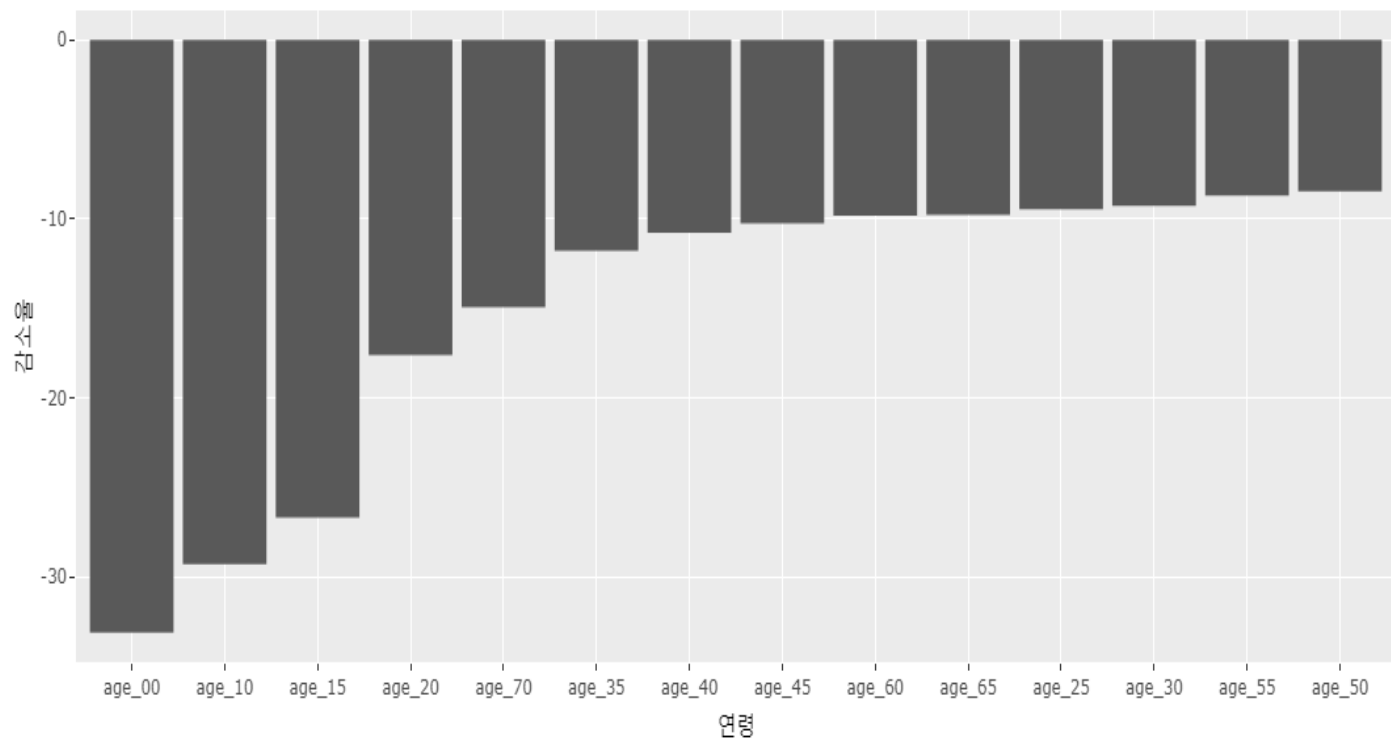
1월 대비 3월 유동인구 지역별 감소 추이



1월 대비  
3월 유동인구  
지역별 감소 추이

지역에 따라 서로 다른  
감소폭을 보임을 확인

1월 대비 3월 유동인구 연령별 감소 추이



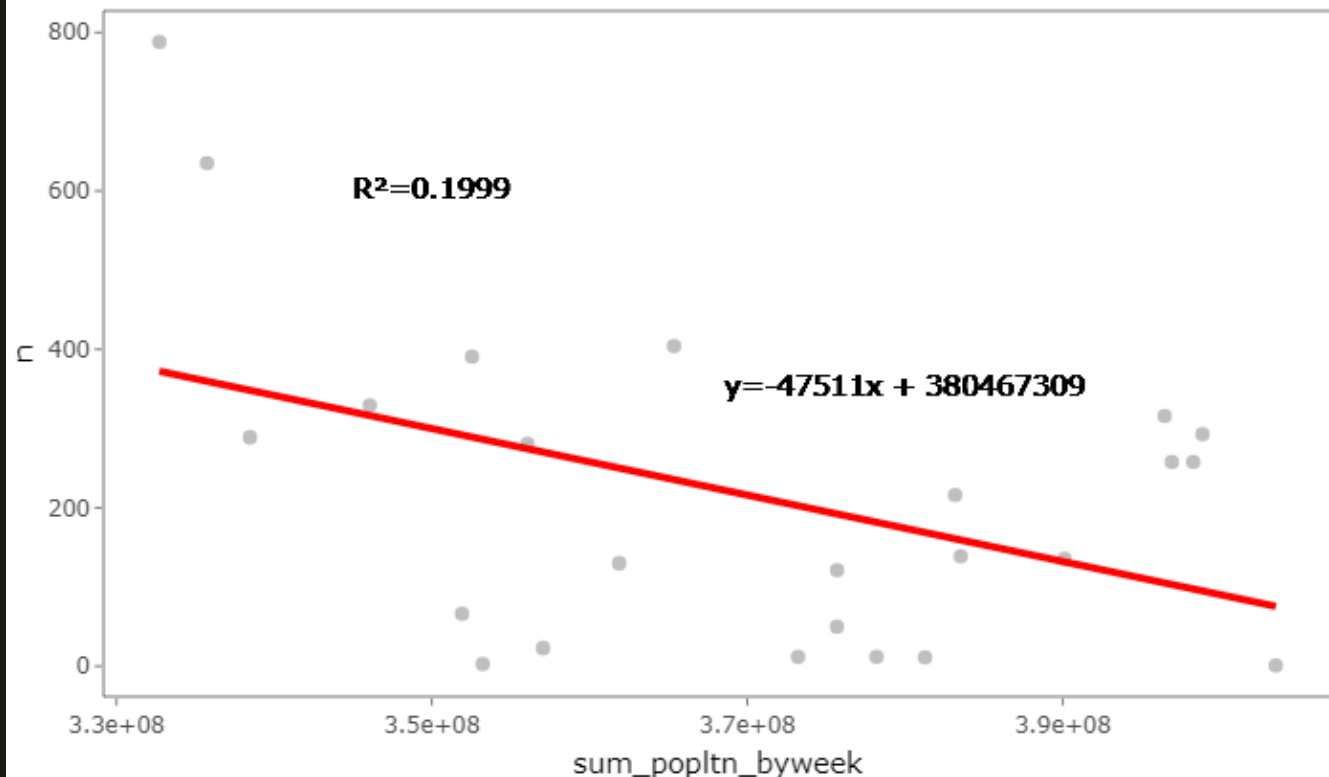
1월 대비  
3월 유동인구  
연령별 감소 추이

아동, 청소년 및 노인 등  
취약계층의 감소율이  
크다는 것을 알 수 있음

# 인사이드 추출

선형회귀와 관계

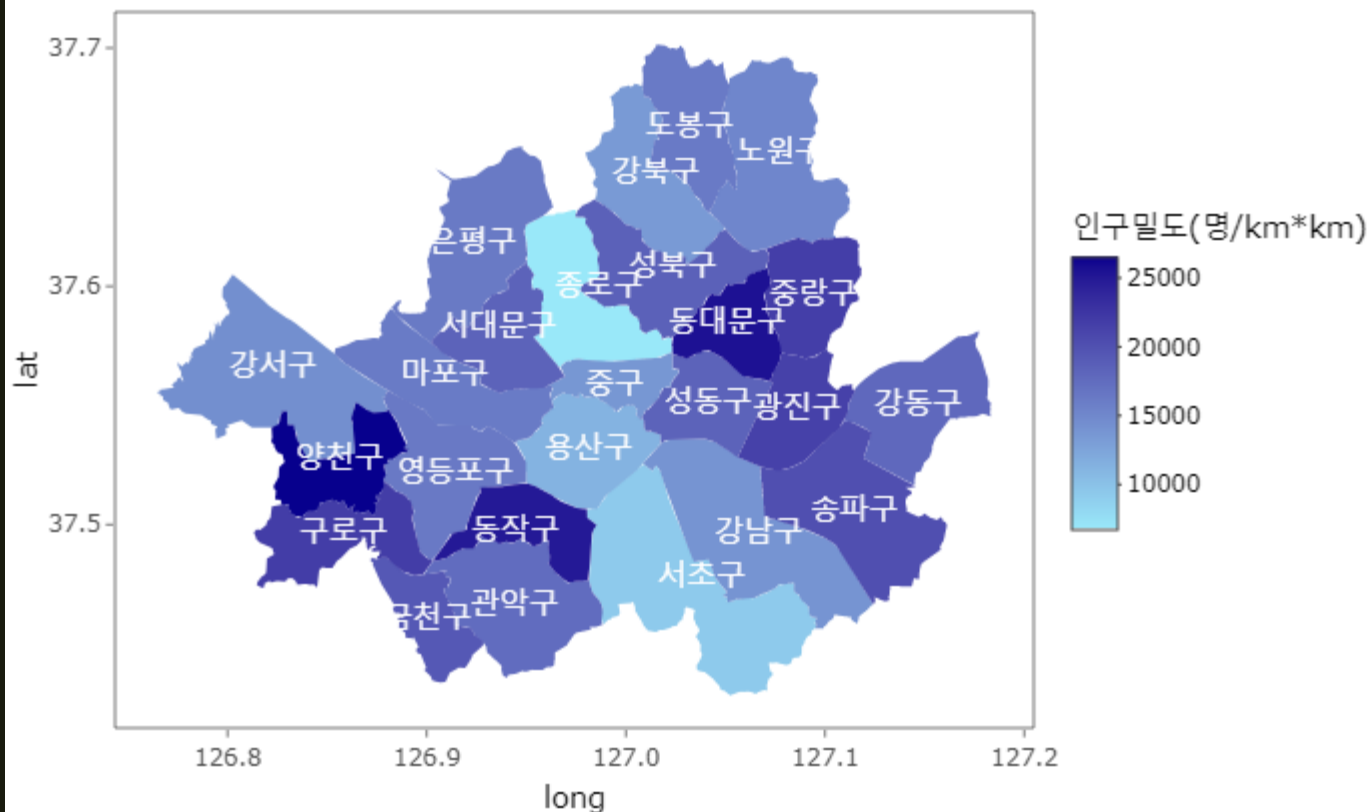
유동인구-확진자 수 상관관계도



유동인구  
-신규 확진자 수  
선형회귀 결과

피어슨 상관계수  
값이 약 -0.45로,  
뚜렷한 음적 상관관계  
를 나타냄

서울시 2019년 자치구별 인구밀도

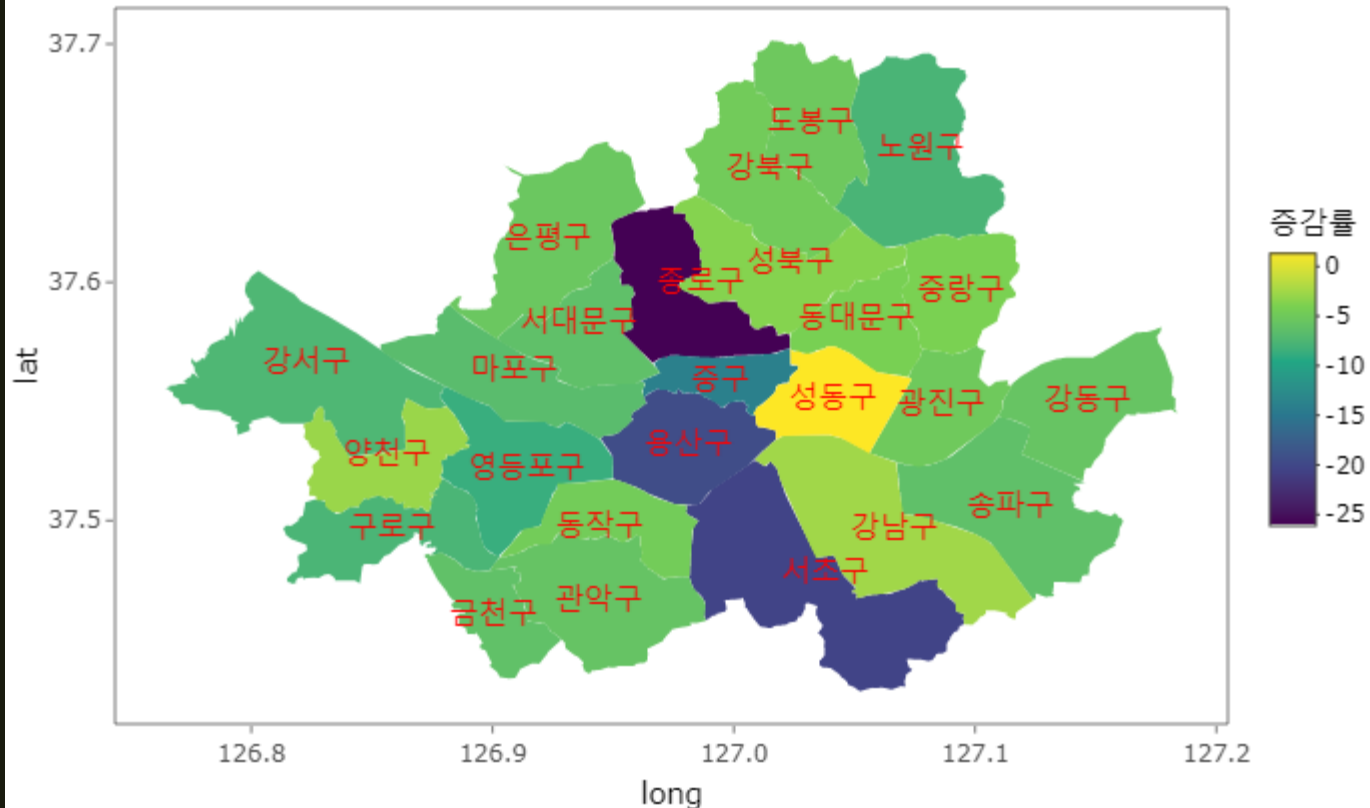


# 서울 자치구별 인구밀도

인구밀도가  
낮은 곳은 비거주지역,  
높은 곳은 거주지역으로  
고려



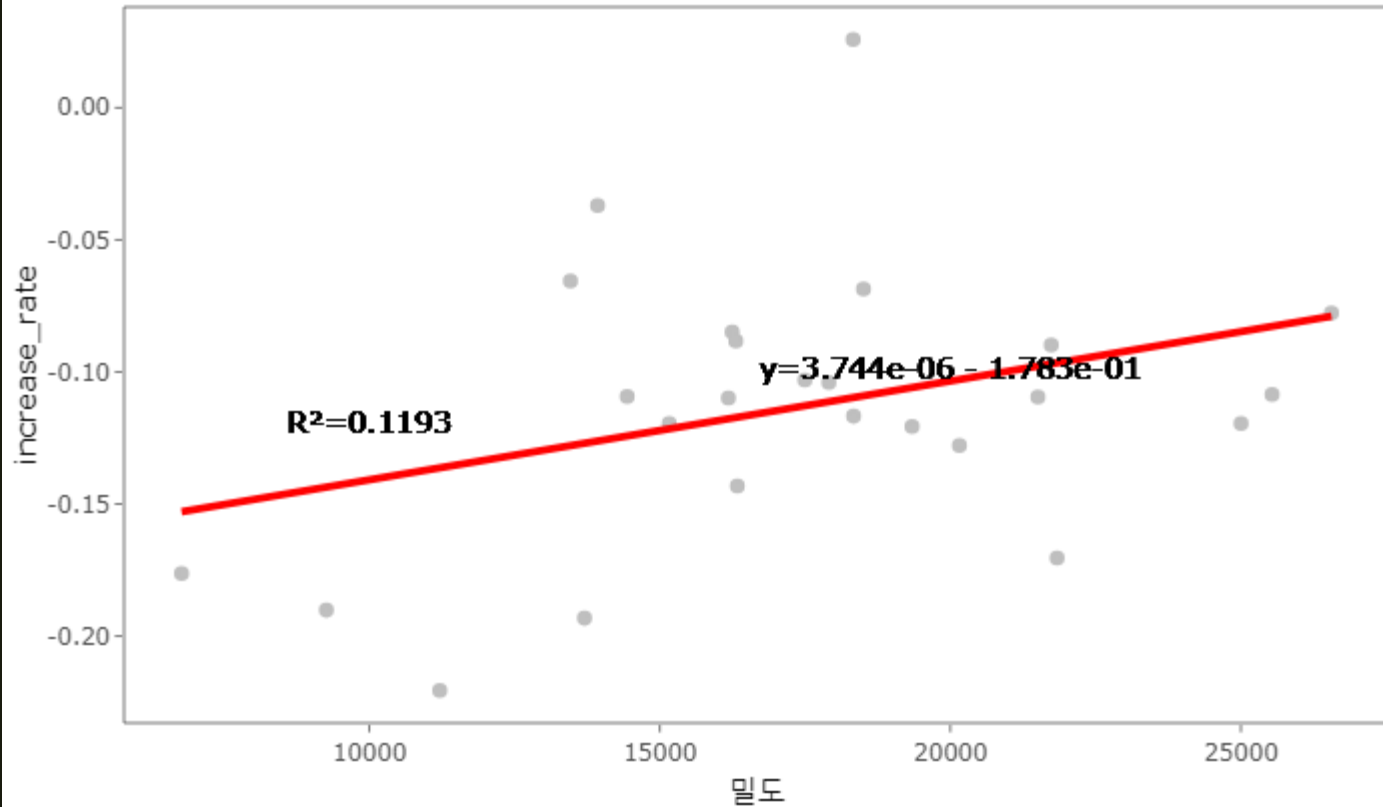
서울시 1월 대비 3월 인구밀도당 유동인구 증감율



서울  
자치구별  
유동인구  
증감률

인구밀도가  
낮은 구의 감소율이  
더 두드러지게 나타남

인구밀도-증감률 상관관계도



# 인구밀도 -유동인구 증감률 선형회귀 결과

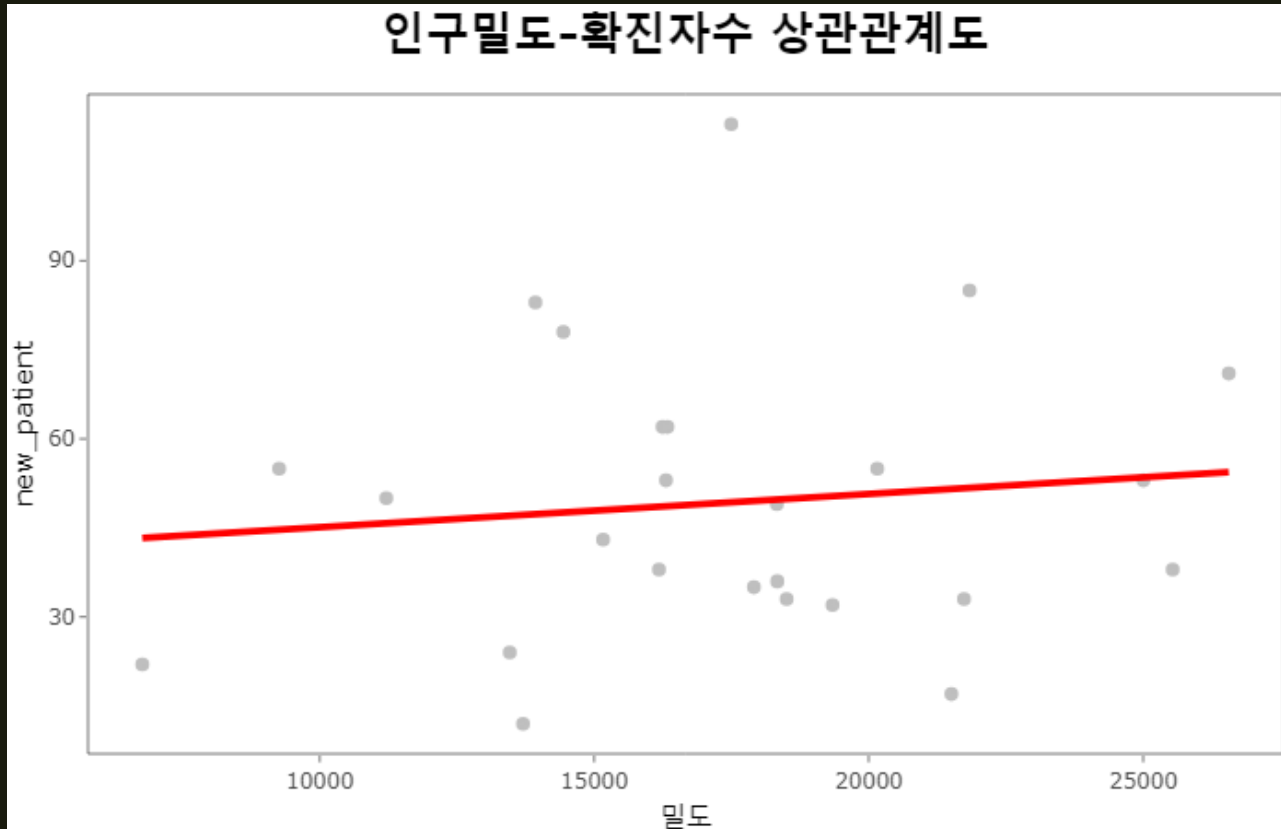
피어슨 상관계수 값이  
약 0.35로, 뚜렷한  
양적 상관관계를 나타냄

# 주차별 신규 확진자 수 & 주별 유동인구 수

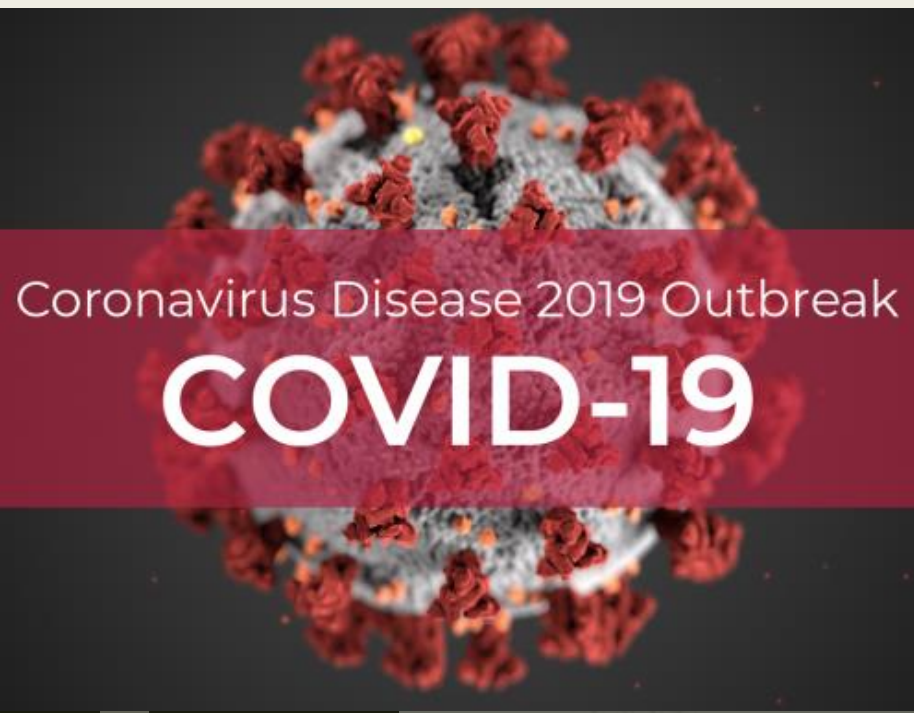


초기에는 확진자수  
-유동인구 수가 뚜렷한  
반전 관계를 보이지만  
시간이 지날 수록  
영향이 줄어들음

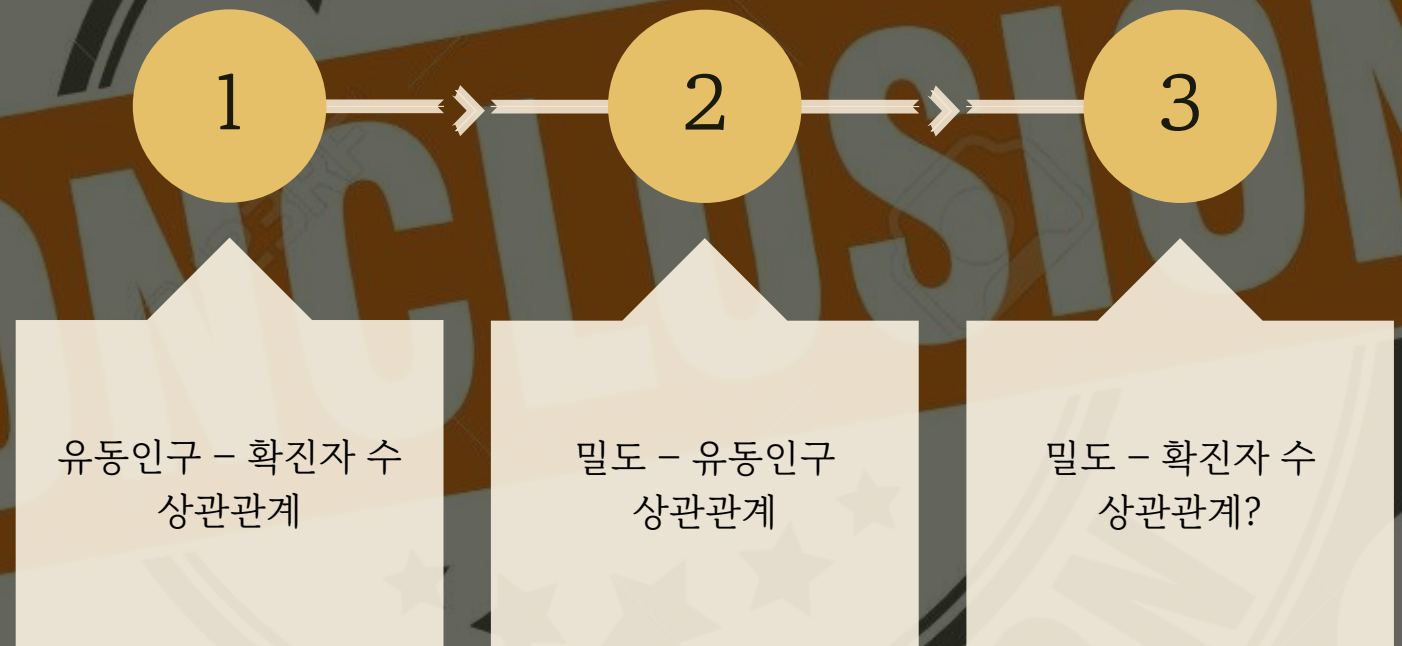
# 인구밀도 -신규확진자 선형회귀결과



피어슨 상관관계수 값이  
약 0.11 로, 상관관계가  
거의 **없다**고 볼 수 있음



## 결론







# 감사합니다!

우평식팀