

TEAM : 위기탈출 NO.1

Covid 19 Data Analysis

About Team

팀명 : 위기탈출 NO.1

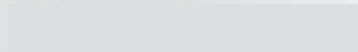
팀원들과 재난에 대한 이야기를 하면서 안전수칙, 위기 상황, 긴급출동, 생활건강등을 알려주는 TV 프로그램 인 위기탈출 NO.1 이라는 방송명을 오마주 하게 되었습니다.

Github

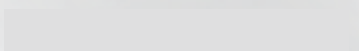
<https://github.com/ch00251/data-analysis-project>

팀원

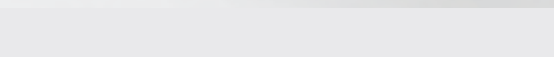
팀장 지윤

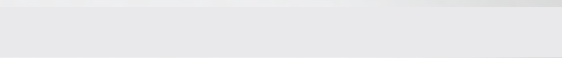
email 

팀원 최우

email 

Blog

지윤 

최우 

CONTENTS

01 프로젝트 목적

- 1 사용한 언어 모듈
- 2 팀원의 역할

02 분석 처리 과정

- 1 사용한 언어 모듈
- 2 API 정의서

03 분석

- 1 확진자가 많은 요일
- 2 확진자가 많은 구
- 3 이상치가 발견된 구
- 4 위도와 경도의 차이에 따른 전염

04 결론

프로젝트 목적

현재 가장 유행하는 Covid19의 데이터를 분석하여
앞으로 다가올 미래의 전염병을 예방 및 피해를 최소화
하는데 지침이 되는 것에 목적이 있습니다.

팀원 역할

팀장 지윤

①

Covid19의 데이터를 이용하여 구단위
감염자 자료 조사 및 분석

②

Covid19의 데이터를 이용하여 Covid19
감염자가 많은 요일 조사하고
자료에 대해서 분석하고 해결방안 탐색

팀원 최우

①

Covid19의 데이터를 이용하여 이미 입
증된 이론을 데이터와 비교 분석

②

Covid19의 데이터를 이용하여 Covid19
이상치의 존재 확인과 이상치의 발생원
인 조사 및 분석

분석 처리 과정

데이터 수집(Crwing)



서울 열린 데이터 광장

DATA 공공데이터포털
. GO . KR

전처리(PreProcessing)



Jupyter

pandas

seaborn

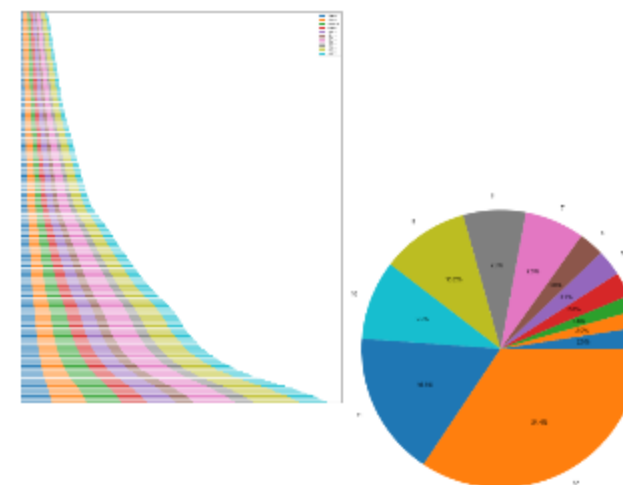
Folium

python

NumPy



분석(Analysis)



데이터를 분석 하고 후에 발생할 전
염병에 대한 피해 예방 및 피해 차단
방향 제시

API 정의서

데이터 출처

서울시 Covid19 확진자 데이터 : <https://data.seoul.go.kr/dataList/OA-20470/S/1/datasetView.do>

서울시 총 인구수 데이터 : <https://data.seoul.go.kr/dataList/10790/S/2/datasetView.do>

코로나 데이터의 범위

2021년 1월 1일 ~ 2021년 12월 31일

corona_original	Covid19 전처리 데이터(Raw Data)	
corona_2021	2021년 Covid19 Data	
han_river_gu	한강 인접 행정구 목록	
han_river_gu_re	한강 인접 행정구 목록 컬럼 변경	
state_geo	서울 행정구 경계 데이터	출처 : https://github.com/southkorea/southkorea-maps
longitude	경선에 따른 행정구 구분	
latitude_gu	위선에 따른 행정구 구분	
Corona_seoul	2020년 ~ 2022년 서울 각 구별 코로나 확진자 데이터(누적, 추가)	
Corona_gu	corona_seoul에서 추가 확진자들만 가져온 데이터	
Corona_2021	2021년도 서울 코로나 확진자 데이터	
Month_2021	서울시 각 구별 2021년 월 기준 총 확진자 수	
Day_2021	서울시 각 구별 2021년 1월 1일 부터 12월 31일 까지 일별 코로나 확진자 수	
Year_2021	1년간 각 구별 총 확진자	

문제



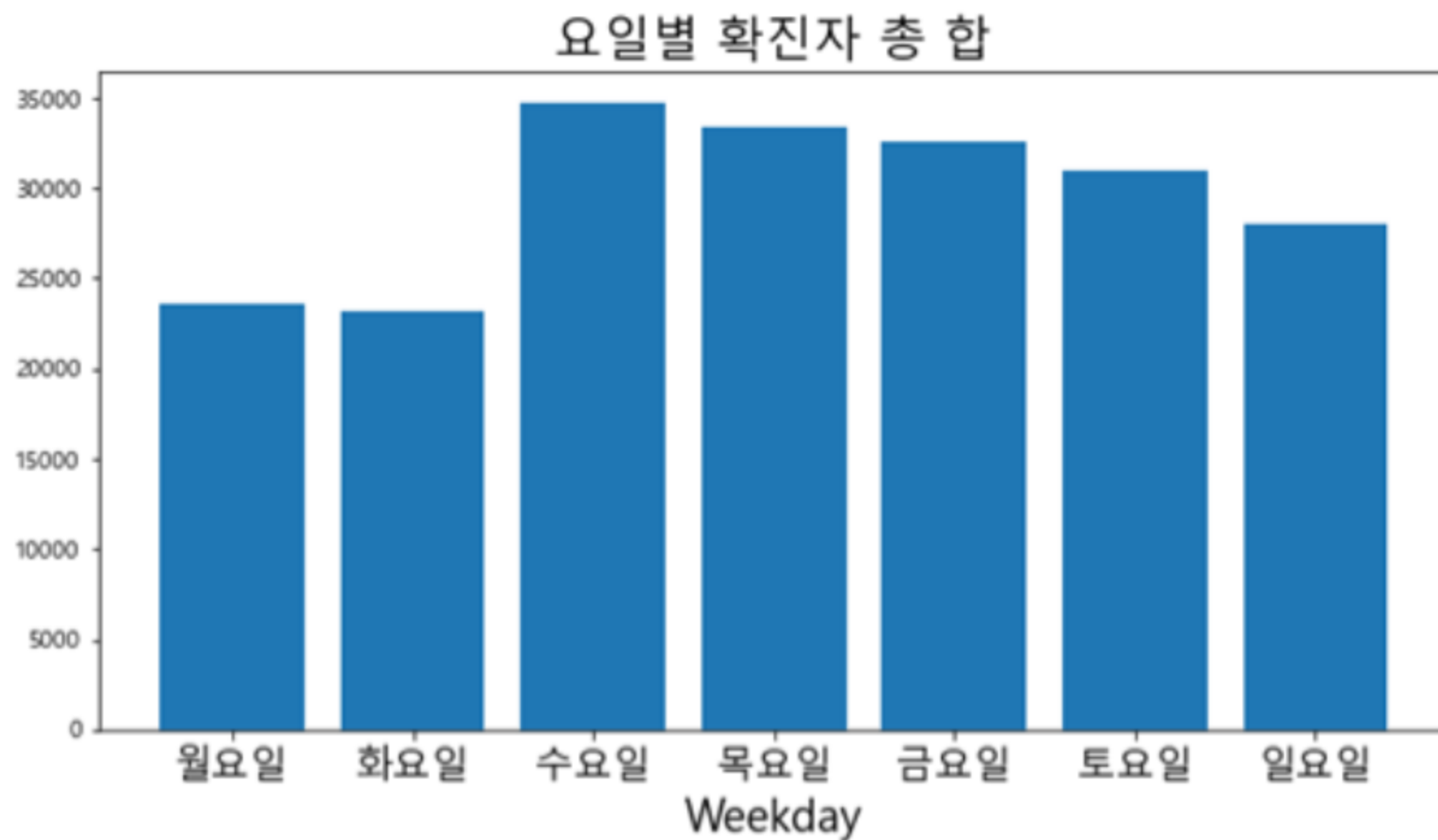
가장 확진자가 많이 나오는 요일은?

데이터 처리

	종로 구	중 구	용산 구	성동 구	광진 구	동대문 구	중랑 구	성북 구	강북 구	도봉 구	...	금천 구	영등포 구	동작 구	관악 구	서초 구	강남 구	송파 구	강동 구	기타	합
weekdays																					
금요일	538	515	746	806	1052	1457	1242	1356	1036	989	...	924	1620	1283	1554	1100	1764	2041	1417	2228	32626
목요일	544	626	763	823	1036	1391	1203	1324	973	978	...	867	1615	1182	1656	1220	1893	1932	1407	2988	33392
수요일	567	601	882	856	1079	1468	1285	1308	970	1078	...	918	1686	1340	1862	1289	1909	2136	1511	2133	34730
월요일	439	387	524	592	784	947	950	1188	706	669	...	625	1128	945	1222	889	1310	1529	961	1442	23604
일요일	492	477	670	721	870	1224	1046	1169	797	788	...	762	1340	1134	1446	1025	1483	1666	1205	2128	28062
토요일	541	549	734	800	986	1315	1104	1196	885	966	...	849	1499	1215	1544	1142	1746	1856	1248	2276	30968
화요일	470	397	559	663	808	977	859	1110	887	743	...	535	962	879	984	828	1429	1294	879	1444	23195

자치구 기준일 별 확진자의 총 합을 구하고 2021년 1월 1일이 금요일인 것을 이용하여 모든 행 데이터의 요일을 구한 후 요일별 확진자 총 합을 구하였습니다.

분석



수요일에 가장 많은 확진자 발생

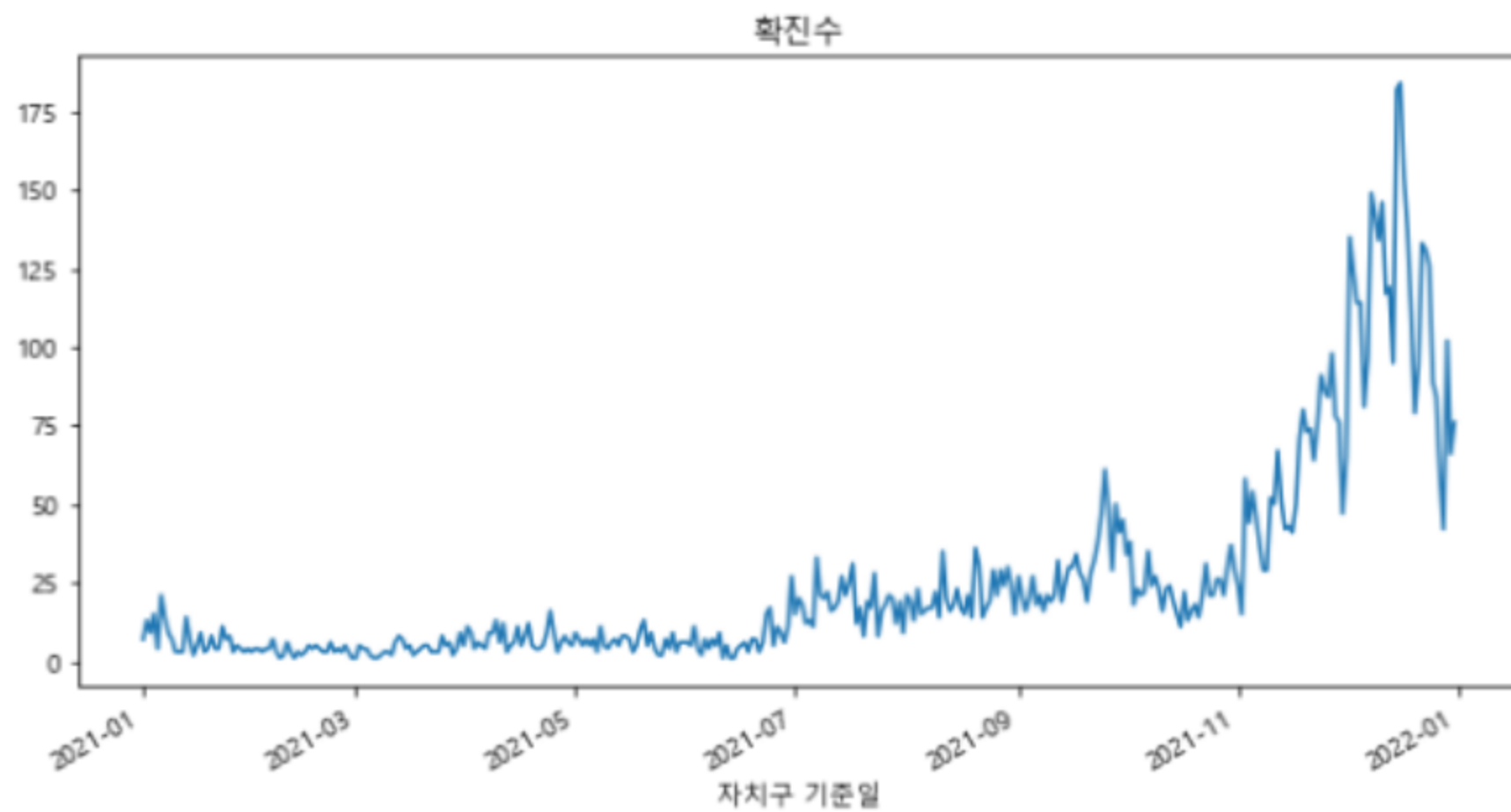
‘주말효과’로 인하여 토,일요일 검사건수 감소에 따라 발표일 기준 월, 화요일 확진자 수도 감소하고 주말효과가 걷히는 수요일에 검사량이 평일 수준으로 회복하면서 가장 최고치를 찍는다는 것을 알 수 있음

참고

<https://www.hani.co.kr/arti/area/cpita/1011392.html>

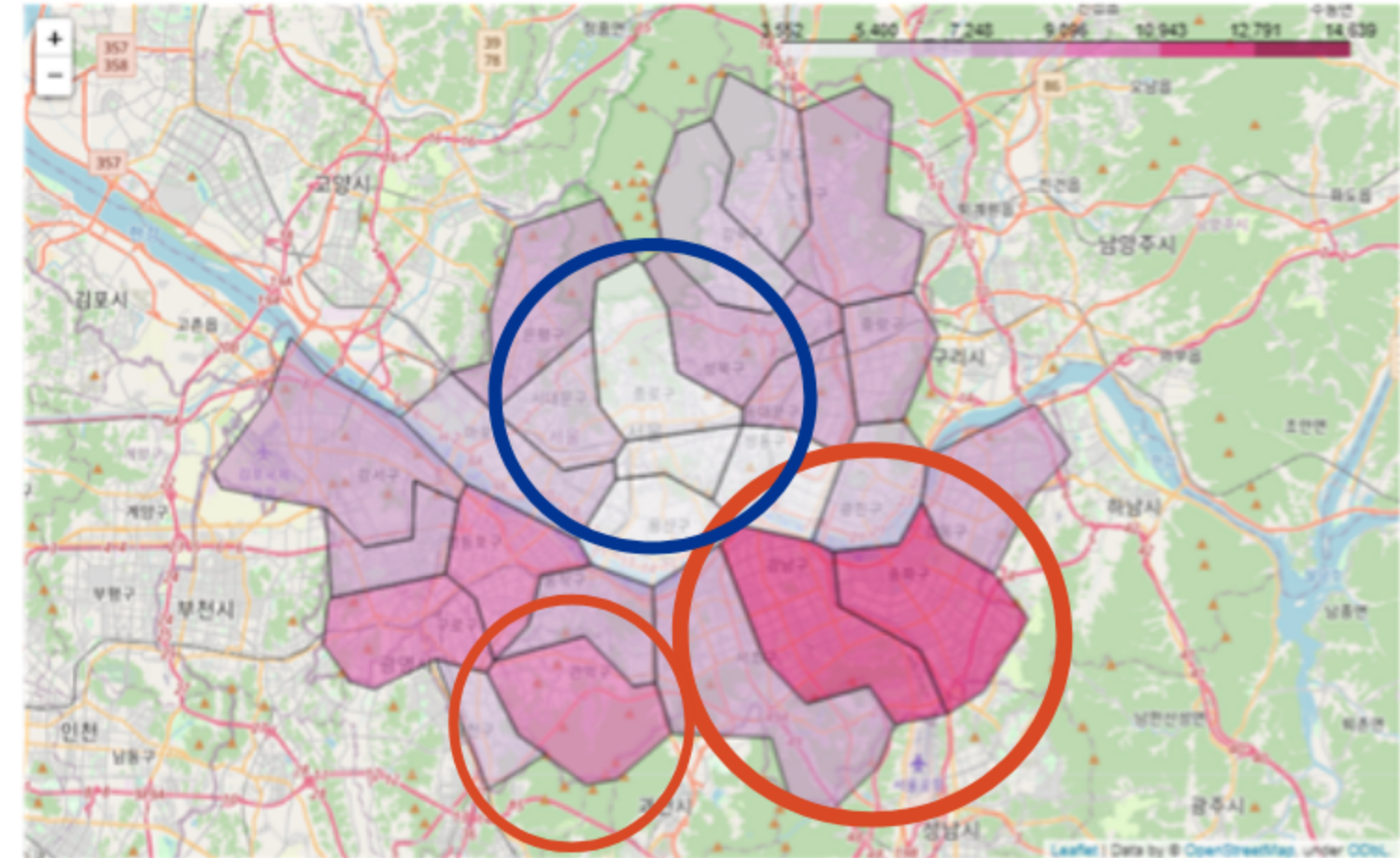
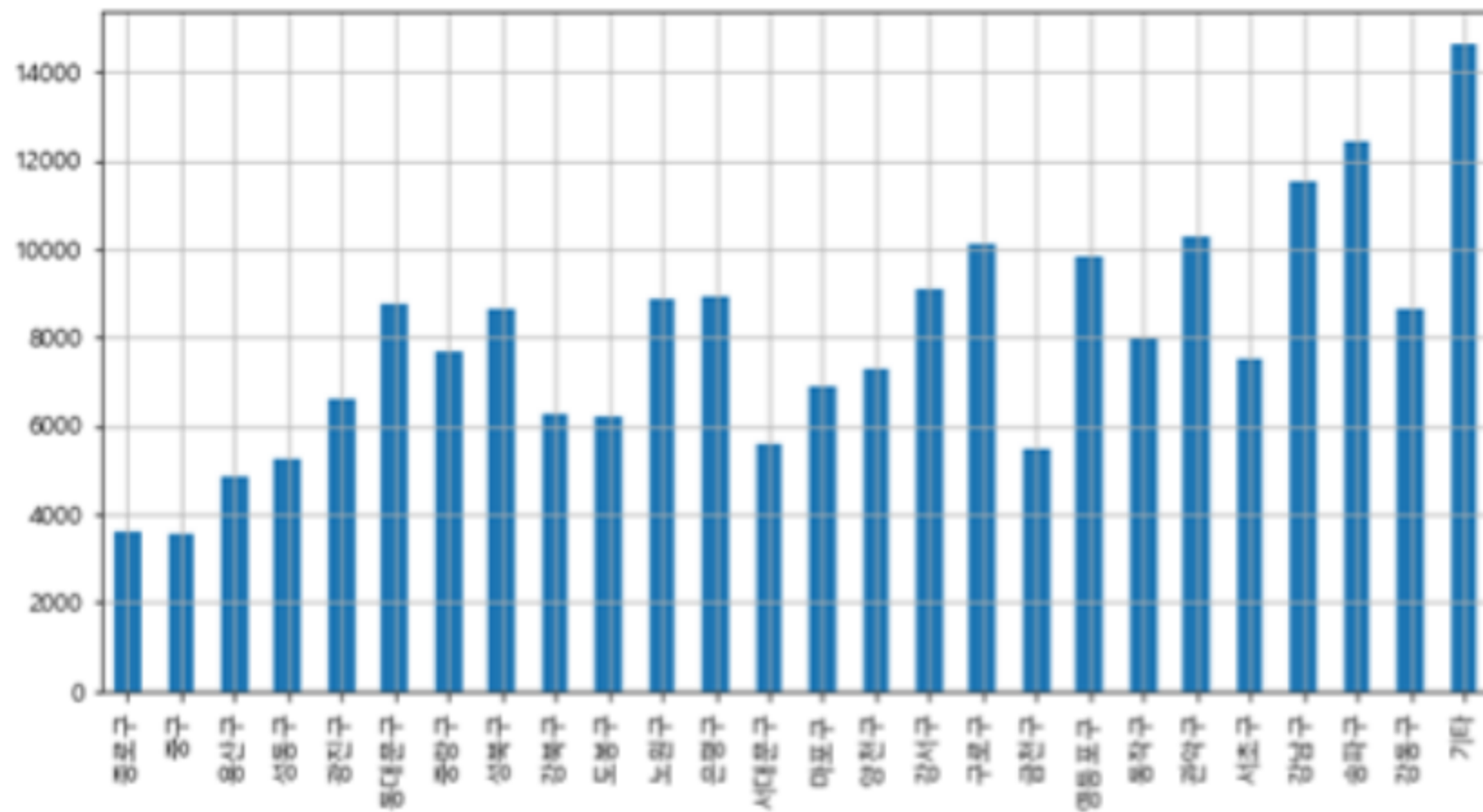
<https://www.joongang.co.kr/article/24116198>

문제



서울시에서 가장 많은
확진자가 발생한 구는 어디일까?

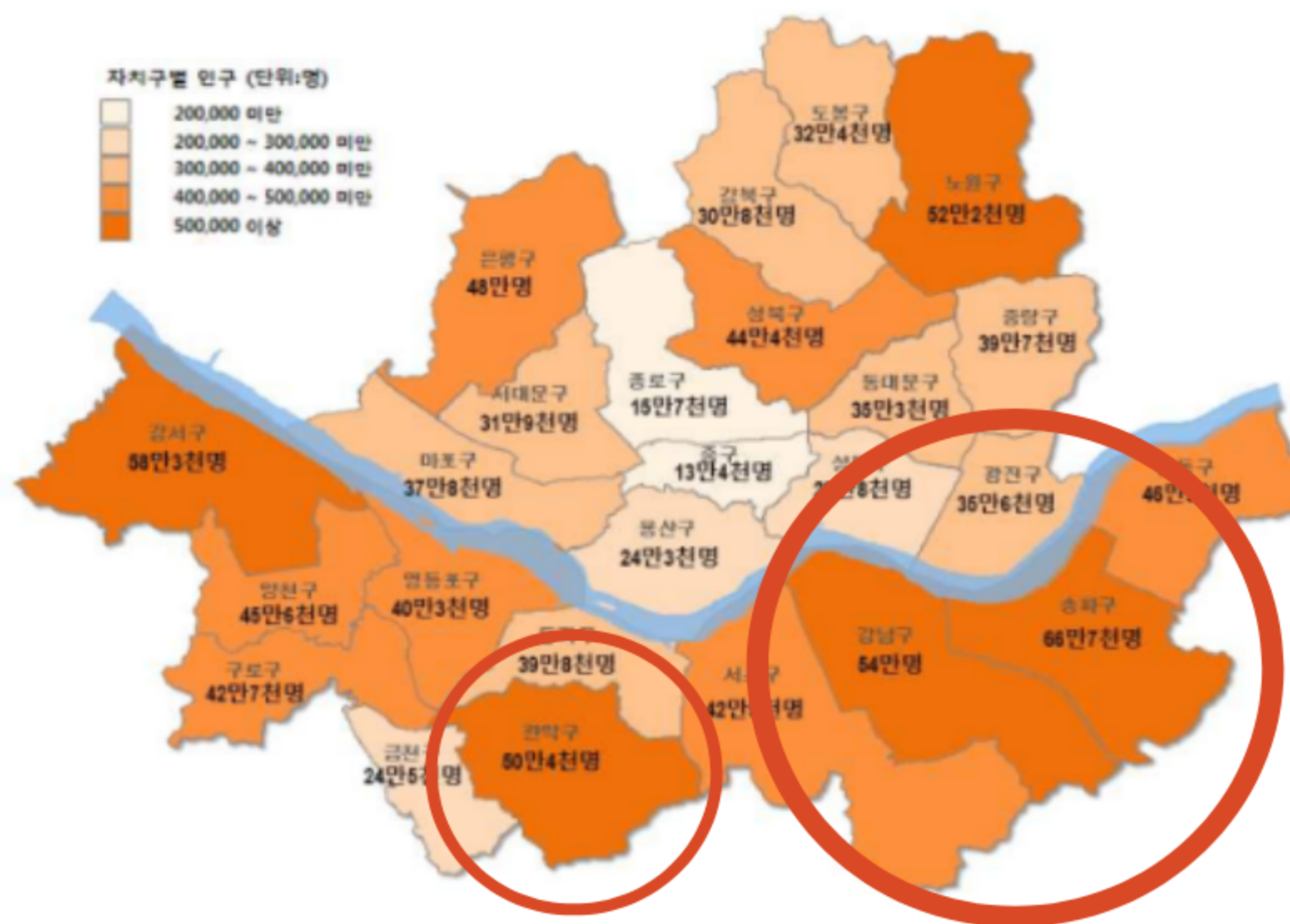
데이터 처리



확진자가 많은 구(빨간원) : 송파구, 강남구, 관악구
확진자가 적은 구(파란원) : 종로구, 중구

확진자 많은 구

【자치구별 인구현황】



송파구, 강남구 확진자가 많은 이유

- 각각 약 67만명, 54만명 정도의 인구가 거주하는것으로 확인되어 확진자 수가 가장 많은 것에 영향이 있다고 보여집니다.

관악구 확진자가 많은 이유?

- 생산가능한 15-64세 인구 비율과 고령자 비율이 다른 구에 비해 상대적으로 높습니다.

【연령계층별 인구 및 구성비】

(단위 : 명, %, 세)

구분	총인구			구성비			중위연령		
	유소년 (0-14세)	생산가능 (15-64세)	고령 (65세 이상)	유소년 (0-14세)	생산가능 (15-64세)	고령 (65세 이상)	전체	남	여
서울시	997,826	7,262,449	1,567,819	10.2	73.9	16.0	43.3	42.6	44.0
관악구	34,981	390,759	78,400	6.9	77.5	15.6	40.1	39.3	41.0
서초구	56,754	308,140	60,209	13.4	72.5	14.2	42.1	41.7	42.4
강남구	65,710	398,226	75,602	12.2	73.8	14.0	42.0	41.6	42.3
송파구	78,893	494,077	94,145	11.8	74.1	14.1	42.0	41.5	42.4
강동구	55,536	342,076	71,203	11.8	73.0	15.2	43.0	42.3	43.8

참조 :

[http://stat.seoul.go.kr/pdf/registration\(2021.1Q\).pdf](http://stat.seoul.go.kr/pdf/registration(2021.1Q).pdf)

확진자 적은 구

【자치구별 인구현황】



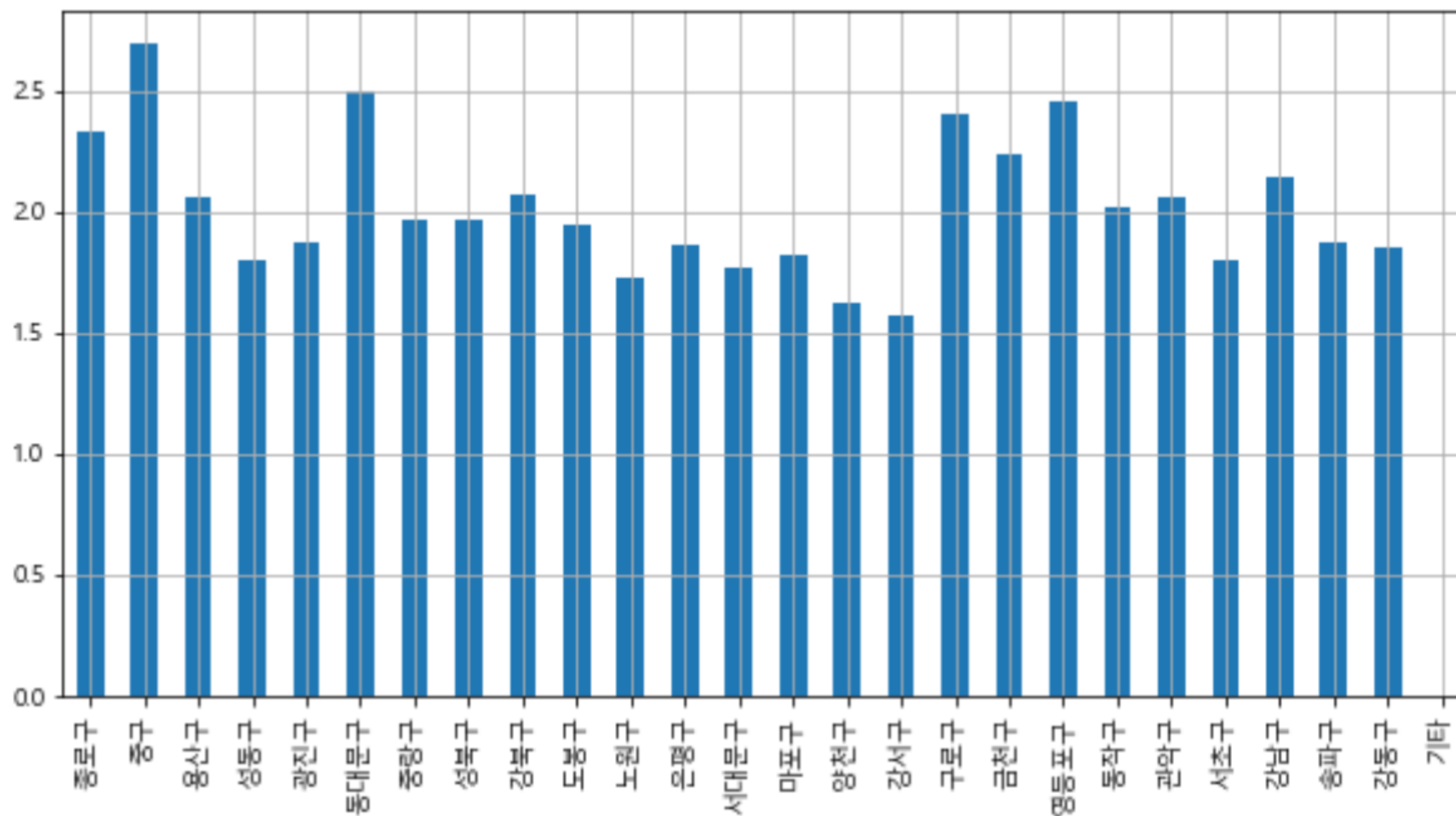
코로나의 확진이 적은 구도 조사해 보았습니다.

종로와 중구가 상대적으로 다른 구에 비해 매우 적은 숫자를 보여 주는데 이는 종로와 중구는 문화, 상업, 행정구로 대부분 유동인 구들이 일(work)이나 문화생활을 즐기고 떠나는 경우가 많고 '실제 거주민이 적은구'이기 때문에 확진자 수가 거주민에 비례하여 적은것으로 보입니다.

참조

<https://www.donga.com/news/Society/article/all/20201113/103945233/1>

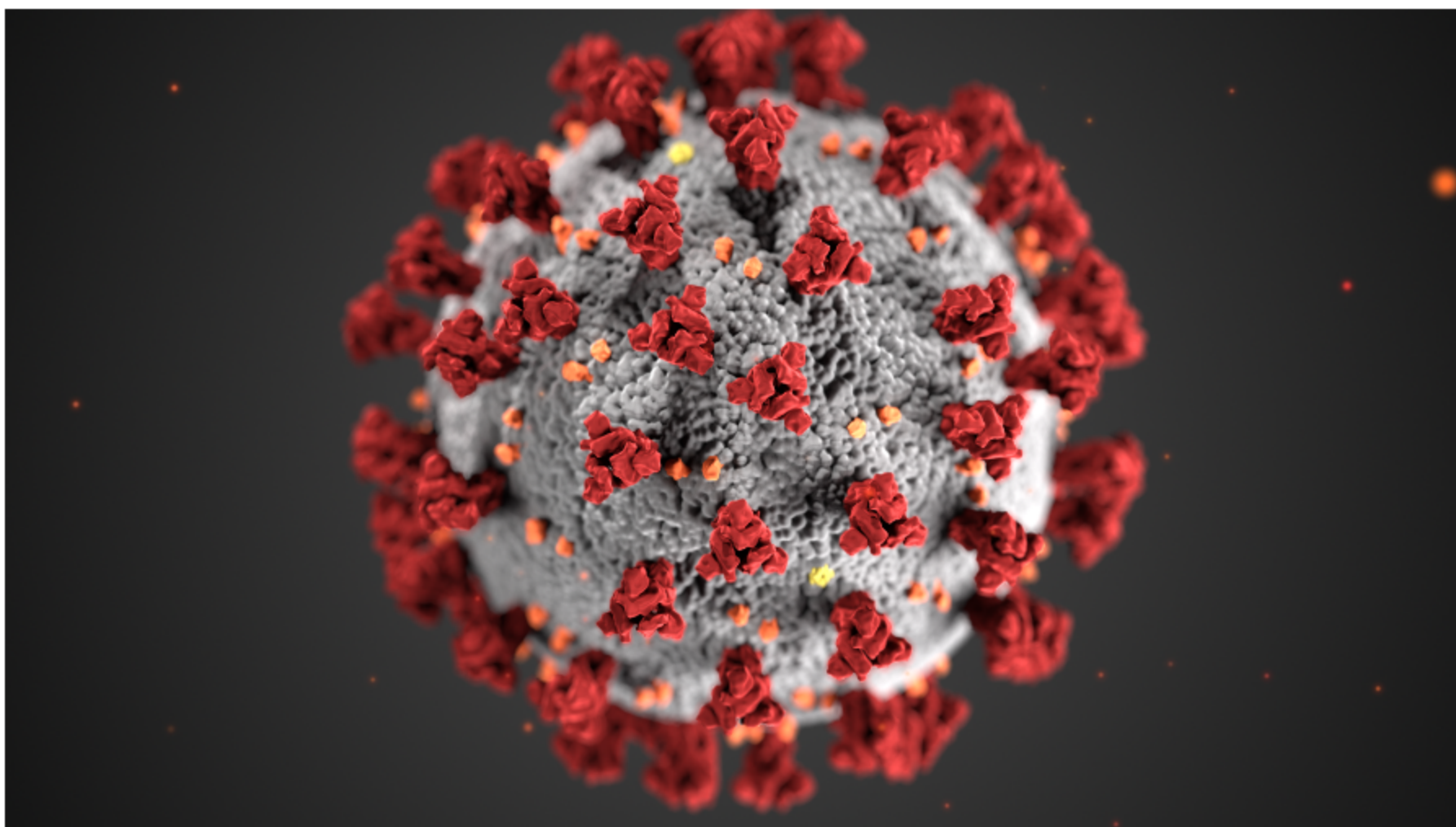
분석



통계적 오류인지 확인하기 위해서 인구 비율에 따른 확진자 수를 구하였습니다.

$$\text{비율 계산} = \text{각 구별 코로나 확진자 수} / \text{구별 총 인구수} \times 100$$

문제



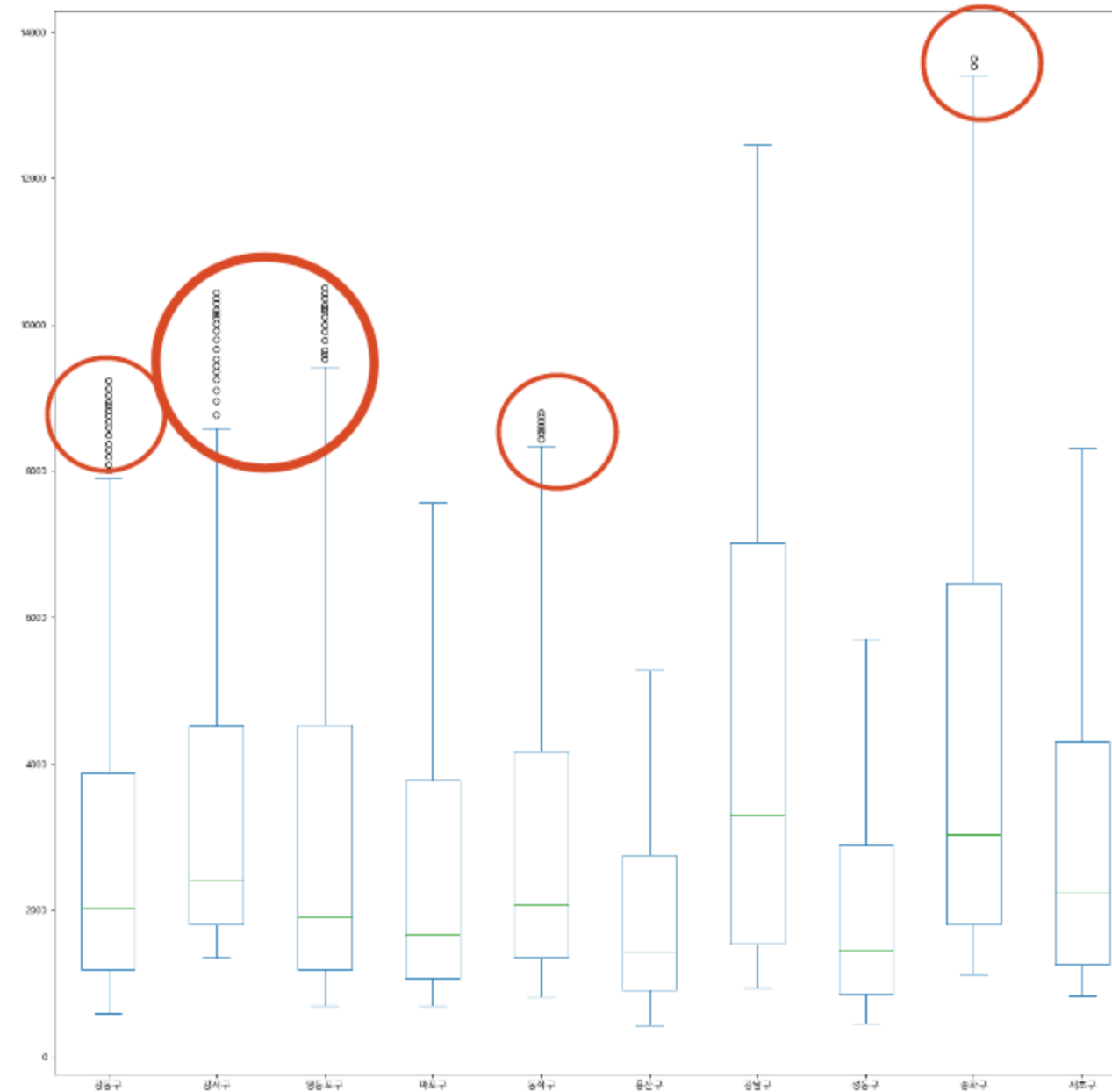
한강이 코로나에 주는 영향을 알아보기 위해 한강 근처 행정구를 조사하였습니다.

데이터 처리

데이터의 비교 행정구를 한강 근처의 행정구만 조사하여 코로나의 특별한 현상을 조사하기로 하였습니다.

데이터를 이용하여 각종 그래프를 그리던중 특이한 현상을 발견하여 박스 그래프로 확인하였습니다.

한강 이남 행정구에서만 발견되는 이상치 데이터를 발견하였습니다.



이상치 데이터의 정체

이상치는 타 행정구 구민 출신의 감염자입니다.

데이터 출처

<https://www.data.go.kr>



분석

분석

여러가지 데이터들을 찾아본 결과
Covid19 확진자 이상치와 공원의 상관관
계를 알게 되었습니다.

여기서 일반 공원이 아닌 '근린공원'과의
상관관계가 나타나 보입니다.

타 행정구 시민이 서울도심 근린공원에서
감염이 발생하는것으로 보여집니다.

참조
2020년 서울연구논문 공모전
텍스트 빅데이터 분석을 통한 COVID-19 전후 서
울시 주요 도시공원의 시민 이용행태 및 관심도
변화



자료: 국토교통부 한국토지정보시스템(KLIS) 용도지역지구도,
<https://www.data.go.kr/> (서울시공공데이터포털)의 서울시 주요 공원 좌표 활용



자료: 국토교통부 한국토지정보시스템(KLIS) 용도지역지구도,
<https://www.data.go.kr/> (서울시공공데이터포털)의 서울시 주요 공원 좌표 활용

문제



총균쇠라는 책에서는 전염병이 인류사에 미친 영향에 대해서 언급되어 있습니다.
코로나 데이터를 통해 이를 확인해보고자 합니다.

책에서의 주장

전염병이 경선을 따라서 전파되는 것보다 위선을 따라서 전파되는 속도가 빠르다!

데이터 처리

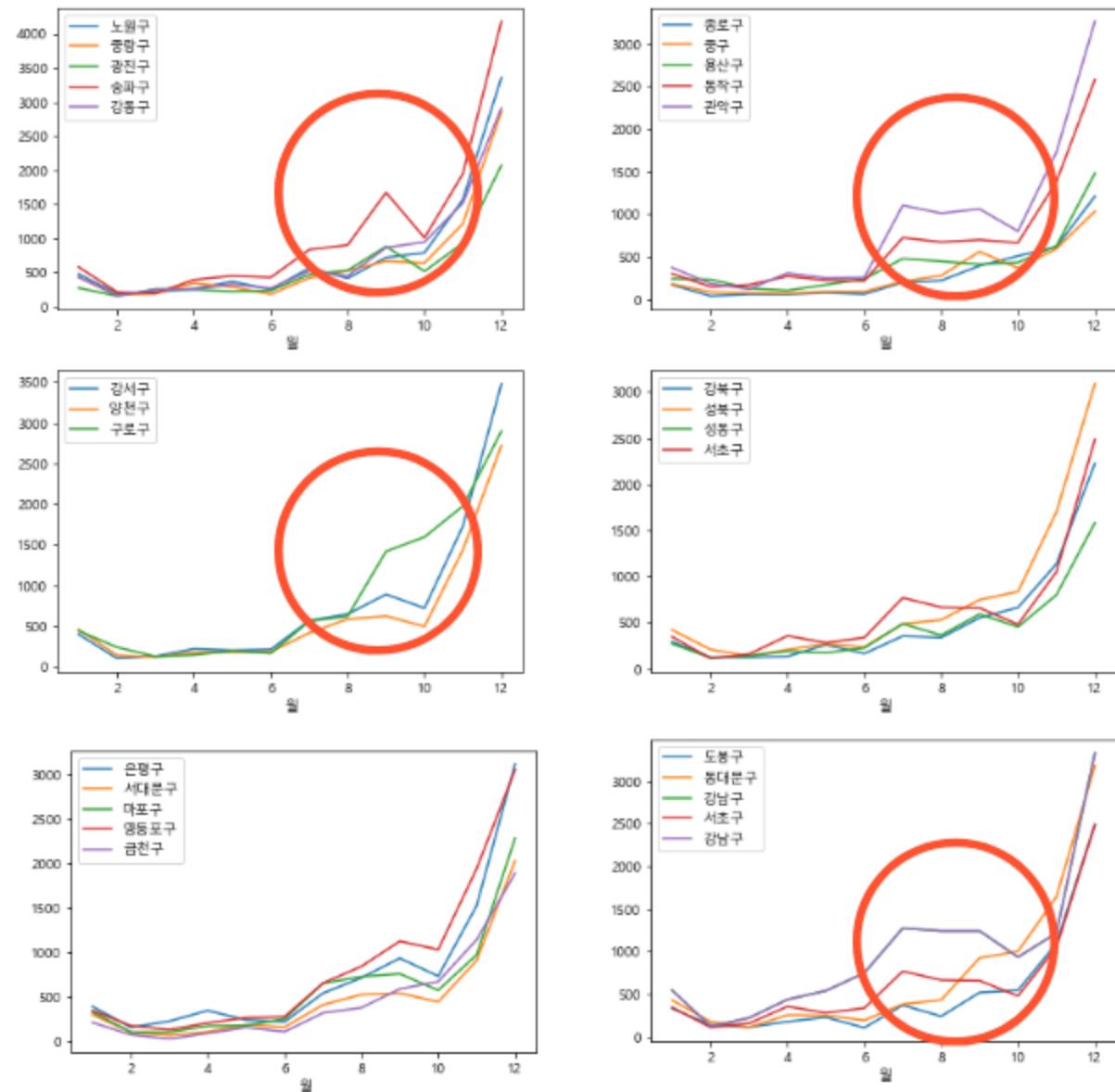


서울시를 가로와 세로로 그룹을 정해
준뒤 데이터를 비교 하였습니다.

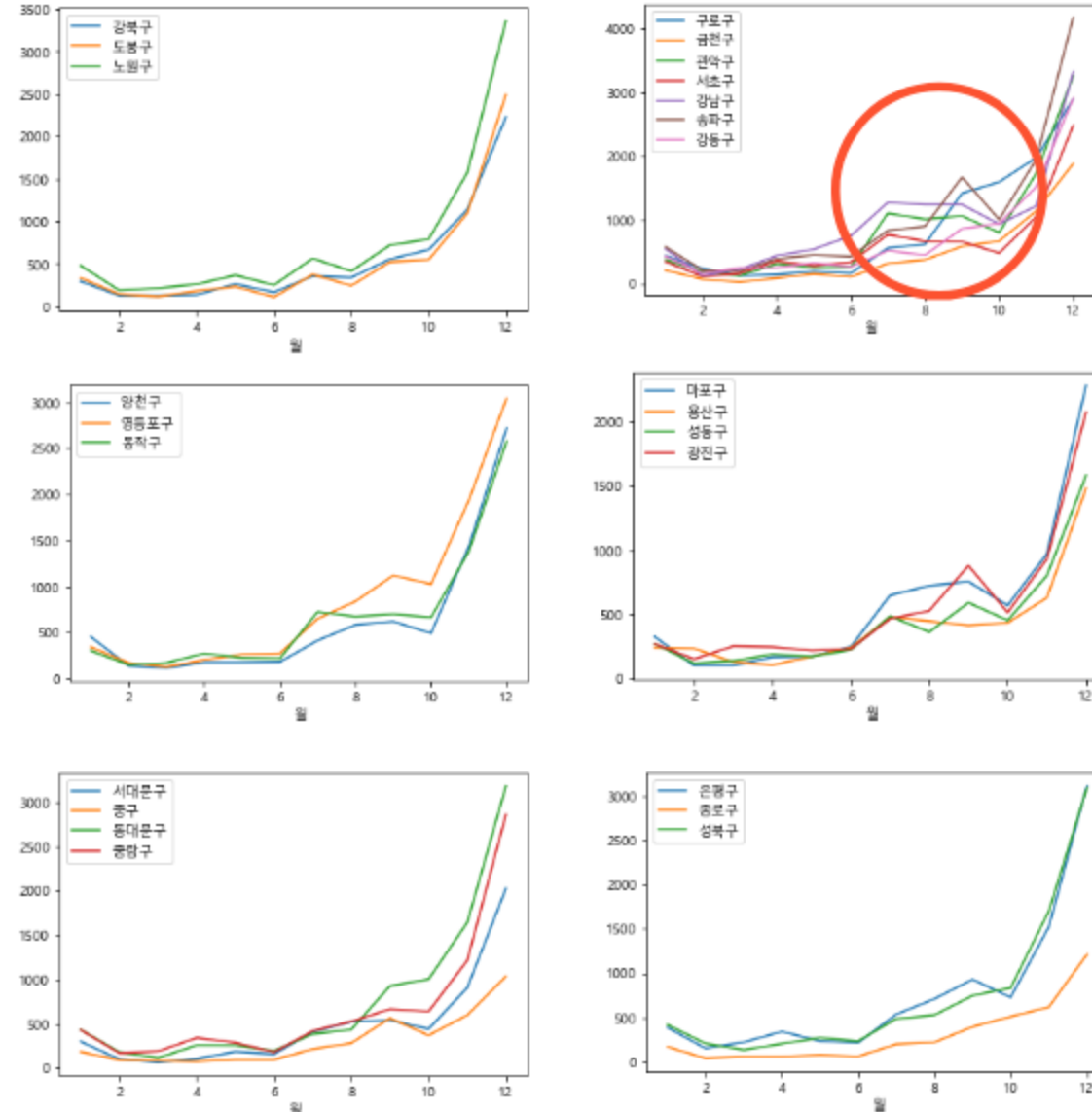
기본 그래프인 plot()이 가장 시각적
으로 보기 편하여 plot()으로 그래프
를 만들었습니다.

분석

경도에 따른 행정구 분류



위도에 따른 행정구 분류



결론
 총균쇠에서 주장하는 전염병 과 위도
 와 경도에 차이의 전파 양상의 경우
 Covid19의 관계의 연관성은 적용하기
 어렵습니다.

분석
 1. 일반적인 전염병과 다른 패턴을 보임
 2. 지역적, 교통적, 시대적인 상황이 달
 라서 적용 어려움

결론



데이터 분석을 통해서 이미 증명된 전염병 이론을 Covid19에 접목하기는 어렵습니다. Covid19는 기존의 전염병과는 다른 양상을 보이고 있습니다.

Covid19의 경우 감염의 경로를 추적하기 어렵고 서울처럼 인구 밀집도가 높고 유동인구가 많은 지역은 사람간의 접촉을 피하고 공원이나 사람이 몰릴것으로 예상되는 지역도 피하는것이 매우 현명합니다.



THE END.

Thank you
