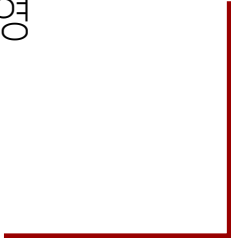




# 코로나 공공데이터 분석

허찬 이정진 이나영  
조민제 지현



# 2020 COVID-19 OUTBREAK

## 소상공인 매출 빅데이터 분석했더니...맞춤 지원 절실 > 뉴스 9...

2020. 4. 28. - [앵커] 코로나19로 손님들 발길이 끊긴 거리, 또 그만큼 많은 사람들이 일터를 떠나야 했습니다. 오늘(28일) 통계가 하나 나왔는데요, 지난달 국내 ...

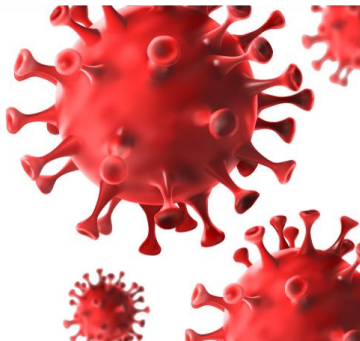
## 해외언론동향 - [Opinion] 빅 데이터 분석 및 자동 데이터 공유가 ...

2020. 3. 12. - 코로나바이러스(COVID-19) 감염으로 인해 하루에 천 건이 넘는 새로운 ... 매일 빅데이터의 분석과 공유를 어떻게 보장할 수 있으며, 또한 뉴스의 ...

## '코로나 맵' 이어 '코로나 알리미'

2020. 2. 2. - 민간에서 코로나 바이러스 관련 정보 지고 있다. 고려대 학... .. 코로나 알리미 사이트

**CORONAVIRUS  
OUTBREAK**  
COVID-19 ALERT



**빅데이터**

## 이태원 사태로 변한 국내 이동량은?...'통계청·SKT 빅데이터 분석 ...

5일 전 - 신종 코로나바이러스 감염증(코로나19)이 확산하면서 작년 대비 58% 수준까지 떨어졌던 국내 이동량이 황금연휴 때 83% 수준까지 회복한 것으로 ...

## 빅데이터로 살펴본 코로나 19의 기록 ① - 이슈&뉴스 | 국토 ...

2020. 4. 24. - 뉴스기사와 유동인구 데이터를 중심으로 - ... 다양한 빅데이터를 발굴·분석하여 코로나19 감염에 대한 국민의 사회적 인식변화 및 지역의 현안, ...

# Topic Selecion

코로나  
관련  
주가분석

주가데이터 분석을 통해  
코로나로 인한 주식시장  
의 영향 분석 및 예측

정부발표에  
따른  
뉴스데이터  
및 검색어  
변화 분석

대중교통, 생활인구 데이  
터, 검색어 기록을 이용해  
뉴스와 정부발표가  
미치는 영향 파악

재난지원금  
이후  
소비분석

카드지출데이터를 이용해  
재난지원금의 사용양상  
분석

지역 별  
최적의  
의료인프라  
예측

지역 별 병상수와 확진자  
수의 상관관계 파악,  
최적의 의료인프라  
예측모델링

코로나  
감염율의  
영향요인  
분석

감염율에 영향을  
미치는 사회적 요인을  
회귀모형을 이용해 분석

# Our Plan

---

1. 뉴스데이터 및 검색어 분석으로 사람들에 대한  
정부정책의 영향력 파악
2. 대중교통 데이터, 생활인구 데이터를 통해  
사람들의 실제 행동변화 분석
3. 정부의 지출액 중 보건의료 분야가 차지하는  
비중이 감염률에 미친 영향 분석
4. **감염률의 원인요소 모델링**

# OUR IDEA

## 위험소통과 정부신뢰:

### 빅데이터 감정분석을 통한 메르스 사례 연구<sup>1</sup>

최미정<sup>2</sup>, 은재호<sup>3</sup>

#### <Abstract>

우리 정부는 물론 재난 상황에 처한 많은 외국 정부도 재난 정  
않음으로써 적시 대응에 실패하는 사례가 비일비재하다. 대중에게  
두려움과 같은 부정 정서로 인해 비합리적 반응을 보일 것이라고  
은폐하는 경향이 있다는 것이다. 그러나 빅데이터 감정분석을 통  
분석한 이 연구 결과에 따르면, 재난 상황에 처한 대중의 집단  
우려하듯이 그렇게 비합리적이지 않으며, 오히려 정보가 투명하게  
때 정부에 대한 신뢰 역시 높아졌다. 즉, 메르스 감염자를 진료한  
혼란과 부정적 반응이 일어날 것이라는 정부의 예측과 달리 감

#### NBER WORKING PAPER SERIES

### DEMOGRAPHIC DETERMINANTS OF TESTING INCIDENCE AND COVID-19 INFECTIONS IN NEW YORK CITY NEIGHBORHOODS

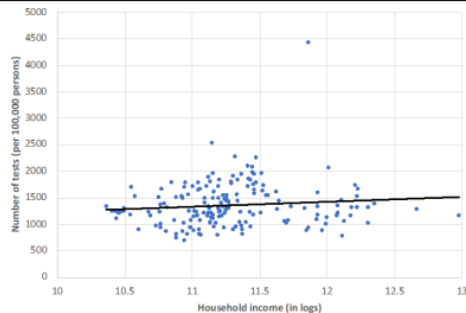
George J. Borjas

Working Paper 26952  
<http://www.nber.org/papers/w26952>

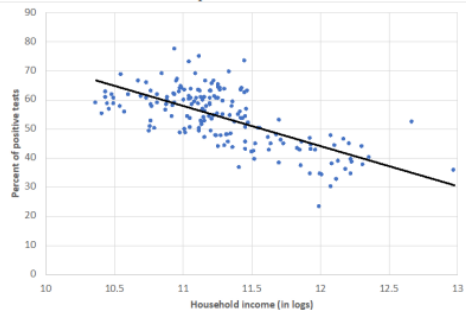
NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
April 2020

Figure 2. Household income and COVID-19

#### A. Incidence of testing (per 100,000 persons)



#### B. Percent of tests that are positive



# 1. 뉴스데이터/ 실시간검색어/ 금통위 의사록

뉴스데이터 및 검색어 분석으로  
사람들에 대한 정부정책의 영향력 파악

네이버 검색어 비교.csv

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 도움말 Easy Document Creator

붙여넣기 클립보드 글꼴 맞춤

A1 date

	A	B	C	D	E	F	G	H	I
1	date	cold	flu	pneumonia	coronavirus				
2	2016-01-01	0.11663	0.0559	0.15726	0.00736				
3	2016-01-02	0.13372	0.17135	0.20826	0.0089				
4	2016-01-03	0.14917	0.22317	0.19326	0.00845				
5	2016-01-04	0.17463	0.18626	0.29008	0.01145				
6	2016-01-05	0.17226	0.15072	0.24562	0.01381				
7	2016-01-06	0.17272	0.14399	0.25081	0.01381				
8	2016-01-07	0.1739	0.12408	0.25135	0.0119				
9	2016-01-08	0.16735	0.12499	0.29044	0.01572				
10	2016-01-09	0.13463	0.1019	0.2439	0.01036				
11	2016-01-10	0.15135	0.33062	0.21108	0.00699				
12	2016-01-11	0.18444	0.50643	0.26353	0.01299				
13	2016-01-12	0.17963	0.34608	0.21962	0.01572				
14	2016-01-13	0.17026	0.23244	0.22717	0.01163				
15	2016-01-14	0.16535	0.40198	0.21962	0.01227				
16	2016-01-15	0.16999	0.40853	0.2199	0.01081				
17	2016-01-16	0.14081	0.22972	0.17163	0.00727				
18	2016-01-17	0.1499	0.20881	0.15617	0.00599				

네이버 검색어 비교

파일명	압축크기	원본크기	압축률	종류
2019.11.29_제22차_금통위 의사...	20,841	100,140	70%	텍스트 문서
2019.12.12_제23차_금통위 의사...				
2019.12.26_제24차_금통위 의사...				
2020.01.17_제1차_금통위 의사록				
2020.02.27_제4차_금통위 의사록				
2020.03.12_제5차_금통위 의사록				
2020.03.16_제6차_금통위 의사록				
2020.03.26_제7차_금통위 의사록				
2020.04.09_제8차_금통위 의사록				
2020.04.16_제9차_금통위 의사록				

**자연어 처리  
워드 클라우드**

파일명	압축	원본	압축률	종류
코로나_뉴스_20200101-2020013...	2,016,181	2,016,181	0%	Microsoft Excel ...
코로나_뉴스_20200201-2020022...	11,876,206	11,873,302	0%	Microsoft Excel ...
코로나_뉴스_20200301-2020033...	12,457,501	12,454,720	0%	Microsoft Excel ...
코로나_뉴스_20200401-2020043...	11,279,722	11,277,415	0%	Microsoft Excel ...
코로나_뉴스_20200501-2020051...	6,033,189	6,032,111	0%	Microsoft Excel ...
Newsdata_description.txt	454	711	36%	텍스트 문서

## 2. 대중교통 데이터/ 생활인구 데이터

대중교통 데이터, 생활인구 데이터를  
통해 사람들의 실제 행동변화 분석

A1							
	A	B	C	D	E	F	G
1	사용일자	노선명	역ID	역명	승차총승객	하차총승객	등록일자
2	20200201	과천선	1451	경마공원	12159	12362	20200204
3	20200201	과천선	1452	대공원	2761	2916	20200204
4	20200201	과천선	1453	과천	3607	3224	20200204
5	20200201	과천선	1454	정부과천청	6872	6620	20200204
6	20200201	과천선	1455	인덕원	16458	16048	20200204
7	20200201	과천선	1456	평촌	9383	9239	20200204
8	20200201	과천선	1457	범계	21039	20529	20200204
9	20200201	분당선	1023	선릉	5028	8499	20200204
10	20200201	일산선	1954	백석	7603	7684	20200204
11	20200201	일산선	1953	대곡	1140	930	20200204
12	20200201	일산선	1952	화정	13309	13872	20200204
13	20200201	일산선	1951	원당	8109	7498	20200204
14	20200201	일산선	1950	삼송	12539	11597	20200204
15	20200201	일산선	1948	원흥	6031	5513	20200204
16	20200201	우이신설선	4713	신설동	1405	1340	20200204
17	20200201	우이신설선	4712	보문	1089	1120	20200204
18	20200201	우이신설선	4711	성신여대입	2129	2350	20200204

2020년\_버스노선별\_정류장별\_시간대별\_승하차\_인원\_정보(02월).csv - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 도움말 Easy Document Creator 어떤 작업을 원하시나요?

보여주기

클립보드

글꼴

맞춤

표시 형식

스타일

11가

자동 줄바꿈

일반

조건부 서식

표

셀 서식

스타일

A1

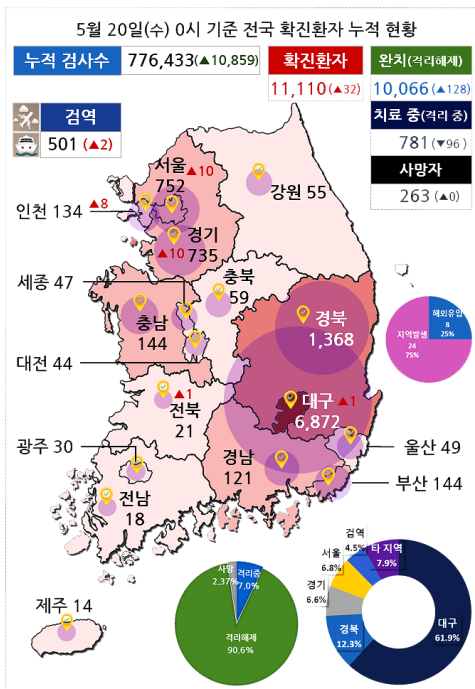
사용년월

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	사용년월	노선ID	노선번호	노선명	표준버스정류장ID	버스정류장명	역ID	역명	00시승차량	00시하차량	1시승차량	1시하차량	2시승차량
2	202002	11110001	100	100번(하)	1E+08	1003	7197	명륜3가.상	0	0	0	0	0
3	202002	11110001	100	100번(하)	1.01E+08	2156	8839	을지로입구	103	21	0	0	0
4	202002	11110001	100	100번(하)	1.02E+08	3267	8878	서빙고역교	1	0	0	0	0
5	202002	11110001	100	100번(하)	1E+08	1198	8952	원남동	0	0	0	0	0
6	202002	11110001	100	100번(하)	1.1E+08	11452	70562	하계극동0	0	0	0	0	0
7	202002	11110001	100	100번(하)	1.1E+08	11434	71603	골마을근로	0	0	0	0	0
8	202002	11110001	100	100번(하)	1.01E+08	2168	72254	을지로5가	78	5	0	0	0
9	202002	11110001	100	100번(하)	1E+08	1219	72259	통신대(이	53	55	0	0	0
10	202002	11110001	100	100번(하)	1.1E+08	11426	71604	하계역	0	0	0	0	0
11	202002	11110001	100	100번(하)	1E+08	1220	72260	혜화역.마	195	47	0	0	0
12	202002	11110001	100	100번(하)	1E+08	1229	72261	혜화역.동	329	88	5	0	0
13	202002	11110001	100	100번(하)	1.1E+08	11339	70563	월계헬스커	0	0	0	0	0
14	202002	11110001	100	100번(하)	1.1E+08	11311	71605	인덕대학	0	0	0	0	0
15	202002	11110001	100	100번(하)	1.01E+08	2206	72394	퇴계로6가	84	16	0	0	0
16	202002	11110001	100	100번(하)	1.01E+08	2219	72410	남대문시장	24	21	0	0	0
17	202002	11110001	100	100번(하)	1.1E+08	11310	71606	월계주공1	0	0	0	0	0
18	202002	11110001	100	100번(하)	1.01E+08	2142	72411	롯데영프라	57	11	0	0	0

2020년\_버스노선별\_정류장별\_시간대별\_승하차\_인원\_정보

# 3. 보건의료데이터

## 지역 별 코로나 감염률과 의료인프라 사이의 관계



1.지역별 중별 입원실 현황.xls [제한된 보기] - Excel

제한된 보기 주의하세요—인터넷에서 가져온 파일에는 바이러스가 있을 수 있습니다. 편집하지 않는다면 제한된 보기에서 여는 것이 안전합니다. 편집 사용(E)

A1 : 1.지역별 중별 입원실 현황

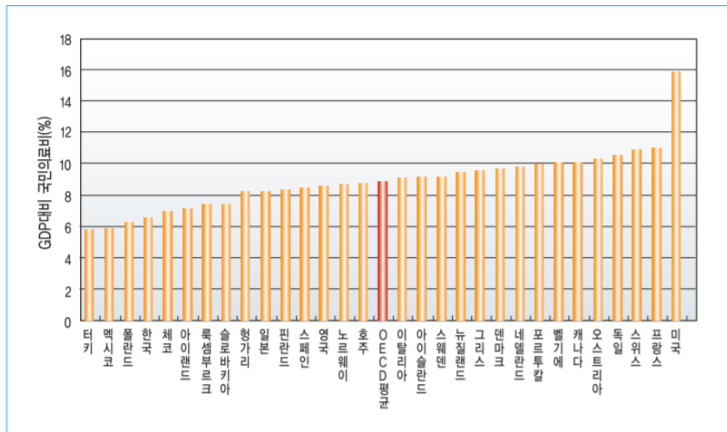
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1.지역별 중별 입원실 현황													
2														
3	지역별 중별 입원실 현황(단위: 실,개)													
4														
5	police_div ('police_main')SHEET_NAME ('지역별 중별 입원실 현황(단위: 실,개)')PARAM1 ('기초분기 ('202001')													
6														
7		구분상위	계	계	일반입원실	일반입원실	일반입원실	일반입원실	일반입원실	일반입원실	정신과폐쇄	정신과폐쇄	정신과폐쇄	정신과폐쇄
8		구분	계	계	일반입원실	일반입원실	일반입원실	일반입원실	일반입원실	일반입원실	정신과폐쇄	정신과폐쇄	정신과폐쇄	정신과폐쇄
9	시도	요양기관중	병상수	병상수	병상수	병상수	병상수	병상수	병상수	병상수	병상수	병상수	병상수	병상수
10	서울	상급종합병	5,307	17,221	1,303	1,303	3,552	13,751	-	-	17	17	81	314
11	서울	종합병원	4,324	16,195	735	735	3,214	13,780	27	196	1	1	32	146
12	서울	병원	5,267	18,967	1,110	1,110	3,713	15,145	-	-	12	12	358	2,406
13	서울	요양병원	4,573	21,984	951	1,568	3,574	20,141	-	-	-	-	19	179
14	서울	의원	4,454	9,795	2,506	3,493	1,859	5,888	-	-	17	41	72	373
15	서울	치과병원	29	83	13	23	16	60	-	-	-	-	-	-
16	서울	치과의원	38	65	32	57	6	8	-	-	-	-	-	-
17	서울	조산원	12	12	12	12	-	-	-	-	-	-	-	-
18	서울	한방병원	1,110	3,766	125	125	985	3,641	-	-	-	-	-	-
19	서울	한의원	324	948	97	176	227	772	-	-	-	-	-	-
20	부산	상급종합병	1,103	3,944	187	187	798	3,212	-	-	3	3	21	72
21	부산	종합병원	2,673	8,987	664	664	1,895	7,702	-	-	-	-	15	81
22	부산	병원	4,316	14,394	1,149	1,149	3,023	12,355	-	-	-	-	127	829
23	부산	요양병원	7,765	38,361	1,538	3,451	5,377	30,012	-	-	4	4	797	4,746
24	부산	의원	1,583	3,594	799	1,122	784	2,472	-	-	-	-	-	-
25	부산	치과병원	4	4	4	4	-	-	-	-	-	-	-	-
26	부산	치과의원	5	5	5	5	-	-	-	-	-	-	-	-
27	부산	조산원	2	2	2	2	-	-	-	-	-	-	-	-
28	부산	한방병원	227	814	36	36	191	778	-	-	-	-	-	-



# 3. 보건의료데이터

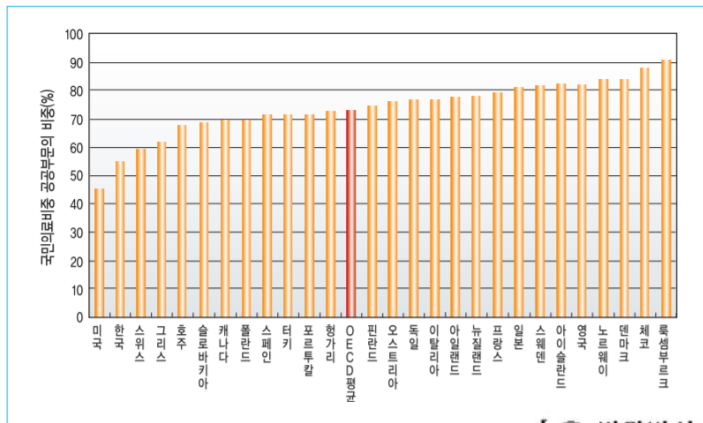
## 공공의료와 전염병 사이의 관계 파악 (전세계적 추이)

[그림 2] OECD 국가들의 GDP 대비 국민의료비 비중(2006년)



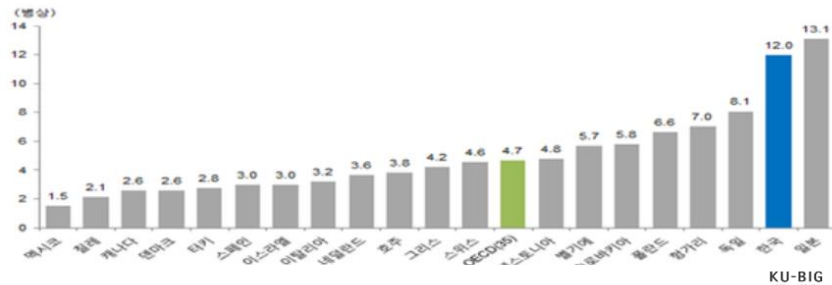
출처 : OECD, OECD Health Data File

[그림 5] OECD 국가의 국민의료비 중 공공부분의 비중(2006년)



출처 : OECD, OECD Health Data File

[ 총 병원병상 수 (2016) ]



## 4. 코로나 감염율의 영향요인 모델링

코로나19의 사회경제적 원인요인 분석

상관관계

소득수준

병상수

복지지출

정부발표

대중교통  
이용빈도

클럽 수

거주지역

인구밀도

AND  
MORE

감염율

QUESTIONS?

THANK YOU