

D&A Visualization Competition

나이썬 조
황건하, 김예향, 김서령, 김해우, 한준규

CONTENTS

01

문제 인식

02

데이터 전처리
&
분석

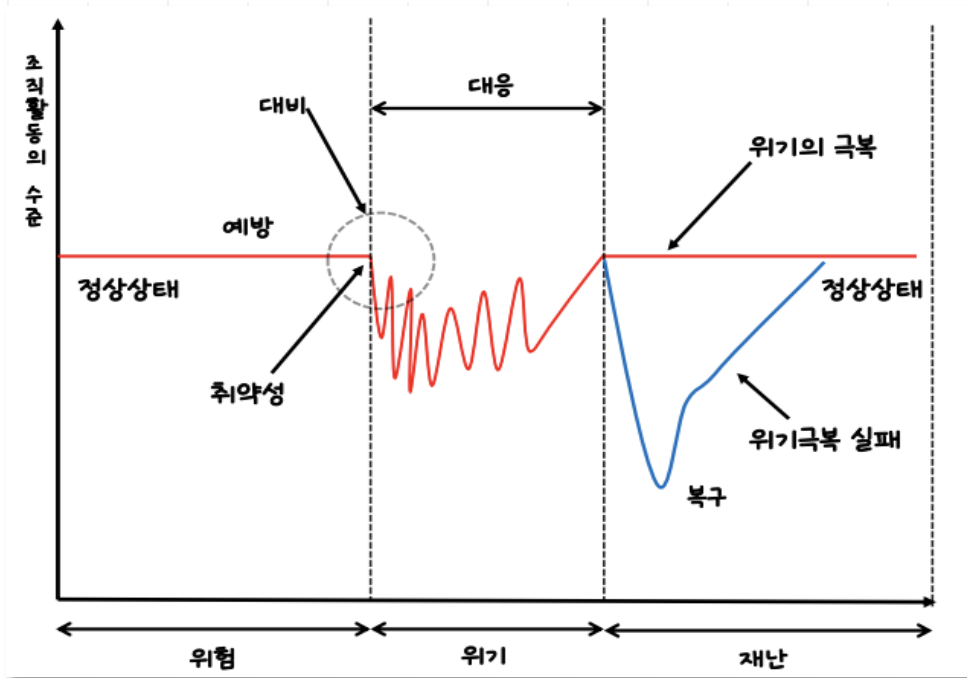
03

데이터 시각화

04

결론 도출

문제 인식



자료출처:
정지범, (2009), "광의와 협의의 위험, 위기, 재난관리의
범위", 한국방재학회논문집 제9권 4호, p64

“

**신종 인플루엔자, 메르스 등
이전 감염병 사례가 있었음에도
코로나 대응 정책이 미흡했다.**

Analyzing

Patient_info.csv 파일

```
patient_info = pd.read_csv('data/PatientInfo.csv')  
df = patient_info
```

```
df = patient_info  
df.head()
```

	patient_id	sex	age	country	province	city	infection_case	infected_by	contact_number	symptom_onset_date	confirmed_month	confirmed_d
0	1000000001	male	50s	Korea	Seoul	Gangseo-gu	overseas inflow	etc	75	2020-01-22	1	
1	1000000002	male	30s	Korea	Seoul	Jungnang-gu	overseas inflow	etc	31	2000-02-02	1	
2	1000000003	male	50s	Korea	Seoul	Jongno-gu	contact with patient	2002000001	17	2000-02-02	1	
3	1000000004	male	20s	Korea	Seoul	Mapo-gu	overseas inflow	etc	9	2020-01-26	1	
4	1000000005	female	20s	Korea	Seoul	Seongbuk-gu	contact with patient	1000000002	2	2000-02-02	1	

Cleansing / Analyzing

데이터 전처리

patientinfo의 'confirmed_date'
결측치 '2000-02-02'로 대체

patientinfo의 'confirmed_date'
타입 datetime으로 변경

데이터 전처리 과정에서 데이터 손실을 막기 위해 결측치를 특이값으로 변환

```
df['confirmed_date'] = df['confirmed_date'].fillna('2000-02-02')
```

```
df['confirmed_date'] = pd.to_datetime(df['confirmed_date'], format = '%Y-%m-%d')
```

Analyzing

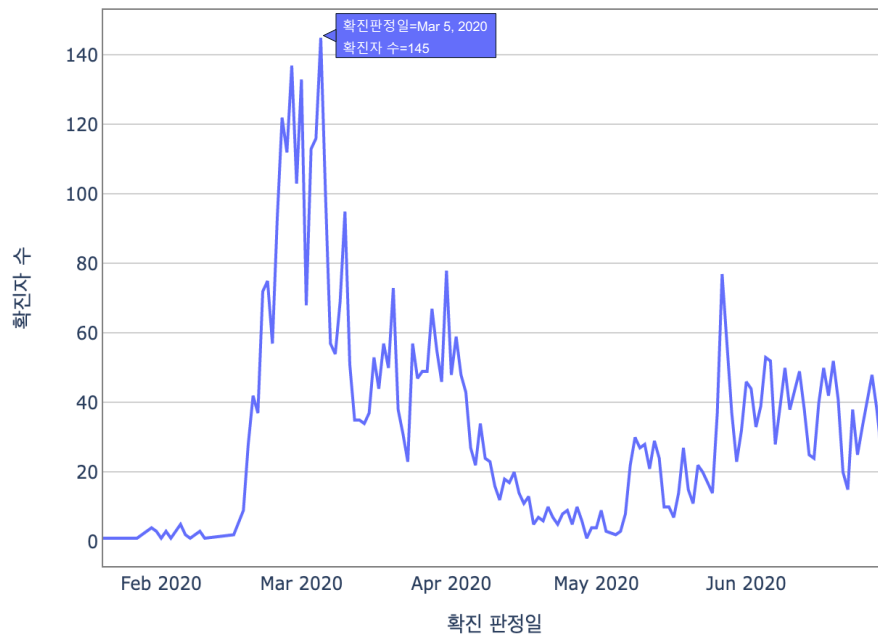
patientinfo의
'confirmed_date'에서
'2000-02-02'를 가지지 않는
값만을 고름

groupby로
'confirmed_date'별로
'patient_id'열을 count

```
c = df['confirmed_date'].isin(['2000-02-02'])  
df_c = df[~c]  
df_c
```

```
df_con_count = df_c.groupby('confirmed_date')[['patient_id']].count().reset_index()  
df_con_count
```

일별 확진자 수 현황



Analyzing

Case.csv 파일

```
case = pd.read_csv('data/Case.csv')
```

	case_id	province	city	group	infection_case	confirmed	latitude	longitude
0	1000001	Seoul	Yongsan-gu	True	Itaewon Clubs	139	37.538621	126.992652
1	1000002	Seoul	Gwanak-gu	True	Richway	119	37.48208	126.901384
2	1000003	Seoul	Guro-gu	True	Guro-gu Call Center	95	37.508163	126.884387
3	1000004	Seoul	Yangcheon-gu	True	Yangcheon Table Tennis Club	43	37.546061	126.874209
4	1000005	Seoul	Dobong-gu	True	Day Care Center	43	37.679422	127.044374
...
169	6100012	Gyeongsangnam-do	-	False	etc	20	-	-
170	7000001	Jeju-do	-	False	overseas inflow	14	-	-
171	7000002	Jeju-do	-	False	contact with patient	0	-	-
172	7000003	Jeju-do	-	False	etc	4	-	-
173	7000004	Jeju-do	from other city	True	Itaewon Clubs	1	-	-

Analyzing

Case의 'group'에서
False가 아닌 행만 추출

```
c = case['group'].isin([False])  
case_c = case[~c]
```

'infection_case'를 기준으로
'confirmed' 열을 sum 으로
groupby

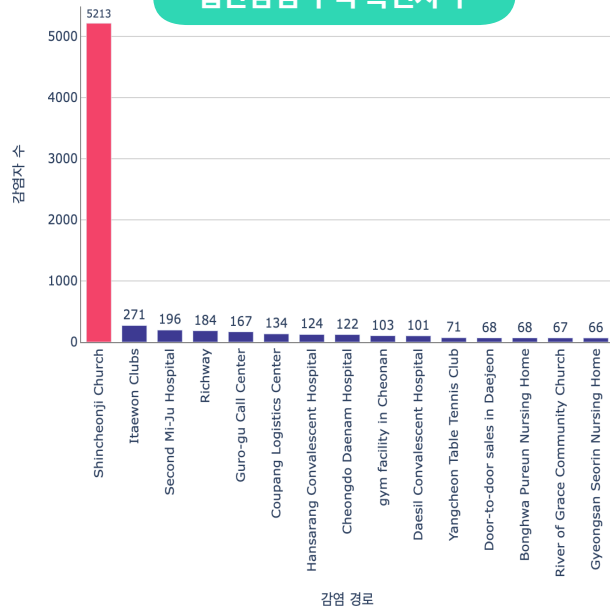
```
case_c_g = case_c.groupby('infection_case')[['confirmed']].sum()  
case_c_g
```

상위 15개의 집단감염
누적 확진자 수 막대 그래프

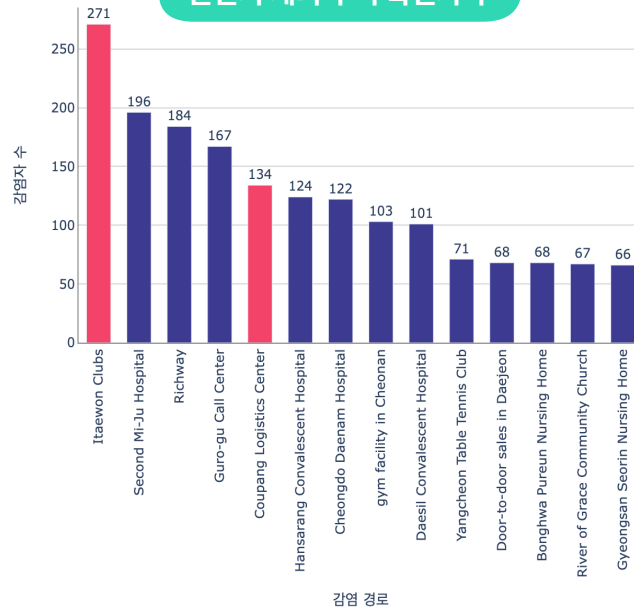
```
sor_case_c_g=case_c_g.sort_values(by='confirmed', ascending=False)[:15]  
sor_case_c_g=sor_case_c_g.reset_index()
```

집단감염별 누적 확진자 수

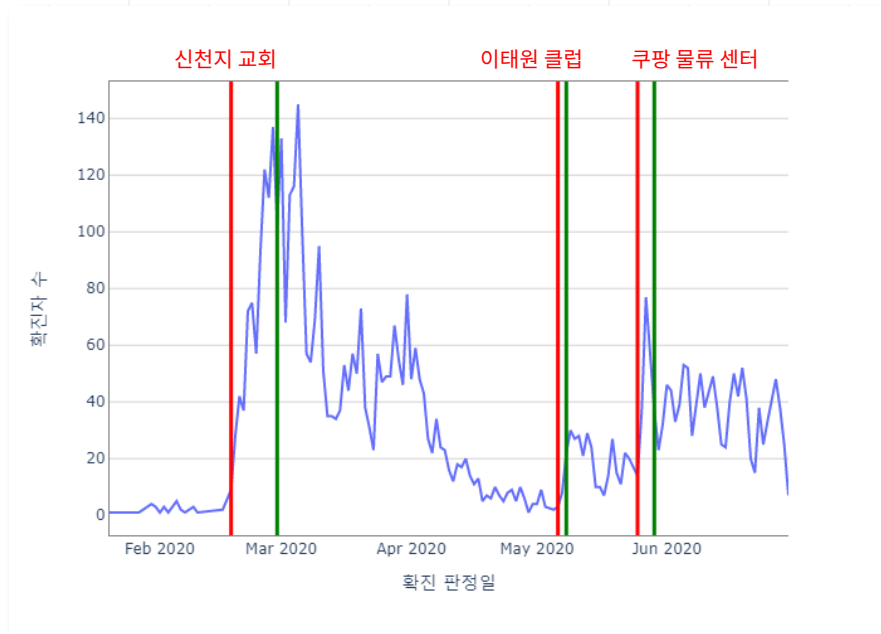
집단감염 누적 확진자 수



신천지 제외 누적 확진자 수



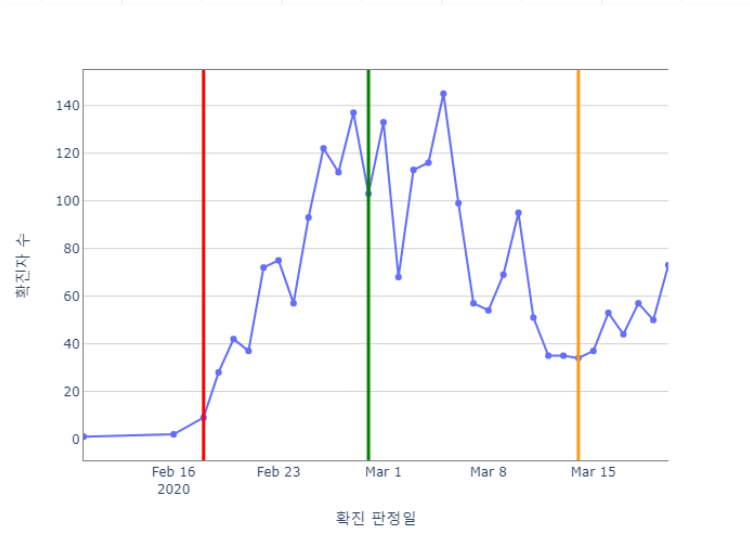
집단감염 시작일과 정책 시행일



— 집단감염 시작일
— 정책 시행일

신천지 교회 집단감염 추세

- 집단감염 시작일
- 정책 시행일
- 정책 시행 2주 후



2월 18일 집단감염 발생



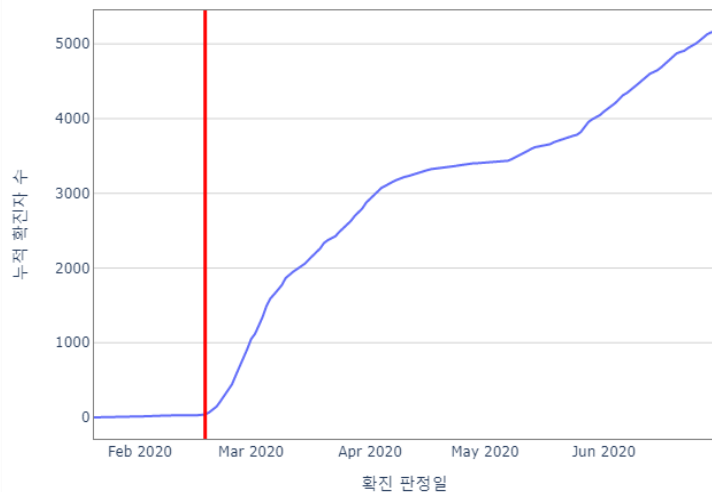
2월 29일 사회적 거리두기 시행



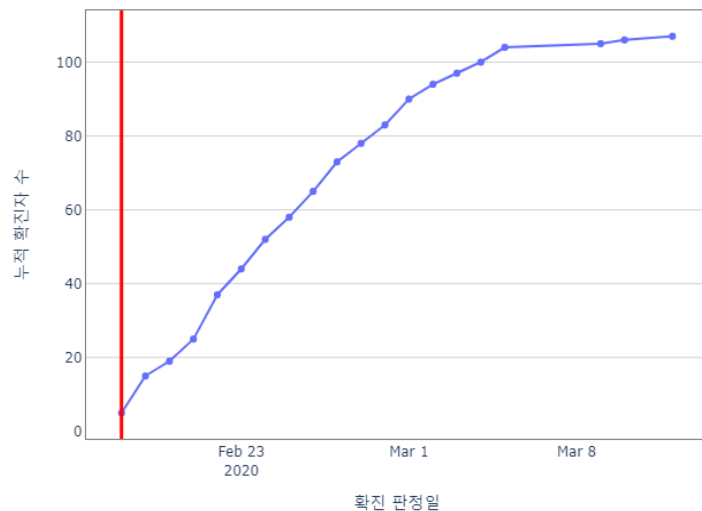
정책 시행 2주 후 증가 추세

31번 확진자 확진 판정일 기준 전후 비교

전체 누적 확진자 수

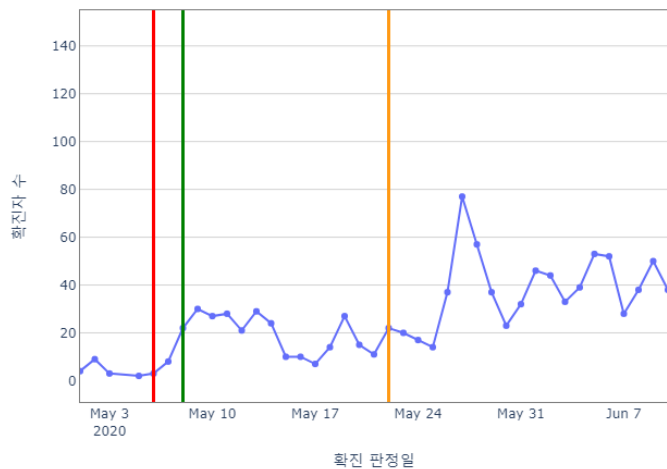


신천지 누적 확진자 수



이태원 클럽 집단감염 추세

- 집단감염 시작일
- 정책 시행일
- 정책 시행 2주 후



5월 6일 집단감염 발생



5월 8일 클럽과 술집 영업제한 시행



정책 시행 2주 후 감소 추세

Analyzing

confirmed_date 열을
인덱스로 변경

```
df_date = df.set_index('confirmed_date')  
df_date
```

age 열의 결측치 제거

```
df_date = df_date[['age']].dropna()  
df_date
```

5/22 기준으로 전후 데이터 추출

```
df_date_1 = df_date.loc['2020-05-08' : '2020-05-22']  
df_date_1
```

```
df_date_2 = df_date.loc['2020-05-22' : '2020-06-05']  
df_date_2
```

* 5/22은 정책 시행 2주 후 날짜

Analyzing

○ 20대

age별 value_counts

```
df_date_1_counting = df_date_1['age'].value_counts()  
df_date_1_counting
```

20s	48
30s	26
50s	17
40s	13
60s	8
10s	8
70s	6
0s	5
90s	1
80s	1

Name: age, dtype: int64

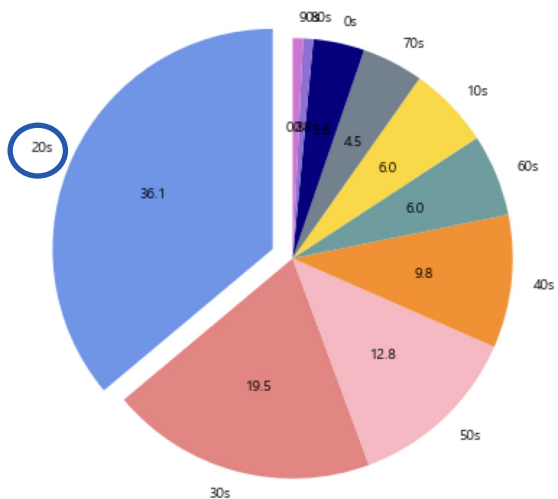
```
df_date_2_counting = df_date_2['age'].value_counts()  
df_date_2_counting
```

30s	37
50s	27
60s	22
20s	22
40s	16
10s	9
70s	6
80s	6
90s	2
0s	1

Name: age, dtype: int64

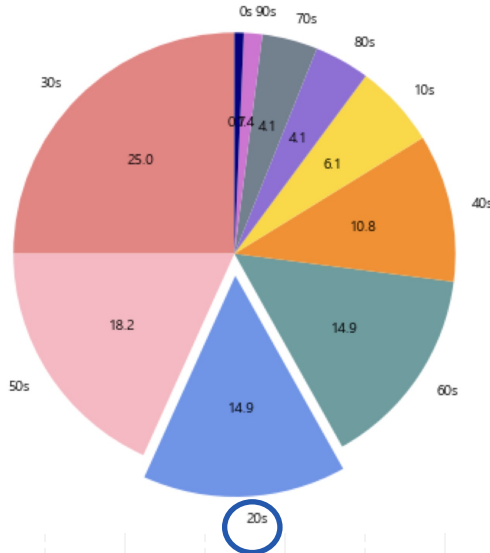
정책 시행 전후 연령대별 확진자 비교

5/22 이전 연령대별 확진자 비율



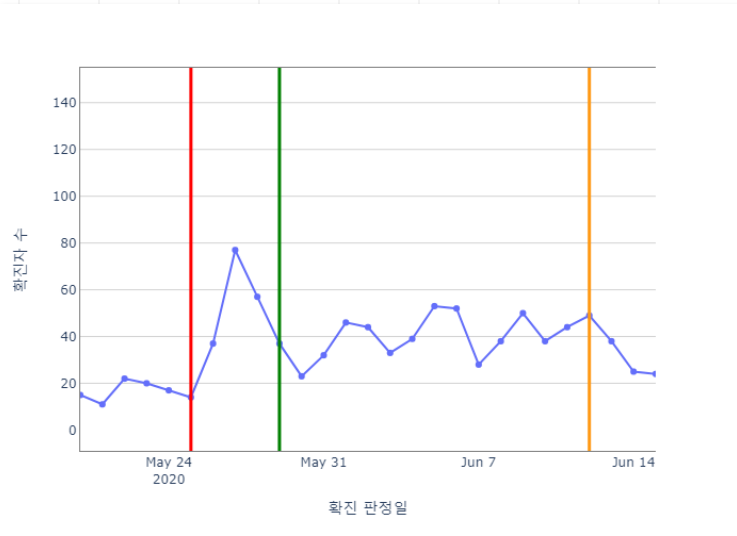
○ 20대

5/22 이후 연령대별 확진자 비율



쿠팡 물류 센터 집단감염 추세

- 집단감염 시작일
- 정책 시행일
- 정책 시행 2주 후



5월 25일 집단감염 발생



5월 29일 물류시설 현장 점검 시행



정책 시행 2주 후 감소 추세

Analyzing

Case의 province 에서
서울, 인천, 경기만 추출

Province를 기준으로 merge

쿠팡 집단감염 사례만 추출

```
pro = case['province'].isin(['Seoul', 'Incheon', 'Gyeonggi-do'])
case_pro = case[pro]
case_pro.reset_index()
```

```
pro_1 = df['province'].isin(['Seoul', 'Incheon', 'Gyeonggi-do'])
df_pro_1 = df[pro_1]
df_pro_1.reset_index()
```

```
new_csv = pd.merge(case_pro, df_pro_1, how='inner', on='province')
```

```
cou = new_csv['infection_case_x'].isin(['Coupang Logistics Center'])
new_csv = new_csv[cou]
```

Analyzing

confirmed_date 열을
인덱스로 변경

6/12일 기준으로 전후
데이터 추출

* 6/12은 정책 시행 2주 후 날짜

```
new_csv = new_csv.set_index('confirmed_date')
```

```
new_csv_1 = new_csv.loc['2020-05-29' : '2020-06-12']  
new_csv_1
```

```
new_csv_2 = new_csv.loc['2020-06-12' : '2020-06-26']  
new_csv_2
```

Analyzing

province 열을 value_counts

```
new_csv_1_counting = new_csv_1['province'].value_counts()  
new_csv_1_counting
```

```
Seoul          256  
Gyeonggi-do    210  
Incheon        104  
Name: province, dtype: int64
```

```
new_csv_2_counting = new_csv_2['province'].value_counts()  
new_csv_2_counting
```

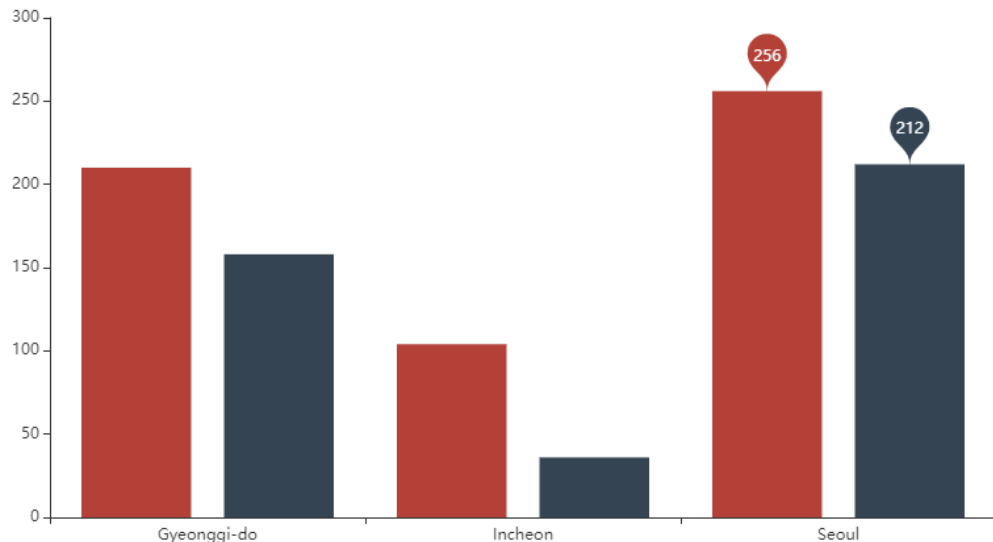
```
Seoul          212  
Gyeonggi-do    158  
Incheon        36  
Name: province, dtype: int64
```

6/12 전후 서울, 인천, 경기 확진자 비교

지역별 확진자

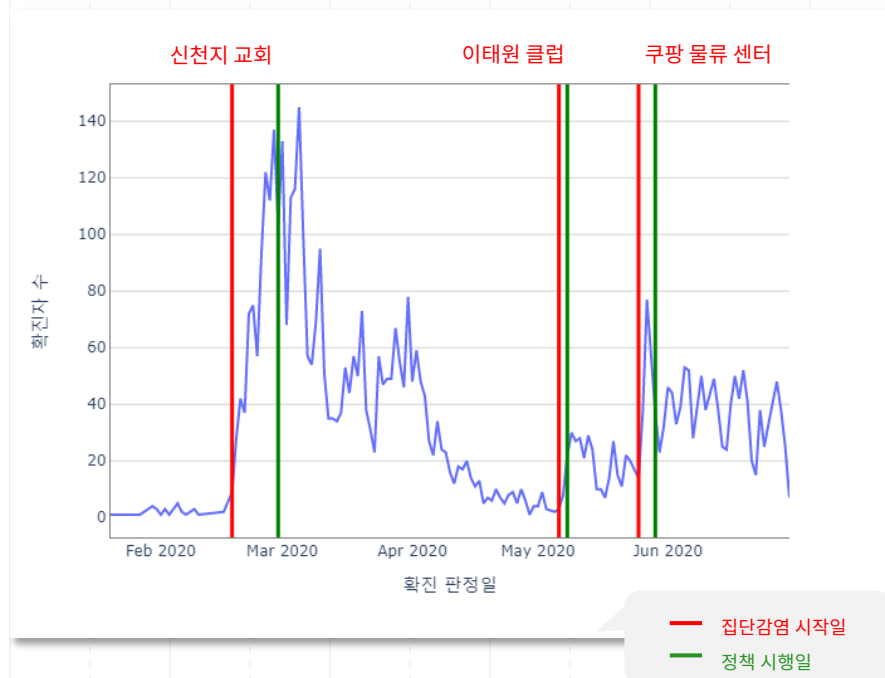
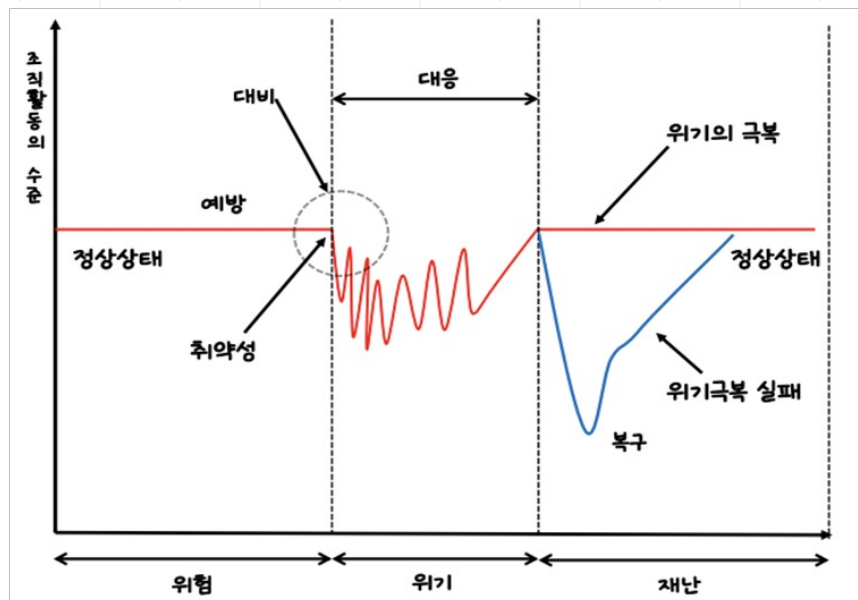
쿠광물류센터

■ 06-12 이전 ■ 06-12 이후



* 6/12은 정책 시행 2주 후 날짜

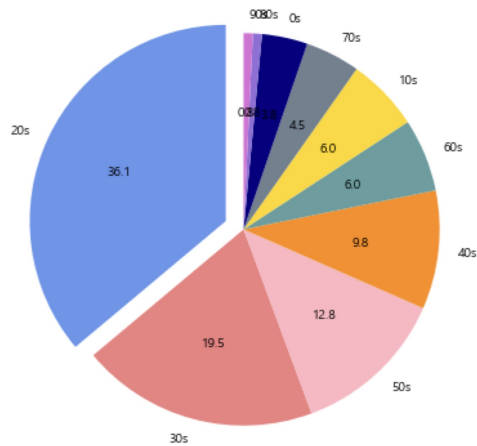
결론



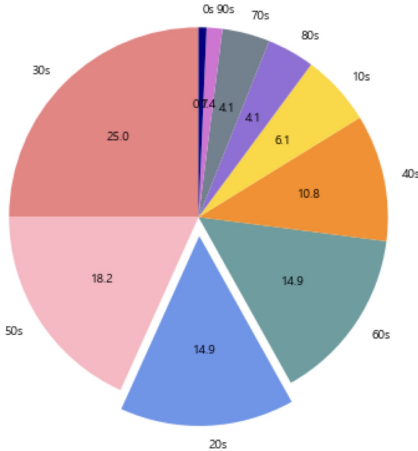
결론

이태원 집단감염

5/22 이전



5/22 이후

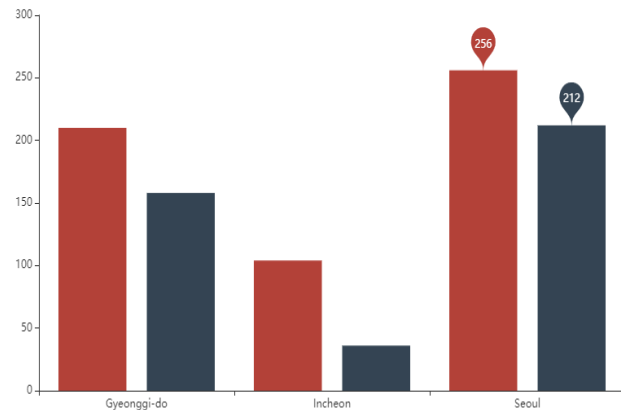


쿠팡 물류센터 집단감염

지역별 확진자

쿠팡물류센터

06-12 이전 06-12 이후



결론



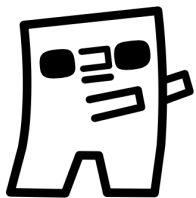
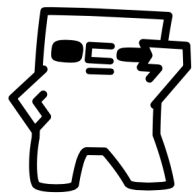
- 자연적인 재난은 막을 수 없음.
- 재난 발생 전 철저히 예방 및 대비를 할 필요가 있음.
- 신종 인플루엔자, 메르스, 코로나 사례와 같이 질병 사례가 많이 있었는데도 불구하고 대비를 못하고 있는 것이 현실.



- 우리 나이썬 팀은 이런 상황을 예방하고 대한민국 국민들과 정부에 ‘안일함’과 ‘경각심’을 주고자 데이터 분석을 실시함.
- 데이터 시각화를 통해 정부와 국민들이 재난 상황에서의 “골든타임”의 중요성을 기억하며 예방 및 대비를 철저히 했으면 좋겠음.

한계점

- 주어진 데이터에 비해 변수를 많이 사용하지 않아서 그래프가 단조로워 짐.
- 공간 빅데이터 분석을 하고 싶었지만 역량 부족으로 시각화를 못함.
- 데이터 병합 능력 부족으로 주제에 맞는 적절한 데이터를 찾아보지 않음.
- GitHub 같은 협업 툴의 중요성을 알 수 있었음.



감사합니다!

