

2011 제 35대- 2021 제 38대

# 서울시장 핵심공약 키워드로 본 사회이슈

김보영 김유진 김채란 남기봉 윤지윤

# INDEX

1

팀원 소개

2

분석 주제 및 목표

3

데이터 수집 방법

4

분석 프로세스

# 5

## 분석 결과



## 1 팀원 소개

김보영 김유진 김채란 남기봉 윤지윤

## 2

분석 주제 및 목표

# 서울시 공약의 효과 분석

# 주제선정 과정

## 선정 배경

현재 진행 중인 서울시장 보궐선거의 주요 이슈를 파악하고자 함.  
더불어 과거에 있었던 서울시장 선거들의 주요 이슈와는 어떤 차이점이 있는지  
그 변화 양상을 알고자 함.

## 주제

서울시장 관련 언론보도를 통해 알아본 당대 서울시 이슈

## 방법

역대 후보자들의 공약에 대한 뉴스 기사의 빈출 단어를 워드 클라우드로 시각화

# 논의 대상

01  
#One

## 선거 범위

이번 선거 + 과거의 선거를 어디까지 다뤄야 유의미 한가

02  
#Two

## 불용어 처리

언론사 관련, 선거 방식 관련 불용어  
기타 불용어 처리 과정 상 주관 개입 문제

03  
#Three

## 키워드 선정

# 논의 결과

## 01 #One

### 10년 간의 선거기간 선정

선거일 기준 한 달

-35대 2011.09.27~2011.10.26

-36대 2014.05.04~2014.06.03

-37대 2018.05.13~2018.06.12

-38대 2021.03.01~2021.03.24

## 02 #Two

### 불용어 구글 스프레드시트 공유파일 생성

## 03 #Three

### ‘서울시장 핵심공약’ 키워드 선정

‘서울시장 공약’ -> 관련 없는 기사 다수

‘서울시장 핵심공약 + 문화/노동/... -> 지나친 주관 개입



# 3 데이터 수집방법

N | 서울시장 핵심공약

통합 VIEW 이미지 지식IN 동영상 쇼핑 뉴스 어학사전 지도 ... 검색옵션 ^

정렬 기간 영역 유형 언론사 기자명 옵션유지 상세검색

전체 1시간  
1일 2시간  
1주 3시간  
1개월 4시간  
6개월 5시간  
1년 6시간

직접입력

2021.03.01

2021.03.26

적용

검색결과 자동고침

뉴스1 1일 전 네이버뉴스

'서울시장' 핵심 부동산정책...박영선 vs 오세훈 '공급-규제' 셈법은?  
(오세훈 국민의힘 서울시장 후보) 오는 4월 서울시장 보궐선거를 앞두고 오세훈 국민의힘 후보가 야권 단일화에 성공하면서 박영선 더불어민주당 후보와 결전을...

박영선 '공공' vs 오세훈 '민간'...서울시 부동산정책 공... 이뉴스투데이 1일 전  
박영선-오세훈, 닮은 듯 다른 부동산 공약... 공공 vs... 아시아투데이 19시간 전



디스커버리뉴스 1일 전

서울시관광협회, 박영선 서울시장 후보에 서울관광 핵심공약 제...  
그 시작으로 서울시관광협회는 지난 19일, 박영선 더불어민주당 서울시장 후보에게 무너진 서울관광시장 재건방안과 글로벌 서울관광을 위한 핵심공약을 제시하...



연합뉴스TV PICK 18시간 전 네이버뉴스

서울시장 후보들 "재건축 규제 완화"...집값 자극받나

달리 공약이 현실화할지는 미지수입니다. <박원갑 / K8국민은행 부동산 수석전문위원> "서울시장의 권한을... 규제의 핵심인 재건축초과이익환수제의 내용을 바...



데일리안 PICK 1일 전 네이버뉴스

[4·7 재보선] 단일화로 '시동' 늦게 걸렸지만...오세훈 자치구별 공...  
국민의힘 서울시장 후보가 자치구별 공약 경쟁에 자신감을 보이고 있다. 오세훈 후보는 공식선거운동기간... 자치구별 핵심 정책을 담은 영상이 올라오고 있다. 19...



코로나19  
확진현황 및 백신-접종 정보

선별 진료소  
내 주변 진료소 찾기

코로나19 팩트체크  
백신 정보 팩트는?

코로나19 접종센터  
내 주변 접종센터 찾기

## 4 분석 프로세스

---

(1) 네이버 뉴스 크롤링 :

선거일 직전 한 달 간의 기사

검색어: '서울시장 핵심공약'

(2) 데이터 프레임 형성 :

Title/ Date/ Article/ URL/ PressCompany

## 4 분석 프로세스

---

(3) 데이터프레임 중 Article 토큰화

Noun 추출

공통 불용어 엑셀 파일 생성하여 불용어처리

(4) 워드 클라우드

빈도수 상위 100개 선정

서울시 지도 모양으로 시각화



## 5 워드 클라우드 결과









## 6 추가 데이터 시각화

# 추가적인 데이터 시각화

**이유** 구체적인 수치를 시각화하는 것이 필요하다고 판단

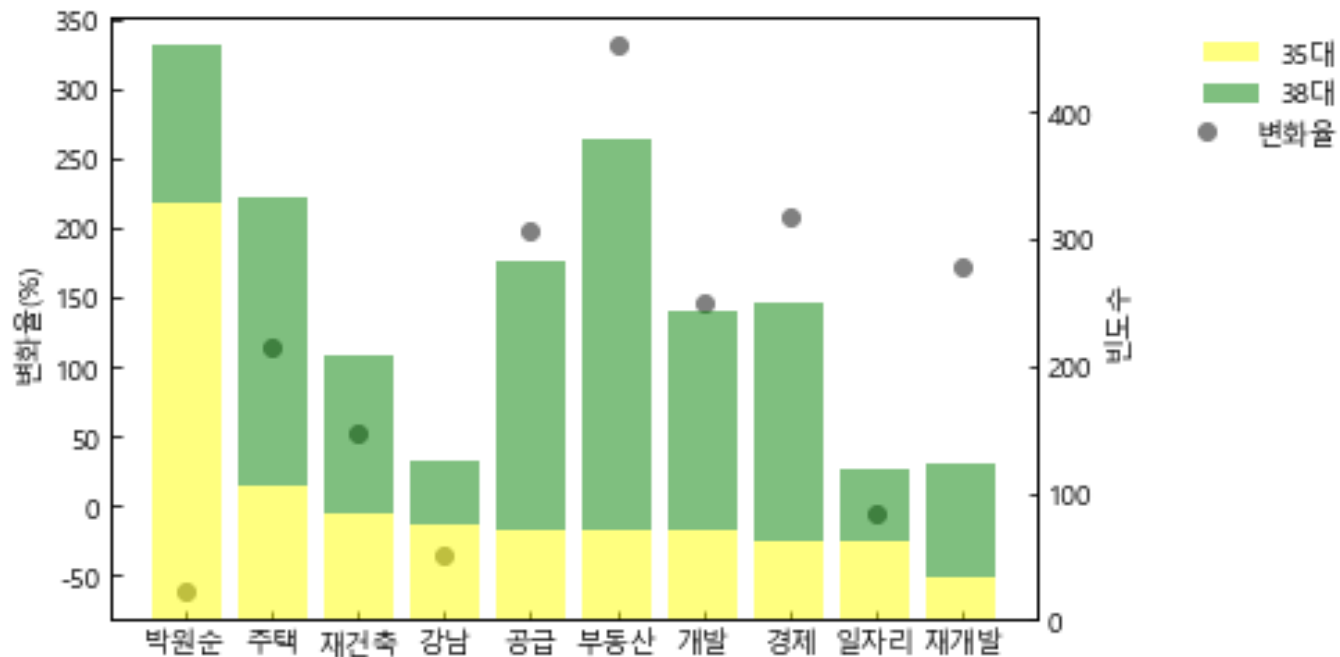
**주제** 공통 단어의 빈도수 변화 양상

**방법** 이전에 전처리했던 데이터를 이용

1. 35, 38대 (10년 전/후) 공통단어 빈도수 상위 10개의 변화
2. 35, 36, 37, 38대 (10년 동안의) 공통단어 빈도수 상위 10개의 변화

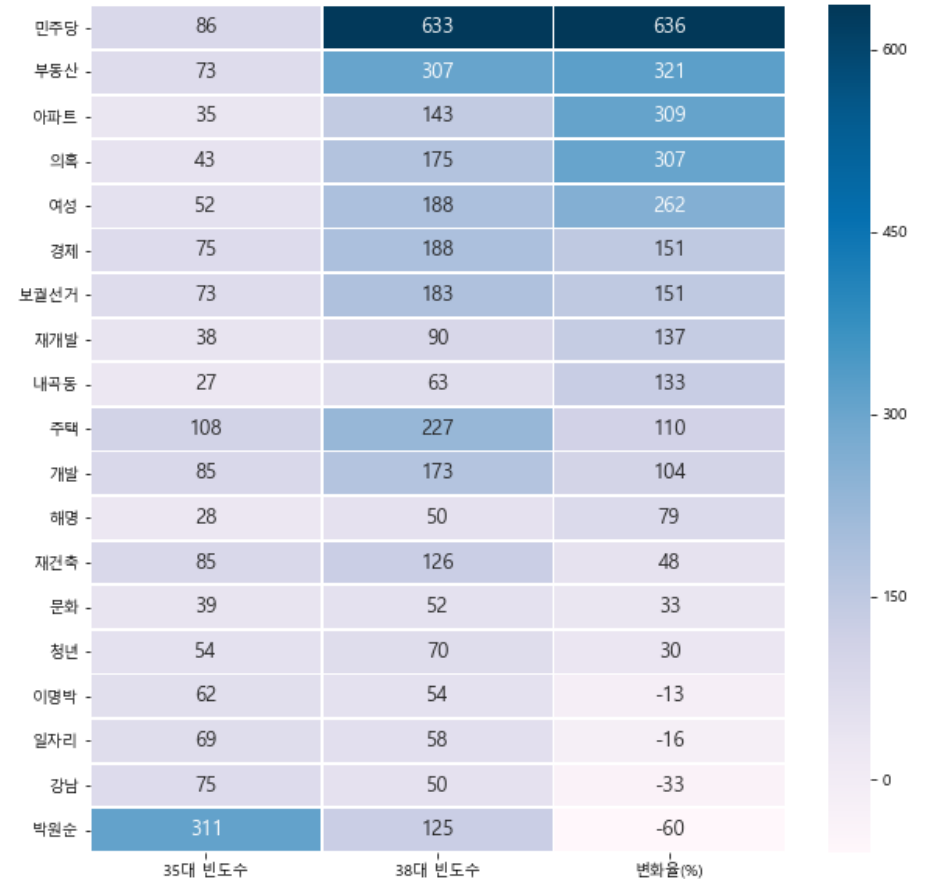


|   | word | 35대 | 36대 | 37대 | 38대 | 변화율        |
|---|------|-----|-----|-----|-----|------------|
| 0 | 박원순  | 328 | 209 | 161 | 125 | -61.890244 |
| 1 | 주택   | 106 | 58  | 23  | 227 | 114.150943 |
| 2 | 재건축  | 83  | 41  | 162 | 126 | 51.807229  |
| 3 | 강남   | 76  | 18  | 30  | 50  | -34.210526 |
| 4 | 공급   | 71  | 58  | 19  | 212 | 198.591549 |
| 5 | 부동산  | 71  | 22  | 62  | 307 | 332.394366 |
| 6 | 개발   | 70  | 128 | 76  | 173 | 147.142857 |
| 7 | 경제   | 61  | 86  | 55  | 188 | 208.196721 |
| 8 | 일자리  | 61  | 104 | 32  | 58  | -4.918033  |
| 9 | 재개발  | 33  | 46  | 86  | 90  | 172.727273 |



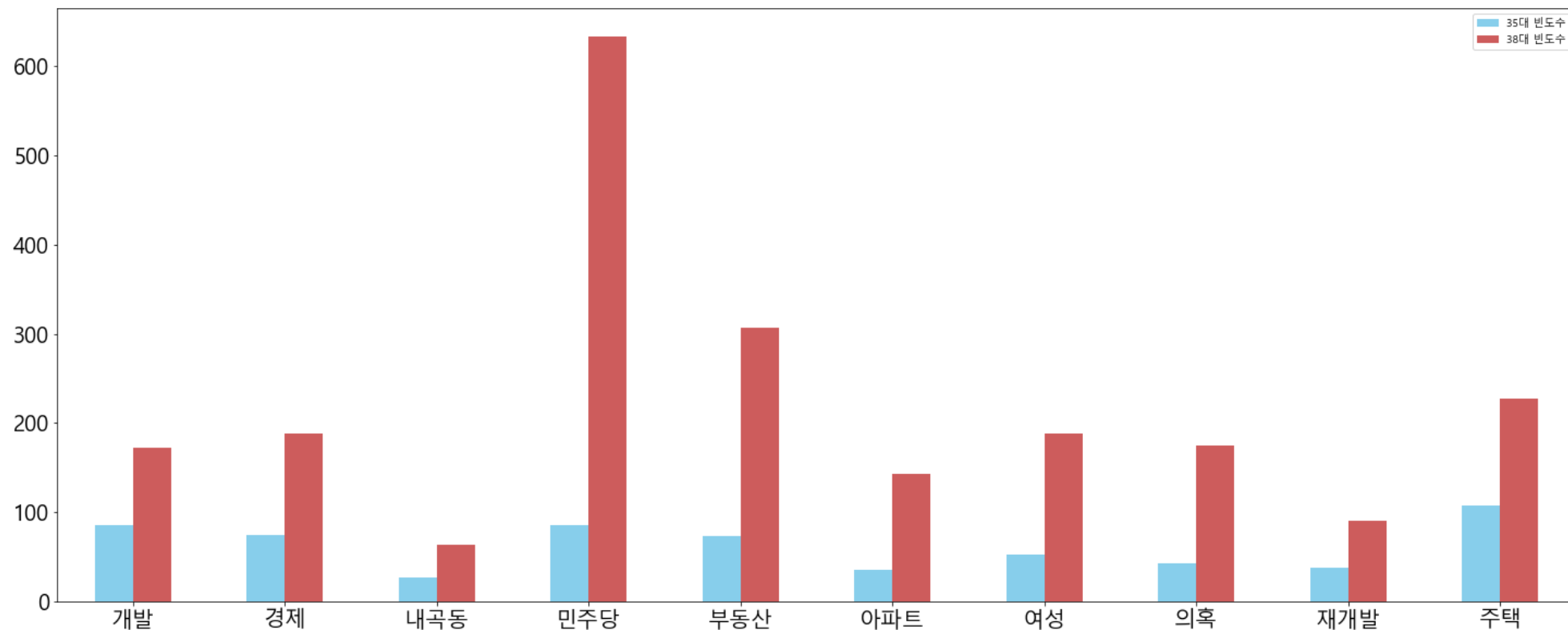
빈도수의 변화와 변화율을  
bar graph+marker로 시각화

## 중간 과정 시각화



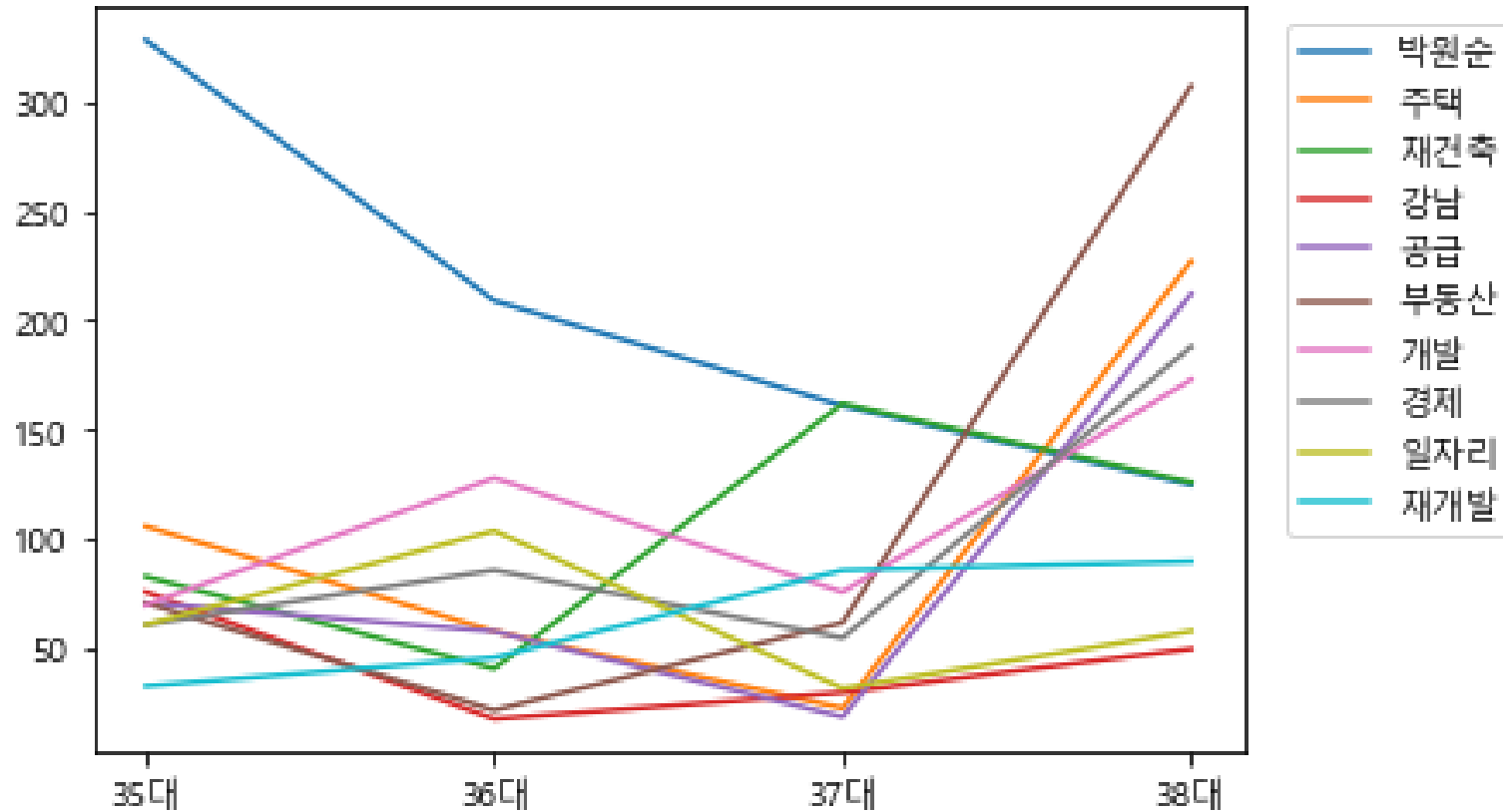
빈도수의 변화와 변화율을  
heatmap으로 시각화

# 최종 시각화 1



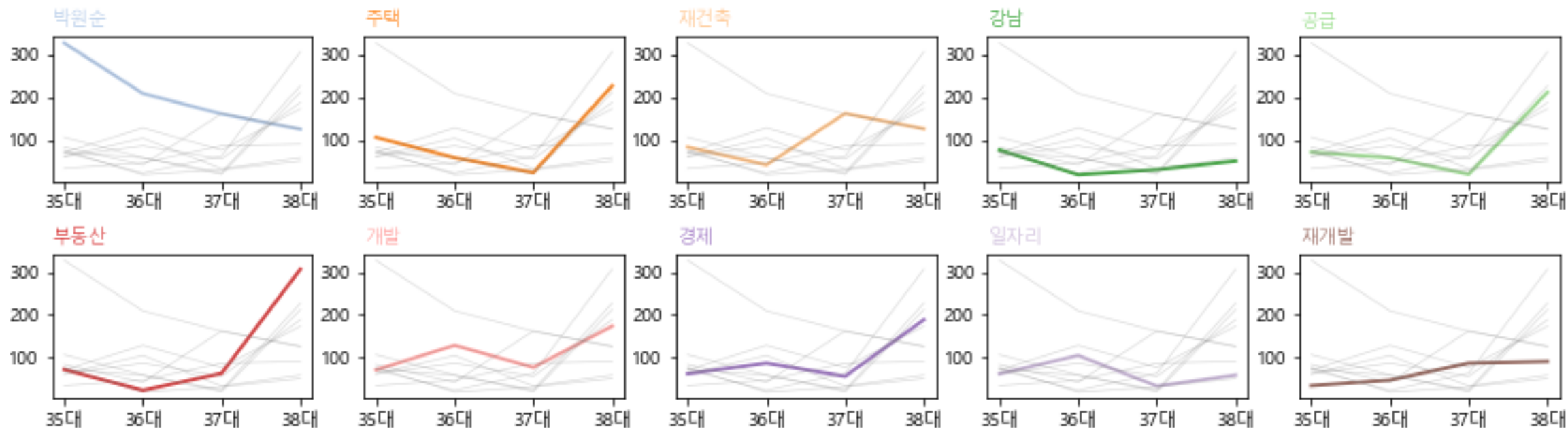
35, 38대 (10년 전/후)공통단어 빈도수 상위 10개의 변화

## 최종 시각화 2



35, 36, 37, 38대 (10년 동안의)공통단어 빈도수 상위 10개의 변화

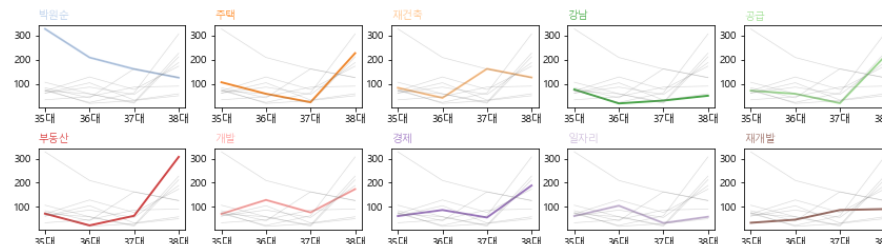
35대-38대 공통 키워드의 빈도수 변화



35, 36, 37, 38대 (10년 동안의)공통단어 빈도수 상위 10개의 변화

# CODE

35대~38대 공통 키워드의 빈도수 변화



```
num=0
for column in df.drop('x', axis=1):
    num+=1

    plt.subplot(2,5, num)

    plt.plot(df['x'], df[column], color=palette(num), linewidth=1.9, alpha=0.9, label=column)
    plt.title(column, loc='left', fontsize=10, fontweight=0, color=palette(num) )

    plt.subplots_adjust(top=0.8, right=2, hspace=0.5)

    for v in df.drop('x', axis=1):
        plt.plot(df['x'], df[v], color='grey', linewidth=0.6, alpha=0.3)

    plt.xticks([35,36,37,38], ['35대 ', '36대 ', '37대 ', '38대 '])

plt.suptitle("35대~38대 공통 키워드의 빈도수 변화", fontsize=13, color='black', x=1.1)
# plt.show()

plt.savefig('rank_result.png', bbox_inches = 'tight')
```

# 아쉬웠던 점

01  
#One

## 불용어 처리 문제

모든 불용어를 완전히 처리하지 못 했고, 수작업이 수반되었으나 한계  
→ 워드클라우드를 구성할 단어에 불용어가 포함된 부분

02  
#Two

## 단어 반영 한계

합성어이나, 고유 명사화 된 '청년 실업' '스마트시티' 등을 반영하지 못하는 한계

03  
#Three

## 시각화 방법 선정

활용할 수 있는 시각화 방법 매우 많아 선택하고, 이해하고, 구현하는 데까지 오랜 시간

## 추가 발전 방향

1

당대 이슈와  
후보자 당선 간의 상관성

한계점 : 데이터 양

2

정치 성향에 따른  
키워드의 변화 비교

한계점 : 키워드의 유사도, 필터링의 어려움



Thank  
you

