

금요일 조

이지안, 윤상진, 원대인, 전영호



1 배경

분석에 앞서 기본적인 인구수 파악을 위해 서브데이터의 주민등록인구 통계를 이용하여 서울시 자치구별인구수 상위/하위 10위를 매겨보았습니다.

자치구별 총인구 상위 10퍼센트 순위가 송파구, 강서구, 강남구, 노원구 순으로 높았는데 이 지역들은 공통적으로 지역내 인프라와 교통이 잘 형성되어있고 무엇보다 교육으로 유명합니다.

사용해야하는 나머지 메인데이터가 학원,학교,유치원 등의 교육에 집중되어 있기에 인구수 상위 지역들의 교육환경을 분석해보기로 하였습니다.

2

사용한 데이터



메인데이터

- ✓ 서울특별시 학원 교습소 정보
- ✓ 서울특별시 학교 기본정보
- ✓ 서울특별시 유치원 일반현황
- ✓ 서울시 주민등록 연양인구(연령별동계) 통계



서브데이터

- ✓ 서울시 주민등록인구 통계
- ✓ 서울시 유치원 통학차량 현황

3

전처리

과정

main1 학원

전처리

```
del_idx=main1[main1.도로명주소.isna() != False].index
del_idx=[i for i in del_idx]
main1.drop(del_idx,axis=0,inplace=True)
for i in range(len(main1)):
    try:
        if pd.isnull(main1['행정구역명'][i])==True:
            main1['행정구역명'][i]=main1['도로명주소'][i].split()[1]
    except KeyError:
        pass
main1.reset_index(drop=True,inplace=True)

del_idx_t=[]
for i in range(len(main1)):
    try:
        main1['등록일자'][i]=pd.to_datetime(str(main1['등록일자'][i])[0:4]+'-'+str(main1['등록일자'][i])[4:6]+'-'+str(
        main1['개설일자'][i]=pd.to_datetime(str(main1['개설일자'][i])[0:4]+'-'+str(main1['개설일자'][i])[4:6]+'-'+str
    except ValueError:
        del_idx_t.append(i)
    except KeyError:
        pass
main1.drop(del_idx_t,axis=0,inplace=True)
main1.reset_index(drop=True,inplace=True)

for i in range(len(main1)):
    try:
        main1['휴원시작일자'][i]=pd.to_datetime(str(main1['휴원시작일자'][i])[0:4]+'-'+str(main1['휴원시작일자'][i])[4:6]
        main1['휴원종료일자'][i]=pd.to_datetime(str(main1['휴원종료일자'][i])[0:4]+'-'+str(main1['휴원종료일자'][i])[4:6]
    except ValueError:
        pass

del_idx_cat=main1[main1.교습계열명.isna() != False].index
del_idx_cat=[i for i in del_idx_cat]
main1.drop(del_idx_cat,axis=0,inplace=True)
main1.reset_index(drop=True,inplace=True)

b=main1['일시수용능력인원합계'].isna()==True
main1[b]
```

...

```
main1['기숙사학원여부'].fillna('N',inplace=True)
main1['인당수강료내용'].fillna('불명',inplace=True)
main1['일시수용능력인원합계'].fillna(main1['정원합계'],inplace=True)
main1.reset_index(drop=True,inplace=True)
```

```
a=main1['휴원시작일자'].isna()!=True
main1[a]
```

3

전처리 과정

main2 학교

main2

...

#전처리

```
main2=main2.fillna(0)
main2['구']=main2['도로명주소'].str.split().str[1]
main2 = main2.drop_duplicates(['학교명'])
```

3

전처리

과정

main3 유치원

##전처리

#1. 중복삭제

```
main3.duplicated().sum()
```

#2. 결측값을 0으로 대체

```
main3=main3.fillna(0)
```

#3. 데이터 타입 변경

```
main3['설립일'] = main3['설립일'].astype('str')
main3['개원일'] = main3['개원일'].astype('str')
main3['만3세학급수'] = main3['만3세학급수'].astype(int)
main3['만4세학급수'] = main3['만4세학급수'].astype('int')
main3['만5세학급수'] = main3['만5세학급수'].astype('int')
main3['특수학급수'] = main3['특수학급수'].astype('int')
main3['만3세유아수'] = main3['만3세유아수'].astype('int')
main3['만4세유아수'] = main3['만4세유아수'].astype('int')
main3['만5세유아수'] = main3['만5세유아수'].astype('int')
main3['특수유아수'] = main3['특수유아수'].astype('int')
```

#4. 데이터 타입을 날짜로 바꿈

```
main3['설립일'] = pd.to_datetime(main3['설립일'])
main3['개원일'] = pd.to_datetime(main3['개원일'])
```

#5. "구" 열 생성

```
main3['구'] = main3['주소'].str.split().str[1]
```

3

전처리 과정

main4 주민등록연양인구

main4

...

#전처리

```
a = main4.groupby("동").get_group("소계")
a = a.set_index("자치구", drop=True)
a = a.drop(columns = '동')
```

```
for i in range(len(a.columns)):
    b = a.columns[i]
    a[b] = a[b].str.replace(",", '')
```

```
for i in range(3, 23):
    b = a.columns[i]
    a[b] = a[b].astype('float')
```

```
al = a.groupby("구분").get_group("계")
al['0~4세'] = al['0~4세'].astype('float')
al = al.drop(columns = ["계", "구분"])
```

```
boy = a.groupby("구분").get_group("남자")
boy = boy.drop(columns = ["계", "구분"])
```

```
girl = a.groupby("구분").get_group("여자")
girl = girl.drop(columns = ["계", "구분"])
```

3

전처리 과정

sub2 서울시 주민등록인구 (동별) 통계

sub2

...

#전처리

```
a = sub2.groupby("동").get_group("소계")
a = a.set_index("자치구")
a['세대'] = a['세대'].str.replace(",", "")
a['세대'] = a['세대'].astype('int')
a = a.drop(columns = '동')

a['총인구'] = a['총인구'].str.replace(",", "")
a['남자인구'] = a['남자인구'].str.replace(",", "")
a['여자인구'] = a['여자인구'].str.replace(",", "")
a['한국인인구'] = a['한국인인구'].str.replace(",", "")
a['한국인남자인구'] = a['한국인남자인구'].str.replace(",", "")
a['한국인여자인구'] = a['한국인여자인구'].str.replace(",", "")
a['외국인인구'] = a['외국인인구'].str.replace(",", "")
a['외국인남자인구'] = a['외국인남자인구'].str.replace(",", "")
a['외국인여자인구'] = a['외국인여자인구'].str.replace(",", "")
a['세대당인구'] = a['세대당인구'].str.replace(",", "")
a['65세이상고령자'] = a['65세이상고령자'].str.replace(",", "")

a['총인구'] = a['총인구'].astype('int')
a['남자인구'] = a['남자인구'].astype('int')
a['여자인구'] = a['여자인구'].astype('int')
a['한국인인구'] = a['한국인인구'].astype('int')
a['한국인남자인구'] = a['한국인남자인구'].astype('int')
a['한국인여자인구'] = a['한국인여자인구'].astype('int')
a['외국인인구'] = a['외국인인구'].astype('int')
a['외국인남자인구'] = a['외국인남자인구'].astype('int')
a['외국인여자인구'] = a['외국인여자인구'].astype('int')
a['세대당인구'] = a['세대당인구'].astype('float')
a['65세이상고령자'] = a['65세이상고령자'].astype('int')
```


3

전처리 과정

sub3 유치원 통학차량

#전처리

```
sub3['탑승가능인원']=0
for i in range(len(sub3)):
    sub3['신고차량수'][i]=sub3['9인승신고차량수'][i]+sub3['12인승신고차량수'][i]+sub3['15인승신고차량수'][i]
    sub3['탑승가능인원'][i]=sub3['9인승신고차량수'][i]*9+sub3['12인승신고차량수'][i]*12+sub3['15인승신고차량수'][i]*15

sub3['행정구역명']=0
for i in range(len(sub3)):
    sub3['행정구역명'][i]=sub3['주소'][i].split()[1]
```

4

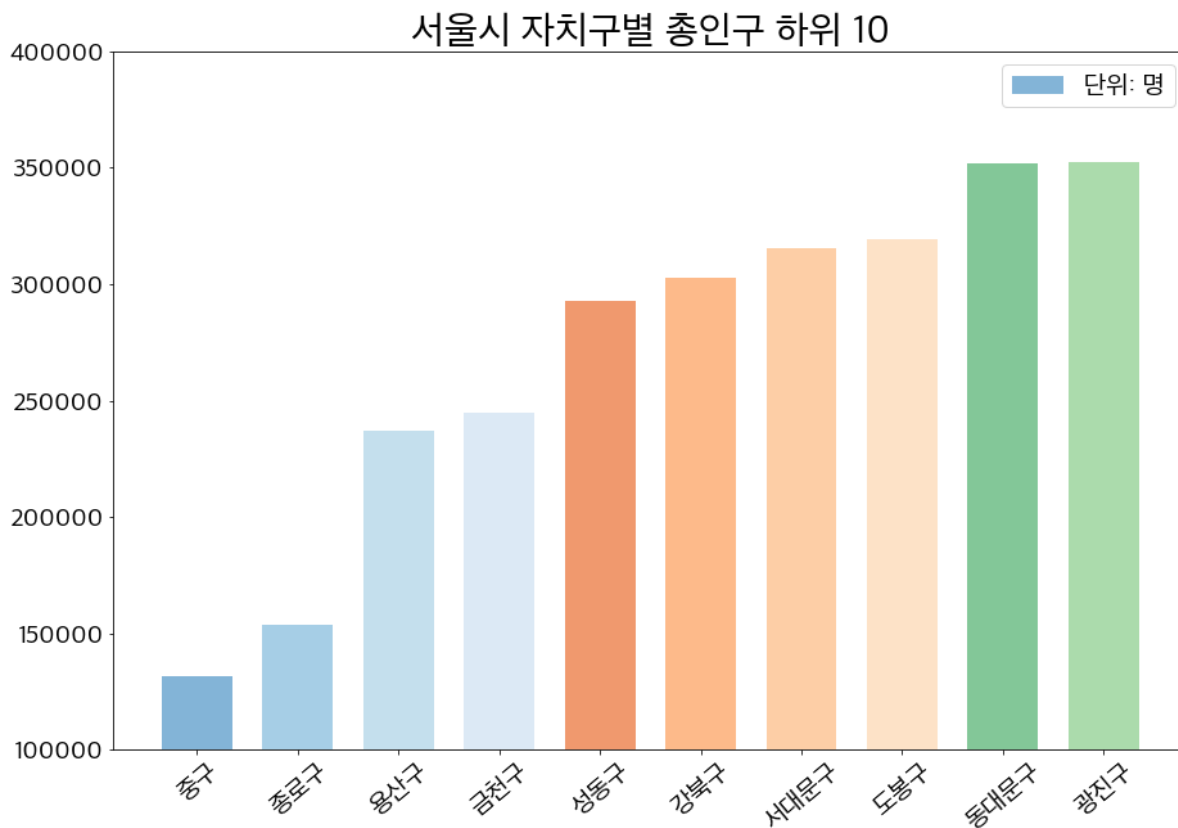
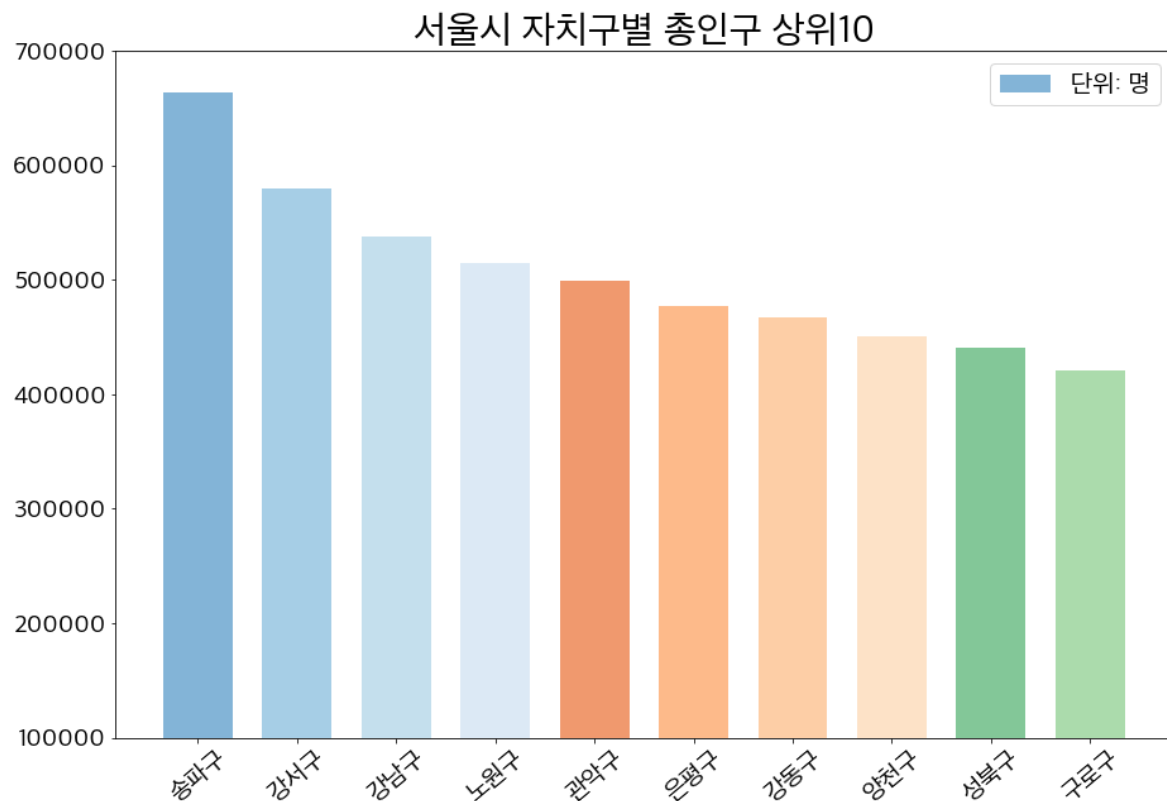
분석 과정



1. 주민등록인구통계 데이터를 이용해 서울시 자치구별 인구 상위/하위 10위 추출
2. 연양인구 데이터 이용 인구수 상위 4위까지 지역내 교육대상 인구수 추출
3. 학원데이터
 - 3-1 자치구별 학원개수 추출 및 대상인구와의 비교
 - 3-2 자치구별 학원 교습과정
4. 학교데이터
 - 4-1 각 자치구별 학교 개수 추출 및 대상인구와의 비교
5. 학원수에 따른 인구이동 분석
6. 유치원데이터
 - 6-1 각 자치구별 유치원 개수 추출 및 대상인구와의 비교
 - 6-2 유치원 공립/사립분석

4 분석과정

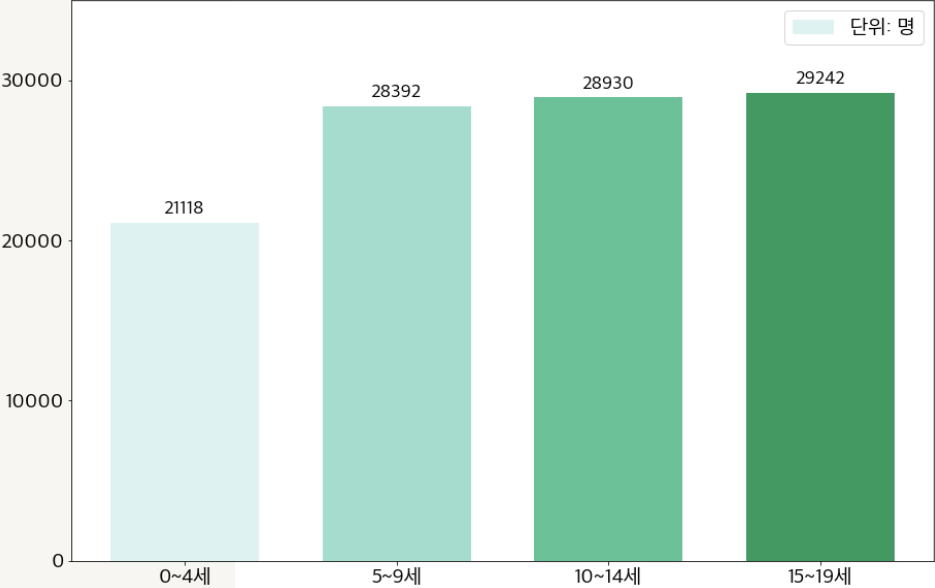
1. 주민등록인구통계 데이터를 이용해 서울시 자치구별 인구 상위/하위 10위 추출



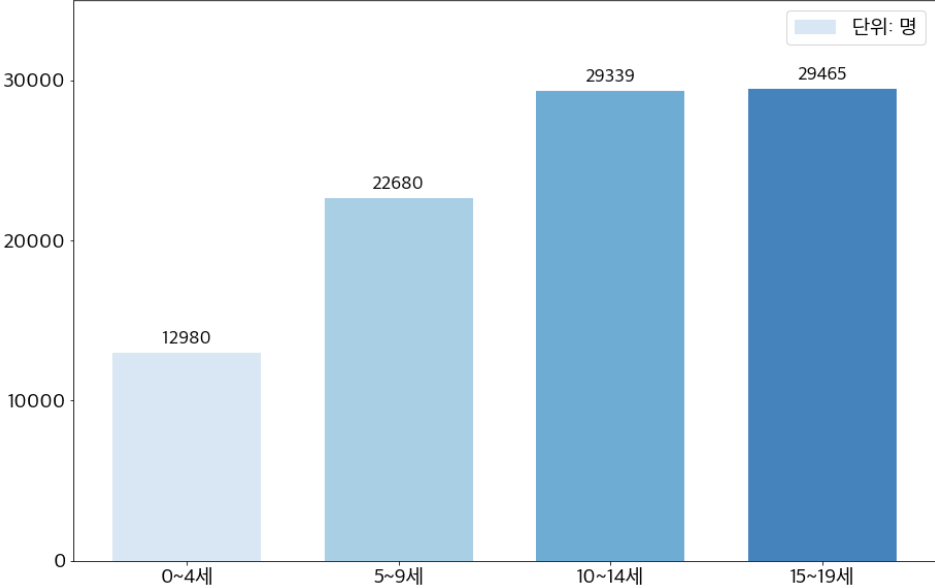
2. 연양인구 데이터 이용 인구수 상위 4위까지 지역내 교육대상 인구수 추출



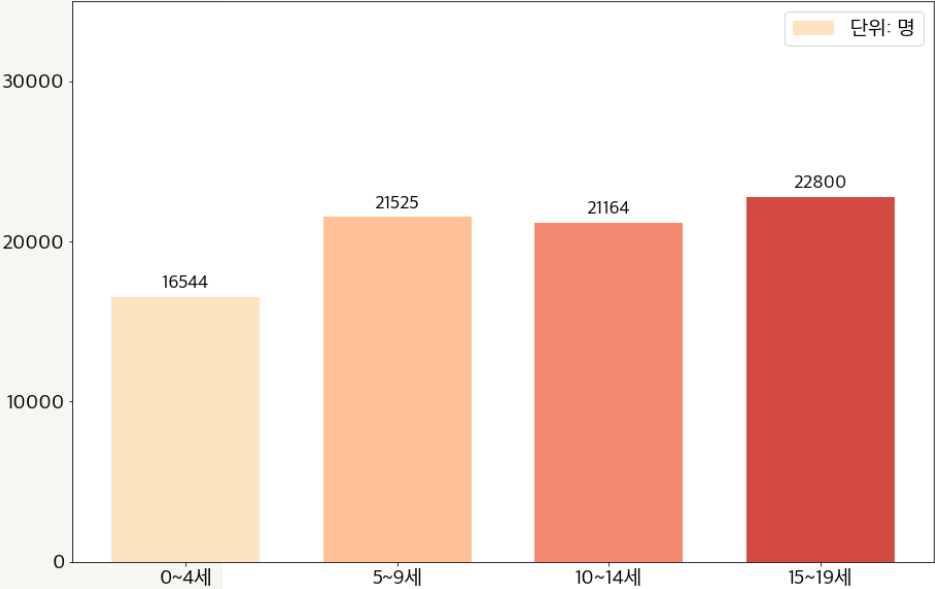
송파구 교육대상 인구 수



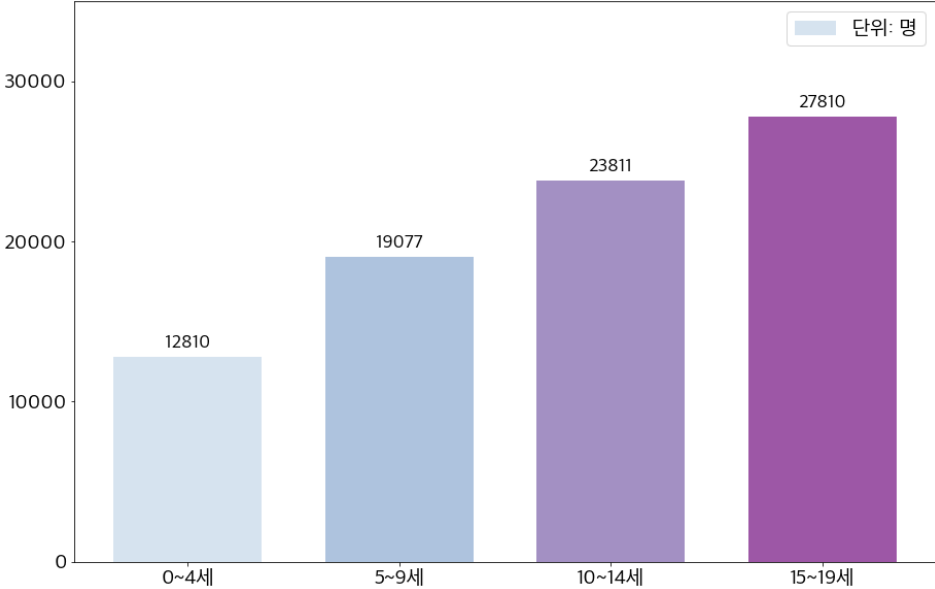
강남구 교육대상 인구 수



강서구 교육대상 인구 수



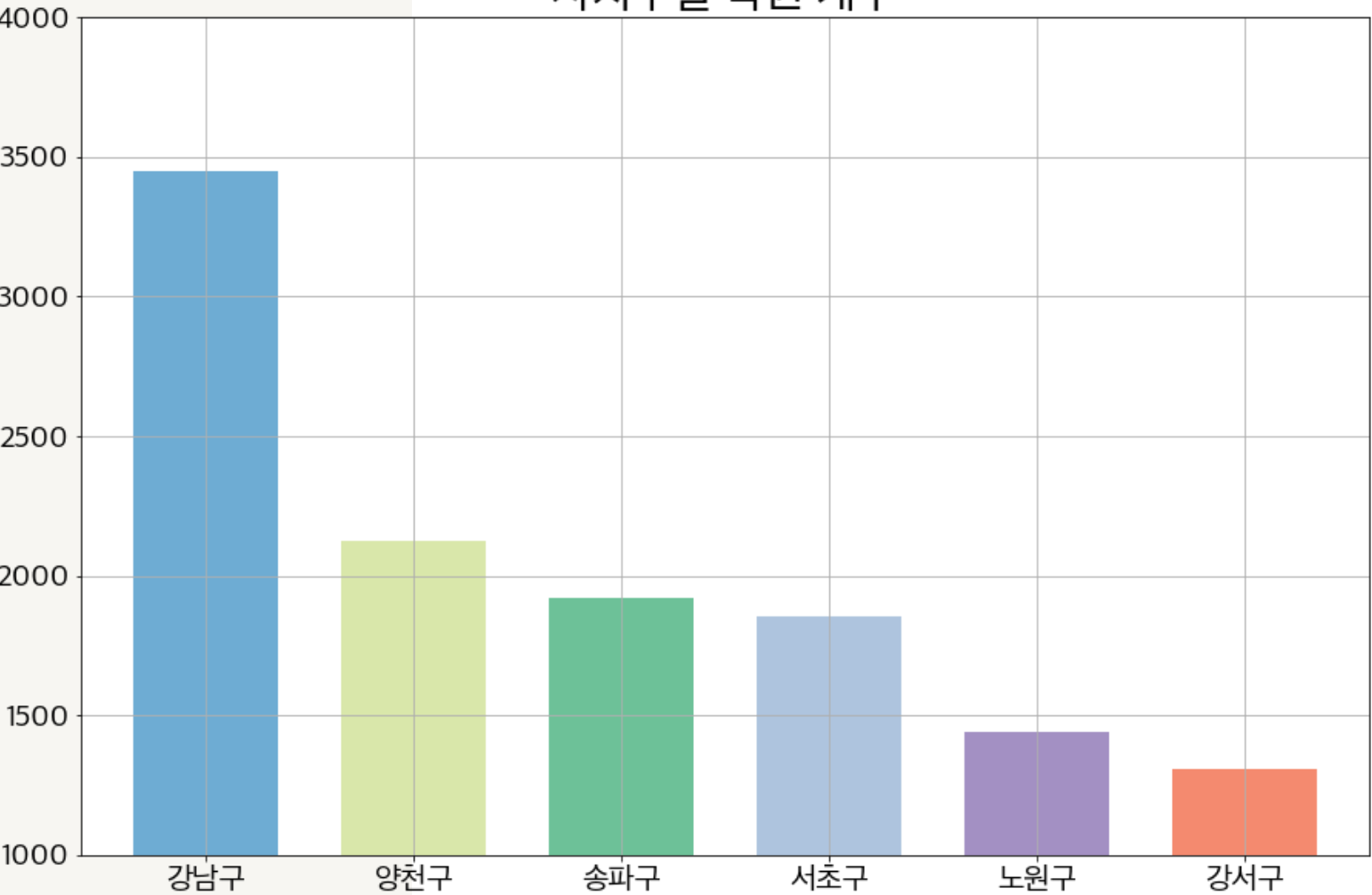
노원구 교육대상 인구 수



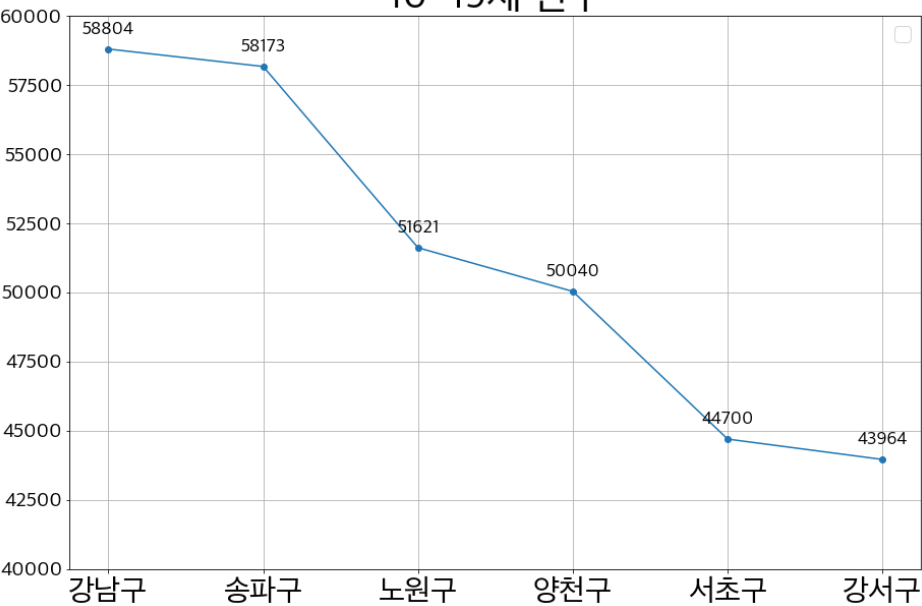
3-1 자치구별 학원 개수



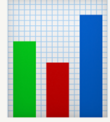
자치구별 학원 개수



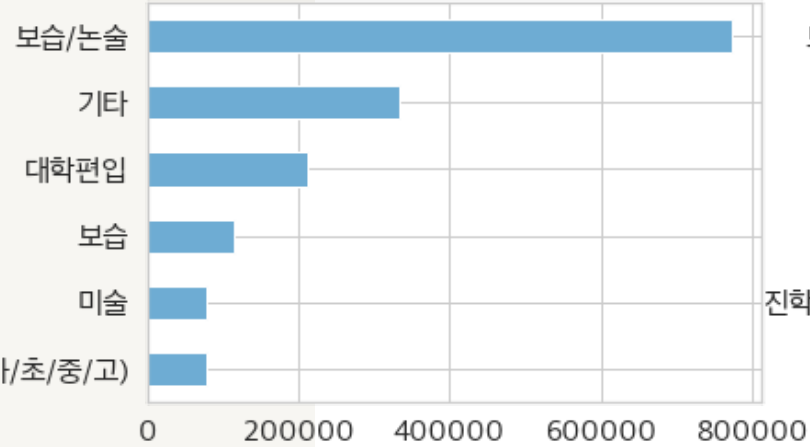
10~19세 인구



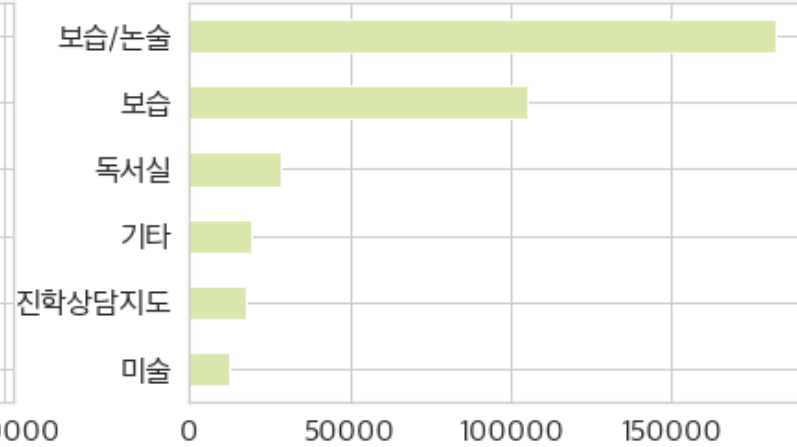
3-2 자치구별 학원 교습과정 순위



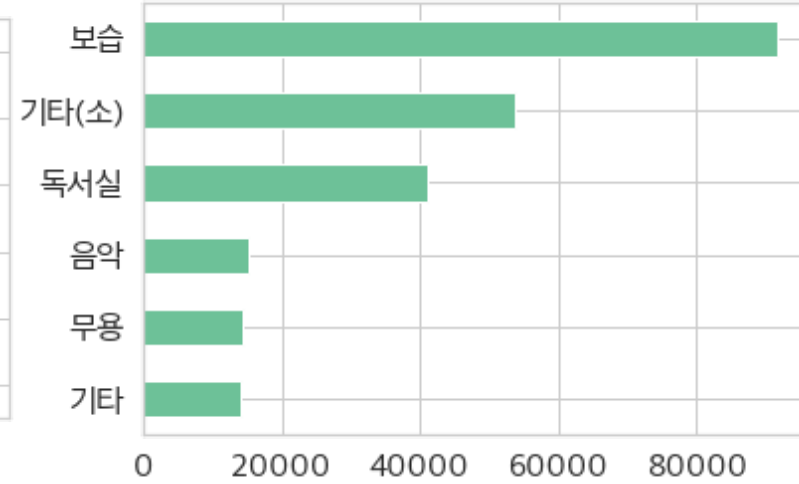
강남구 학원 교습과정 순위



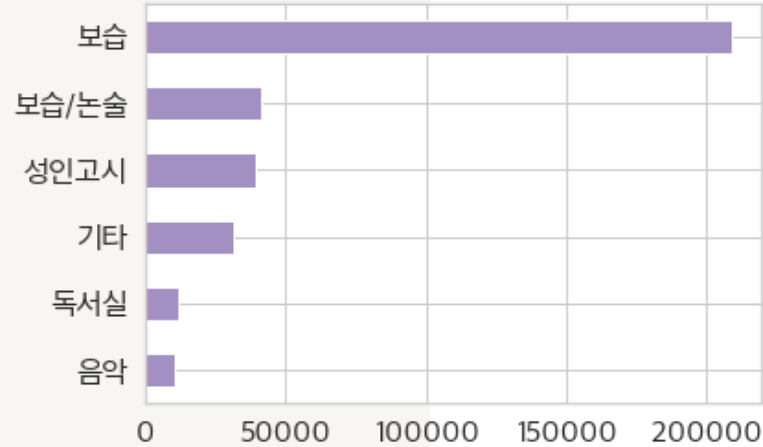
양천구 학원 교습과정 순위



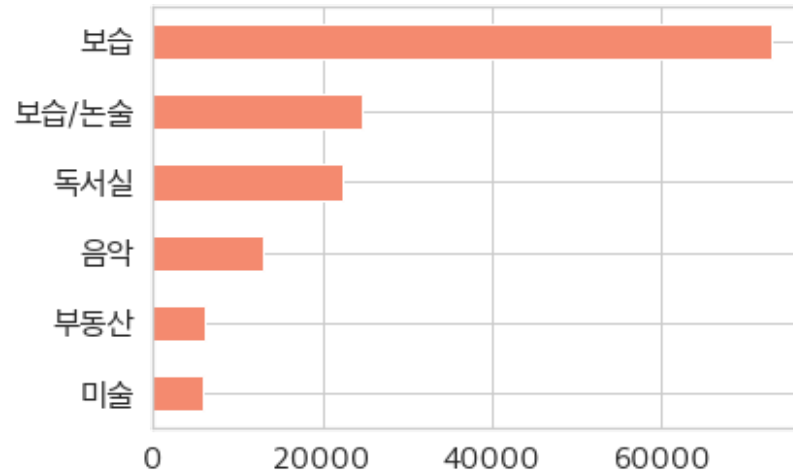
송파구 학원 교습과정 순위



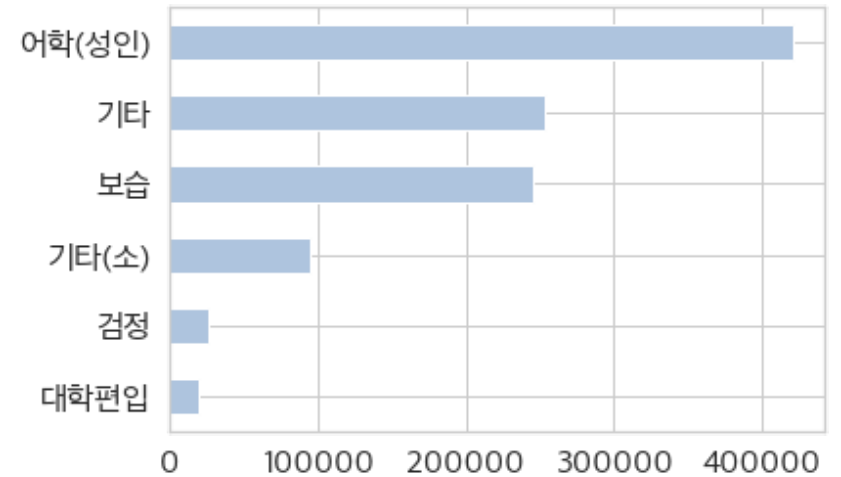
노원구 학원 교습과정 순위



강서구 학원 교습과정 순위



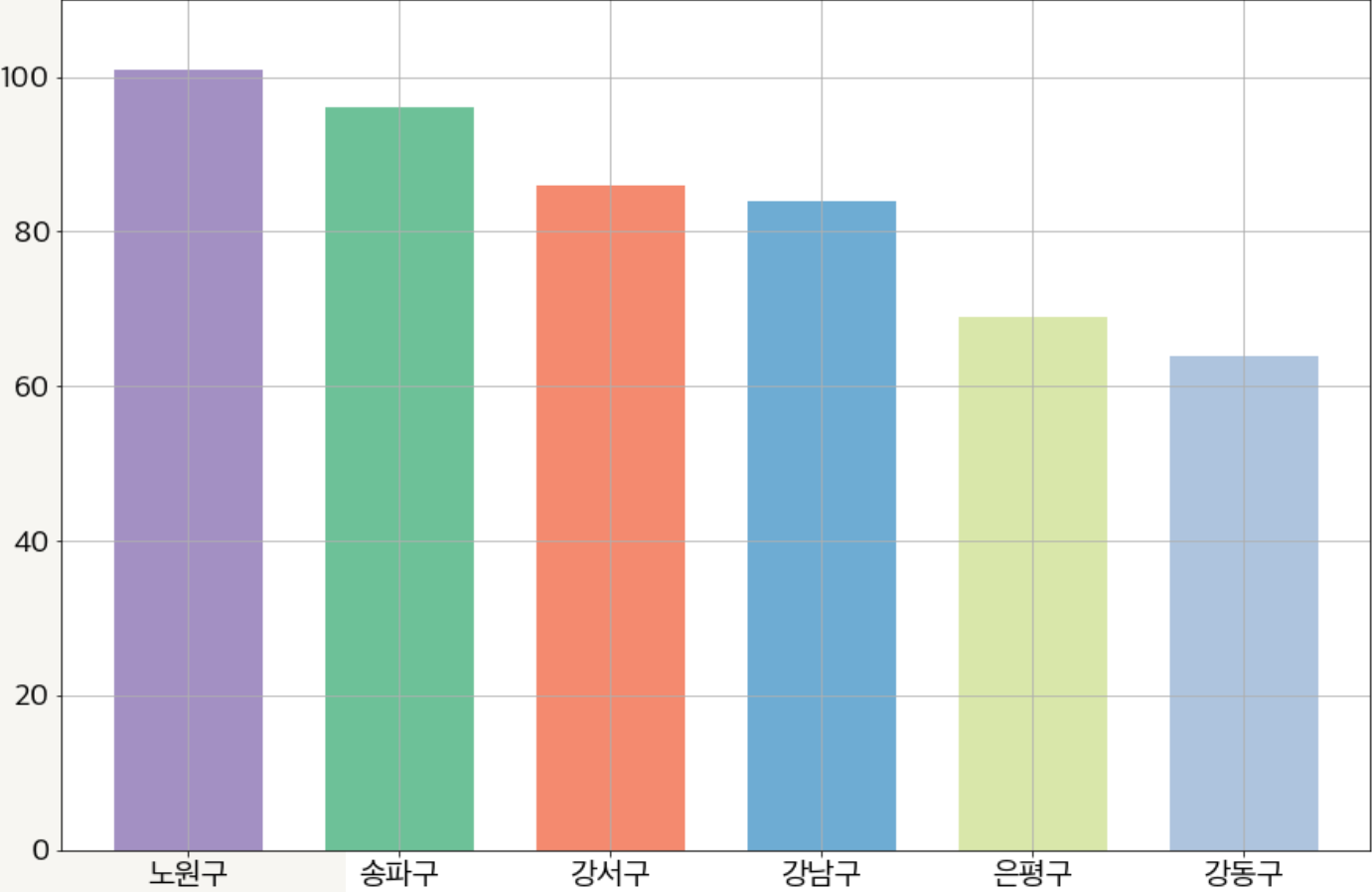
서초구 학원 교습과정 순위



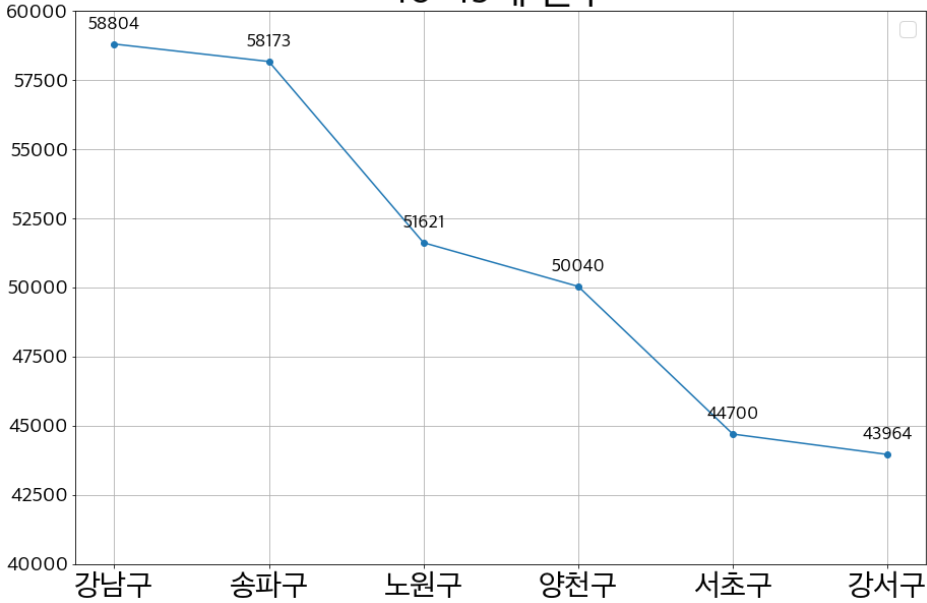
4-1 각 자치구별 학교 개수



자치구별 학교 개수



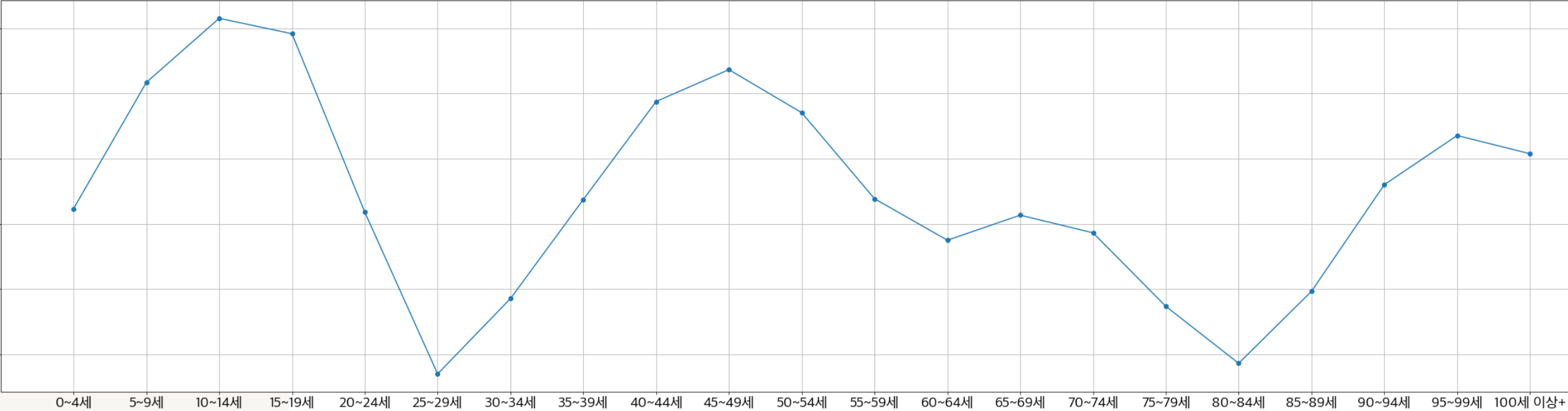
10~19세 인구



5. 학원수에 따른 인구이동 분석



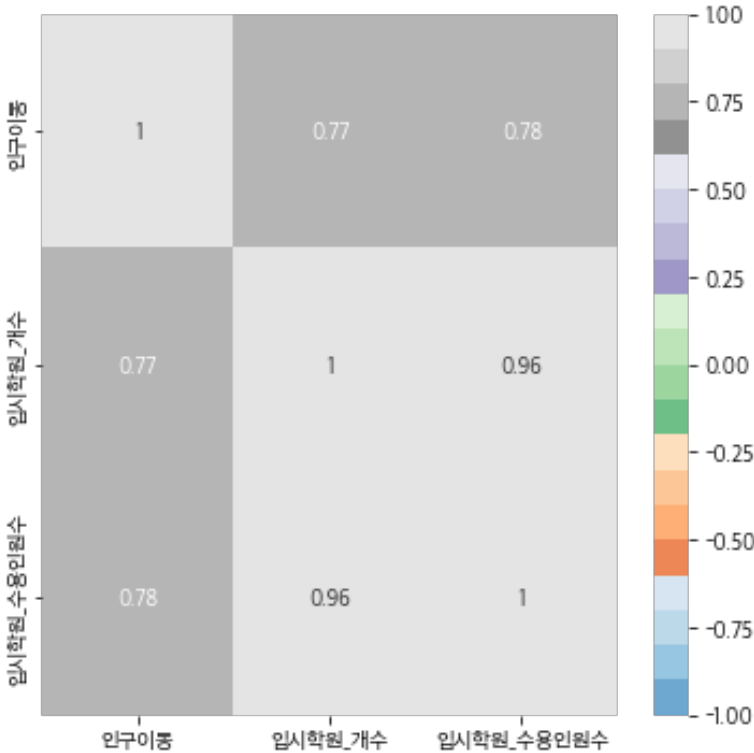
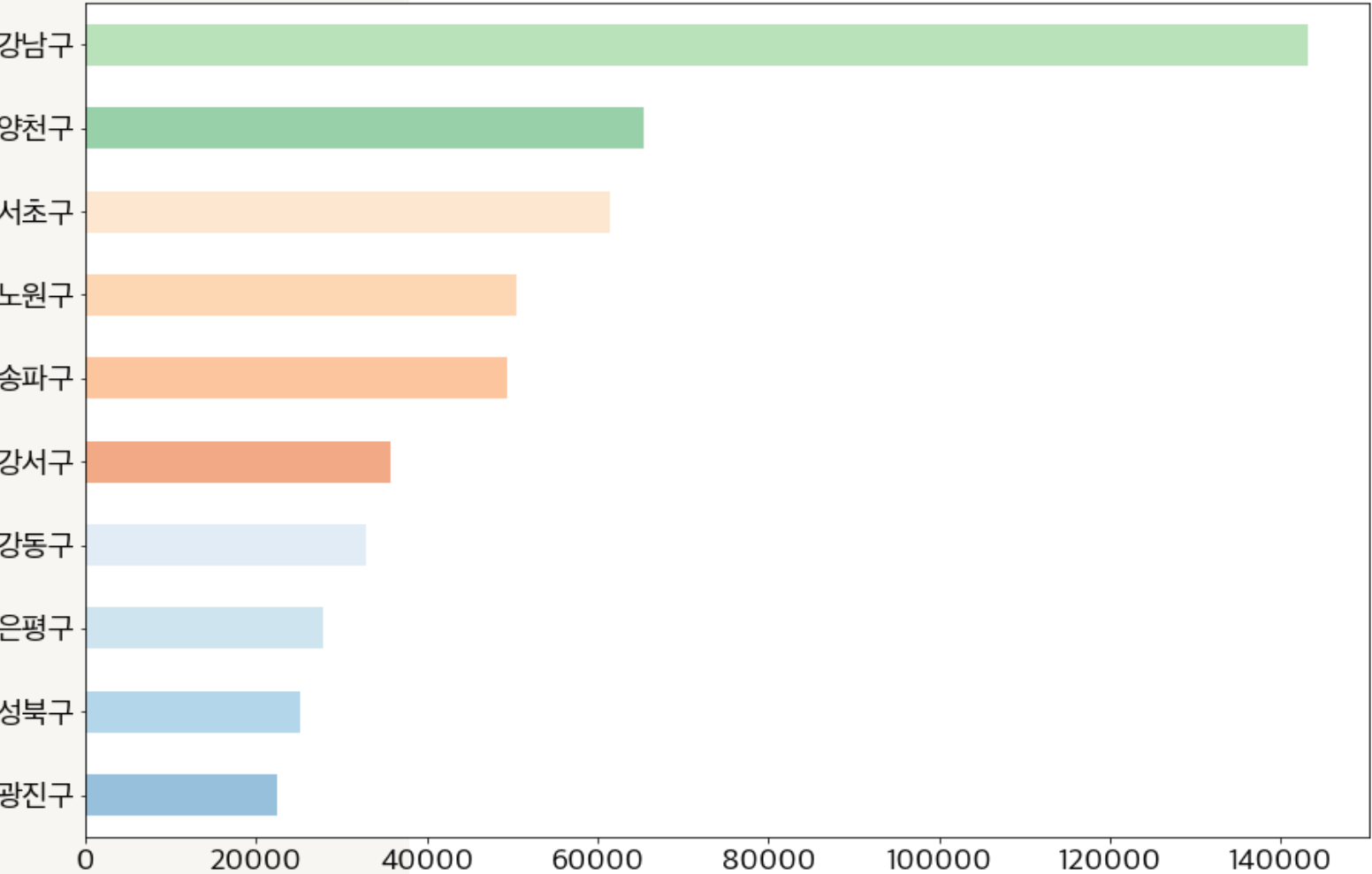
<연령대별 인구/입시학원수 & 수용인원수 상관도>



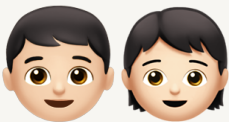
5. 학원수에 따른 인구이동 분석



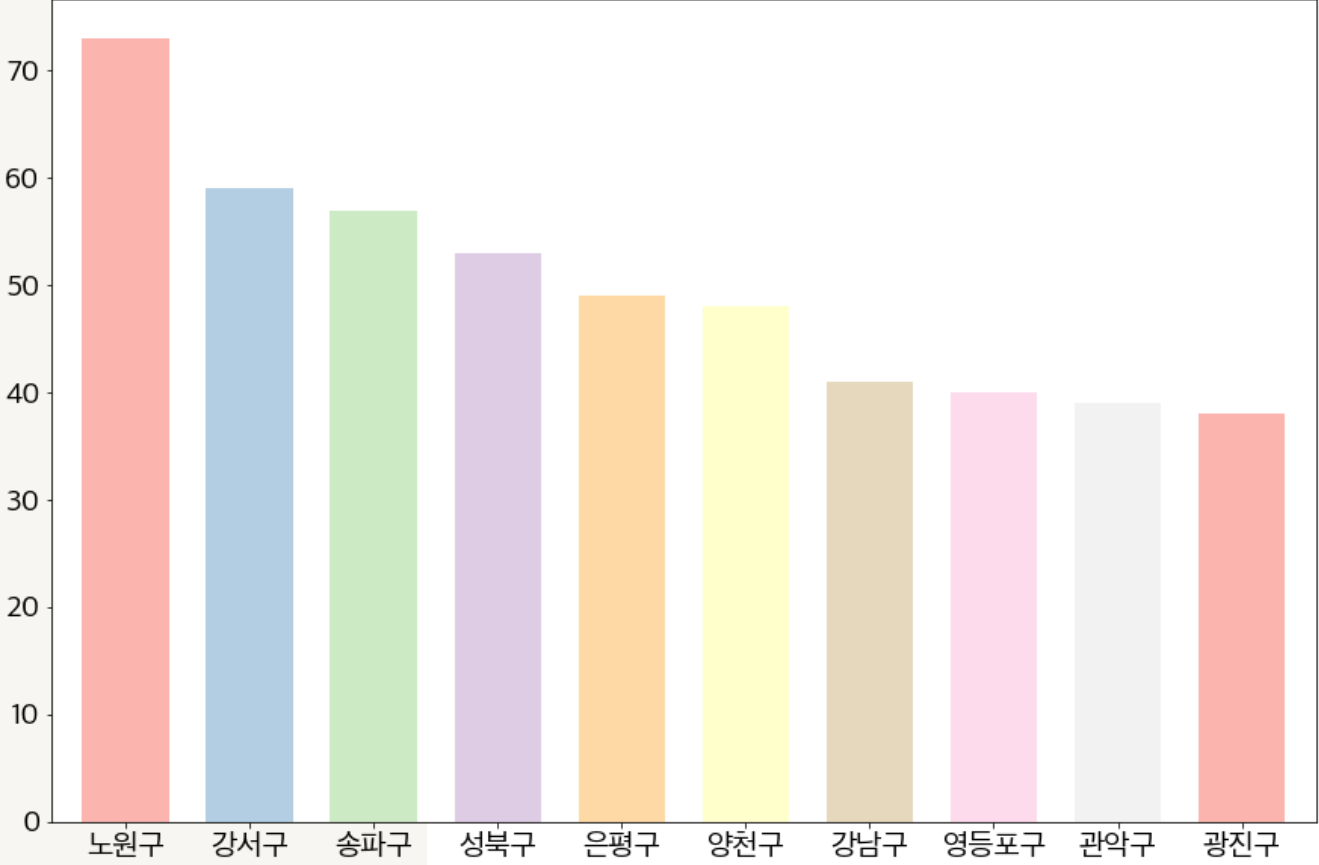
입시학원 수용인원수



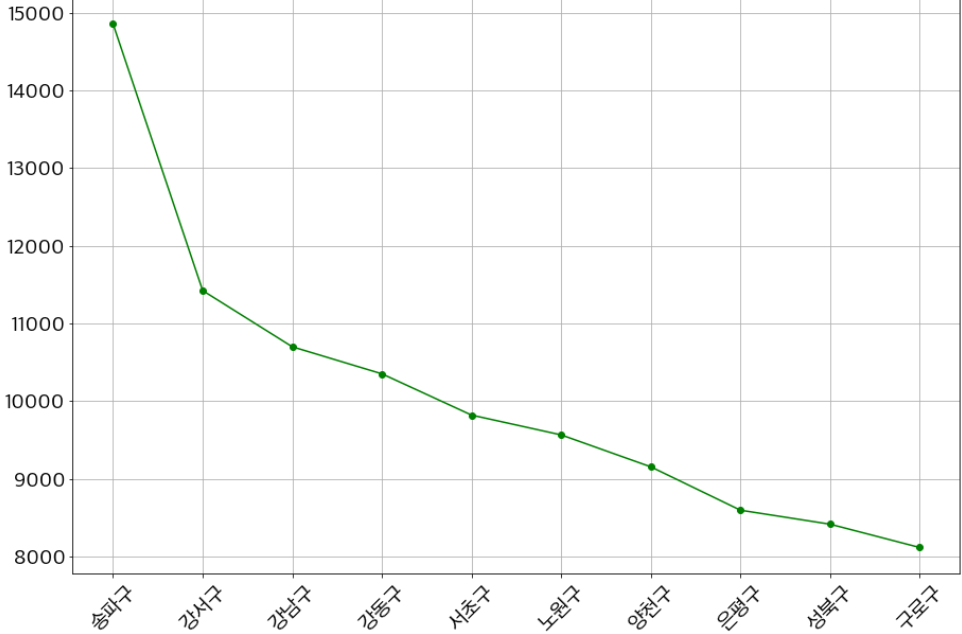
6-1 각 자치구별 유치원 수



자치구별 유치원 수



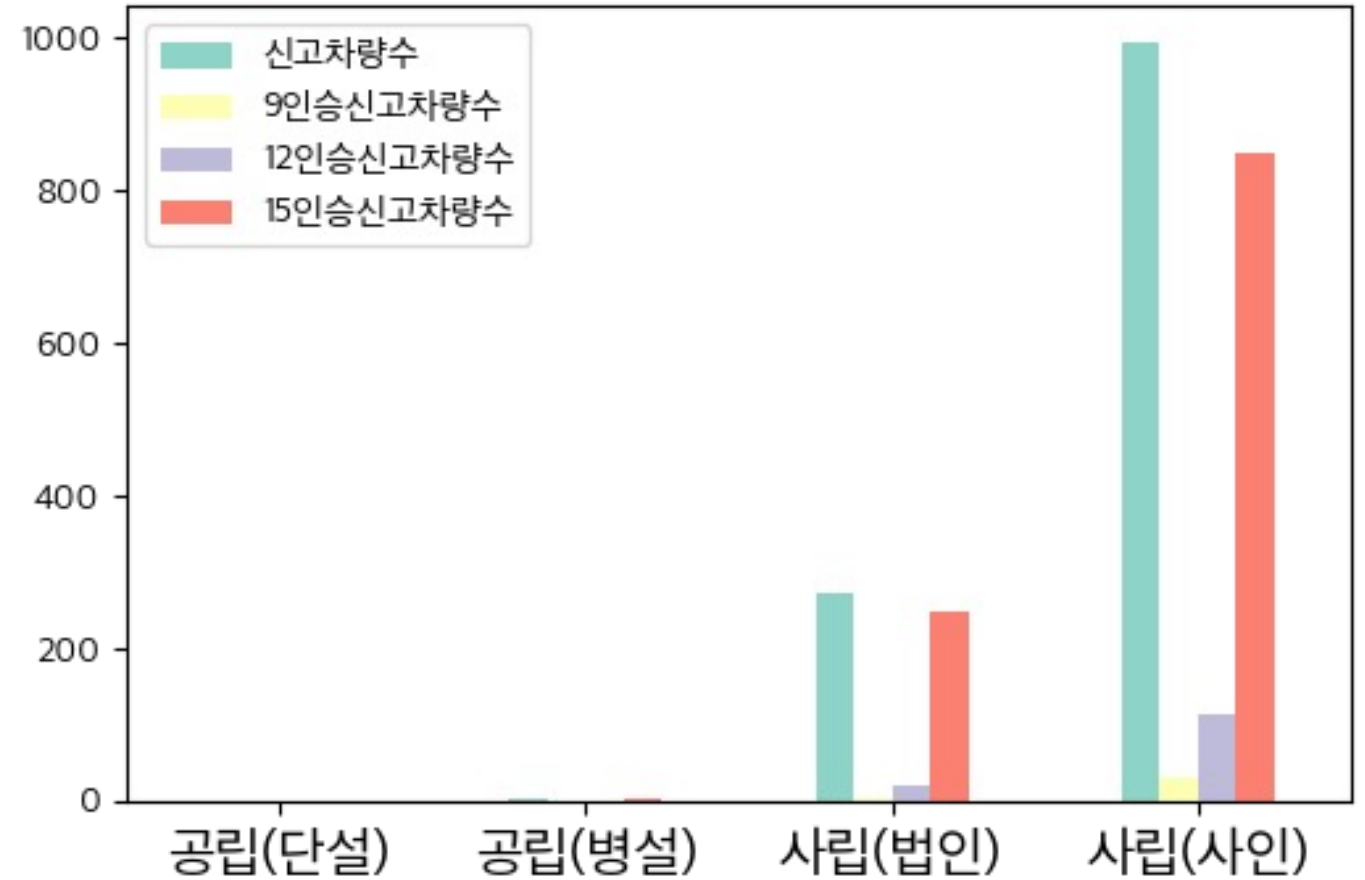
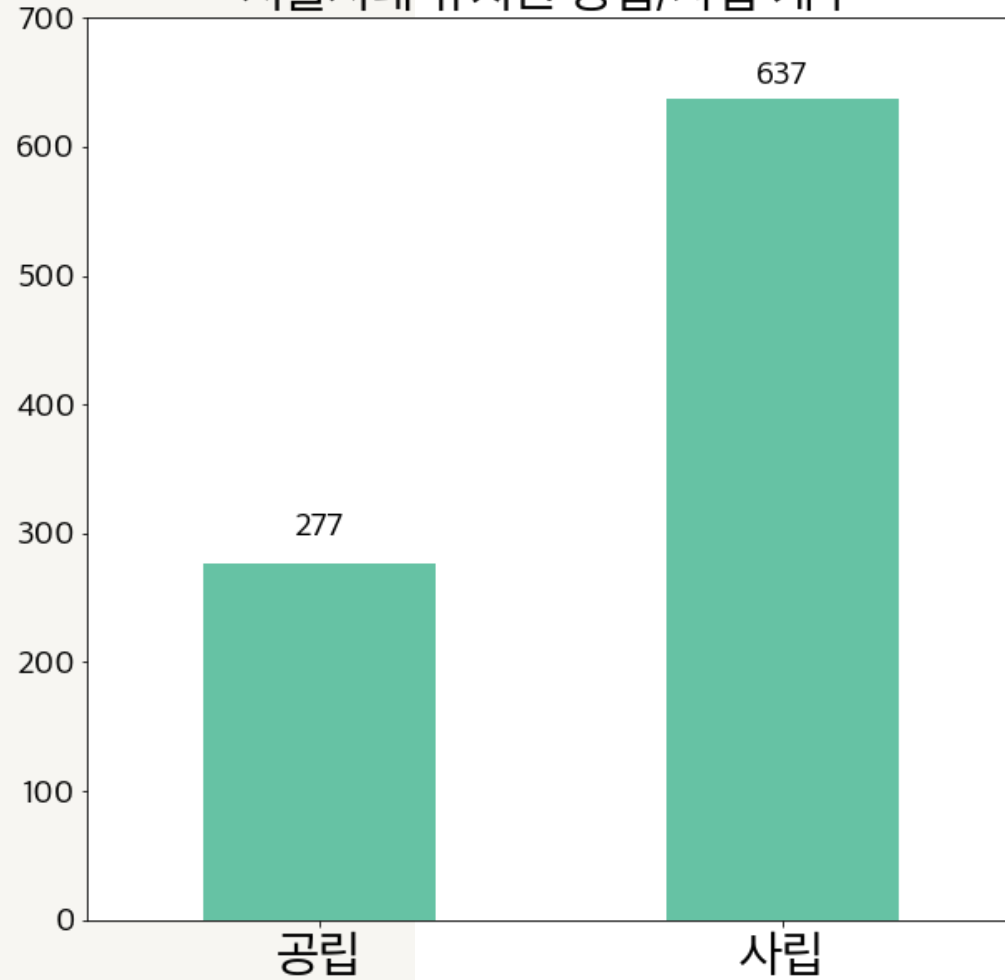
유치원대상인구



6-2 공립/사립분석



서울시내 유치원 공립/사립 개수



6 결론 및 활용방안

- 인구수와 각 단계별 교육기관 수 비례
- 학교보다 학원에 치중되어 있는 인구밀집
- 유치원 대상인구와 비례하지 않는 유치원시설
- 공립유치원의 시설/가용차량 증설 필요

감사합니다

