



생활인구 데이터를 이용한 세종대학교 학생의 통학시간 분석

전민재
물리천문학과, 응용통계학과
2023 자연과학대학 학술제

1. 분석 배경 및 서론

경기도민은 인생의 3분의 1을 지하철에서 보낸다는 말이 있다. 본 연구자도 세종대학교 까지 90분의 통학길을 매일 감수한다. 지하철 이동거리는 23.4km로 400원의 추가요금이 발생한다. 어린이 대공원 하차시 앞사람의 기록을 보면 대부분 추가요금이 100~300원에 해당 하지만 가끔 500원 이상인 사람도 존재한다. 이에 본 연구자보다 더 장거리 통학을 감수하는 사람의 분포를 알아본다. 또한 전체 학생들의 통학시간 분포를 알아보고, 건국대학교생 통학 시간분포와 비교한다. 전체 통학시간 중 90분의 누적비율을 알아보고 통학시간에 따른 학업, 삶의 만족도에 알아본다.

통학시간에 관한 연구에 따르면, 통학시간이 길수록 대학 몰입이 떨어지며 학업중단의사가 높다 [1]. 긴 통학시간으로 인해 시간 압박감을 초래해 스트레스를 느낀다 [2]. 긴 통학시간은 학생들의 시간을 제약해 인간관계 형성을 방해한다[3].

2. Data (데이터)

(1) 서울 생활이동 인구 데이터 (서울 공공데이터)

데이터 수집방법: KT 기지국 데이터 통신 기반 이동경로 기록.

(출발행정동), (도착 행정동), 시간(1시간 단위), 나이(20~25), 이동인원, 평균이동시간(분)

위 데이터는 방향성이 존재하는 데이터로 출발 행정동을 통해 위치를 파악할 수 있다. 시간은 1시간 단위로 00시~23시까지 나뉘어 있으며, 나이는 5단위로 20~25, 25~30으로 범주화 되어 있다. 이동인원은 방향과 시간, 나이가 같은 인원이 기록되어 있으며 개인정보 보호를 위해 소수의 인원은 비식별화 되어 있다. 평균이동시간은 이동인원의 이동시간의 평균을 나타낸다.

(2) 행정동경계 데이터 (국가공간정보포털)

모든 읍/면/동 경계에 대한 공간정보 및 속성정보를 가지고 있다. 각 행정동은 이름과 행정동 코드, 위도와 경도 정보를 가지고 있다. 데이터를 시각화하기 위해 쓰인다.

3. 연구 방법

(1) 세종대학교 학생 정의

위 데이터에서 (도착 행정동) = 군자동, (나이) = 20~25, (시간) = 08~17시 인 단방향 데이터를 필터링 하여 세종대 재학생으로 정의한다. 군자동의 행정구역은 그림 1과 같이 정의된다. 군자동은 학교와 주거지로 분류가 된다. 서울 상권분석서비스에 의하면 군자동의 일평균 유동인구 수는 66,098명이며 연령별 유동인구는 20대가 30%이며, 20대의 상주인구는 전체 나이대 중 5%이다. 상주인구에 비하여 유동인구가 매우 많으며 대학생의 주된 나이대인 20~25로 필터링 하여 통학생으로 가정한다. (출발 행정동)은 서울, 경기, 인천의 행정동으로 제한한다.

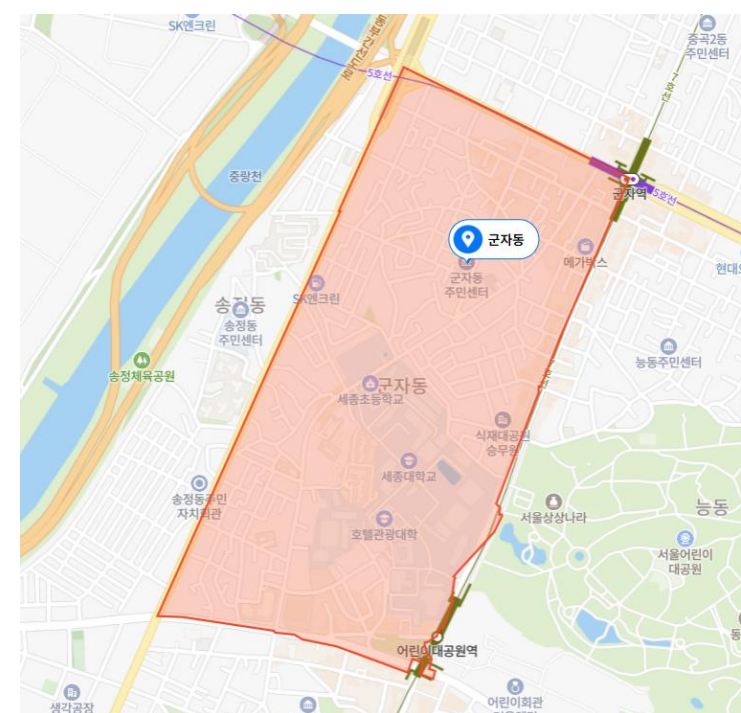


그림1 < 군자동 행정동 경계 >
출처: 네이버 지도

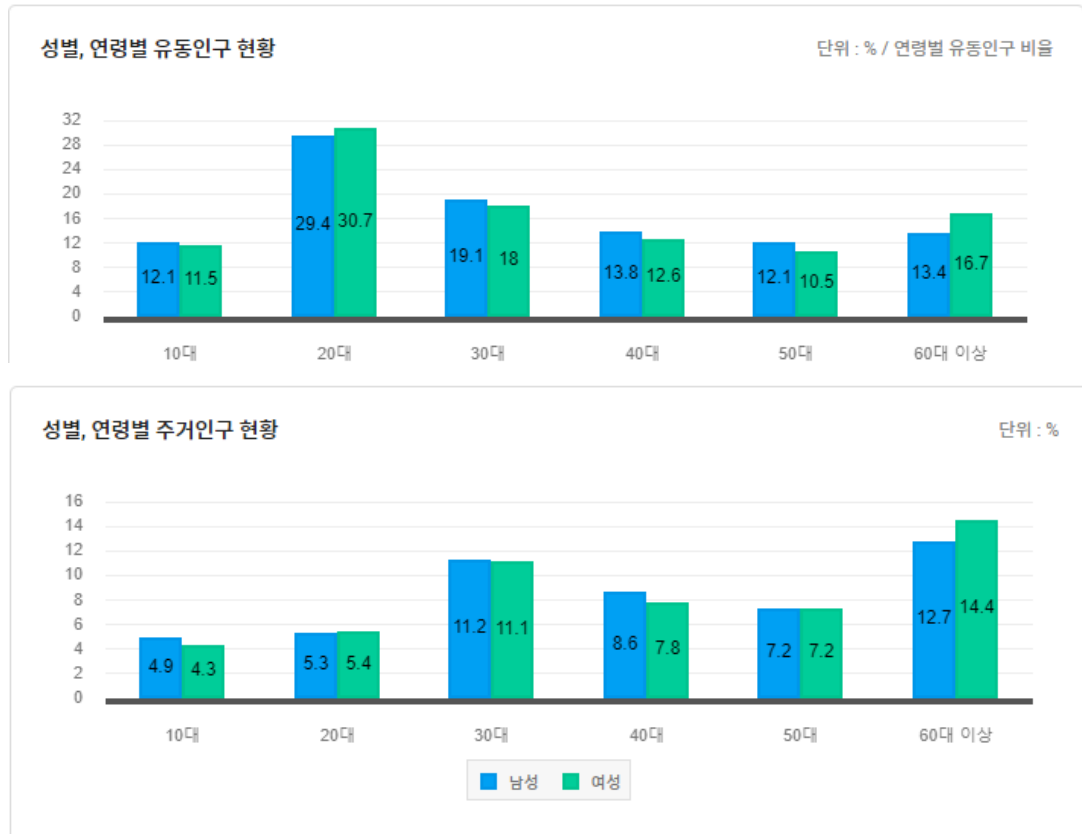


그림2 < 군자동 유동인구 및 주거인구 현황 >
출처: 서울 상권분석서비스

(2) 데이터 전처리 방법

서울 생활인구 3월 데이터에서 도착 행정구가 군자동인 20~25대, 요일은 월, 화, 수, 목, 금인 데이터를 추출한다. 이 때 군자동은 '서울특별시 광진구 군자동'과 '경기도 시흥시 군자동' 이름이 중복되기 때문에 행정동 코드 '1105054'을 이용한다. 건국대학교 학생은 위와 동일하게 도착 행정구가 화양동인 행정동 코드 '1105053'을 이용한다. 각 데이터의 평균이동시간과 인원수, 출발 행정동을 피쳐로 생성한다. Boxplot을 이용하여 이상치를 제거한다.

(3) 이동시간 및 이동거리 분포 추정

평균이동시간에 따른 인원 수를 빈도 수(Frequency)로 하여 분포 그래프를 생성한다. 생성된 분포에 대하여 로그정규 분포로 추정한다. 로그정규 분포는 확률 변수의 로그가 정규분포를 따르는 분포로 다음과 같다.

$$f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{((\ln(x)-\mu))^2}{2\sigma^2}}$$

분포그래프의 평균과 최빈값을 표시하고 90분의 누적분포값을 구한다. 행정동 사이 이동거리는 Haversine 공식을 이용하여 군자동의 중심 좌표와 각 행정동 사이의 좌표를 이용하여 거리를 구한다.

(4) 지도 시각화

Geopandas 라이브러리를 이용하여 .shp 형태의 시도 경계를 나타낸다. 법정동 코드인 'ADM_CD'를 이용하여 서울, 경기, 인천을 나눈다. 서울의 법정동 코드는 1100000000, 인천의 법정동 코드는 2300000000, 경기의 법정동 코드는 1100000000 이다. 생활인구의 행정동 코드와 지도 데이터의 법정동 코드를 기준으로 병합하여 나타낸다.

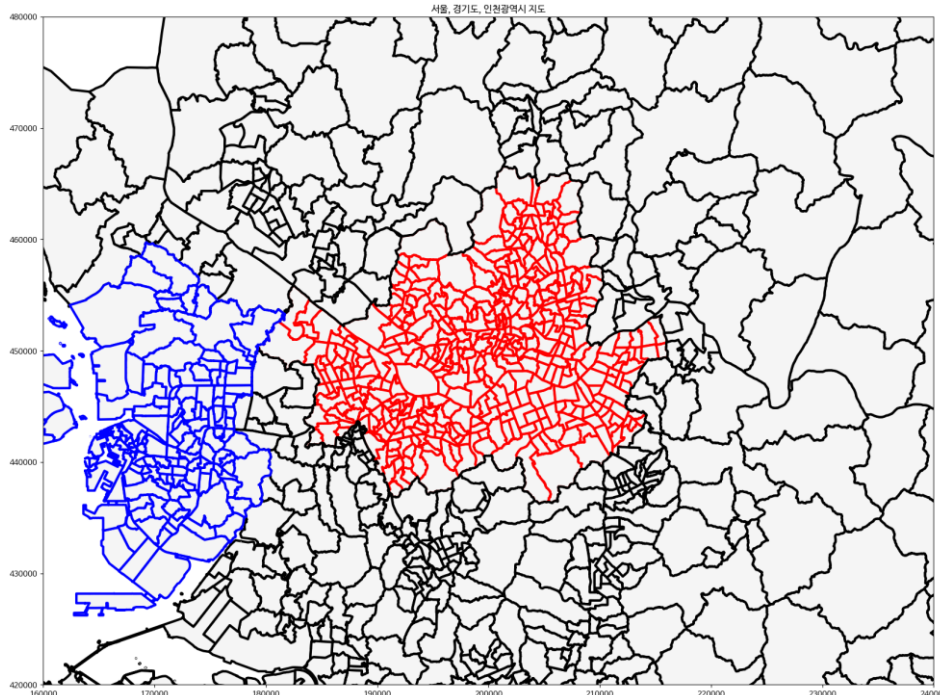
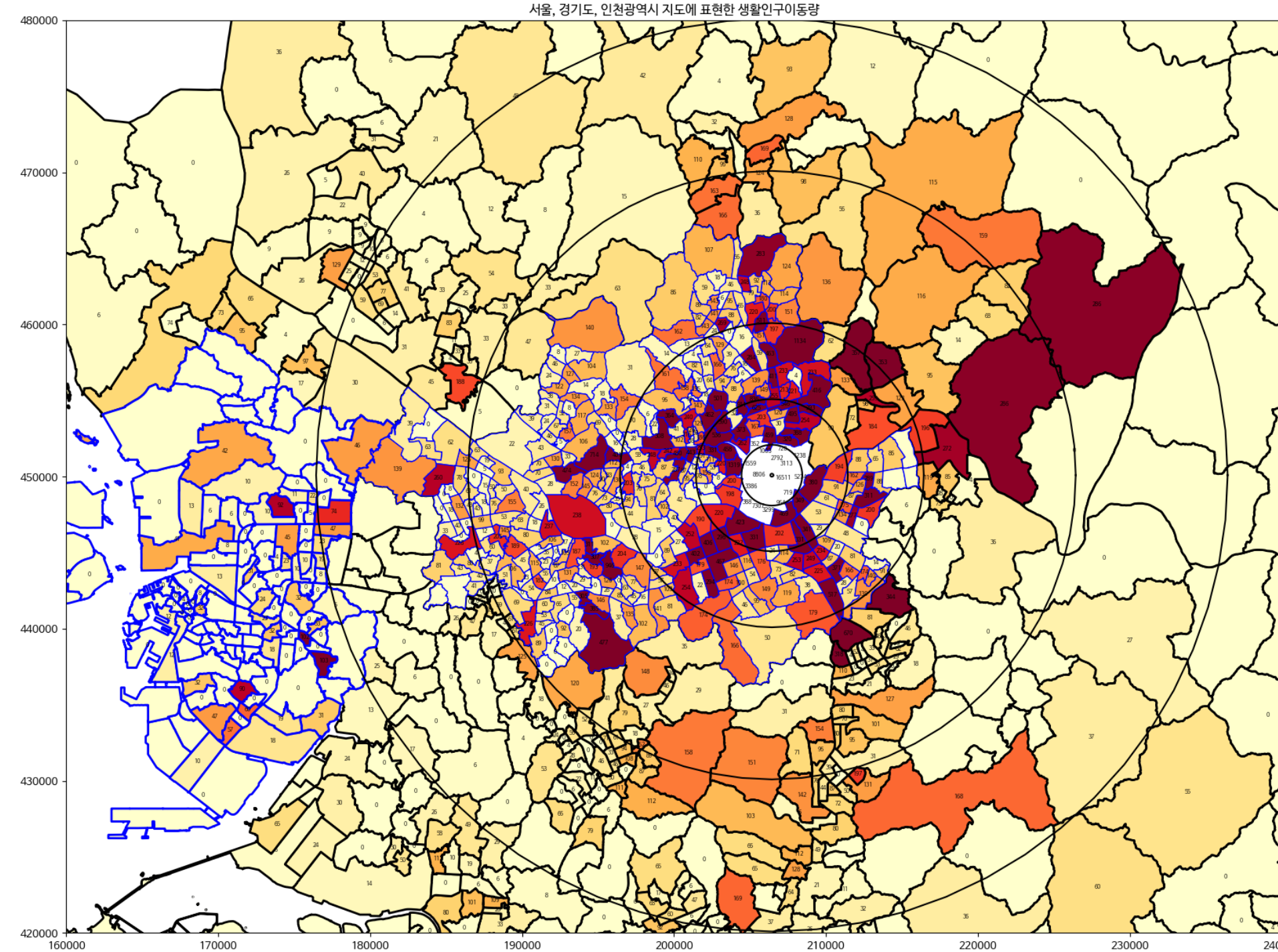


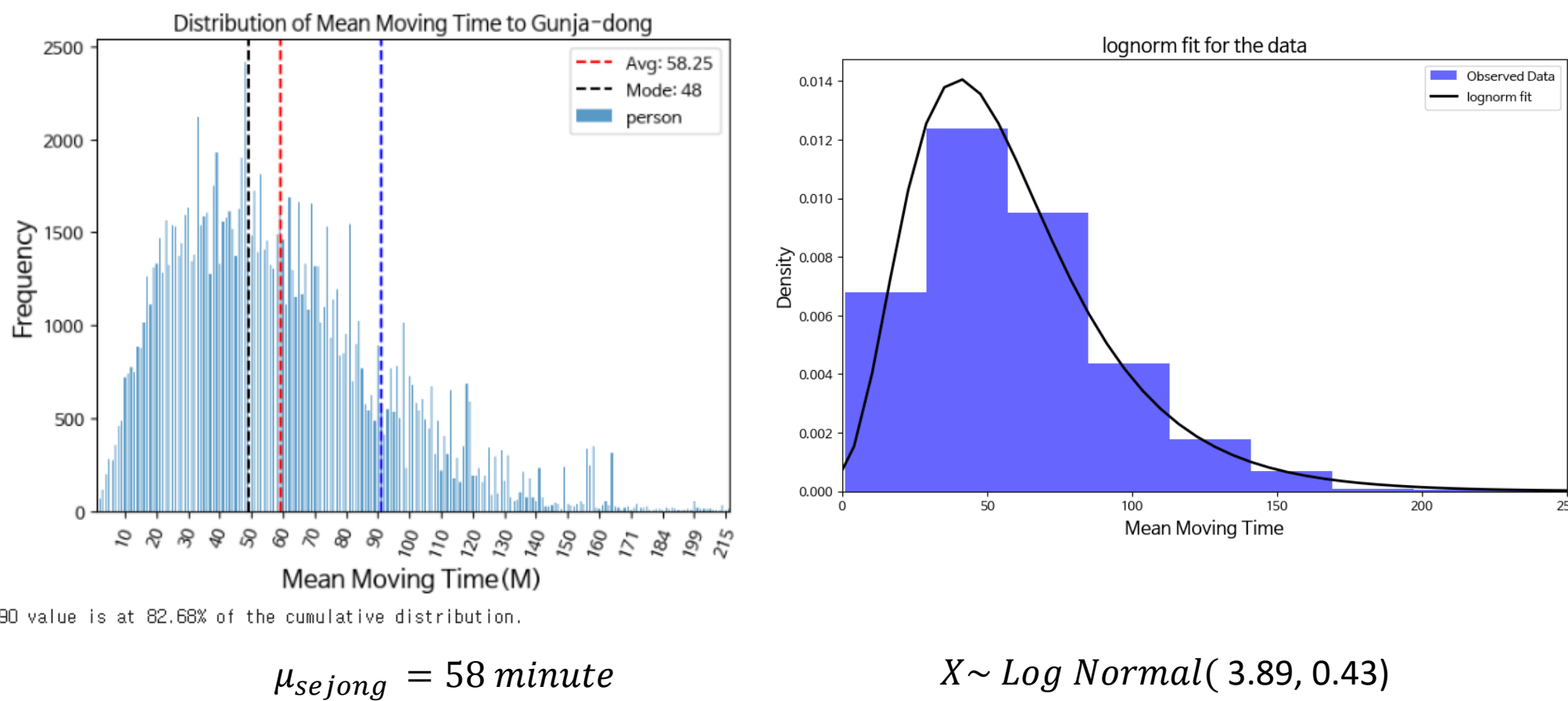
그림3 < 서울, 경기, 인천 지도 시각화 >

4. 결과

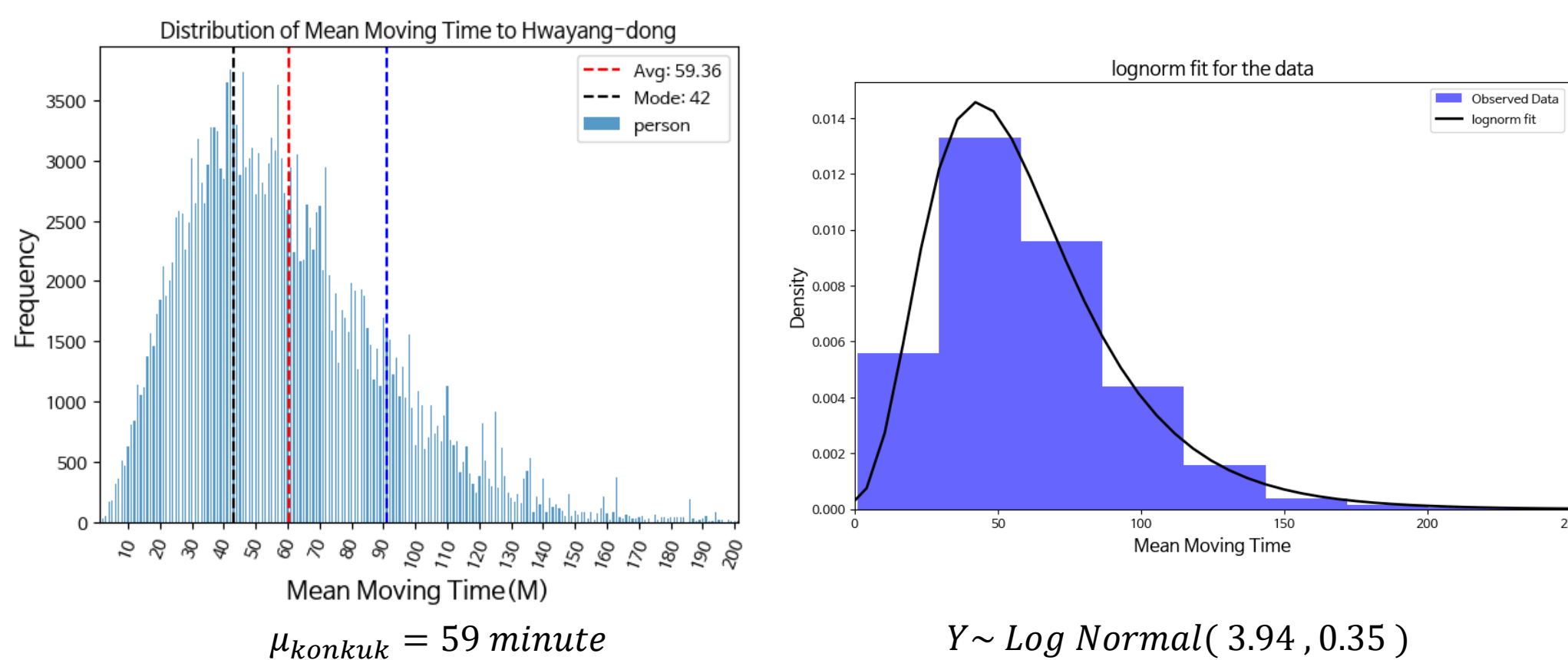
(1) 행정동별 통학 인원 수 히트맵



(2) 세종대학교 학생의 통학시간 분포 (군자동)



(3) 건국대학교 학생의 통학시간 분포 (화양동)



(4) 세종대학교 분포와 건국대학교 분포 T-test

$$H_0 : \mu_{sejong} = \mu_{konkuk} \quad H_1 : \mu_{sejong} \neq \mu_{konkuk}$$

$$T\text{test statistic} = -2.29 \quad p\text{-value} = 0.0221 \quad 95\% \text{ C.I} = (-0.381, -0.029)$$

6. Conclusion

행정동별 인구유입경로는 군자동을 중심으로 퍼져 있음을 알 수 있다. 인원이 많은 행정동은 서울의 경우 사근동, 상도1동, 공릉2동 회기동이다. 경기도는 복정동, 태평1동, 위례동, 다산1동이며, 인천은 만수6동, 효성1동, 연수1동, 동춘2동이다.

세종대 학생의 통학시간의 평균은 58분 최빈값은 48분이며, 분포는 평균 3.89, 분산 0.43, 인 로그 정규분포를 따른다. 3월 통학인원은 129,040 명이고, 방학기간인 8월의 통학인원은 8,593 명으로 현저히 떨어진다. 이는 세종대 학생 재학생 정의가 유효함을 나타낸다. 전체 분포에서 본 연구자의 통학시간 90분은 전체 분포중 83%에 해당했다. 이는 전체 학생 중 긴 통학시간을 가진 것으로 분류된다.

건국대 학생은 평균이 59분, 최빈값은 42분 이었으며, 분포는 평균 3.94, 분산 0.35인 로그정규분포를 따른다. 건국대학교의 전체 빈도 수가 228,699 명으로 세종대학교 학생 129,040 명보다 많다. 이는 건국대학교는 인구유입 원인이 많기 때문이다.

두 분포의 t-test결과 유의확률 0.02에서 귀무가설을 기각할 수 있다. 따라서 두 분포의 평균은 다르며 다른 모집단임을 확인할 수 있다.

본 연구는 생활인구이동량을 통해 세종대학교 학생의 통학시간의 분포를 나타내었다. 세종대 학생의 경우 평균 통학시간이 1시간으로 매우 긴 것을 알 수 있다. 긴 통학시간은 대학 몰입도를 방해해 각종 활동 및 학업에 영향을 준다. [1] 따라서 학생의 학교생활 복지를 위해 긴 통학시간에 대한 대책이 필요함을 알 수 있다. 행정동 별로 유입인원이 많은 곳을 묶어 셔틀버스 운행 등 학생복지를 위한 대책 등이 필요하다.

[1] 신혜숙, 김미현. (2018). 대학생의 통학시간이 대학 몰입을 매개로 학업중단의사에 미치는 영향. 한국교육문제연구, 36(1), 25-44.
[2] Ackerman,D.S., & Gross,B.L. (2003). Some choices, solittime: measuring the effect of free choice and enjoyment on perception of free time. e, time pressure and time deprivation. Advances in Consumer Research, 30, 290-294.
[3] 김선영.(2015).통학대학생들의대학사회내통합증진,생활만족도증가,학업성취도향상을위한기초연구:OO사립대학장시간통학생들을중심으로.청소년복지연구,17(3),187-214.