

접수번호	
------	--

## 「통계빅데이터 자료분석·활용대회」 분석 결과보고서

신청자명	최종욱(대표), 김동현, 박소담
분석주제명	신도시 타당성 요인분석

### 1. 주제 선정 및 자료분석 배경

본 분석은 기존 1, 2기 신도시를 분석하여 최근 발표한 3기 신도시 개발의 방향성에 대해 제언하고자 한다. 신도시의 성공기준은 기존 계획인구와 현 거주인구의 일치 또는 상회 여부로 생각한다. 성공에 영향을 미치는 요인은 크게 3가지로 ①교통, ②상권, ③자급능력을 채택했다.

우리나라의 수도권 신도시의 대부분은 주택시장 안정과 주택문제를 해결하기 위한 목적으로 건설되어왔다. 신도시의 본래 취지는 자족 기능을 갖춘 도시조성이었으나, 기존 신도시는 서울의 주택가격 상승으로 인한 수요분산을 통해 서울 주택가격 하락을 주목적으로 하고 있어 자족 기능을 갖춘 도시보다 베드타운이 된 도시가 상당히 많은 편이다.

현재 1기 신도시 분당, 일산, 부천(중동), 평촌, 산본 5개의 도시가 구축 완료, 2기 신도시는 판교, 동탄1, 동탄2, 한강, 운정, 광교, 옥정, 위례, 고덕, 검단 10개의 도시가 구축 완료 혹은 개발 중이다. 여기에 더해 문재인 정부는 계속되는 서울지역 부동산 시장의 과열화를 막기 위해 주택 30만 호 공급계획을 발표했다. 대규모 택지개발 대상 지역과 함께 ‘3기 신도시’를 발표했는데, 2018년 12월 19일 남양주 왕숙지구, 하남 교산지구, 인천 계양 계양테크노밸리, 2019년 5월 7일 고양 창릉지구, 부천 대장지구가 그 대상지이며 총 5개의 신도시를 추가 발표했다.

3기 신도시는 서울시계에서 이격된 지역에 조성한다는 2기 신도시의 원칙을 폐지하고 1기 신도시보다도 서울과 가까운 지역에 조성이 예정됐다. 2기 신도시의 경우 판교나 동탄 정도를 제외하면 인근 대도시 상권에 의존하는 경향을 보였으며 도시계획 대부분이 도시 인프라 구축보단 아파트 건축에 치중하고 있어 결과적으로 부동산 시장의 서울 선호 현상을 가중시켰다.

본 연구는 3기 신도시의 목표인 ‘수도권 인구 흡수를 통한 주택시장 안정화’를 고려해 타당성 기준을 ‘신도시 지역의 인구’로 채택했으나 신도시의 자족 능력이 밑바탕 되지 않는다면 경제활동인구의 일자리를 찾아 도심으로 회귀하는 현상이 지속적으로 발생할 것이다. 이로 인해 수도권 밀집 현상이 발생할 것이며, 기존 신도시의 노후화가 예측되는 만큼 추가적인 연구와 정책 보완이 필요해 보인다.

## 2. 분석 내용

### (1) 활용 데이터

#### <통계빅데이터센터>

- 통계기업등록부DB(사업자 기준)  
2015년 1월 ~ 2016년 12월  
: 신도시별 상용근로자 수 전처리
- 성·연령별 유동인구(전국) 2015년 1월 ~ 2016년 12월  
: 신도시별 성·연령별 유동인구 전처리
- 업종별 매출액(현대카드)(전국) 2015년 1월 ~ 2016년 12월  
: 신도시의 상권분석을 위한 신도시 업종별 매출액 전처리
- 1000대 상권 정보(전국)  
: 신도시 소재 지역 중 1000대 상권 확인

#### <공공·민간데이터>

- 다음 웹사이트 크롤링  
: 신도시를 경유하는 버스 노선 및 정류장 정보 크롤링
- (경기버스정보) 버스노선정보, 노선경유정보, 정류소정보<sup>1)</sup>  
: 경기도 지역 버스노선정보, 노선경유정보, 정류소정보 데이터를 통해 신도시의 버스 정보 전처리
- (서울 열린데이터 광장)<sup>2)</sup>
  - ① 서울특별시 버스정류장 위치정보
  - ② 서울시 역코드로 지하철역 위치 조회  
: 서울시 버스정류장, 지하철역 위치

1) <http://www.gbis.go.kr/>

2) <https://data.seoul.go.kr/dataList/datasetView.do?infId=OA-15067&srvType=S&serviceKind=1&currentPageNo=1>  
<https://data.seoul.go.kr/dataList/datasetView.do?infId=OA-118&srvType=S&serviceKind=1&currentPageNo=1>

### 좌표 확인

- (코레일) 2018년 수송실적<sup>3)</sup>  
: 신도시 지하철 이용객 및 승하차 인원 확인
- (코레일) 2017 철도통계 연보<sup>4)</sup>  
: 신분당선 이용객 및 승하차 인원 확인
- (KB 부동산) 월간 KB주택가격동향<sup>5)</sup>  
: 서울의 주택가격 동향을 확인하고 시계열 작성
- (공공데이터포털) 버스노선 조회서비스-REST<sup>6)</sup>  
: 버스 노선조회
- (구글 클라우드 플랫폼) 지도 API  
: 지도 API 사용

3)

<http://info.korail.com/mbs/www/jsp/board/view.jsp?boardId=10224927&boardSeq=14674693>

4)

<http://www.kric.go.kr/jsp/handbook/sta/statisticsTchList.jsp>

5)

<https://onland.kbstar.com/quics?page=C059744>

6)

<https://www.data.go.kr/dataset/15000430/openapi.do?mypageFlag=Y>

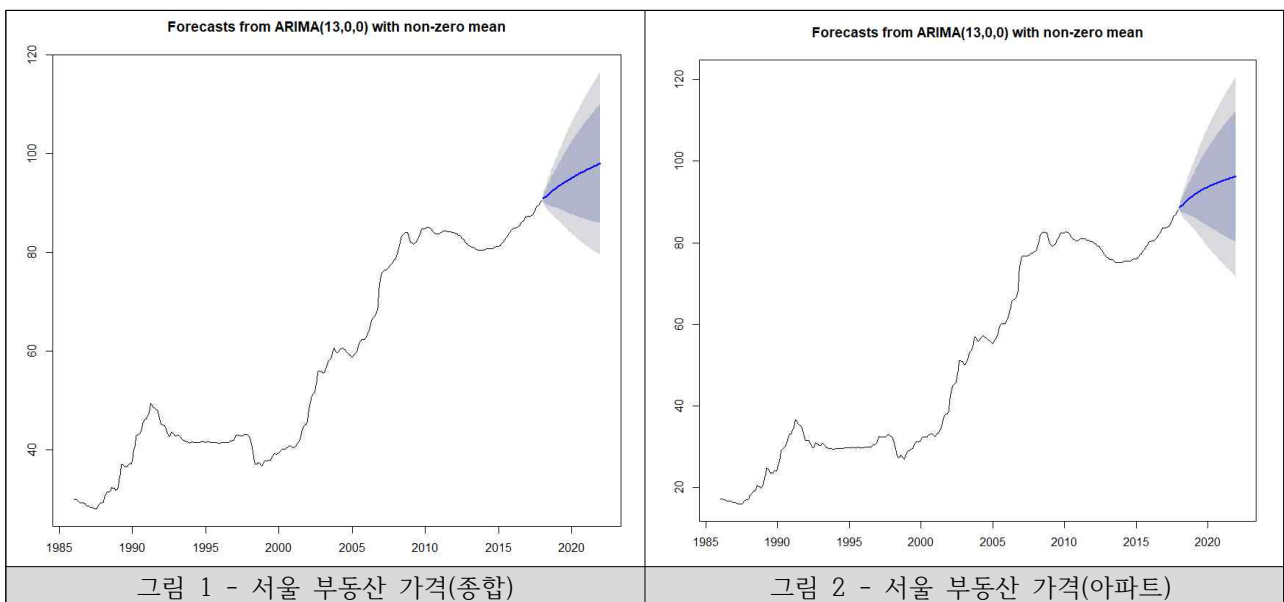
## (2) 데이터 분석 결과

### 1) 변수설명

변수명	변수 의미
city	각 도시의 신도시 구분
city_n	신도시 이름
success	인구수 기준 신도시 성공 여부
pop_n	총인구수
pop_p	계획 인구수
pop_pr	$\text{pop\_n} / \text{pop\_p}$
bus	버스 개수
seoulbus	신도시 소재 버스가 경유하는 서울정류장 개수
subway	신도시 소재 지하철역 개수
pop3040	30, 40대의 인구비율
pop_float	일평균 유동인구 수
pop_rework	상용근로자 수
work_qual	상용근로자 / 전체근로자
school	학교 개수
subOD_day	일평균 지하철 승하차 인원
subOD_year	연평균 지하철 승하차 인원
sale	신도시 소재 1000대 상권 포함 상권 개수
edu	교육매출
medic	의료매출

## 2) 변수 분석 및 시각화

### • 1. 서울 부동산 가격 시각화



: 1기 신도시는 88올림픽으로 인해 과열된 1990년대~2000년대 초반의 서울 집중화 현상을 완화하는데 결정적인 역할을 했다. 하지만 2기 신도시는 부동산 시장의 과열화와 서울 인구집중 현상을 완화시키지 못했음을 서울 주택가격 데이터의 시계열 예측을 통해 확인할 수 있었다.

• 2. 버스 노선, 서울정류장 시각화

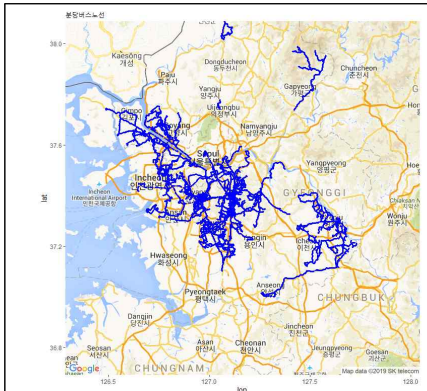


그림 3 - 분당신도시 버스 노선

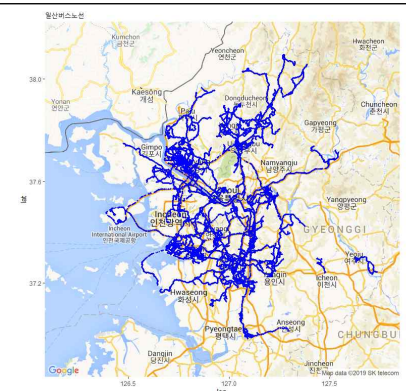


그림 4 - 일산신도시 버스 노선

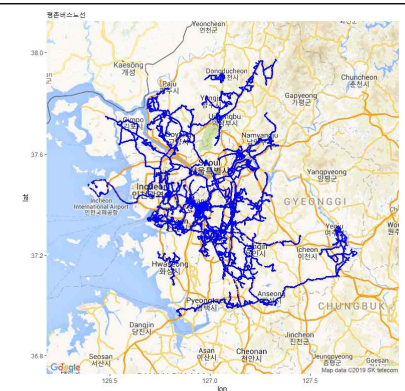


그림 5 - 평촌신도시 버스 노선

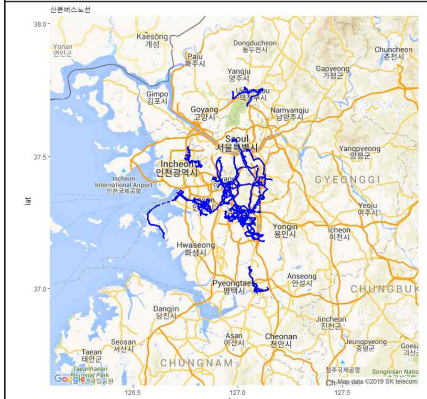


그림 6 - 산본신도시 버스 노선

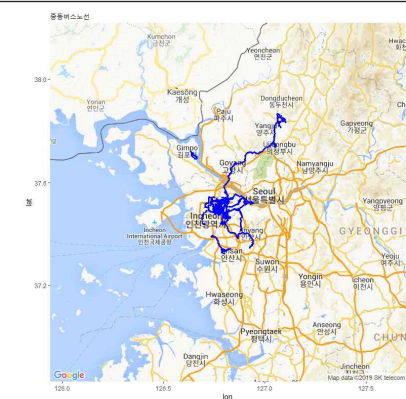


그림 7 - 중동신도시 버스 노선

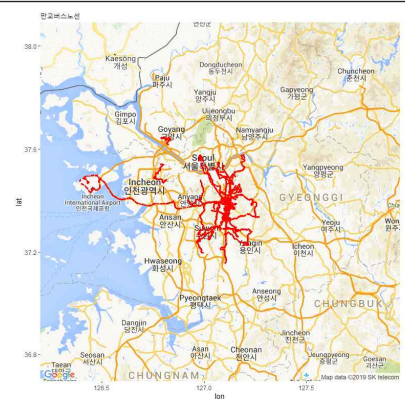


그림 8 - 판교신도시 버스 노선

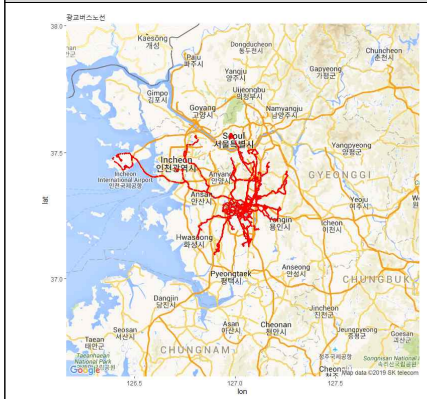


그림 9 - 광교신도시 버스 노선

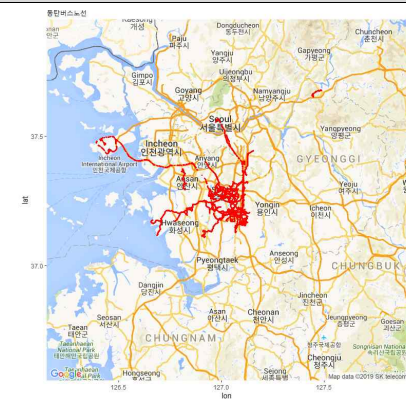


그림 10 - 동탄신도시 버스 노선

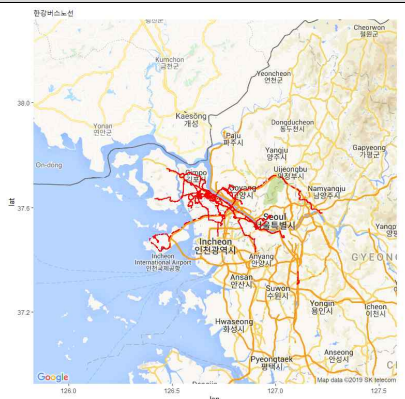


그림 11 - 한강신도시 버스 노선

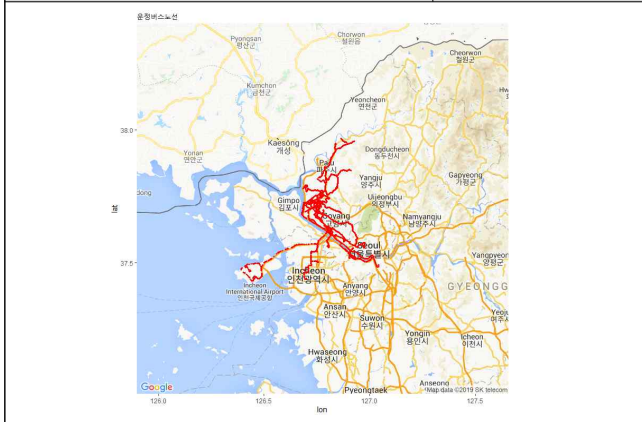


그림 12 - 운정신도시 버스 노선

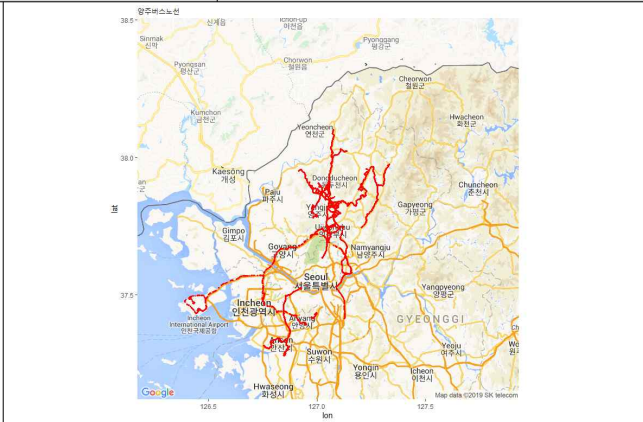


그림 13 - 양주신도시 버스 노선



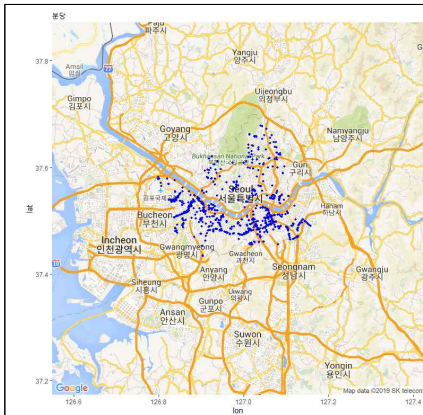


그림 14 - 분당신도시 서울정류장

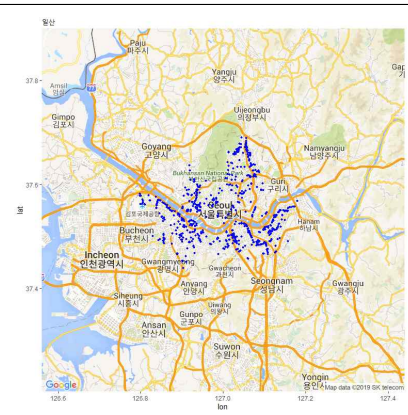


그림 15 - 일산신도시 서울정류장

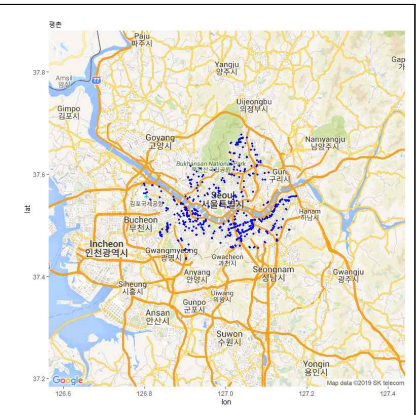


그림 16 - 평촌신도시 서울정류장

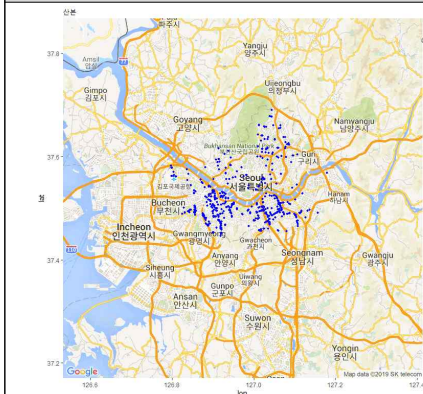


그림 17 - 산본신도시 서울정류장

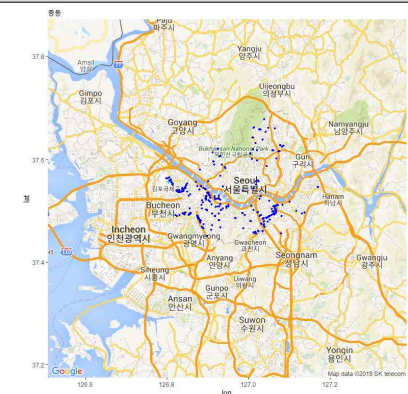


그림 18 - 중동신도시 서울정류장

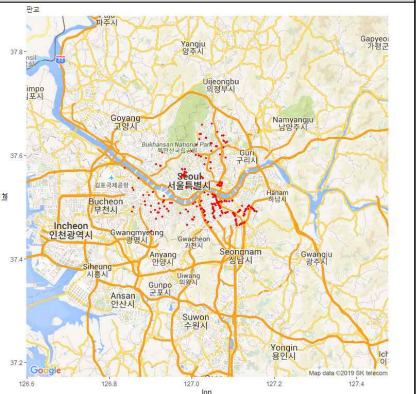


그림 19 - 판교신도시 서울정류장

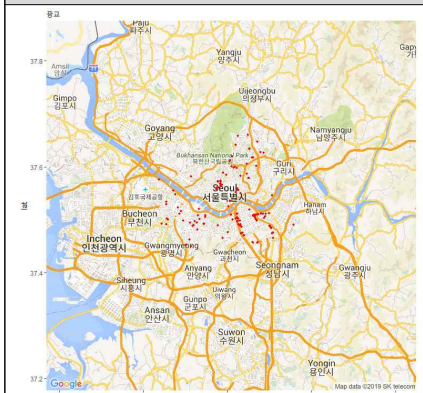


그림 20 - 광교신도시 서울정류장

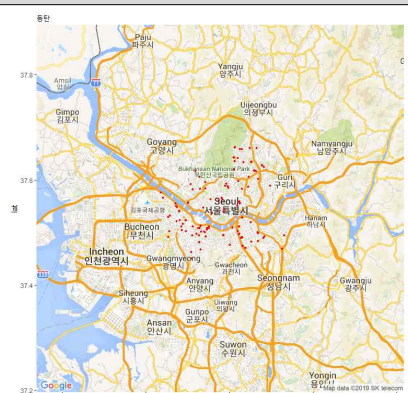


그림 21 - 동탄신도시 서울정류장

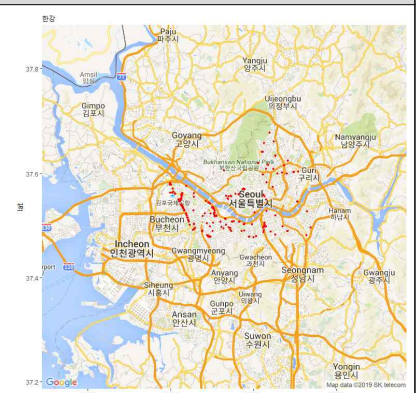


그림 22 - 한강신도시 서울정류장

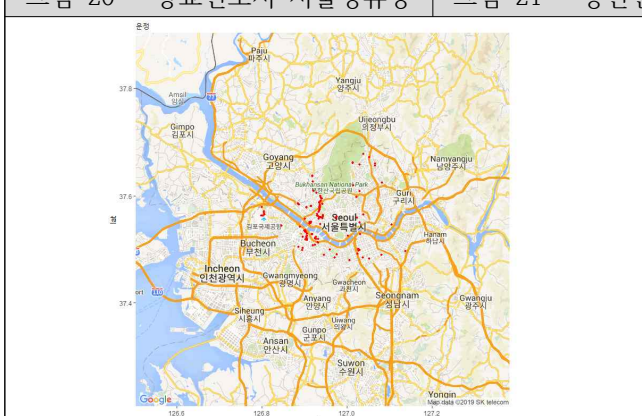


그림 23 - 운정신도시 서울정류장

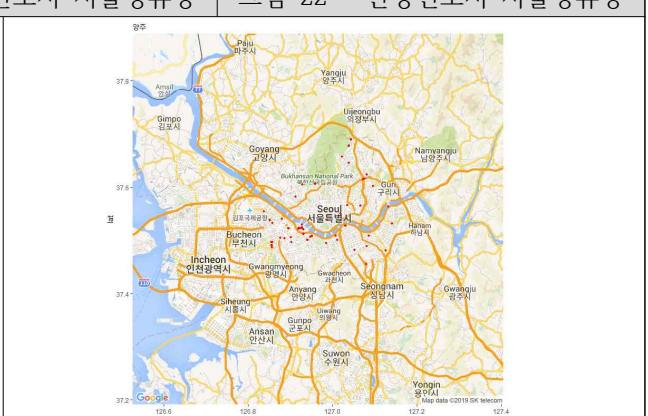
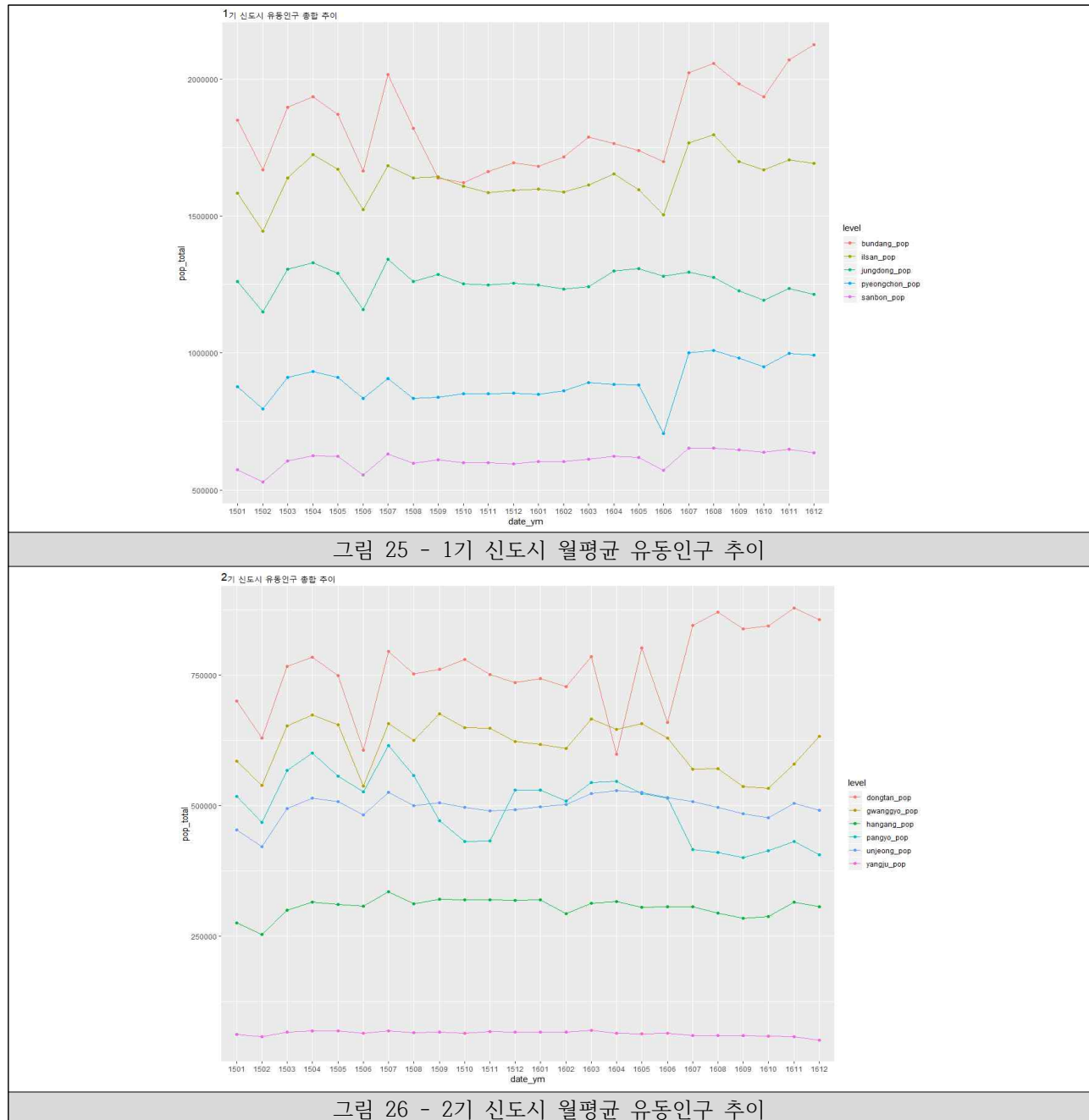


그림 24 - 양주신도시 서울정류장

: 신도시의 성공 요인 중 ‘교통’ 이 주요인인지 판단하기 위해 신도시별 버스 노선과 서울정류장 시각화를 통해 교통이 신도시의 성공에 미치는 영향성을 판단했다. 시각화한 그림을 보면 성공한 신도시는 교통과 서울과의 접근성이 모두 양호했음을 확인할 수 있다. 따라서 교통요인은 신도시에 성공에 인과관계가 있다고 판단하고 버스(bus)와 지하철역의 개수(subway) 승하차 인원(subOD\_day, subOD\_year)을 변수에 추가했다.

### • 3. 유동인구 시각화

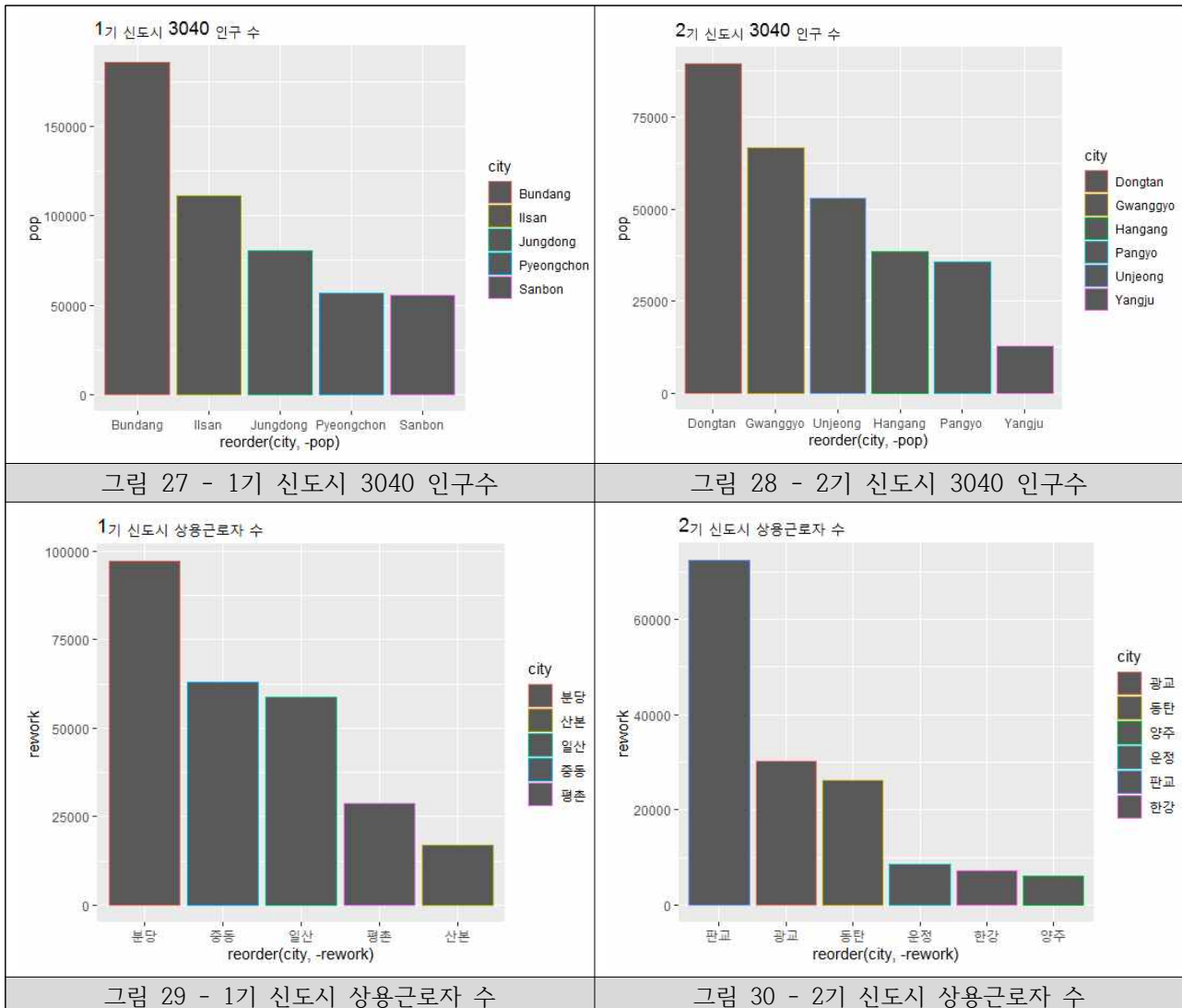


: 신도시의 성공 요인 중 ‘상권’ 은 유동인구가 주요인이라 판단, 내부 데이터의 유동인구에 대해 분석했다. 유동인구의 월평균, 연령별, 남녀비율에 대해 종합적으로 고려해 시각화 해보았다. 다각도에서 유동인구를 시각화해 분석한 결과, 유동인구의 연령별, 남녀비율 기준은 의미 있는 정보를 확인할 수 없었고 월평균 유동인구만이 신도시



분석의 변수로 채택했다. 따라서 상권 요인을 고려하기 위해 도시별 월평균 유동인구(pop\_float), 1000대 상권의 개수(sale)를 변수에 추가했다.

#### • 4. 3040 인구비율, 교육, 의료 시각화



: 신도시의 성공 요인 중 ‘자급능력’을 판단하기 위해 다양한 변수를 고려했다. 3040 인구비율(pop3040), 상용근로자 수(pop\_rework), 일자리 질(work\_qual), 학교 수(school), 교육매출(edu), 의료매출(medic)이 자급능력을 평가할 변수라 생각했고 이들을 추가했다. 시각화 결과, 상용근로자 수를 제외한 대부분 변수는 성공한 신도시의 기준을 의미 있게 제시하고 있지 않아 보인다. 하지만 자급능력에 임의로 중요도를 부여해 의미 없어 보이는 변수도 분석에 추가했다.

### 3) 모델링

#### ① 주성분 분석(PCA)

신도시의 특성을 고려하여 다양한 변수를 선택했으나 변수들이 서로 상관성이 높은 다중 공선성이 매우 크다는 문제점을 발견하였다. 이에 따라, 다중 공선성을 제거하기 위해 ‘주성분 분석(PCA)’을 진행하였다.

#### ② k-근방 분류(knn)

이 분석방법은 주성분 분석을 토대로 진행하였다. 앞서 언급한 주성분 분석을 통해 로지스틱 모형, 판별분석 모형을 만들 수 있었으나, 본 분석에서 사용한 데이터는 선형성이 없는 것으로 판단되었다. 따라서, ‘거리’라는 개념을 사용하여 분류할 수 있는 ‘k-근방 분류(knn)’ 모형을 채택하였다. 본 분석에서 사용한 거리는 ‘표준화 거리’이다. 이 거리를 토대로 k 개수만큼의 이웃(거리로 가까운 개체)을 선택한 후 다수결을 통해 분류하는 모델이다.

#### ③ 본 분석의 모델링 장점

실제 데이터를 돌린다고 가정하자. 해당 도시의 성공 및 실패 여부 판단 혹은 신도시 보장사항을 어떻게 할지 살펴보는 경우, 모든 개체를 이용하여 분석하는 것은 어려운 일이며 시간과 비용이 많이 드는 단점이 존재한다. 그러나 ‘knn’을 이용하면 최적 이웃의 개수를 찾아 분석에 필요한 개체의 수가 한정되므로 더욱 효율적인 분석이 될 것이라는 점을 고려하여 ‘주성분 분석(PCA)’과 ‘k-근방 분류(knn)’ 모형을 채택하였다.

#### • 1. 전체 데이터를 활용한 주성분 분석 및 knn

##### <주성분 분석>

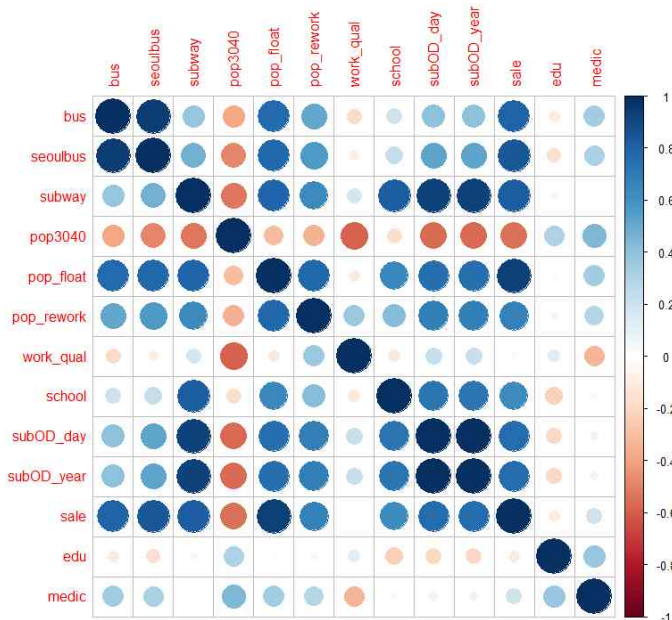
##### ① 변수 요약

```
> str(town2_1)
'data.frame': 11 obs. of 14 variables:
 $ success : num 1 1 0 0 1 1 1 0 0 0 ...
 $ bus      : num 280 273 220 171 58 59 58 89 44 31 ...
 $ seoulbus : num 845 914 679 649 351 323 175 145 243 162 ...
 $ subway   : num 6 7 2 2 10 1 3 1 0 2 ...
 $ pop3040   : num 0.343 0.33 0.347 0.339 0.328 ...
 $ pop_float : num 43923898 39219439 21409197 14647505 30191760 ...
 $ pop_rework : num 97146 58813 28609 17079 62825 ...
 $ work_qual : num 0.657 0.647 0.645 0.669 0.69 ...
 $ school    : num 57 91 26 29 124 18 23 78 24 32 ...
 $ subOD_day : num 234266 139742 92082 91873 347569 ...
 $ subOD_year : num 8.55e+07 5.10e+07 3.36e+07 3.40e+07 1.27e+08 ...
 $ sale      : num 8 10 4 2 6 0 1 0 0 0 ...
 $ edu       : num 40489388 32882949 57237093 21879915 21553949 ...
 $ medic     : num 1.50e+08 1.24e+08 1.86e+08 9.28e+07 1.13e+08 ...
```



## ② 변수 간의 상관관계 확인

```
> x <- cor(town2_1[, -1])
> library(corrplot)
> corrplot(x)
```



▶ 결과 : 변수 간의 상관관계가 크게 나타났음을 확인할 수 있다.

## ③ 주성분 분석(PCA)

```
> town.pca1 <- prcomp(town2_1[, -1], center = T, scale = T)
> summary(town.pca1)
```

```
> summary(town.pca1) # 3번째 주성분 부터 누적설명력이 80%를 넘기 때문에 PCA3까지 선택
Importance of components:
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Standard deviation	2.615	1.4892	1.1882	1.1515	0.77474	0.57022	0.41964	0.24235	0.20803	0.05817	4.483e-16
Proportion of Variance	0.526	0.1706	0.1086	0.1020	0.04617	0.02501	0.01355	0.00452	0.00333	0.00026	0.000e+00
Cumulative Proportion	0.526	0.6966	0.8052	0.9072	0.95333	0.97835	0.99189	0.99641	0.99974	1.00000	1.000e+00

▶ 결과 : 누적설명력(Cumulative Proportion)이 80%를 넘는 PC3(주성분 3)까지 선택한다.

## ④ 주성분들의 고유벡터

```
> town.pca1$rotation
```

```
> town.pca1$rotation
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
bus	-0.27066333	0.30100550	-0.42269911	0.094951634	-0.11034539	-0.08344118	-0.10771771	-0.52252573	0.38043623	-0.05739478	-0.190147609
seoulbus	-0.29791789	0.22464956	-0.41534341	0.078303727	-0.04130055	-0.09689829	0.01419864	-0.09448630	-0.75805915	-0.04180462	-0.276972296
subway	-0.34332062	-0.14093199	0.26721383	-0.048317468	-0.25286157	-0.03693934	0.03433895	0.27032057	-0.23137916	0.37288886	-0.131179751
pop3040	0.22222728	0.40650710	0.40343779	0.004011427	0.20927038	-0.26266540	0.21109677	-0.24474850	-0.36455817	-0.23301249	-0.454301915
pop_float	-0.35544775	0.20199566	0.04015898	-0.013592843	-0.02549855	-0.32375289	0.19687795	-0.03906619	0.10580768	-0.60560286	-0.187377885
pop_rework	-0.30075324	0.01648135	-0.02224158	-0.303099479	0.54941185	-0.46390453	0.16509363	0.16411471	0.13109821	-0.20080701	0.394191628
work_qual	-0.04623636	-0.47049500	-0.19423518	-0.540843318	0.20677189	-0.05034682	-0.35312302	-0.22356457	-0.21162628	0.04915566	-0.414449887
school	-0.26448670	-0.05704291	0.50621642	0.220248341	-0.07510889	-0.19271690	-0.61853683	-0.30364575	-0.03935447	-0.10130611	0.278068305
subOD_day	-0.34692523	-0.15598403	0.19756409	-0.025203734	0.07143631	0.37191196	0.30879981	-0.18800335	0.05825410	-0.16294121	-0.080702408
subOD_year	-0.34690882	-0.15512512	0.19864404	-0.023882063	0.07223381	0.37141867	0.30687563	-0.20645106	0.05338096	-0.16968843	-0.006085943
sale	-0.36358309	0.10222106	-0.08991189	0.054612331	-0.22769477	-0.10789501	-0.16929093	0.53087958	0.06234526	-0.51966846	-0.400732920
edu	0.06789727	0.22028706	0.13326347	-0.685093808	-0.59110023	-0.05142166	0.10351207	-0.12071377	0.04024298	-0.14128087	0.237425780
medic	-0.06587941	0.54841110	0.11927242	-0.275473309	0.34047073	0.51287120	-0.37449917	0.20012561	0.08780472	0.19105029	-0.020877797

▶ 결과

- PC1은 ‘pop3040’, ‘edu’ 변수 이외의 모든 계수가 음수값을 나타내고 있다.

따라서, PC1 점수가 클수록 30~40대 인구비율과 교육매출이 크다는 것을 나타낸다.

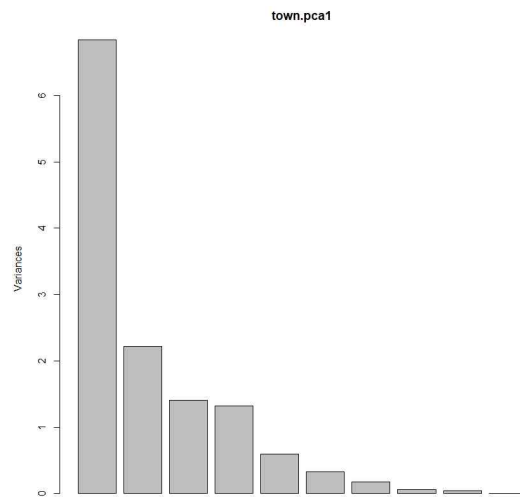
- PC2는 ‘subway’, ‘work\_qual’, ‘school’, ‘subOD\_day’, ‘subOD\_year’ 변수들의

계수만 음수값을 나타내고 있다. 따라서, PC2 점수가 클수록 지하철 개수, 상용근로자의 비율, 학교 수, 지하철 일평균 및 연평균 이용객 수가 적다는 것을 나타낸다.

- PC3는 'bus', 'seoulbus', 'pop\_rework', 'work\_qual', 'sale' 변수들의 계수만 음수값을 나타내고 있다. 따라서, PC3 점수가 클수록 지역 내 버스 개수, 서울을 지나가는 버스 정류장 개수, 근로자의 수, 상용근로자의 비율, 상권 수가 적다는 것을 나타낸다.

## ⑤ 고유값(Eigenvalues) 확인

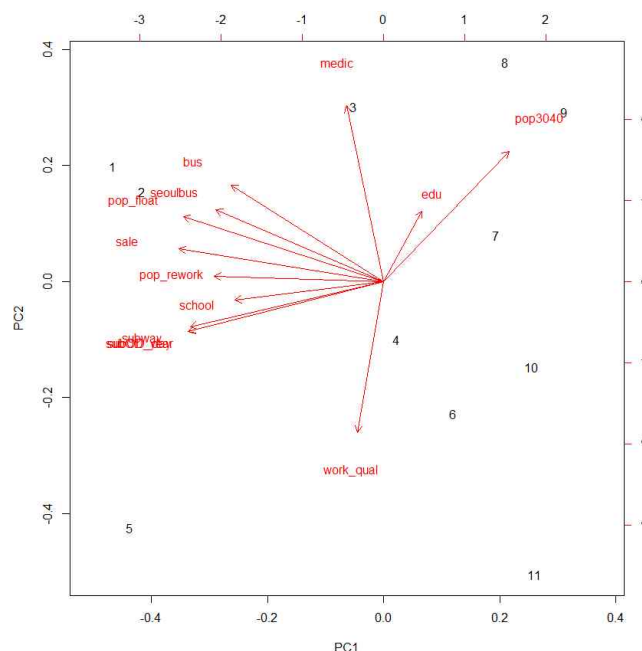
> plot(town.pca1)



- ▶ 결과 : 위 그림은 PC1~PC11까지 고유값 크기를 나타낸다. 분산이 가장 큰 주성분이 원 데이터의 흠어짐에 대한 정보를 가장 많이 가진다.

## ⑥ 주성분 1(PC1), 주성분 2(PC2) 점수 그림

> biplot(town.pca1)



## ▶ 결과

### - 1기 신도시

1 : 분당 / 2 : 일산 / 3 : 평촌 / 4 : 산본 / 5 : 중동

### - 2기 신도시

6 : 판교 / 7 : 광교 / 8 : 동탄 / 9 : 한강 / 10 : 운정 / 11 : 양주

- 분당, 일산신도시(1, 2)는 30~40대 인구비율과 교육매출이 상대적으로 적은 편임을 알 수 있다. 그러나 버스 개수, 서울을 지나가는 버스 정류장 개수가 상당히 많은 편임을 보아 교통 부분이 잘 갖추어져 있는 도시임을 확인할 수 있다. 유동인구가 많으므로 상권도 많은 편임을 볼 수 있다.

- 중동신도시(5)도 역시 30~40대 인구비율과 교육매출이 상대적으로 적은 편임을 알 수 있다. 중동은 특이하게도 앞서 언급한 분당, 일산신도시(1, 2)보다 버스가 상대적으로 적은 편임을 확인할 수 있다. 그러나 지하철이 다른 신도시에 비해 잘 갖추어져 있는 것을 볼 수 있다. 이에 따라 지하철 일평균 및 연평균 이용객 수도 압도적으로 많음을 볼 수 있다. 상용근로자 비율도 다른 신도시에 비해 높은 편에 속한다.

## <주성분 분석을 통한 knn>

### ① 주성분 계수와 데이터 내적(inner product)

```
> data <- as.matrix(town2_1[, -1]) %*% town.pca1$rotation
> data1 <- as.data.frame(cbind(town2_1$success, as.data.frame(data)))
> colnames(data1)[1] <- 'success'
> data1$success <- as.factor(data1$success)
```

### ② PC3까지 선택한 데이터 만들기

```
> library(dplyr)
> pcadata1 <- data1 %>% select(success, PC1, PC2, PC3)
```

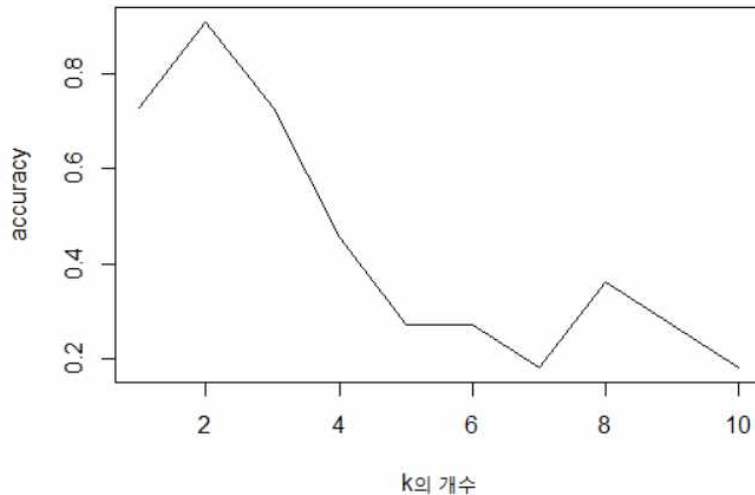
### ③ 교차 타당성 기법을 이용한 knn 분류기 만들기(적정 k값 찾기)

```
> library(class)
> acc <- matrix(data = NA, nrow = 11, ncol = 10)
> for(i in 1:nrow(town2_1)) {
+   for(j in 1:10) {
+     train <- pcadata1[-i, ]
+     test <- pcadata1[i, ]
+
+     pred <- knn(train = train[, -1], test = test[-1], cl = train[, 1], k = j, prob = T)
+     acc[i, j] <- mean(pred == test[, 1])
+   }
+ }
> acc <- as.data.frame(acc)
```



#### ④ 정확성(Accuracy) 확인

```
> accuracy <- NULL  
> for(i in 1:10) {  
+   accuracy[i] <- mean(acc[, i])  
+ }  
> plot(1:10, accuracy, type = "l")
```



▶ **결과** : k의 적정값을 찾기 위해 k를 1~10까지 설정했을 때, k = 2인 경우 정확도가 약 91%로 가장 높았다.

#### • 2. 변수 축약 후 주성분 분석 및 knn

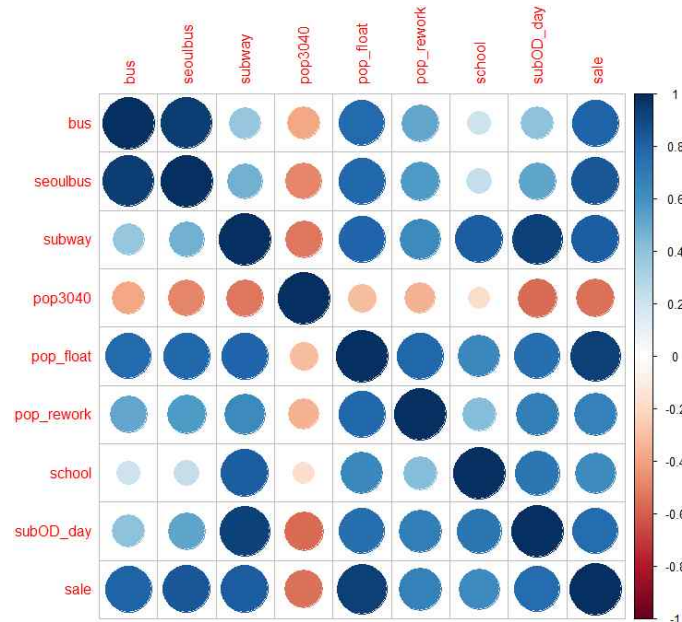
독립변수 중 다른 변수들과의 관계가 모호한 변수들( 'work\_qual' , 'edu' , 'medic' )이 발견되었으며, 이 변수들은 다른 변수들과 상관성이 작았다. 이로 인해 주성분의 설명력을 떨어뜨리는 것을 확인할 수 있었으며 이에 따라 해당 변수들을 제거하기로 하였다.

또한, 'subOD\_day' , 'subOD\_year' 변수는 상관성이 1인 것을 확인하였다. 본 분석에서는 지하철 일평균 이용객 수에 초점을 맞추기 위해 지하철 연평균 이용객 수를 나타내는 'subOD\_year' 변수는 제거하기로 하였다.

## <주성분 분석>

### ① 변수 선택

```
> town2_2 <- town2_1[, -c(8, 11, 13, 14)]
> x <- cor(town2_2[, -1])
> corrplot(x)
```



▶ 결과 : 대부분 변수에서 상관성이 큰 것을 확인할 수 있다.

### ② 주성분 분석(PCA)

```
> town.pca2 <- prcomp(town2_2[, -1], center = T, scale = T)
> summary(town.pca2)
> summary(town.pca2) # 2번째 주성분 부터 누적설명력이 80%를 넘기 때문에 PC2)까지 선택
Importance of components:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9
Standard deviation  2.4442  1.1816  0.92410  0.70363  0.36280  0.30395  0.19880  0.1161  0.06234
Proportion of Variance 0.6638  0.1551  0.09488  0.05501  0.01463  0.01026  0.00439  0.0015  0.00043
Cumulative Proportion 0.6638  0.8189  0.91378  0.96879  0.98342  0.99368  0.99807  0.9996  1.00000
```

▶ 결과 : 누적설명력(Cumulative Proportion)이 80%를 넘는 PC2(주성분 2)까지 선택한다.

### ③ 주성분들의 고유벡터

```
> town.pca2$rotation
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9
bus      -0.3072054  0.51936071 -0.14690116  0.211693028 -0.09941513  0.23685776 -0.54351076  0.06969769 -0.450831183
seoulbus -0.3316184  0.46496864 -0.02836565  0.131347369  0.24641059  0.20285067  0.70046626  0.23933319  0.077309408
subway    -0.3600458 -0.36213280  0.09770490  0.078501526  0.13165256 -0.52432393  0.07359883  0.42622897 -0.494147970
pop3040    0.2339259 -0.08468819 -0.87120384 -0.004745815  0.32411718 -0.06027488  0.09968043 -0.11637629 -0.216617851
pop_float  -0.3867825  0.04261916 -0.32204486 -0.040117018  0.03910524 -0.27395347 -0.32794003  0.29488972  0.687150873
pop_rework -0.3194612 -0.01072036 -0.14905197 -0.847679359 -0.31894652  0.12249236  0.11731539 -0.04221961 -0.157404520
school     -0.2745114 -0.52303990 -0.22744308  0.392165999 -0.44363579  0.47642664  0.12784192  0.05763694  0.045690311
subOD_day  -0.3565220 -0.30362840  0.17460116 -0.118452033  0.69468449  0.37331487 -0.20032441 -0.27123506  0.023163053
sale       -0.3964583  0.08980039 -0.03045921  0.209175027 -0.15607291 -0.40963416  0.14425363 -0.75961462  0.009610031
```

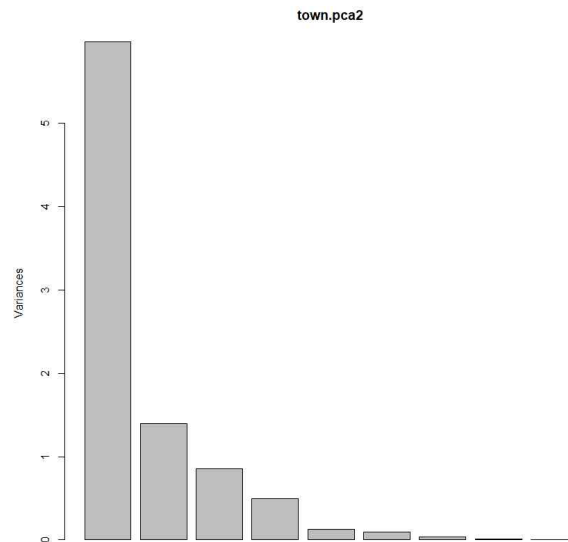
### ▶ 결과

- PC1은 'pop3040' 변수 이외의 모든 계수가 음수값을 나타내고 있다. 따라서, PC1 점수가 클수록 30~40대 인구비율이 크다는 것을 나타낸다.

- PC2는 'subway', 'pop3040', 'pop\_rework', 'school', 'subOD\_day' 변수들의 계수만 음수값을 나타내고 있다. 따라서, PC2 점수가 클수록 지하철 개수, 30~40대 인구비율, 근로자 수, 학교 수, 지하철 일평균 이용객 수가 적다는 것을 나타낸다.

#### ④ 고유값(Eigenvalues) 확인

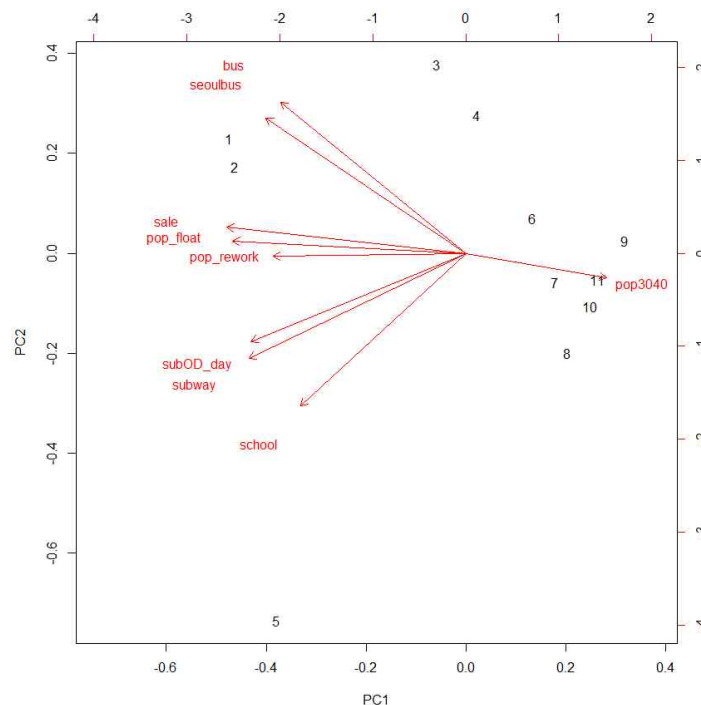
> plot(town.pca2)



- ▶ 결과 : 위 그림은 PC1~PC9까지 고유값 크기를 나타낸다. 분산이 가장 큰 주성분이 원 데이터의 흠어짐에 대한 정보를 가장 많이 가진다.

#### ⑥ 주성분 1(PC1), 주성분 2(PC2) 점수 그림

> biplot(town.pca2)



#### ▶ 결과

- 1기 신도시

1 : 분당 / 2 : 일산 / 3 : 평촌 / 4 : 산본 / 5 : 중동



- 2기 신도시

6 : 판교 / 7 : 광교 / 8 : 동탄 / 9 : 한강 / 10 : 운정 / 11 : 양주

- 분당, 일산신도시(1, 2)는 전체 데이터로 주성분 분석을 했을 때처럼 30~40대 인구비율이 상대적으로 적은 편임을 알 수 있다. 그러나 버스 개수, 서울을 지나가는 버스 정류장 개수가 상당히 많은 편이다. 지하철은 중동신도시(5)보다는 적은 편이다. 유동인구가 많으므로 상권도 많은 편임을 볼 수 있다.

- 중동신도시(5)도 마찬가지로 30~40대 인구비율이 상대적으로 적은 편임을 알 수 있다. 앞서 언급한 분당, 일산신도시(1, 2)보다도 버스가 상대적으로 적은 편임을 보여주고 있다. 그러나 지하철이 다른 신도시에 비해 잘 갖춰져 있는 것을 볼 수 있다. 이에 따라 지하철 일평균 및 연평균 이용객 수도 상대적으로 많음을 볼 수 있다. 상용근로자 비율도 다른 신도시에 비해 높은 편에 속한다.

- 2기 신도시(6~11)은 30~40대 인구비율은 높지만, 나머지는 전반적으로 낮은 편이다.

### <주성분 분석을 통한 knn>

① 주성분 계수와 데이터 내적(inner product)

```
> data <- as.matrix(town2_2[, -1]) %*% town.pca2$rotation
> data1 <- as.data.frame(cbind(town2_2$success, as.data.frame(data)))
> colnames(data1)[1] <- "success"
> data1$success <- as.factor(data1$success)
```

② PC2까지 선택한 데이터 만들기

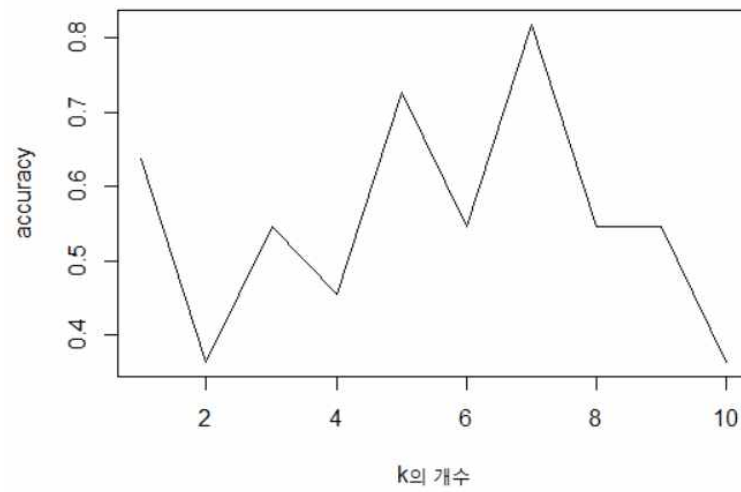
```
> pcadata2 <- data1 %>% select(success, PC1, PC2)
```

③ 교차 타당성 기법을 이용한 knn 분류기 만들기(적정 k값 찾기)

```
> acc <- matrix(data = NA, nrow = 11, ncol = 10)
> for(i in 1:nrow(town2_2)) {
+   for(j in 1:10) {
+     train <- pcadata2[-i, ]
+     test <- pcadata2[i, ]
+     pred <- knn(train = train[, -1], test = test[-1], cl = train[, 1], k = j, prob = T)
+     acc[i, j] <- mean(pred == test[, 1])
+   }
+ }
> acc <- as.data.frame(acc)
```

④ 정확성(Accuracy) 확인

```
> accuracy <- NULL
> for(i in 1:10) {
+   accuracy[i] <- mean(acc[, i])
+ }
> plot(1:10, accuracy, type = "l")
```



- ▶ **결과** : k의 적정값을 찾기 위해 k를 1~10까지 설정했을 때, k = 7인 경우 정확도가 약 82%로 가장 높았다.

### 3. 분석 결과 활용

#### 1) 현황분석을 통한 시사점 / 정책 제안

##### • 1. 현황분석

1기 신도시에 반해 2기 신도시는 성공한 신도시의 요인인 교통, 상권, 자급능력 부분에서 아직 상대적으로 부족함을 확인했다. 특히, 2기 신도시는 서울과 이격된 지역에 지어지는 만큼 자체적인 상권과 자급능력이 매우 중요했음에도 불구하고 대다수 도시는 기존 1기 신도시의 상권에 의존하거나 자체적인 상권이 갖춰지지 않았다. 심지어 가장 성공한 신도시라 평가받는 판교조차 자체적인 상권 없이 강남지역의 상권과 분당의 상권에 의존하는 모습을 보였다. 그중 한강과 양주신도시는 상권 부분은 물론이고, 교통 부분에서 매우 취약한 모습을 보여주었다.

2기 신도시의 주요 보완점을 요약해 보자면, 판교, 동탄, 운정은 자체적인 상권의 부재, 한강과 양주는 교통망 부족이라 판단된다. 2기 신도시 중 서울과 이격된 지역에 건설됨에도 불구하고 상권, 교통, 일자리 질이 모두 상위권에 속하는 광교신도시만이 본 연구에서는 이상적인 신도시라고 판단하였다.

##### • 2. 시사점

대다수의 2기 신도시조차 완전히 정착되지 않은 현재, 3기 신도시의 발표는 시기상조라고 판단된다. 하지만 이미 3기 신도시의 선정과 발표가 완료된 현재 3기 신도시의 성공 방안에 대해 논의가 필요하다.

3기 신도시는 지리적, 정책적으로 부천(중동)신도시와 연관 관계가 있다. 기존 2기 신도시와는 다르게 모두 서울과 비교적 가까운 지역에 선정되었기 때문이다. 중동신도시는 여타 1기 신도시와 다르게 버스에 대한 의존도가 매우 낮고 지하철역수와 승하차 인원이 월등히 높은 것으로 나타났다.

따라서, 3기 신도시의 최종목표인 ‘수도권 인구 흡수를 통한 주택시장 안정화’를 성공적으로 완수하기 위해서 중동신도시에 대한 집중적인 분석이 필요해 보인다.

##### • 3. 정책 제안

###### ① 부천(중동)신도시 연구

3기 신도시의 성공적인 조성을 위해서는 ‘부천(중동)신도시’의 연구가 우선시되어야 한다고 판단된다. 우선시 되어야 하는 사항은 부천의 지리적 특징, 교통, 상권, 자급능력이라고 생각된다.

본 분석에서 부천지역의 연구가 우선시되어야 한다고 한 근거는 해당 지역이 문제점을 가지고 있기 때문이다. 실제로 2000년대 이후, 부천지역의 인구 유입은 감소하였으며 이는 1기 신도시의 노후화에 따른 예견된 현상이라고 보여진다. 신도시는 주로 신혼부부가 많이 간다는 점을 고려했을 때, 부천처럼 1기 신도시의 건축이 오랜 시간이 지난다면 해당 신도시에 거주하고 있는 신혼부부가 낳은 자식들은 향후 성인이 되고 취업을 하기 위해 서울로 가게 되는 아이러니가 발생할 수 있다.



이에 따라, 부천신도시의 문제점을 파악하고 이를 사전에 대비해 3기 신도시에서는 발생하지 않도록 해야 할 것이다.

## ② 지하철 신설 및 확충 또는 기존노선 연장

앞서 시사점에서 언급하였듯이 지리적, 정책적으로 ‘부천(중동)신도시’와 상당한 연관성이 있는 것으로 보이므로 해당 신도시처럼 지하철이 잘 갖춰져 있어야 한다고 판단된다. 본 연구에서 부천의 교통 특징에 대해 분석한 결과, 수도권 인근 지역은 버스 노선 확보보다 지하철의 개통이 더 주요한 요인이라고 판단하였기 때문이다.

## 2) 실현 가능성

### ① 부천(중동)신도시 연구

부천지역이 가지고 있는 신도시로서의 문제점은 이미 많은 분석을 통해 연구되고 있다. 앞서 정책 제안에서 언급하였듯이 인구 유입이 감소한 점은 1기 신도시 노후화에 따른 예견된 현상이라고 보여진다. 따라서, 부천신도시가 크게 성장할 수 있었던 원인과 급성장으로 인한 문제점에 대해 분석하는 것은 다양한 데이터가 존재하므로 어려운 일이 아니라고 판단된다.

### ② 지하철 신설 및 확충 또는 기존노선 연장

지하철 정책은 중앙정부의 도움 없이는 지자체가 나서서 할 수 있는 일은 아니다. 현 정부가 계속되는 서울지역 부동산 시장의 과열화를 막기 위해 주택 30만 호 공급계획을 발표한 만큼, 적극적으로 지하철 확충에 힘을 써야 할 것으로 보인다.

## 3) 기대효과

### ① 신도시 자급능력 성장

정책 ①인 부천신도시의 문제점을 연구를 통해 파악하고 난 후 3기 신도시를 지을 때 재발하지 않도록 한다면, 신혼부부의 자식이 일자리를 찾기 위해 서울로 올라가는 아이러니가 발생하지 않을 것이다. 즉, 3기 신도시가 자급능력을 기를 수 있을 것으로 보인다.

### ② 서울지역의 인구 분산 및 주택가격 안정화

정책 ②인 지하철 신설 및 확충 혹은 기존노선의 연장을 통해 교통 인프라가 잘 갖춰졌다면, 직장인들의 출퇴근이 편리해지게 되므로 서울지역의 인구가 분산되면서 그에 따라 주택가격도 안정화될 것으로 보인다.

### ③ 해당 신도시 소재 지자체 성장

남양주 왕숙지구, 하남 교산지구 등 5개의 3기 신도시에 해당되는 지역인 남양주, 하남 등의 지자체에 인구가 증가하게 될 것이다. 그에 따라 해당 지역의 시민들을 위한 문화 및 여가시설, F&B(Food and Beverage) 시설이 잘 갖춰지게 되면서 상권이 발달하게 될 것이다. 즉, 상권이 발달하게 되면서 해당 신도시 소재의 지자체가 크게 성장할 수 있는 기회가 될 것으로 보인다.