

데이터 마이닝 7조 중간 보고서



12132591 정동호

12131820 이건도

12161890 하나영

<

목

차

>

1. 주제

2. EDA

1. 주제

- 주제

인천광역시 아파트 단지 데이터를 이용한 아파트 제곱미터당 평균 매매 가격 예측 및 영향을 주는 변수 분석

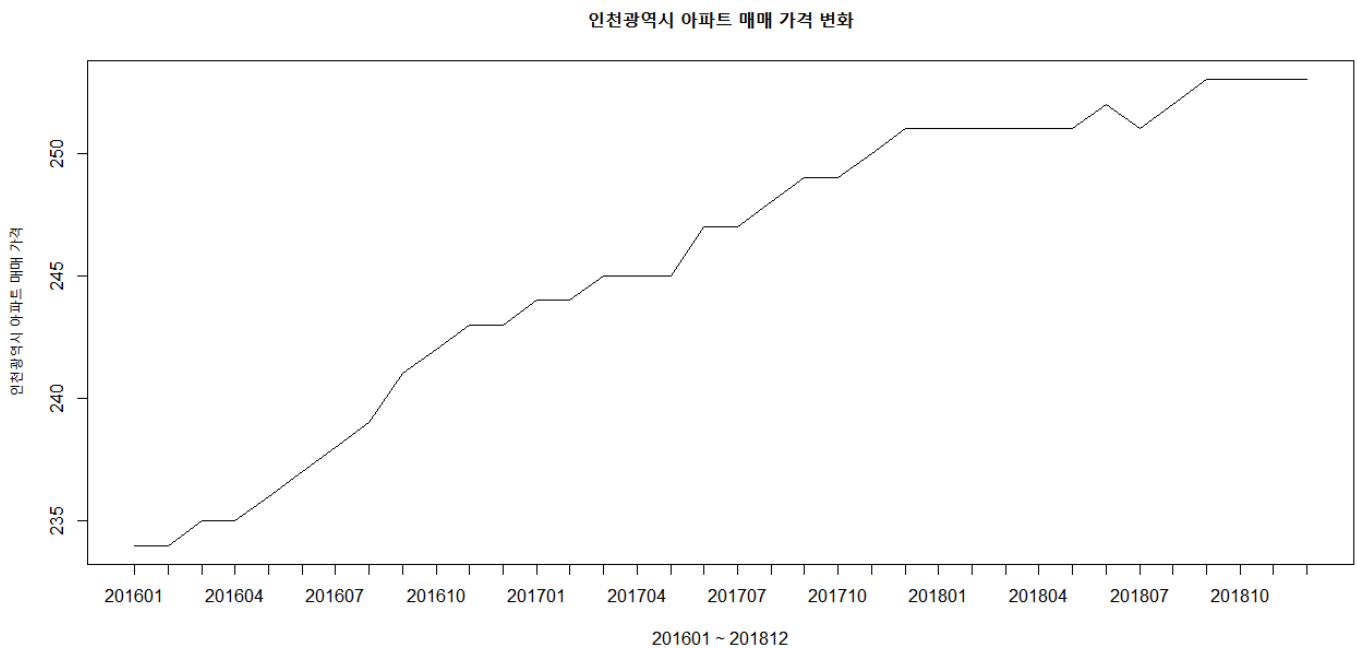
- 분석 단위

최근 3년(16년 1월부터 18년 12월까지의 월 기준) 인천광역시 단지별 아파트 평균 매매 가격 기준

2. EDA 결과

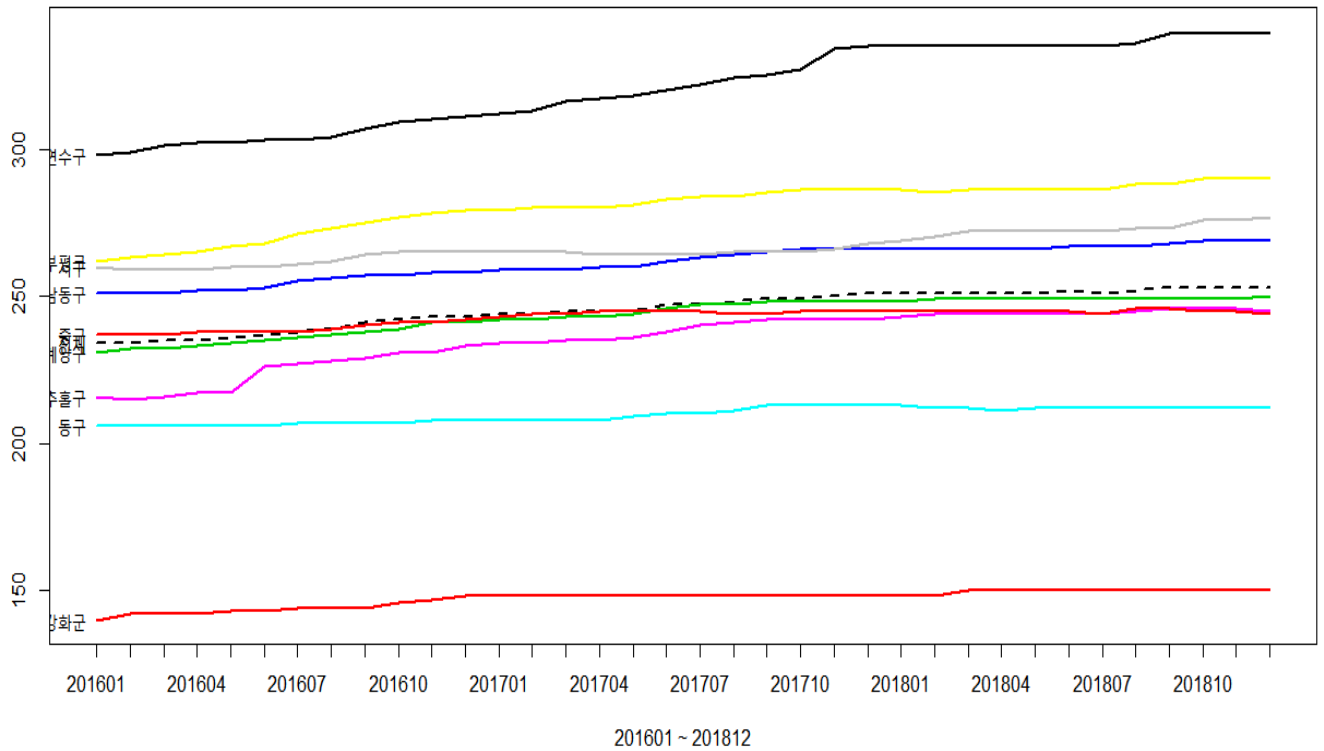
- 인천광역시 행정구역별 아파트 매매 가격 변동(16년 1월 ~ 18년 12월)

- 시계열 데이터

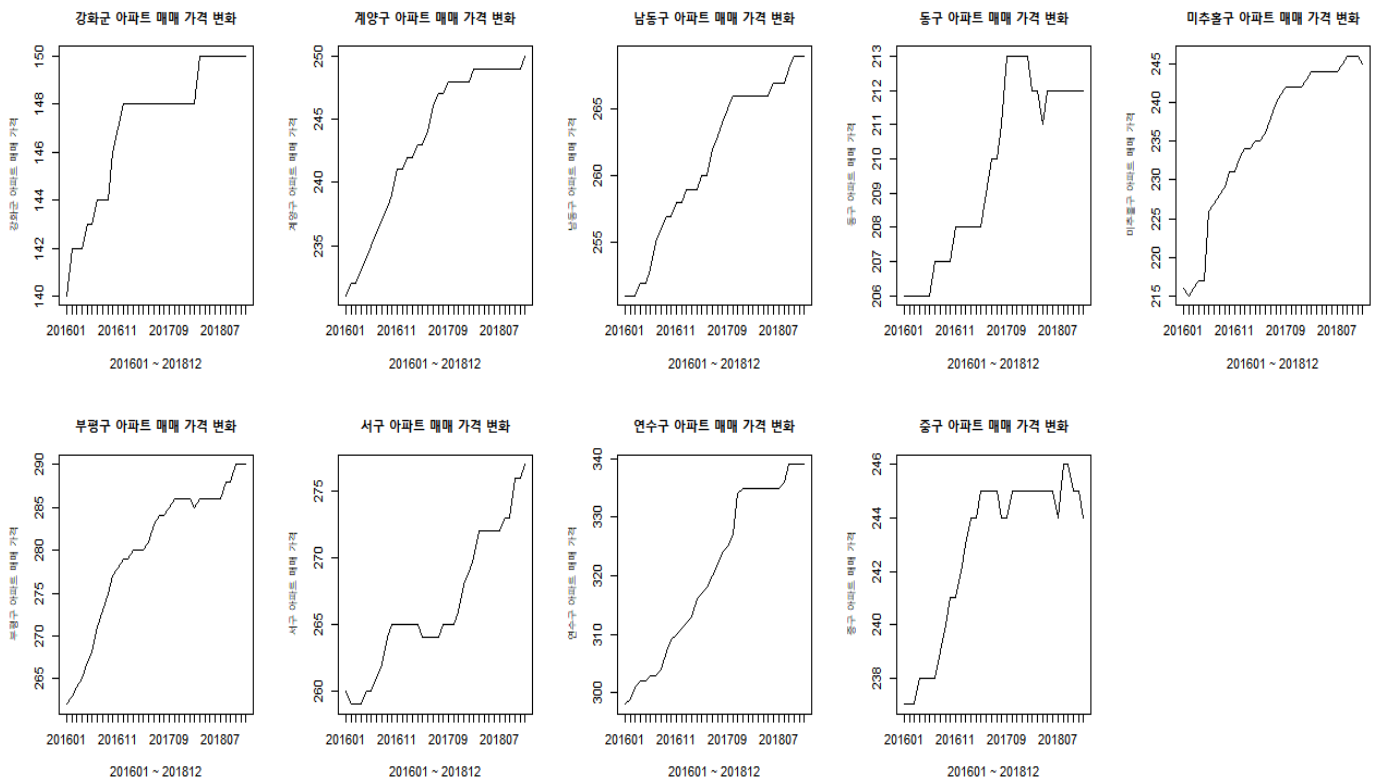


- 전반적으로 증가하는 추세를 보인다.

인천광역시 아파트 매매 가격 변화



- 연수구가 독보적으로 인천광역시 중에서 제일 높고, 강화군이 제일 낮다는 것을 확인할 수 있다.
- 미추홀구의 경우, 7위로 하위권이였지만, 꾸준히 증가하여 중구와 계양구 수준으로 올라간 것을 확인할 수 있다.

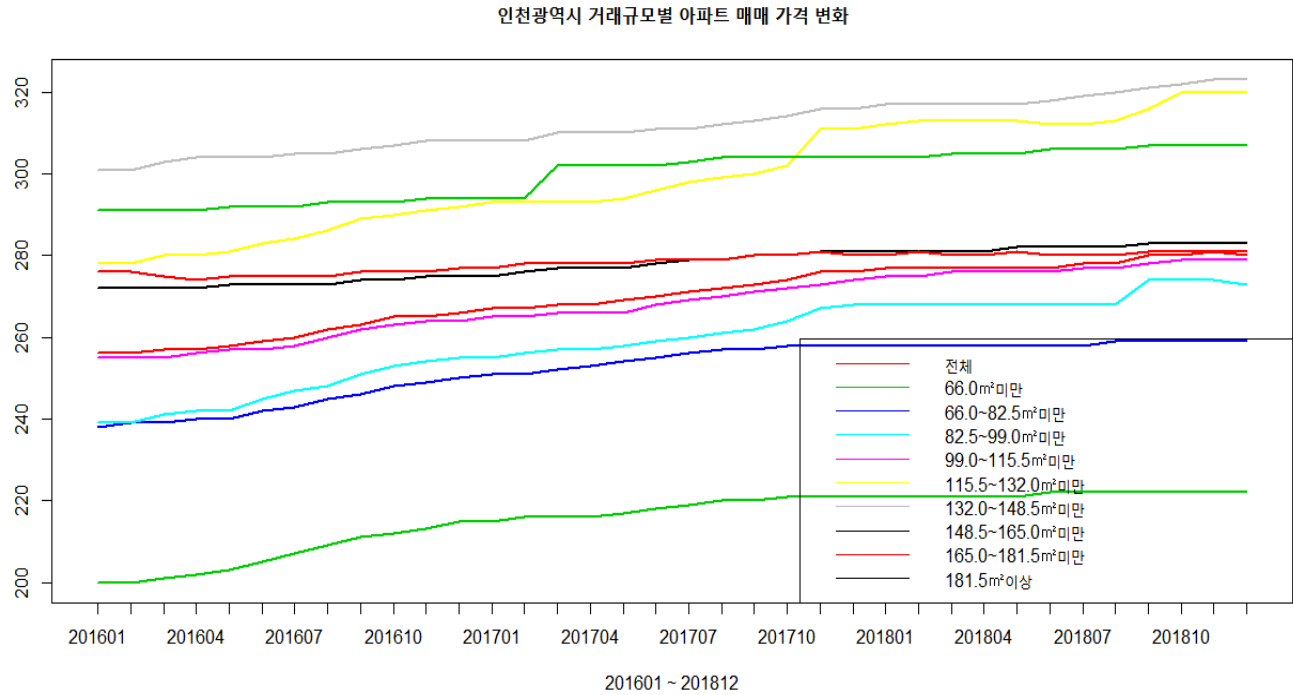


- 모든 행정구역이 전체적으로 증가한 것을 확인할 수 있다.

- 다만 동구의 경우, 최고점 이후 감소했다는 것을 확인할 수 있다.

● 인천광역시 거래규모별 아파트 매매 가격 변동(16년 1월 ~ 18년 12월)

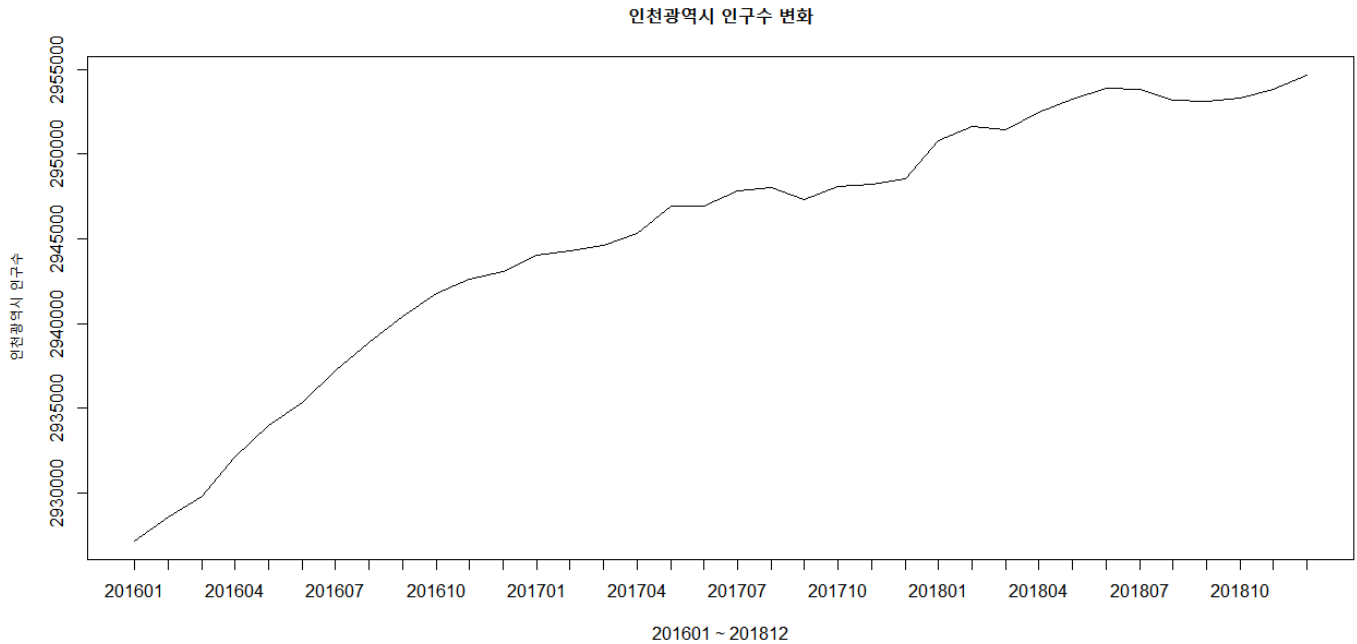
■ 시계열 데이터



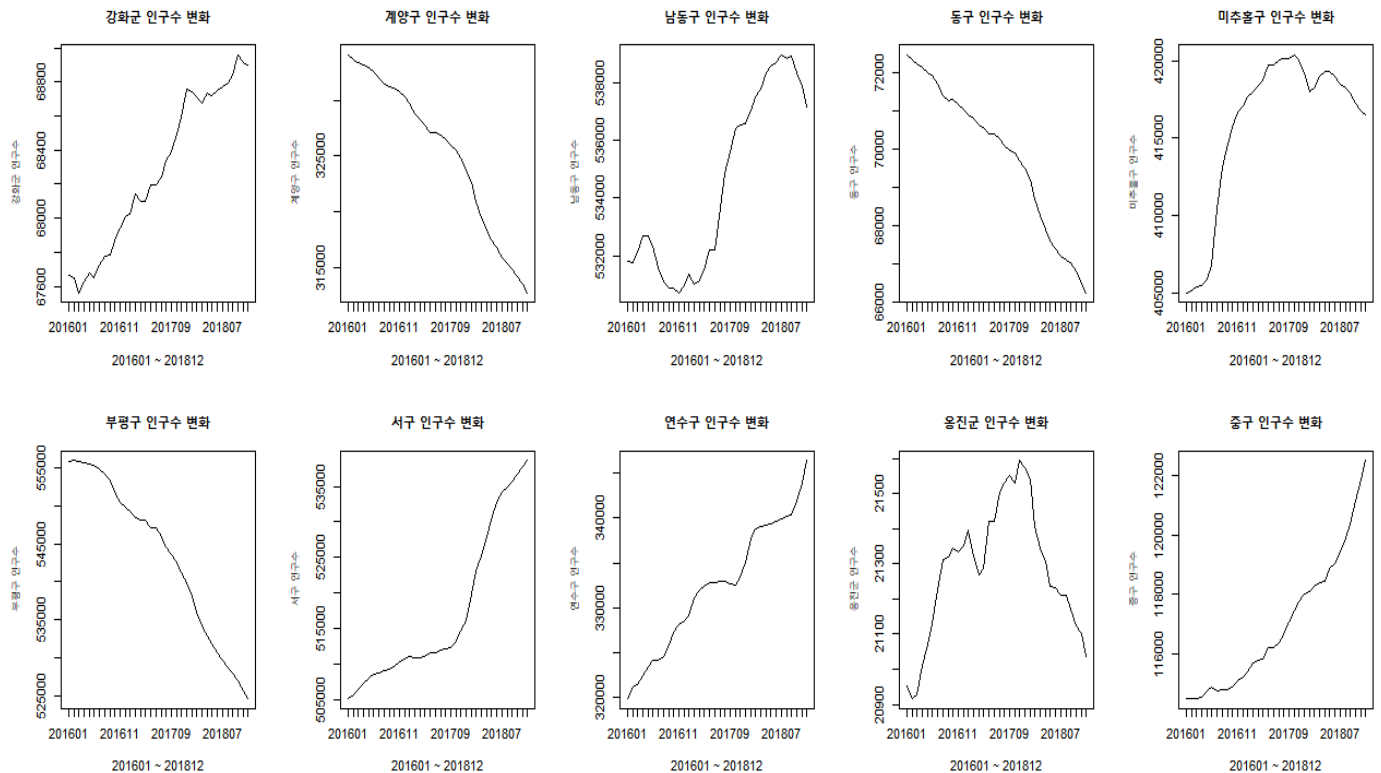
- 전반적으로 증가 추세를 보이고 있고, 규모가 큰 아파트보다는 중간 규모 수준의 아파트 가격대가 더 높은 것을 확인할 수 있다.

● 인천광역시 인구수 변화(16년 1월 ~ 18년 12월)

■ 시계열 데이터



- 우선 인천광역시 인구수는 전체적으로 증가한 것을 알 수 있다.



- 세부적으로 본다면, 강화군, 서구, 연수구, 중구는 전체적으로 증가하는 것을 알 수 있다.
- 계양구, 동구, 부평구의 경우는 지속적으로 감소한 것을 확인할 수 있다.
- 남동구의 경우, 16년에는 증가하다 감소하고 이후 다시 증가를 하다가 다시 감소하는 형세를 보인다.

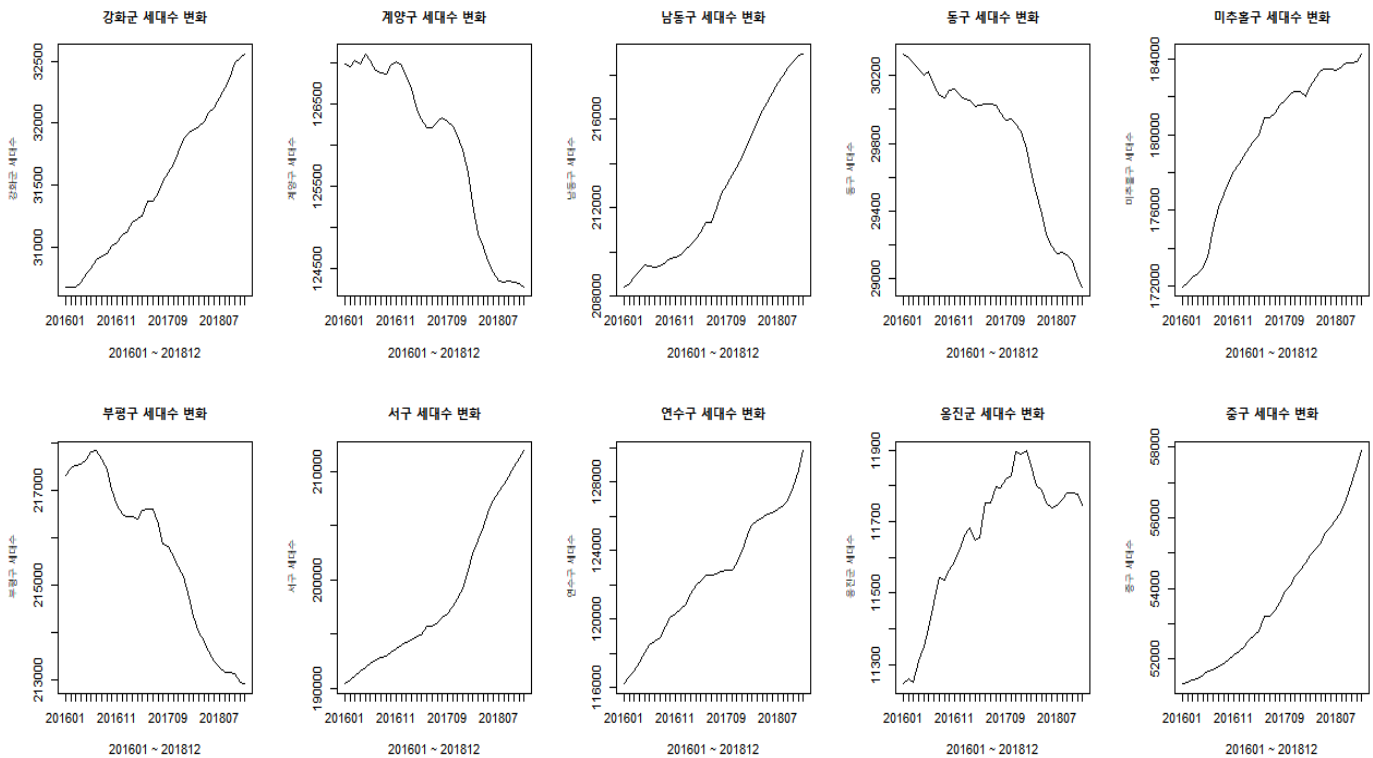
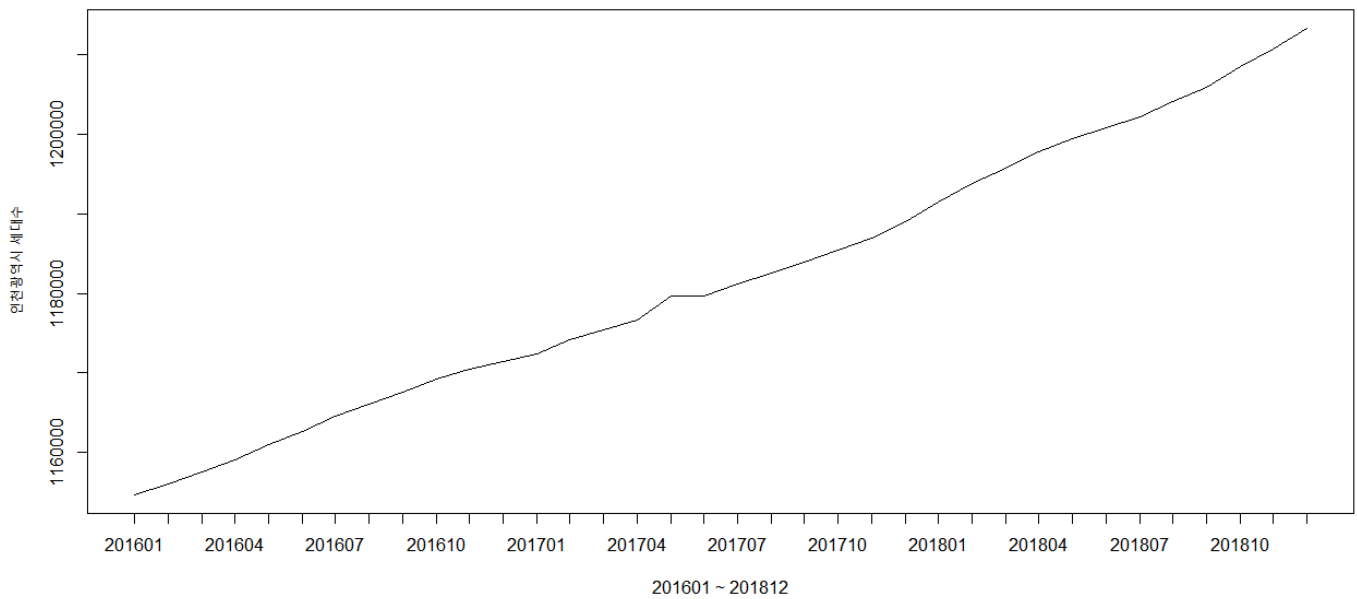
- 미추홀구의 경우, 급격하게 증가하다가 최고점 이후 감소세를 보이고 있다.
- 용진군은 전체적으로 증가했다가 감소세를 보이고 있다.

● 인천광역시 세대수 변화(16년 1월 ~ 18년 12월)

■ 시계열 데이터

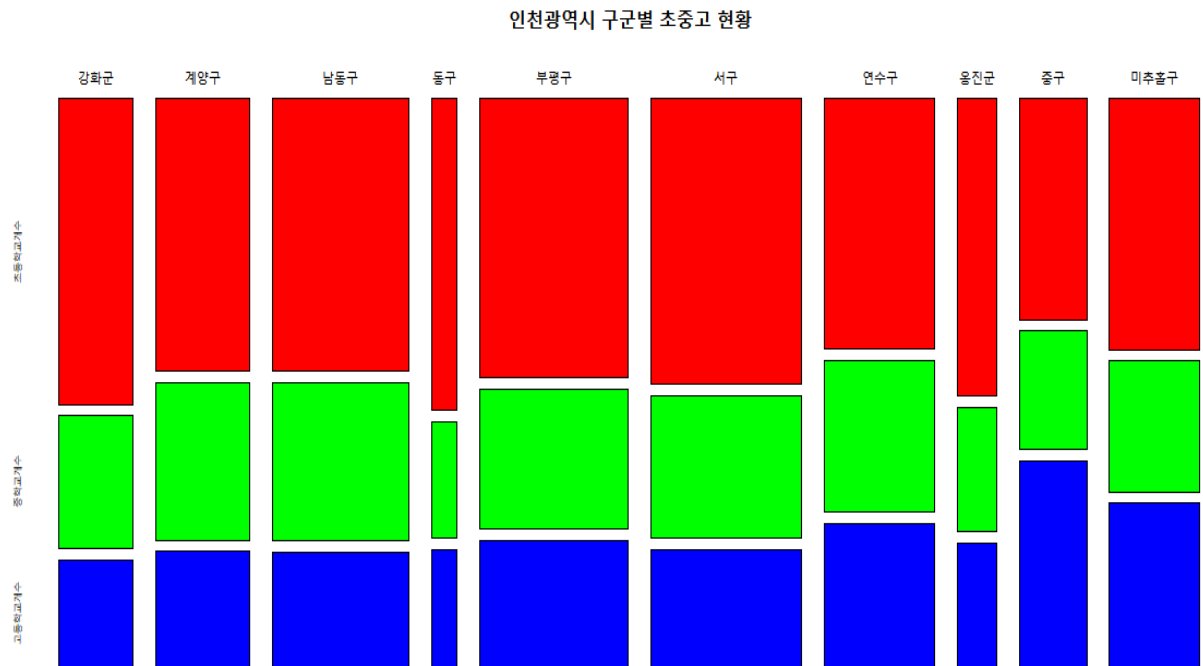
■ 세대수는 직접적으로 인구수 변화에 영향을 받기 때문에 그 변화가 유사하다

인천광역시 세대수 변화



● 인천광역시 초중고 현황

■ 범주형 데이터



- 빨간색(초등학교 수), 녹색(중학교 수), 파란색(고등학교 수)
- 모든 행정구역이 초등학교 수가 비중이 크다는 것을 확인할 수 있다.
- 중구의 경우, 초등학교와 고등학교 비중이 비슷하다
- 중학교와 고등학교 비중은 거의 비슷하거나 중학교 비중이 약간 큰 것을 확인할 수 있다.
-

	초등학교개수	중학교개수	고등학교개수	Row Total
강화군	23 0.315	10 0.041	8 0.356	41
계양구	26 0.000	15 0.161	11 0.185	52
남동구	38 0.000	22 0.252	16 0.290	76
동구	8 0.147	3 0.113	3 0.041	14
부평구	42 0.028	21 0.005	19 0.028	82
서구	44 0.103	22 0.001	18 0.246	84
연수구	28 0.196	17 0.081	16 0.117	61
용진군	12 0.095	5 0.091	5 0.017	22
중구	15 0.650	8 0.274	14 2.906	37
미추홀구	23 0.153	12 0.078	15 0.726	50
Column Total	259	135	125	519

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 7.697246 d.f. = 18 p = 0.9828108

- 분할표의 경우
- p-value가 0.05보다 크므로 초중고와 인천광역시 구군 변수 간의 연관성이 없다고 볼 수 있다.

- 인천 3년 평균 세대수(16년 1월 ~ 18년 12월의 세대수의 3년 평균)

■ 연속형 데이터

- 해당 EDA경우, 첨부1_인천3년평균세대수.pdf에 있습니다.

- 인천 3년 평균 인구수(16년 1월 ~ 18년 12월의 인구수의 3년 평균)

■ 연속형 데이터

- 해당 EDA경우, 첨부2_인천3년평균인구수.pdf에 있습니다.

- 인천아파트매매가격 vs. 인천 3년 평균인구수

■ 연속형 vs. 연속형

- 산점도와 박스플롯의 경우 그룹화된 범주형 변수와 연속형으로 보인다.

- 해당 EDA경우, 첨부3_인천아파트매매평균vs인천3년평균인구수.pdf에 있습니다.

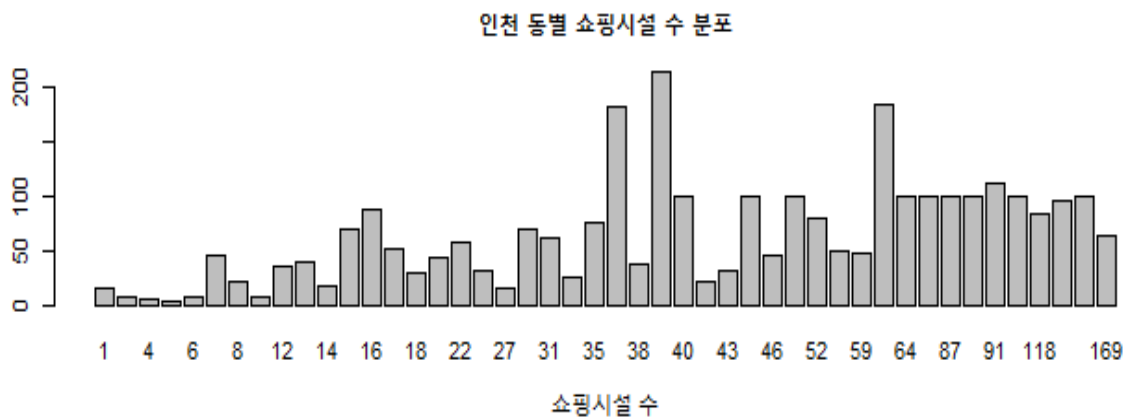
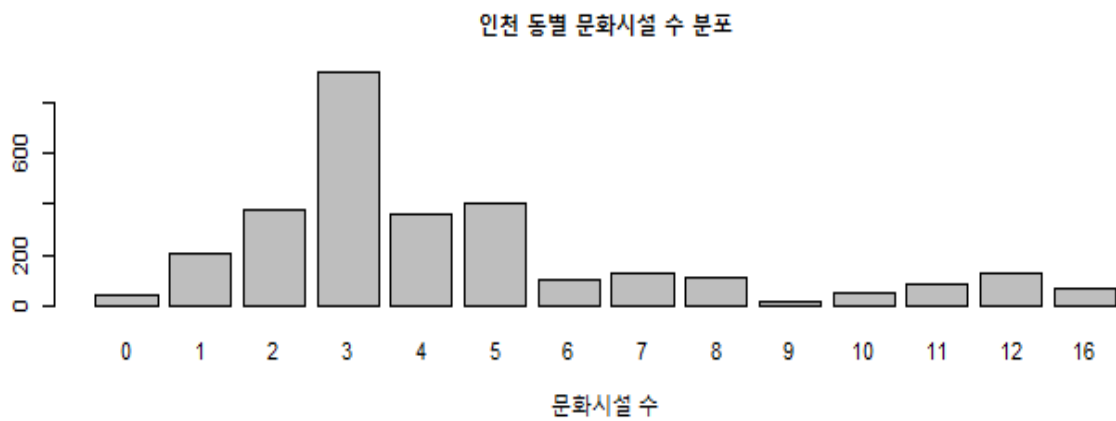
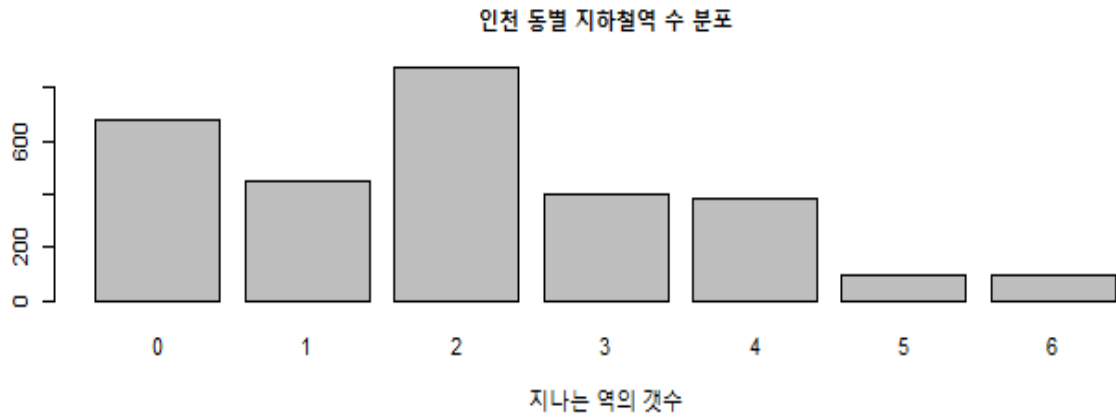
- 인천아파트매매가격 vs. 인천초중고현황

- 연속형 vs. 연속형

- 산점도와 박스플롯의 경우 그룹화된 범주형 변수와 연속형으로 보인다.

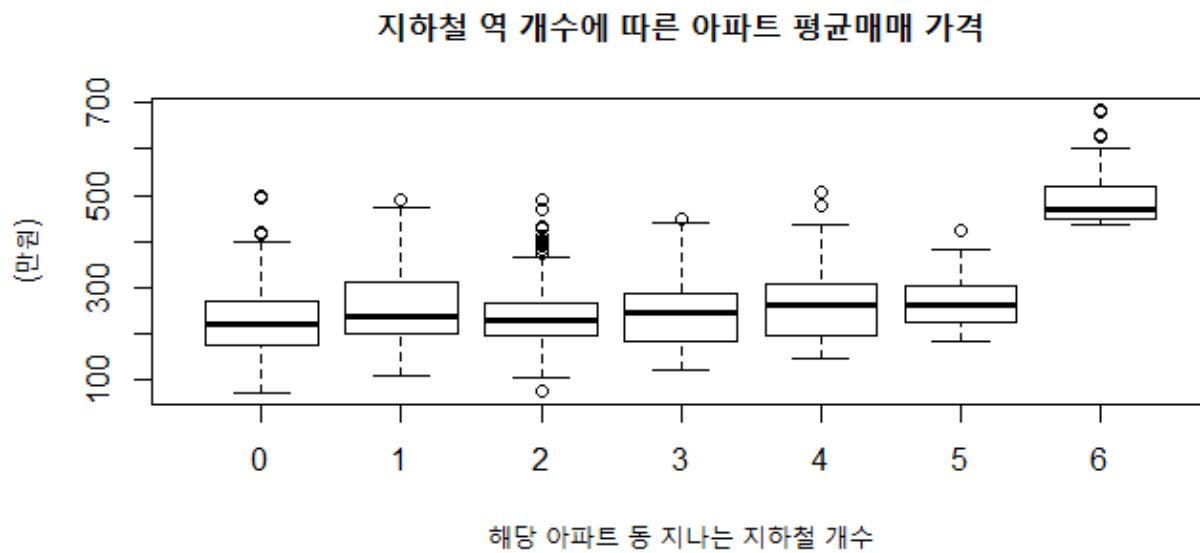
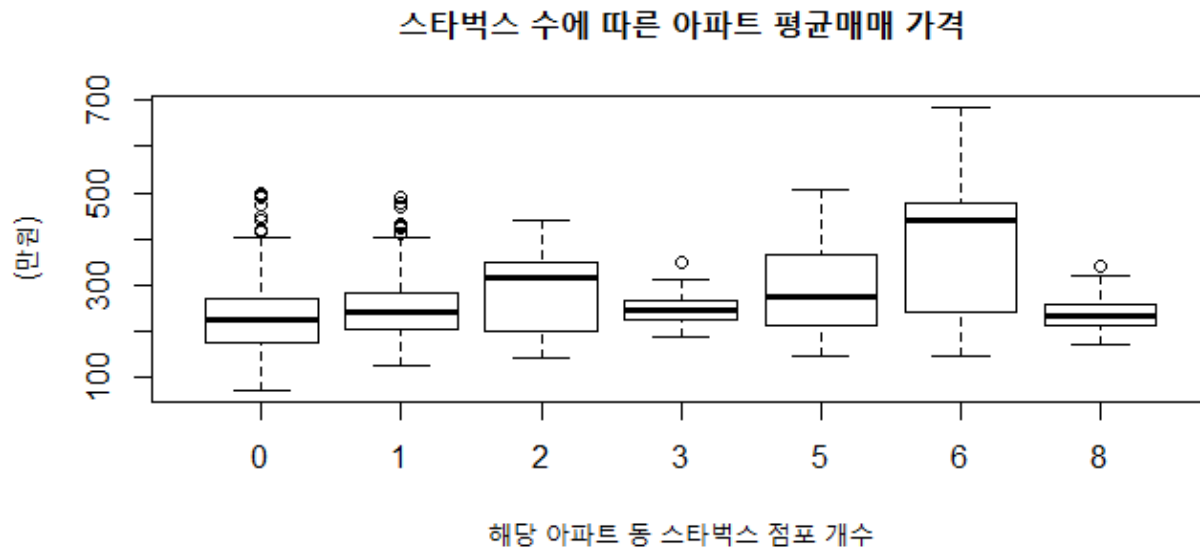
- 해당 EDA경우, 첨부4_인천아파트매매평균vs인천초중고현황.pdf에 있습니다.

● 인천 동별 지하철역



- Barplot of '인천 동 별 지하철역 개수, 문화시설 수, 쇼핑시설 수 '

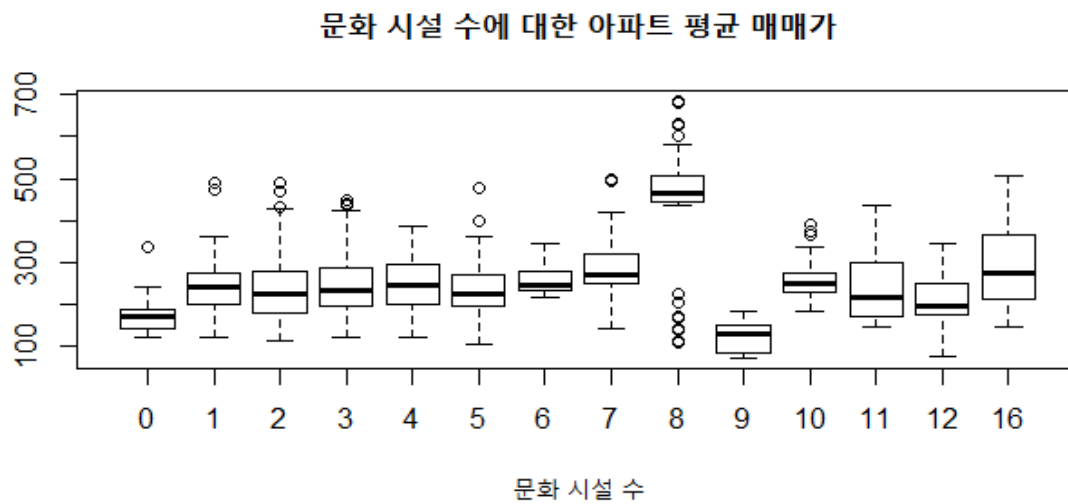
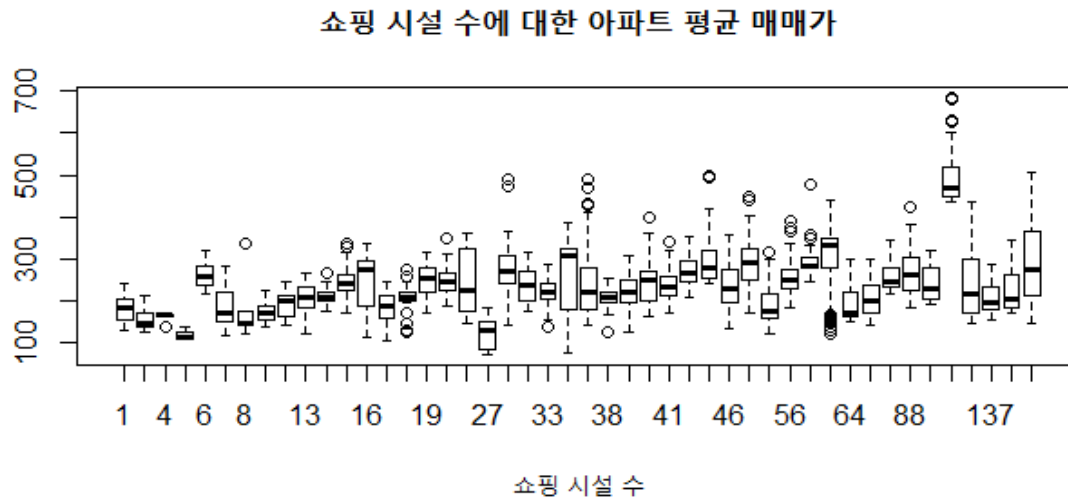
- 인천아파트평균매매 vs 스타벅스 수, 지하철 역 개수



- Boxplot of ' 스타벅스, 지하철 개수에 따른 아파트 평균 매매가격 '

첫번째 그림에서 스타벅스 8개점포의 동은 인천공항이 있는 중구 운서동.

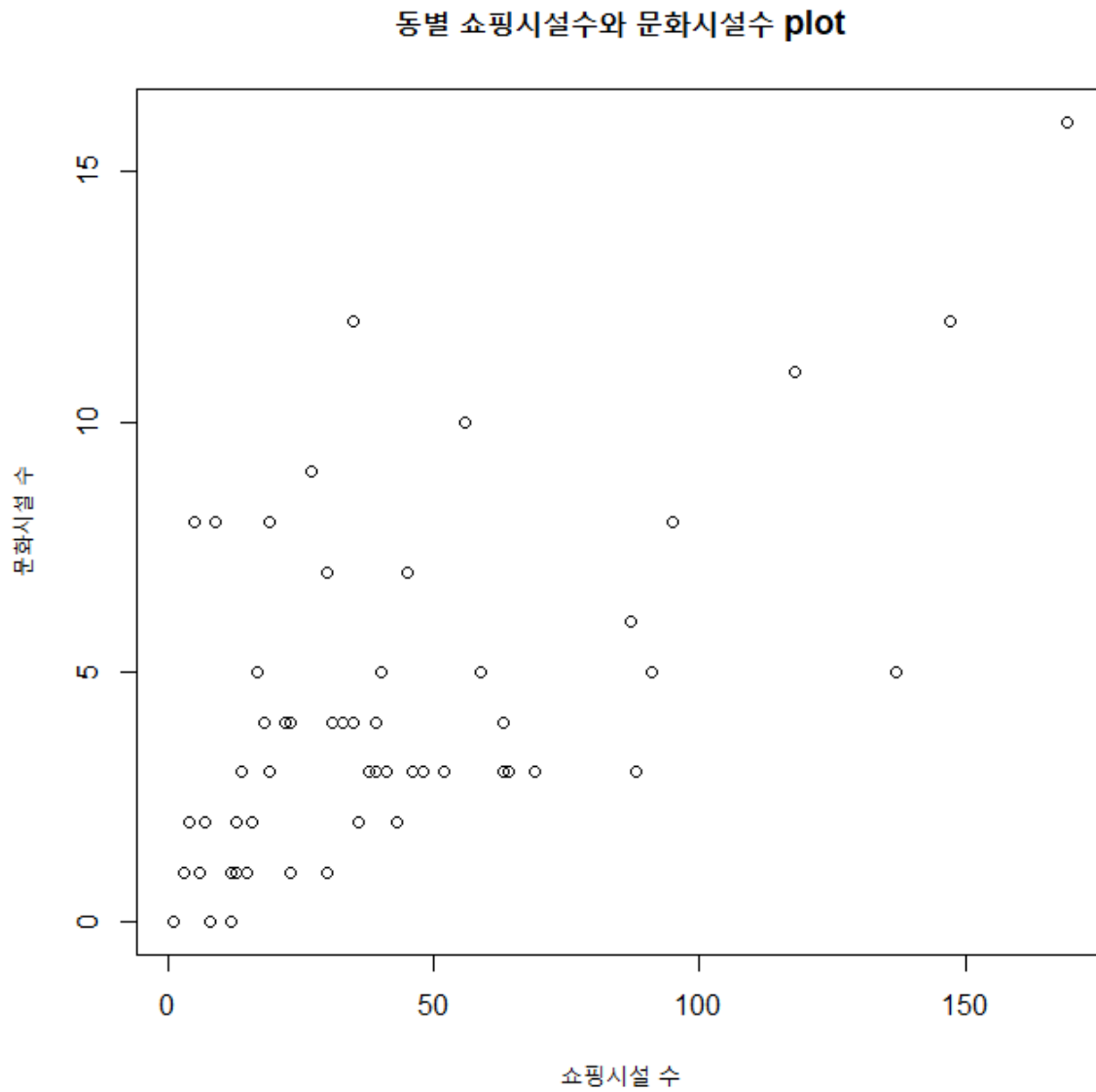
- 인천아파트평균매매 vs 쇼핑시설 수, 문화시설 수



- Boxplot of ' 쇼핑시설 수, 문화시설 수에 따른 아파트 평균 매매가격' .

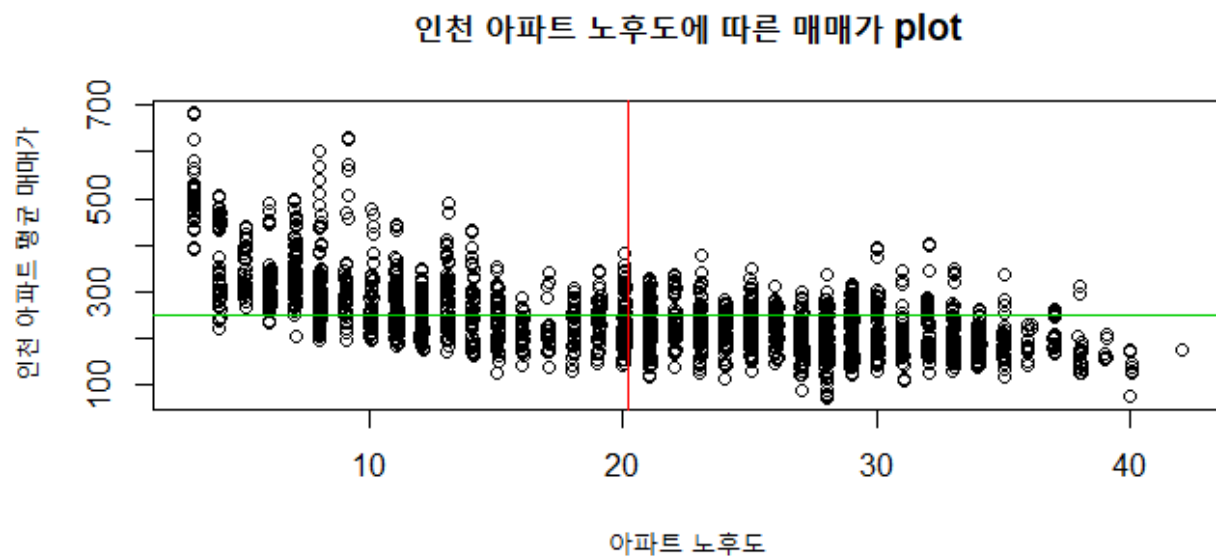
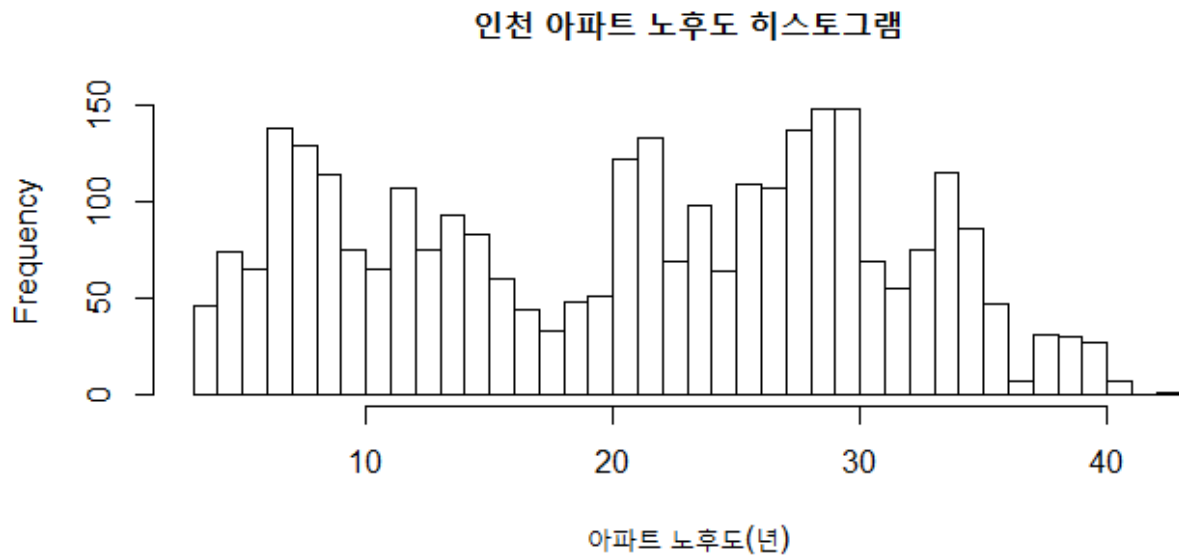
- 쇼핑 시설 : 대형마트, 백화점, 편의점 ,마켓 등
- 문화 시설 : 영화관, 박물관, 공원 등

- 쇼핑시설 수 vs 문화시설 수



- 인천 동 별 쇼핑 시설 수에 대한 문화시설 수 plot (cor = 0.71)

- 인천 아파트 노후도 & 인천아파트 평균 매매가 vs 인천아파트 노후도



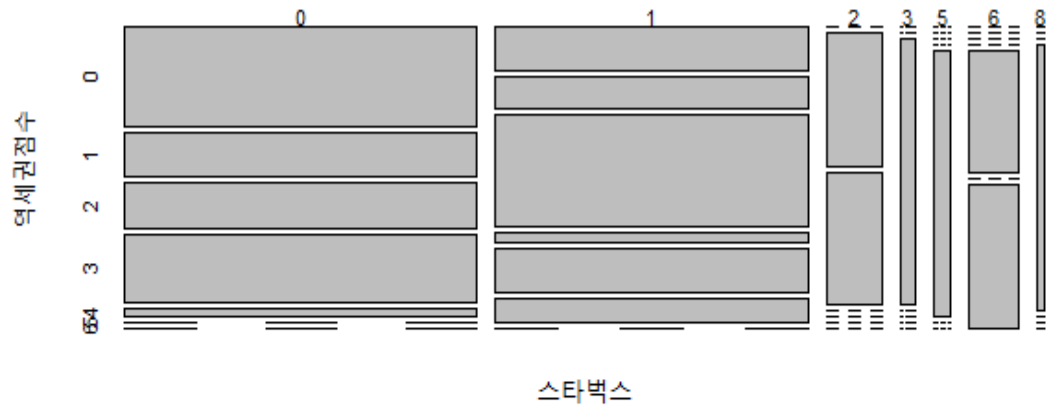
- Histogram of '아파트 노후도' & Scatter plot of '아파트 노후도에 대한 아파트 평균매매가'

빨간 선 : 노후도 평균(약 20년)

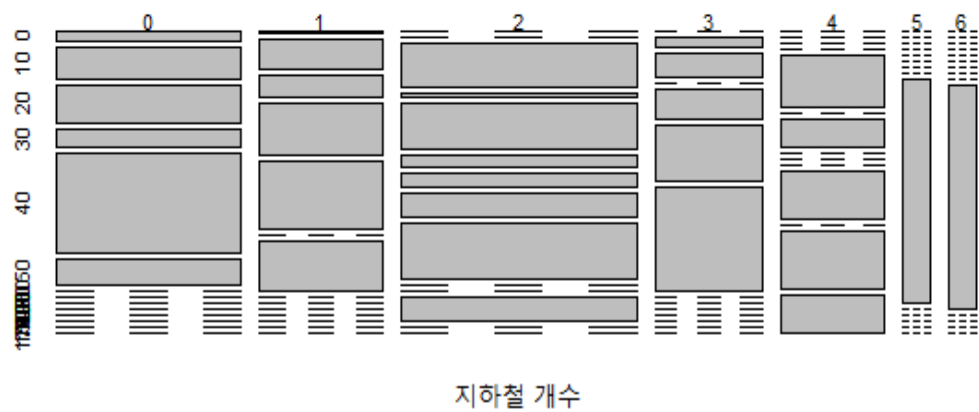
녹색 선 : 인천 아파트 최근3년 평균 매매가의 평균(약 249만원)

- 스타벅스 수 vs 지하철 역 수, 지하철 역 수 vs 쇼핑시설 수

스타벅스 수와 지하철역 수의 모자이크 그림

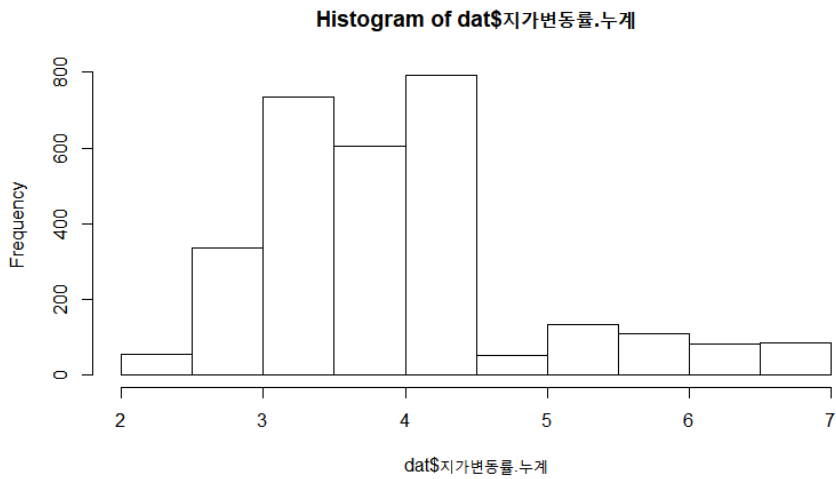


지하철역 수와 쇼핑시설 수의 모자이크그림

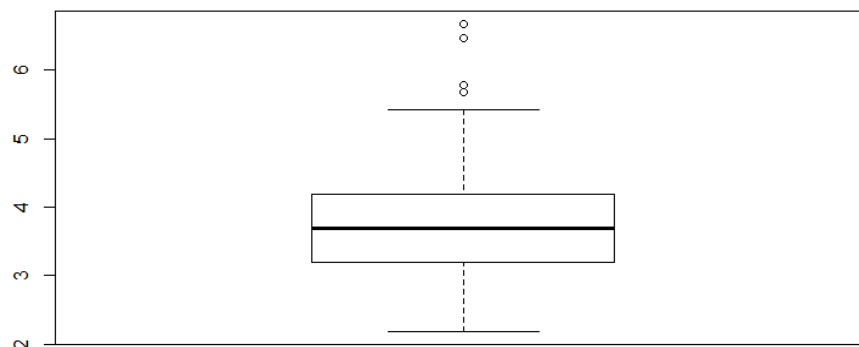
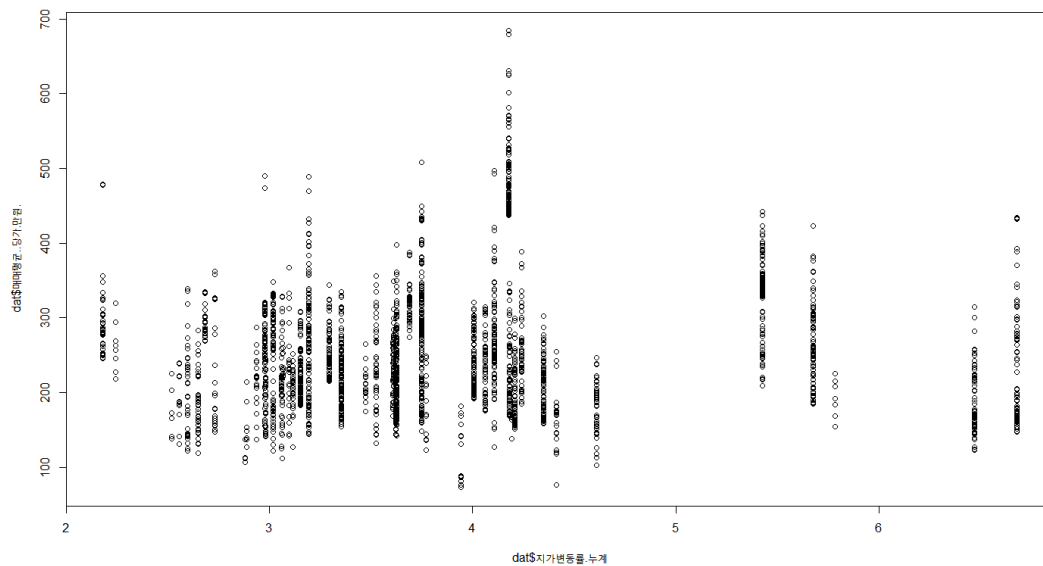


- Mosaic plot of '스타벅스 수와 지하철역의 수' & '지하철역의 수와 쇼핑시설 수'

● 지가변동률, 지가변동률 ~ 매매평균가



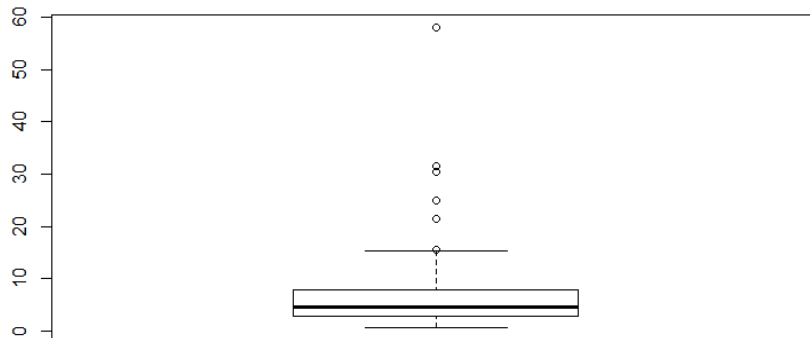
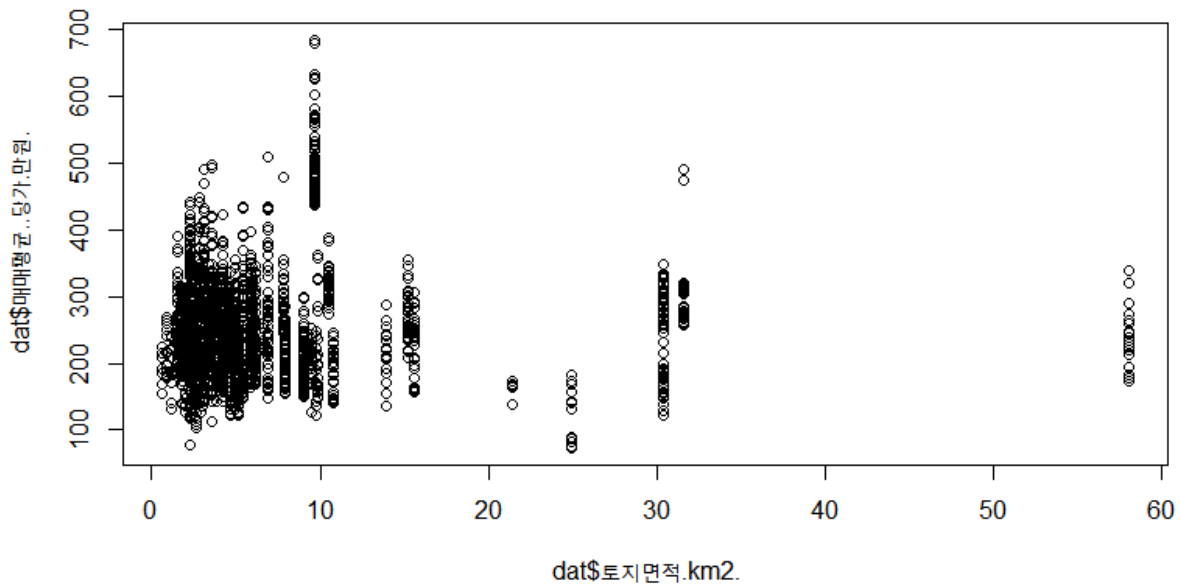
지가변동률~매매평균가 (산점도, boxplot)



제1사분위수가 3.194, 중앙값이 3.689, 제3사분위수가 4.182이다. 사분위수에서 알 수 있듯이 대체로 지가변동률 누계값 3~4사이에 많이 분포하는 것으로 보인다. 5.6이상 되는 outlier가 4개 정도의 동에서 나타나는 것으로 보

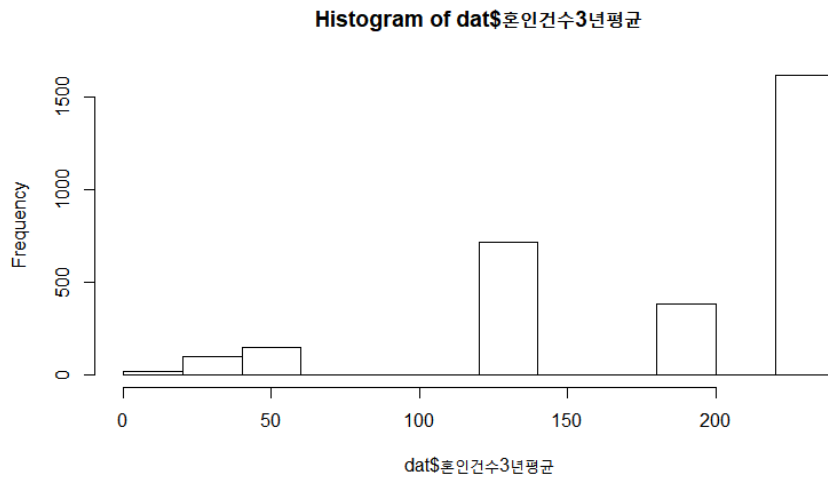
인다. 의외로 자가변동률이 가장 높은 동에서의 매매평균가는 그리 높지 않은 것을 알 수 있다. 또한 자가변동률이 가장 낮은 동에서 매매가격의 최저가 200 후반으로 꽤 높은 것으로 보아 안정된 입지를 가지고 있는 것으로 예측할 수 있다. 산점도의 자가변동률 4.3에서 유독 최고가의 매매평균가를 가진 동이 나타난다.

- 매매가격~토지면적(산점도,boxplot)

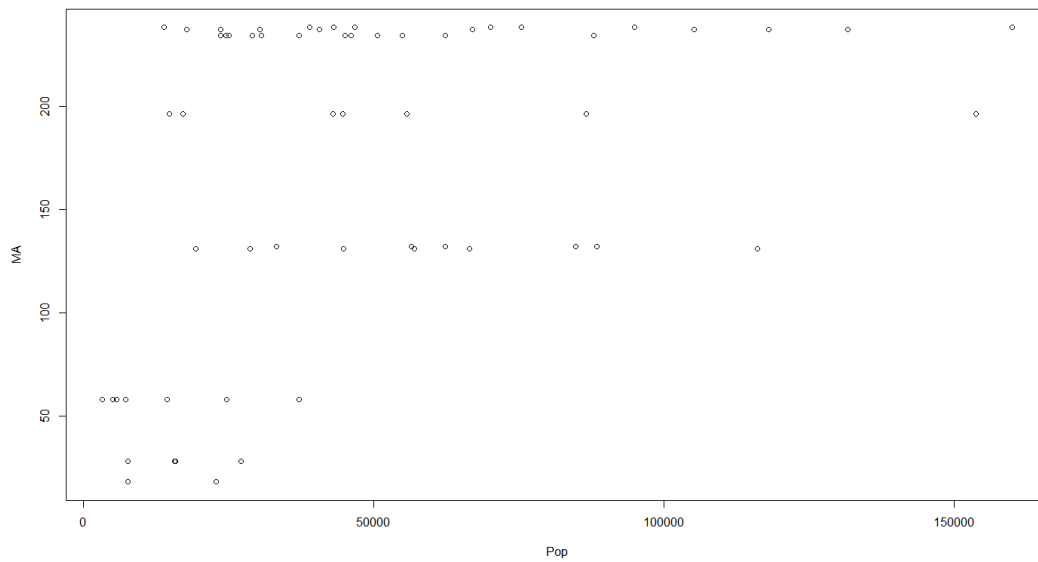


토지면적의 제1사분위수는 2.79, 중앙값은 2.79, 제3사분위수는 4.55이고 범위내 최대 관측값 15.23다. 왼쪽으로 굉장히 많이 쏠린 것으로 보아 최대치와의 차이는 굉장히 크다고 볼 수있다. 위 자가변동률에서 본 것과 같이 토지면적 10에서 다른 것과는 비교될 만큼 높은 가격대를 보이는 동이 존재함을 알수있다.

- 혼인건수, 평균혼인건수 ~ 인구수

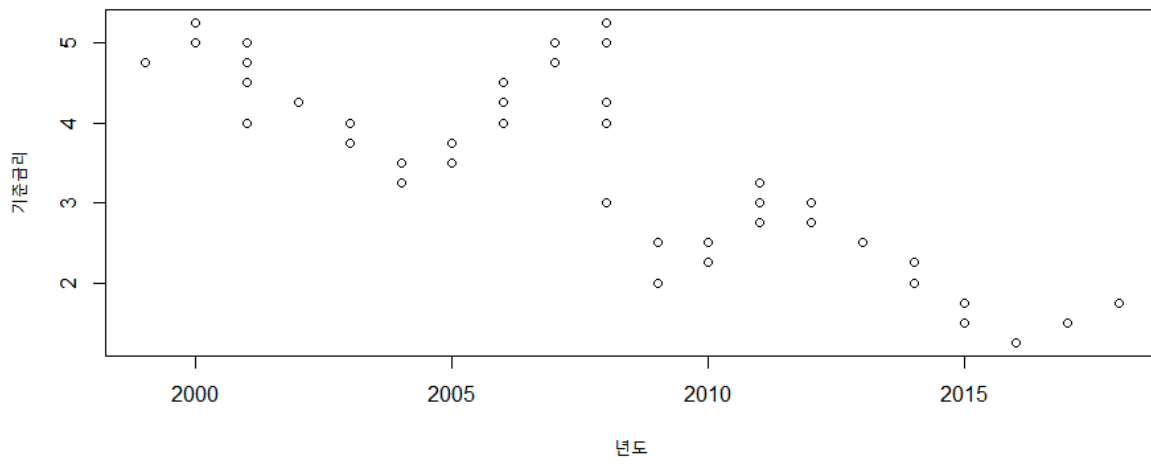


구별 평균혼인건수~인구수(산점도)



인구수가 적은 곳에서 혼인건수(100건 이하)가 적은 것을 알 수 있다.

● 기준금리 흐름 (1999년도~2018년도)



- 전체적인 추세가 감소하는 형태임을 확인할 수 있다.