데이터 마이닝 7조 중간 보고서

개체이(가) 표시된 사진

자동 생성된 설명

12132591 정동호

12131820 이건도

12161890 하나영

<목차>

1. 주제
2. EDA
3. **EDA 결과**
   1. 인천광역시 아파트 제곱미터당 평균 매매 가격 변동

여기서 인천광역시 아파트 제곱미터당 평균 매매 가격은 부동산114에서 얻은 인천에 있는 모든 아파트를 대상으로한 가격이다.

시계열 데이터이고 기간은 2016년 1월부터 2018년 12월의 자료를 이용해서 시계열 그래프와 분석을 진행했다. 인천광역시 전체 아파트 제곱미터당 평균 매매가격 변화는 다음과 같다.

지도, 텍스트이(가) 표시된 사진

자동 생성된 설명

전반적으로 지속적으로 아파트 제곱미터당 평균 매매 가격은 상승한다는 것을 확인할 수 있다. 2016년 1월부터 2018년 12월까지 제곱미터당 25만원 정도 상승한 것을 확인할 수 있다.

이 변동이 향후 미래에는 어떻게 변화할 지에 대해 시계열 분석을 이용하여 시계열 예측과 실제 2019년 1월부터 3월까지의 변동을 비교해보도록 한다.

정상성의 만족을 하는지를 확인하기 위해 ACF와 PACF 그래프를 통해 확인했다.

스크린샷이(가) 표시된 사진

자동 생성된 설명

ACF의 경우, 지수적으로 급격하게 감소하지 않고, 일정한 수준으로 감소하는 모양세이다. 정상성을 만족한다고 볼 수 없다.

스크린샷이(가) 표시된 사진

자동 생성된 설명

PACF를 보면, lag 1에서 절단값을 보이지만, ACF가 정상성을 만족하지 않으므로 1차 차분을 진행한다.

스크린샷이(가) 표시된 사진

자동 생성된 설명

1차 차분 진행 후, ACF를 보면 lag 3이후로 0에 가깝다고 볼 수 있으므로 lag 3 이후로 절단값을 갖고, 또한 정상성을 만족하는 것으로 보인다. lag 3이후로 절단값을 가지므로 MA(3)이라 판단할 수 있다.

스크린샷이(가) 표시된 사진

자동 생성된 설명

또한 1차 차분 PACF를 보면 lag 2에서 절단값을 가지므로 정상성도 만족하고 AR(2)로 판단할 수 있다.

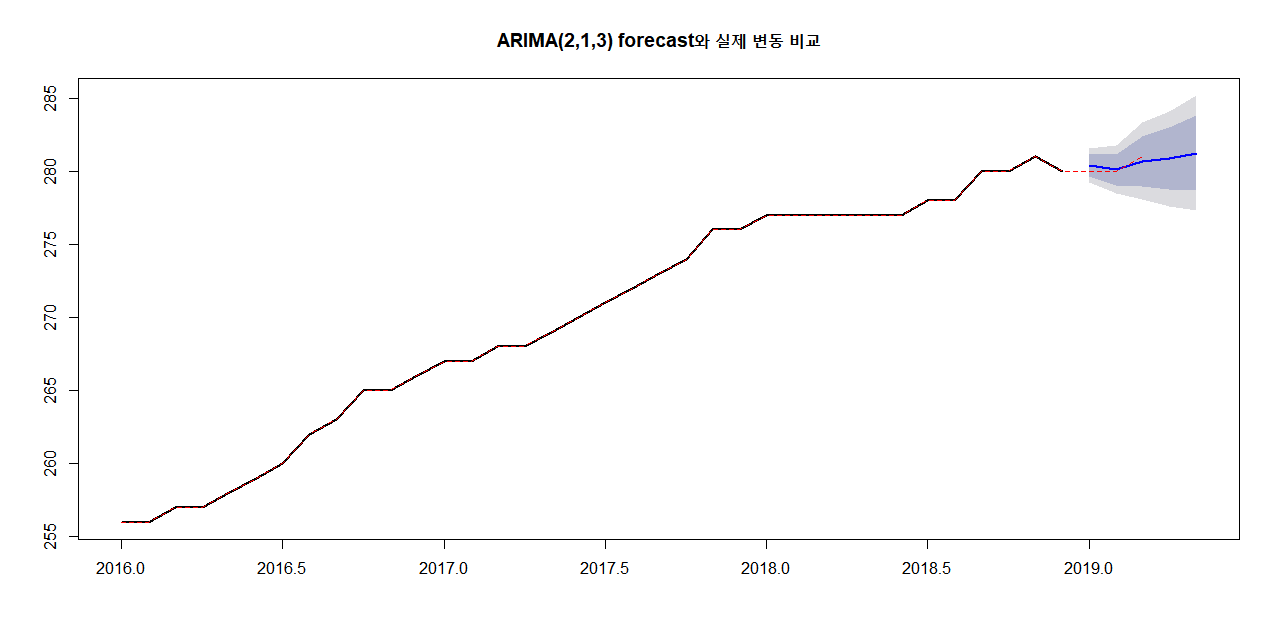
이러한 결과를 토대로 ARIMA 모형을 생각하면, AR(2), 1차차분 진행, MA(3) 이므로 ARIMA(2,1,3)으로 생각해볼 수 있다.

ARIMA(2,1,3)으로 적합한 결과를 보면 다음과 같다.

지도이(가) 표시된 사진

자동 생성된 설명

상승세를 보인다는 것을 예측할 수 있고, 실제 2019년 1월부터 2019년 3월의 변동 데이터와 비교를 해보면



빨간색 점선이 실제 데이터 변동이고, 그림상 어느정도 비슷하다는 것을 확인할 수 있다.

* 1. 인천광역시 구군별 제곱미터당 매매가격 변화

스크린샷이(가) 표시된 사진

자동 생성된 설명

연수구가 독보적으로 인천광역시 중에서 제일 높고, 강화군이 제일 낮다는 것을 확인할 수 있다. 미추홀구의 경우, 2016년 1월에는 7위로 하위권이었지만, 꾸준히 증가하여 중구와 계양구와 비슷한 수준으로 올라간 것을 확인할 수 있다. 꾸준히 증가할 수 있었던 요인은 인하대학교 주변 아파트 단지가 집값 상승에 영향을 주었다고 판단할 수 있다. 전체적으로는 모든 구군에서 조금씩 상승하는 것을 확인할 수 있고, 독보적으로 연수구는 크게 상승한다는 것을 확인할 수 있다. 인천국제도시라는 송도의 프리미엄 입지와 대규모 쇼핑시설, 아파트 단지 등으로 인해 짐작할 수 있다.

각 행정구역별로 세부적인 제곱미터당 평균 매매가 변동을 보면

텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

모든 행정구역이 전체적으로 증가한 것을 확인할 수 있다. 다만 동구의 경우, 최고점 이후 감소했다는 것을 확인할 수 있다.

* 1. 인천광역시 거래규모별 아파트 제곱미터당 평균 매매 가격 변동

텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

전반적으로 증가 추세를 보이고 있고, 최상위권은 132.0~148.5이고, 최하위권은 66.0미만이다.

규모가 클수록 아파트의 가격이 높아지는 것은 아닌 것을 확인할 수 있으며, 중간 규모 수준의 아파트 가격대가 더 높은 것을 확인할 수 있다.

지도, 텍스트이(가) 표시된 사진

자동 생성된 설명

다음의 개별의 그래프를 보면, 모든 그래프가 상승세를 보이는 것을 확인할 수 있다.

* 1. 인천광역시 인구수 변화

2016년 1월부터 2018년 12월까지의 인천광역시 인구수 변화는 다음과 같다.

지도이(가) 표시된 사진

자동 생성된 설명

우선 인천광역시 인구수는 전체적으로 증가한 것을 알 수 있다. 또한 이러한 상승세를 유지한다면 3백만명 돌파는 시간 문제이다. 또한 위의 인천 전체 아파트 제곱미터당 평균 매매가의 시계열 그래프와 같은 추세를 보이므로, 인구수 증가가 집값에 영향을 준다고 짐작할 수 있다.

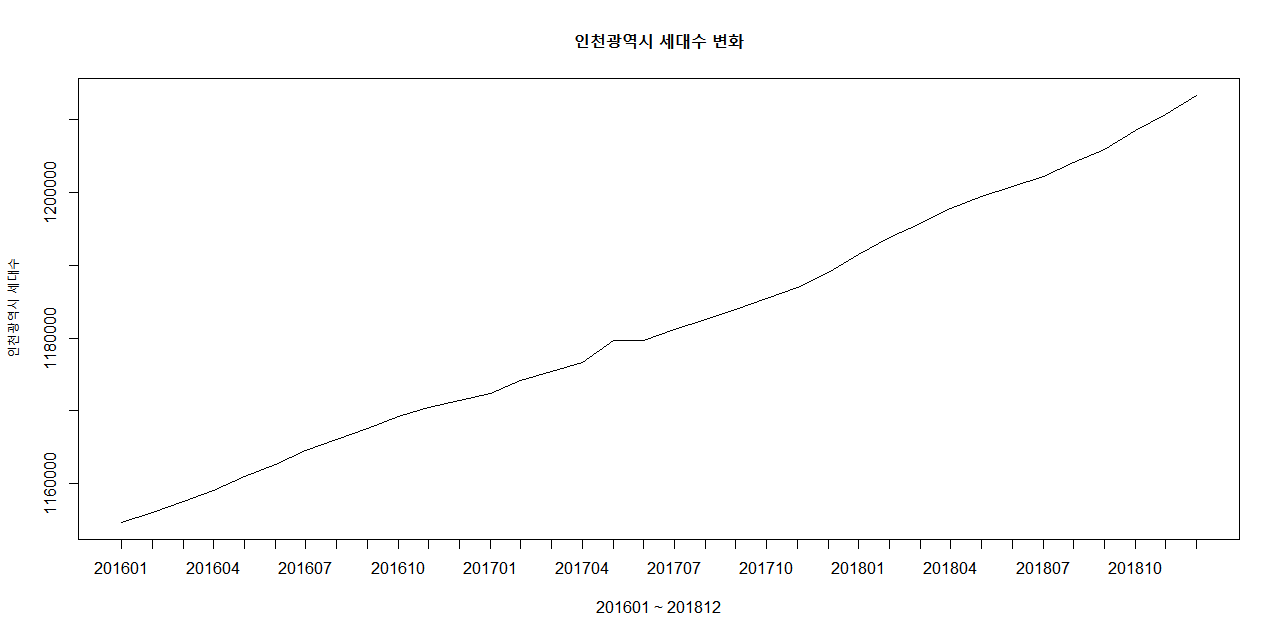
텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

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 최고점이후 감소세를 보이고 있다. 옹진군은 전체적으로 증가했다가 감소세를 보이고 있다.

* 1. 인천광역시 세대수 변화

2016년 1월부터 2018년 12월 세대수는 인구수 변화와 유사하다

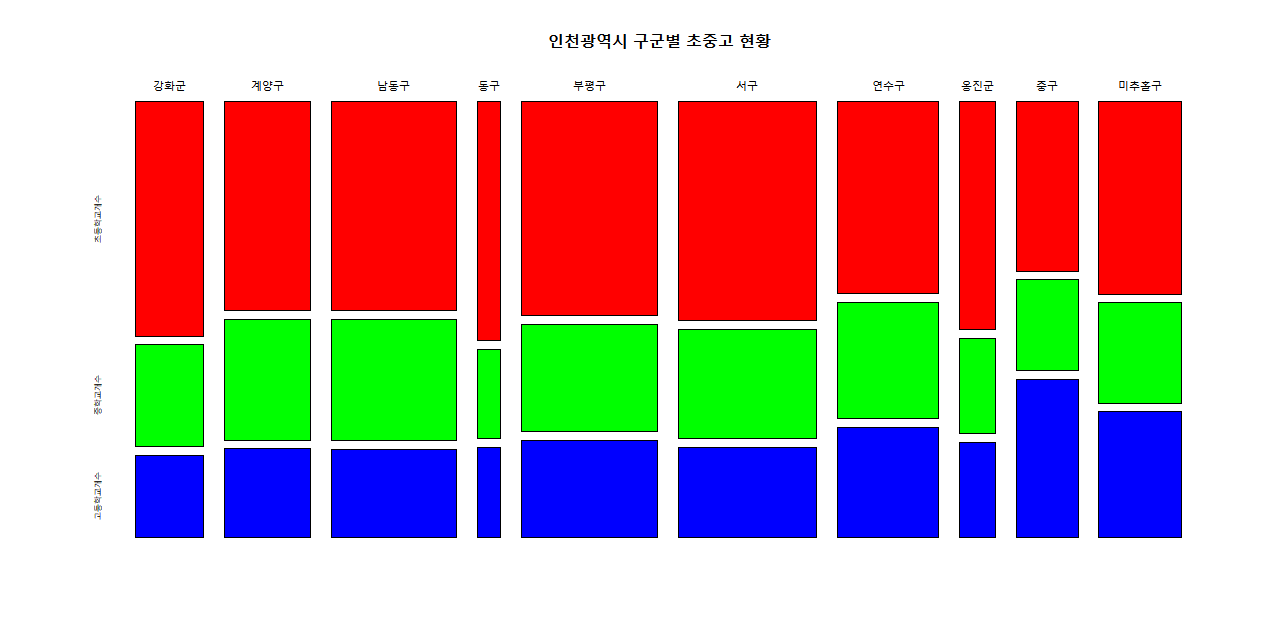


텍스트, 지도이(가) 표시된 사진

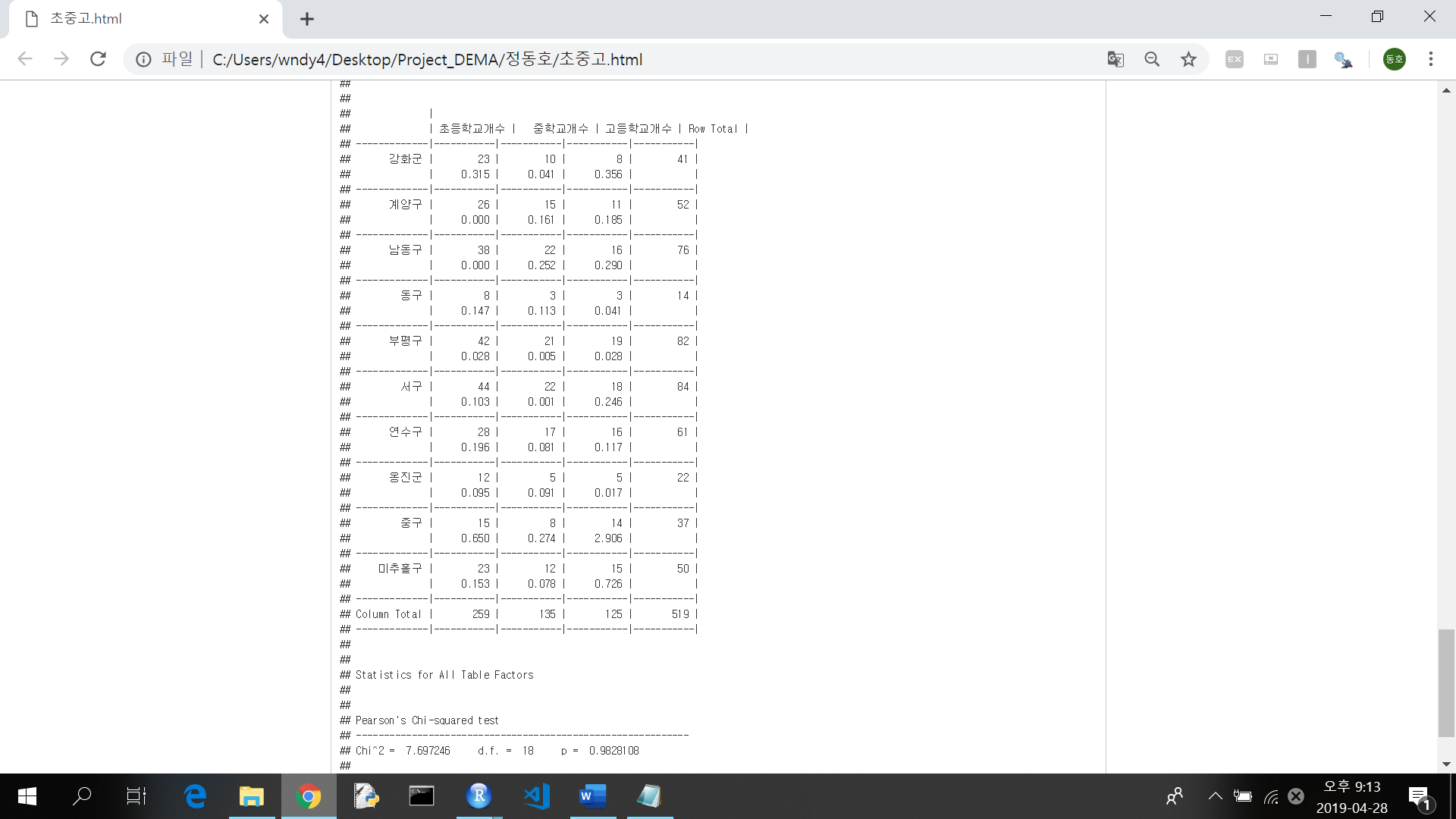
자동 생성된 설명

개별로 보아도 인구수와 유사한 추세를 보인다는 것을 확인할 수 있다.

* 1. 인천광역시 교육시설 수 현황

인천광역시에 있는 초등학교, 중학교, 고등학교 현황을 각 행정구역별로 구분한 그래프이다.빨간색(초등학교 수), 녹색(중학교 수), 파란색(고등학교 수)이고 모든 행정구역에서 초등학교 수의 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.

분할표를 보면



P-value가 0.05보다 크므로 교육시설 수와 인천광역시 구군 변수 간의 연관성이 없다고 볼 수 있다.

1. 인천아파트 제곱미터당 평균 매매가 vs. 인천 3년 평균인구수

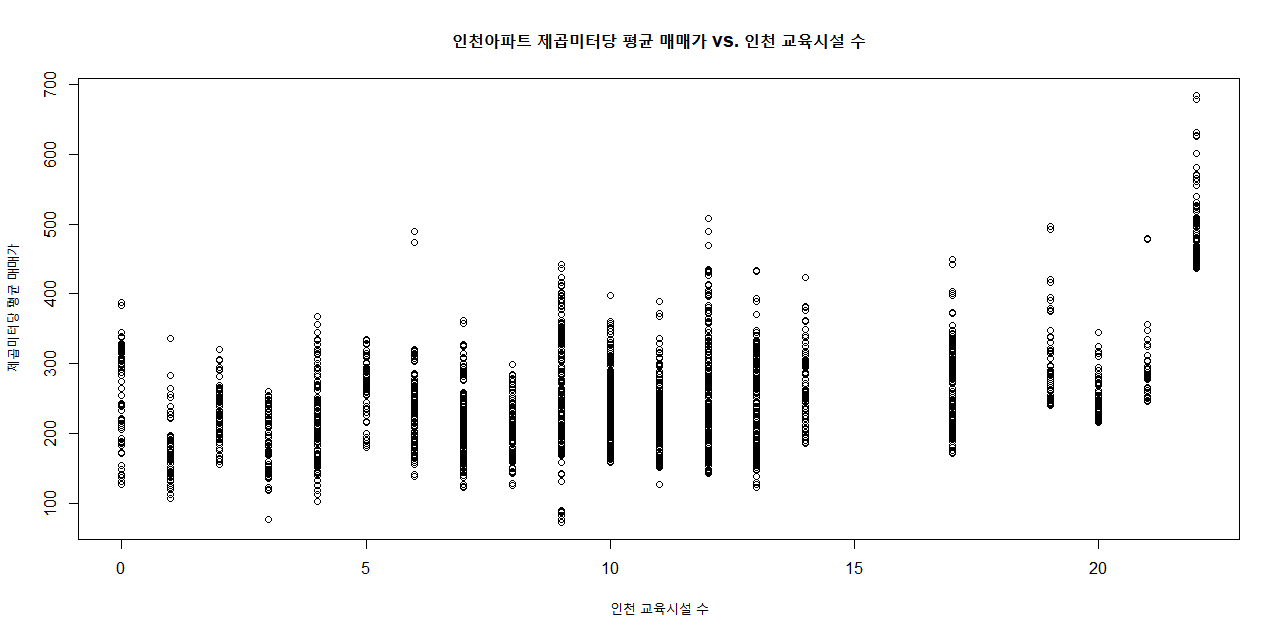
스크린샷이(가) 표시된 사진

자동 생성된 설명

전체적으로 3년 평균 인구수가 증가할수록 인천아파트 제곱미터당 평균 매매가도 증가하는 추세를 보인다고 할 수 있다. Spearman 상관관계를 통해 분석하면, 정도로 어느정도 양의 상관관계를 보인다고 할 수 있다. 특정 3년 평균 인구수에서 집값의 가격대가 높은 것을 확인할 수 있는데 앞서 보인 송도동의 특징을 고려한다면 송도동의 가격대일 수 있다.

1. 인천아파트 제곱미터당 평균 매매가 vs. 인천 교육시설 수

텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

Spearman 상관관계 분석 결과, 로 어느정도 양의 상관관계를 보인다고 볼 수 있다. 인천아파트 제곱미터당 평균 매매가와 교육시설 수 관계로 알 수 있는 점은 교육시설 수가 무작정 많다고 해서 집값이 크게 상승하지 않는다는 점이다. 다만 교육시설 수가 9에서 13 사이의 경우, 같은 교육시설 수내에서 가격대가 넓게 분포되어 있는 것을 확인할 수 있다. 또한 교육시설 수가 22인 경우, 다른 경우에 비해 훨씬 높은 가격대가 형성되어 있다는 것을 확인할 수 있다. 22개의 교육시설 수를 가지고 있는 동은 송도동이다.

1. 강남구아파트 제곱미터당 평균 매매가격과 인천아파트 전체 제곱미터당 평균 매매가격 관계

인천광역시 전체 아파트 제곱미터당 평균 매매가에 서울에 대표적인 집값을 상징하는 강남구 아파트 제곱미터당 평균 매매가격과의 관계를 파악한다.

지도이(가) 표시된 사진

자동 생성된 설명

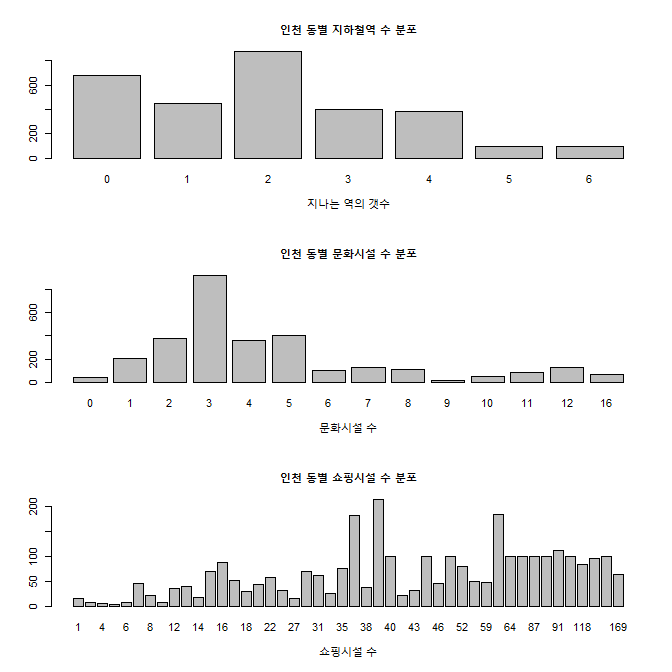
2016년 1월부터 2018년 12월까지의 강남구 전체 아파트 제곱미터당 평균 매매가격의 시계열 그래프이다. 전체적으로 가파른 성장세를 보이고 있다. 3년 동안 제곱미터당 약 500만원정도가 상승했다. 2018년 말에 들어서 약간 상승세가 감소세를 보이는 데 그것은 정부의 부동산 관련 규제로 인한 효과로 볼 수 있다.

지도, 스크린샷이(가) 표시된 사진

자동 생성된 설명

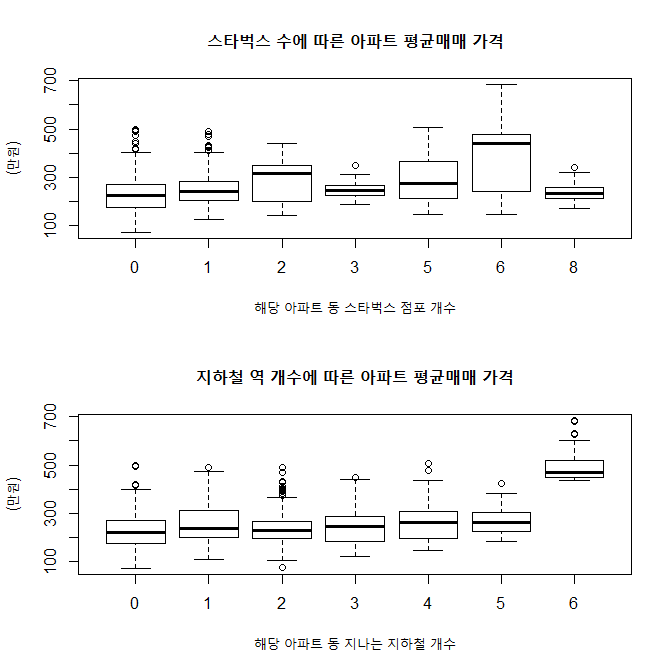
강남아파트 집값의 변동이 얼마나 인천 전체 아파트 집값과 상관이 있는지 파악하기 위해 산점도를 그렸다. Correlation은 0.9439로 매우 강하다. 또한 spearman correlatin에서 이므로 매우 강한 양의 상관관계를 보인다고 볼 수 있다.

1. 인천 동별 지하철역



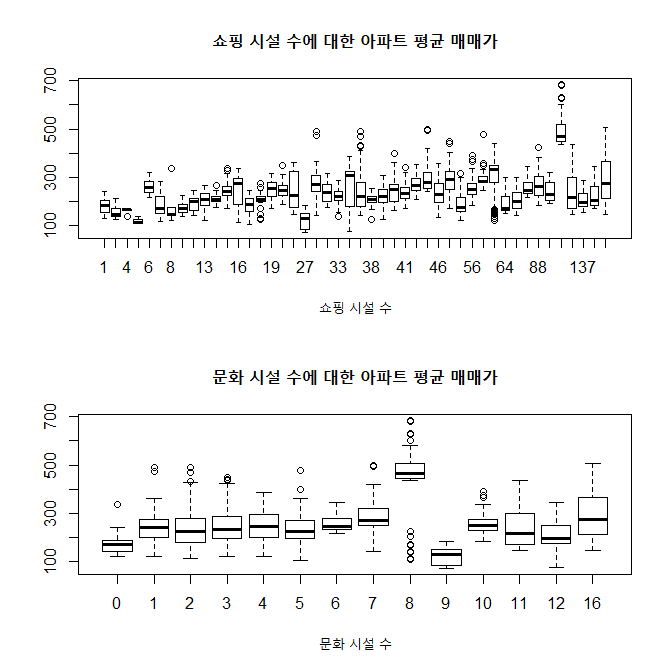
인천광역시 동별로 지하철 역, 문화시설 수, 쇼핑시설 수의 경우, 산정방법은 해당 동 범위에 지나가는 지하철 역의 수, 포함된 문화시설 수, 쇼핑시설 수를 계산한 것이다.

1. 인천아파트 제곱미터당 평균매매가 vs 스타벅스 수, 지하철 역 개수



첫번째 그림에서 스타벅스 8개점포의 동은 인천공항이 있는 중구 운서동이다. 대체적으로 스타벅스 수가 많다고 해서 인천아파트 제곱미터당 평균매매가가 급격하게 증가하는 것으로 보이지는 않는다. 지하철 역 개수에 따라 인천아파트 제곱미터당 평균매매가는 0개에서 5개까지는 수준이 비슷하지만, 6개부터는 많이 증가한 것으로 보인다. 6개가 지나는 동은 모두 송도동이므로 단순히 지하철 역의 개수가 많다고 높은 것이 아니라 다른 요인이 있다는 것을 생각할 수 있다.

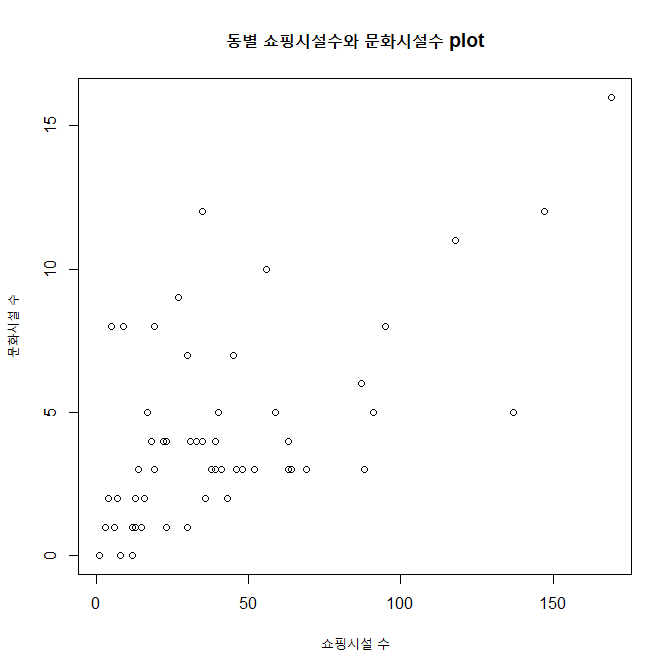
1. 인천아파트 제곱미터당 평균매매가 vs 쇼핑시설 수, 문화시설 수



쇼핑 시설 수는 해당 동에 분포하고 있는 대형마트, 백화점, 편의점, 마켓 등을 합산한 값이다. 쇼핑 시설 수가 많을수록 인천아파트 제곱미터당 평균 매매가격도 상승하는 추세를 보인다. 하지만 급격한 변화는 보인다고 볼 수 없으며, 쇼핑 시설이 95개인 경우, 다른 곳에 비해 가격대가 높게 형성되어 있는 것을 확인할 수 있다. 그곳은 바로 송도동이다.

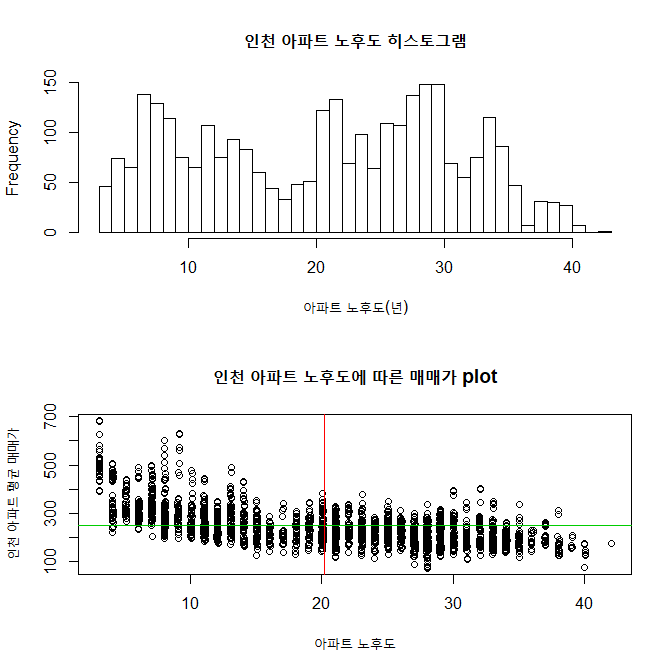
문화 시설 수는 해당 동에 분포하고 있는 영화관, 박물관, 공원 등을 합산한 값이다. 문화 시설 수가 많을수록 인천아파트 제곱미터당 평균 매매 가격이 증가한다고 볼 수 없다. 다만, 문화 시설 수가 8인 경우 가장 높은 가격대를 형성했다. 문학동, 송도동 등이 있는데, 주로 송도동이 많이 포함되어 있다.

1. 쇼핑시설 수 vs 문화시설 수



인천 동별 쇼핑 시설 수에 대한 문화 시설 수 산점도이다. 대체적으로 우상향하는 분포를 보이고 있으며, 쇼핑 시설 수가 0에서 50사이에서 문화 시설 수가 많이 분포되어 있다는 것을 확인할 수 있다.

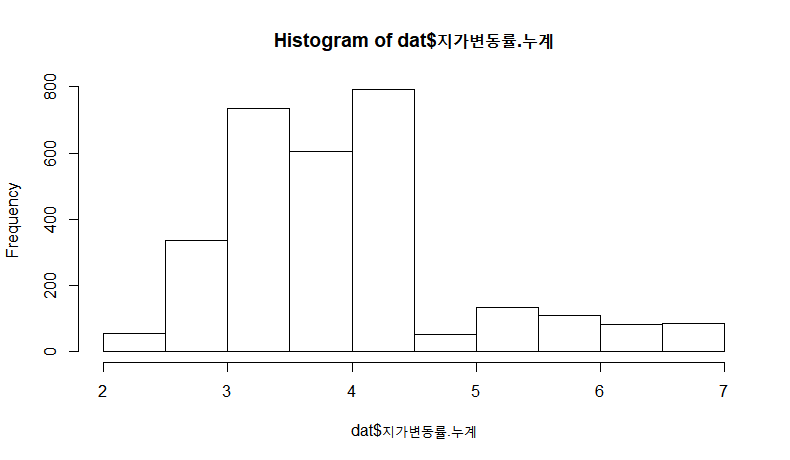
1. 인천 아파트 노후도 & 인천아파트 제곱미터당 평균 매매가 vs 인천아파트 노후도



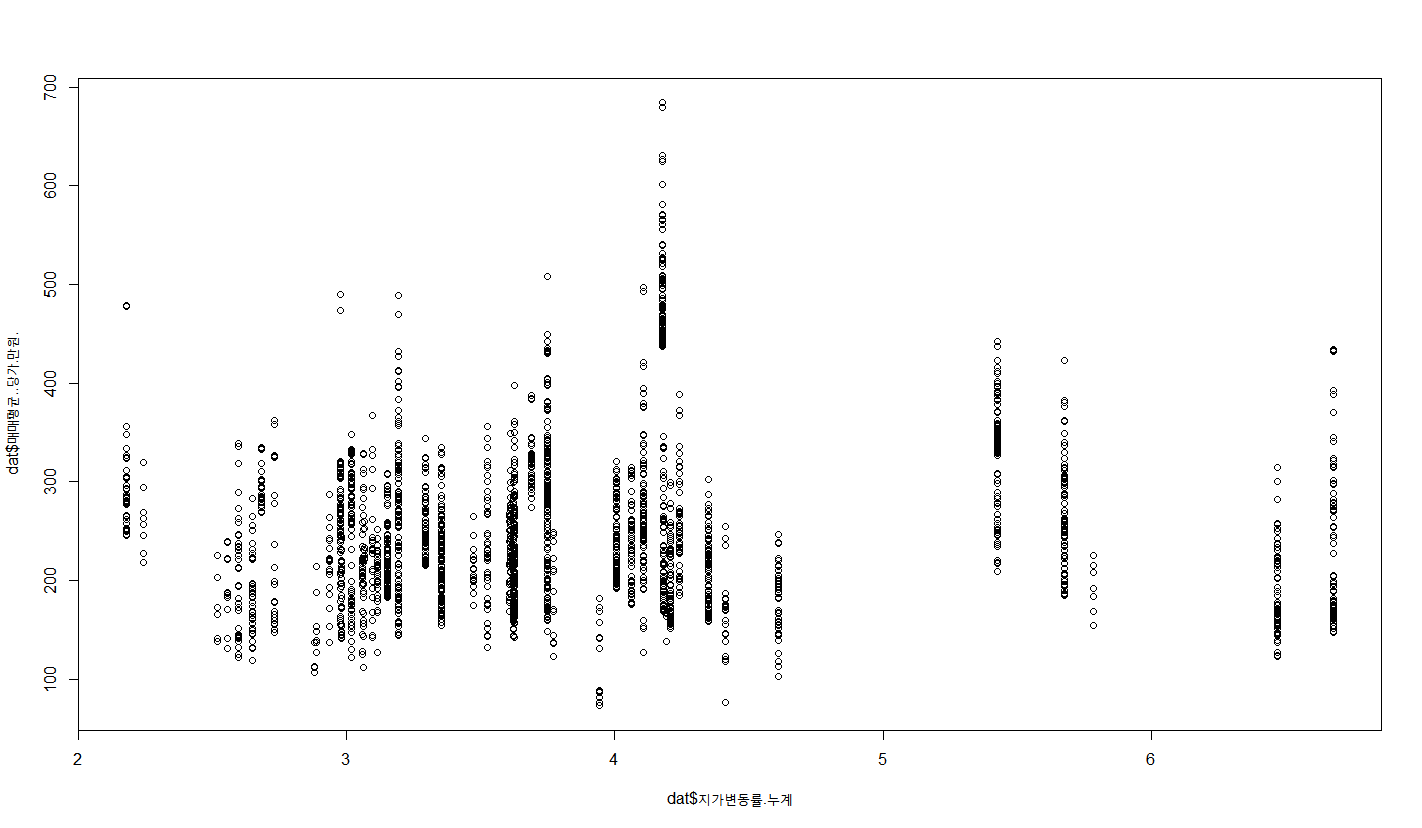
두번째 산점도에서 빨간색 선은 노후도 평균(약 20년)을 의미하고 녹색 선은 인천아파트 최근 3년 제곱미터당 평균 매매가의 평균값(약 249만원)이다.

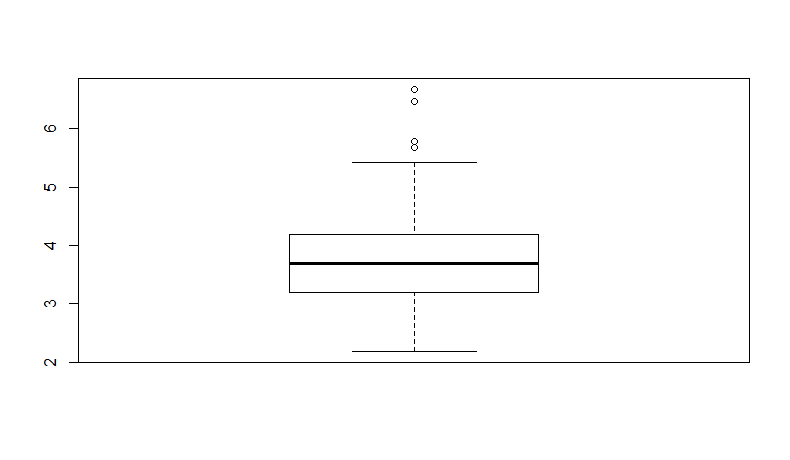
아파트 노후도에 따라 노후도가 높을수록 인천아파트 제곱미터당 평균 매매가는 낮아지는 추세를 확인할 수 있다.

1. 지가변동률과 인천아파트 제곱미터당 평균매매가 분석



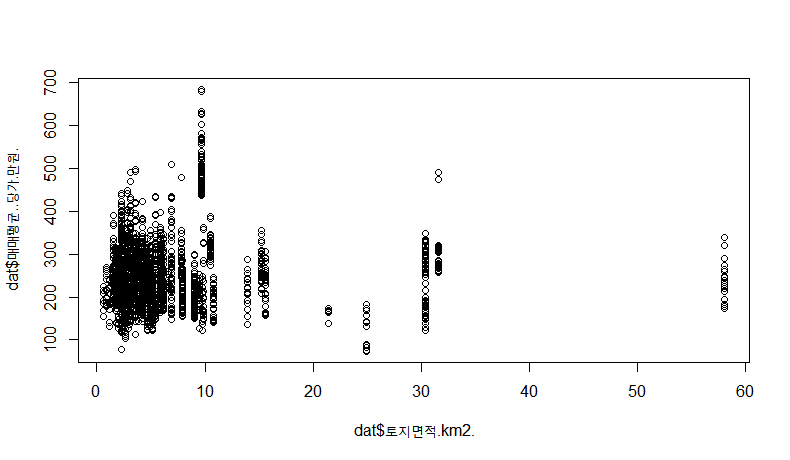
지가변동률~매매평균가 (산점도, boxplot)

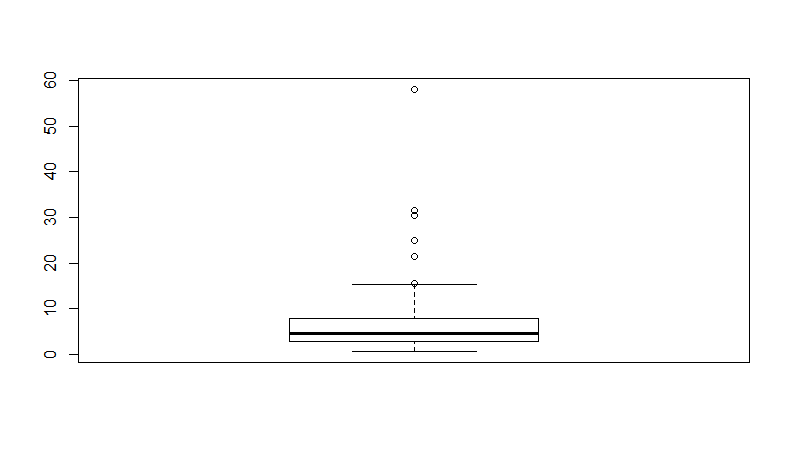




제1사분위수가 3.194, 중앙값이 3.689, 제3사분위수가 4.182이다. 사분위수에서 알 수 있듯이 대체로 지가변동률 누계값 3~4사이에 많이 분포하는 것으로 보인다. 5.6이상 되는 outlier가 4개 정도의 동에서 나타나는 것으로 보인다. 의외로 지가변동률이 가장 높은 동에서의 제곱미터당 평균 매매가는 그리 높지 않은 것을 알 수 있다. 또한 지가변동률이 가장 낮은 동에서 제곱미터당 평균 매매가격의 최저가 200후반으로 꽤 높은 것으로 보아 안정된 입지를 가지고 있는 것으로 예측할 수 있다. 산점도의 지가변동률 4.3에서 유독 최고가의 제곱미터당 평균매매가를 가진 동이 나타난다.

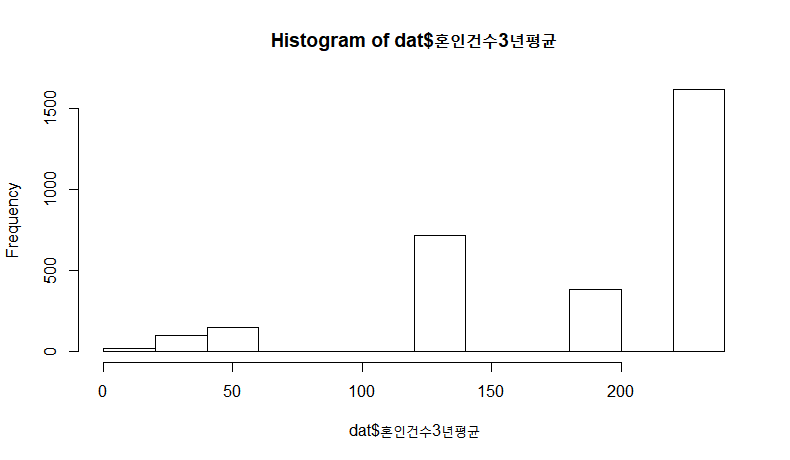
1. 인천아파트 제곱미터당 평균 매매가와 토지면적 분석



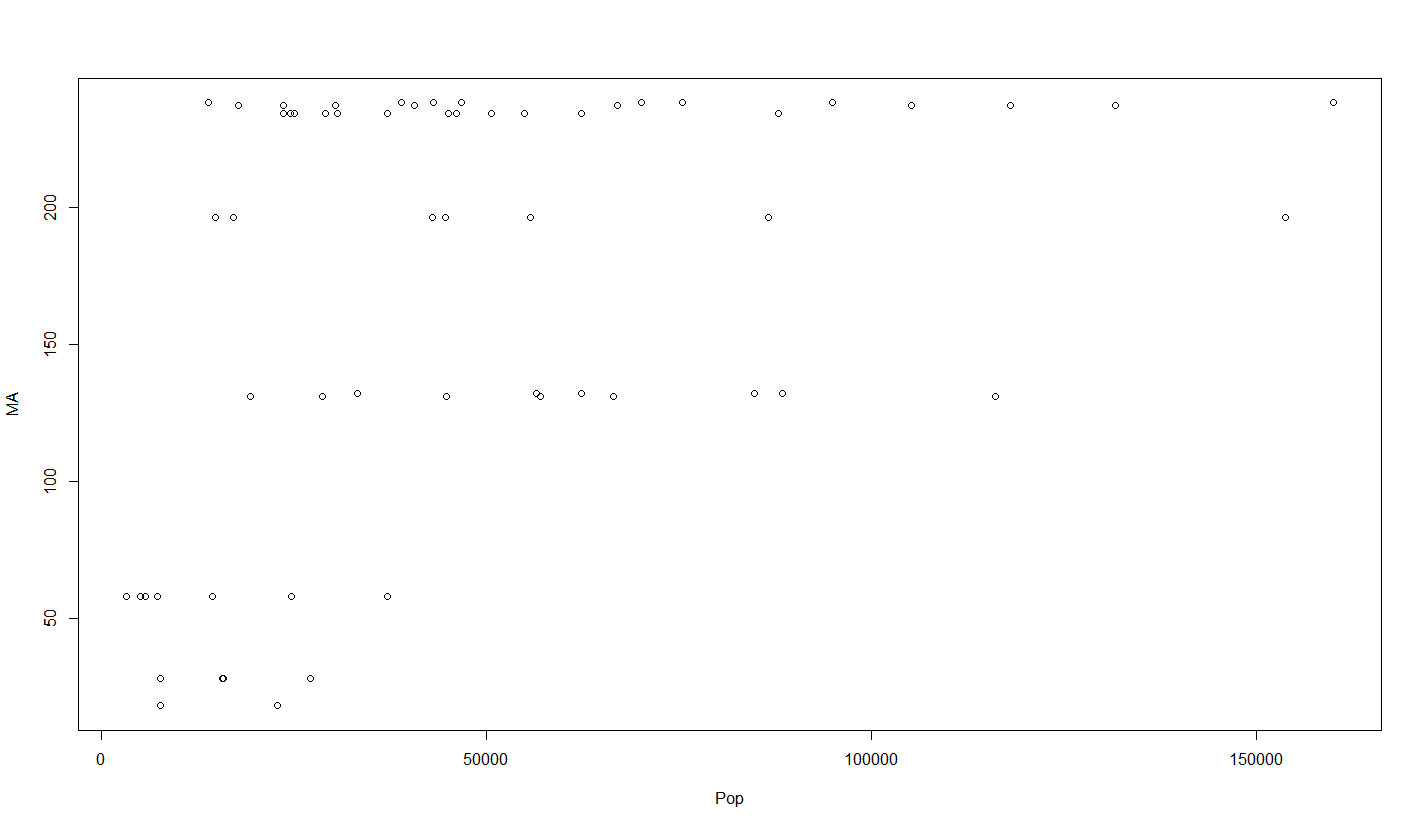


토지면적의 제1사분위수는 2.79, 중앙값은 2.79,제3사분위수는 4.55이고 범위내 최대 관측값 15.23다. 왼쪽으로 굉장히 많이 쏠린 것으로 보아 최대치와의 차이는 굉장히 크다고 볼 수 있다. 위 지가변동률에서 본 것과 같이 토지면적 10에서 다른 것과는 비교될 만큼 높은 가격대를 보이는 동이 존재함을 알 수 있다.

1. 3년 평균 혼인건수와 인구수 관련 분석

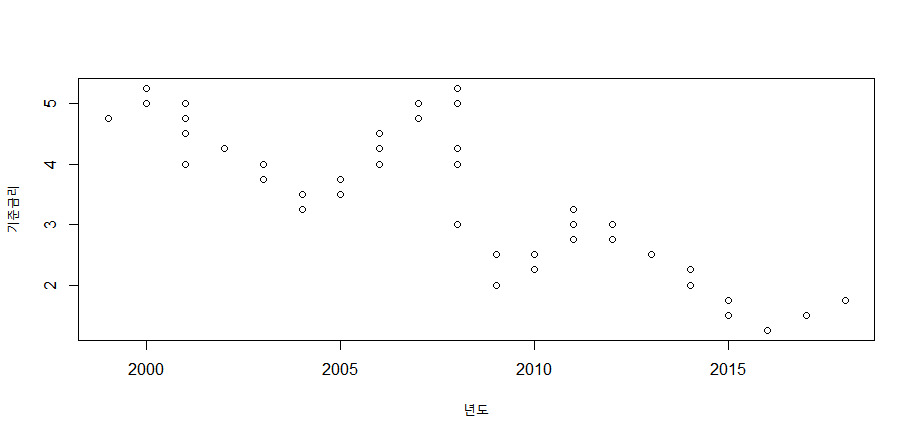


구별 평균혼인건수~인구수(산점도)



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

1. 기준금리 흐름 (1999년도~2018년도)



전체적인 추세가 감소하는 형태임을 확인할 수 있다.