

<분석방향>

- 전체적인 분석방향: 매도가격이 정해진 요인 분석하기 – 단지 전체적인 quality 또는 condition 으로 매겨지는 매도 가격이 아닌 소비자 입장에서 통계적으로 접근 (매도가격: SalePrice)
- 1. 지어진 년도 / 리모델링 년도
 - # 1: YearBuilt & Functional – 언제 지어졌는지, 하자의 정도
 - # 2: YearRemodAdd – 리모델링의 유무
- 2. 집의 전체 크기 - 방의 갯수 (화장실 불포함)
 - # 1: HouseStyle – 전체 층; 11stFlrSF / 2ndFlrSF / GrLivArea
 - # 2: LotFrontage & LotArea – 집 전체 적용 면적
- 3. 주차장의 유무
 - # 1: GarageArea – 주차장 크기와 매도가격의 관계
 - # 2: GarageQual – 주차장 상태와 매도가격의 관계

• 1. 지어진 년도 / 리모델링 년도

- # 1: YearBuilt & Functional – 언제 지어졌는지, 하자의 정도

- **가설1**: 최근에 지어진 집일 수록 매도의 가격이 높을 것이다.

- $y_{saleprice}$

$$= -2.530 * 10^6 + 1.375 * 10^3 * x_{YearBuilt}$$

- $R^2 = 0.2734$

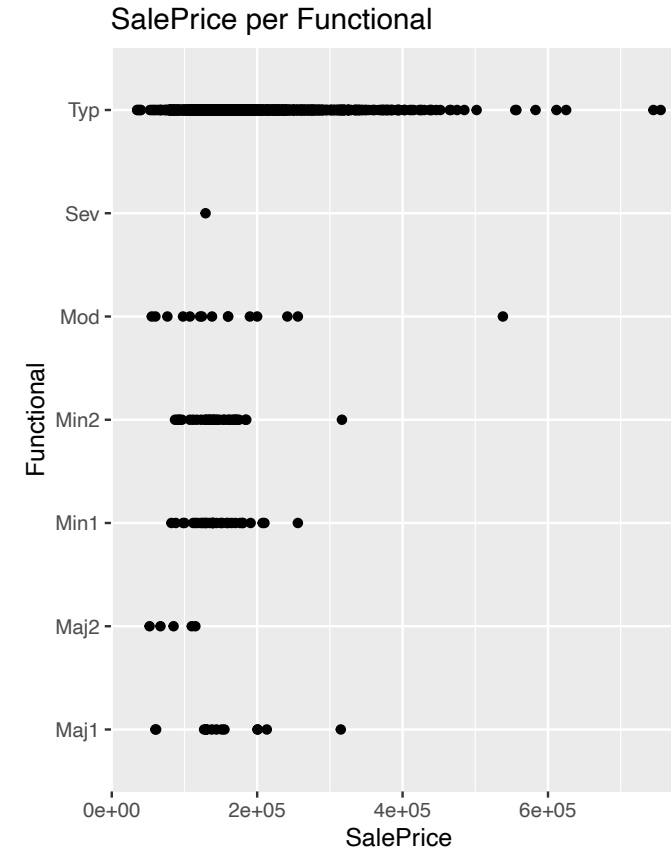
- $Adj R^2 = 0.2729$

- $F = 548.7$ (on 1 and 1458 DF)

- $p - value < 2.2 * 10^{-6}$

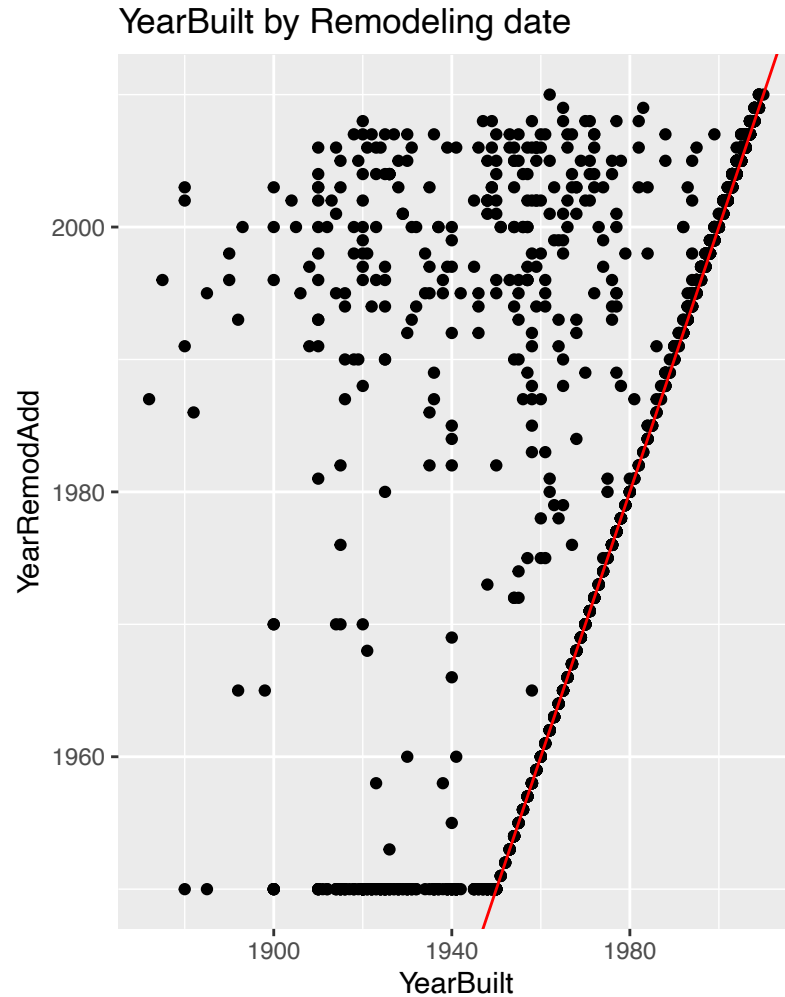
- 결론: 리그레션 결과 년도가 높아질수록 매도 가격이 높아짐을 알 수 있다.

- **가설2**: 하자의 정도가 낮은 집의 매도가격이 높을 것이다.



Minor damage 작은 하자가 있거나 Typical Functionality 하자가 거의 없는 주택이 매도가 되고 있음을 확인할 수 있음. Major damage 큰 하자가 있는 주택은 매도가격이 낮고 매도된 주택이 적음을 확인할 수 있음.

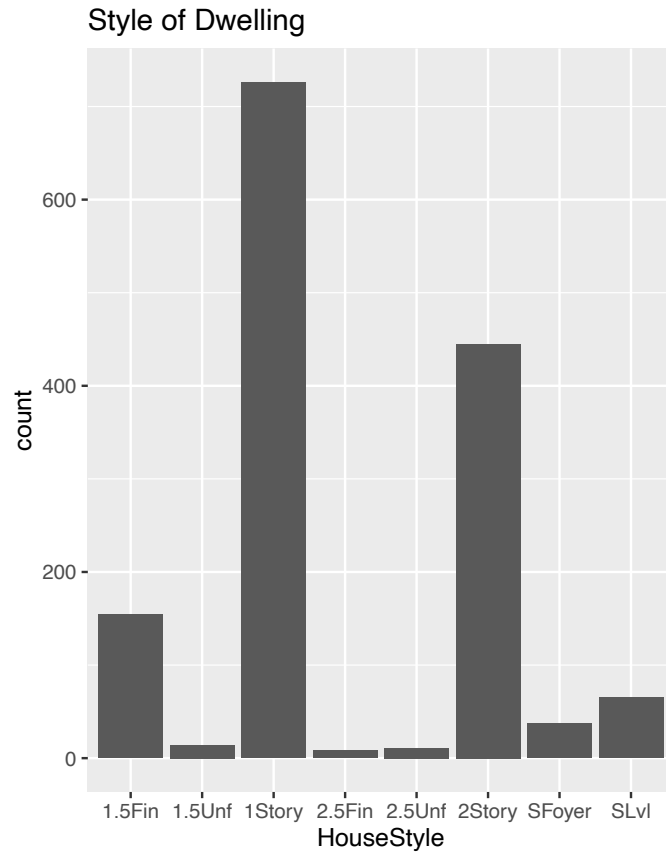
- # 2: YearRemodAdd – 리모델링의 유무
 - YearBuilt 와 YearRemodAdd 의 관계 파악



- red abline: 주택이 지어진 년도와 리모델링 년도가 같음을 나타냄
- 리모델링은 1980년도 이후에 나타남
- 리모델링이 되어진 주택이 단순히 오래된 집만 되어왔다고 결론 내릴 수 없음
- **결론: YearBuilt 와 YearRemodAdd 의 관계가 뚜렷하지 않음**

- 2. 집의 전체 크기 - 방의 갯수 (화장실 불포함)

- # 1: HouseStyle – 몇층인지

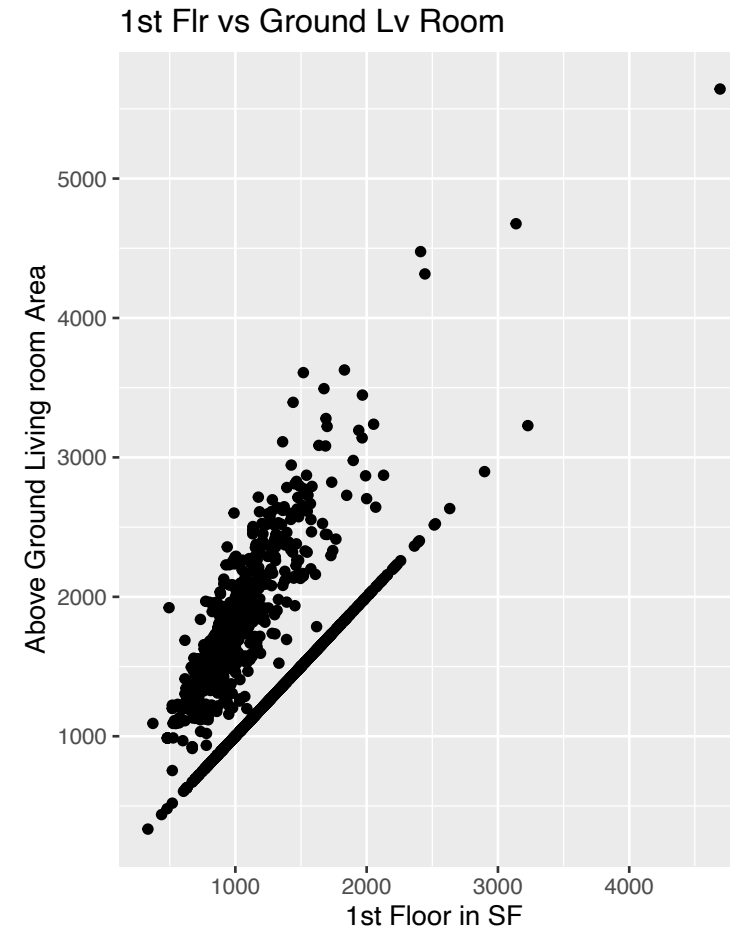


- 1층, 2층, 1.5층 순 – SalePrice 와 관계가 있는지?

- #2: 1stFlrSF / 2ndFlrSF / GrLivArea

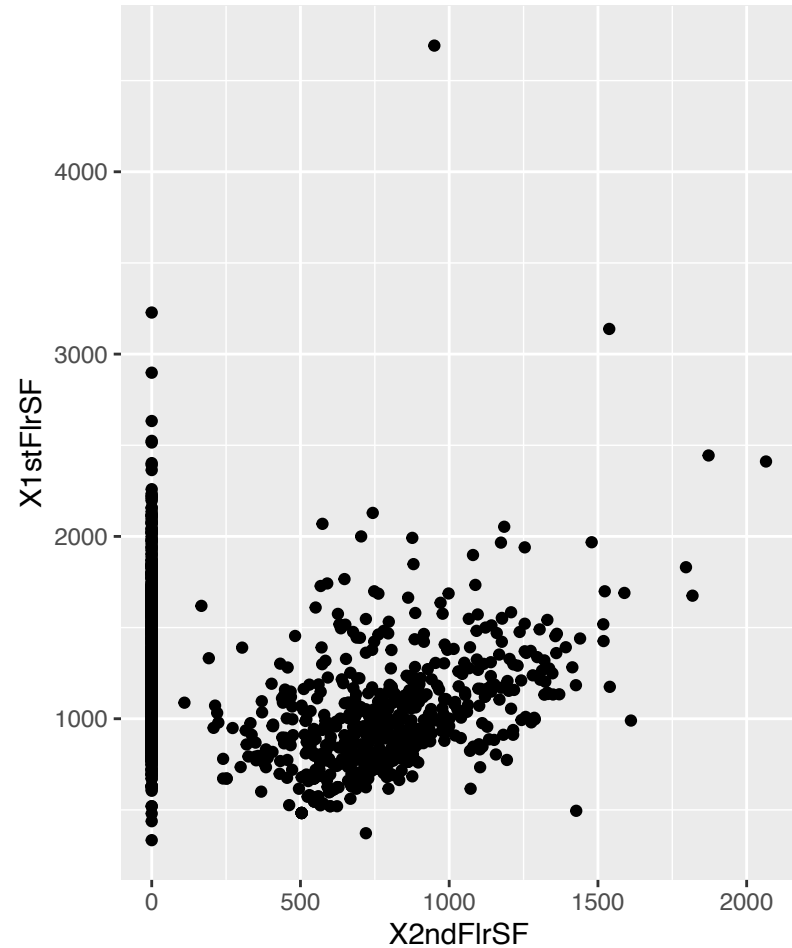
- 1. 1층의 크기와 거실의 크기 관계

(1stFlrSF vs. GrLivArea)



- #2: 1stFlrSF / 2ndFlrSF / GrLivArea 이어서
- 2. 1stFlrSF vs. 2ndFlrSF

2nd Floor vs 1st Floor in SF



- 1층과 2층을 비교해 본 결과, 1층의 크기와 상관없이 2층의 크기는 0인 경우가 많음

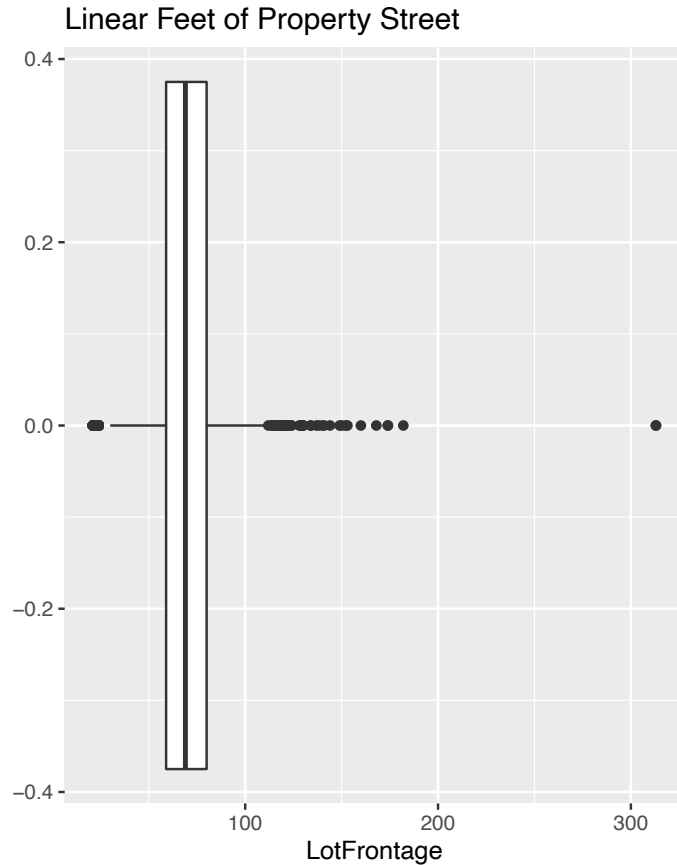
- 2층 통계량

Min.	Q.1	Median	Mean	Q.3	Max.	Stdev
0	0	0	347	728	2065	436.5284

- 표의 결과를 뒷받침하여 통계량을 비교해본 결과, 최솟값 과 중간값이 0이지만 평균이 347 인 것으로 보아 2층 집이 대다수임을 알 수 있음.

3: LotFrontage & LotArea – 집 전체 적용 면적

1. LotFrontage: Linear Feet of street connected to property



Outliers 가 너무 많음 - 부적절

2. LotArea: Lot Size in Square Feet 와 매도가격의 관계

➤ $Y_{\text{saleprice}} = 1.588 * 10^5 + 2.10 * X_{\text{LotArea}}$

➤ $R^2 = 0.06961$

➤ $\text{Adj } R^2 = 0.06898$

➤ $F = 109,1$ (on 1 and 1458 DF)

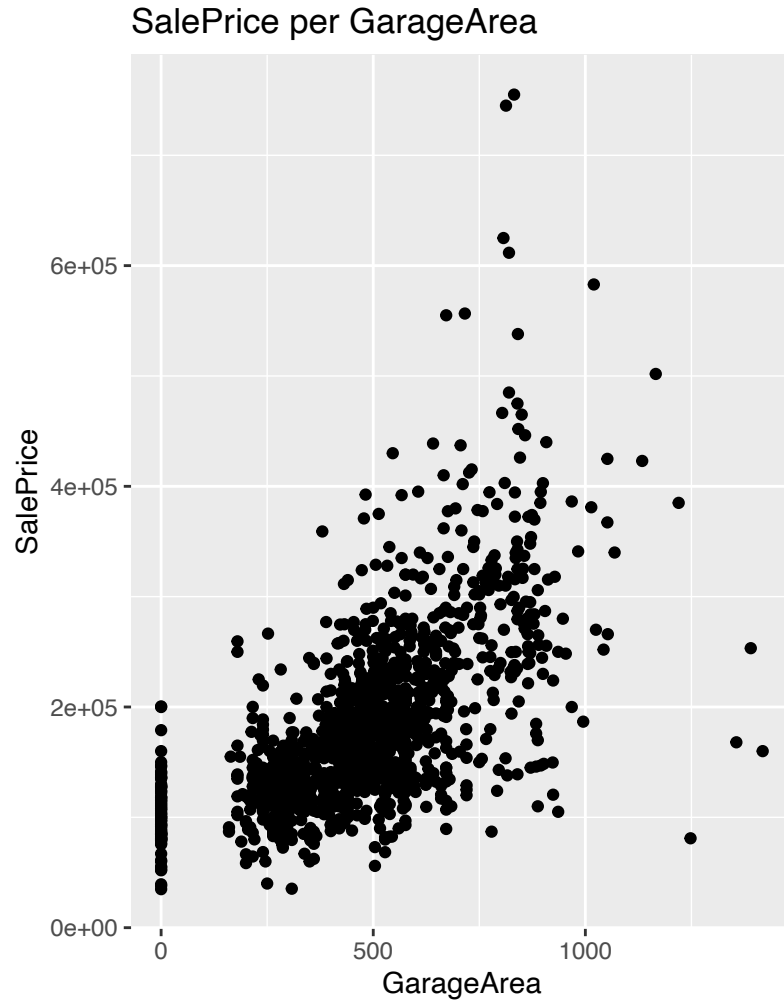
➤ $P\text{-value} < 2.2e-16$

결론:

집과 마당 전체 면적이 넓을 수록 매도가격이 높음을 알 수 있습니다. p-value 가 매우 작음. (상관관계가 있음)

• 3. 주차장의 유무와 크기

- #1: GarageArea 와 SalePrice 와의 관계



- 주차장의 크기가 커질수록 매도가격이 올라감을 확인할 수 있음
- 하지만 주차장의 크기가 존재하지 않음에도 매도가격이 존재함을 확인할 수 있음

➤ $y_{saleprice} = 71357.421 + 231.646 * x_{GarageArea}$

➤ $R^2 = 0.3887$

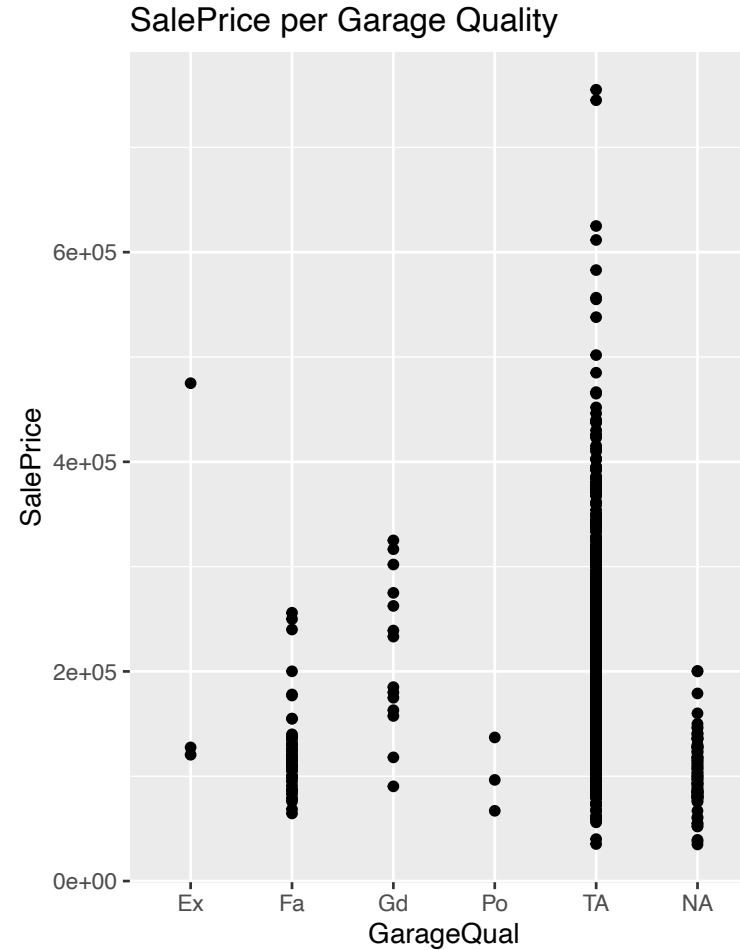
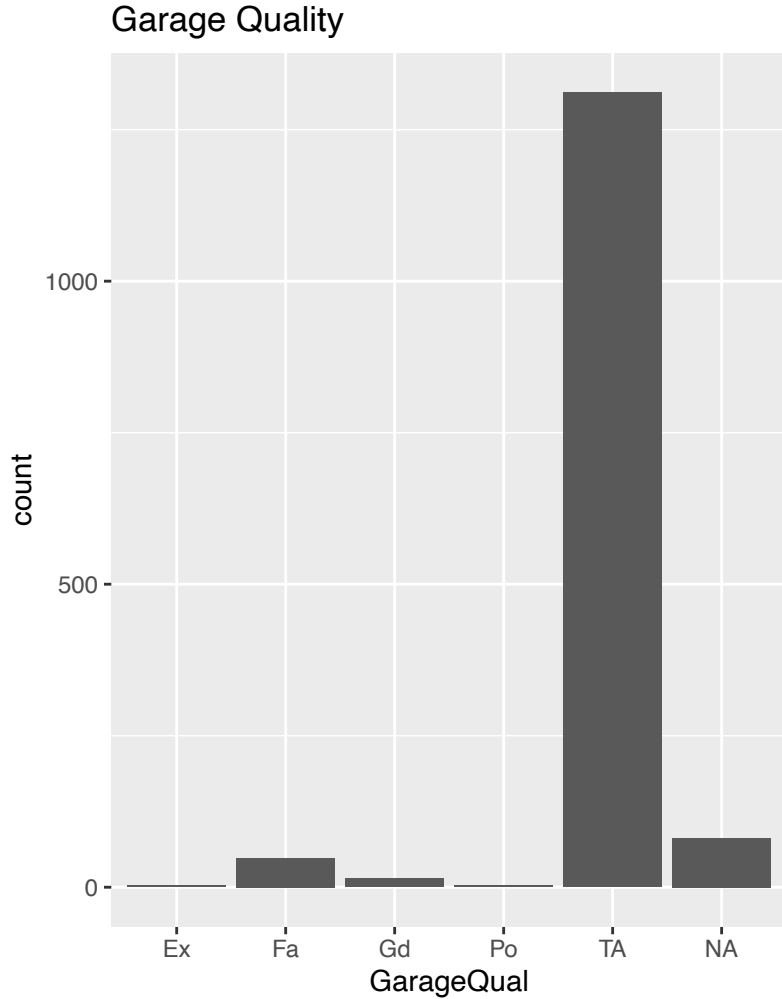
➤ $Adj R^2 = 0.3882$

➤ $F = 927$ (on 1 and 1458 df)

➤ $p - value < 2.2 * 10^{-16}$

- 리그레션 결과:
- 주차장의 크기가 커질수록 매도가격이 올라감

- 속성 2: GarageQual : 주차장 상태와 매도 가격의 관계



- 대부분의 주차장의 상태는 Typical / Average 보통임을 확인할 수 있음
- 주차장의 상태가 보통임에도 가격이 높음을 알 수 있음
- 하지만 주차장의 상태가 Excellent 최고임에도 매도 가격이 높지 않음을 확인할 수 있음. 또한 주차장이 NA 없음에도 매도가격이 낮지만 존재함을 확인가능
- 결론: 주차장의 유무와 상태가 소비자의 입장에서 집을 사고 또 매도가격이 정해짐에 큰 차이가 없음을 확인할 수 있음**

<결론>

***Bold 된 것 – 분석 가능한 가설**

- 1. 지어진 년도 / 리모델링 년도
 - # 1: **YearBuilt & Functional** – 언제 지어졌는지, 하자의 정도
 - # 2: YearRemodAdd – 리모델링의 유무
- 2. 집의 전체 크기 - 방의 갯수 (화장실 불포함)
 - # 1: **HouseStyle** – 전체 층; 11stFlrSF / 2ndFlrSF / GrLivArea
 - # 2: LotFrontage & **LotArea** – 집 전체 적용 면적
- 3. 주차장의 유무
 - # 1: **GarageArea** – 주차장 크기와 매도가격의 관계
 - # 2: GarageQual – 주차장 상태와 매도가격의 관계
- 앞으로 더 분석할 주제
 - 1. 매도가격과 집이 지어진년도, 하자의 유무가 통계적인 관계가 있다.
 - 2. 주차장의 크기와 매도가격이 통계적으로 관계가 있다.
 - 3. 주택의 적용면적과 층의 갯수, 거실의 크기, 집 전체 크기의 상관관계를 더 파악하고 매도가격과 관련이 있는지 소비자의 입장에서 분석