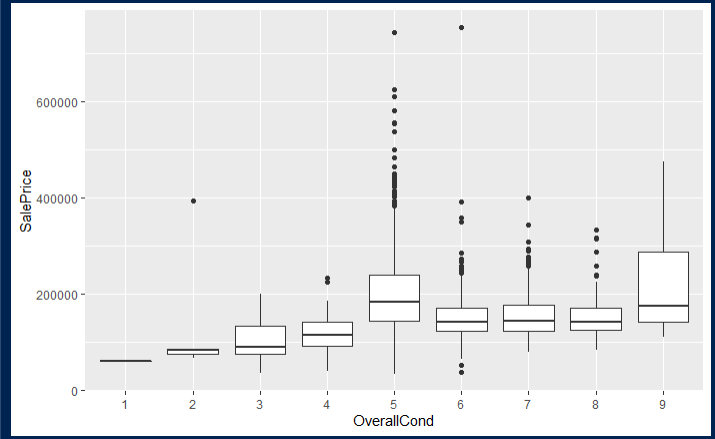
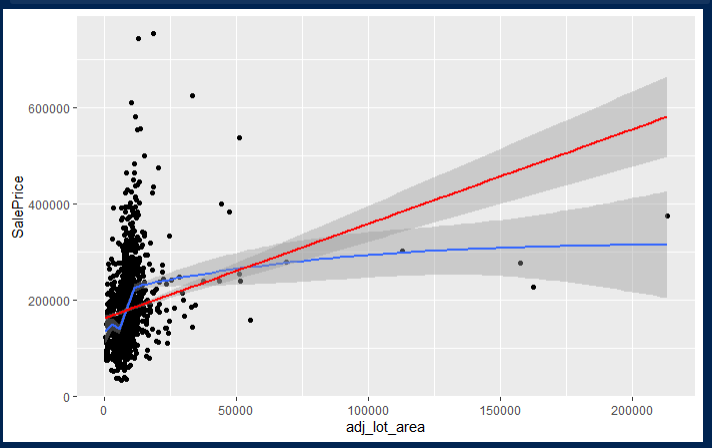
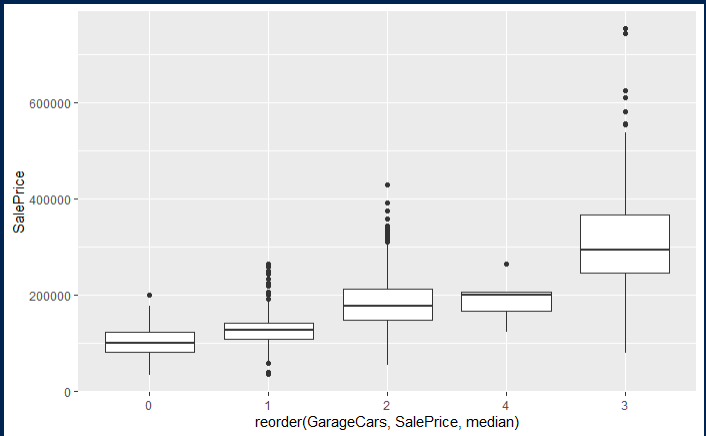
1. Overall Condition
   * Adjusted to be a binary for negative overall condition – 1 if it fell below a 5 rating and 0 if it was at a five or above.
   * This dummy feature is ***neg\_ovrll\_cond***
     1. The median value tended to be in the same general range for 5 and above and also generally the same for 1 and below
     2. Unlike Overall Quality field, the medican Sale Price did not linearly trend with the rating in Overall Condition – thus I opted to make into a dummy variable
     3. 
2. Lot Area
   * I think we may want to cap the “adjusted” lot area because there are a couple of points that have a lot of lot area even after taking out the outside sf and 1st floor – interested to see the effect on the model (e.g., adjusted lot area greater than 150,000 -> which are probably farms)
   * May be appropriate to bin these areas (e.g., 0 to 2000, 2000 to 50000, 50000 +) if we want to lessen the linear affect of having more lot area
   * 
3. Garage Cars
   * Introduced a dummy variable for more than two cars because 4 car garages have same median value as 2 car garage.
     1. Although there are only 5 houses with 4-car garages in the sample - so may be fine to leave as is b/c home price appears to linearly scale from 0 to 1 to 2 to 3.
   * Just wanted a different option in case
   * The dummy >= 2 car garage is ***two\_plus\_cr\_garg***. The original GarageCars is still in the column set of the pre\_processed data.



1. Building Type Binary re: Single Family homes
   * Changed this binary column to be named as sgl\_famly\_hm instead of BldgType
2. Home Age
   * There are two columns – one for remodel age (YrSold – YearRemodAdded) and one for homeage(YrSold – YrBuilt)
     1. Have both in there right now but will likely end up dropping one.