

# Assignment 6

Steven Spiegel

2023-09-29

## Assignment 6

Below are my answers to Assignment 6

Question 1: To start this assignment, let's take a look at the relationship between a home's overall condition (Overall\_Cond) and its sale price

Part A: Use ggplot to make a boxplot that illustrates the relationship between Sale\_Price and Overall\_Cond

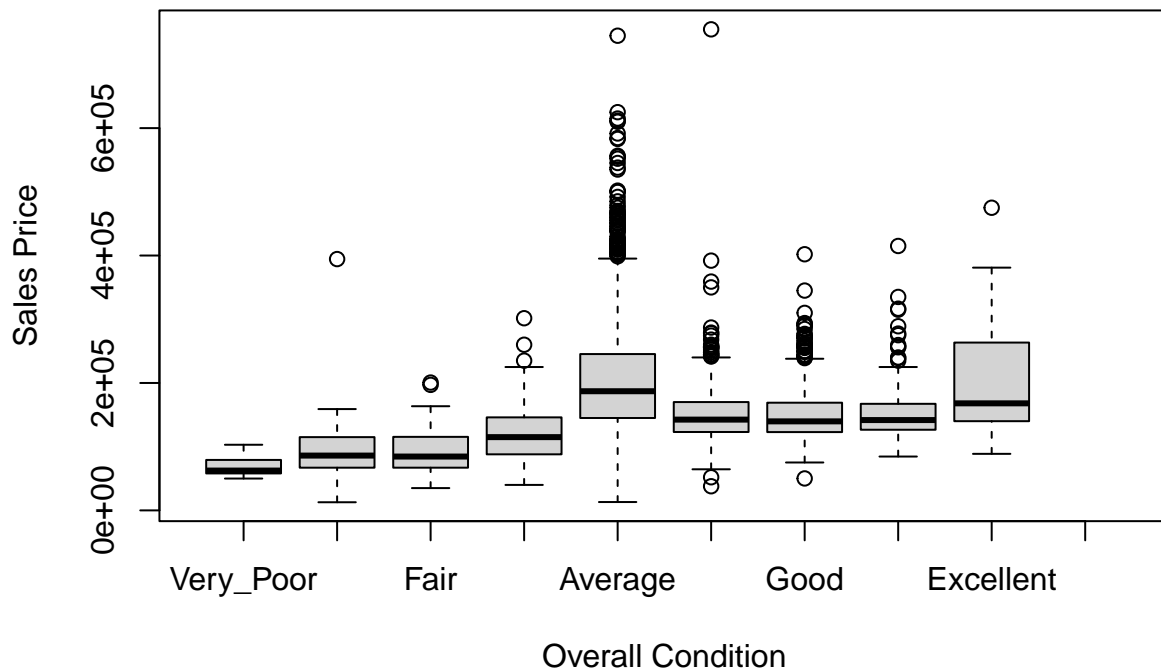
Response: see code below

```
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(AmesHousing)
ames <- make_ordinal_ames()
boxplot(ames$Sale_Price ~ ames$Overall_Cond, xlab = "Overall Condition", ylab = "Sales Price", main = "Overall Condition vs Sales Price")
```

## Condition vs Sales Price



Part B: Describe the general relationship between overall condition and sale price. How does this relationship break down for the “Average” category?

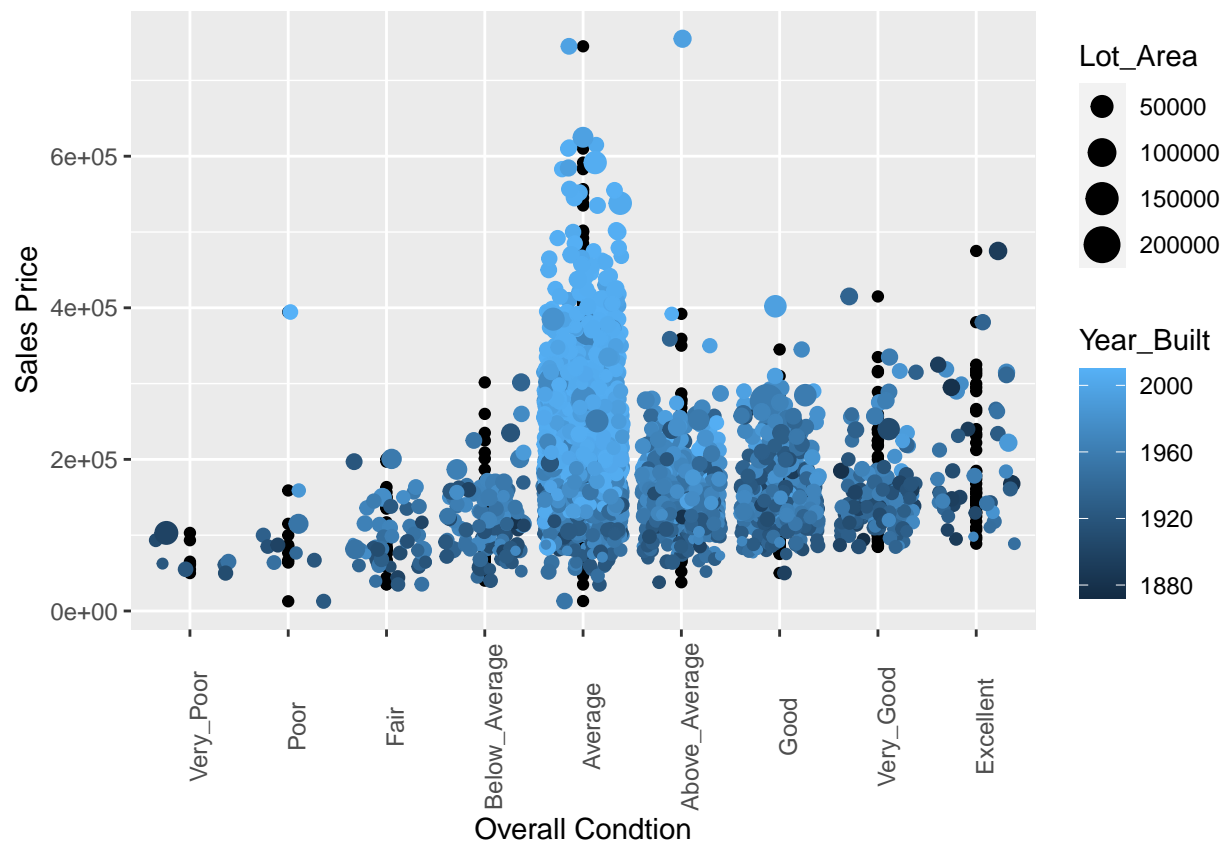
Response: The Sales price is steadily increasing from very poor to Excellent. Average seems to have a similar mean to that of “excellent” and higher than the “good” condition. It also has more outliers that seem to be pulling the median and mean sales price higher.

Question 2: Let’s explore what might be leading to the breaking of the pattern for the “Average” condition group?

Part A: It might be the case that the way home condition is recorded is not the same for older homes and newer homes. That could especially be the case if the ratings were not all done at the same time. On the other hand, perhaps some average-condition homes sell for more if they’re on bigger lots.

Response: See code below

```
ggplot(ames, aes(Overall_Cond, Sale_Price)) +
  geom_point() +
  # geom_jitter(width = 0.5, height = 0.5) +
  # theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  geom_jitter(aes(colour = Year_Built, size = Lot_Area)) +
  theme(axis.text.x = element_text(angle = 90)) +
  xlab("Overall Condition") +
  ylab("Sales Price")
```



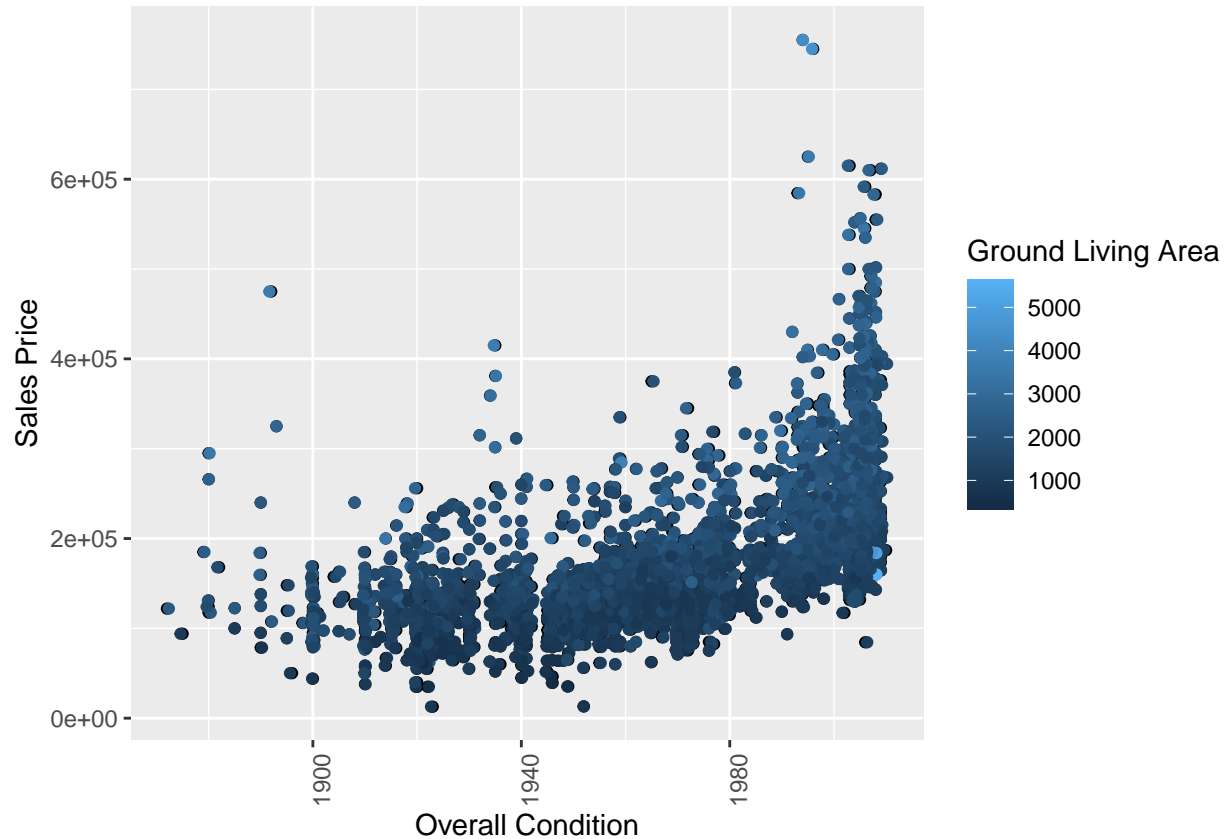
Part B: Briefly describe what your graph shows you about the relationship of Year\_Built and Lot\_Area to Sale\_Price, especially among the “average” condition homes. In other words, which seems to have the stronger relationship to Sale\_Price?

Response: Year built seems to have a stronger relationship to higher sales price, as well as lot area. Larger lots are more expensive in the *Average* Condition among the newer homes.

Part C: For this part, choose one other variable that you believe will be highly related to Sale\_Price, and create a graph that displays Sale\_Price, the variable you identified in (b), and the new explanatory variable you’ve identified in part (c). Then use ggplot to create a graph with appropriate mappings to display the relationship between all three variables. Make sure to pay attention to graphical detail, as you did in part (a).

Response: see code below

```
ggplot(ames, aes(Year_Built, Sale_Price)) +
  geom_point() +
  geom_jitter(aes(colour = Gr_Liv_Area)) +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(x = "Overall Condition", y = "Sales Price",
       colour = "Ground Living Area")
```



Part D: Now let's sum it up. Write a short paragraph that explains your hypothesis from part (c), describes what your graph showed about the relationship between the three variables, and concludes with whether your graph seems to be consistent with your hypothesis from part (c).

Response: I state that house price is correlated to both Year built and ground living area. As we saw in previous questions, we saw that higher priced homes in the "average" overall condition were homes that were more recently built. Another factor is ground floor area. The more living space, the higher the price of the home. From the graph we can see a steady increase in home price for more recent homes, and the newer homes that are lower price also have lower living space.