Garfield County Sales Data Processing

Casey Bates

1/24/2019

Motivation

Somebody you know is starting a home inspection business in Garfield County, Colorado. You would like to help them launch their business by providing some insights about the real estate market in the county. A dataset is publically available on the county assessor website that contains 2 years of data from summer 2014 through summer 2016.

Processing of the dataset

- Read-in two Excel files: (1) single family home sales, and (2) condo & townhome sales
- Replace spaces in column names with underscore and make lowercase
- Rename some columns
- Add `classification` column to `single_family` dataset
  - Set all values to "Single Family"
- Use `bind_rows()` to combine the datasets into one
- Remove "Garage Only" observations

Glimpse of the dataset

```
glimpse(home_sales)
```

```
## Observations: 1,967
## Variables: 14
## $ account              <chr> "R340967", "R340073", "R112
## $ parcel_number        <chr> "239334401005", "2393342000
## $ reception            <chr> "879240", "870778", "869383
## $ sale_date            <chr> "6/29/2016", "11/24/2015", "
## $ sale_price           <dbl> 650000, 560000, 2750000, 63
## $ situs_address        <chr> "000066 N 2ND ST", "000276 1
## $ location             <chr> "CARBONDALE", "CARBONDALE",
## $ architectural_style  <chr> "ONE STORY", "ONE STORY", "O
## $ year_built           <dbl> 1970, 1971, 2002, 1999, 2008
## $ bedrooms             <dbl> 0, 1, 0, 1, 2, 1, 1, 1, 2, 2
## $ baths                <dbl> 0.00, 1.00, 0.75, 1.00, 1.00
## $ square_feet          <dbl> 0, 480, 680, 710, 764, 804,
## $ legal                <chr> "Section: 34 Township: 7 Ran
## $ classification       <chr> "Single Family", "Single Fam
```

Locations (cities) in Garfield County

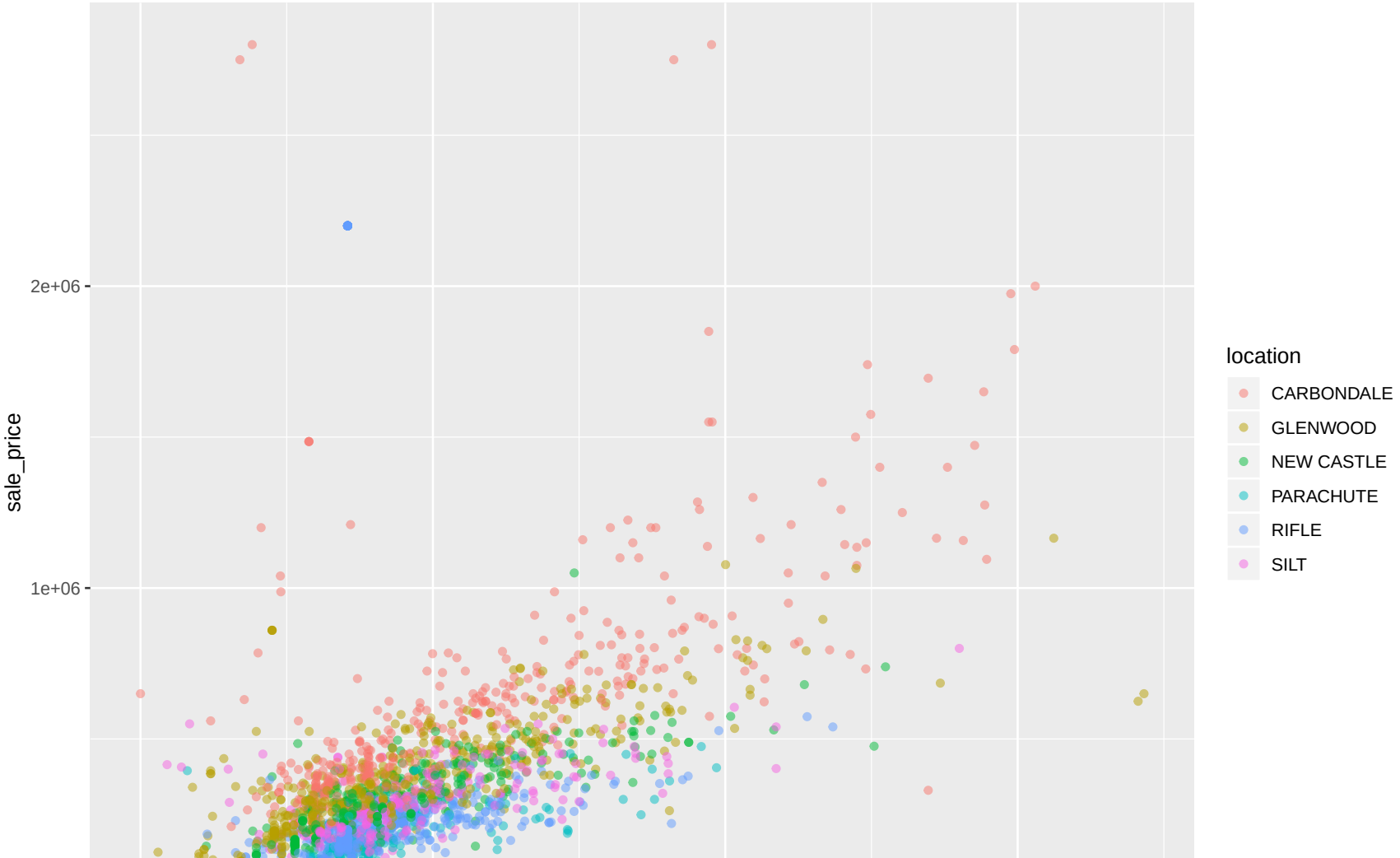
```
unique(home_sales$location)
```

```
## [1] "CARBONDALE" "GLENWOOD"    "NEW CASTLE" "PARACHUTE"
## [6] "SILT"
```

Scatterplot of square_foot and sale_price

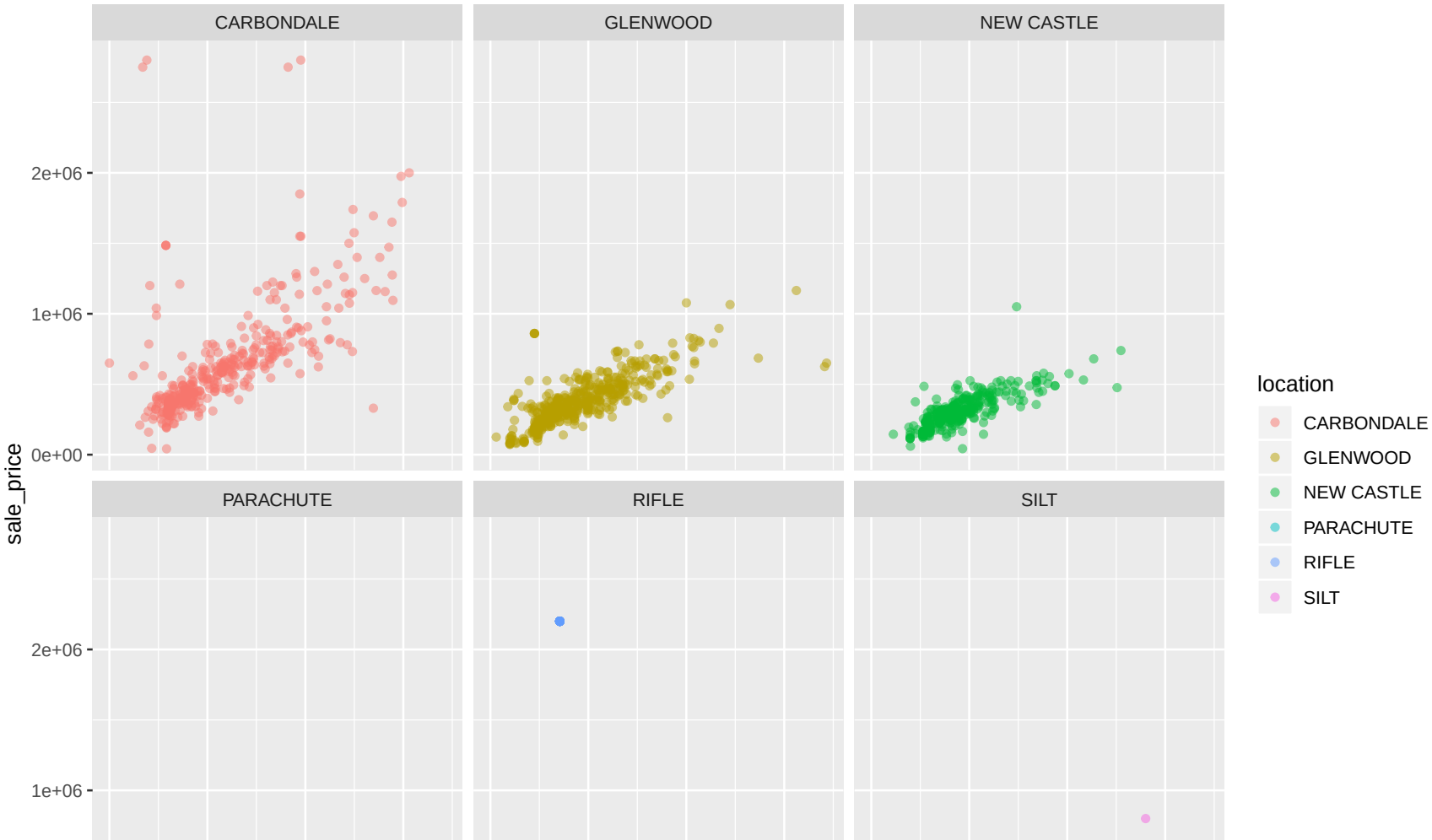
```
ggplot(home_sales, aes(square_feet, sale_price, color = loc
  geom_point(alpha = 0.5)
```

```
## Warning: Removed 1 rows containing missing values (geom_
```



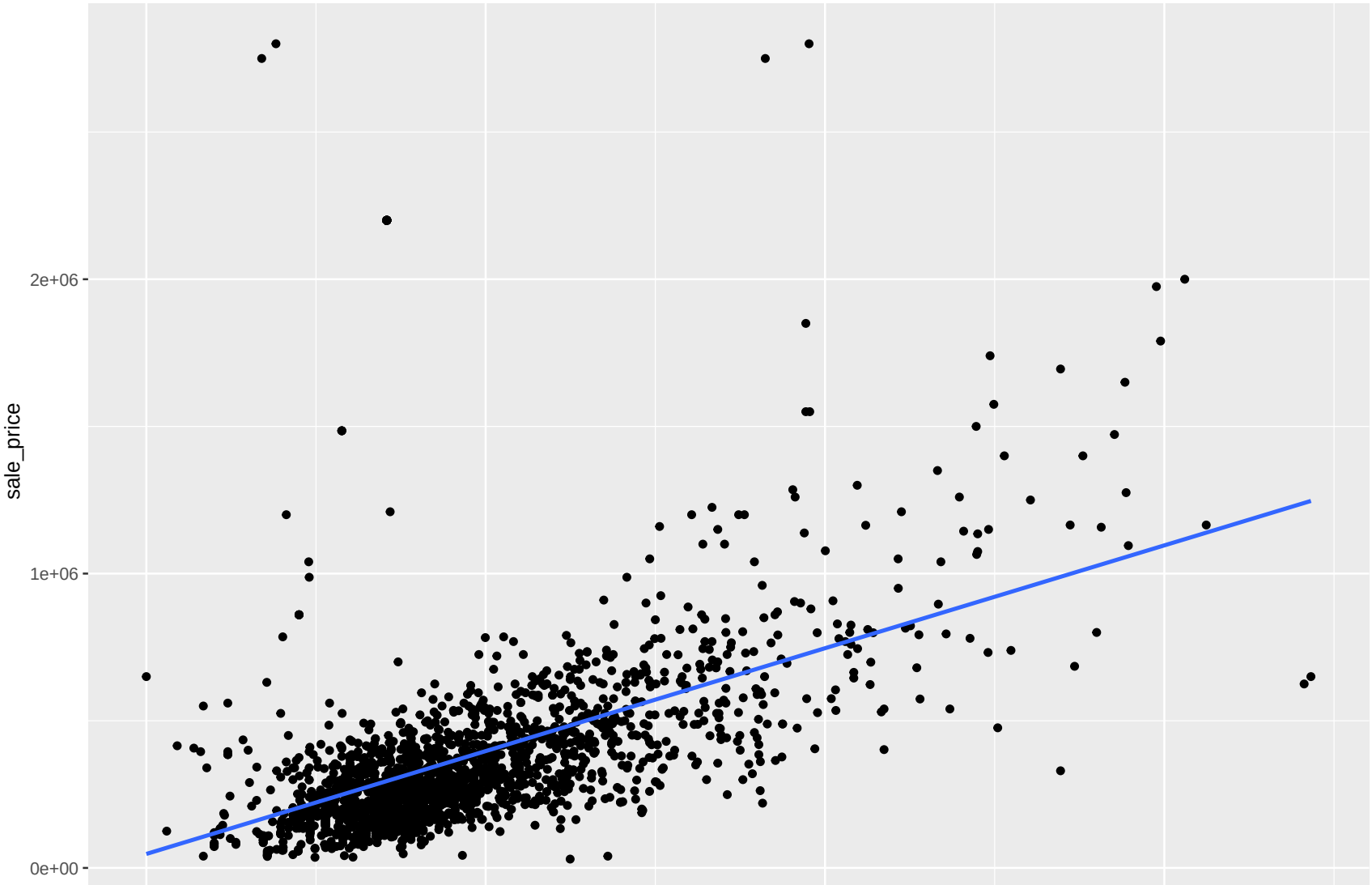
Scatterplot of square_foot and sale_price, faceted by location

```
ggplot(home_sales, aes(square_feet, sale_price, color = loc
  geom_point(alpha = 0.5, position = "jitter") +
  facet_wrap(~ location)
```



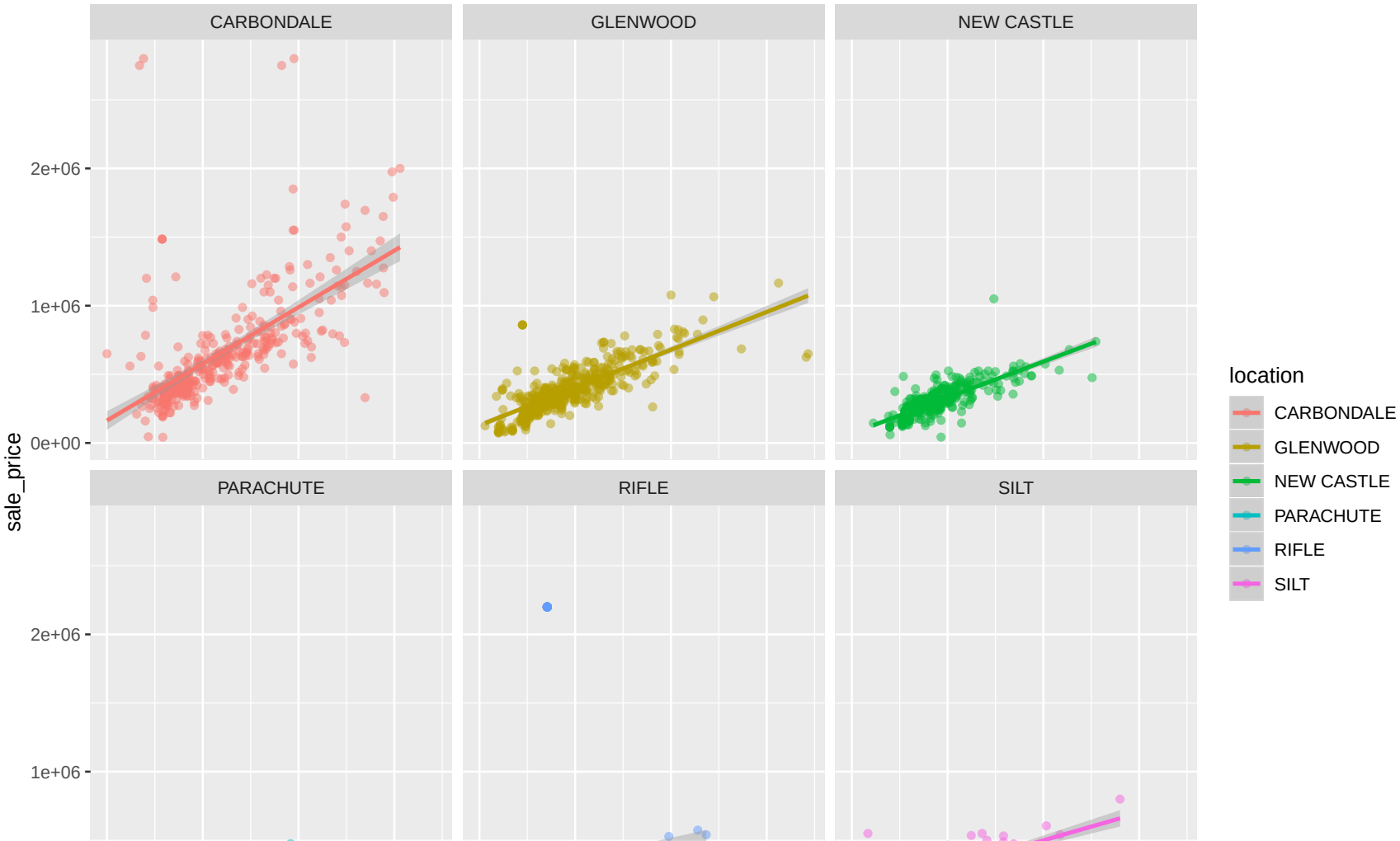
Linear model of sale price vs. square feet

```
ggplot(data = home_sales, aes(x = square_feet, y = sale_pri
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) #Method set to lm
```



Linear model facetted by location

```
ggplot(home_sales, aes(square_feet, sale_price, color = loc
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm") +
  facet_wrap(~ location)
```



High sale price observations:

```
home_sales %>% arrange(desc(sale_price)) %>% select(c("sale
```

```
## # A tibble: 30 x 5
##    sale_price location   classification bedrooms square_
##         <dbl> <chr>      <chr>             <dbl>       <
##  1    2800000 CARBONDALE Single Family         2
##  2    2800000 CARBONDALE Single Family         2
##  3    2750000 CARBONDALE Single Family         0
##  4    2750000 CARBONDALE Single Family         3
##  5    2200000 RIFLE      Townhome              3
##  6    2200000 RIFLE      Townhome              3
##  7    2200000 RIFLE      Townhome              3
##  8    2200000 RIFLE      Townhome              3
##  9    2200000 RIFLE      Townhome              3
## 10    2200000 RIFLE      Townhome              3
## # ... with 20 more rows
```