Honors Project Part 1

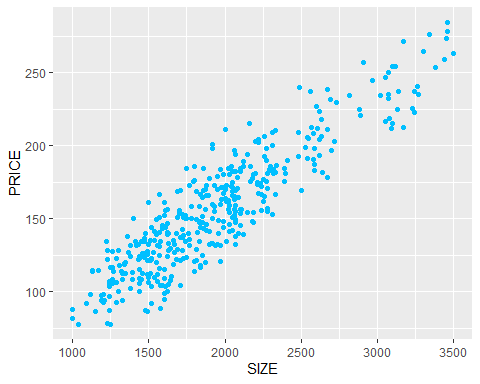
Amit Grewal

2/13/2021

# R Plots and Commands

## Plot 4

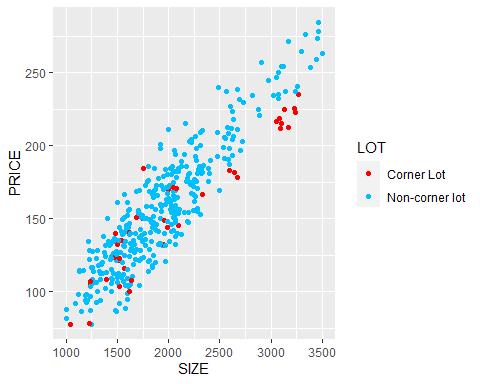
ggplot(data=HOUSEDATA) +  
 geom\_point(mapping=aes(x=SIZE, y=PRICE), color="deepskyblue")



This plot graphs the datapoints from the “HOUSEDATA” dataset on the x~y planes SIZE~PRICE. SIZE corresponds to a datapoint/house’s overall size in Sq. Ft., while PRICE corresponds to each house’s list price in thousands of dollars.

## Plot 7

ggplot(data=HOUSEDATA) +  
 geom\_point(mapping=aes(x=SIZE, y=PRICE, color=LOT)) +  
 scale\_color\_manual(values=c("red", "deepskyblue"))



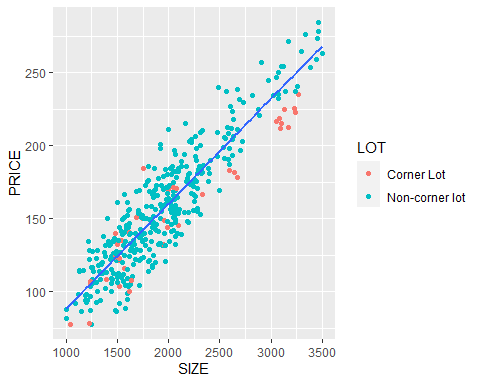
This plot graphs houses on the X~Y planes SIZE~PRICE which are the house’s size in Sq. Ft. and its list price in thousands of dollars, respectively. The plot also differentiates among lot type with red dots corresponding to a corner lot and red dots to a non-corner lot.

## Plots 10-B, 10-C, and 10-D

Plot 10-A is not included as the R command had an error and thus did not compile properly.

### Plot 10-B

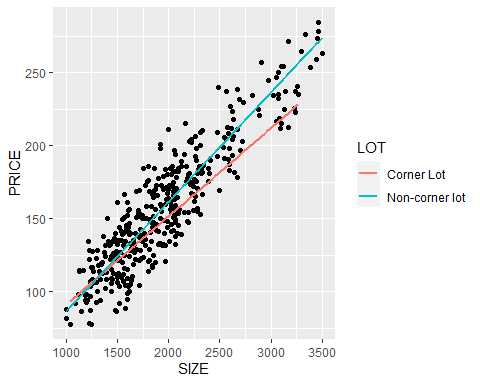
ggplot(data=HOUSEDATA) +  
 geom\_point(mapping=aes(x=SIZE, y=PRICE, color=LOT)) +   
 geom\_smooth(mapping=aes(x=SIZE, y=PRICE), method="lm", se=FALSE)



This plot is incorrect as it only has one overall trend line instead of two trendlines, one for each lot type.

### Plot 10-C

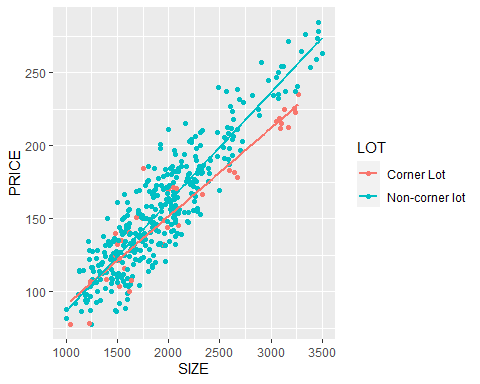
ggplot(data=HOUSEDATA) +  
 geom\_point(mapping=aes(x=SIZE, y=PRICE)) +   
 geom\_smooth(mapping=aes(x=SIZE, y=PRICE, color=LOT), method="lm", se=FALSE)



This plot is incorrect because the datapoints are not differentiated by lot type, and thus are not colored (other than the default black).

### Plot 10-D

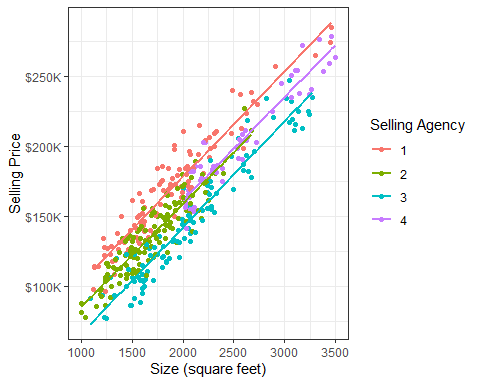
ggplot(data=HOUSEDATA) +  
 geom\_point(mapping=aes(x=SIZE, y=PRICE, color=LOT)) +   
 geom\_smooth(mapping=aes(x=SIZE, y=PRICE, color=LOT), method="lm", se=FALSE)



This plot is correct as it differentiates the datapoints/houses by lot type and includes a trend line for each the two groups of houses (corner and non-corner lots).

## Plot 17

ggplot(data=HOUSEDATA) +  
 geom\_point(mapping=aes(x=SIZE, y=PRICE, color=as.factor(AGENCY))) +   
 geom\_smooth(mapping=aes(x=SIZE, y=PRICE, color=as.factor(AGENCY)), method="lm", se=FALSE) +   
 labs(x="Size (square feet)", y="Selling Price", color="Selling Agency") +  
 scale\_y\_continuous(labels= label\_dollar(prefix = "$", suffix = "K"))+  
 theme\_bw()



### Line 1

(ggplot…) plots the HOUSEDATA canvas.

### Line 2

(geom\_point…) plots the datapoints/houses on the X~Y planes SIZE~PRICE (house size in Sq. Ft. and list price in thousands of dollars) and differentiates among them with color based on the selling agency.

### Line 3

(geom\_smooth…) plots trend lines for each group of houses/datapoints (grouped according to selling agency) and colors them with the same color scheme as geom\_point…

### Line 4

(labs..) changes the X-axis label to “Size (square feet)” and the Y-axis label to “Selling Price”

### Line 5

(scale\_y…) changes the Y-axis units to “$YYYK” where “YYY” is the house’s list price. I.e. the command adds the prefix “$” and the suffix “K” to denote dollars and thousand, respectively

### Line 6

(theme…) changes the plot background from white border/gray cells/white grid lines to black border/white cells/gray grid lines

## Plot 18

ggplot(data=HOUSEDATA) +  
 geom\_point(mapping=aes(x=SIZE, y=PRICE\*1000, color=as.factor(AGENCY))) +   
 geom\_smooth(mapping=aes(x=SIZE, y=PRICE\*1000, color=as.factor(AGENCY)), method="lm", se=FALSE) +   
 labs(x="Size (square feet)", y="Selling Price", color="Selling Agency") +   
 scale\_y\_continuous(labels=scales::dollar\_format()) +   
 theme\_bw()

