# Module 4 - Assignment 2

## Broach, Stuart

### Exploratory Data Analysis

library("tidyverse")

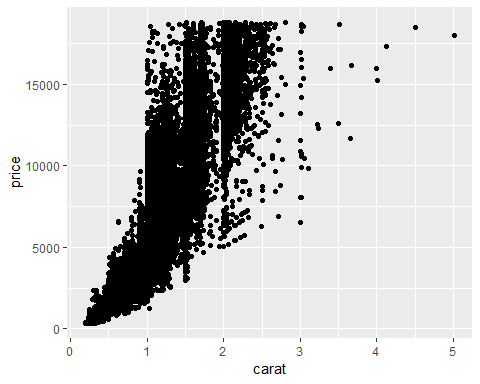
## -- Attaching packages ----------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.7  
## v tidyr 0.8.2 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts -------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

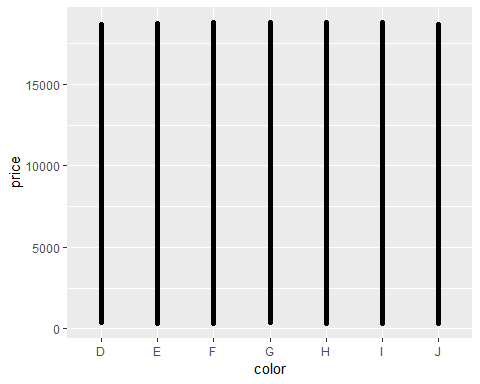
#### Diamond Color and Price

ggplot(diamonds, aes(x=carat,y=price)) + geom\_point()



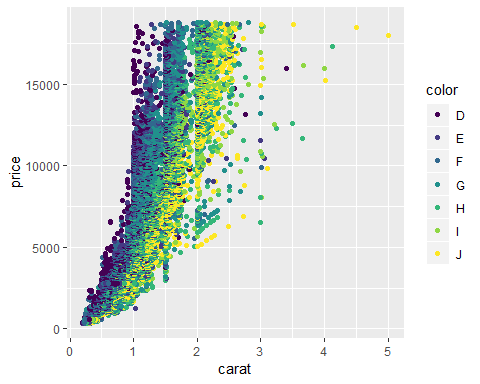
1.What do you notice from the scatterplot as the carat size increases? As the carat size increases, the price increases with it. 2.From the scatterplot, what carats are most represented within the diamonds dataset? 0.25 to 2.5 carats.

ggplot(diamonds, aes(x=color,y=price)) + geom\_point()



Based on this scatter plot, its hard to tel if the color affects the price. All I’m sure of is that all of the colors max out at around the same price.

ggplot(diamonds, aes(x=carat,y=price, color=color)) + geom\_point()



1.) Does color impact the price? I dont think the color impacts the price so much because there appears to be an even amount of colors in all the price ranges. 2.) Are certain colors associated with carat size? Provide an example. Yes, this is absolutley true. The larger carats were mostly the color J.

dsample <- diamonds[sample(nrow(diamonds), 100), ]  
ggplot(dsample, aes(x=carat,y=price, color=color)) + geom\_point() + geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at 0.36525

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 0.35475

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : There are other near singularities as well. 0.83677

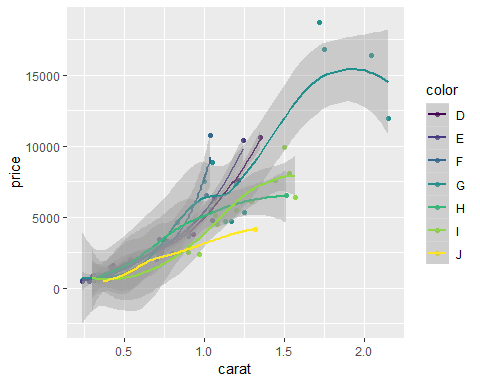
## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : span too small.  
## fewer data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used  
## at 0.36525

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius  
## 0.35475

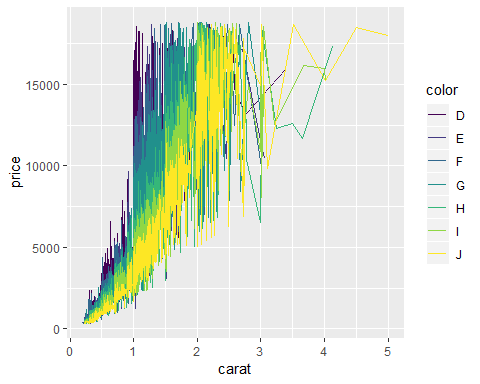
## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal  
## condition number 0

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : There are other  
## near singularities as well. 0.83677



From this plot I can gain little insight based on color, price, and carat other than my initial findings of price increasing with size. Color D has a sudden jump around .8 and then dips back down before returning up. This could be a trend or it could be unique to my small sample. Edit: I had to reenter the dsample and now my chart is different.

ggplot(diamonds, aes(carat,price,color=color)) + geom\_line()



I think the same thing is happening using the line chart as was with the scatter plot. It does make the color more pronounced though and can see a clearer distinction in where the difference between price and carat is.