Visualization Exercises

Kim Roth & Johnathan Clementi

2/3/2020

These are slightly modified exercises from parts of the Visualization Chapter in R for Data Science by Wickham and Grolemund.

The website where the book appears is (and will always be) **free to use**, and is licensed under the [Creative Commons Attribution-NonCommercial-NoDerivs 3.0](http://creativecommons.org/licenses/by-nc-nd/3.0/us/) License.

library(tidyverse)

## -- Attaching packages --------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

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

To add more r chunks to this document you can either pick R from the Insert menu or hit option-command-I(mac) or control-alt-I(pc) ### Exercises

1. How many rows are in diamonds? How many columns?

?diamonds

## starting httpd help server ... done

1A. There are 53940 rows and 10 columns in the diamonds dataset.

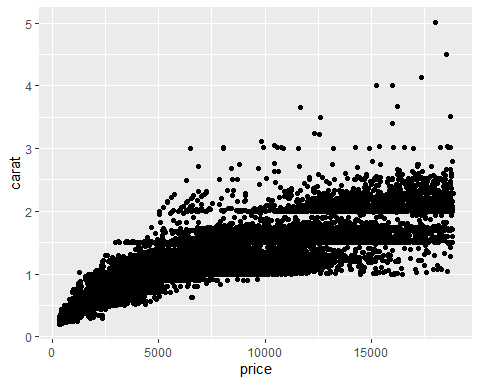
1. What does the carat variable describe? Read the help for ?diamonds to find out. You could also type diamonds in the search bar on the help tab in the Tools window.

2A. The carat variable describes the weight of the diamond (with a range of 0.2 - 5.01).

1. Make a scatterplot of price vs carat.

3A.

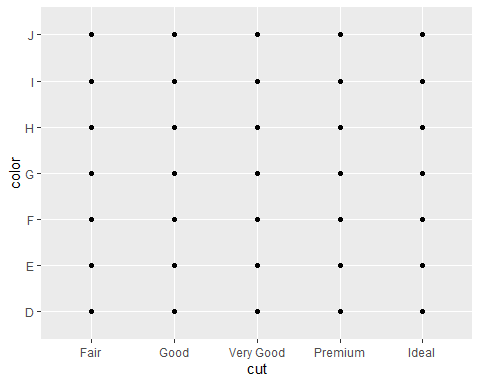
ggplot(data = diamonds) + geom\_point(aes(x = price, y = carat))



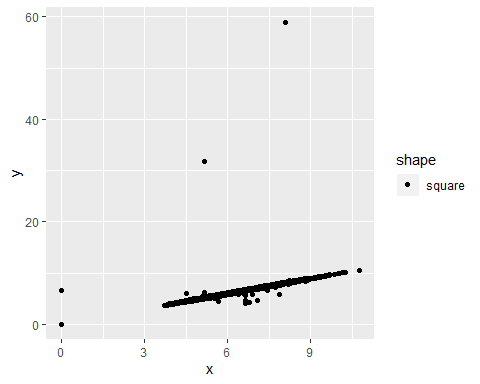
1. What happens if you make a scatterplot of cut vs color? Why is the plot not useful?

4A. This figure is not useful because it plots two unrelated categorical variables against eachother.

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



1. What’s gone wrong with this code? Why are the points not square?

* ggplot(data = diamonds) +   
   geom\_point(mapping = aes(x = x, y = y, shape="square"))
* 

5A.  
The code should be as follows, ensuring that the shape tag is outside of the aes tag

```r  
ggplot(data = diamonds) +   
 geom\_point(mapping = aes(x = x, y = y), shape = "square")  
```  
  
![](ClementiVisualizationExercisesR4DS\_files/figure-docx/unnamed-chunk-5-1.png)<!-- -->

1. Which variables in diamonds are categorical/discrete? Which variables are quantiative/continuous? (Hint: type ?diamonds or type diamonds in the search bar on the help tab in the Tools window to read the documentation for the dataset). How can you see this information when you run diamonds?

6A. Price: continuous; Carat: continuous; Cut: categorical; color: categorical; Clarity: categorical; x: continuous; y: continuous; z: continuous; Depth: continuous; Table: continuous; Table: continuous

When ‘diamonds’ is run in the console, the data type of the variables is displayed at the top of the column - In this case, indicates ordinal (categorical) data, and and indicate numeric (continuous) data types.

1. Map a continuous variable to color, size, and shape. How do these aesthetics behave differently for categorical vs. continuous variables?

7A. ggplot can visualize color and size aesthetics for continuous variables on the fly, meaning that it defines a spectrum of color or size for the variable to be visualized by. In the case of assigning shape to a continuous variable, shape is not something that can be mapped on a spectrum, and therefore cannot be used to visualize continuous data.

# Price by Color  
 ggplot(data = diamonds) + geom\_point(aes(x = price, y = carat, color = price))  
  
 # Price by Size  
 ggplot(data = diamonds) + geom\_point(aes(x = price, y = carat, size = price))  
   
 # Price by Shape  
 ggplot(data = diamonds) + geom\_point(aes(x = price, y = carat, shape = price))

1. What happens if you map the same variable to multiple aesthetics?

8A. In some cases, this can enhance the figure. In the example below, It is easier to decipher the carat difference of diamonds at the higher price range because of the color difference.

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

1. What happens if you map an aesthetic to something other than a variable name, like aes(colour = depth < 60)? Note, you’ll also need to specify x and y.

9A. In this case, it will create a binary coloring system - one color marks values that meet the criteria, one marks values that do not meet the criteria.

ggplot(data = diamonds) + geom\_point(aes(x = price, y = carat, colour = depth < 60))

1. What happens if you facet on a continuous variable?

10A. If a continuous variable is used for a facet, ggplot will create a subplot for each value based on significant figures provided.

ggplot(data = diamonds) + geom\_point(aes(x = price, y = carat)) +  
 facet\_wrap(~ x)



1. What does the se argument to geom\_smooth() do?

11A. The se argument is a Boolean argument that determines if the confidence intervals are displayed on a figure or not.

1. Will these two graphs look different? Why/why not?

12A. These graphs will look the same because the x and y variables are the same for the aes, point, and smooth arguments.

```r  
ggplot(data = diamonds, mapping = aes(x = depth, y = table)) +   
 geom\_point() +   
 geom\_smooth()  
  
ggplot() +   
 geom\_point(data = diamonds, mapping = aes(x = depth, y = table)) +   
 geom\_smooth(data = diamonds, mapping = aes(x = depth, y = table))  
```

1. What plots does the following code make? What does . do?

13A. The following plots are different uses of facets. The ‘.’ operator is a place holder which tells ggplot not to facet in a row or column dimension (depending on argument order)

```r  
 ggplot(data = diamonds) +   
 geom\_point(mapping = aes(x = depth, y = table)) +  
 facet\_wrap(cut ~ .)  
  
ggplot(data = diamonds) +   
 geom\_point(mapping = aes(x = depth, y = table)) +  
 facet\_grid(cut ~ .)  
  
ggplot(data = diamonds) +   
 geom\_point(mapping = aes(x = depth, y = table)) +  
 facet\_grid(. ~ color)  
```

1. Read ?facet\_wrap or type facet\_wrap in the search bar on the help tab in the Tools window. What does nrow do? What does ncol do? What other options control the layout of the individual panels? Why doesn’t facet\_grid() have nrow and ncol arguments?

14A. The ncol argument is the number of columns in a facet\_wrap. There is also an nrow, which controls the number of rows in a facet\_wrap. These arguments are not used in facet\_grids because facet grids are meant to have two variables displayed in rows and columns.

1. When using facet\_grid() you should usually put the variable with more unique levels in the columns. Why?

15A. It will be easier to view in column form than in a row form (books/documents/webpages are usually in portrait view).