ADS2 Practical 6: Visualizing Data

Chaochen Wang

Semester 1, 2019/20

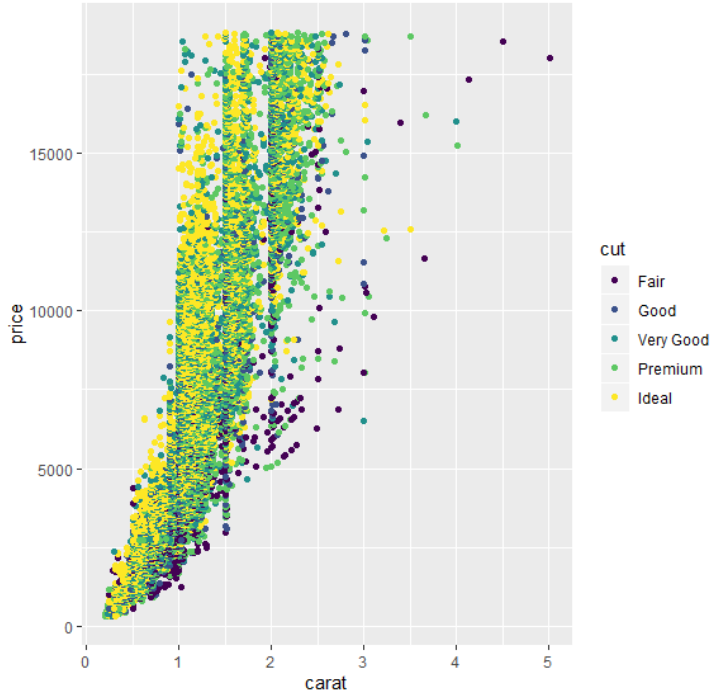
Work through this guide alone or in groups. Facilitators are here to help. The time it takes to complete this practical can vary between individuals - this is OK. Do not worry if you do not finish within the session.

Learning Objectives

- Use ggplot2 for data visualization
- Think critically about data visualization choices

1. Overplotting

In the lecture, you have learnt how to use *ggplot2* to generate a scatter plot from the dataset "diamonds". Please repeat it and generate the following plot. You can get the dataset by data("diamonds").



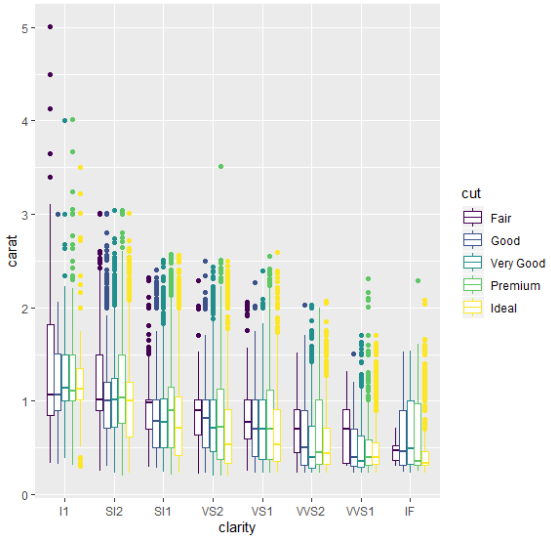
1. You can see that the dots(points) are overlaid. Please try to resize the point size to 0.1 or change the transparency to alpha 1/5 to make it look better.
2. Each observation is presented as a round dot. You can also try another shape by set the shape argument to another number, such as 18.

2. Rewrite the code

In the #18 of lecture slides, we use the stat_bin to generate the two plots. Please try to rewrite the code and use geom_XXX to generate the same plots.

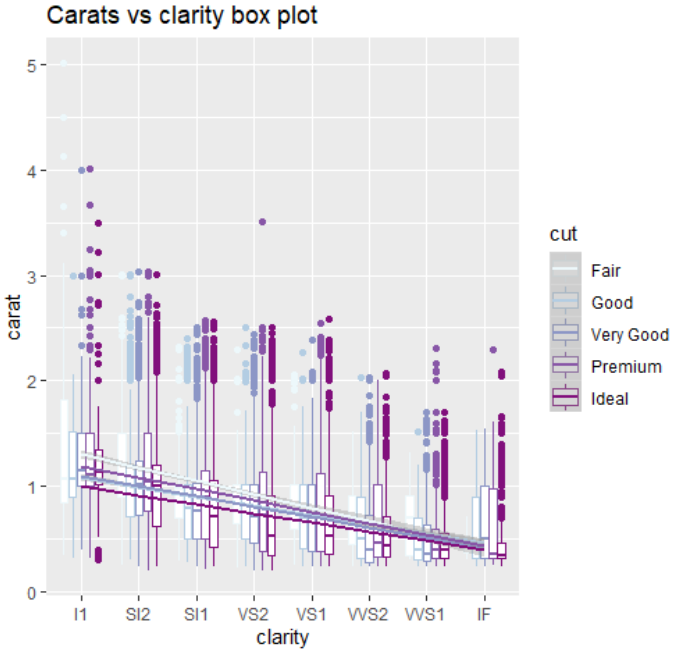
3. Build plots layer by layer

1. Plot the following boxplot from the dataset “diamonds”.



2. generate another layer of linear fitting using geom_smooth, use method “lm”. Save the geom_smooth to a new object sm.
3. faceting the plot by cut, color and cut~color.
4. add another layer to add a title to the plot using labs
5. Change the color key using scale_color_brewer. Now you should get a blot similar to this.
6. Save the plot to a png file.

The file should be similar to this.

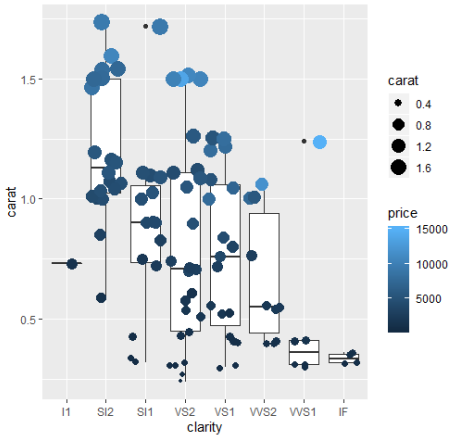


4. Scale the y axis

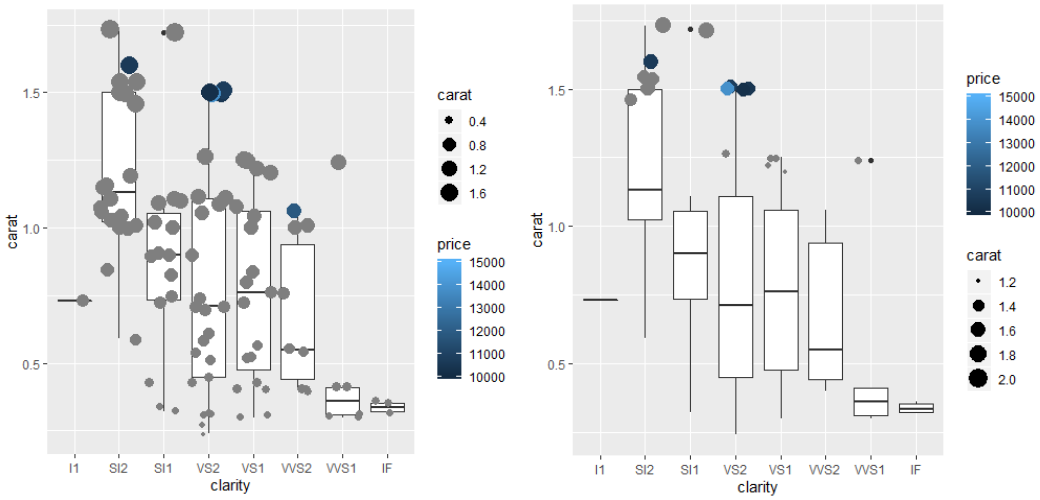
1. Start from the plot in 3.2, transform the y-axis scale to log10 using scale_y_continuous. Pay attention to the change of y-axis. What should be the unit? Please change the y-axis label to include the unit using ylab.
2. redo the boxplot in 3.1 by changing the y aesthetics to log10(carat). Compare the y-axis here with the one in 3.1. Please change the y-axis label to include the unit.
3. Add a layer of linear fitting to the plot in 4.1 by +sm from 3.2, do you see a problem? Please fix the problem by change the aesthetics in sm.
4. change the range of y-axis in the plot 4.1. set the limits to 0.3 to 3.0 using scale_y_continuous.

4. Jitter plot and scales.

1. Sample out 100 cases from diamonds dataset. Generate a plot with a layer of boxplot and a layer of jitter plot like this using geom_jitter.



2. Reset the color scale to only color the diamonds in the price range (10000, 15000). Try scale_color_gradient.
3. Reset the size scale to only show the diamonds in the carat range (1.2, 2). Try scale_size_continous.



5. Position

Generate the plots in lecture slide #21. Use the position argument in geom_bar.