

Practical 9: Visualizing data

Dmytro Shytikov and Xushen Xiong

2023-11-16

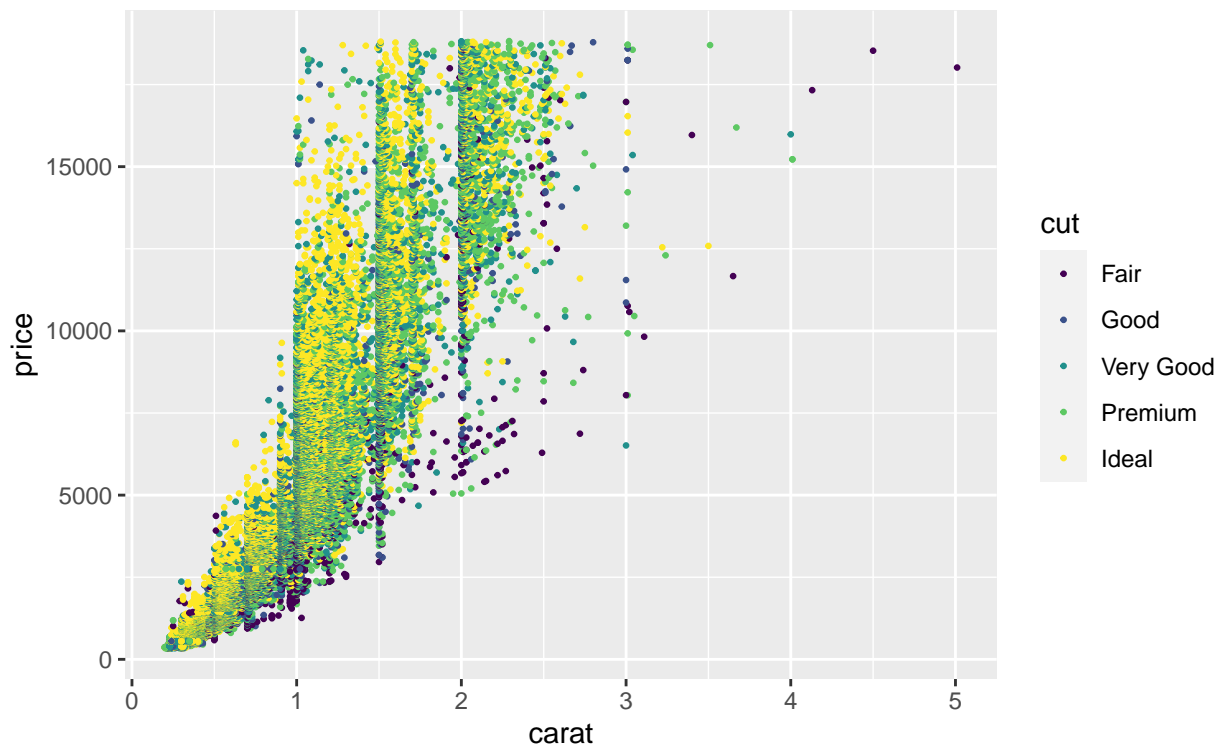
Work through this guide alone or in groups. Facilitators are here to help. The time it takes to complete this practical can vary between individuals – this is OK. Do not worry if you do not finish within the session.

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.

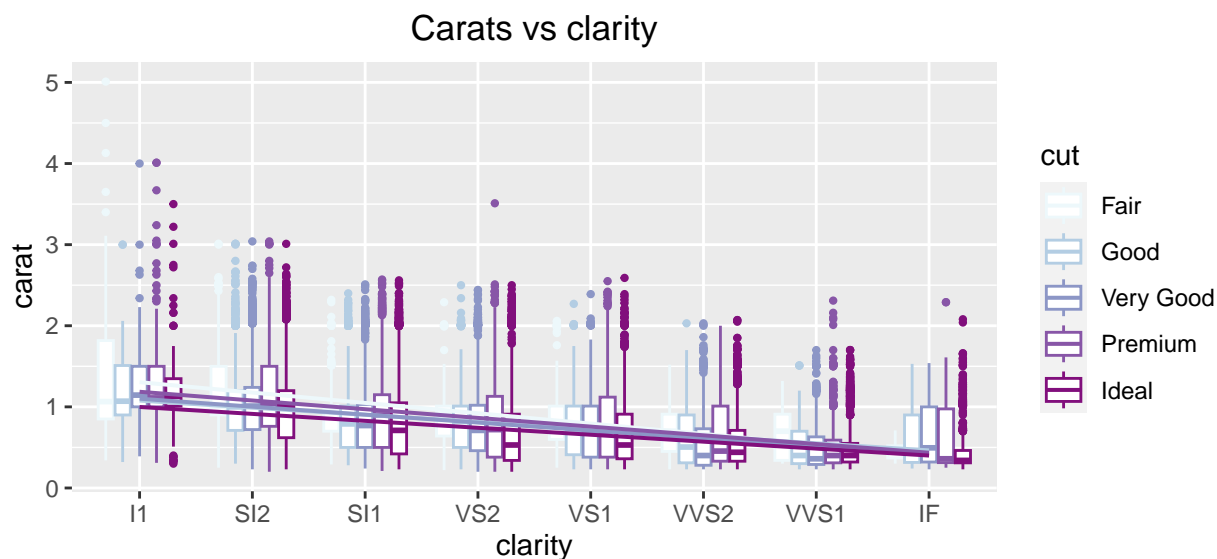
- Each observation is presented as a round dot. You can also try another shape by set the `shape` argument to another number, such as 18.

2. Rewrite the code

In the second of lecture slides (choosing different `geom_XXXX` and `stat_XXXX` functions), we use the `stat_bin` function to generate two plots. Please try to rewrite the code and use `geom_XXX` and `stat_XXX` to generate the same plots.

3. Build plots layer by layer

- Plot the following boxplot from the dataset `diamonds`.
- generate another layer of linear fitting using `geom_smooth`, use method `lm`. Save the `geom_smooth` to a new object `sm`.
- Apply faceting to the plot by `cut`, `color` and `cut~color`.
- Add another layer to add a title to the plot using `labs`
- Change the color key using `scale_color_brewer`. Now you should get a plot similar to this.



- Save the plot to a png file.

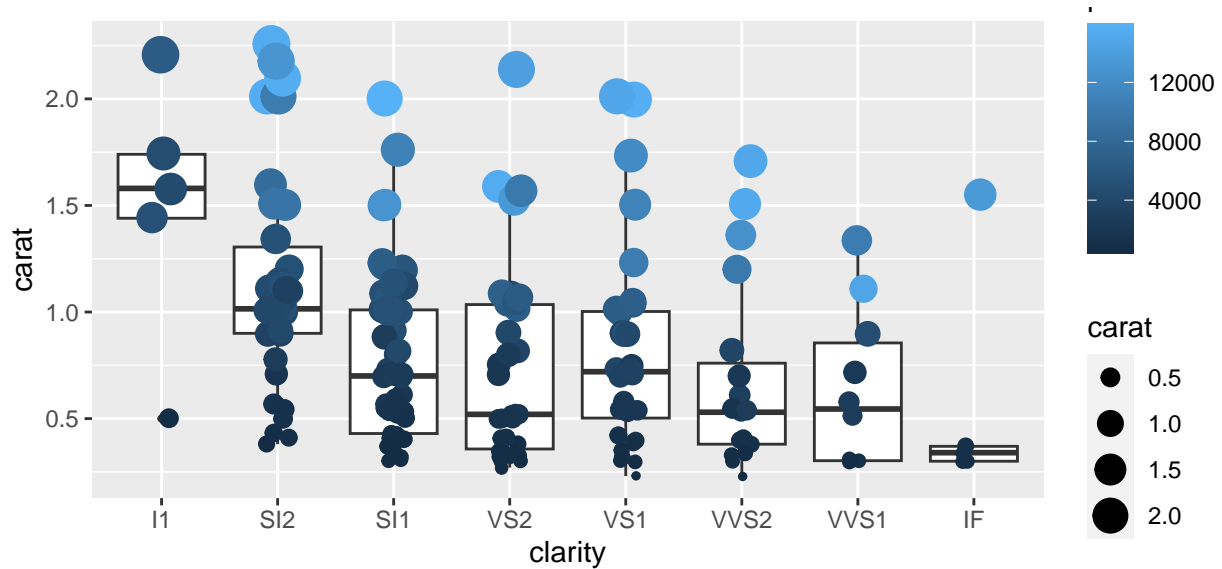
4. Scale the y axis

- 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`.
- 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. Change the y-axis label to include the unit.

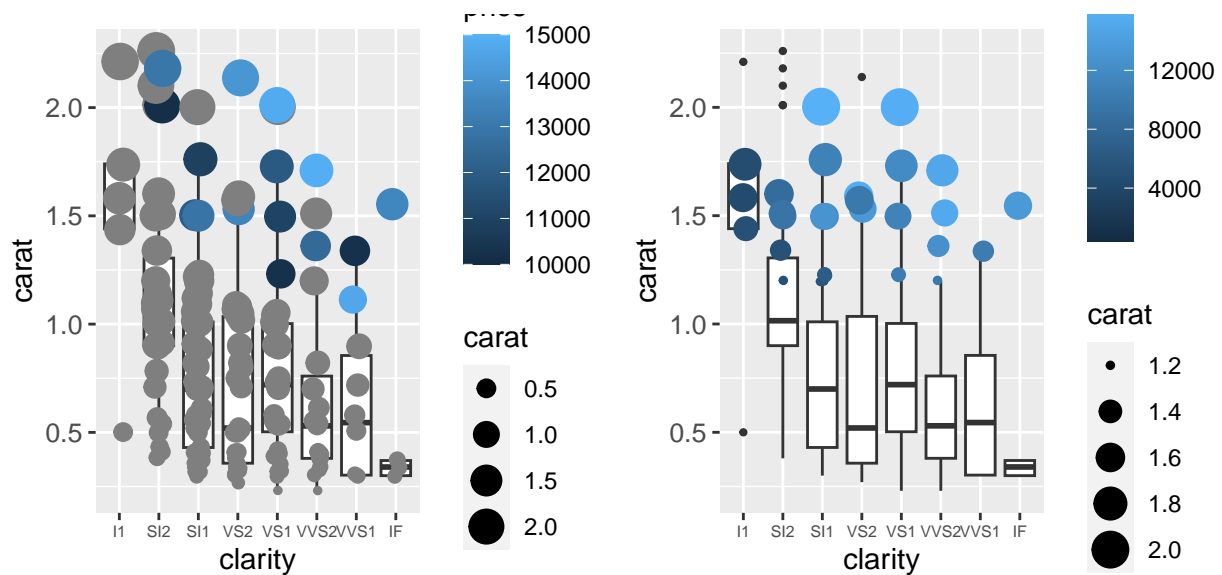
3. Add a layer of linear fitting to the plot in 4.2 by + `sm` from point 3.2. Do you see a problem? Please fix the problem by change the aesthetics in `sm`.
4. change the range of y-axis in the plot 4.1. set the limits to 0.3 to 3.0 using `scale_y_continuous`.

5. 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`.

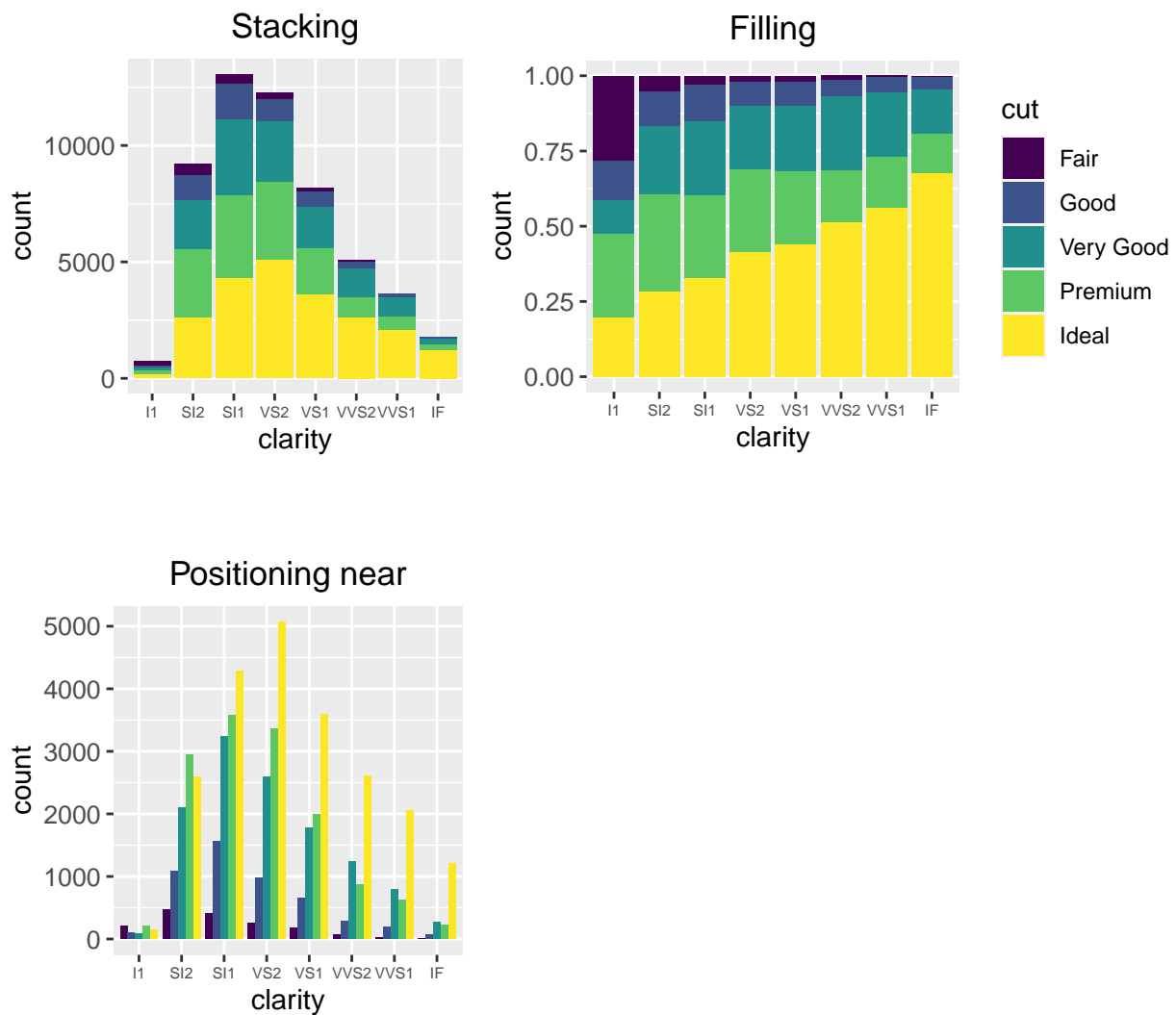


2. Reset the color scale to only color the diamonds in the price `range(10000, 15000)`. Try `scale_color_gradient`.
3. Reset the size scale to only show the diamonds in the carat `range(1.2, 2)`. Try `scale_size_continuous`.



6. Position.

Generate the plots as were in the third part of lecture (slide about the position adjustment). Use the position argument in `geom_bar`.



Originally created by Chaochen Wang in 2021, CC-BY-SA 3.0

Last update by Dmytro Shytikov in 2023