STK 353

Practical 1: Basics

Question sheet

Submission: 23 July 2018, 17:30

This practical is based on a pre-loaded dataset in the R package tidyverse. The dataset can be accessed by loading tidyverse in R. For more information, consult the chapter on Data Visualisation in the STK 353 prescribed textbook 'R for Data Science' (http://r4ds.had.co.nz/data-visualisation.html)

1. Run ggplot(data = mpg). What do you see?
2. How many rows are in mpg? How many columns?
3. What does the `drv` variable describe? Read the help for `?mpg` to find out.
4. Make a scatterplot of `hwy` vs `cyl`.
5. What happens if you make a scatterplot of `class` vs `drv`? Why is the plot not useful?
6. Which variables in `mpg` are categorical? Which variables are continuous? (Hint: type `?mpg` to read the documentation for the dataset). How can you see this information when you run mpg?
7. Map a continuous variable to color, size, and shape. How do these aesthetics behave differently for categorical vs. continuous variables?
8. What happens if you map the same variable to multiple aesthetics?
9. What does the stroke aesthetic do? What shapes does it work with? (Hint: use `?geom_point`)
10. What geom would you use to draw a line chart? A boxplot? A histogram? An area chart?