Lab Sheet 10

STAT260, University of Otago

Sem. 2, 2021

Today we will use the full `diamonds` data set which is part of the `ggplot2` package. The variables we will use are:
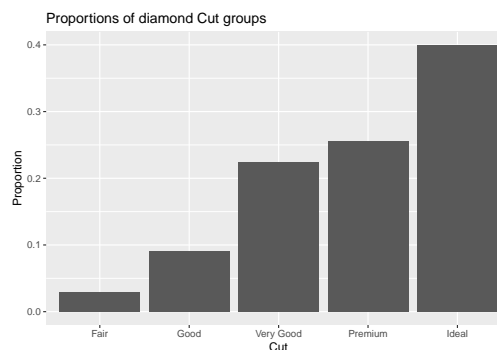
- **carat:** weight of the diamond
- **cut:** quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- **color:** diamond colour, from J (worst) to D (best)

Remember to load the `ggplot2` package via `library(ggplot2)` The data set is called `diamonds`, and is loaded along with `ggplot2`.
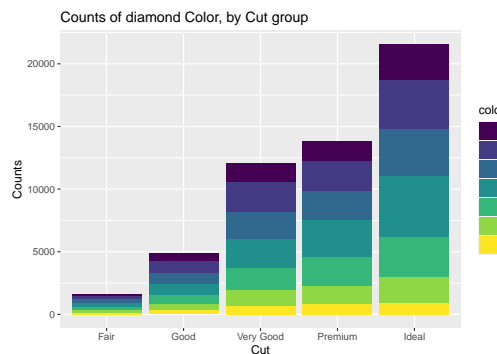
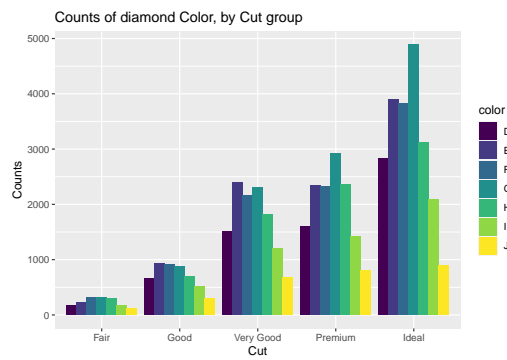## Today's tasks: write R code to reproduce the plots below.

Plot 1

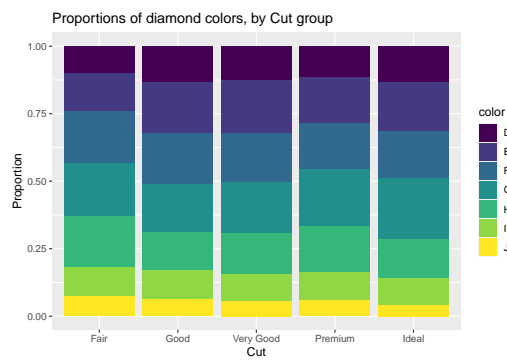HINT: for this plot, use the `..prop..` notation when setting the `aes` value in `geom_bar`.
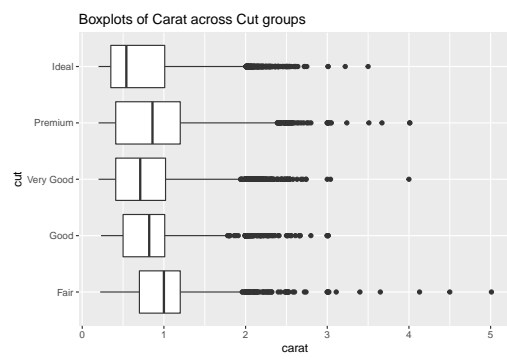

Plot 2

Plot 3

Counts of diamond Color, by Cut group

Plot 4

Proportions of diamond colors, by Cut group

Plot 5

Boxplots of Carat across Cut groups

Plot 6

Range of Carat values, across Cut groups

Plot 7

Violin plots of Carat across Cut groups

Plot 8

Freqpoly representation of Carat histograms by Cut
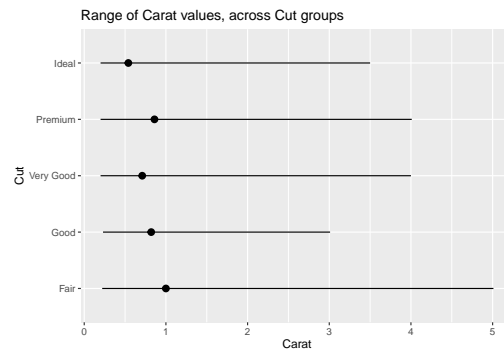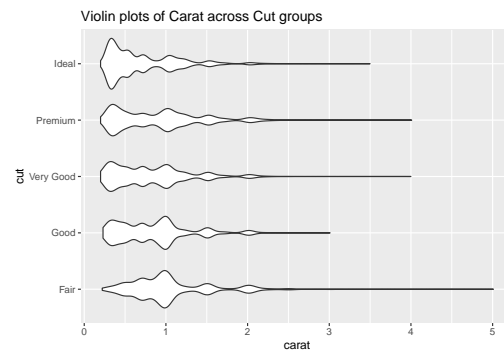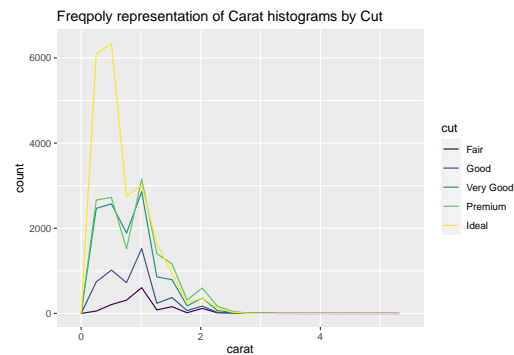
Thinking

The above four plots show different ways of presenting the same data. Think about the pros and cons of each. Which do you think is the most effective means of displaying the data? Why?