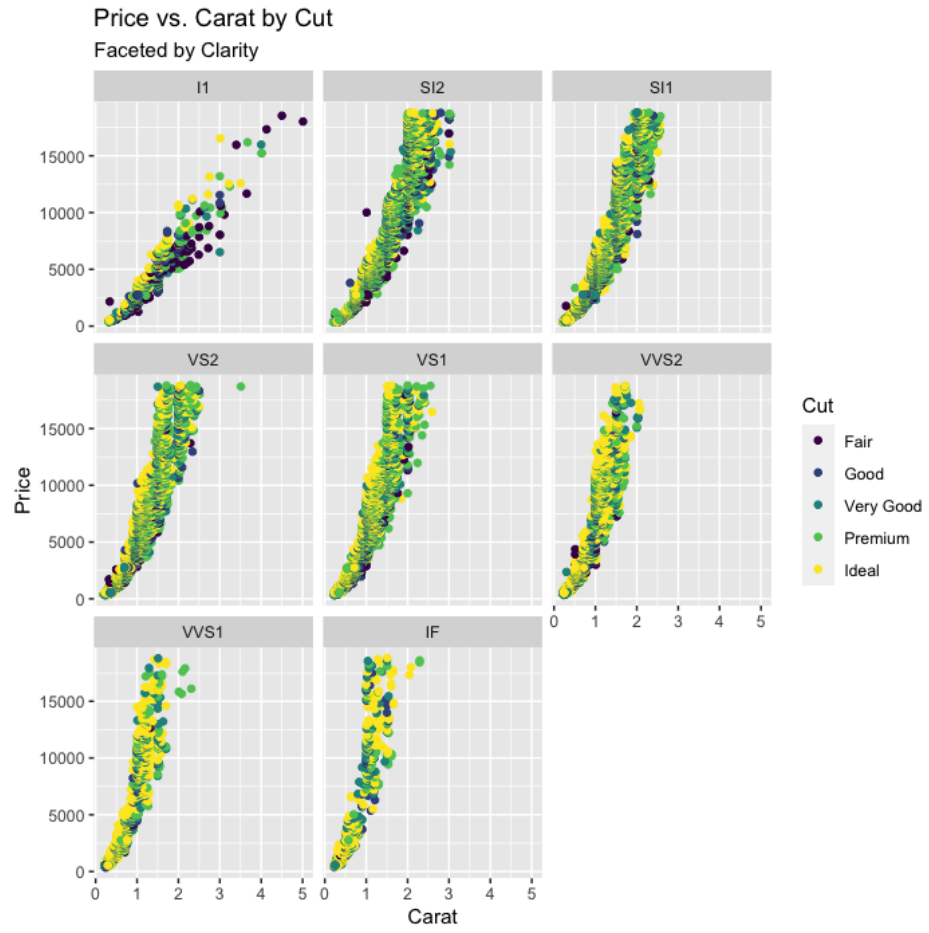
# Visualizing data with ggplot2

Maria Tackett

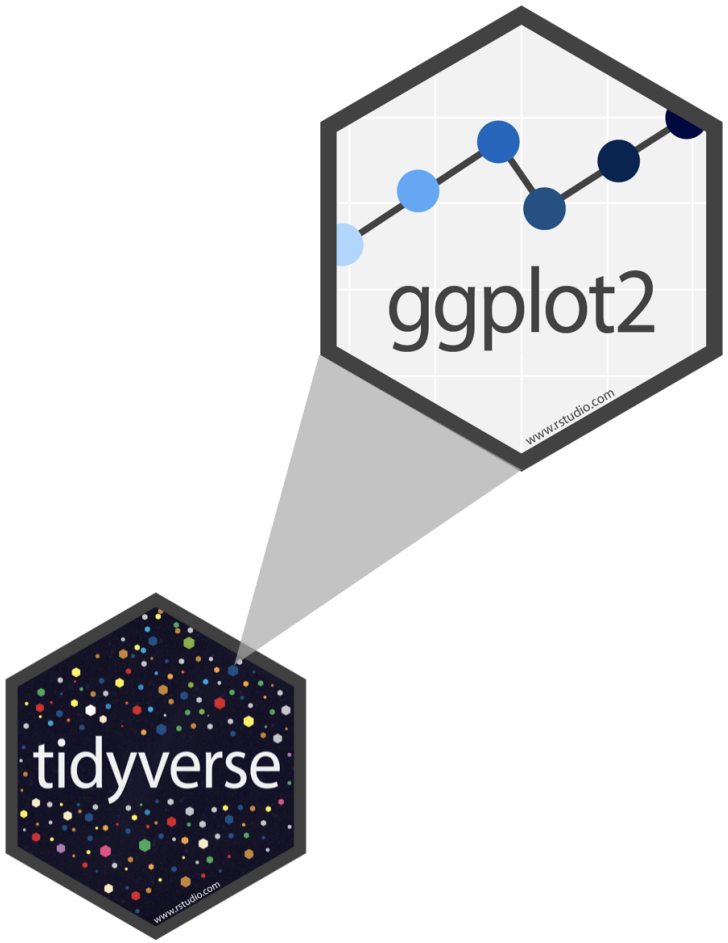
05.12.20

# Learning objectives

- Create a scatterplot using **`ggplot`**
- Add aesthetics to a plot
- Create smaller plots for subsets of data



# ggplot2 in tidyverse



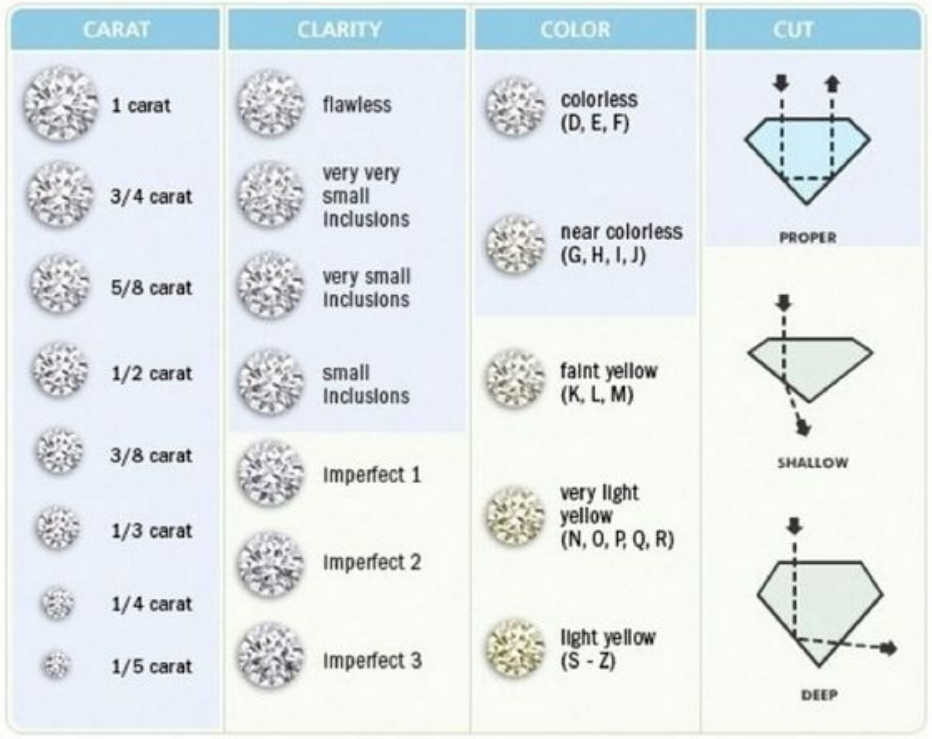

- ggplot2 is tidyverse's data visualization package.
- The **gg** in "ggplot2" stands for "grammar of graphics".
- It is inspired by the book **Grammar of Graphics** by Leland Wilkinson.

# Grammar of Graphics

A **grammar of graphics** is a tool that enables us to concisely describe the components of a graphic.

# Today's data

- We will use the **diamonds** data set in the **`ggplot2`** package
- Contains price and other attributes for ~ 54,000 diamonds
- Variables:
  - **`cut`**
  - **`clarity`**
  - **`carat`**
  - **`color`**
  - **`price`**



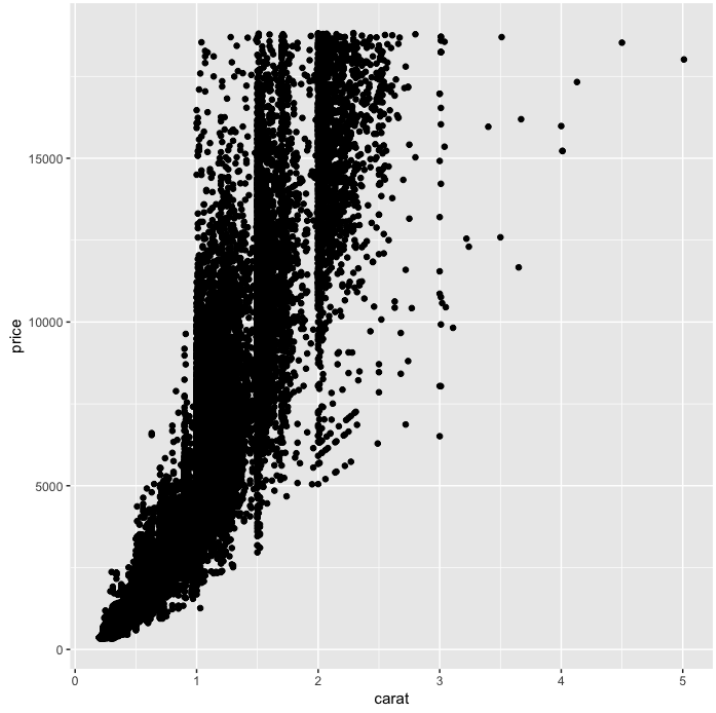
# Basic `ggplot` syntax

```
ggplot(data = [dataset], aes(x = [x-var], y = [y-var])) +
  geom_xx() +
  other options
```

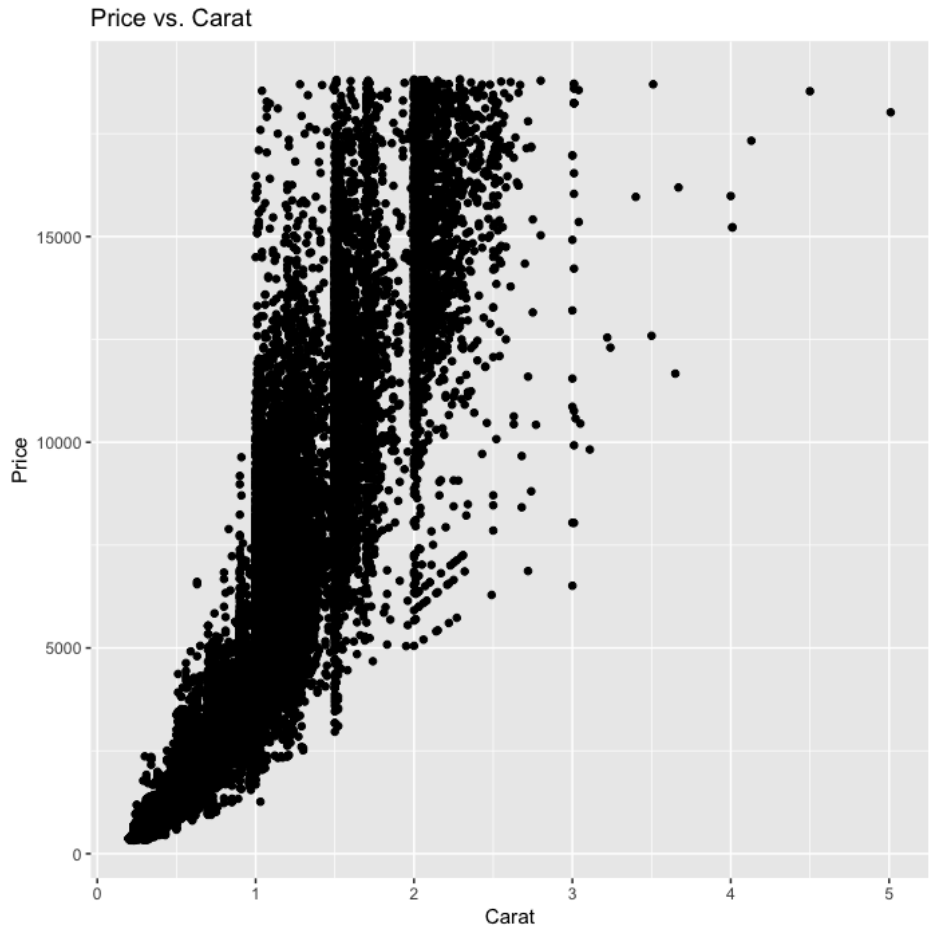
# Basic `ggplot` syntax

```
ggplot(data = [dataset], aes(x = [x-var], y = [y-var])) +
  geom_xx() +
  other options
```

```
ggplot(data = diamonds,
       aes(x = carat,
           y = price)) +
  geom_point()
```



# Our first plot



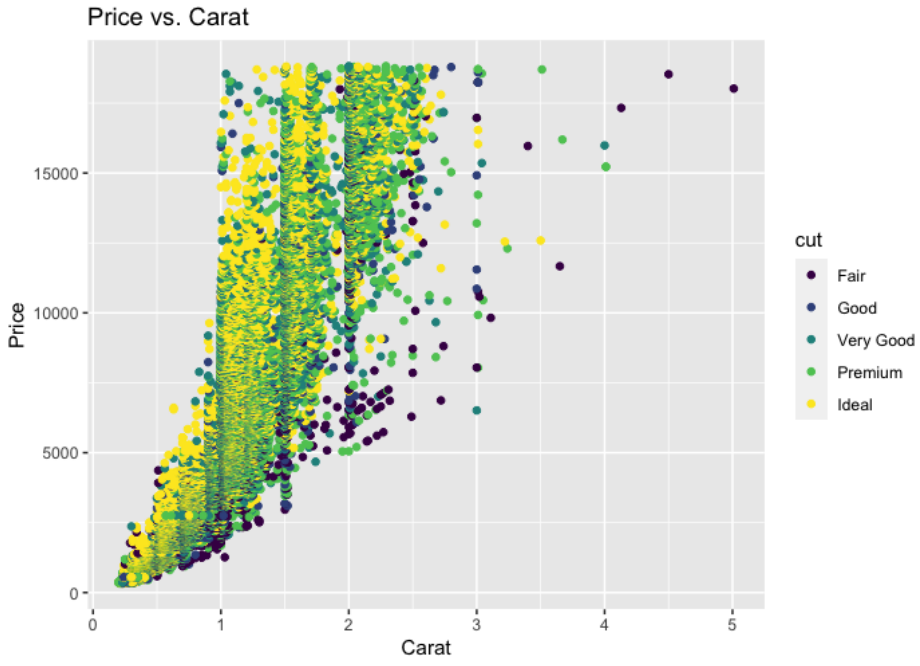

Let's create this plot in R.

# Aesthetics

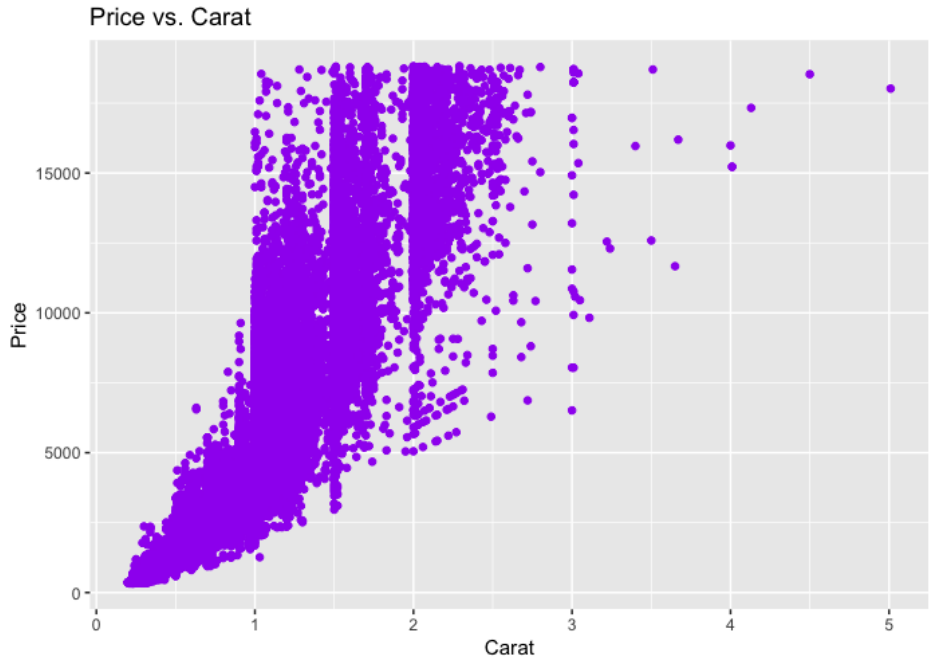
We can add **aesthetics** (features) to our plot to incorporate additional variables or to customize the plot. These aesthetics include

- **`color`**
- **`shape`**
- **`size`**
- **`alpha`** (transparency)

# Adding aesthetics




- Aesthetic based on value of **`cut`**.
- Aesthetic defined inside **`aes()`**.




- Same aesthetic for all observations.
- Aesthetic defined outside of **`aes()`**.

Let's add color and shape aesthetics to our plot.

# Exercise

Consider the following code:

```
ggplot(data = diamonds, aes(x = carat, y = price)) +
  geom_point(aes(shape = cut), color = "blue")
```

Which of the following best describes the points on the plot?

**a.** The shape is the same for all points.

**b.** The color is the same for all points.

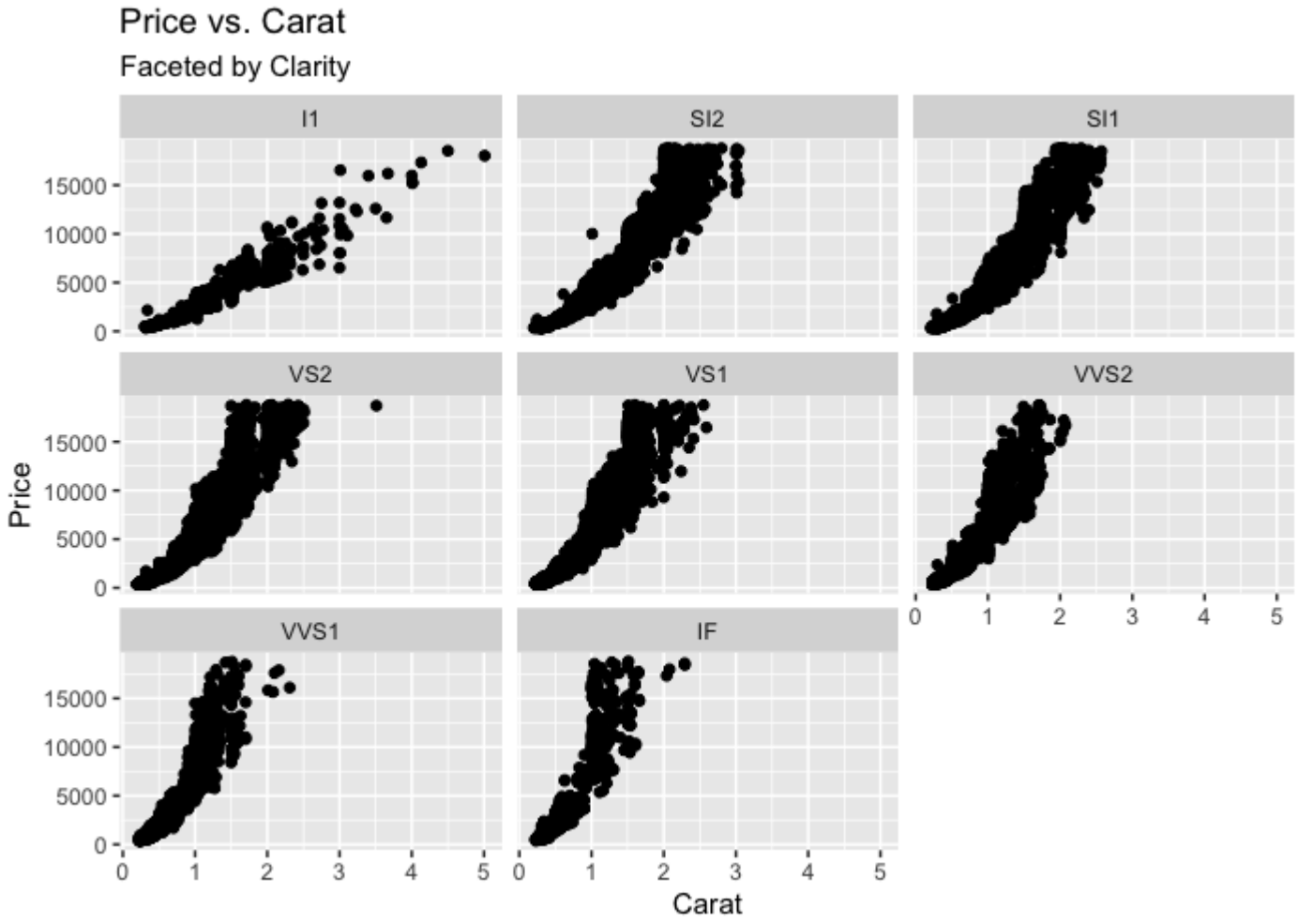
**c.** The color and shape are the same for all points.

# Faceting

Use **faceting** to create smaller plots that display different subsets of the data.

```
ggplot(data = diamonds, aes(x = carat, y = price)) +
  geom_point() +
  labs(x = "Carat",
       y = "Price",
       title = "Price vs. Carat") +
  facet_wrap(~ clarity)
```

# Faceting by `clarity`



Let's put it all together. Create a plot of **`price`** by **`carat`**, with an aesthetic by **`cut`**, faceted by **`clarity`**.

# Your turn

It's your turn to create data visualizations using **`ggplot2`**!

- Go to https://matackett.shinyapps.io/data-viz/.
- Click on *Exercise 2: Putting it all together*.

# `ggplot2` references

- "Data visualisation" in *R for Data Science*
- ggplot2 Reference
- Data Visualization Cheatsheet
- *ggplot2 : Elegant Graphics for Data Analysis*