Data visualization Lab

Bibek Sapkota

Data Visualization Lab (faceting qplot)

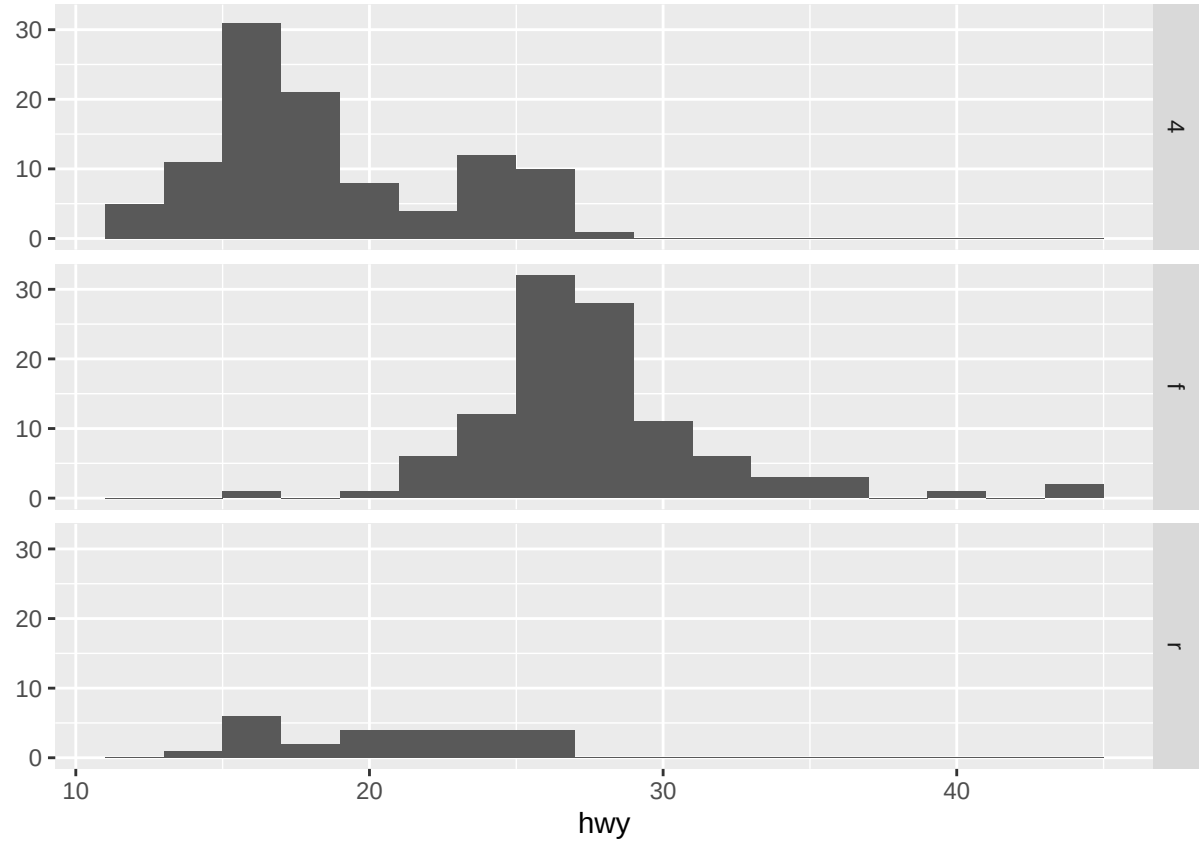
Task: Load ggplot2 package

```
library(ggplot2)
```

Task 1: creating a faceted histogram of highway miles per gallon (hwy) from the mpg dataset. The data is segmented into facets based on drive type (drv), with each facet showing histograms binned at a width of 2.

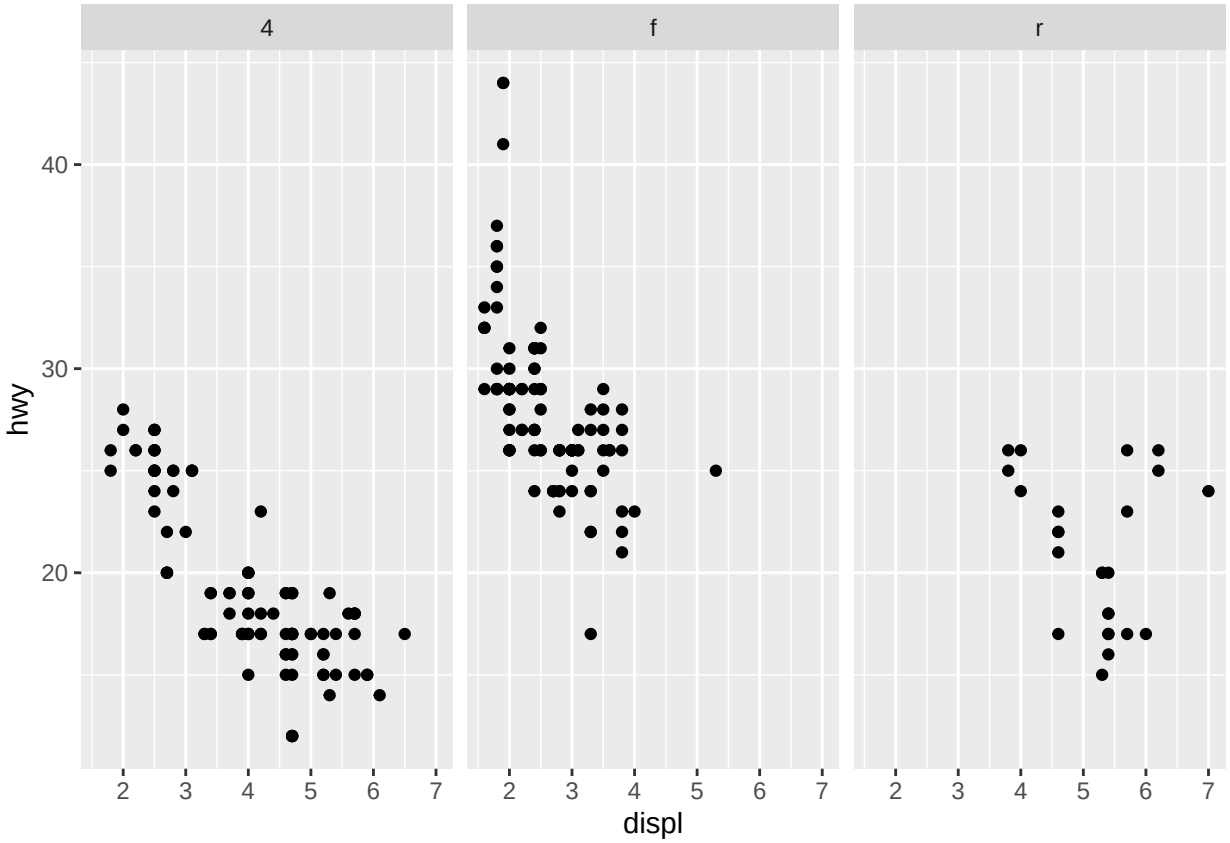
```
qplot(hwy,data = mpg,facets = drv~.,binwidth=2)
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Task 2: creating a faceted scatter plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset. The data is segmented into facets for each drive type (drv).

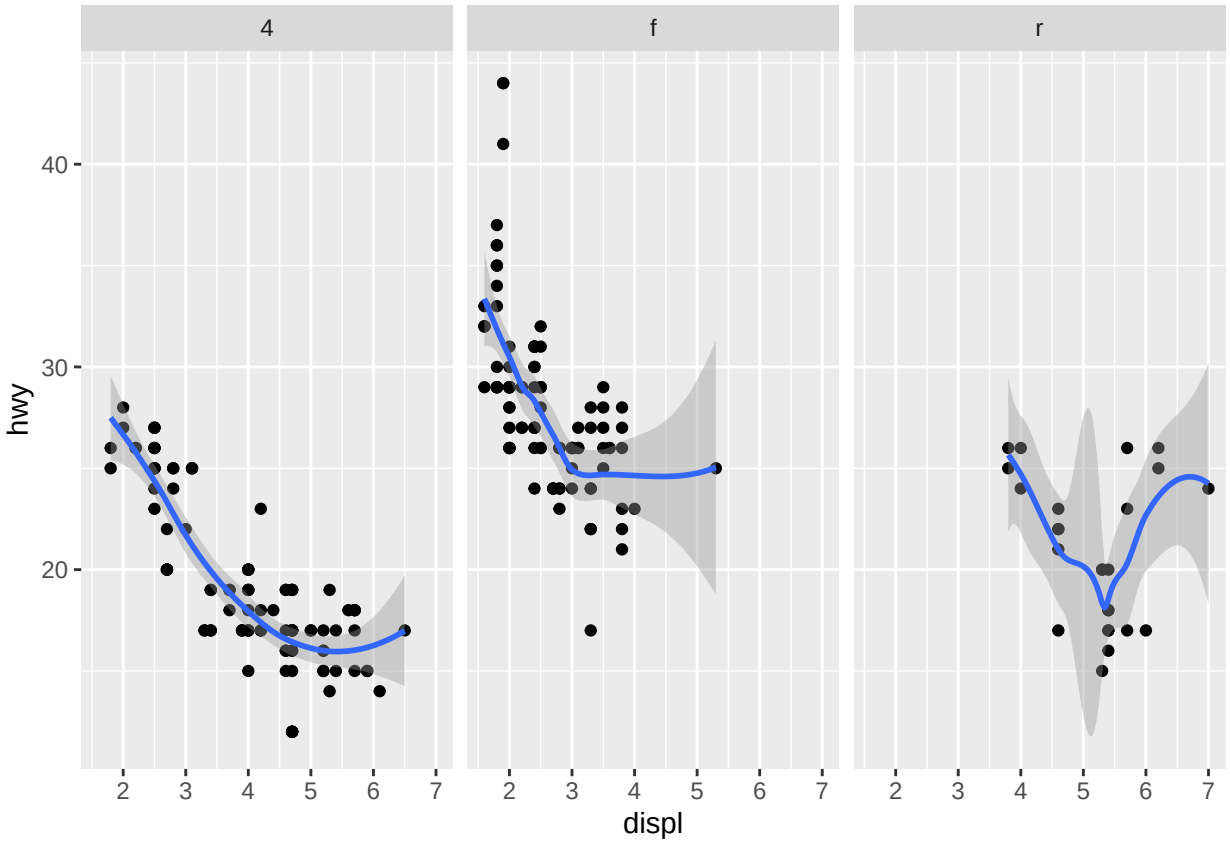
```
qplot(displ,hwy,data = mpg,facets = .~drv)
```



Task 3: creating a faceted scatter plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset. It includes both points and a smoothed trend line (geom = c("point", "smooth")), with data facets arranged by drive type (drv).

```
qplot(displ,hwy,data = mpg,facets = .~drv,geom = c("point","smooth"))
```

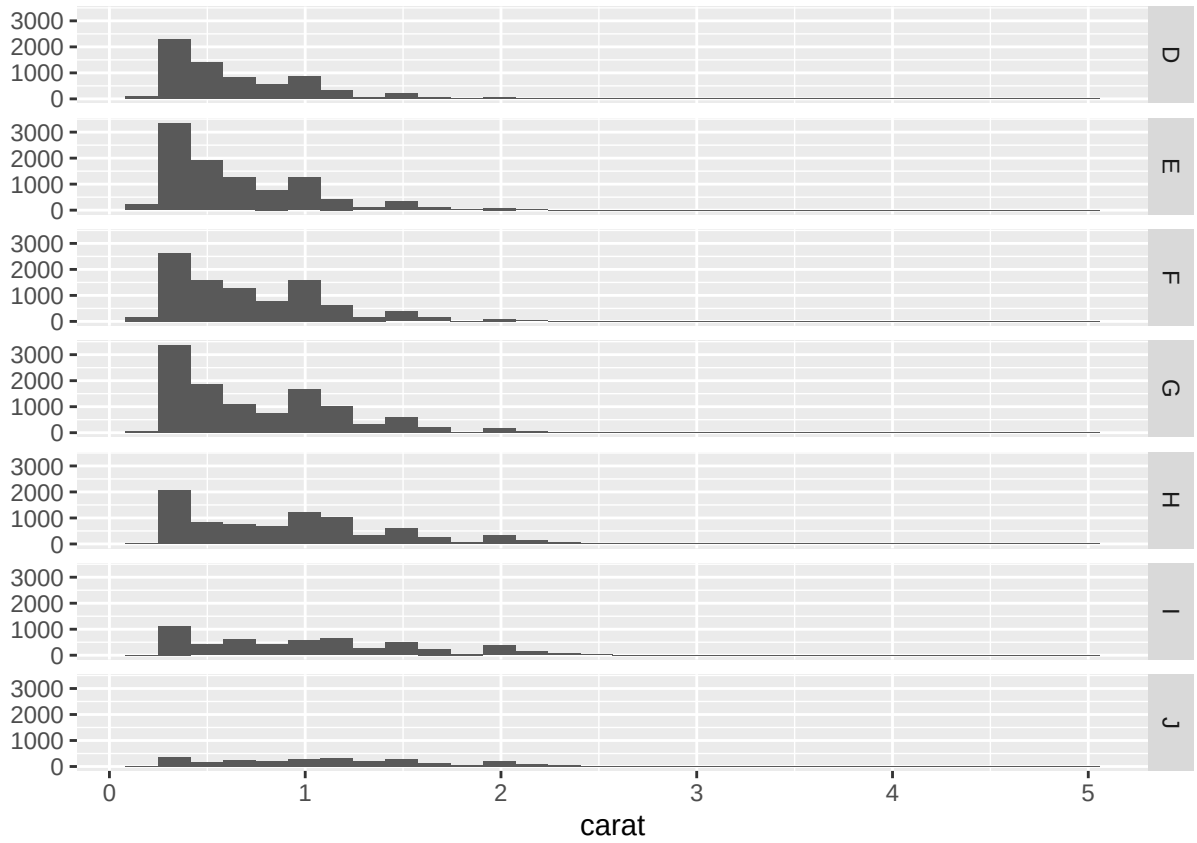
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Task 4:creating a histogram of diamond carat weights (carat) from the diamonds dataset. The data is segmented into facets based on diamond color (color).

```
qplot(carat,data = diamonds,facets = color~.,geom = "histogram")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

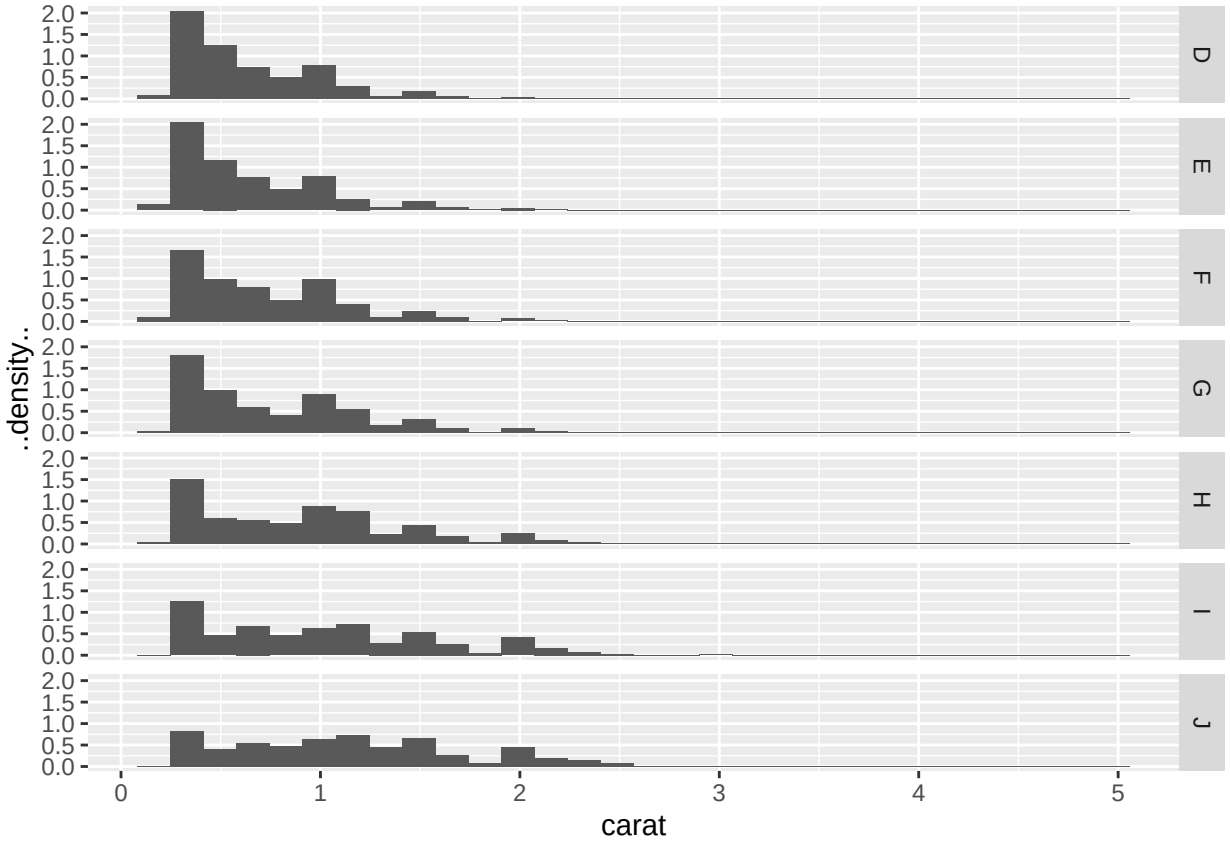


Task 5:Creating a faceted histogram of diamond carat weights (carat) from the diamonds dataset. The histograms show density (normalized count) on the y-axis (..density..), and the data is segmented into facets based on diamond color (color).

```
qplot(carat, ..density.., data = diamonds, facets = color ~ ., geom = "histogram")
```

```
## Warning: The dot-dot notation (‘..density..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(density)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

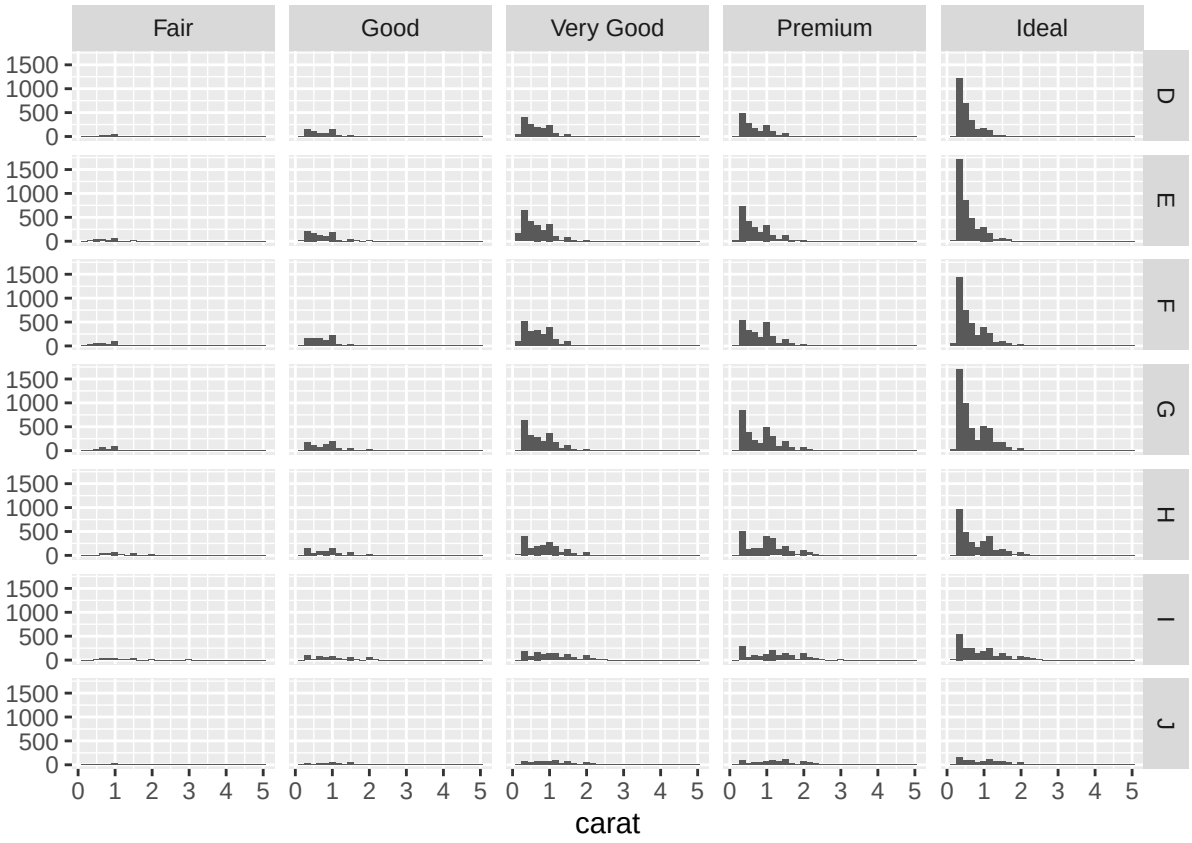
```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```



Task 6:Creating a faceted histogram of diamond carat weights (carat) from the diamonds dataset. The data is segmented into facets based on both diamond color (color) and cut quality (cut). Each facet displays a separate histogram

```
qplot(carat, data = diamonds, facets = color ~ cut, geom = "histogram")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Exercise

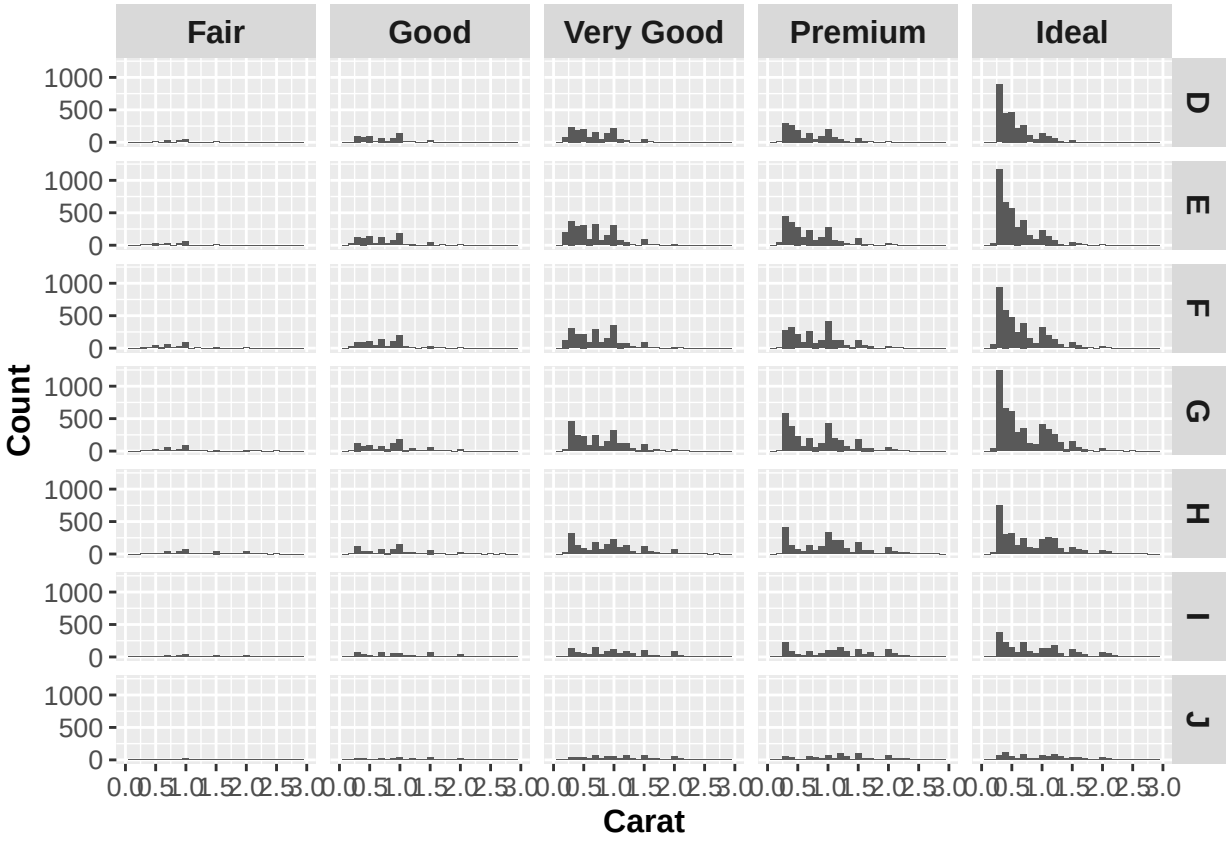
1. Adjust axis limits of last 3 plots to make them clearer

Task 1:Specifies the x-axis variable and creates a histogram. Facets the grid by diamond color and cut. Adjusts the x-axis with a label, limits, and tick marks. Enhances clarity by adjusting the size of axis text, and the size and boldness of axis titles and facet labels.

```
qplot(carat, data = diamonds, geom = "histogram", facets = color ~ cut, binwidth = 0.1) +
  scale_x_continuous(name = "Carat", limits = c(0, 3), breaks = seq(0, 3, by = 0.5)) +
  scale_y_continuous(name = "Count", breaks = seq(0, 3000, by = 500)) +
  theme(
    axis.text.x = element_text(size = 10),
    axis.text.y = element_text(size = 10),
    axis.title.x = element_text(size = 12, face = "bold"),
    axis.title.y = element_text(size = 12, face = "bold"),
    strip.text = element_text(size = 12, face = "bold")
  )
```

```
## Warning: Removed 32 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

```
## Warning: Removed 70 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



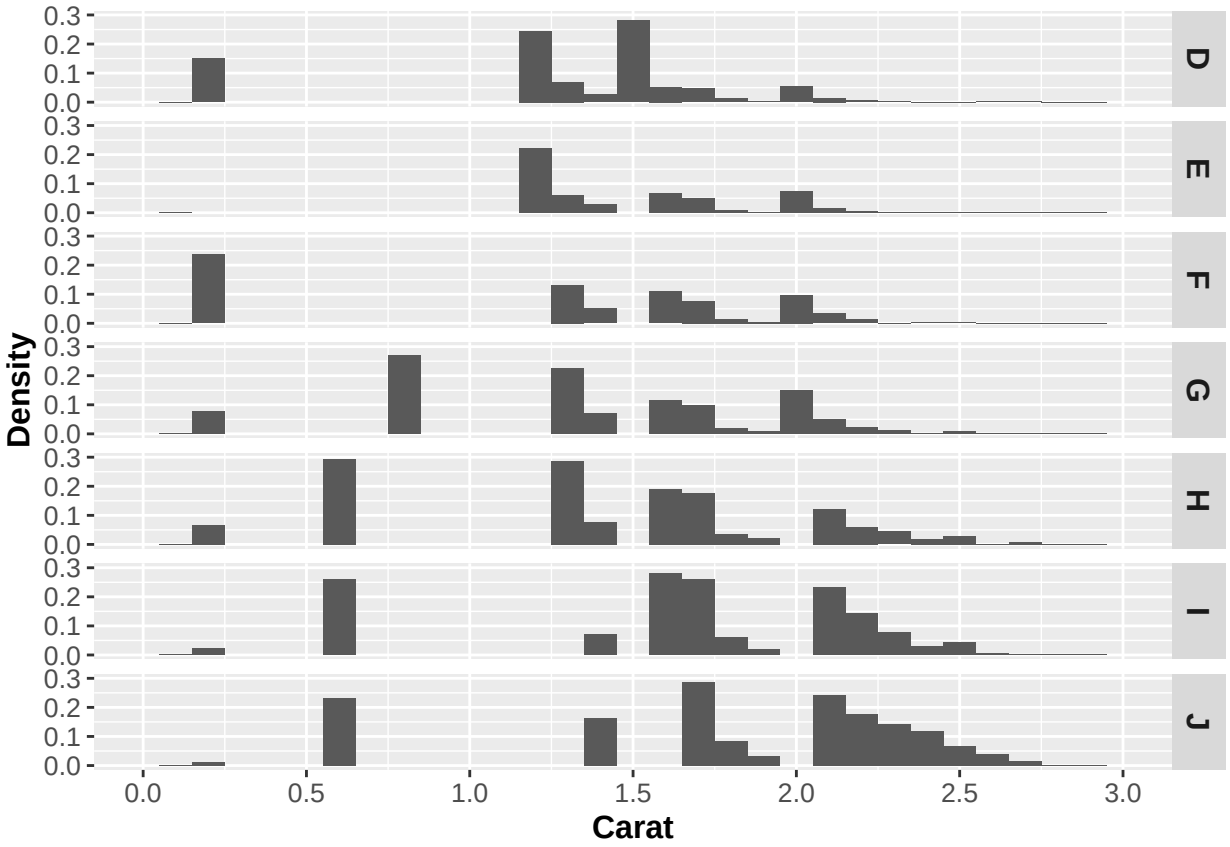
Task 2:Adjustments were made to the x-axis by labeling it "Carat," setting limits from zero to three carats, and adding breaks at regular intervals. The y-axis was labeled "Density," with limits set to ensure density visibility and breaks at regular intervals. The theme was adjusted to improve readability by setting the axis text size appropriately for both axes, and making the axis titles and facet labels bold and clear

```
library(ggplot2)

qplot(carat, ..density.., data = diamonds, geom = "histogram", facets = color ~ ., binwidth = 0.1) +
  scale_x_continuous(name = "Carat", limits = c(0, 3), breaks = seq(0, 3, by = 0.5)) +
  scale_y_continuous(name = "Density", limits = c(0, 0.3), breaks = seq(0, 0.6, by = 0.1)) +
  theme(
    axis.text.x = element_text(size = 10),
    axis.text.y = element_text(size = 10),
    axis.title.x = element_text(size = 12, face = "bold"),
    axis.title.y = element_text(size = 12, face = "bold"),
    strip.text = element_text(size = 12, face = "bold")
  )
```

```
## Warning: Removed 32 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

```
## Warning: Removed 91 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



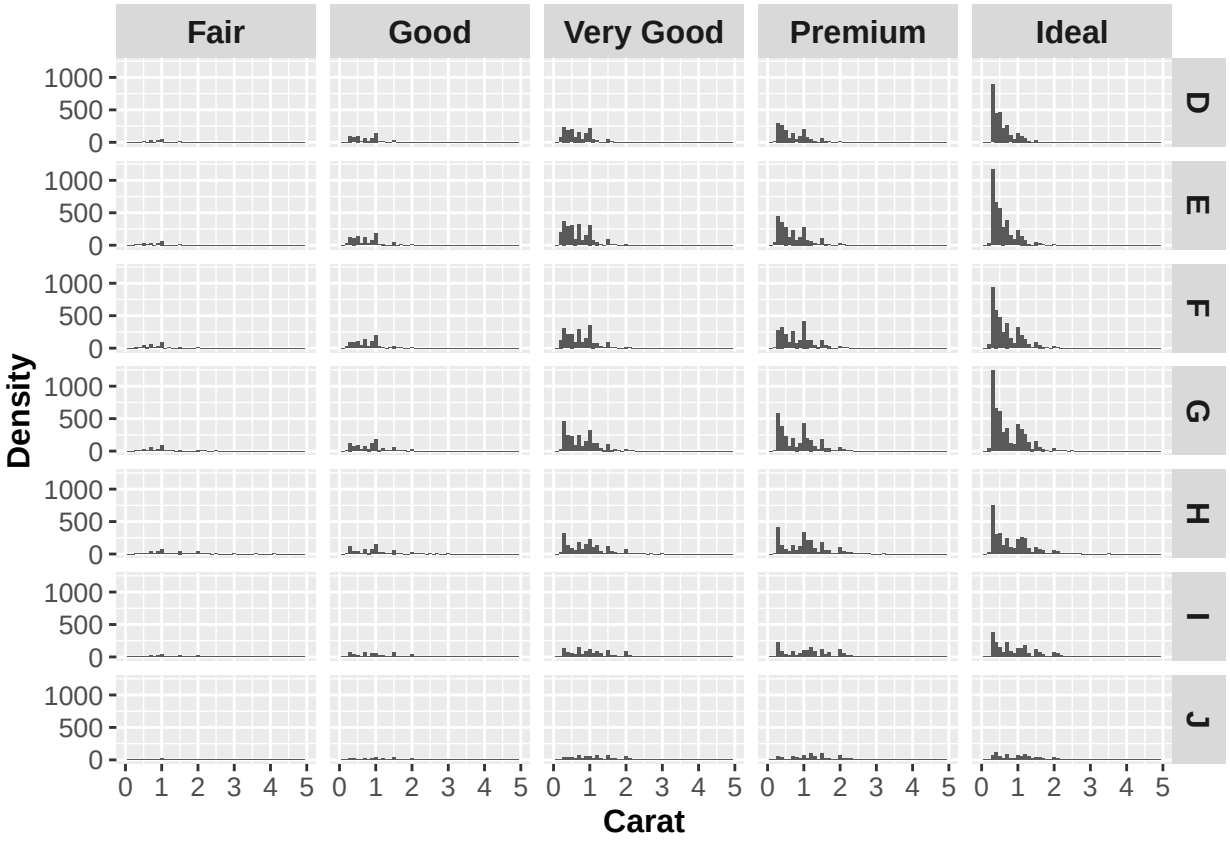
Task 3:To enhance the visibility of density in the histogram plot using qplot, adjust the x-axis to label "Carat" with limits set from 0 to 5 carats and breaks at every 0.5 carats. For the y-axis labeled "Density," ensure breaks are set at every 200 units. Improve clarity with axis text size set to 10 for both axes, and emphasize axis titles in size 12 and bold. Facet labels are also set to size 12 and bold for clear identification across diamond color and cu

```
library(ggplot2)

qplot(carat, data = diamonds, facets = color ~ cut, geom = "histogram", binwidth = 0.1) +
  scale_x_continuous(name = "Carat", limits = c(0, 5), breaks = seq(0, 5, by = 1)) +
  scale_y_continuous(name = "Density", breaks = seq(0, 1000, by = 500)) +
  theme(
    axis.text.x = element_text(size = 10),
    axis.text.y = element_text(size = 10),
    axis.title.x = element_text(size = 12, face = "bold"),
    axis.title.y = element_text(size = 12, face = "bold"),
    strip.text = element_text(size = 12, face = "bold")
  )
```

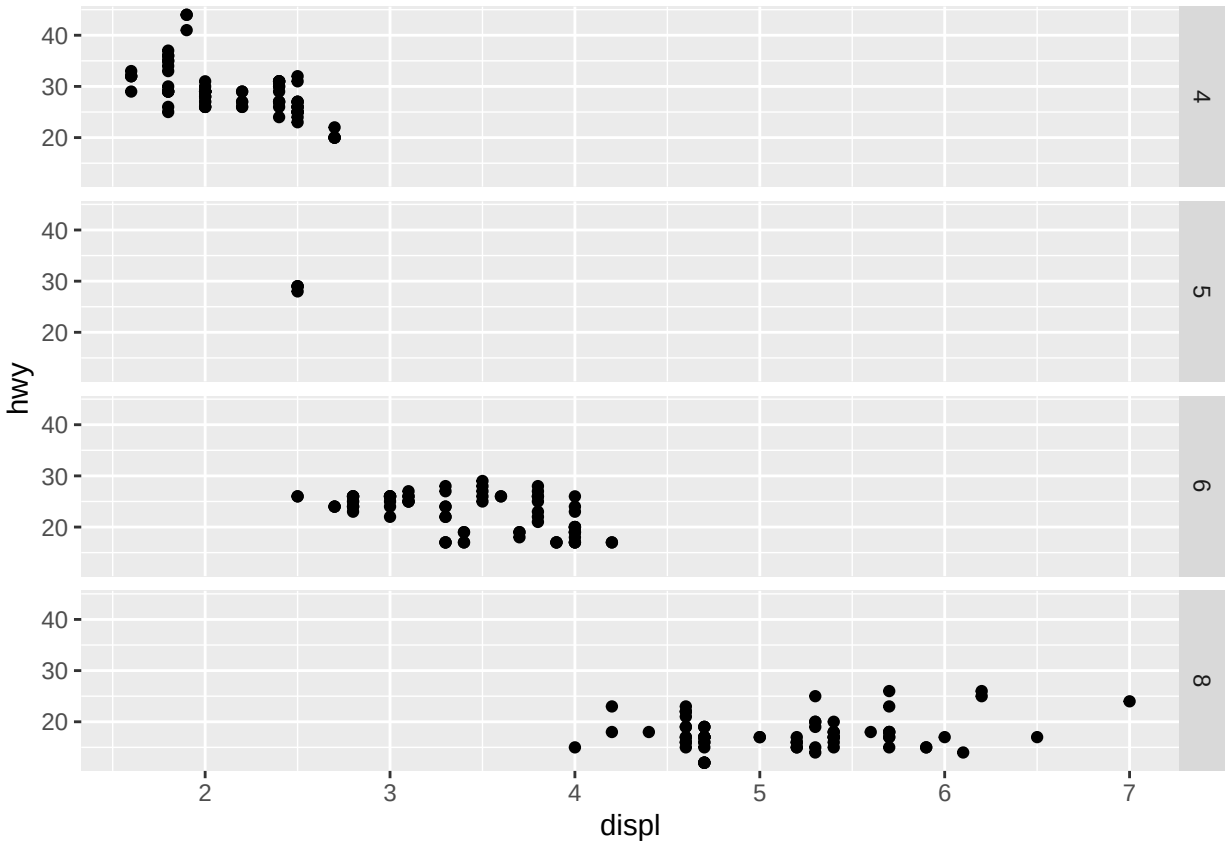
```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```

```
## Warning: Removed 70 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



Task 7:creating a faceted scatter plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset. The data is segmented into facets based on the number of cylinders (cyl), with each facet showing data points corresponding to different cylinder counts.

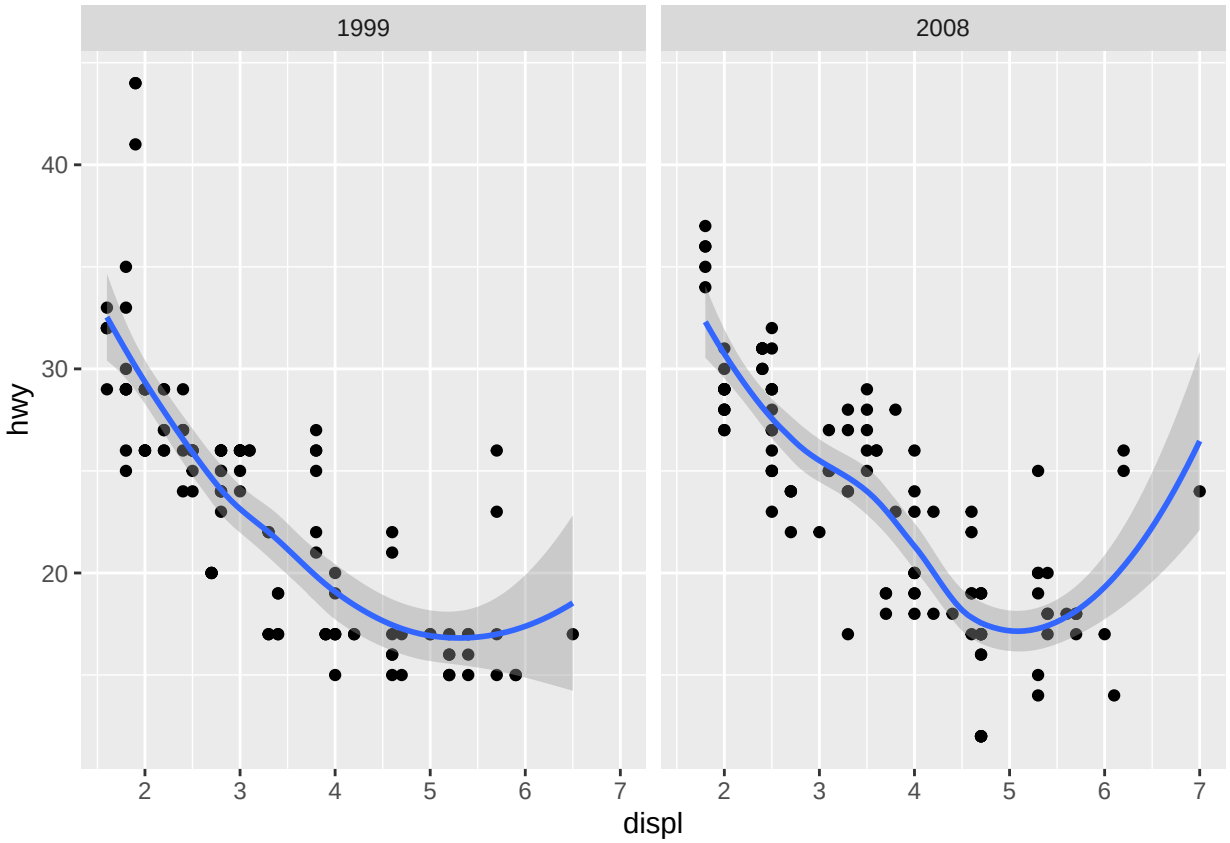
```
qplot(displ, hwy, data = mpg, facets = factor(cyl) ~.)
```



Task 8:creating a faceted scatter plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset. The data is segmented into facets based on the variable year, and each facet includes a smoothed trend line (geom_smooth()) to visualize trends over time.

```
qplot(displ, hwy, data=mpg, facets = . ~ year) + geom_smooth()
```

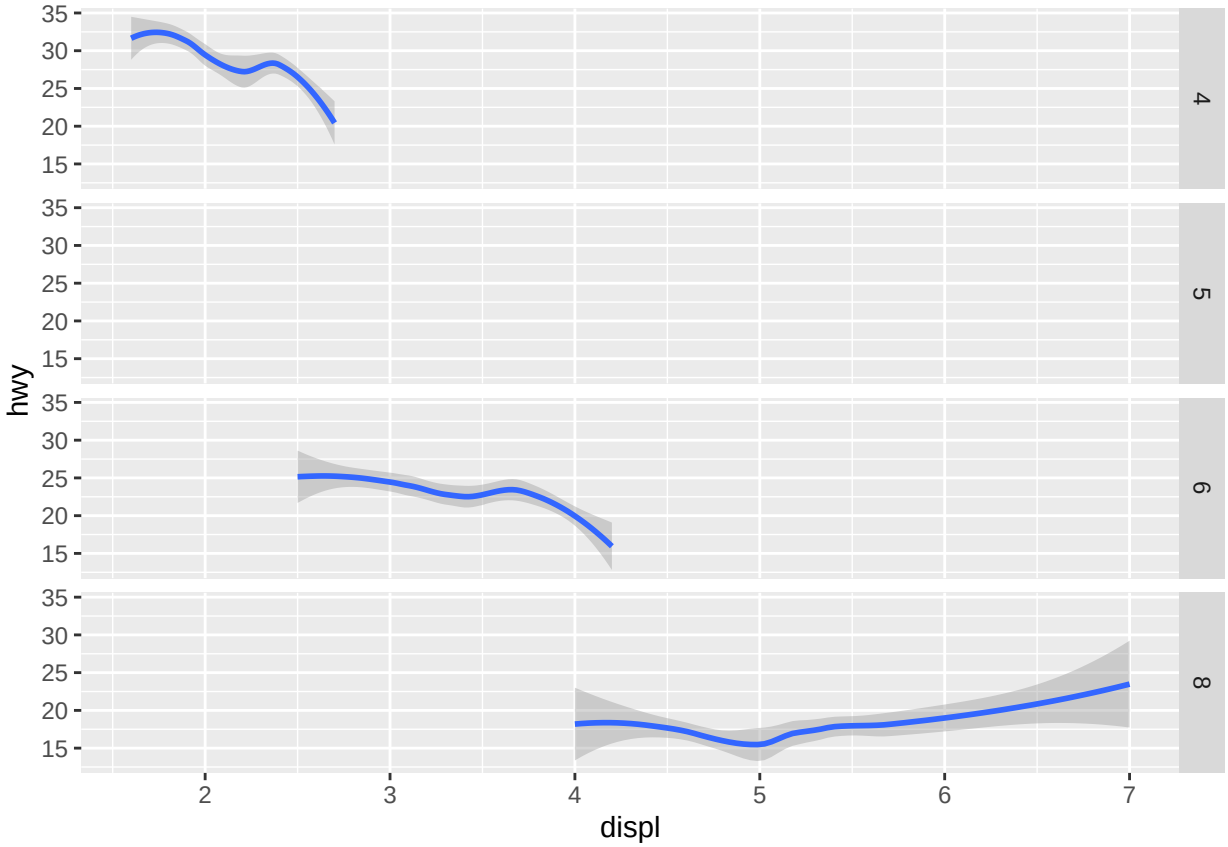
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Task 9:creating a faceted scatter plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset. The data is segmented into facets based on the number of cylinders (cyl), with each facet displaying a smoothed trend line (geom="smooth").

```
qplot(displ, hwy, data = mpg, facets = factor(cyl) ~., geom="smooth")
```

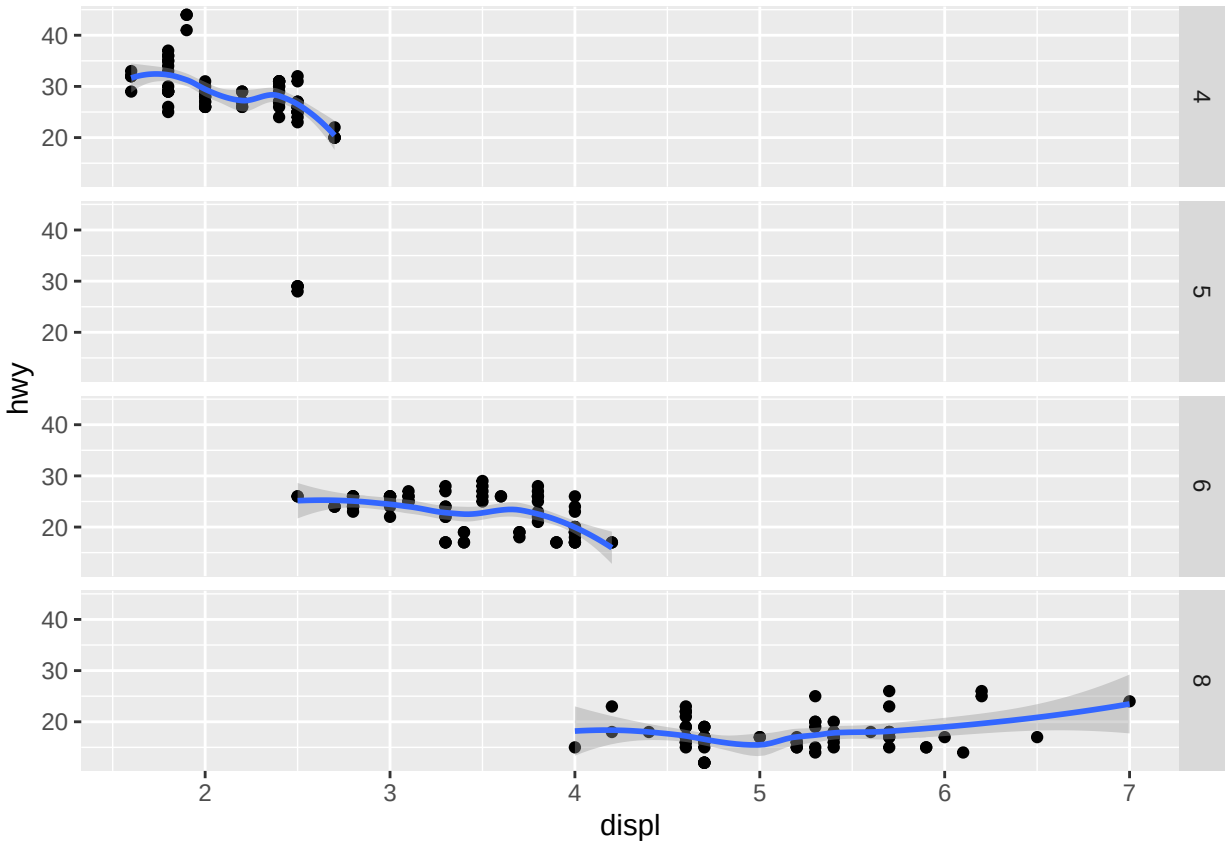
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Task 10:creating a faceted plot of engine displacement (displ) vs highway miles per gallon (hwy) from the mpg dataset. The data is segmented into facets based on the number of cylinders (cyl), and each facet includes both points and a smoothed trend line (geom=c("point","smooth")).

```
qplot(displ, hwy, data = mpg, facets = factor(cyl) ~., geom=c("point","smooth"))
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```


hwy
displ
4
5
6
8
40
30
20
2
3
4
5
6
7