lab-6(faceting)

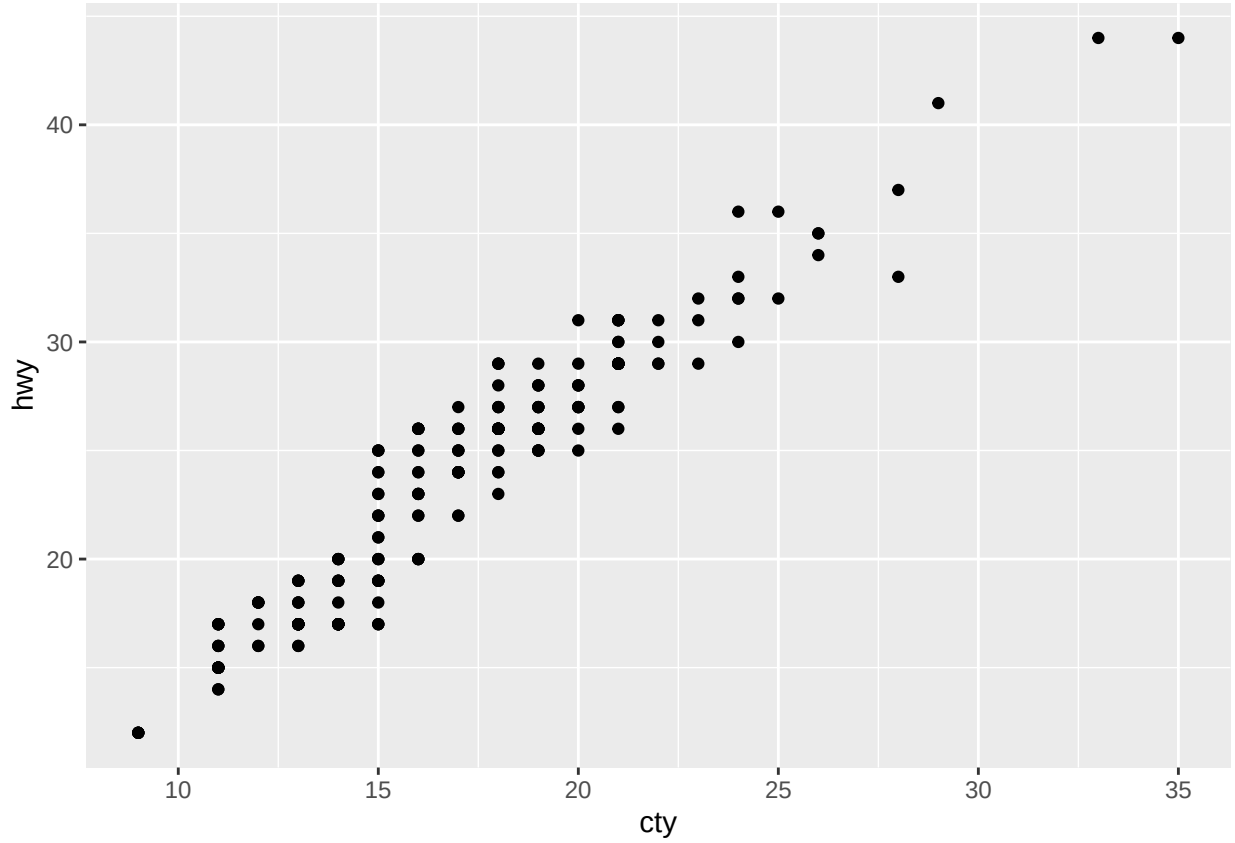
Pradip Basnet

Facet grid

```
library(ggplot2)
```

the code generates sactter plot of cty on the x-axis and hwy on the y-axis however, The facet_grid(.~.) part is unnecessary and doesn't affect the plot since it specifies no faceting.The plot shows the relationship between city and highway miles per gallon for different car models in the mpg dataset

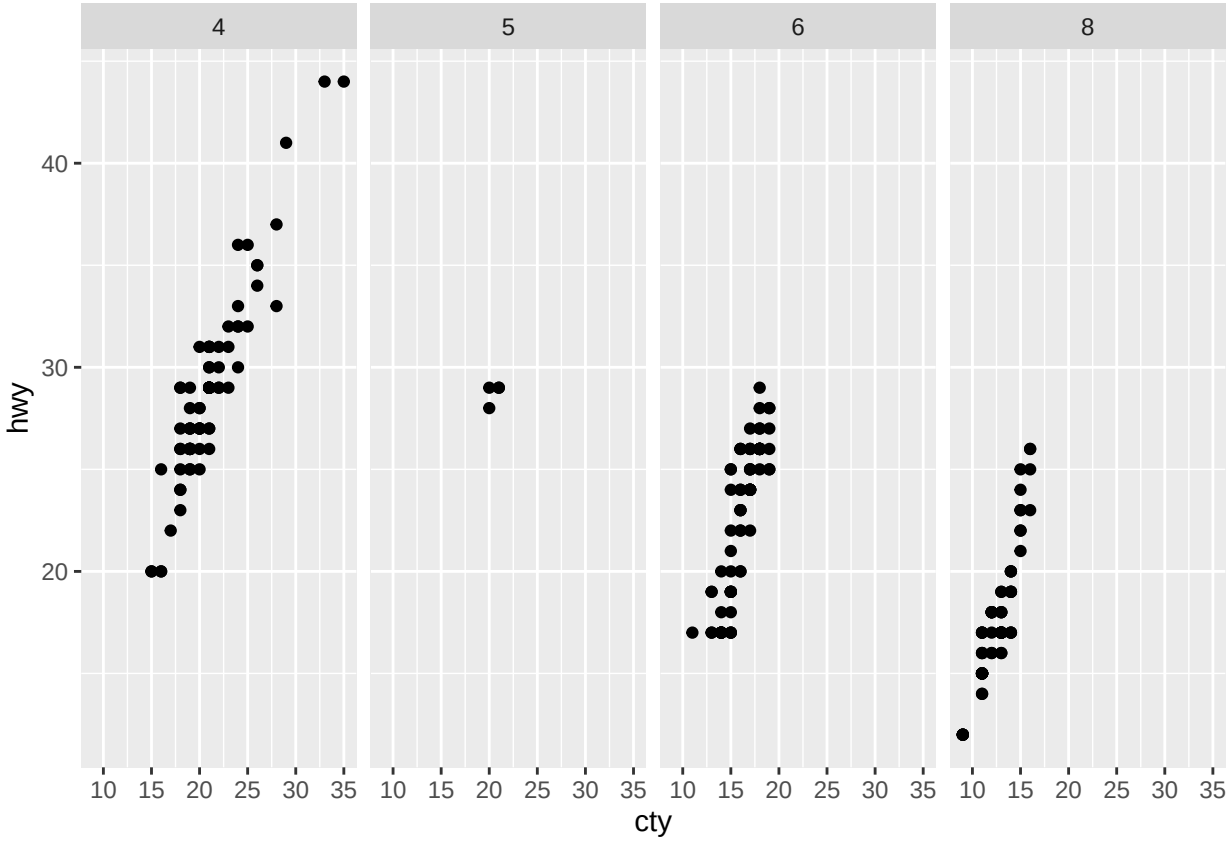
```
ggplot(data = mpg,aes(cty,hwy))+
  geom_point()+
  facet_grid(.~.)
```



code creates a scatter plot of the mpg dataset, where the cty (city miles per gallon) is plotted on the x-axis, and hwy (highway miles per gallon) is plotted on the y-axis. The distributions of "hwy" values vary across the

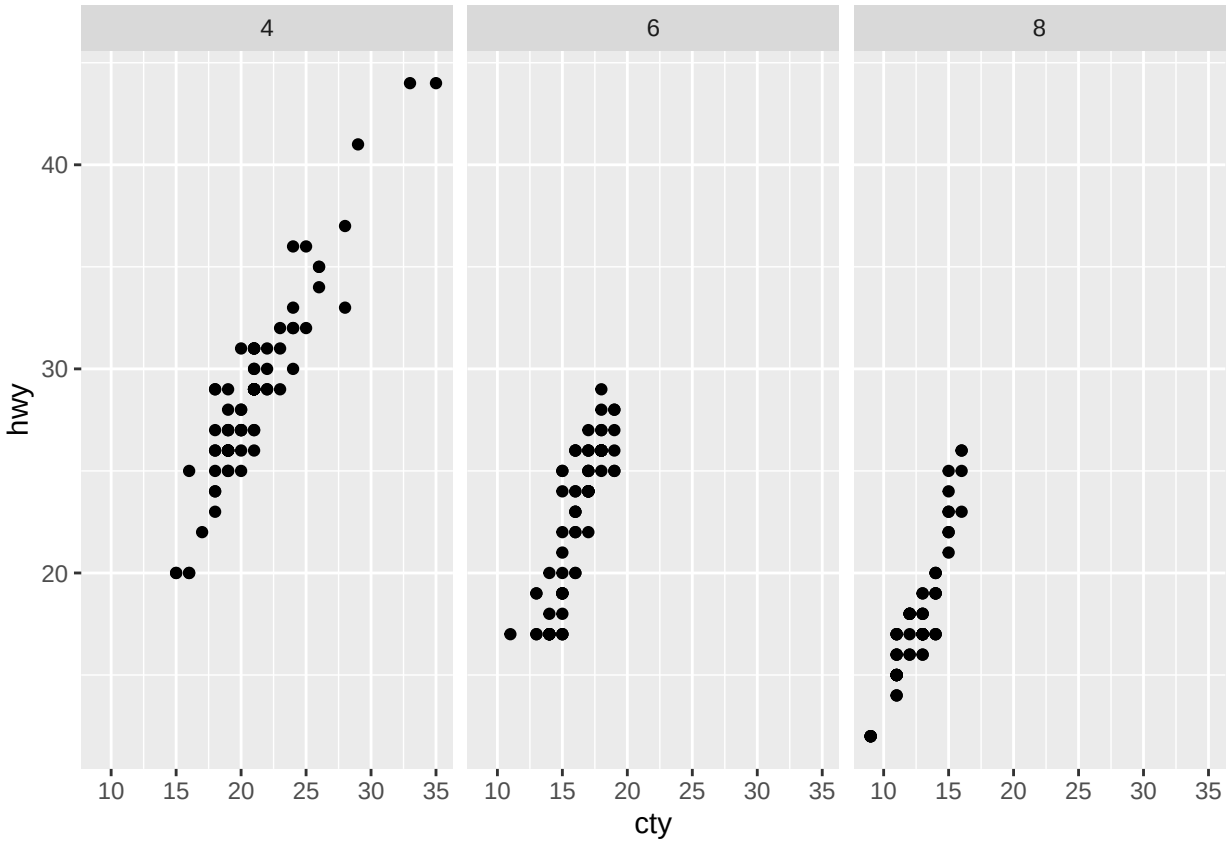
different facets, and there are a few outliers present. This plot can be useful for visualizing the relationship between "hwy" and "city" within each category and for identifying potential patterns or differences between the categories

```
ggplot(data = mpg,aes(cty,hwy))+
  geom_point()+
  facet_grid(.~cyl)
```



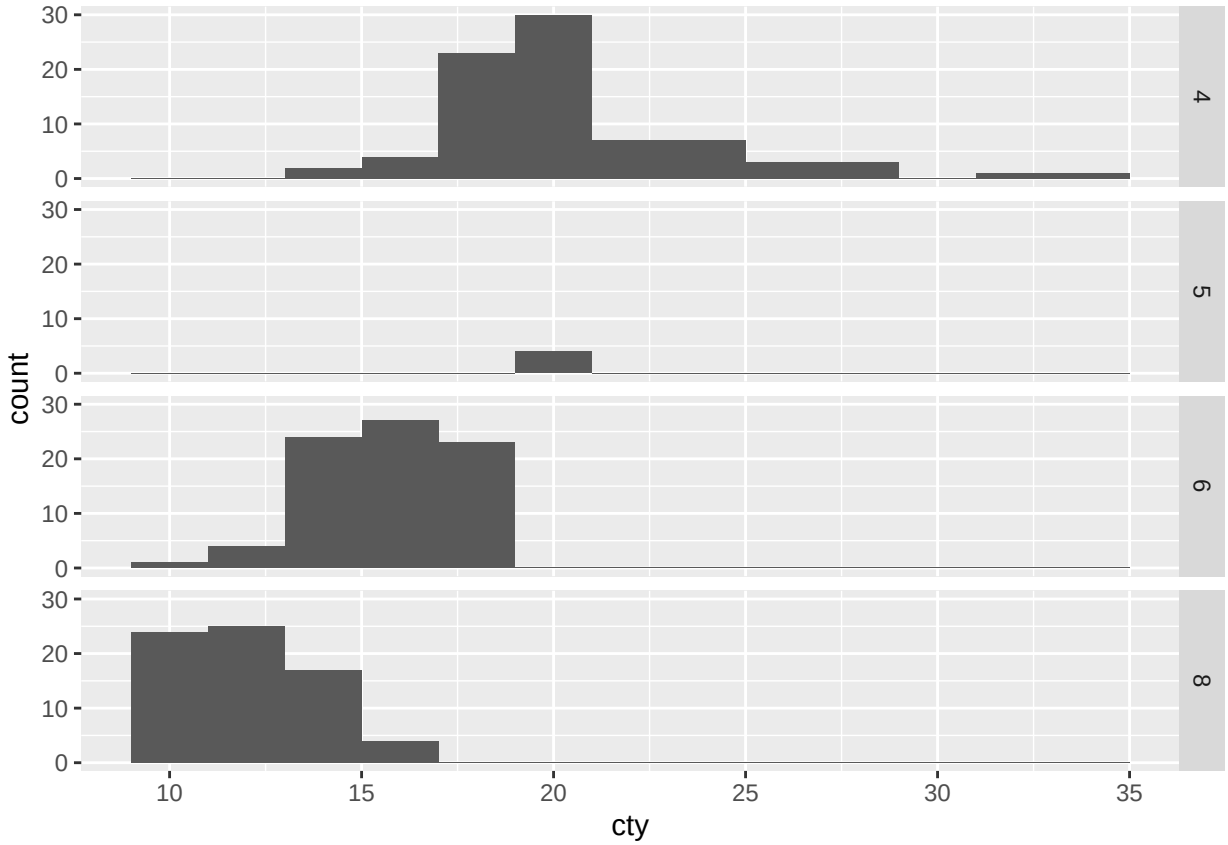
This code generates a scatter plot of city vs. highway miles per gallon using the mpg dataset, excluding cars with 5 cylinders. The plot is split into separate subplots based on the number of cylinders (4, 6, or 8), allowing for a comparison of mileage across different cylinder configurations

```
mpg2 <- subset(mpg, cyl!=5)
ggplot(data = mpg2, aes(cty,hwy))+
  geom_point()+
  facet_grid(.~cyl)
```



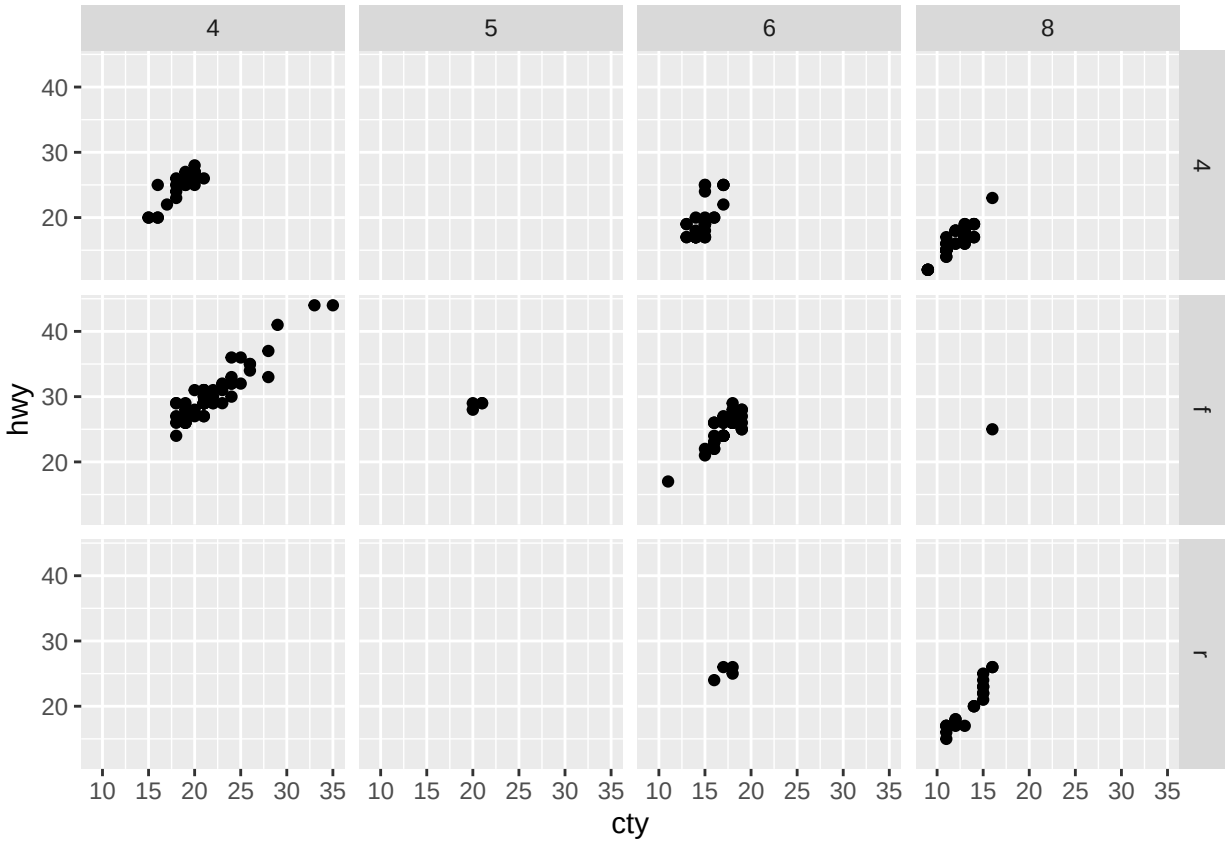
code generates histograms showing the distribution of city miles per gallon (cty) for cars in the mpg dataset, with separate vertical histograms for each cylinder count (cyl). Each histogram displays the frequency of cars within specific cty ranges, using a bin width of 2 miles per gallon plot is a facet grid showing histograms of "city" values for different categories or groups of data. The distributions of "city" values vary across the different facets, and there are no apparent outliers.

```
ggplot(data = mpg, aes(cty))+
  geom_histogram(binwidth = 2)+
  facet_grid(cyl~.)
```



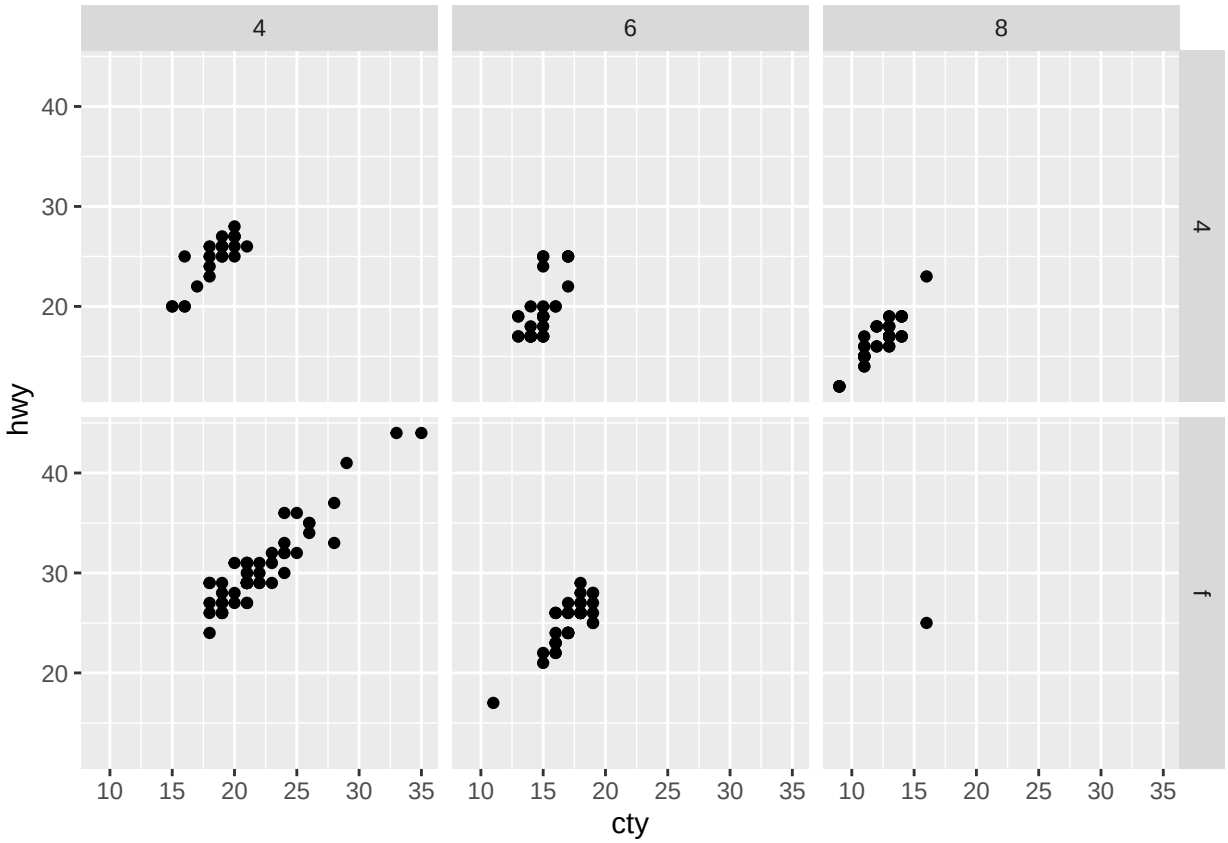
the code generates a scatter plot of city vs. highway miles per gallon (cty vs. hwy) from the mpg dataset. The plot is split into a grid, with separate subplots for each combination of drive type (drv) and cylinder count (cyl) The plot is a facet grid showing scatter plots of "hwy" values against "city" values, likely for different categories or groups of data. The distributions of "hwy" values vary across the different facets, and there are a few outliers present.

```
ggplot(data = mpg, aes(cty,hwy))+
  geom_point()+
  facet_grid(drv~cyl)
```



the given code filters the mpg dataset to exclude cars with 5 cylinders and to include only those with four-wheel or front-wheel drive. It then creates a scatter plot of city vs. highway miles per gallon (cty vs. hwy), with the plot divided into a grid based on drive type (drv) and cylinder count (cyl). Each subplot represents a specific combination of drive type and cylinder count

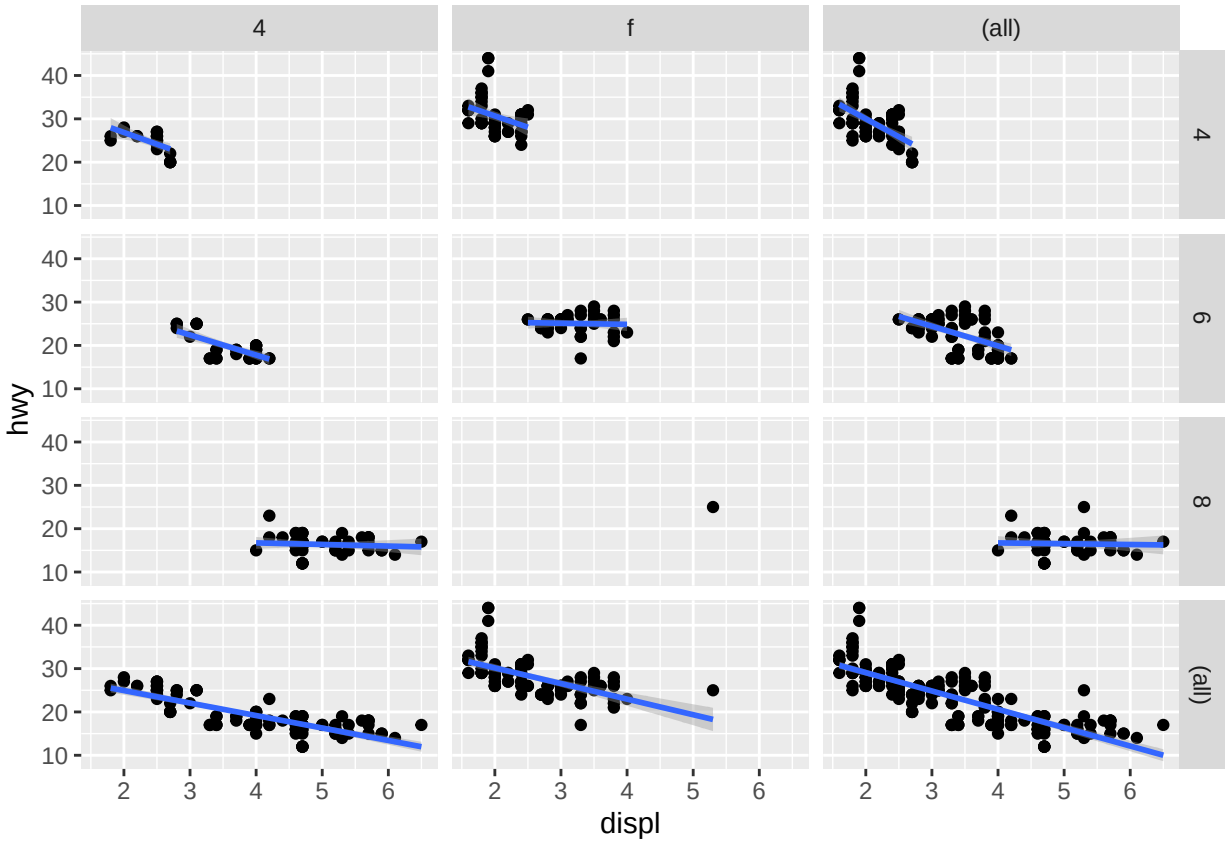
```
mpg1 <- subset(mpg, cyl != 5 & drv %in% c("4", "f"))
ggplot(data = mpg1,aes(cty,hwy))+
  geom_point()+
  facet_grid(drv~cyl)
```



This code filters the mpg dataset to include only cars without 5 cylinders and with either four-wheel or front-wheel drive. It then creates a faceted scatter plot of engine displacement (displ) vs. highway miles per gallon (hwy), adding linear regression lines to show the trend. The plot is divided into a grid where each subplot represents a specific combination of cylinder count (cyl) and drive type (drv). Additionally, it includes marginal plots that summarize the trends across all cylinder counts and drive types

```
mpg2 <- subset(mpg, cyl!=5 & drv %in% c("4","f"))
p <- ggplot(data = mpg2, aes(displ, hwy)) + geom_point() + geom_smooth(method = "lm")
p + facet_grid(cyl ~ drv, margins = T)
```

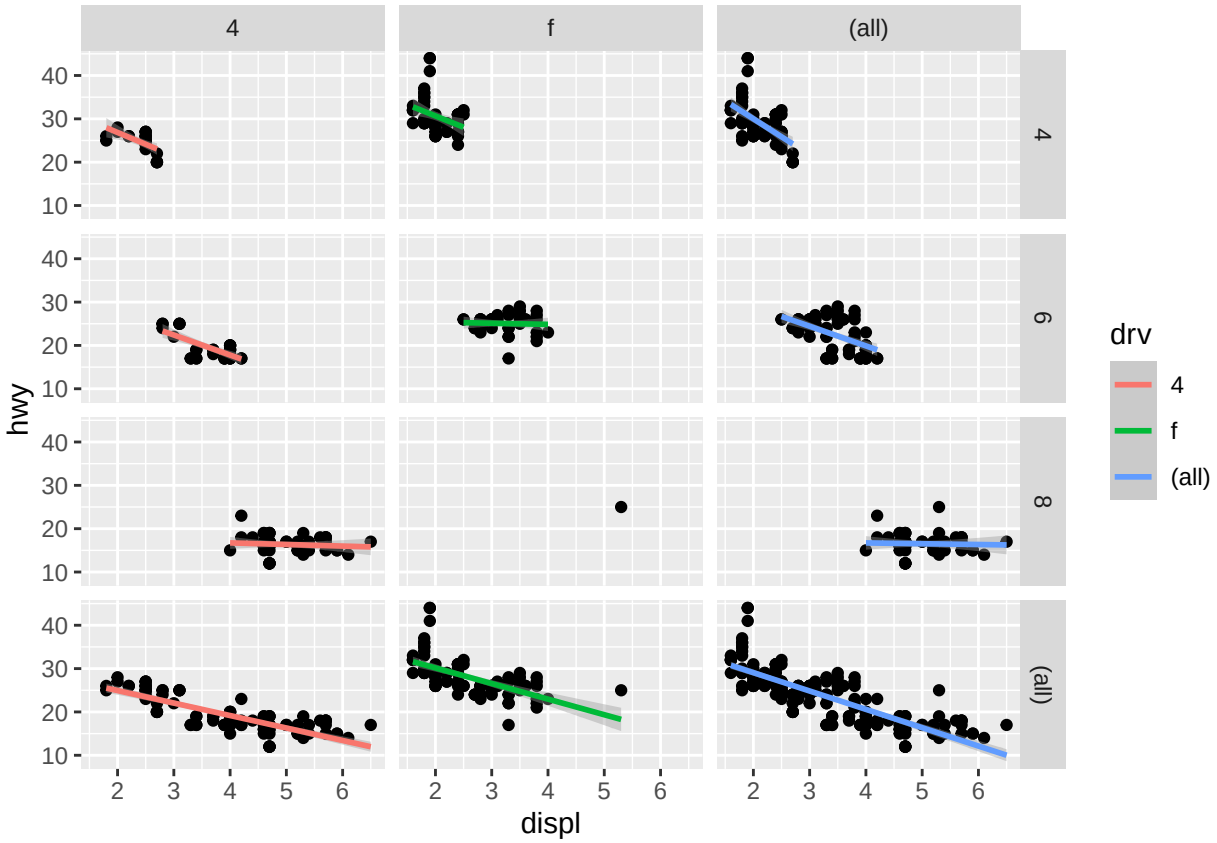
```
## `geom_smooth()` using formula = 'y ~ x'
```



This code filters the mpg dataset to exclude cars with 5 cylinders and to include only those with four-wheel or front-wheel drive. It then creates a faceted scatter plot of engine displacement (displ) vs. highway miles per gallon (hwy), adding color-coded linear regression lines for each drive type (drv). The plot is divided into a grid based on cylinder count (cyl) and drive type, with marginal plots that summarize trends across all cylinder counts and drive types

```
mpg2 <- subset(mpg, cyl != 5 & drv %in% c("4", "f"))
p <- ggplot(data = mpg2, aes(displ, hwy)) + geom_point() + geom_smooth(aes(colour = drv), method = "lm")
p + facet_grid(cyl ~ drv, margins = T)
```

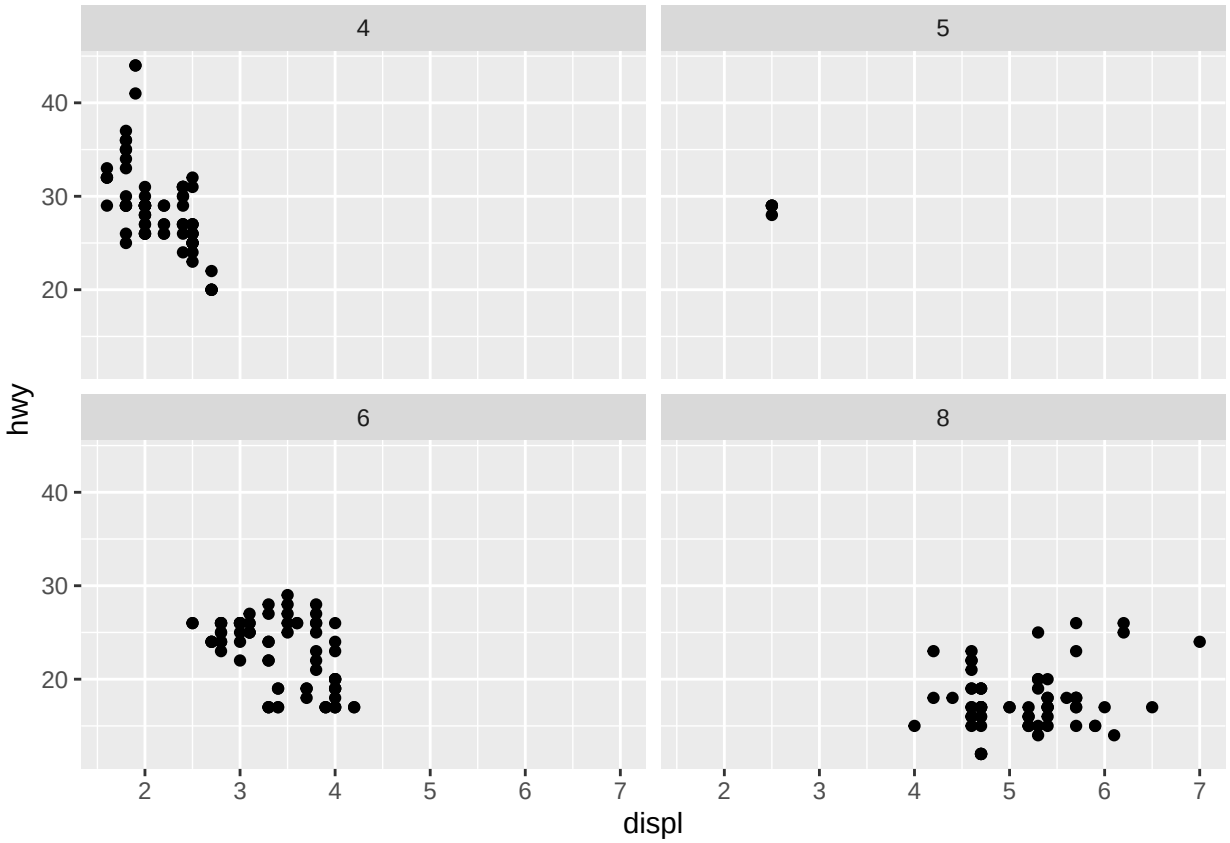
```
## `geom_smooth()` using formula = 'y ~ x'
```



Facet_wrap

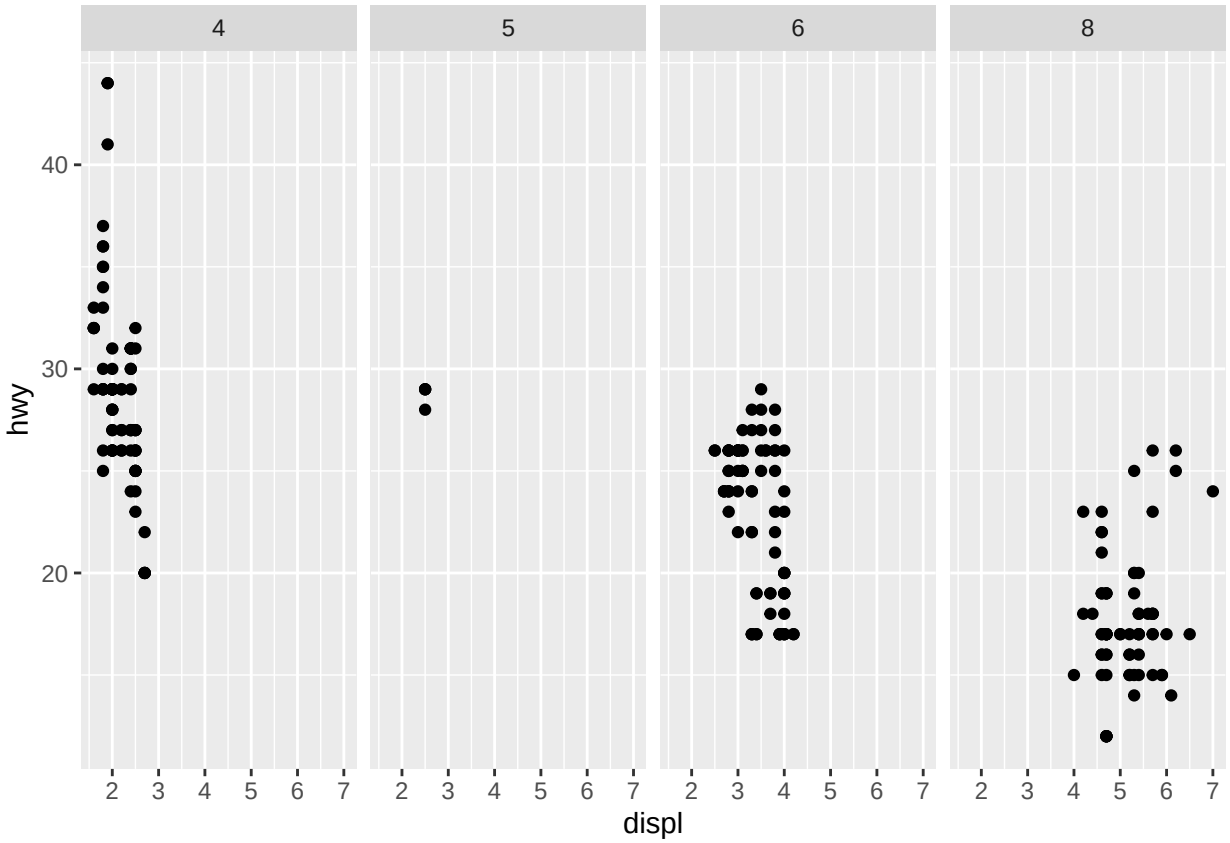
This code generates a scatter plot of engine displacement (displ) vs. highway miles per gallon (hwy) using the mpg dataset. The plot is split into multiple facets, with each facet representing a different number of cylinders (cyl).

```
p <- ggplot(data = mpg, aes(x = displ, y = hwy)) + geom_point()
p + facet_wrap(~cyl)
```



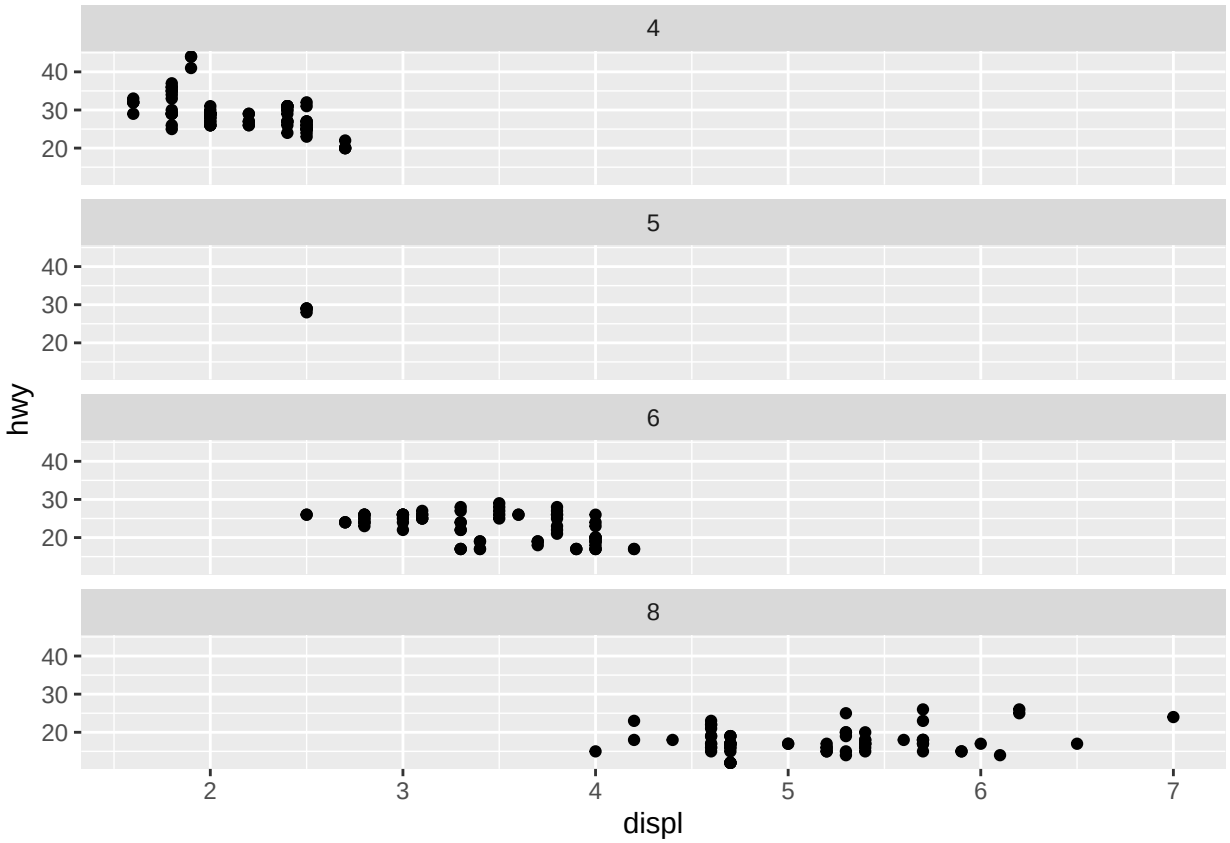
This code modifies the faceted scatter plot to arrange the facets in a single row.This code generates a scatter plot of engine displacement vs. highway miles per gallon, divided into facets based on the number of cylinders. The facets are arranged in a single row, so all plots are displayed horizontally in one continuous line.

```
p + facet_wrap(~cyl, nrow = 1)
```



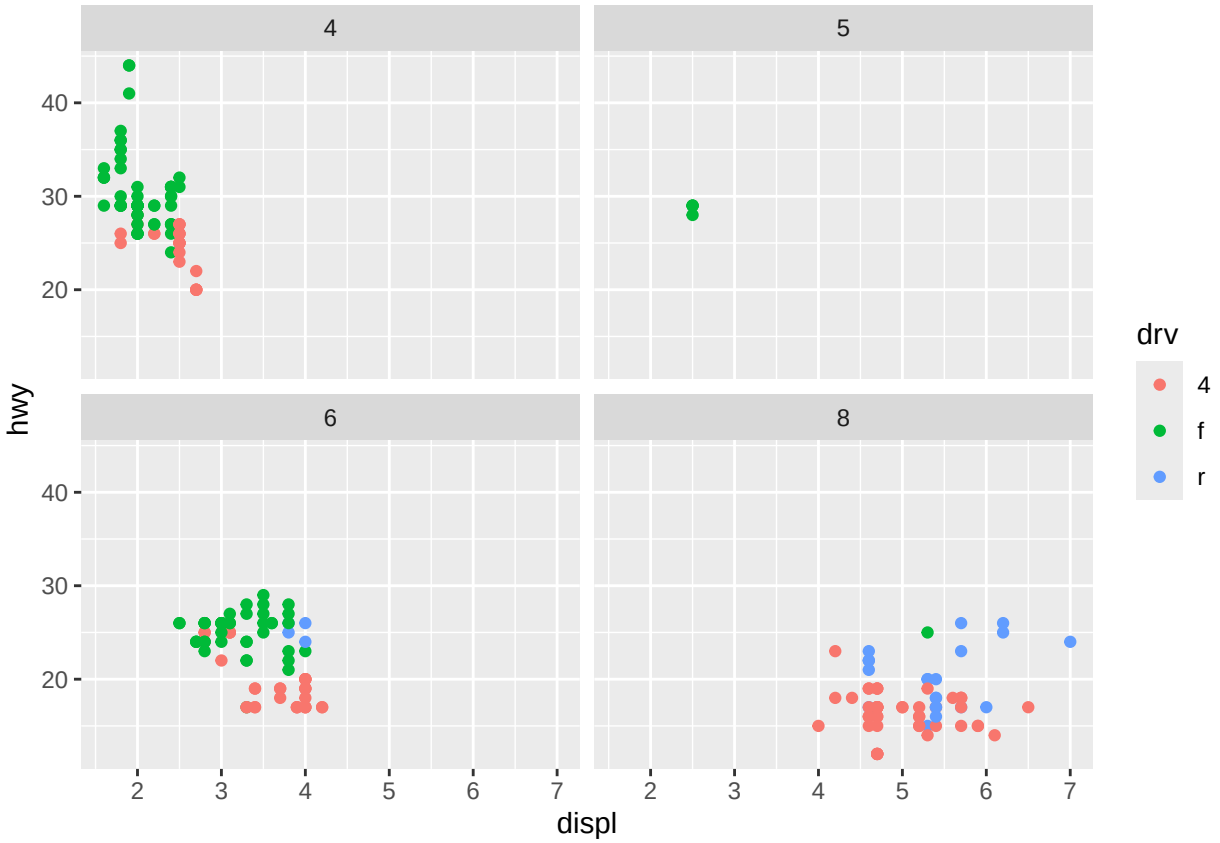
This code generates a scatter plot of engine displacement vs. highway miles per gallon, divided into facets based on the number of cylinders. The facets are arranged in a single column, so all plots are displayed vertically in one continuous line

```
p + facet_wrap(~cyl, ncol = 1)
```



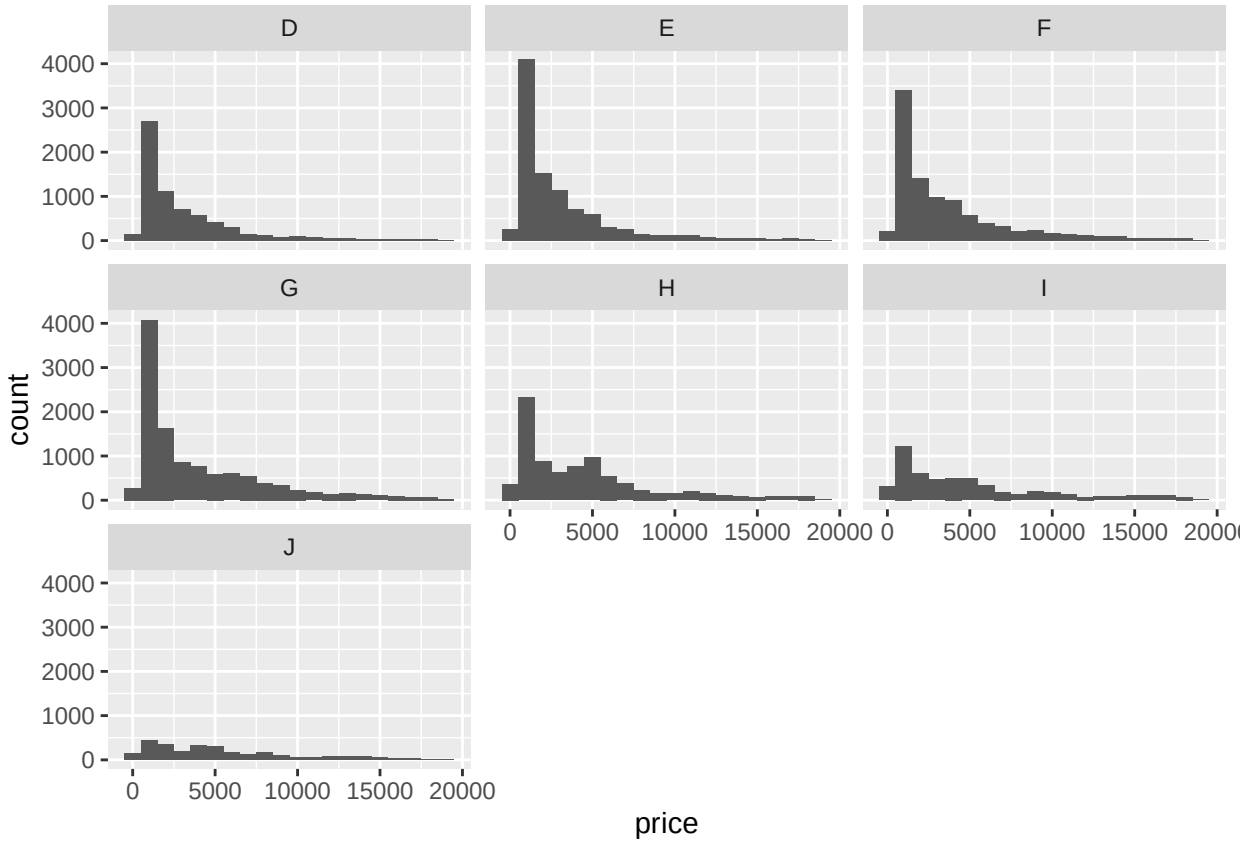
This code creates a faceted scatter plot of engine displacement (displ) vs. highway miles per gallon (hwy) from the mpg dataset. The facets are organized by the number of cylinders (cyl), and within each facet, points are colored according to the drive type (drv).

```
p <- ggplot(data = mpg, aes(x = displ, y = hwy, color = drv)) + geom_point()
p + facet_wrap(~cyl)
```



This code generates a faceted histogram of diamond prices from the diamonds dataset. Each facet represents a different diamond color, and within each facet, the histogram shows the distribution of prices for that color. The binwidth = 1000 argument groups prices into bins of 1000 units, allowing for an examination of how price distributions differ by diamond colo

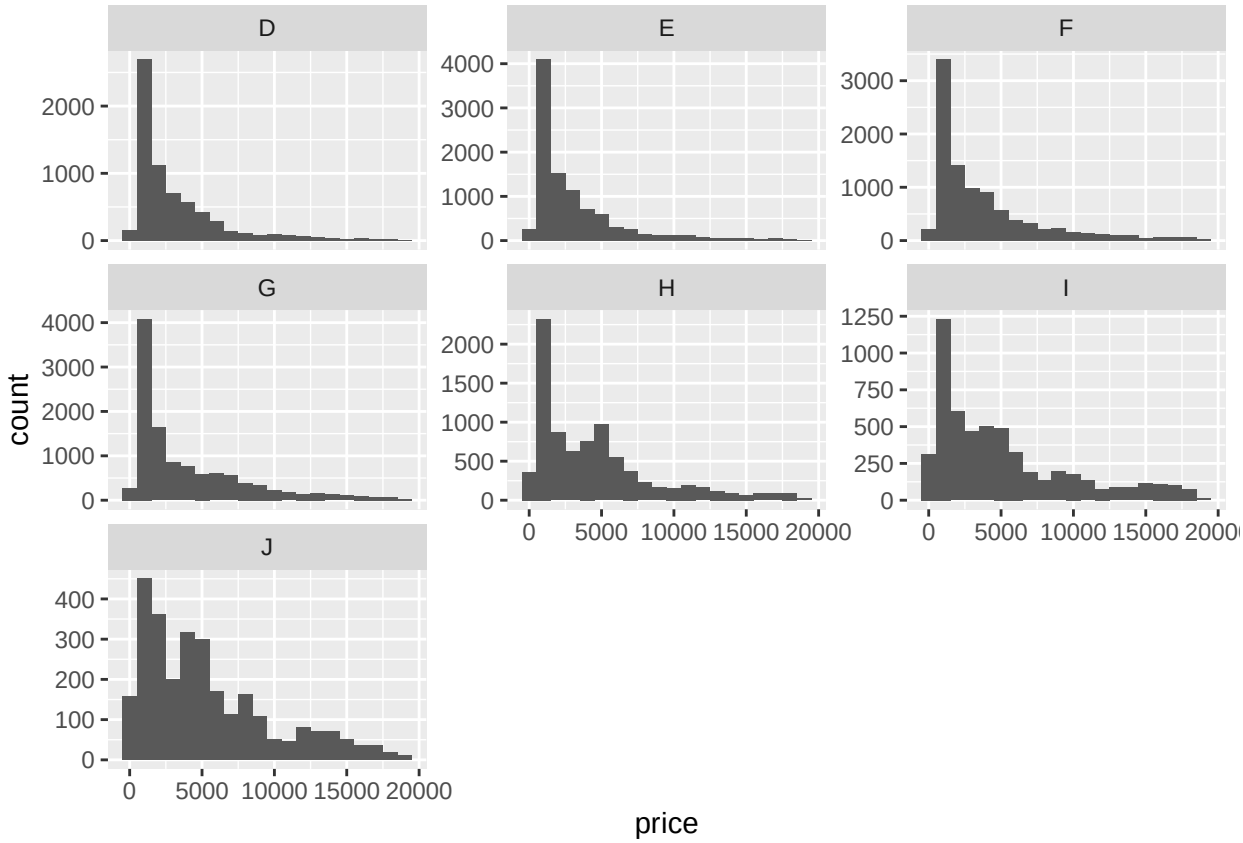
```
p <- ggplot(data = diamonds,aes(x=price))+
  geom_histogram(binwidth = 1000)
  p+facet_wrap(~color)
```



Position scales

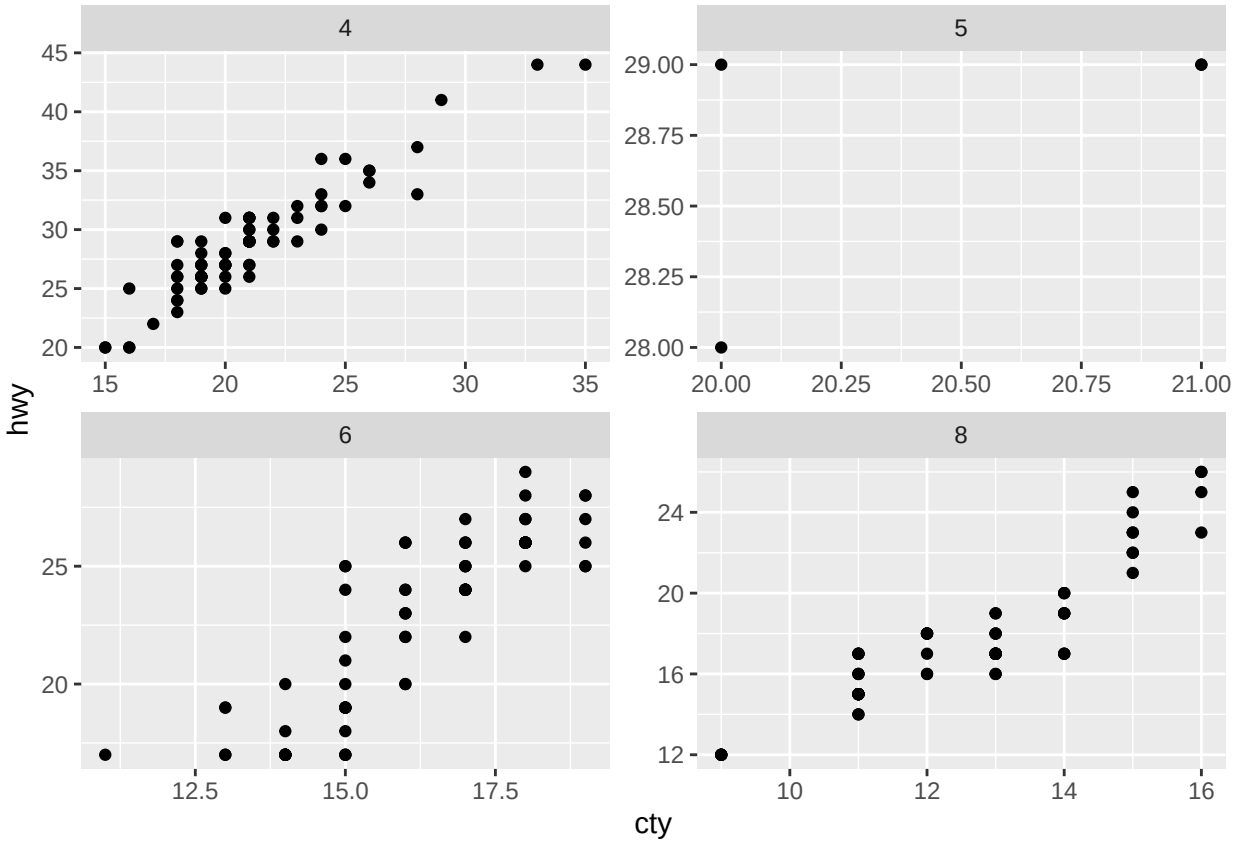
This code creates a faceted histogram of diamond prices, with each facet representing a different diamond color. The scales = "free_y" argument ensures that each facet has its own y-axis scale, allowing for a more detailed and accurate visualization of the distribution of diamond prices across different colors. This is particularly useful when the number of observations for each color varies significantly, as it prevents skewed visual representations due to a fixed y-axis range

```
p + facet_wrap(~color, scales = "free_y")
```



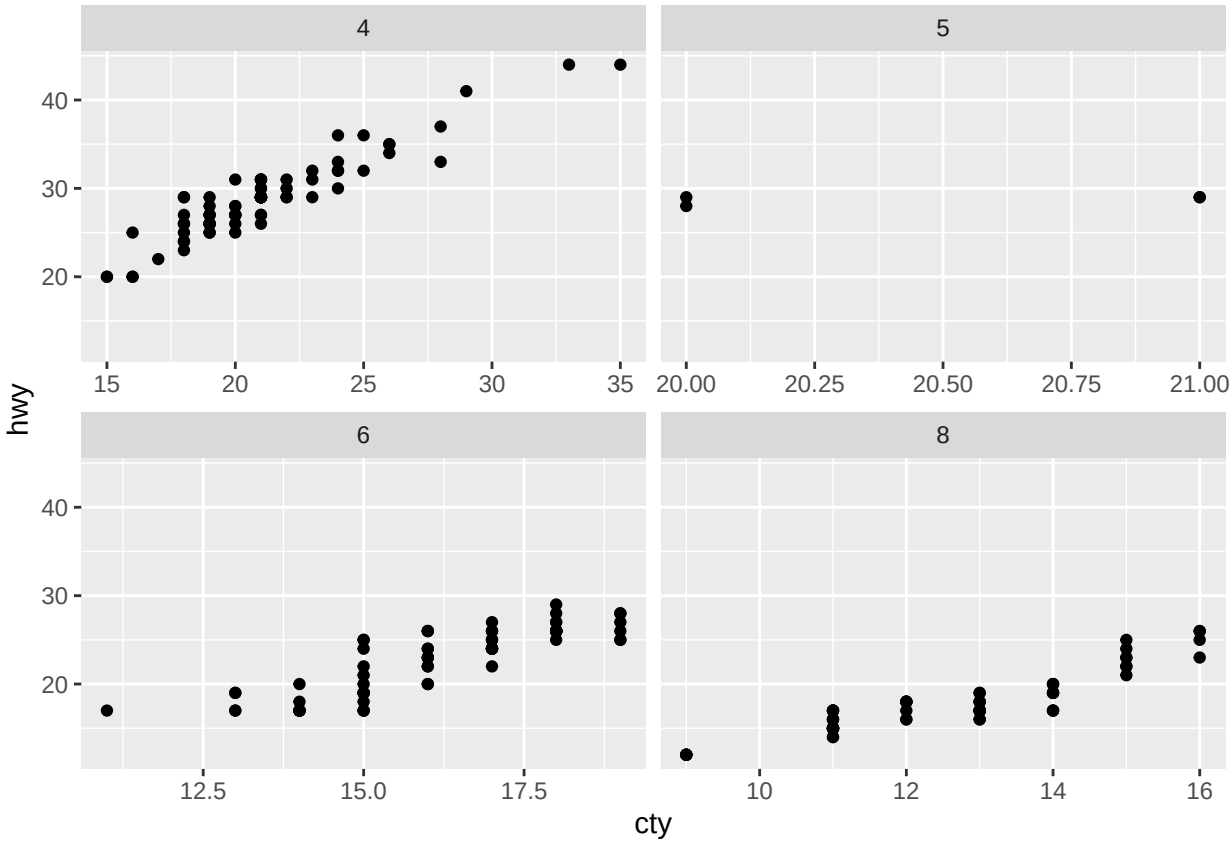
This code generates a scatter plot of city miles per gallon (cty) vs. highway miles per gallon (hwy) from the mpg dataset. The plot is divided into facets based on the number of cylinders (cyl), and each facet has its own x-axis and y-axis scales.

```
ggplot(data = mpg, aes(cty, hwy)) + geom_point() + facet_wrap(~ cyl, scales = "free")
```



This code generates a scatter plot of city miles per gallon (cty) vs. highway miles per gallon (hwy) from the mpg dataset. The plot is divided into facets based on the number of cylinders (cyl), with each facet having its own x-axis scale (city mileage) but sharing the same y-axis scale (highway mileage).

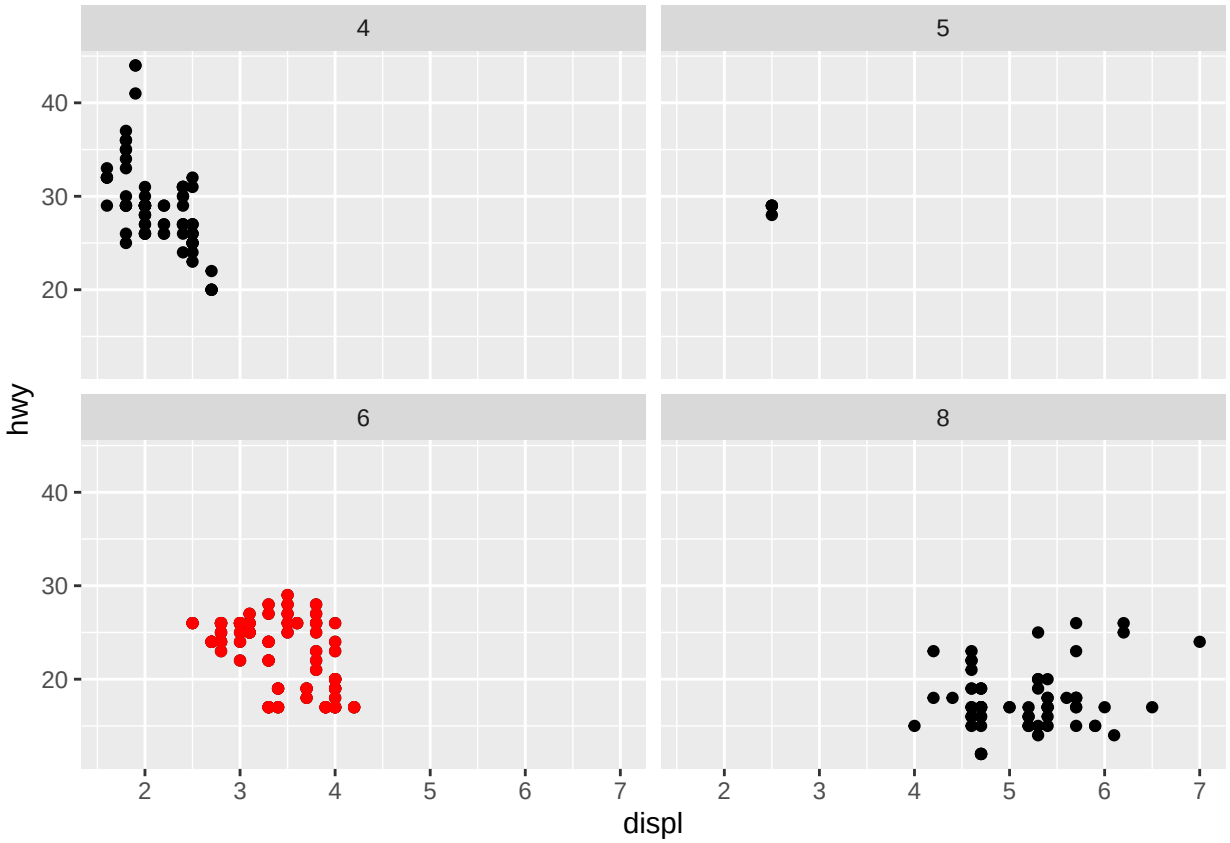
```
ggplot(data = mpg, aes(cty, hwy)) + geom_point() + facet_wrap(~ cyl, scales = "free_x")
```



Decorating facets

This code generates a faceted scatter plot of engine displacement (displ) vs. highway miles per gallon (hwy), divided by cylinder count (cyl). It then overlays additional red points representing cars with 6 cylinders on this plot, highlighting their distribution and allowing for a visual comparison with cars of other cylinder counts

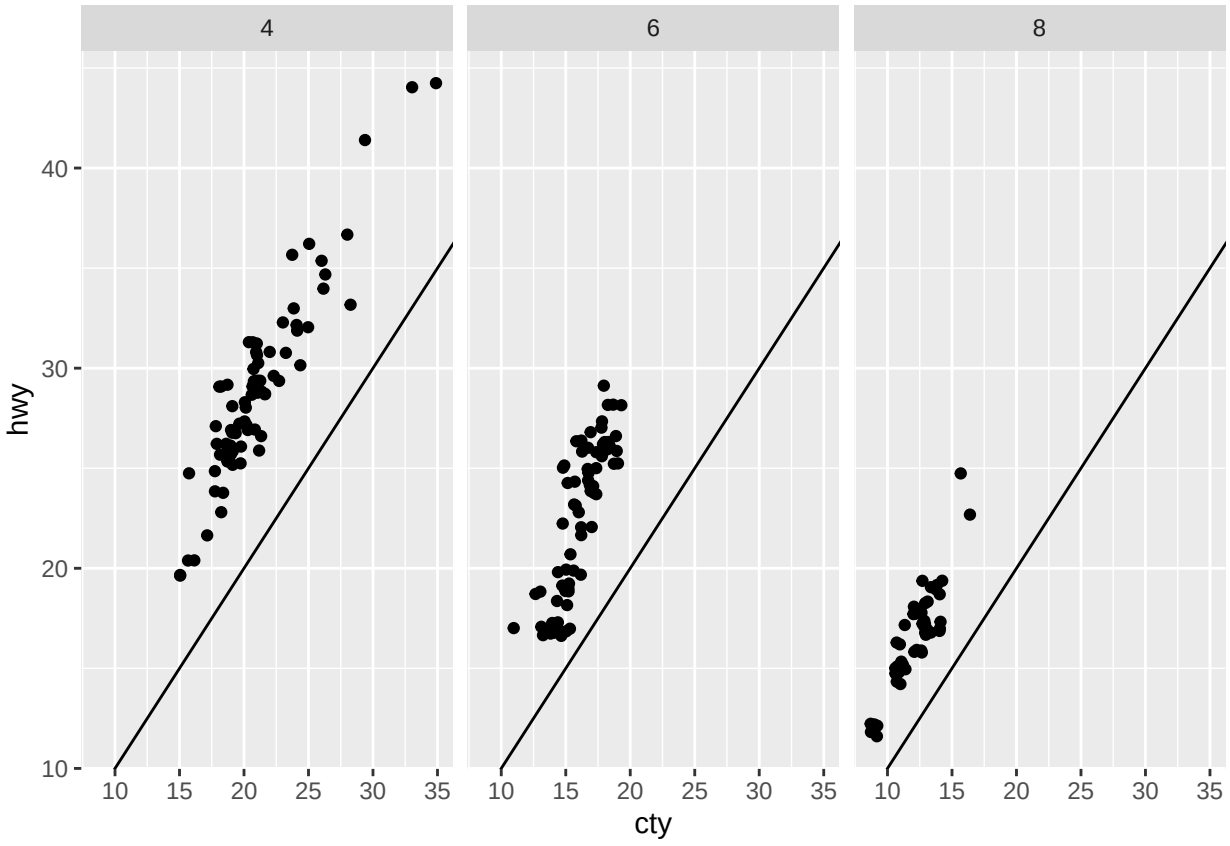
```
p <- ggplot(data = mpg, aes(x = displ, y = hwy)) + geom_point()
q <- p + facet_wrap(~cyl)
cyl6 <- subset(mpg, cyl == 6)
q + geom_point(data = cyl6, color = "red")
```



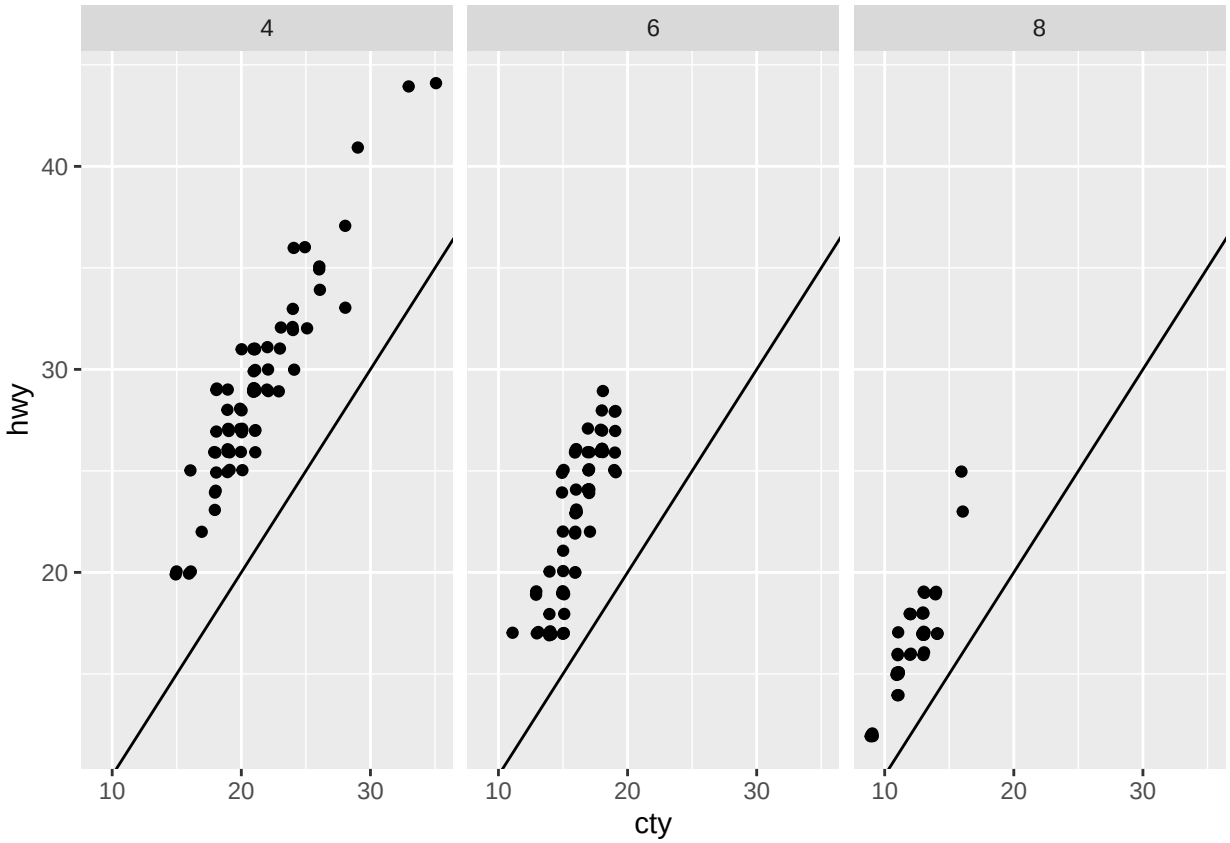
Faceting with abline and jitter

This code generates a faceted scatter plot of city miles per gallon versus highway miles per gallon using the mpg2 dataset, with each facet representing a different cylinder count. The plot includes a reference diagonal line (slope of 1) and jittered points to enhance visibility and reduce overplotting

```
p <- ggplot(mpg2, aes(cty, hwy)) +
 geom_abline() +
 geom_jitter()
p + facet_wrap(~cyl)
```

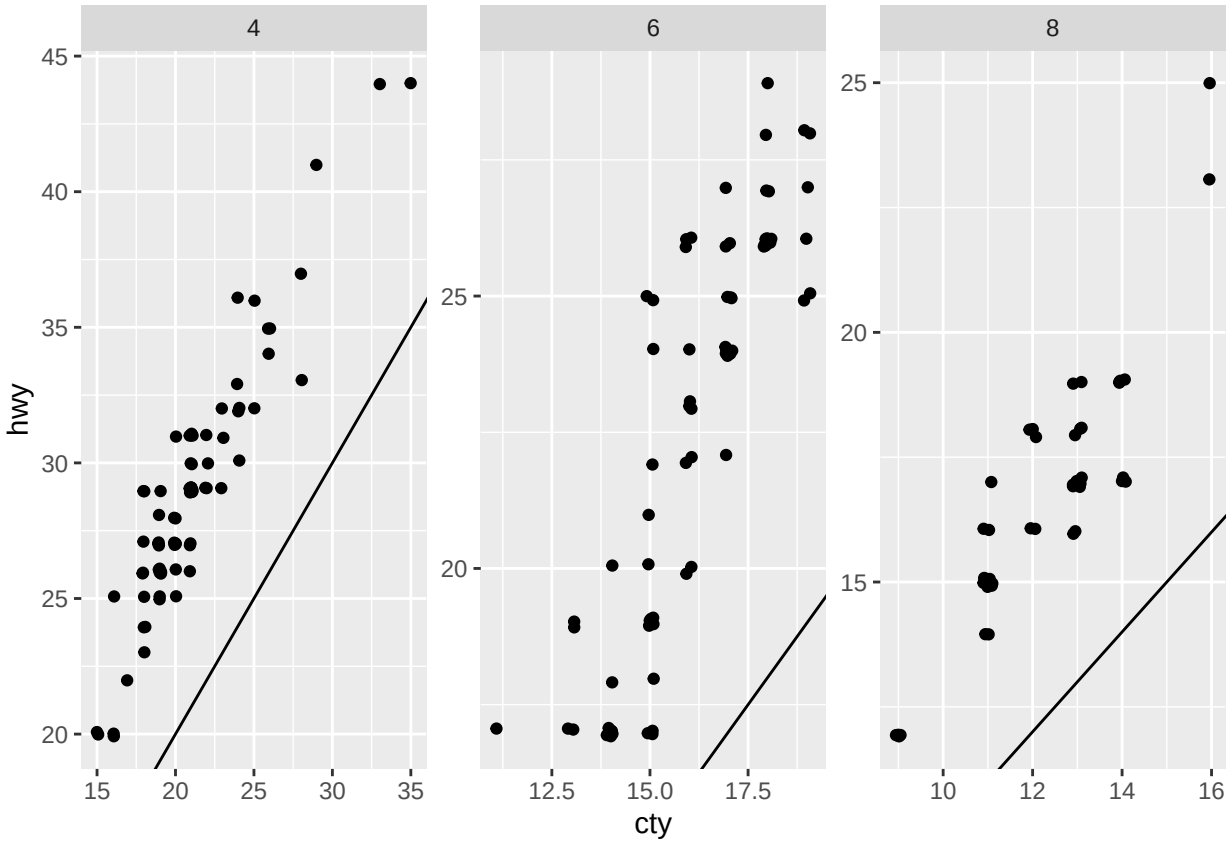



```
p <- ggplot(mpg2, aes(cty, hwy)) +
 geom_abline() +
 geom_jitter(width = 0.1, height = 0.1)
p + facet_wrap(~cyl)
```



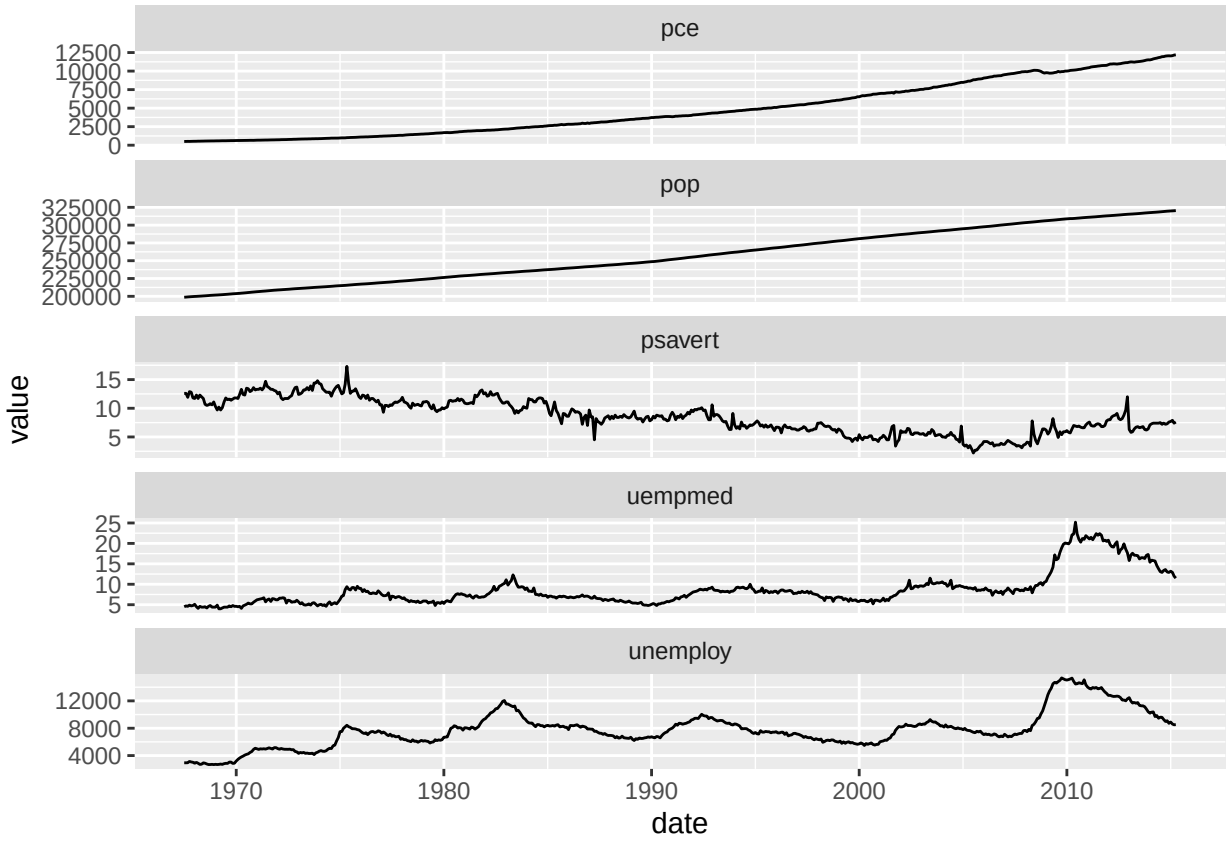
This code generates a faceted scatter plot of city miles per gallon versus highway miles per gallon from the mpg2 dataset. The plot includes a diagonal reference line and jittered points with specified horizontal and vertical jitter to reduce overplotting. Each facet corresponds to a different number of cylinders (cyl)

```
p + facet_wrap(~cyl, scales = "free")
```



This code generates a faceted line plot of various economic variables over time from the economics_long dataset. Each facet represents a different variable and is displayed in a single column. The y-axis scales are adjusted independently for each facet to accurately represent the data, and lines connect data points to show trends over time, The trends in these variables vary, with some showing growth, fluctuations, or cyclical patterns

```
ggplot(economics_long, aes(date, value)) +
 geom_line() +
 facet_wrap(~variable, scales = "free_y", ncol = 1)
```



Facets with boxplot

This code converts the dose column in the ToothGrowth dataset from numeric to a factor, creates a copy of the modified dataset as df, and then displays the first six rows of this modified dataset.

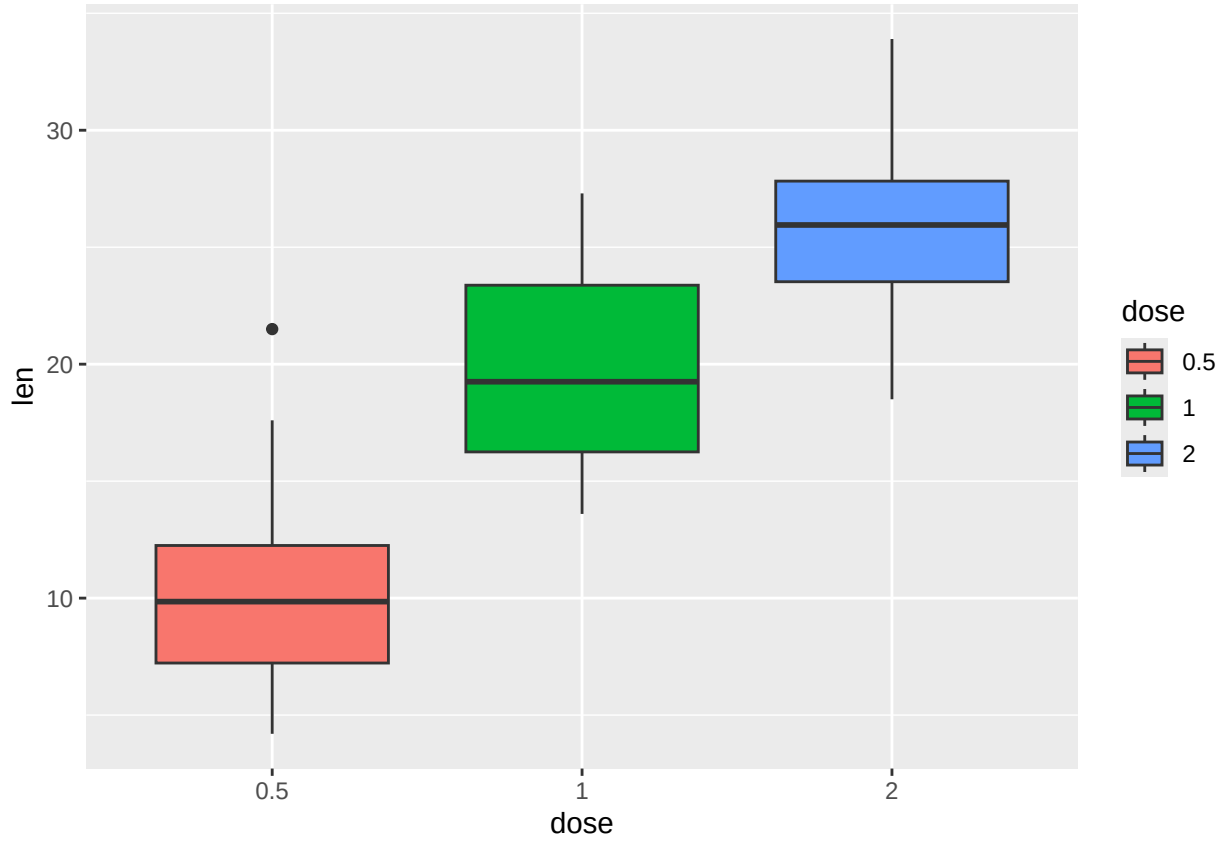
```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
df <- ToothGrowth
head(df)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

Creating a boxplot

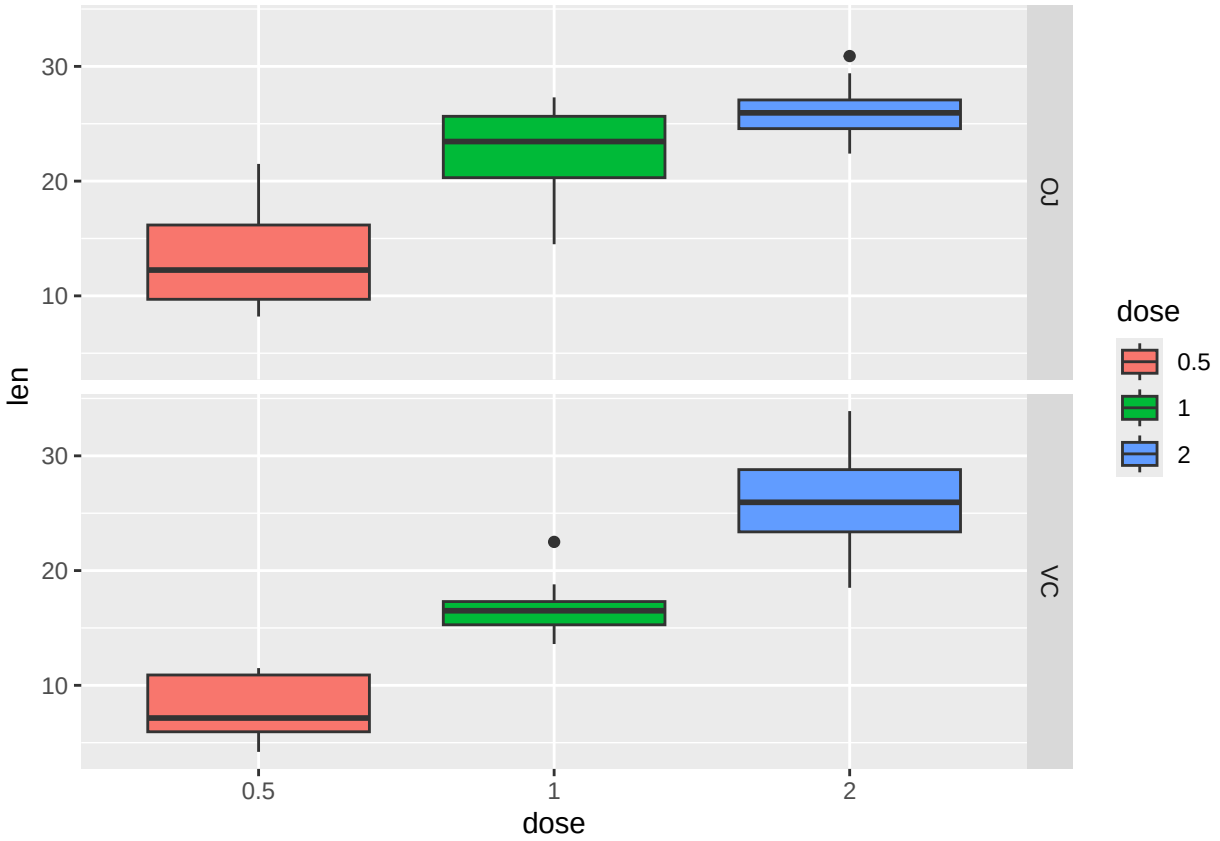
This code creates and displays a boxplot of tooth length (len) against dose (dose) from the df dataset. Each dose level is represented as a separate boxplot, and the boxes are filled with different colors according to the dose levels. The box plot shows the distribution of "len" across different levels of "dose." It reveals that "len" tends to increase with higher doses, and there are some outliers present in the data.

```
bp <- ggplot(df, aes(x=dose, y=len, group=dose)) + geom_boxplot(aes(fill=dose))
print(bp)
```



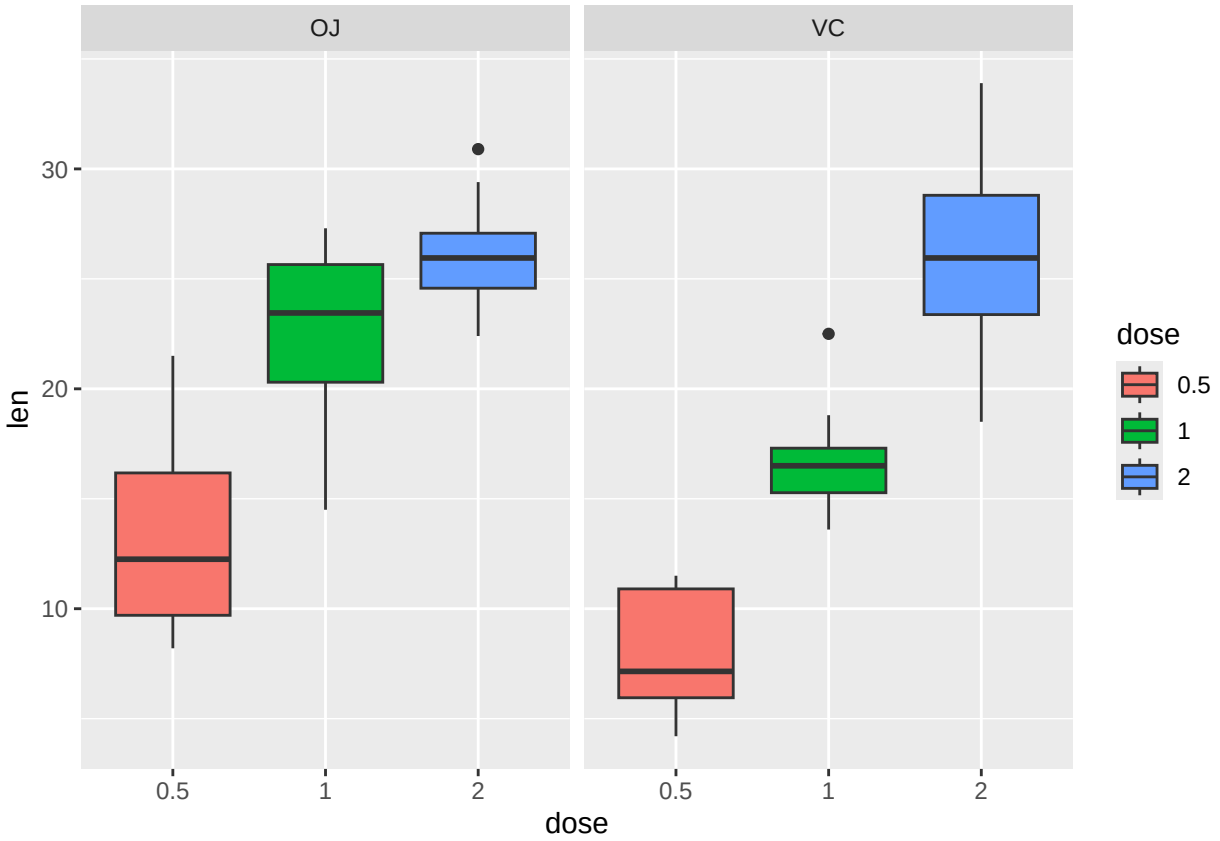
The code adds faceting to the boxplot of tooth length versus dose. The facets are organized into rows based on the supp variable, with each row showing the boxplots for different supp levels. This provides a way to compare the distribution of tooth lengths for different doses separately for each supp category.

```
bp + facet_grid(supp ~ .)
```



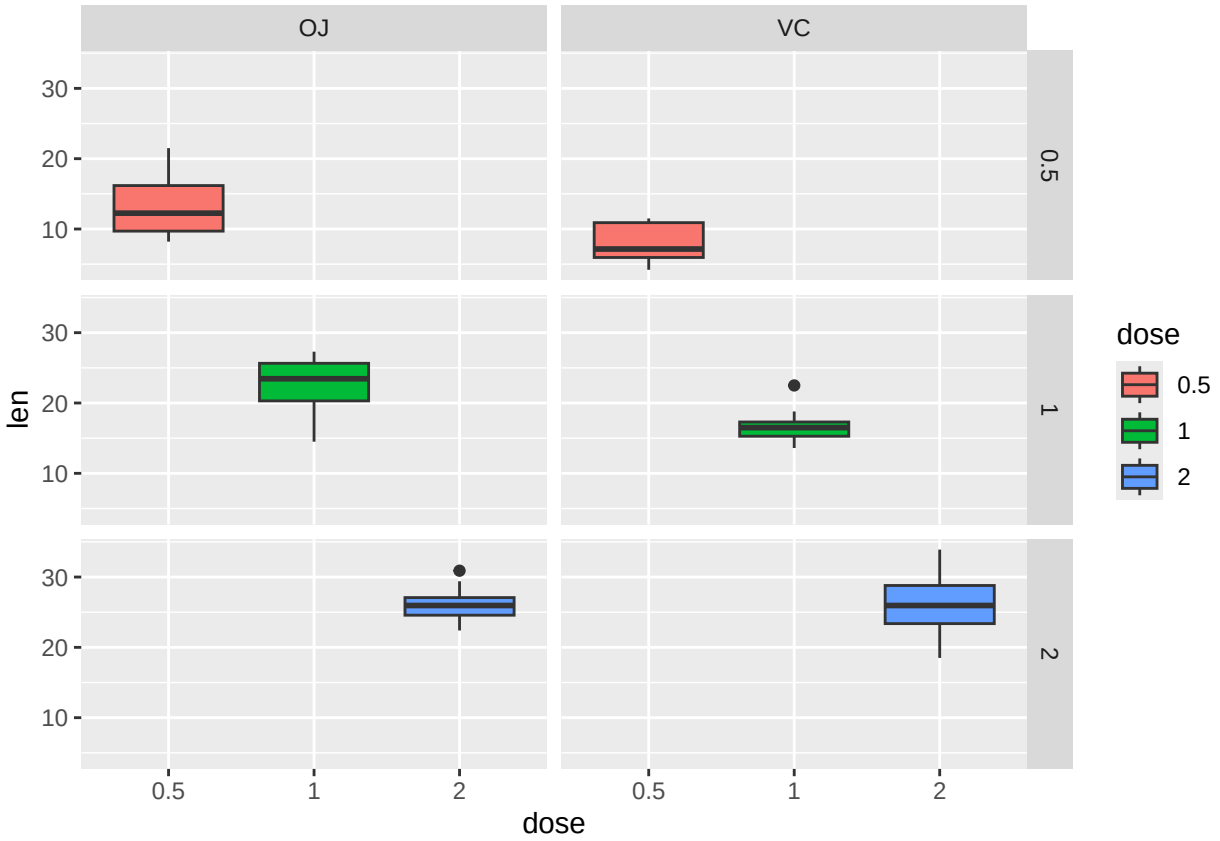
This code adds faceting to the boxplot of tooth length versus dose, organizing the facets into columns based on the supp variable. Each column represents a different level of supp, enabling a comparison of tooth length distributions across different doses for each supp category

```
bp + facet_grid(. ~ supp)
```



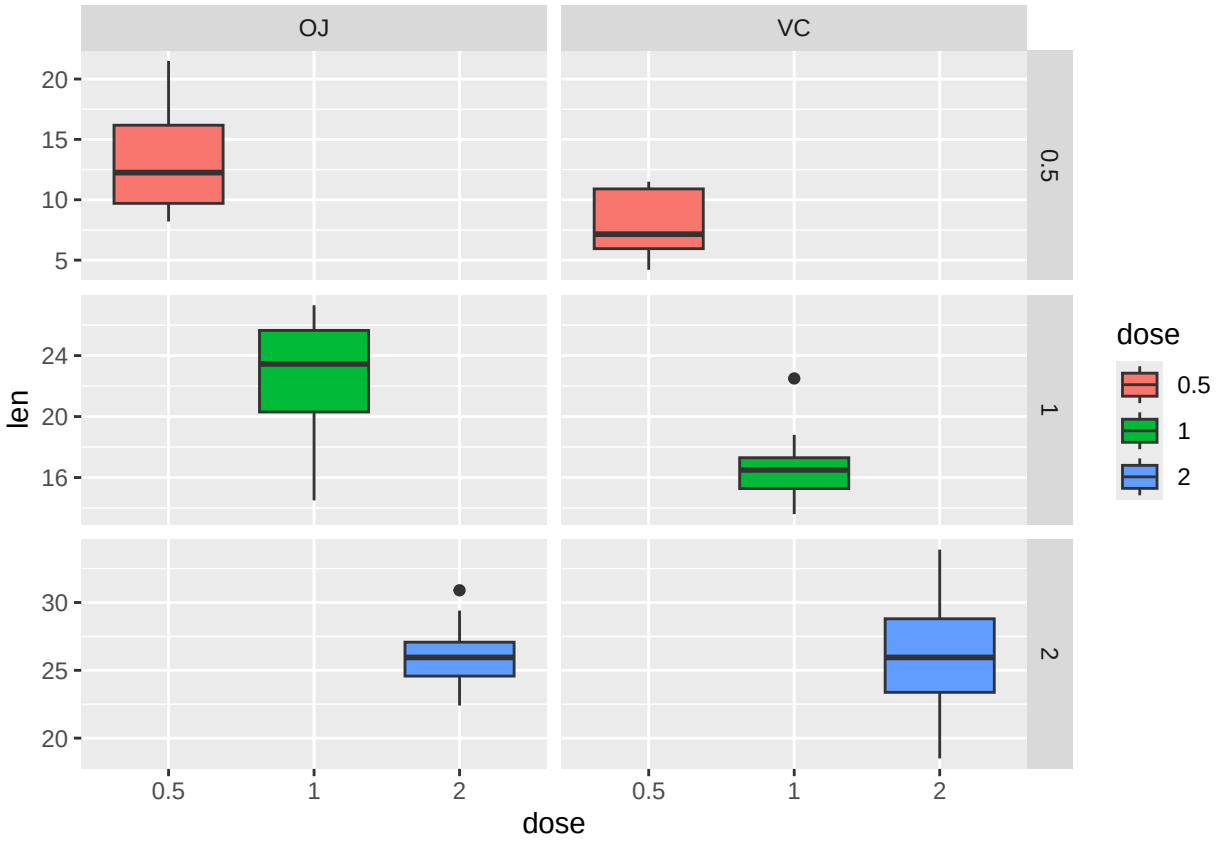
This code adds faceting to the boxplot of tooth length versus dose, arranging the facets into a grid based on the dose and supp variables. The grid layout allows for a comprehensive comparison of tooth length distributions for each combination of dose and supp.

```
bp + facet_grid(dose ~ supp)
```



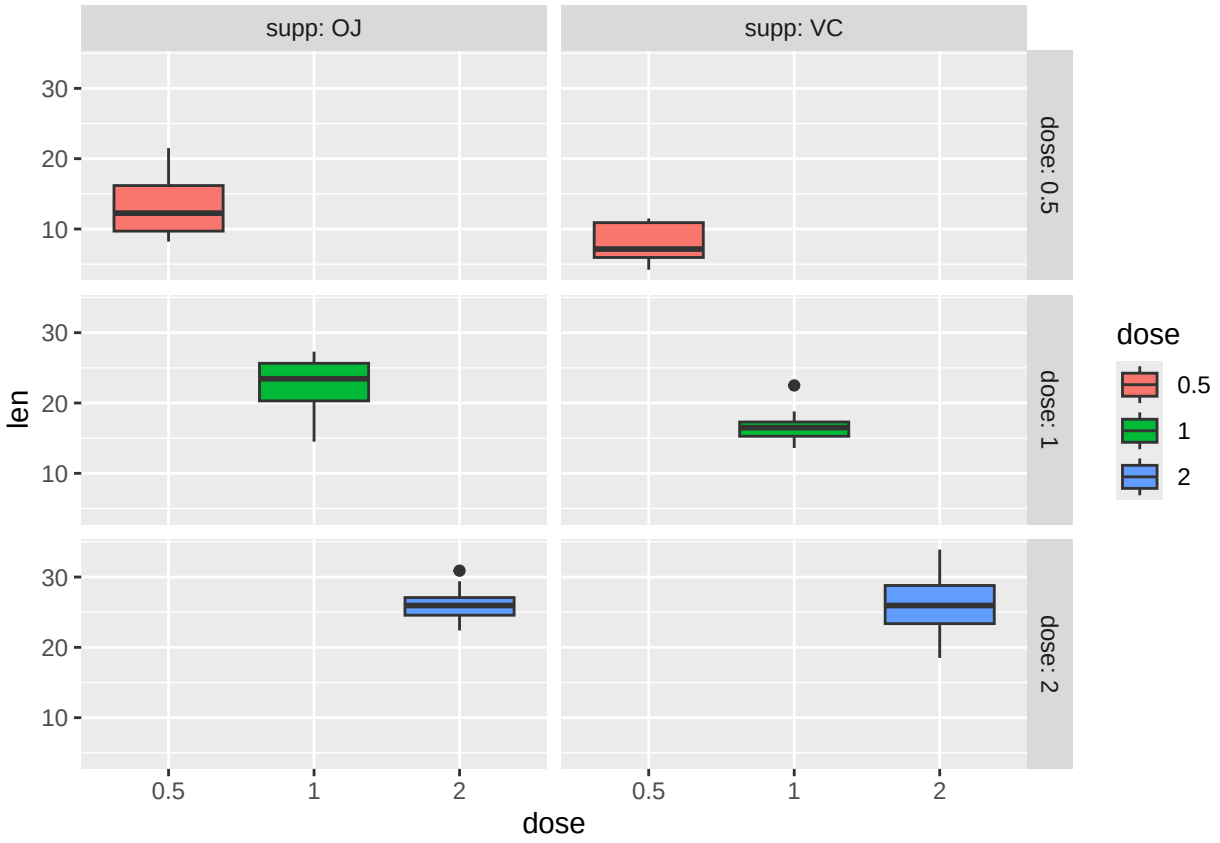
This code generates a faceted boxplot with a grid layout based on dose and supp, where each facet has its own y-axis scale. This allows for a clearer comparison of tooth length distributions across different doses and supplements, especially when the ranges of tooth lengths vary between facets.

```
bp + facet_grid(dose ~ supp, scales='free')
```



The code adds faceting to the existing plot (bp), creating a grid of subplots. Each subplot represents a combination of the dose and supp variables. The label_both labeller is used to show both the row (dose) and column (supp) variable names in the facet labels, making it easier to identify which combination of dose and supp each subplot corresponds to

```
bp + facet_grid(dose ~ supp, labeller=label_both)
```

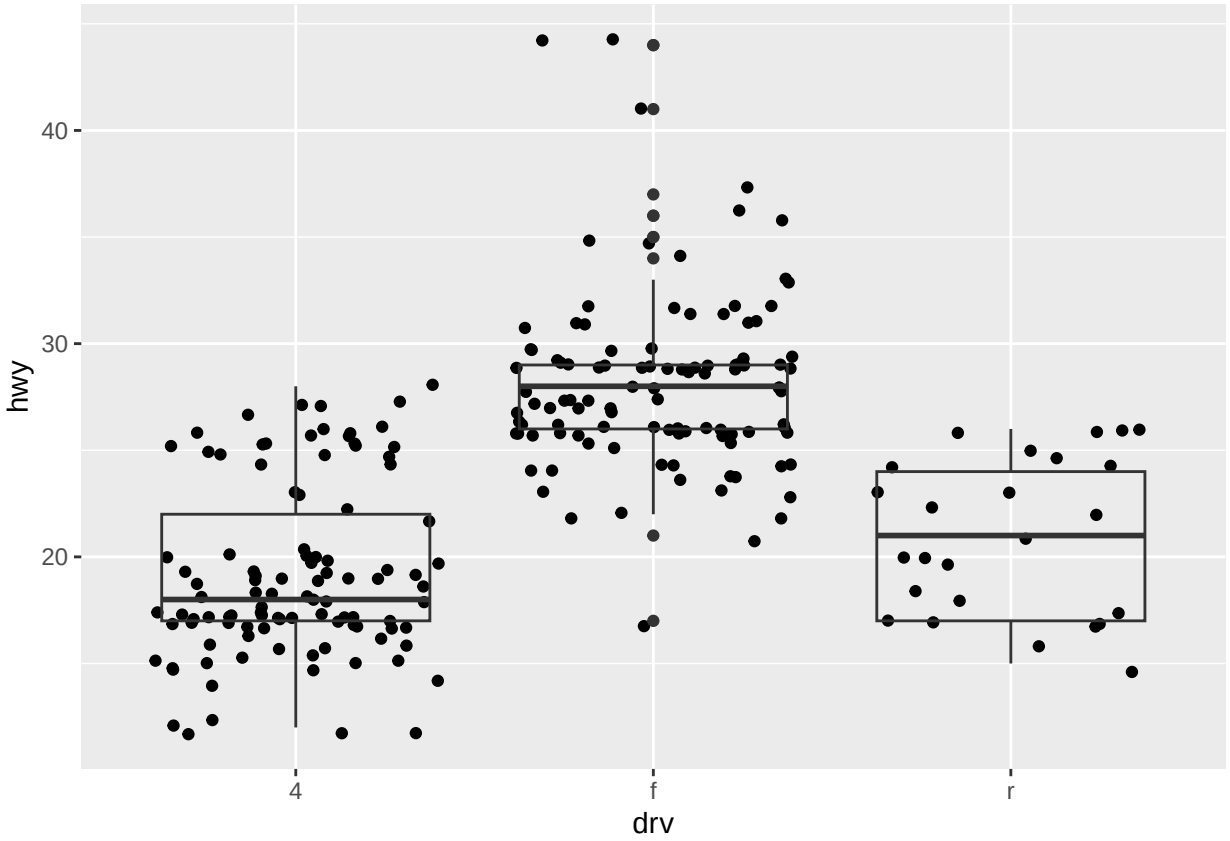


Grouping

Default grouping in ggplot2

The code creates a plot using the mpg dataset that combines a scatter plot and a boxplot. The scatter plot shows individual data points with jitter to prevent overplotting, while the boxplot overlays a summary of the distribution of hwy values for each drv category, with no fill inside the boxes.

```
p1 <- ggplot(mpg, aes(drv, hwy)) +
 geom_jitter() +
 stat_boxplot(fill = NA)
p1
```



making the new dataset

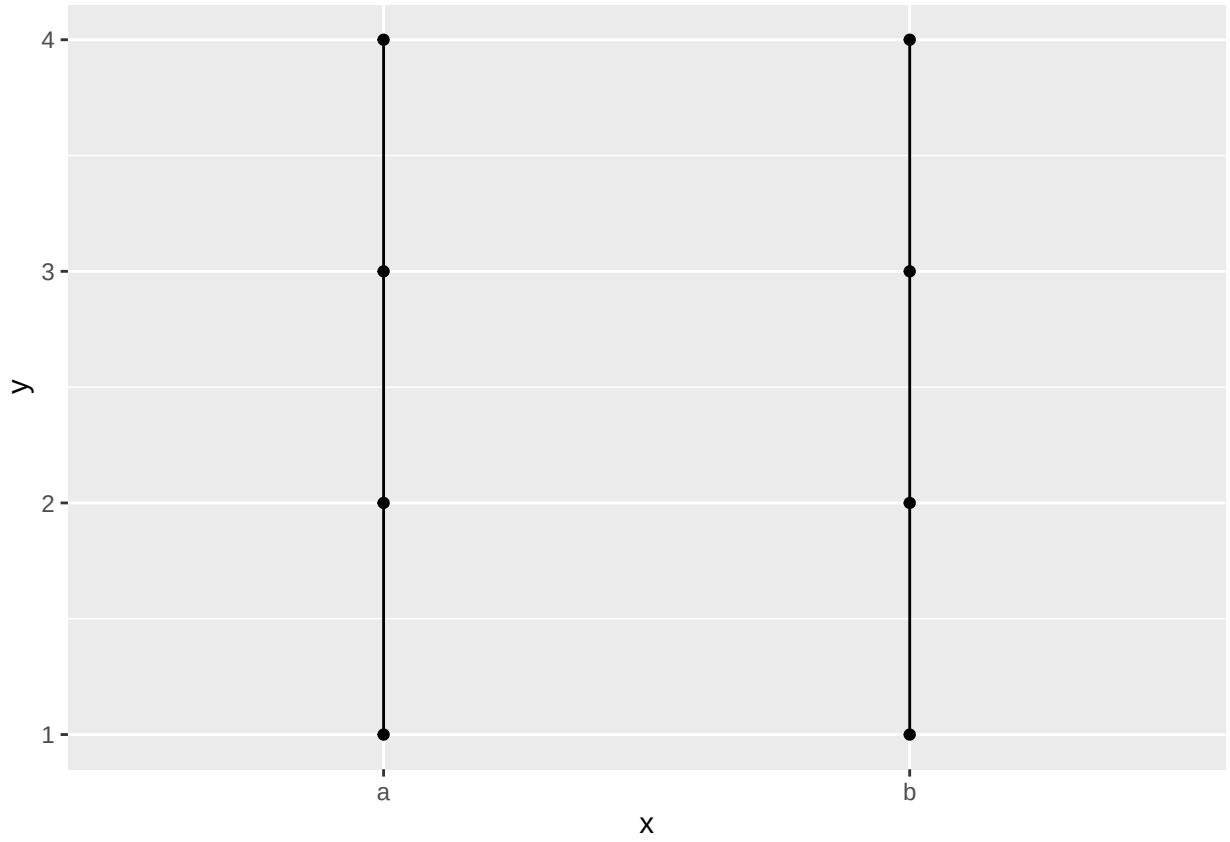
this code creates a data frame with specified values for three columns, where x and z are categorical variables and y is a numerical variable

```
df <- data.frame(x = c("a", "a", "a", "a", "b", "b", "b", "b"),
 y = c(1, 2, 3, 4, 4, 3, 2, 1),
 z = c("A", "A", "B", "B", "B", "B", "A", "A"))
df
```

```
##   x y z
## 1 a 1 A
## 2 a 2 A
## 3 a 3 B
## 4 a 4 B
## 5 b 4 B
## 6 b 3 B
## 7 b 2 A
## 8 b 1 A
```

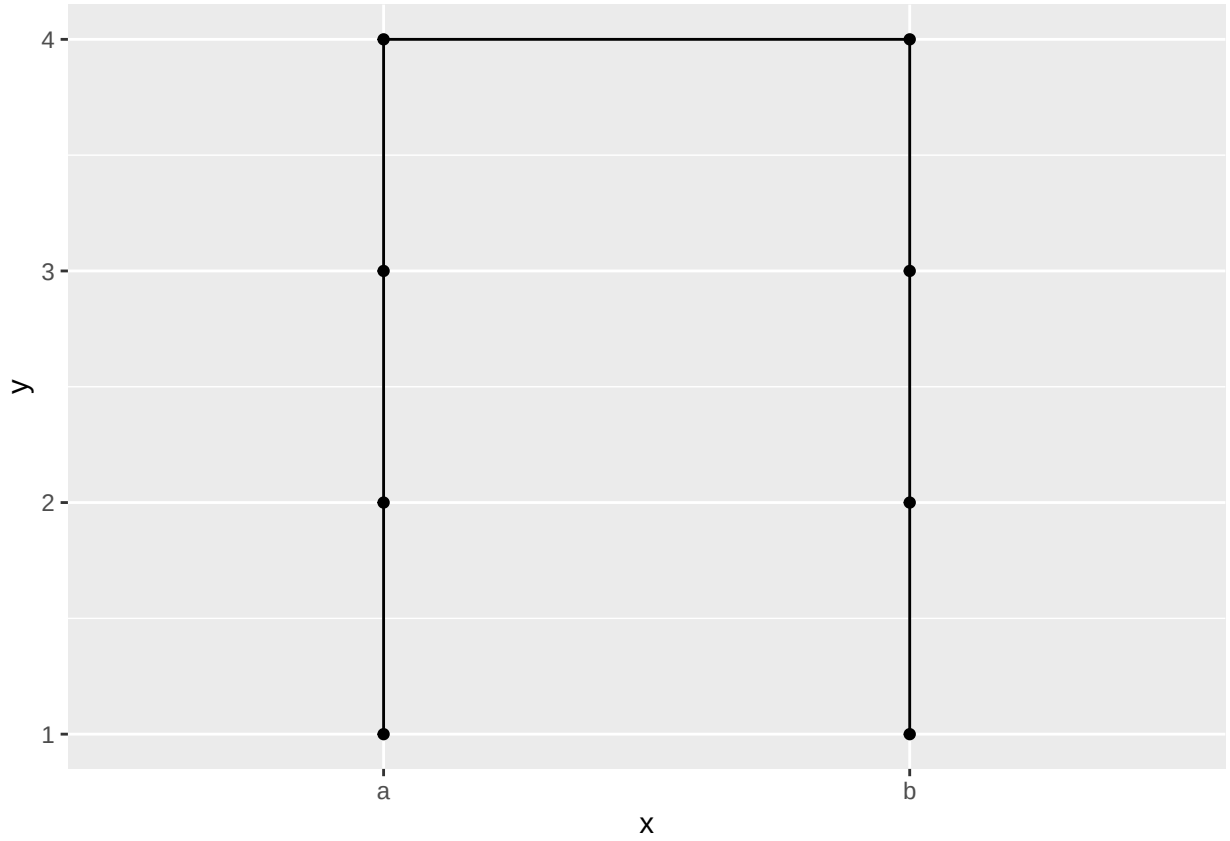
the plot display a scatter plot with points connected by lines in the order they are listed in the data frame, effectively showing the progression or sequence of the data.

```
ggplot(df, aes(x, y)) +
 geom_point() +
 geom_path()
```



the plot display scatter plot with a single line connecting all the points in the order they are listed in the data frame, indicating the sequence of the data as a continuous path

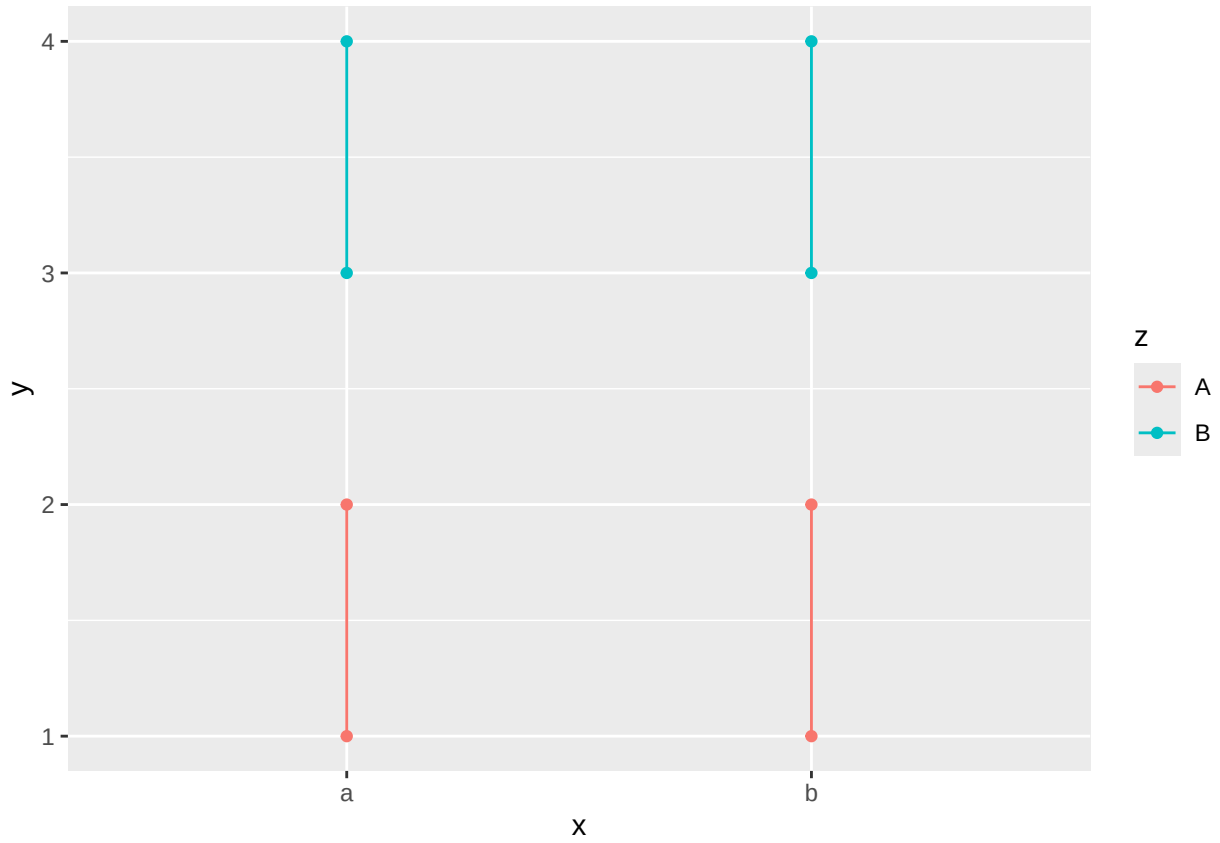
```
ggplot(df, aes(x, y)) +
 geom_point() +
 geom_path(aes(group = 10))
```



Grouping for collective geoms

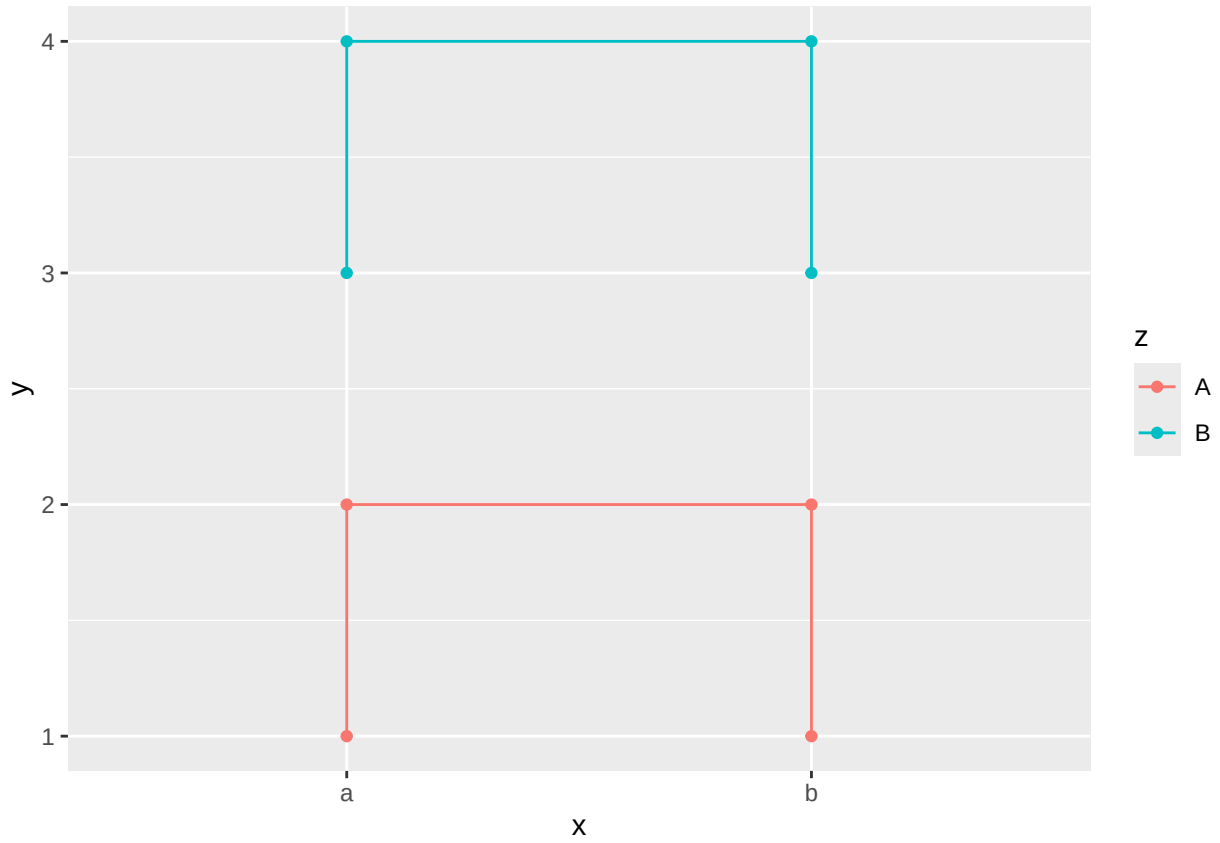
this code generates a scatter plot with points colored according to the variable z and connects these points with lines to show the sequence or trend based on their order in the data.

```
ggplot(df, aes(x, y, color = z)) +
 geom_point() +
 geom_path()
```



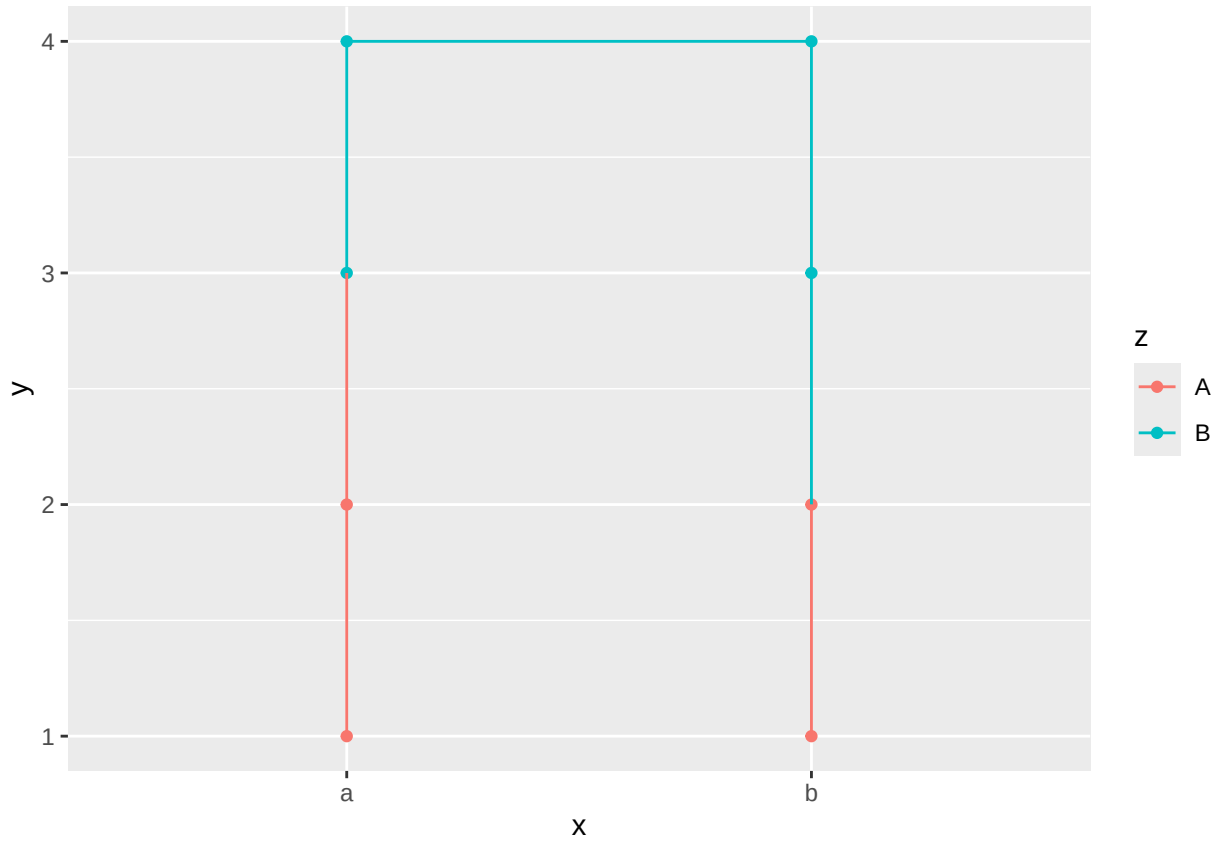
this code creates a scatter plot where points are colored by z, and lines connect the points within each group defined by z. Each group will have a distinct line connecting its points, which helps visualize trends or paths for each category.

```
ggplot(df, aes(x, y, color = z)) +
 geom_point() +
 geom_path(aes(group = z))
```



this code creates a scatter plot where points are colored based on the variable z, and a single line is drawn connecting all points in the dataset, since all points are grouped together under group = 10

```
ggplot(df, aes(x, y, color = z)) +
 geom_point() +
 geom_path(aes(group = 10))
```


4
3
2
1
y
a
b
x
z
A
B