1HW\_Mattern - Visualization class I

Rebecca Mattern

September 21, 2017

#### 3.1.1 Prerequisites:

library( ggplot2 )  
# library( tidyverse )

#### 3.2.4 Examine dataset mpg: (excluding #4)

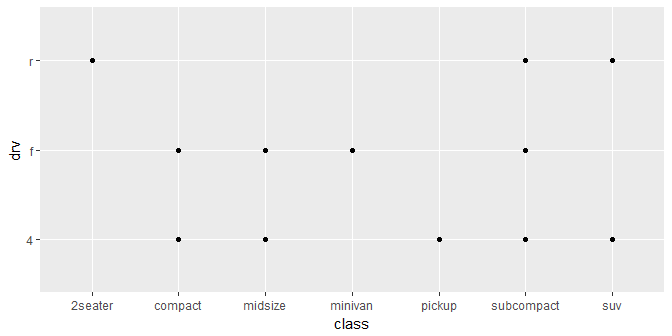
1. I see a table of data
2. There are 234 rows and 11 columns.
3. The drv variable describes the type of wheel drive (front, rear, or four).

mpg

## # A tibble: 234 x 11  
## manufacturer model displ year cyl trans drv cty hwy  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int>  
## 1 audi a4 1.8 1999 4 auto(l5) f 18 29  
## 2 audi a4 1.8 1999 4 manual(m5) f 21 29  
## 3 audi a4 2.0 2008 4 manual(m6) f 20 31  
## 4 audi a4 2.0 2008 4 auto(av) f 21 30  
## 5 audi a4 2.8 1999 6 auto(l5) f 16 26  
## 6 audi a4 2.8 1999 6 manual(m5) f 18 26  
## 7 audi a4 3.1 2008 6 auto(av) f 18 27  
## 8 audi a4 quattro 1.8 1999 4 manual(m5) 4 18 26  
## 9 audi a4 quattro 1.8 1999 4 auto(l5) 4 16 25  
## 10 audi a4 quattro 2.0 2008 4 manual(m6) 4 20 28  
## # ... with 224 more rows, and 2 more variables: fl <chr>, class <chr>

1. (excluded)
2. It is not useful because they are both categorical variables and thus cannot show any correlation.

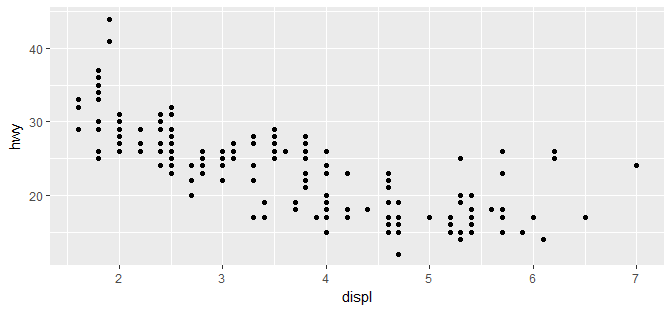
# nPlot1\_Height  
 ggplot( data = mpg ) +  
 geom\_point( mapping = aes( x = class, y = drv ) )



#### 3.3.1 Plot relationship between engine size and highway mileage: (excluding #3)

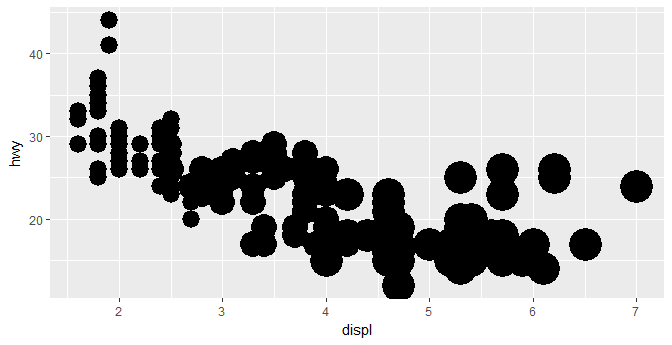
1. The parenthesis needs to go before color="blue", not afterwards.
2. The categorical variables are: model, trans, drv, fl, and class. The continuous variables are: disp, year, cyl, cty, and hwy. The ?mpg commands shows up in the help tab on the lower right hand corner of RStudio.
3. (excluded)
4. RStudio advises against using size for a discrete variable and only the color variable shows up.
5. It shapes the size of the data points based on numerical value, it uses circles.

ggplot( data = mpg ) +  
 geom\_point( mapping = aes( x = displ, y = hwy ) )



1. It is an incomplete expression.

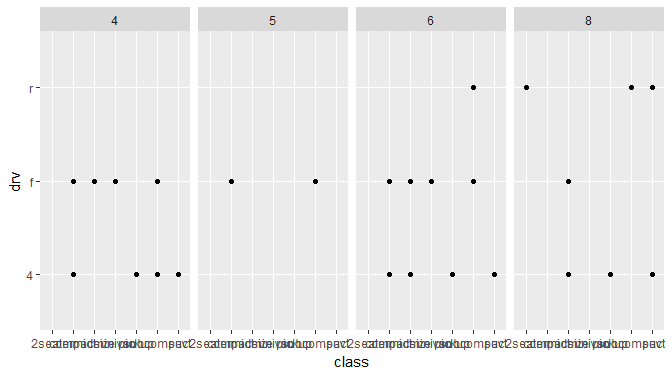
ggplot( data = mpg ) +  
 geom\_point( mapping = aes( x = displ, y = hwy, stroke = cyl ) )



#### 3.5.1 Add Facets: (excluding #4)

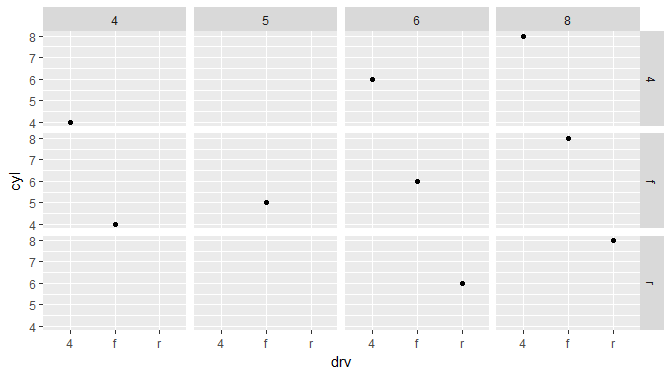
1. There are boxes for each of the continous values.

ggplot( data = mpg ) +  
 geom\_point( mapping = aes( x = class, y = drv ) ) + facet\_grid( . ~ cyl )



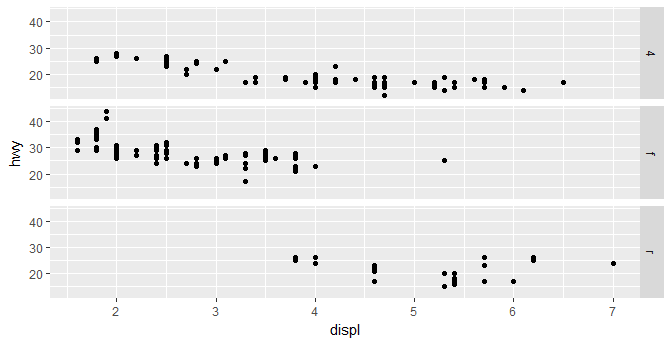
1. The empty cells mean that there are no instances when the rear wheel drive had 4 or 5 cylinders.

ggplot( data = mpg ) +  
 geom\_point( mapping = aes( x = drv, y = cyl ) ) + facet\_grid( drv ~ cyl )



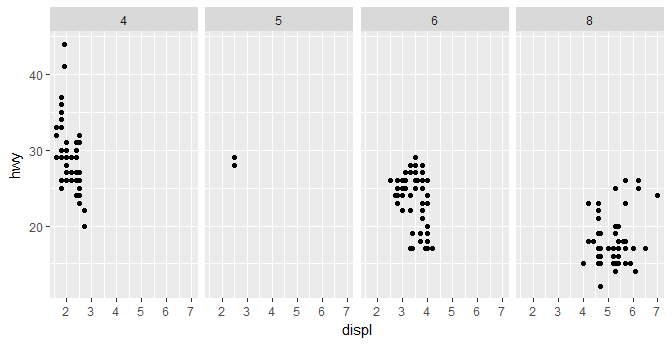
1. The period (.) serves as a filler so that drv is only compared to itself. It is only a one-direction grid.

ggplot( data = mpg ) +  
 geom\_point( mapping = aes( x = displ, y = hwy ) ) + facet\_grid( drv ~ . )



1. (excluded)
2. nRow and nCol stands for the number of rows and columns. The other option controls are scales, shrink, labeller, as.table, switch, drop, dir, and strip.position. Facet\_grid doesn't have nrow or ncol option, because its row and column parameters are set by a variable.
3. When using facet\_grid you should usually put the variable with more unique levels in the columns to prevent from there being too many boxes. It needs to be specific to only a few, so that they can be compared more easily.

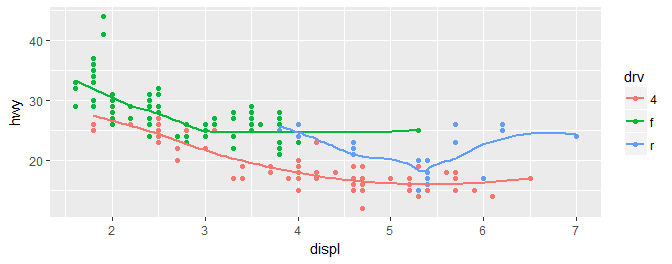
ggplot( data = mpg ) +  
 geom\_point( mapping = aes( x = displ, y = hwy ) ) + facet\_grid( . ~ cyl )



#### 3.6.1 Add more geometric layers / objects: (excluding #6)

1. To draw a line chart you would use geom\_abline.
   * Boxplot: geom\_boxplot()
   * Histogram: geom\_histogram Area
   * Chart: geom\_area.
2. The visualization matched my predictions.
3. show.legend=FALSE takes away the key. Taking it away puts the key back.
4. It took away the shading.
5. The graphs looked the same because both adequately established the inputs.
6. (excluded)

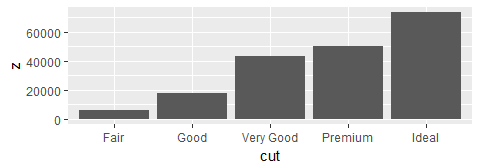
ggplot( data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +  
 geom\_point() +  
 geom\_smooth(se=FALSE)



#### 3.7.1 Statistical Transformations (excluding #3)

1. The default geom associated with stat\_summary is geom\_pointrange. You could rewrite it with that function.
2. Geom\_col allows you to add a y aesthetic. Geom\_bar only allows for one variable and hence shows the count.

ggplot( data = diamonds ) +  
 geom\_col( mapping = aes( x = cut, y = z ) )

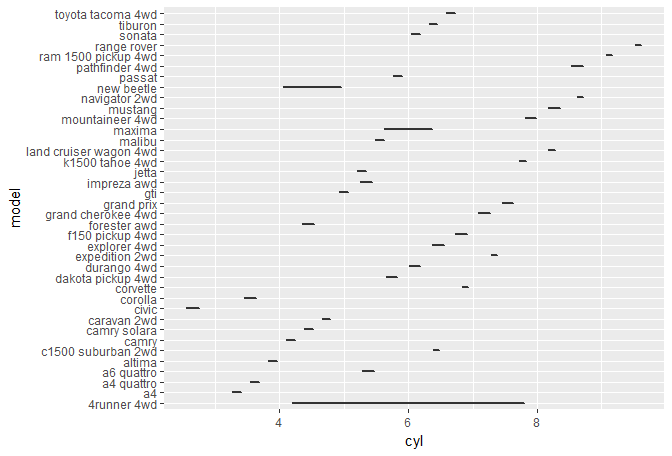


1. (excluded)
2. Stat\_smooth uses se, x, y, ymin, and ymax variables.
3. Without group = 1, the variables all look the same, they are evenly distributed.

#### 3.8.1 Position Adjustments

1. The problem is that there are a few outliers. You could improve it by cutting off the domain at cty=30.
2. Height and width.
3. Geom\_jitter scatters the points around the grid, where geom\_count places one point on every intersection of gridlines.
4. The default position is "dodge".

ggplot( data = mpg ) + geom\_boxplot( mapping = aes( x = cyl, y = model ) )



#### 3.9.1 Coordinate Systems

1. labs() adds labels by adding text to the axis, plot title, or caption below the plot.
2. Coord\_map shows a 3D image on a 2D space, and coord\_quickmap aligns the image straightly.
3. The plot shows that city and highway are positively correlated. Coord\_fixed changes the ratio dimensions of the plot. Geom\_abline adds the line beside the plotted points.

ggplot( data = mpg) + geom\_bar( mapping = aes( x = displ ) ) + coord\_polar( )

