Explore and Visualize

Prof. Eric A. Suess

# Chapter 3 Data Visualization

1. ggplot2
2. scatterplots, points, color, shape
3. faceting

Today we are going to try some of the code from Chapter 3 Data Visualization.

To start we will load the tidyverse. Note that *ggplot2* is the first package loaded!

library(tidyverse)

We will start the *mpg* dataset that is in the ggplot2 package.

mpg

## # A tibble: 234 x 11
## manufacturer model displ year cyl trans drv cty hwy fl cla…
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <ch>
## 1 audi a4 1.8 1999 4 auto… f 18 29 p com…
## 2 audi a4 1.8 1999 4 manu… f 21 29 p com…
## 3 audi a4 2 2008 4 manu… f 20 31 p com…
## 4 audi a4 2 2008 4 auto… f 21 30 p com…
## 5 audi a4 2.8 1999 6 auto… f 16 26 p com…
## 6 audi a4 2.8 1999 6 manu… f 18 26 p com…
## 7 audi a4 3.1 2008 6 auto… f 18 27 p com…
## 8 audi a4 q… 1.8 1999 4 manu… 4 18 26 p com…
## 9 audi a4 q… 1.8 1999 4 auto… 4 16 25 p com…
## 10 audi a4 q… 2 2008 4 manu… 4 20 28 p com…
## # ... with 224 more rows

Make a scatterplot of highway miles per gallon (hwy) and engine size (displ).

Start with an empty graph and add to it.

ggplot(data = mpg)



Add points using geom\_point mapping and aesthetic.

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



Add color.

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



Change size of the points.

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

## Warning: Using size for a discrete variable is not advised.



Change the transparentcy or shape.

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

## Warning: Using alpha for a discrete variable is not advised.



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

## Warning: The shape palette can deal with a maximum of 6 discrete values
## because more than 6 becomes difficult to discriminate; you have 7.
## Consider specifying shapes manually if you must have them.

## Warning: Removed 62 rows containing missing values (geom\_point).



Specify color.

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



Faceting, spliting a plot into separate plots for different levels of a categorical variable.

ggplot(data = mpg) +
 geom\_point(mapping = aes(x = displ, y = hwy)) +
 facet\_wrap(~ class, nrow = 2)

