Quick introduction to ggplot2

DSC 105, Introduction to Data Science, Lyon College Fall 2025

Grammar of graphics with `ggplot2` (from DSC 302)

- Grammar of graphics construction based on human perception
- Better support for multipanel conditioning plots
- Highly extensible, complex, steep learning curve (see: ggplot2.tidyverse.org, with cheatsheets)

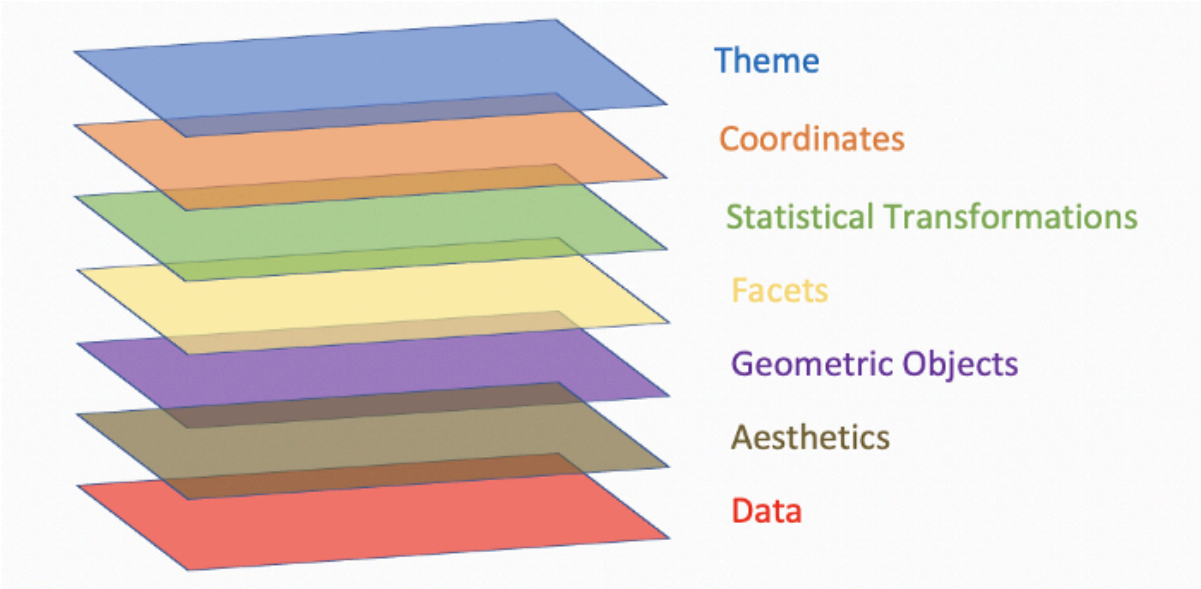


Figure 1: Grammar of Graphics (gg) philosophy

Examples: simple plots

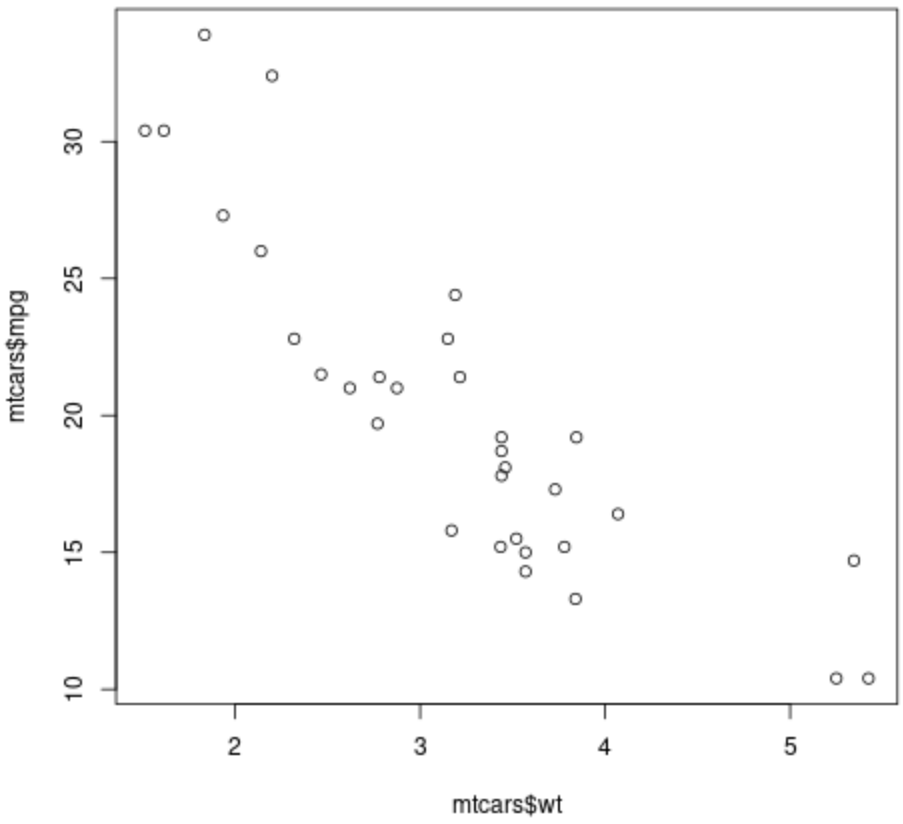
- A few examples contrasting base R and ggplot2 plotting:

```
str(mtcars)
```

```
'data.frame':   32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

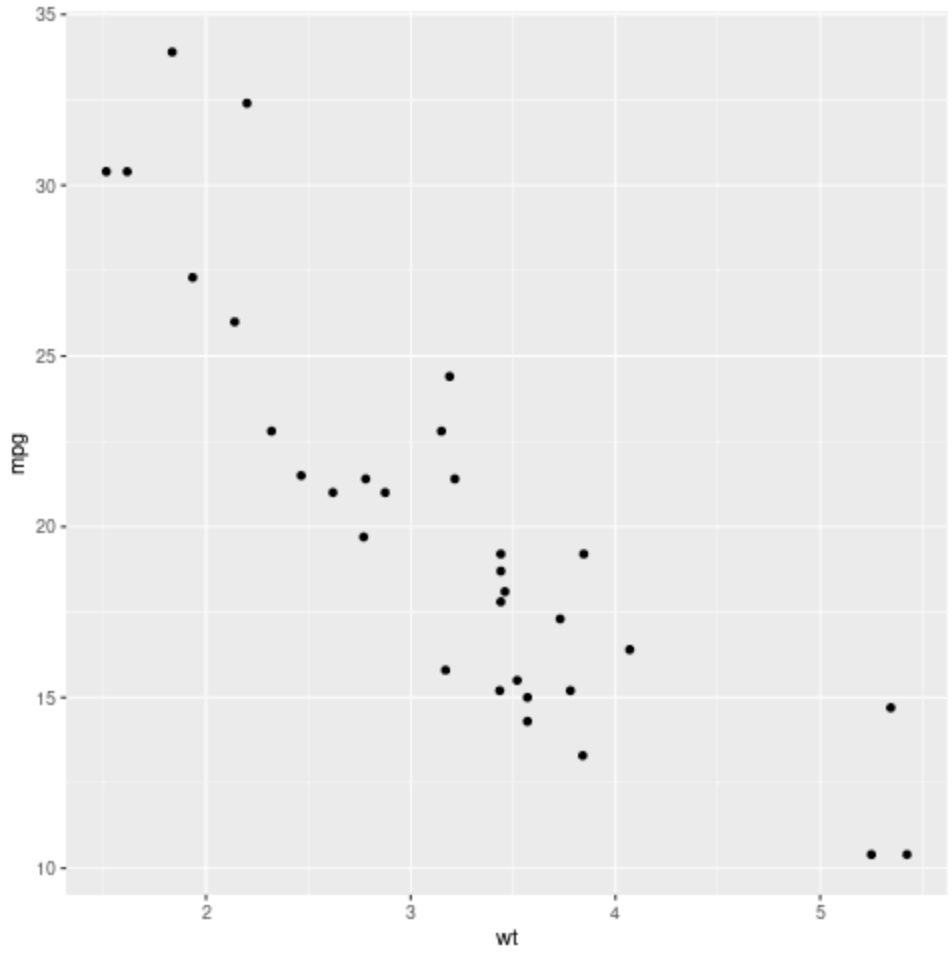
- Plotting miles-per-gallon vs. weight of the cars in `mtcars` with `base::plot`

```
plot(mtcars$mpg ~ mtcars$wt)
```



- With `ggplot`:

```
library(ggplot2)
ggplot(mtcars, aes(wt, mpg)) +
  geom_point()
```

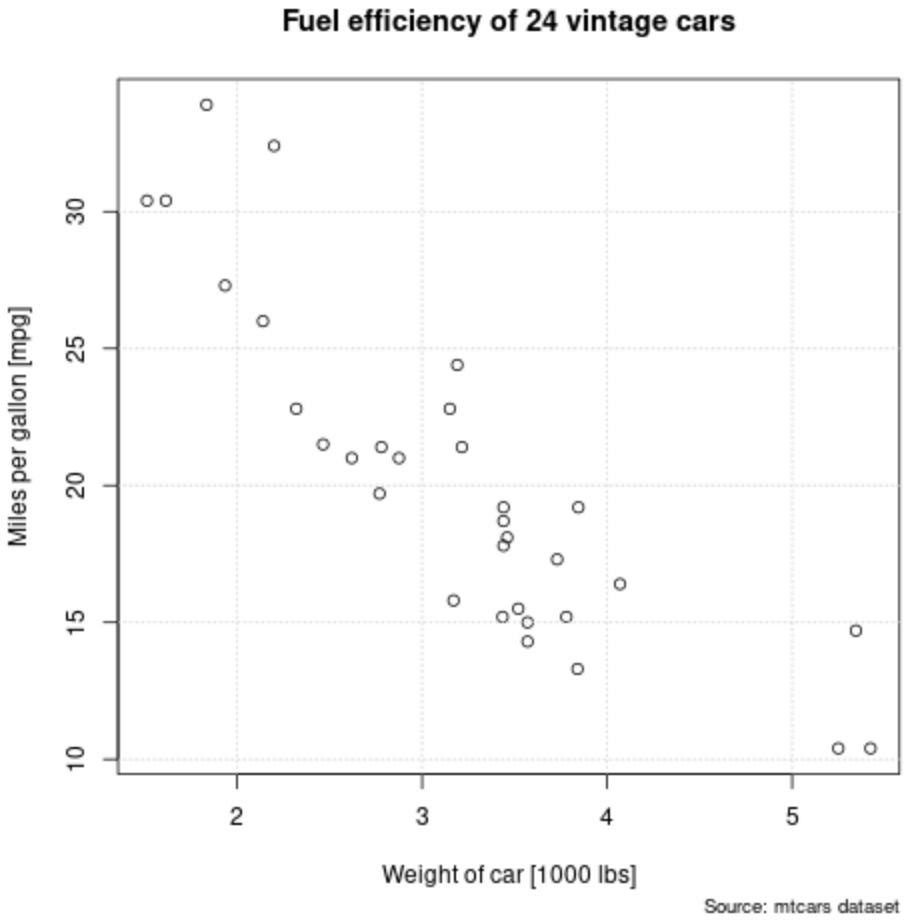


- You can also use the pipe operators `%>%` and `|>` to do this.

Simple customization

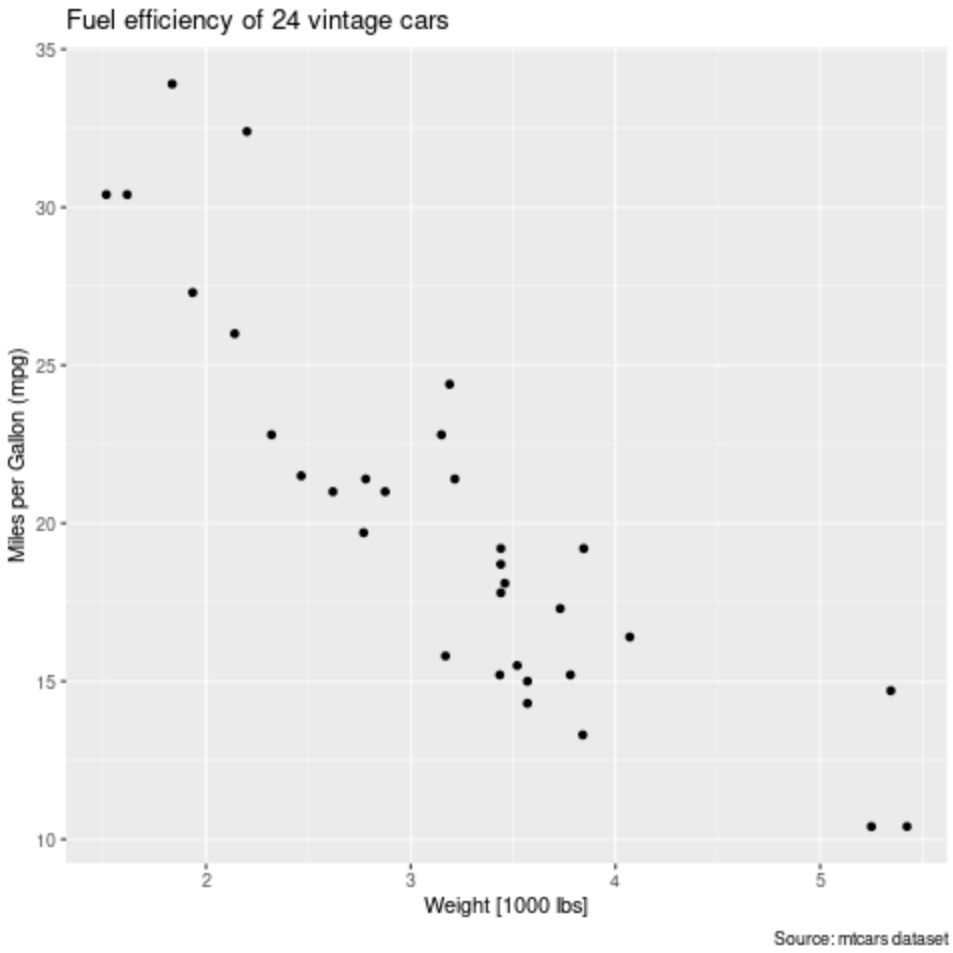
- With customization: Add labels, title, gridlines.
- Base R:

```
plot(mtcars$mpg ~ mtcars$wt,
     xlab="Weight of car [1000 lbs]",
     ylab="Miles per gallon [mpg]",
     main="Fuel efficiency of 24 vintage cars")
mtext("Source: mtcars dataset", side = 1, line = 4, adj = 1, cex = 0.8)
grid()
```



- ggplot:

```
library(ggplot2)
ggplot(mtcars, aes(wt, mpg)) +
  geom_point() +
  labs(
    title = "Fuel efficiency of 24 vintage cars",
    x = "Weight [1000 lbs]",
    y = "Miles per Gallon (mpg)",
    caption = "Source: mtcars dataset"
  )
```

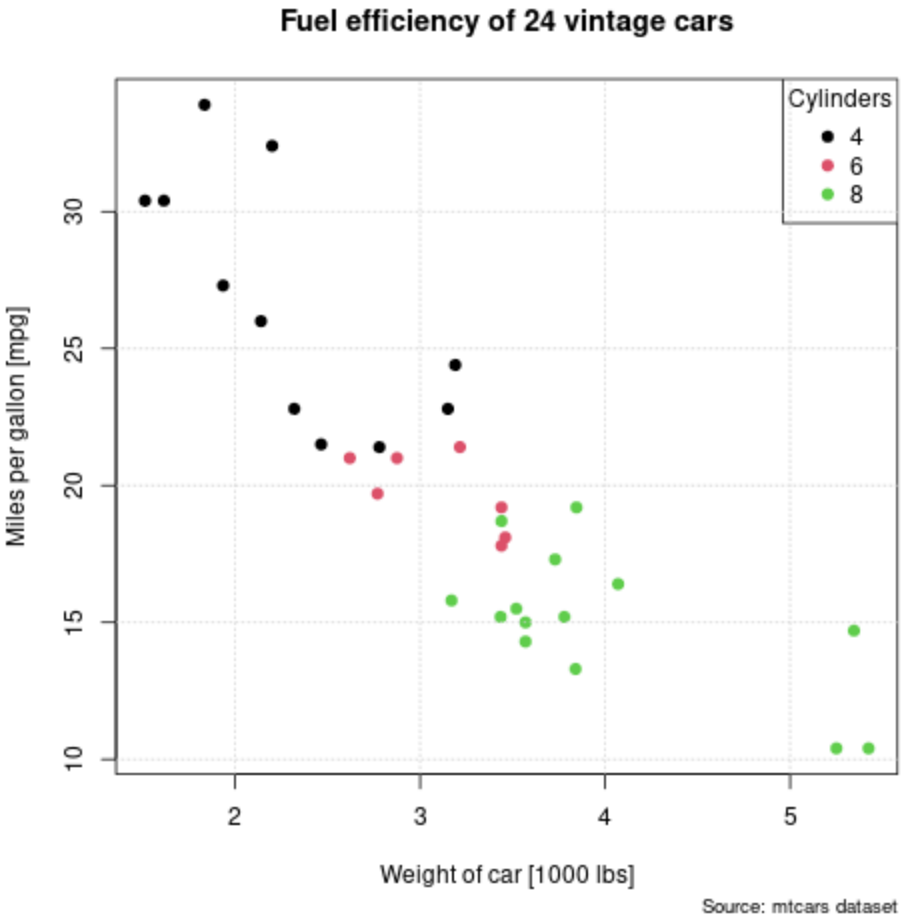


Multi-dimensional plots

- In R, factor vectors are often used to add a categorical dimension to a numeric plot.
- Add the number of cylinders (`mtcars$cyl`) to the plot.

- Base R:

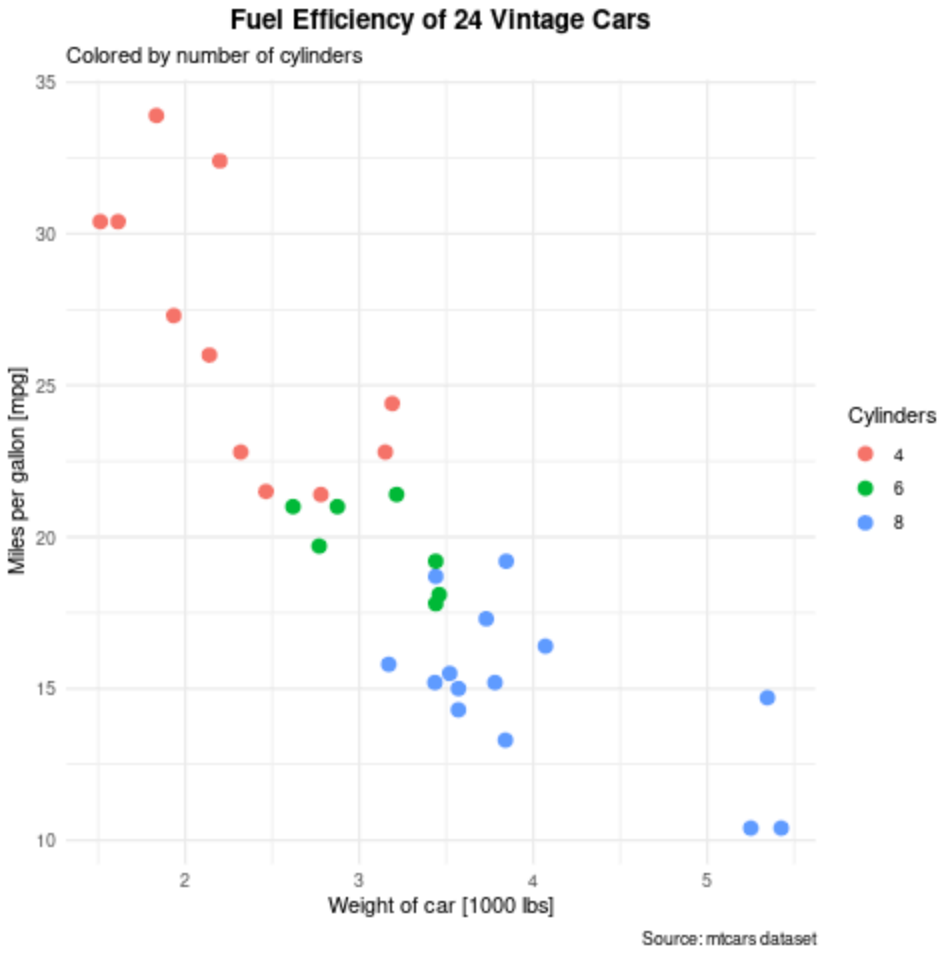
```
plot(mtcars$mpg ~ mtcars$wt,
     xlab="Weight of car [1000 lbs]",
     ylab="Miles per gallon [mpg]",
     main="Fuel efficiency of 24 vintage cars",
     pch = 19,
     col = as.factor(mtcars$cyl))
legend("topright",
       legend = levels(as.factor(mtcars$cyl)),
       col = 1:length(levels(as.factor(mtcars$cyl))),
       pch = 19,
       title = "Cylinders")
mtext("Source: mtcars dataset", side = 1, line = 4, adj = 1, cex = 0.8)
grid()
```



- ggplot:

```
library(ggplot2)

ggplot(mtcars, aes(x = wt, y = mpg, color = as.factor(cyl))) +
  geom_point(size = 3) +
  labs(
    title = "Fuel Efficiency of 24 Vintage Cars",
    subtitle = "Colored by number of cylinders",
    x = "Weight of car [1000 lbs]",
    y = "Miles per gallon [mpg]",
    color = "Cylinders",
    caption = "Source: mtcars dataset"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))
```



Author: Marcus Birkenkrahe (pledged)
Created: 2025-11-13 Thu 20:00