

Graphical critique & theory

Hadley Wickham

Exploratory graphics

Are for **you** (not others). Need to be able to create rapidly because your first attempt will never be the most revealing.

Iteration is crucial for developing the best display of your data.

Gives rise to two key questions:

What should I plot?
How can I plot it?

Two general tools

Plot critique toolkit:

“graphics are like pumpkin pie”

Theory behind ggplot2:

“A layered grammar of graphics”

plus lots of practice...

Graphics are like
pumpkin pie

The four **C**'s of critiquing a graphic

Content



Construction





Context



Consumption

Content

The background of the slide features a close-up, slightly blurred image of several pumpkins. The pumpkins are a warm, orange-brown color and are arranged in a cluster, with one pumpkin in the foreground being more prominent than the others. The lighting is soft, highlighting the texture of the pumpkin skin.

What data (variables) does the graph display? What non-data is present?

What is pumpkin (essence of the graphic) vs what is spice (useful additional info)?

Your turn

Identify the data and non-data on “Napoleon’s march” and “Building an electoral victory”.

Results

**Minard's march: (top) latitude,
longitude, number of troops,
direction, branch, city name
(bottom) latitude, temperature, date**

**Building an electoral victory: state,
number of electoral college votes,
winner, margin of victory**

Construction

How many layers are on the plot? What data does each layer display? What sort of geometric object does it use? How are variables mapped to aesthetics?

Your turn

Answer the following questions for “Napoleon’s march” and “Flight delays”:

How many layers are on the plot?

What data does the layer display? How does it display it?

Results

Napoleon's march: (top) (1) path plot with width mapped to number of troops, colour to direction, separate group for each branch (2) labels giving city names (bottom) (1) line plot with longitude on x-axis and temperature on y-axis (2) text labels giving dates

Flight delays: (1) white circles showing 100% cancellation, (2) outline of states, (3) points with size proportional to percent cancellations at each airport.

Can the explain
composition of a graphic
in words, but how do we
create it?



“If any number of magnitudes are each the same multiple of the same number of other magnitudes, then the sum is that multiple of the sum.”

Euclid, ~300 BC



“If any number of magnitudes are each the same multiple of the same number of other magnitudes, then the sum is that multiple of the sum.”

Euclid, ~300 BC

$$m(\sum x) = \sum(mx)$$

The grammar of graphics

An abstraction which makes it easier to create, understand and communicate graphics.

Developed by Leland Wilkinson, particularly in “The Grammar of Graphics” 1999/2005

ggplot2 adapts for use within R and builds to create “A Layered Grammar of Graphics” (in press, Journal of Computational and Graphical Statistics).

What is a layer?

- Data
- Mappings from variables to aesthetics (**aes**)
- A geometric object (**geom**)
- A statistical transformation (**stat**)
- A position adjustment (**position**)

```
layer(geom, stat, position, data, mapping, ...)
```

```
layer(  
  data = mpg,  
  mapping = aes(x = displ, y = hwy),  
  geom = "point",  
  stat = "identity",  
  position = "identity"  
)
```

```
layer(  
  data = diamonds,  
  mapping = aes(x = carat),  
  geom = "bar",  
  stat = "bin",  
  position = "stack"  
)
```

```
# A lot of typing!
```

```
layer(  
  data = mpg,  
  mapping = aes(x = displ, y = hwy),  
  geom = "point",  
  stat = "identity",  
  position = "identity"  
)
```

```
# Every geom has an associated default statistic  
# (and vice versa), and position adjustment.
```

```
geom_point(aes(displ, hwy), data = mpg)  
geom_histogram(aes(displ), data = mpg)
```

```
# To actually create the plot
```

```
ggplot() +
```

```
  geom_point(aes(displ, hwy), data = mpg)
```

```
ggplot() +
```

```
  geom_histogram(aes(carat), data = diamonds)
```

```
# Multiple layers
```

```
ggplot() +
```

```
  geom_point(aes(displ, hwy), data = mpg) +
```

```
  geom_smooth(aes(displ, hwy), data = mpg)
```

```
# Avoid redundancy:
```

```
ggplot(mpg, aes(displ, hwy)) +
```

```
  geom_point() +
```

```
  geom_smooth()
```



```
# Different layers can have different aesthetics
ggplot(mpg, aes(displ, hwy)) +
  geom_point(aes(colour = class)) +
  geom_smooth()
```

```
ggplot(mpg, aes(displ, hwy)) +
  geom_point(aes(colour = class)) +
  geom_smooth(aes(group = class), method = "lm",
             se = F)
```

Your turn

Perform the same process for the flight delays data, and try and come up with the ggplot2 code that would create it.

```

usa <- map_data("state")

ggplot(feb13, aes(long, lat)) +
  geom_point(aes(size = 1), colour = "white") +
  geom_polygon(aes(group = group), data = usa,
    colour = "grey70", fill = NA) +
  geom_point(aes(size = ncancelw / ntot),
    colour = alpha("black", 1/2))

last_plot() +
  scale_area("% cancelled", to = c(1, 8),
    breaks = seq(0, 1, by = 0.2), limits = c(0, 1))
scale_x_continuous("", limits = c(-125, -67)),
scale_y_continuous("", limits = c(24, 50))

```

Back to Minard

```
troops <- read.csv("minard-troops.csv")
cities <- read.csv("minard-cities.csv")

ggplot(cities, aes(long, lat)) +
  geom_path(aes(size = survivors, colour = direction,
    group = interaction(group, direction)), data = troops) +
  geom_text(aes(label = city), hjust = 0, vjust = 1, size = 4)

# Polish appearance
last_plot() +
  scale_x_continuous("", limits = c(24, 39)) +
  scale_y_continuous("") +
  scale_colour_manual(values = c("grey50", "red")) +
  scale_size(to = c(1, 10))
```

Other components

Scales. Used to override default perceptual mappings. Mainly useful for polishing plot for communication.

Coordinate system. Rarely useful, but when needed are critical.

Facetting. Have seen in use already. Only other feature of importance is interaction with scales.

Themes: control presentation of non-data elements.