



Visualization: Grammar of Graphics

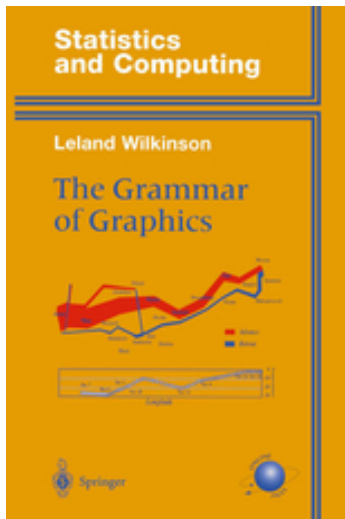
Data Science: Jordan Boyd-Graber
University of Maryland

FEBRUARY 11, 2018

Understanding Data

- After you've cleaned (wrangled) data
- Need to tell a story with the data
- Often a first step before you can build a model
- Necessary afterward to explain model works

Grammar of Graphics



- Don't focus on pixels
- Focus on data
- Easy combination / switches

Tiny Dataset

<i>x</i>	<i>y</i>	Shape
2	4	a
1	1	a
4	15	b
9	80	b

Data Tell a Story

x	y	Shape
25	11	circle
0	0	circle
75	53	square
200	300	square

What visualization helps you tell your story?

Components of a Plot

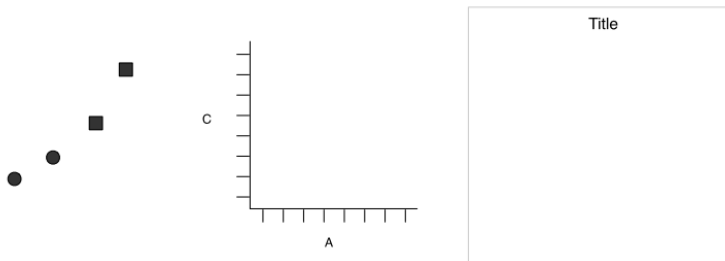
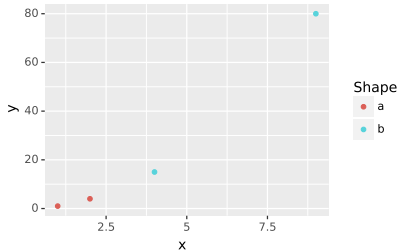


Figure 1. Graphics objects produced by (from left to right): geometric objects, scales and coordinate system, plot annotations.

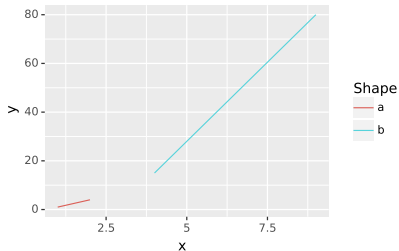
- Geom: How data turn into shapes
- Scales: Relative positioning
- Annotations: Text, explanations

Putting it Together



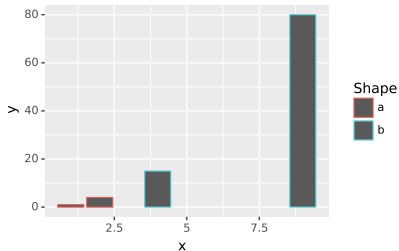
```
simple_point = (ggplot(demo,  
                      aes(color='Shape', y='y', x='x')) +  
               geom_point())  
simple_point.save("simple_point.pdf", scale=0.6,  
                 height=6, width=8)
```

Geometry Options: Line



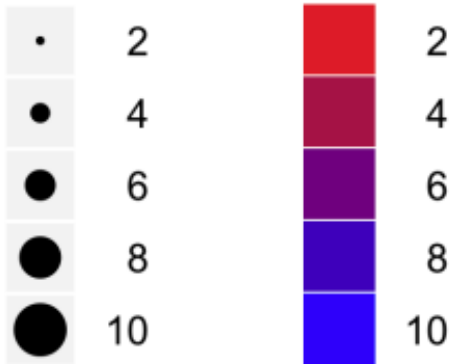
```
simple_point = (ggplot(demo,  
                      aes(color='Shape', y='y', x='x')) +  
               geom_line())  
simple_point.save("simple_line.pdf", scale=0.6,  
                 height=6, width=8)
```


Geometry Options: Bar



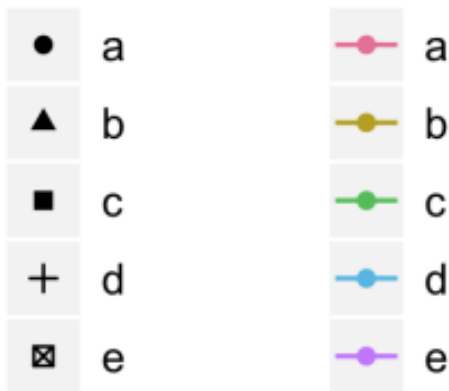
```
simple_point = (ggplot(demo,  
                    aes(color='Shape', y='y', x='x')) +  
              geom_bar(stat="identity"))  
simple_point.save("simple_bar.pdf", scale=0.6,  
                height=6, width=8)
```

Aesthetic Options: Continuous Data



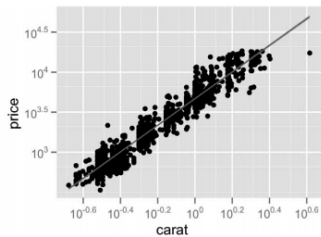
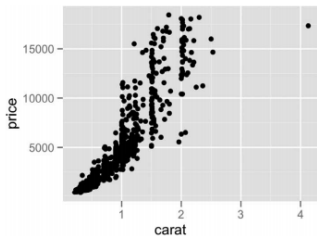
Size, color

Aesthetic Options: Discrete Data



Shape, color

Rescaling Data



```
+ scale_x_log10()
```