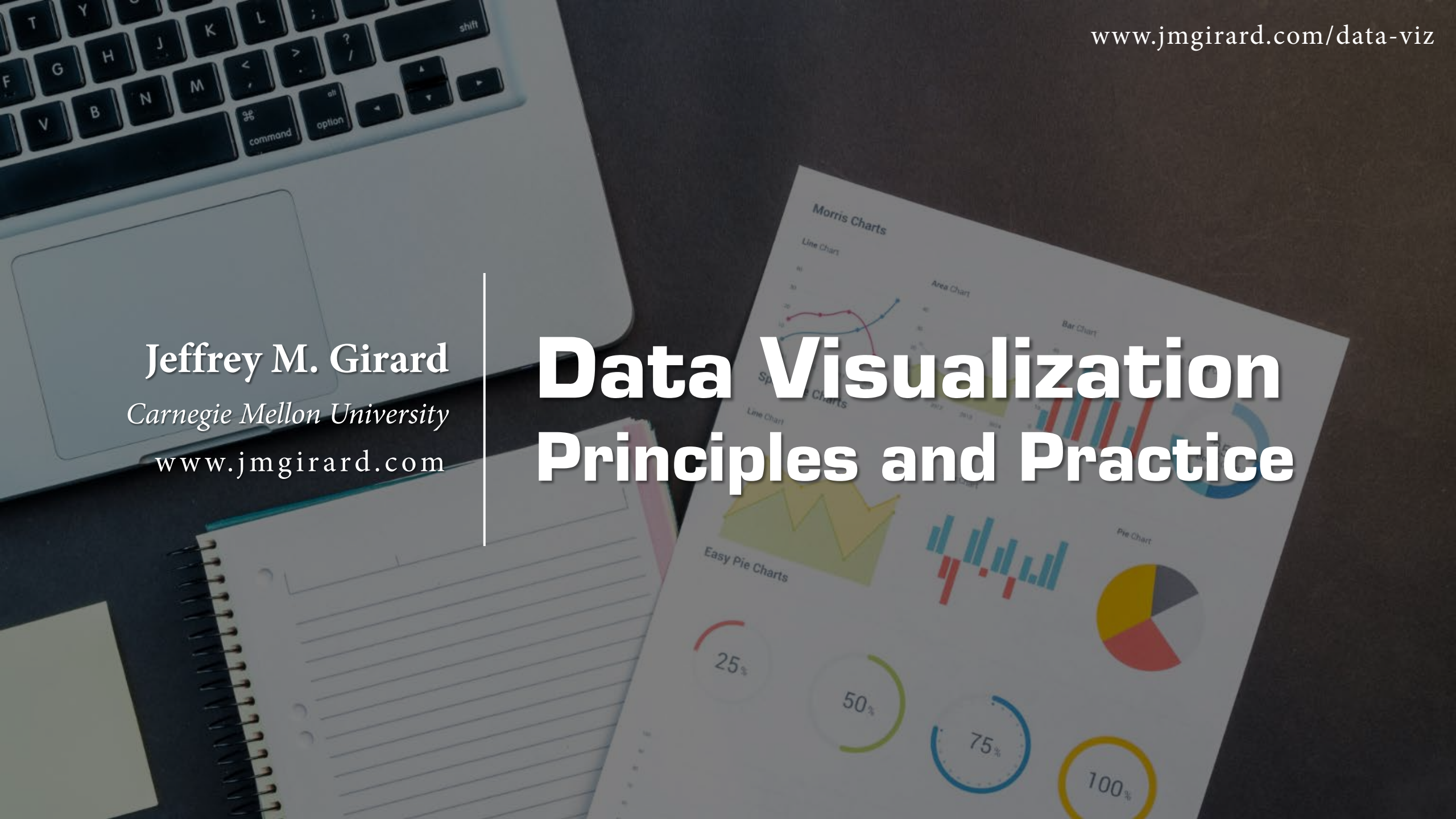
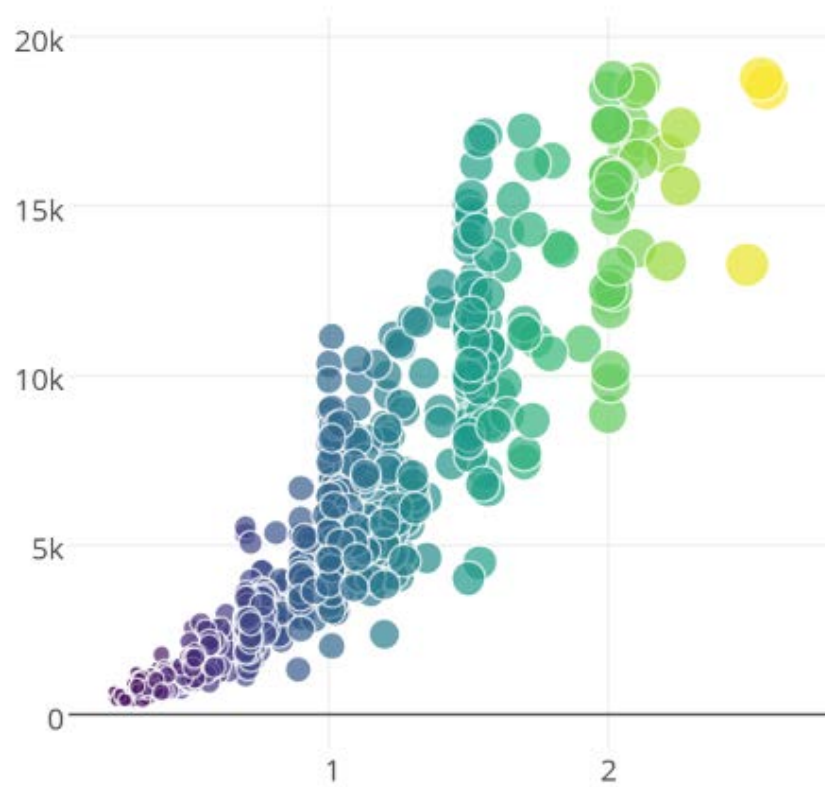


Jeffrey M. Girard
Carnegie Mellon University
www.jmgirard.com

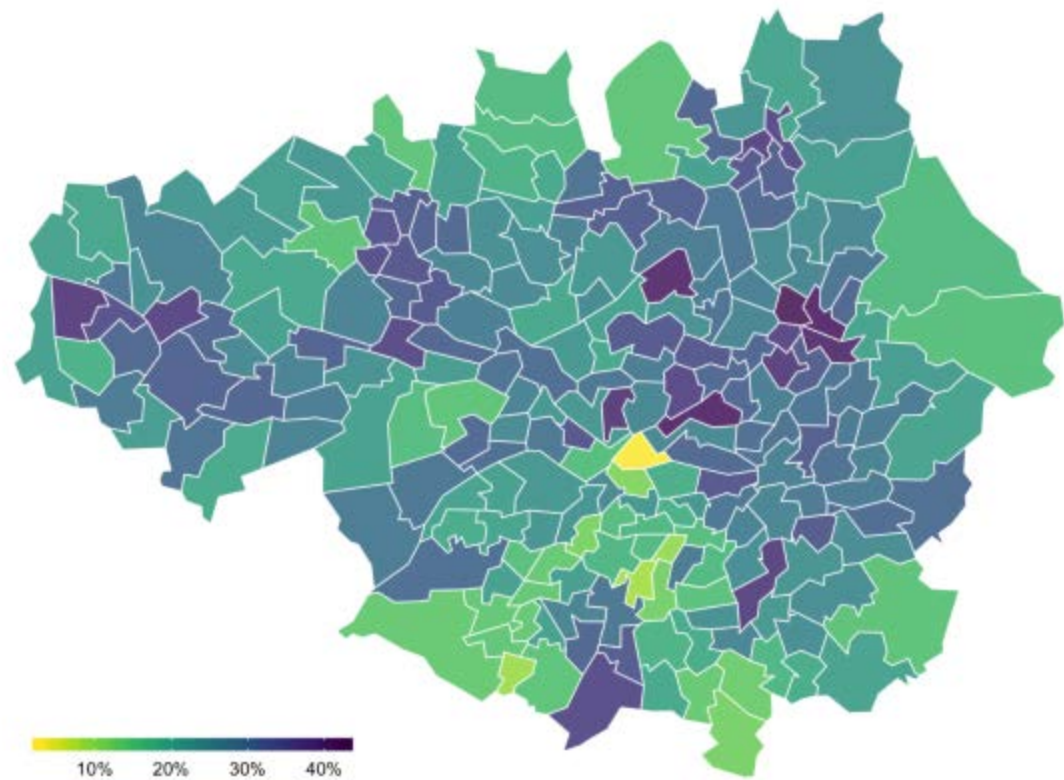
Data Visualization Principles and Practice



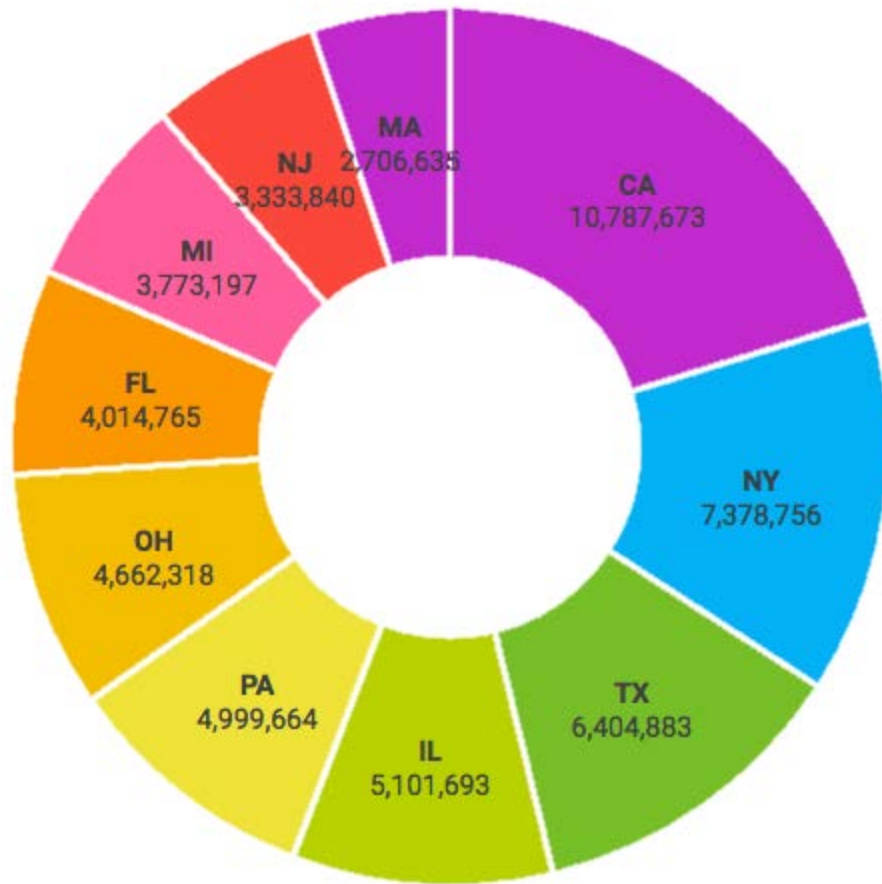
What is a graphic?



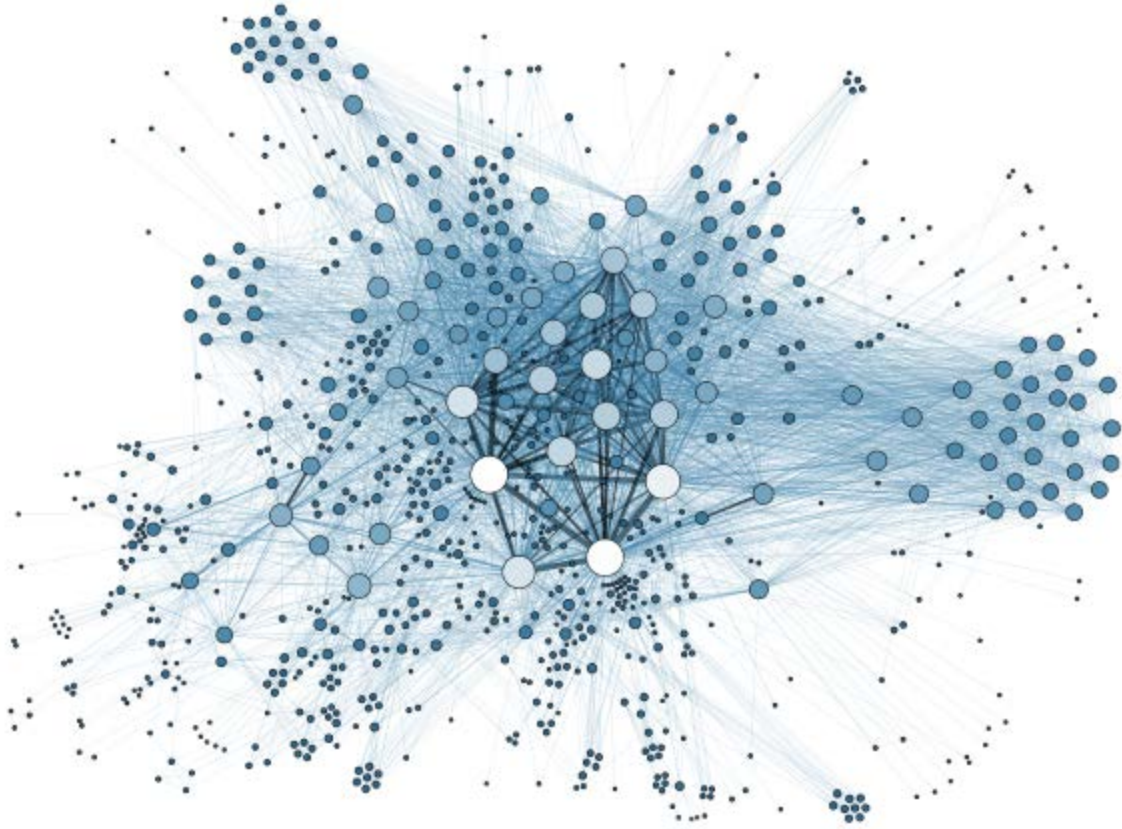
What is a graphic?



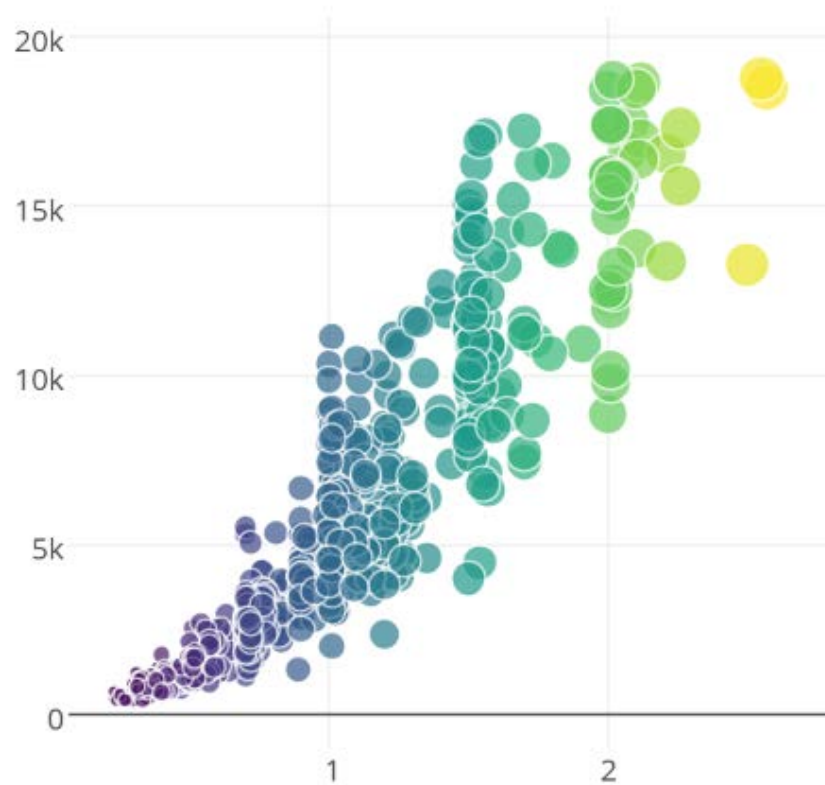
What is a graphic?



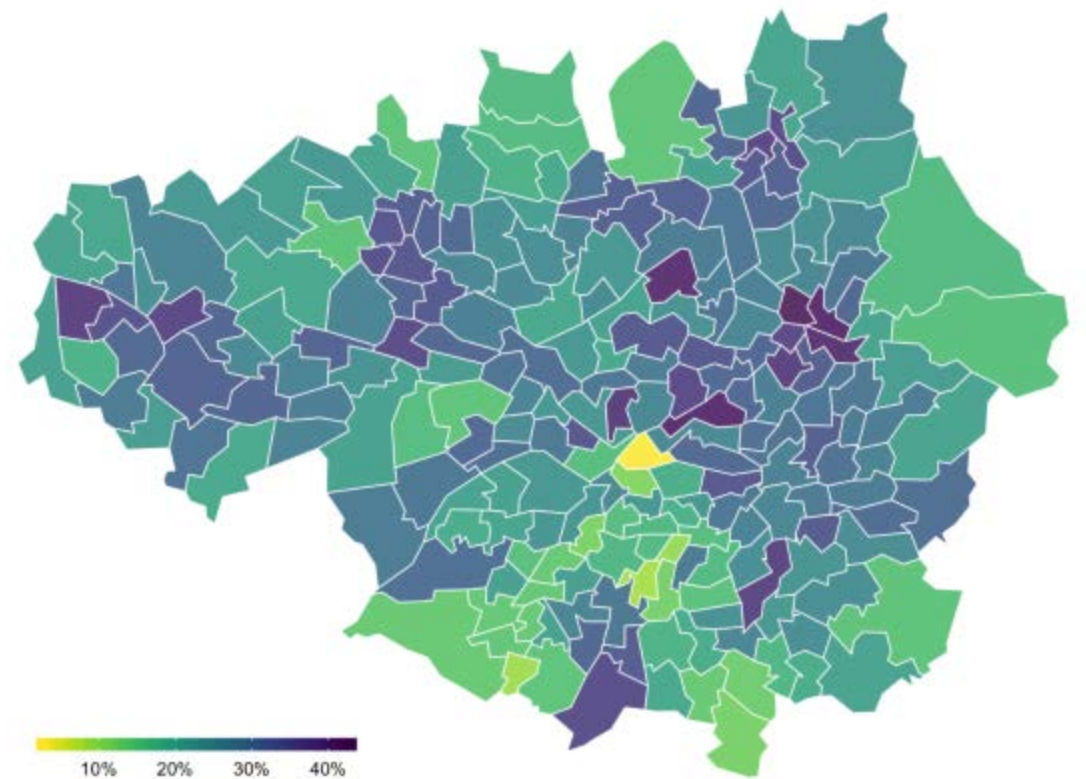
What is a graphic?



How to understand graphics?

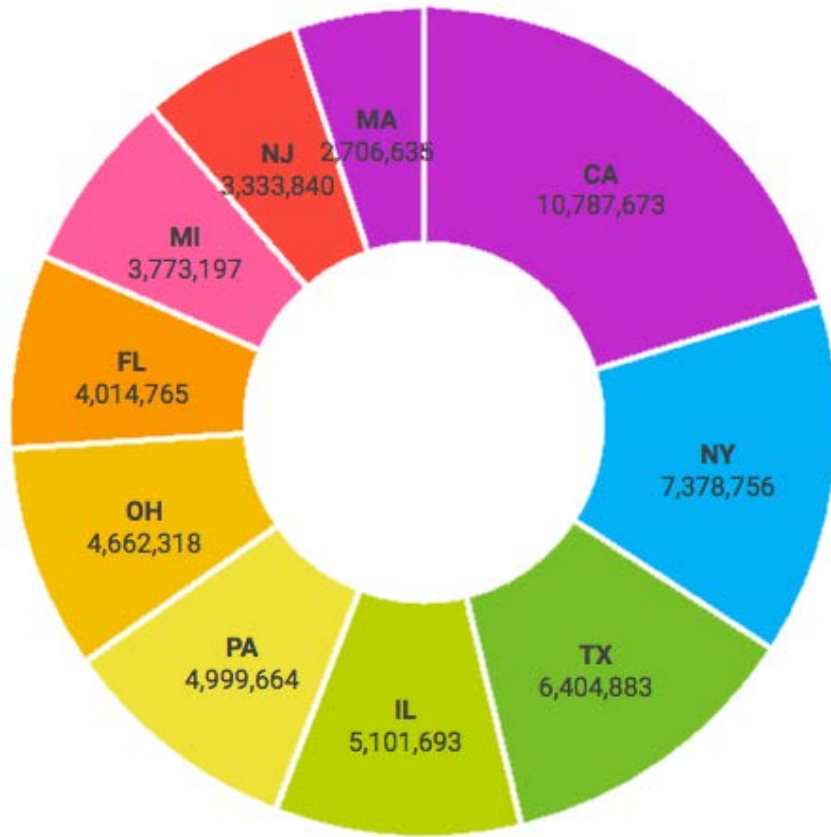


Bubble Chart

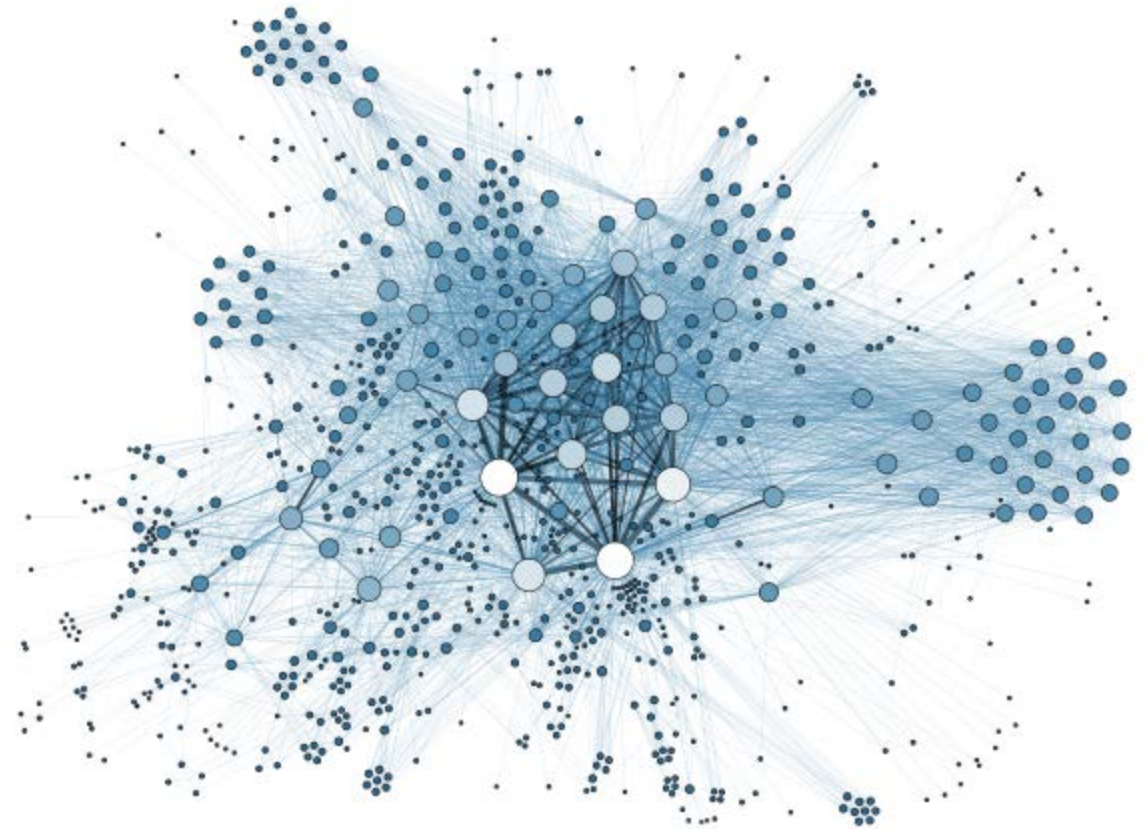


Choropleth Map

How to understand graphics?



Donut Chart

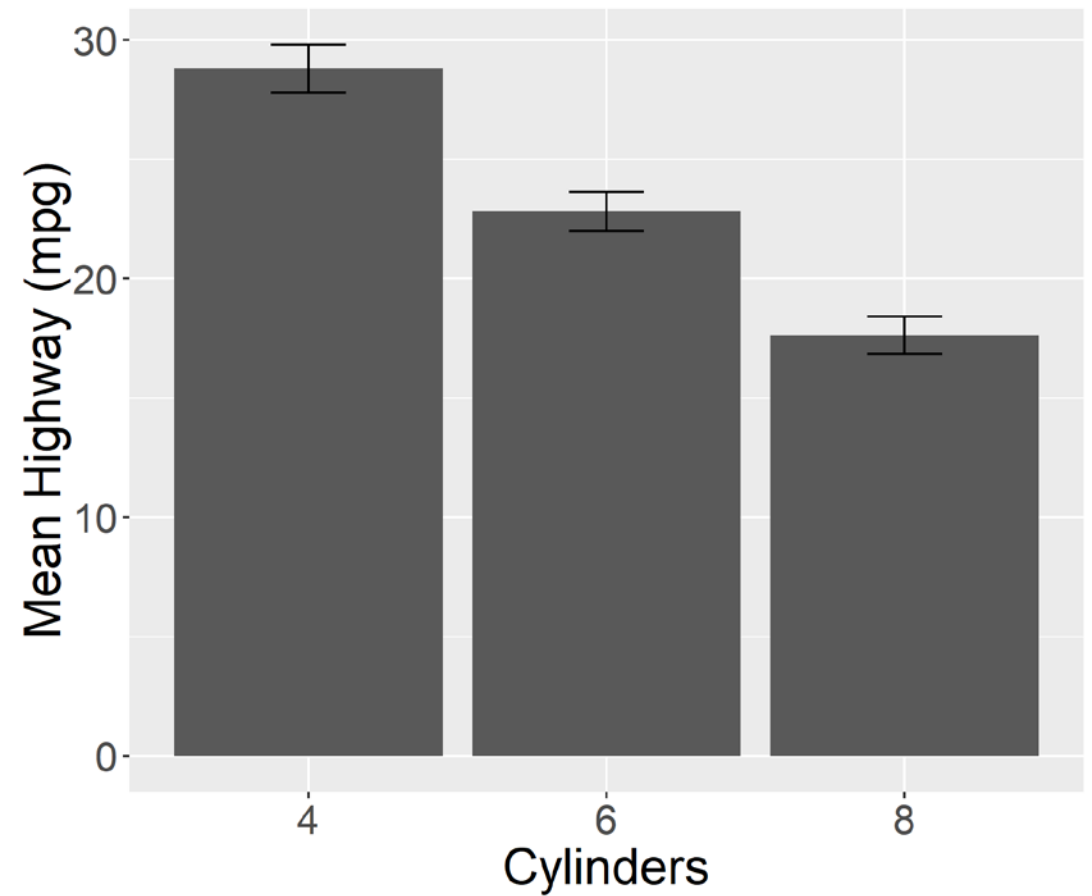
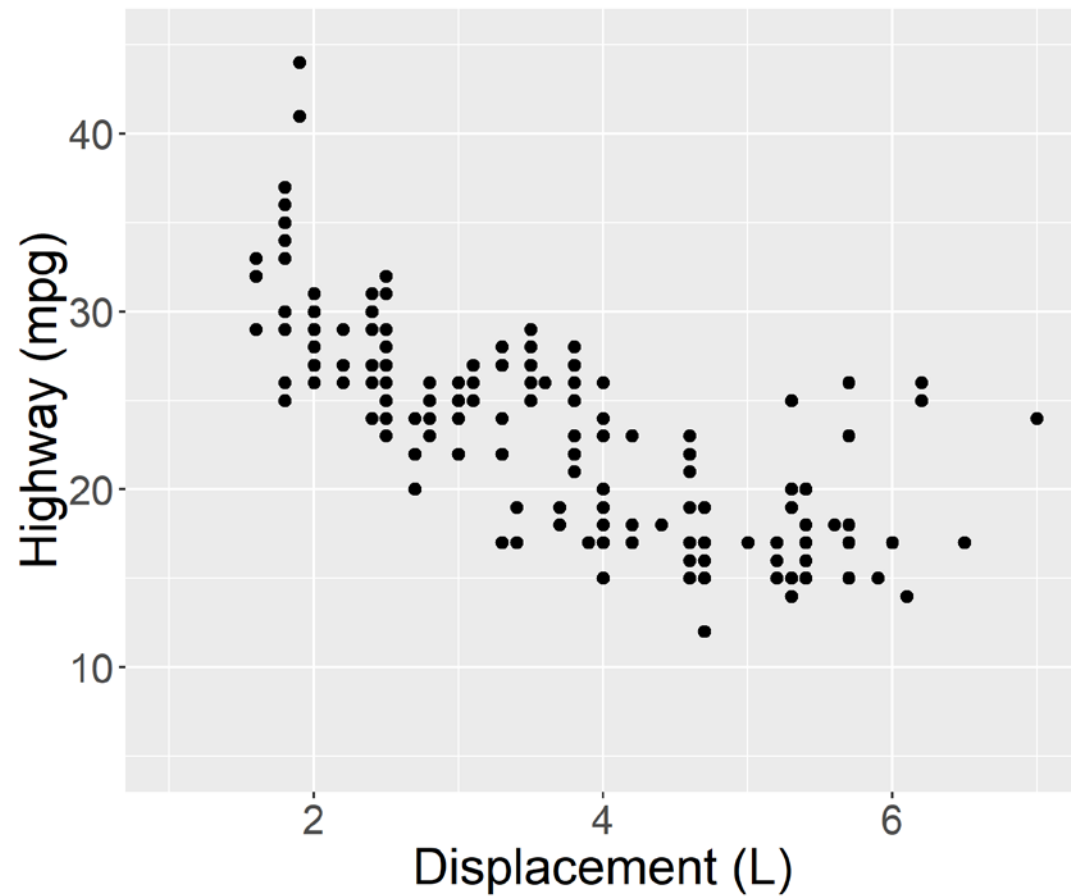


Network Diagram

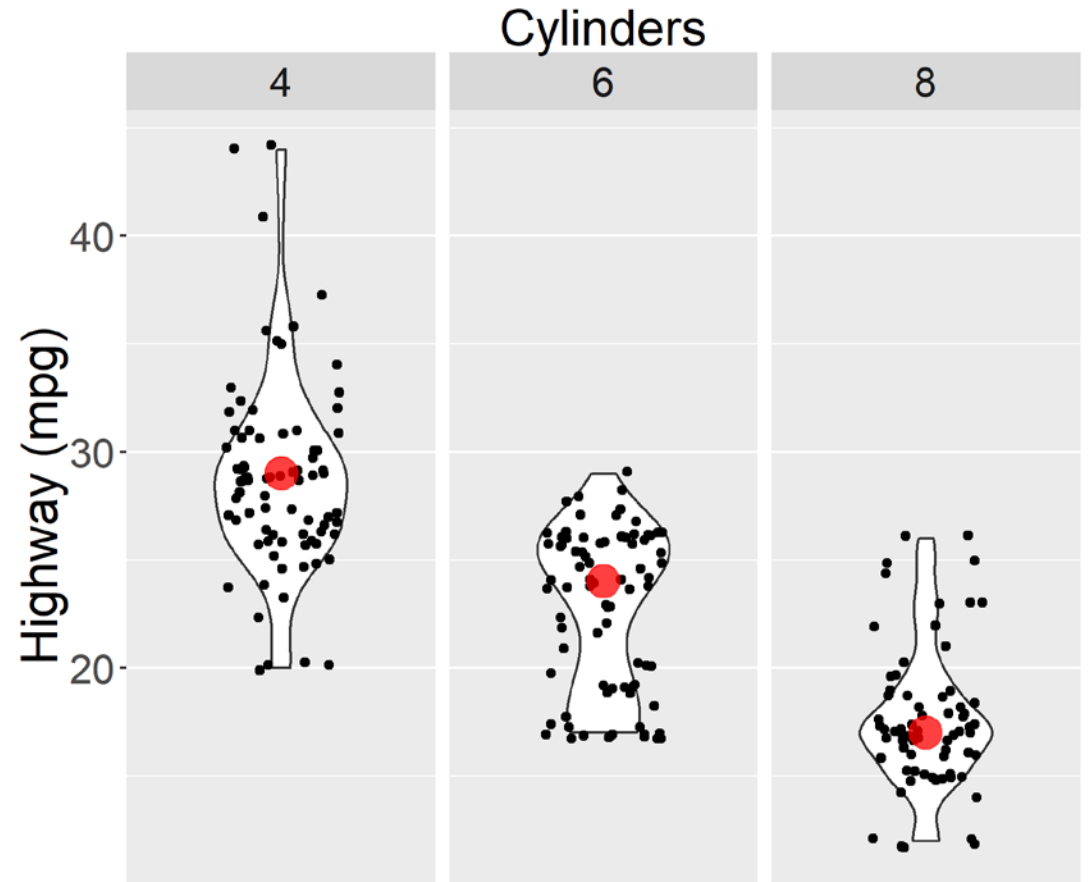
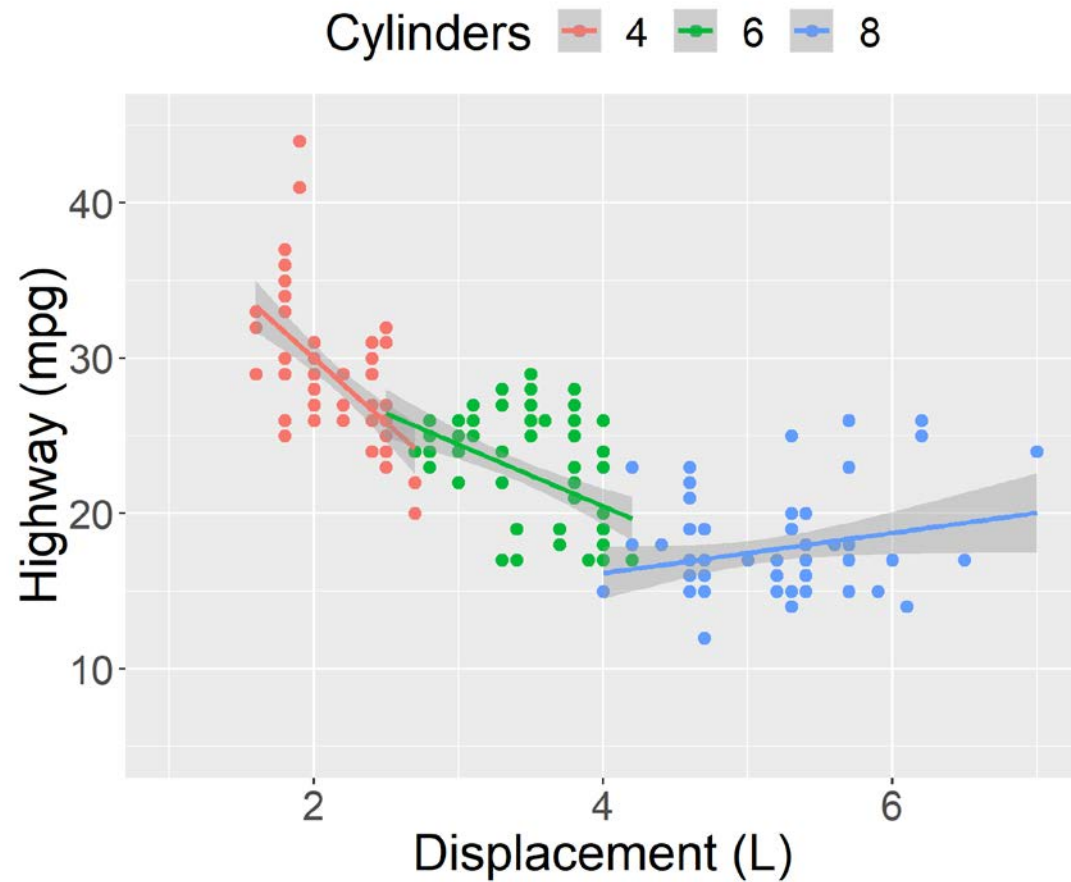
How to understand graphics?



How to understand graphics?



How to understand graphics?



How to understand graphics?

- To understand graphics in general and individual graphics in particular...
- We need a **grammar of graphics**: *fundamental principles or rules of an art or science*
- This will provide a strong foundation to **understand graphics** of diverse types
- Grammar can help us **create graphics** of high quality but is not a guarantee
- After all, you can be grammatically correct and still be speaking nonsense
- We will focus on Wickham's (2010) **layered grammar of graphics**
- This grammar is implemented in R through the **ggplot2** package

Basic Elements of the Grammar

- **Data**

describe observations using variables

- **Aesthetic Mappings**

map data variables to visual qualities

- **Scales**

map values in data space to values in aesthetic space (create axes and legends)

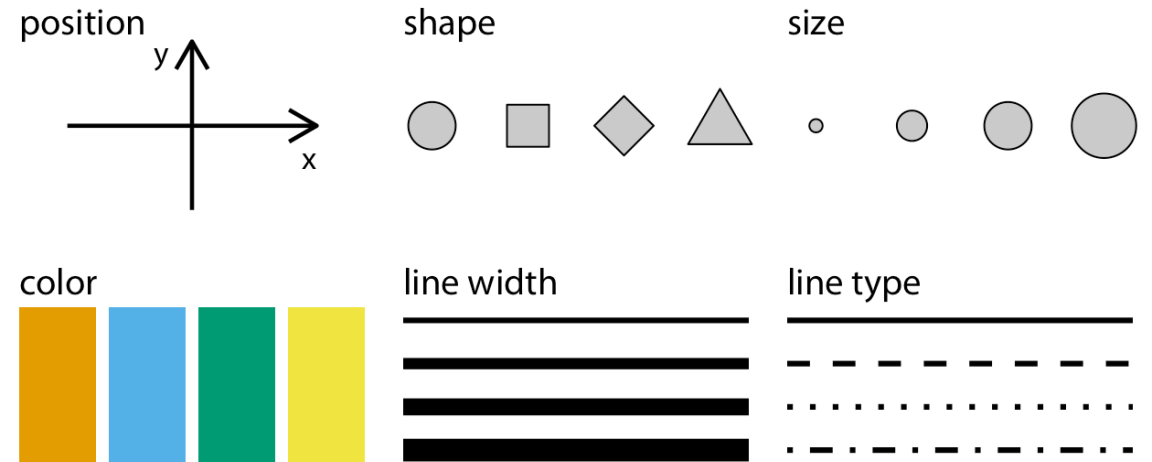
- **Geometric objects (geoms)**

constitute the objects seen on a plot

```
> cars
# A tibble: 230 x 8
  make  model  year displ  cyl  class  hwy  cty
  <chr> <chr>   <int> <dbl> <fct> <fct> <int> <int>
1 audi  a4      1999  1.8  4     compact 29    18
2 audi  a4      1999  1.8  4     compact 29    21
3 audi  a4      2008  2    4     compact 31    20
4 audi  a4      2008  2    4     compact 30    21
5 audi  a4      1999  2.8  6     compact 26    16
6 audi  a4      1999  2.8  6     compact 26    18
7 audi  a4      2008  3.1  6     compact 27    18
8 audi  a4 quattro 1999  1.8  4     compact 26    18
9 audi  a4 quattro 1999  1.8  4     compact 25    16
10 audi a4 quattro 2008  2    4     compact 28    20
# ... with 220 more rows
>
```

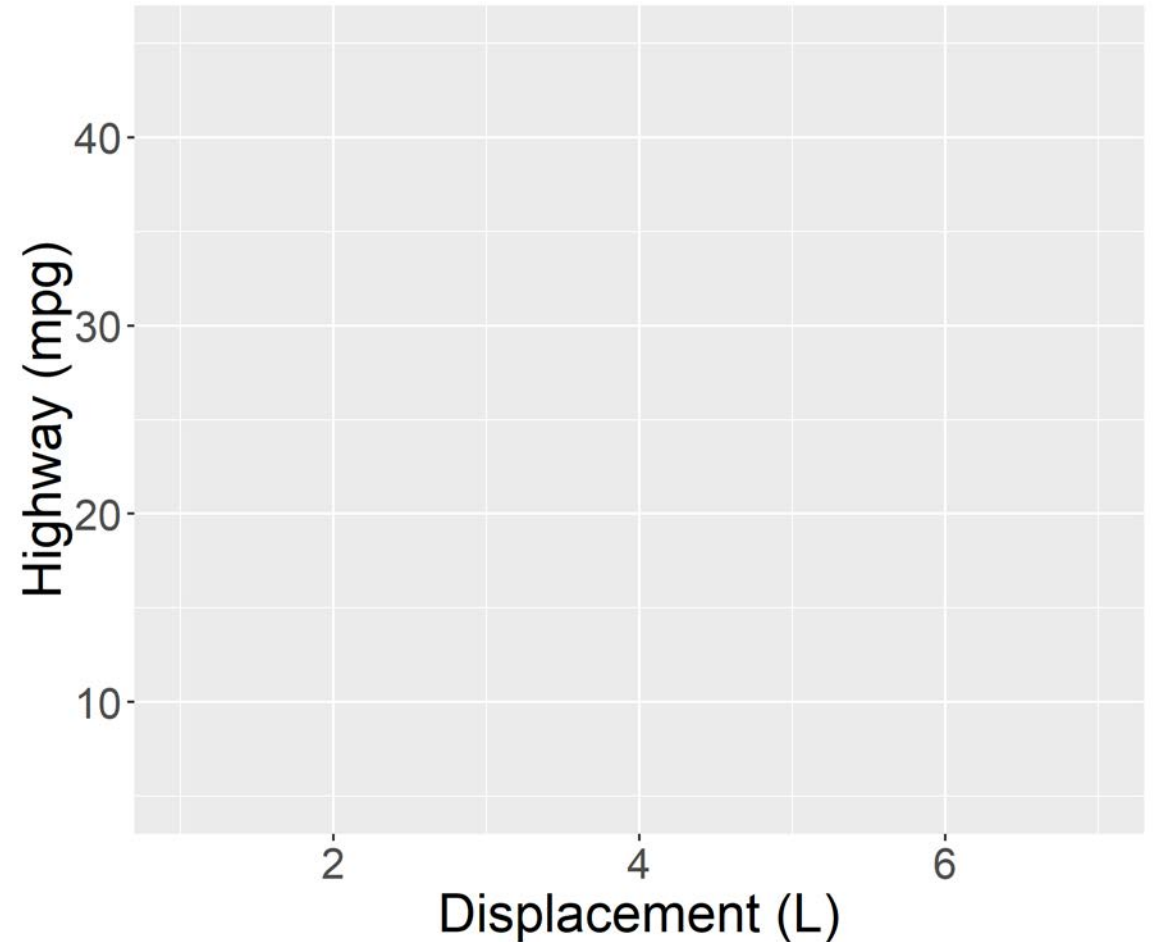

Basic Elements of the Grammar

- **Data**
describe observations using variables
- **Aesthetic Mappings**
map data variables to visual qualities
- **Scales**
map values in data space to values in aesthetic space (create axes and legends)
- **Geometric objects (geoms)**
constitute the objects seen on a plot



Basic Elements of the Grammar

- **Data**
describe observations using variables
- **Aesthetic Mappings**
map data variables to visual qualities
- **Scales**
map values in data space to values in aesthetic space (create axes and legends)
- **Geometric objects (geoms)**
constitute the objects seen on a plot



Basic Elements of the Grammar

- **Data**
describe observations using variables
- **Aesthetic Mappings**
map data variables to visual qualities
- **Scales**
map values in data space to values in aesthetic space (create axes and legends)
- **Geometric objects (geoms)**
constitute the objects seen on a plot

Cylinders  4  6  8

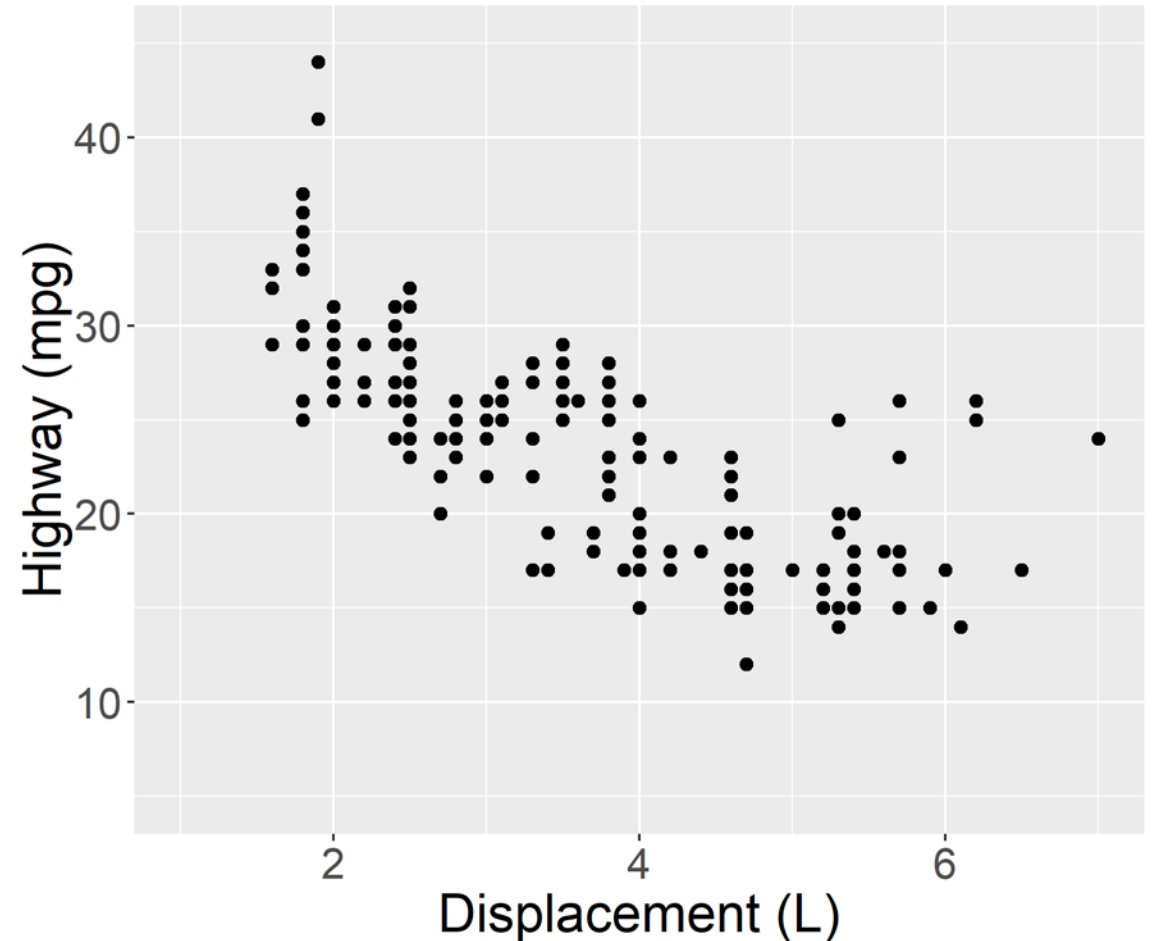
Displacement 
2 3 4 5 6 7

Cylinders  4  6  8

Displacement  2.0  4.5  7.0

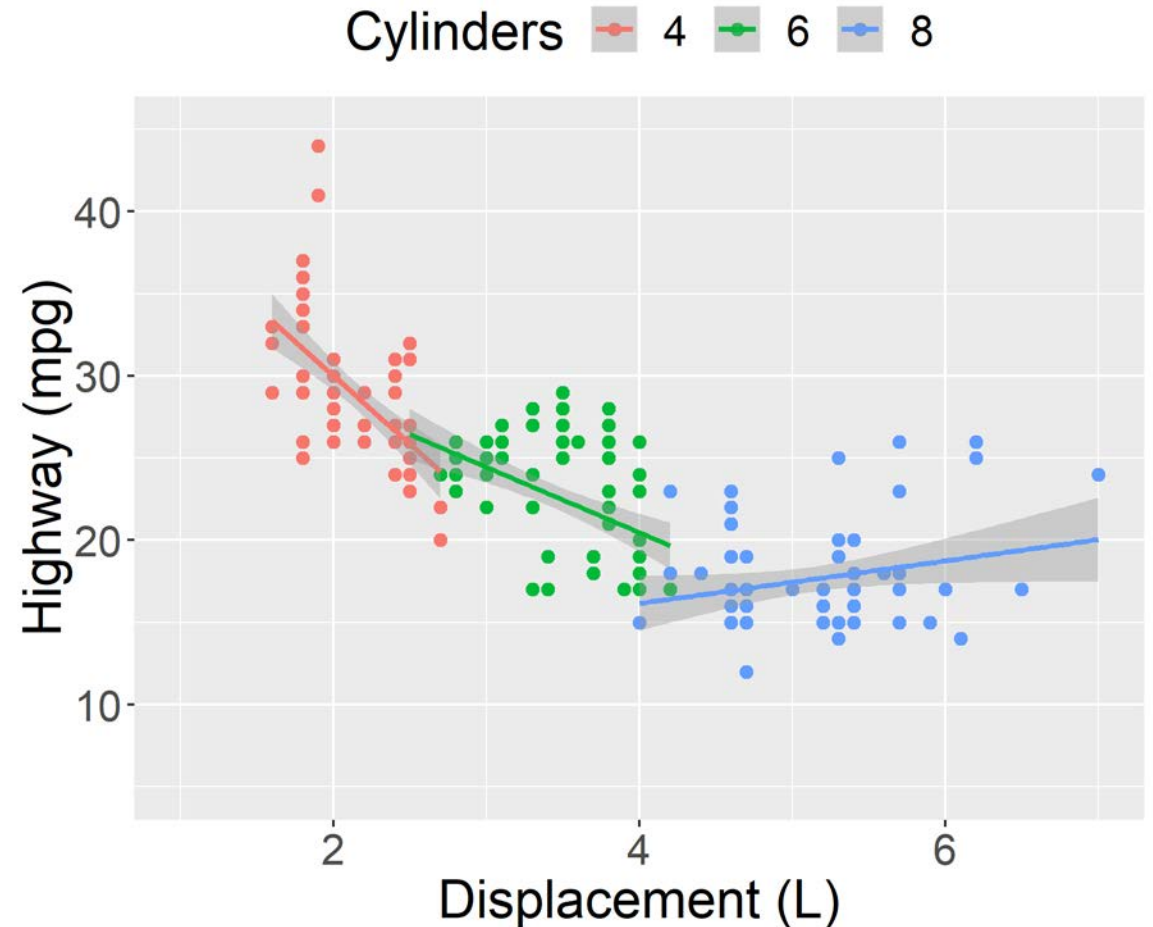
Basic Elements of the Grammar

- **Data**
describe observations using variables
- **Aesthetic Mappings**
map data variables to visual qualities
- **Scales**
map values in data space to values in aesthetic space (create axes and legends)
- **Geometric objects (geoms)**
constitute the objects seen on a plot



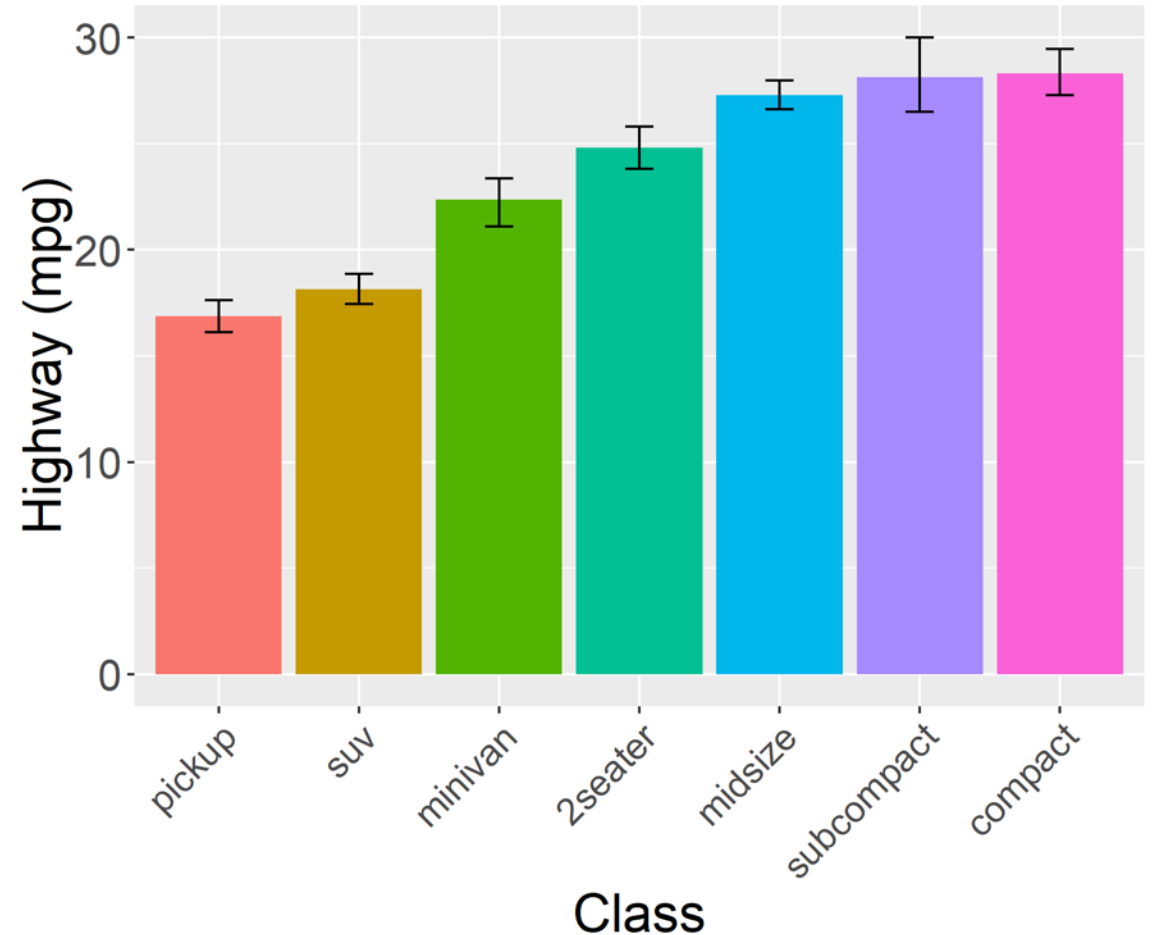
Basic Elements of the Grammar

- **Data**
describe observations using variables
- **Aesthetic Mappings**
map data variables to visual qualities
- **Scales**
map values in data space to values in aesthetic space (create axes and legends)
- **Geometric objects (geoms)**
constitute the objects seen on a plot



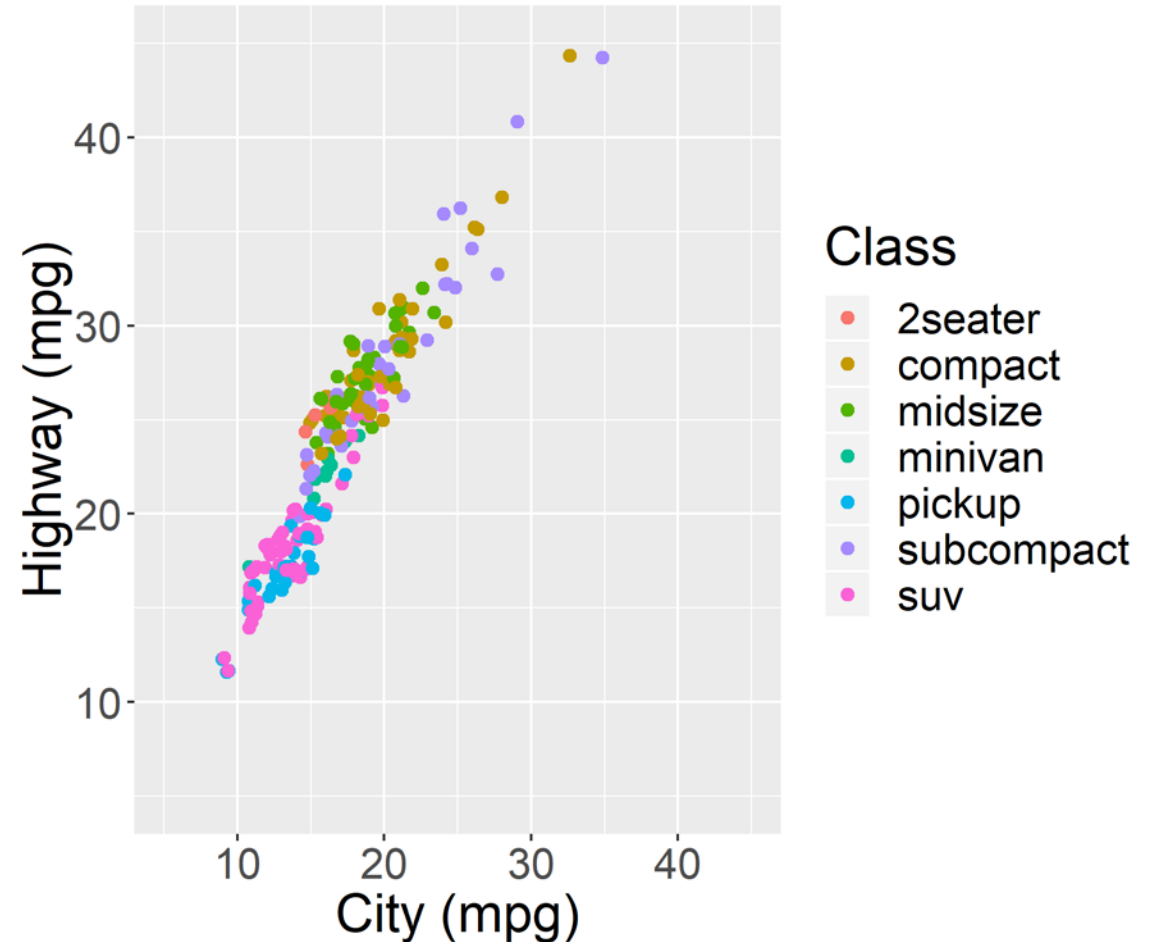
Advanced Elements of the Grammar

- **Statistical Transformations (stats)**
summarize and manipulate data values
- **Coordinate System**
controls scale and geom positioning
- **Faceting Specification**
displays subsets of data in separate axes
- **Theme**
controls the finer points of display



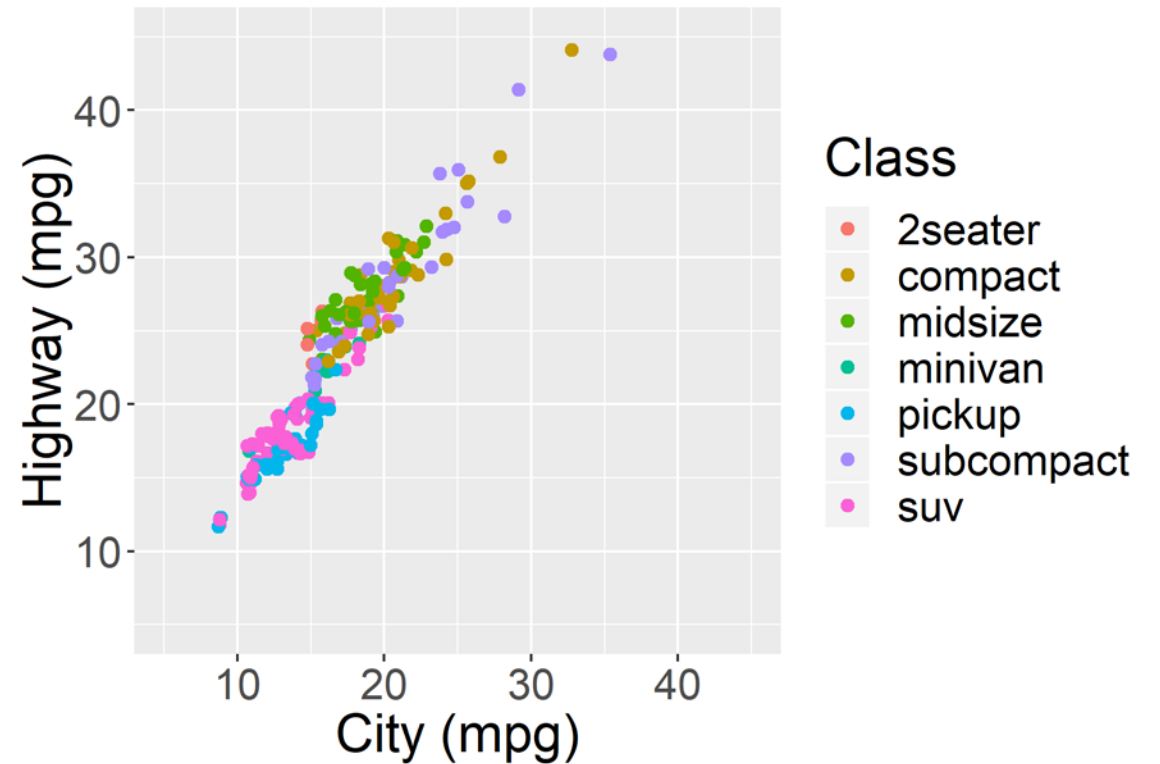
Advanced Elements of the Grammar

- **Statistical Transformations (stats)**
summarize and manipulate data values
- **Coordinate System**
controls scale and geom positioning
- **Faceting Specification**
displays subsets of data in separate axes
- **Theme**
controls the finer points of display



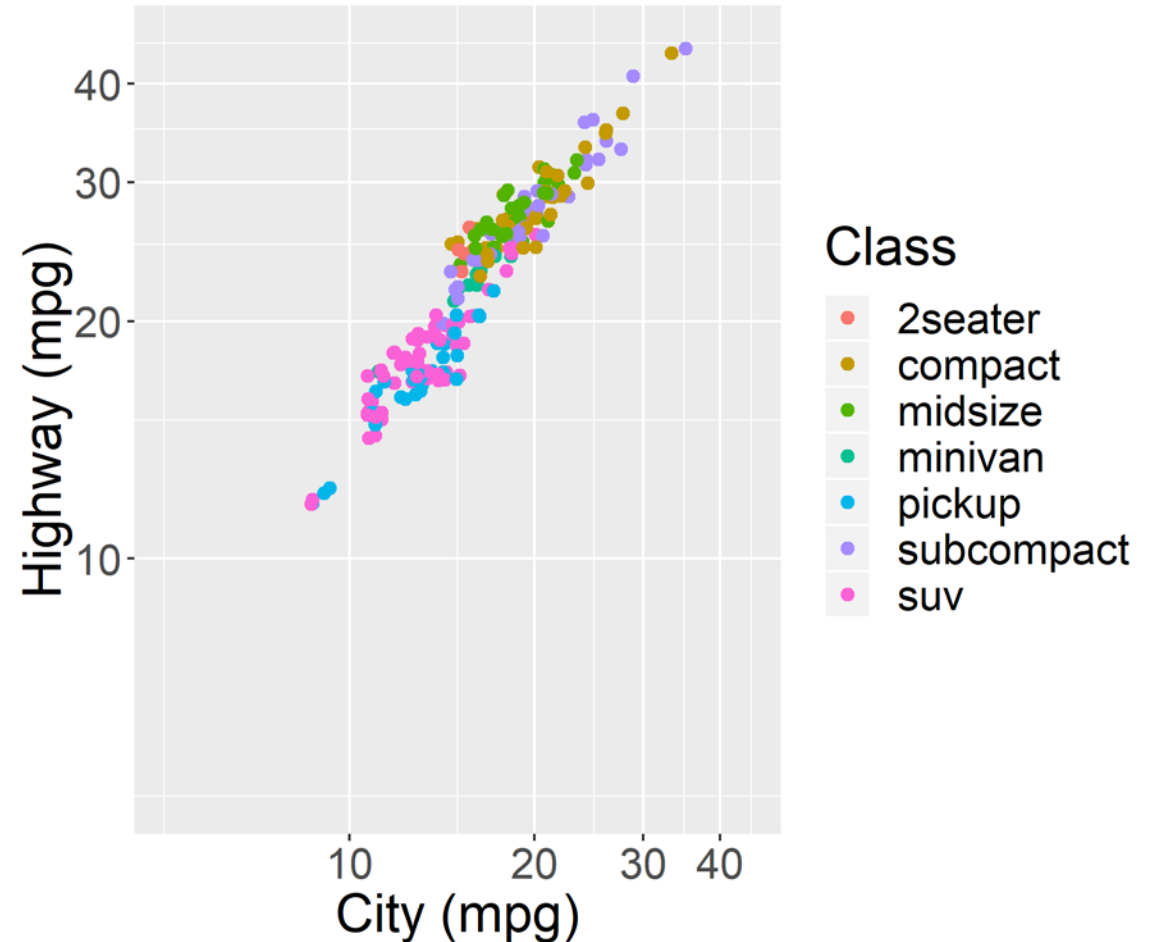
Advanced Elements of the Grammar

- **Statistical Transformations (stats)**
summarize and manipulate data values
- **Coordinate System**
controls scale and geom positioning
- **Faceting Specification**
displays subsets of data in separate axes
- **Theme**
controls the finer points of display



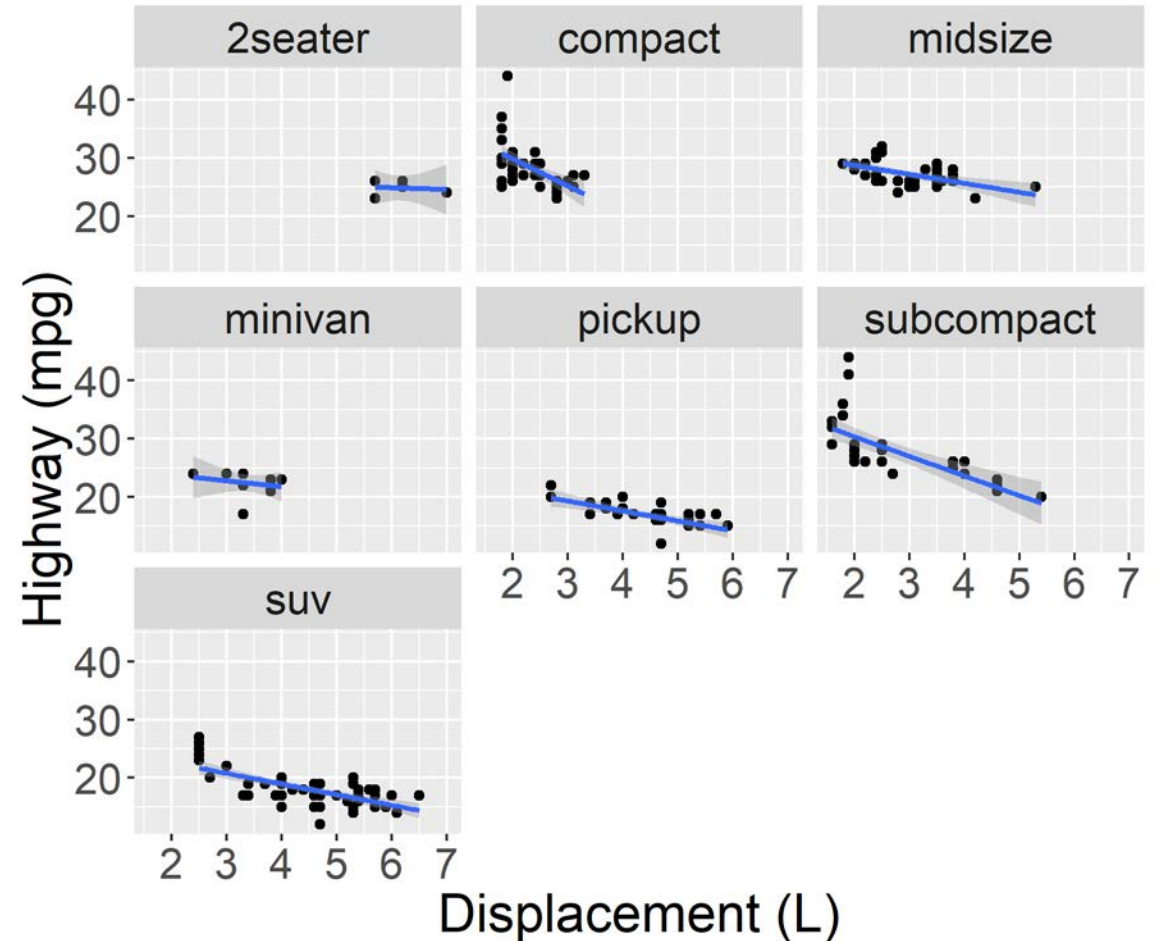
Advanced Elements of the Grammar

- **Statistical Transformations (stats)**
summarize and manipulate data values
- **Coordinate System**
controls scale and geom positioning
- **Faceting Specification**
displays subsets of data in separate axes
- **Theme**
controls the finer points of display



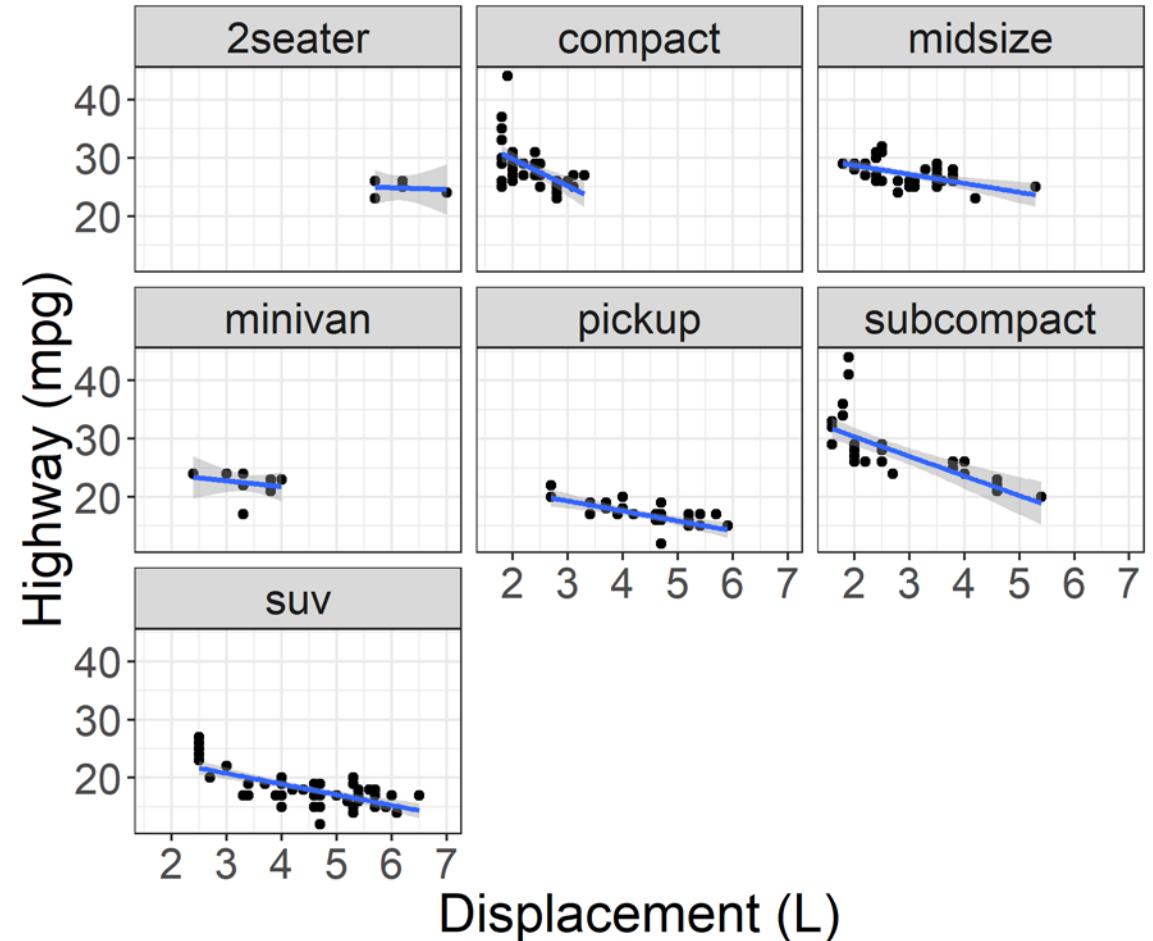
Advanced Elements of the Grammar

- **Statistical Transformations (stats)**
summarize and manipulate data values
- **Coordinate System**
controls scale and geom positioning
- **Faceting Specification**
displays subsets of data in separate axes
- **Theme**
controls the finer points of display



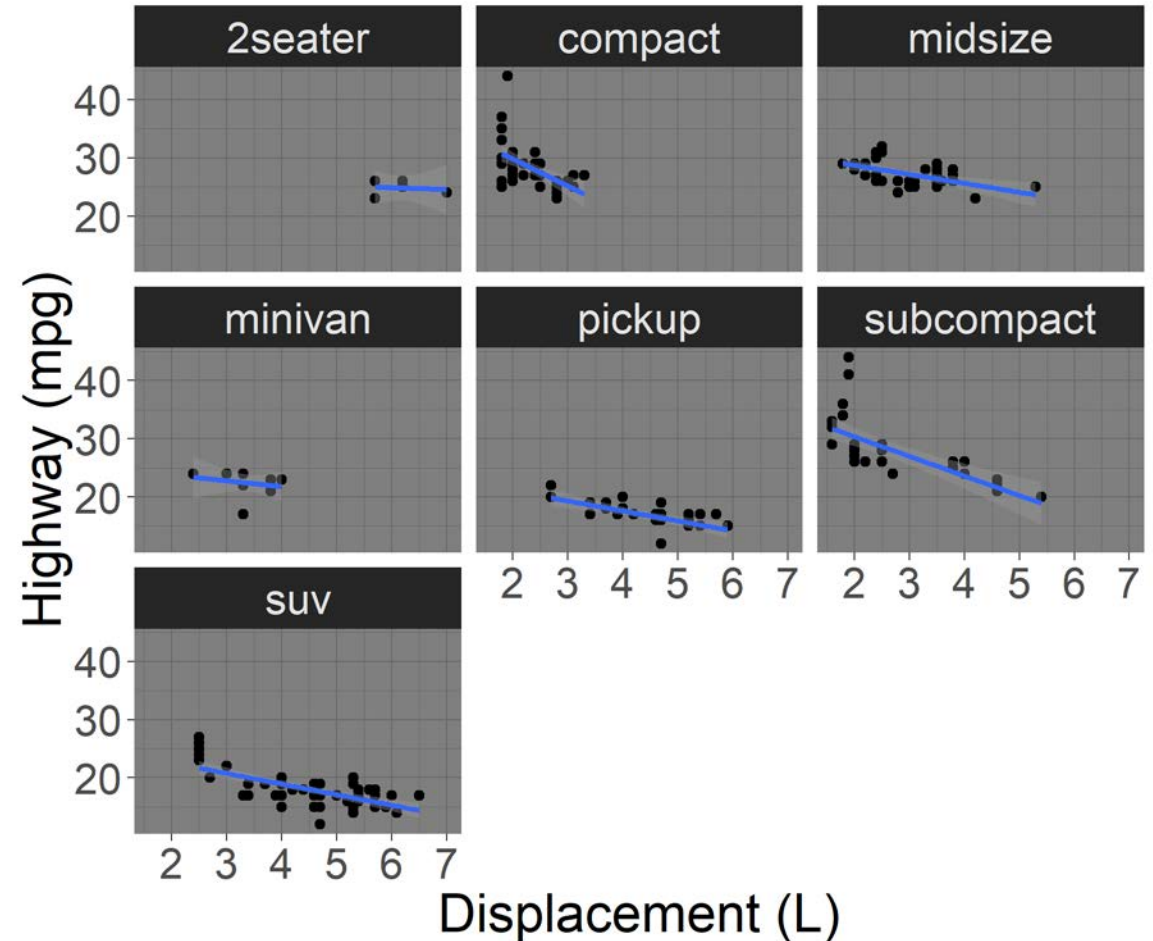
Advanced Elements of the Grammar

- **Statistical Transformations (stats)**
summarize and manipulate data values
- **Coordinate System**
controls scale and geom positioning
- **Faceting Specification**
displays subsets of data in separate axes
- **Theme**
controls the finer points of display



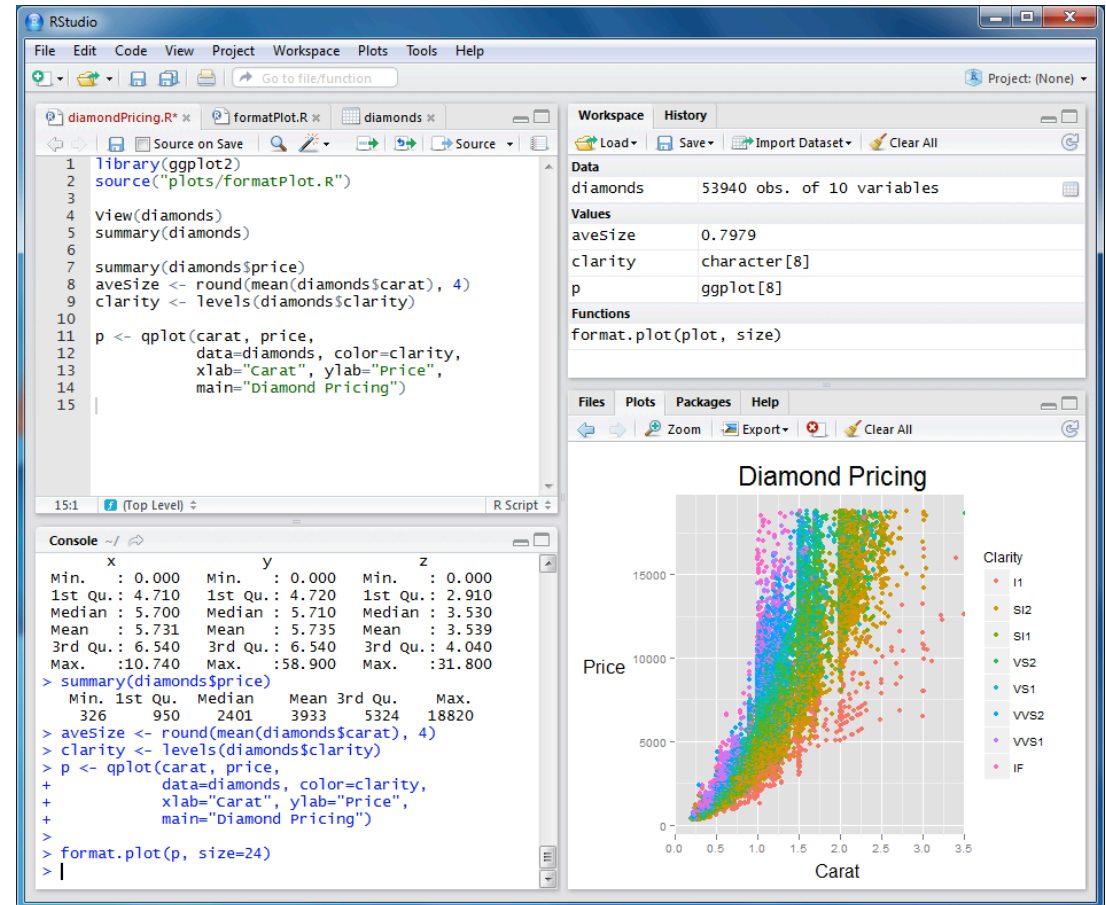
Advanced Elements of the Grammar

- **Statistical Transformations (stats)**
summarize and manipulate data values
- **Coordinate System**
controls scale and geom positioning
- **Faceting Specification**
displays subsets of data in separate axes
- **Theme**
controls the finer points of display



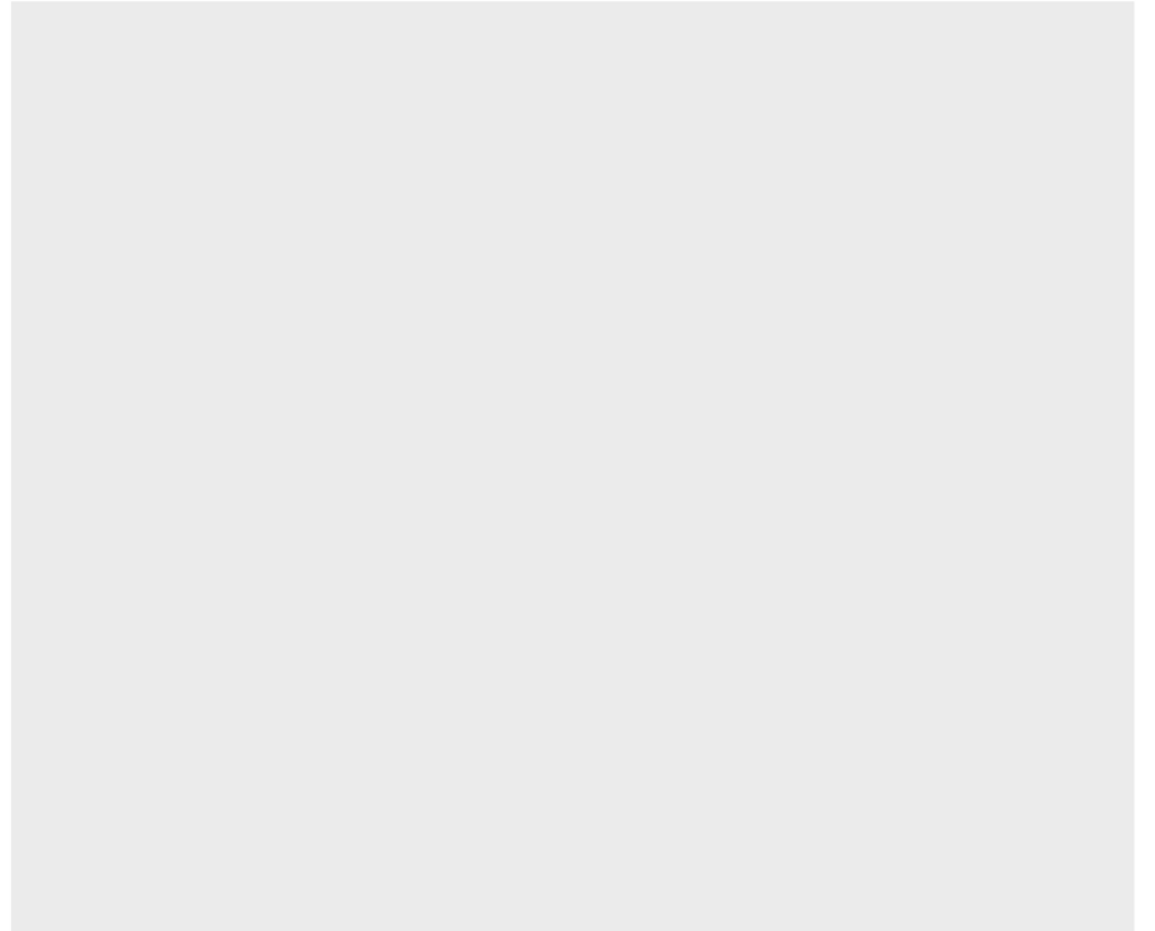
Introduction to ggplot2

- Download and install R
 - cloud.r-project.org
- Download and install RStudio Desktop
 - www.rstudio.com/download
- Open RStudio Desktop
- Install the **ggplot2** package (first time)
 - `install.packages("ggplot2")`
- Load the **ggplot2** package (every time)
 - `library(ggplot2)`



Introduction to ggplot2

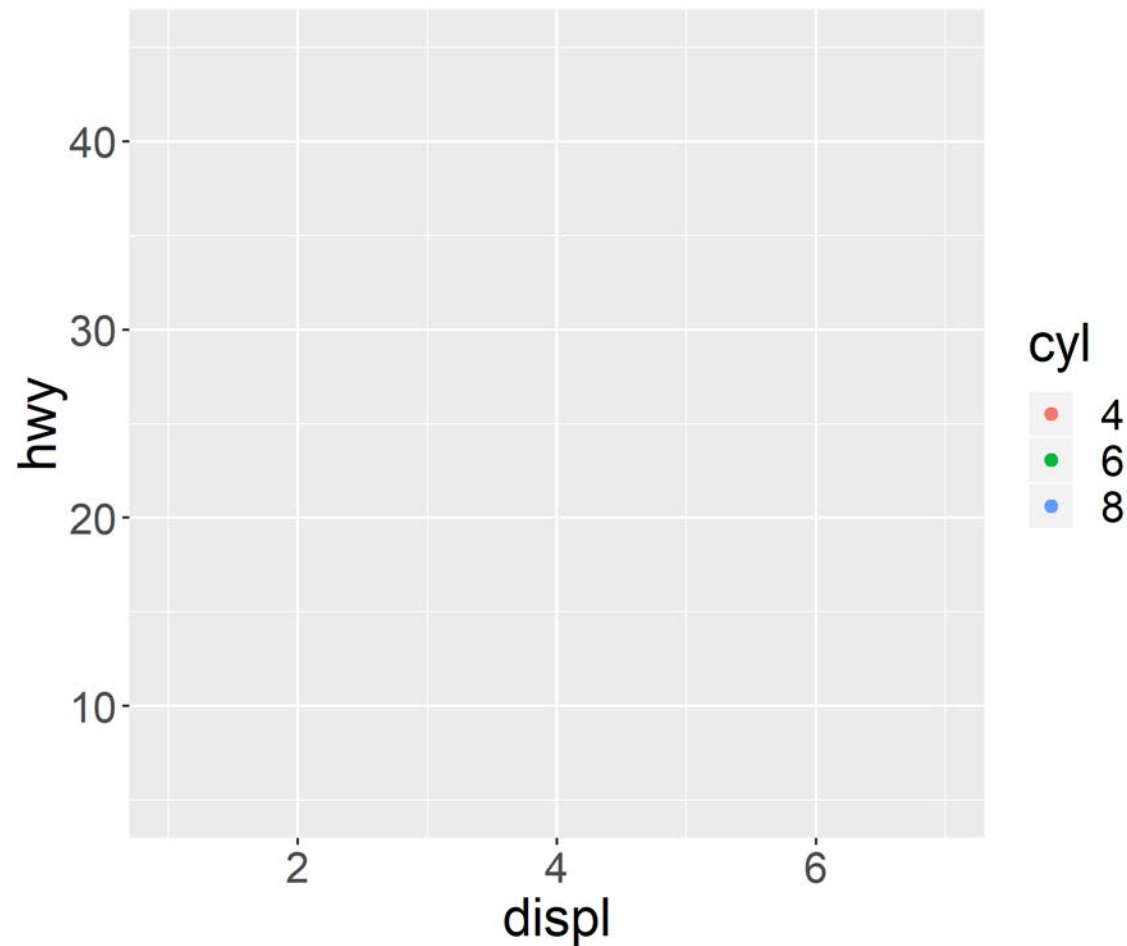
```
# Create graphic and add data  
layer0 <- ggplot(data = cars)
```



Introduction to ggplot2

Add aesthetic mappings

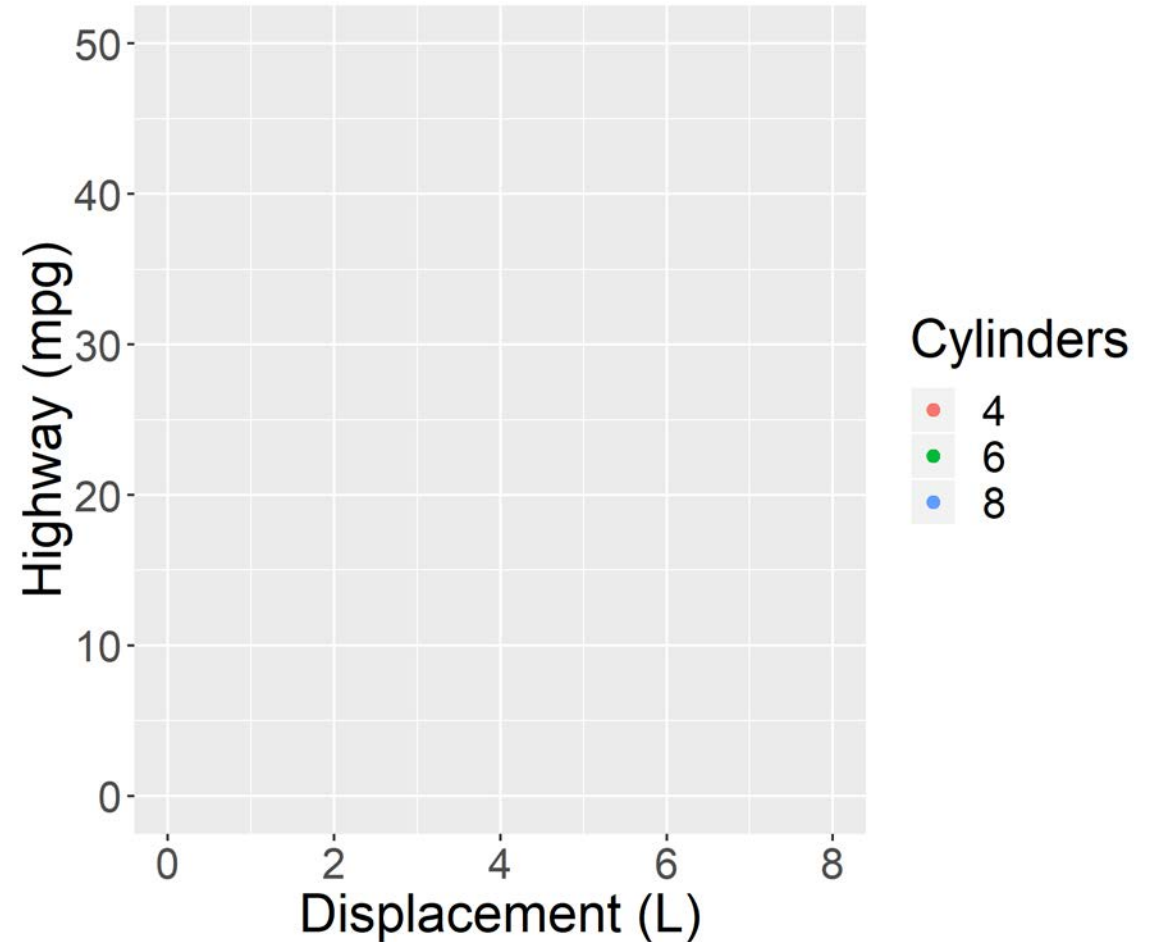
```
layer1 <- ggplot(  
  data = cars,  
  mapping = aes(  
    x = displ,  
    y = hwy,  
    color = cyl  
  )  
)
```



Introduction to ggplot2

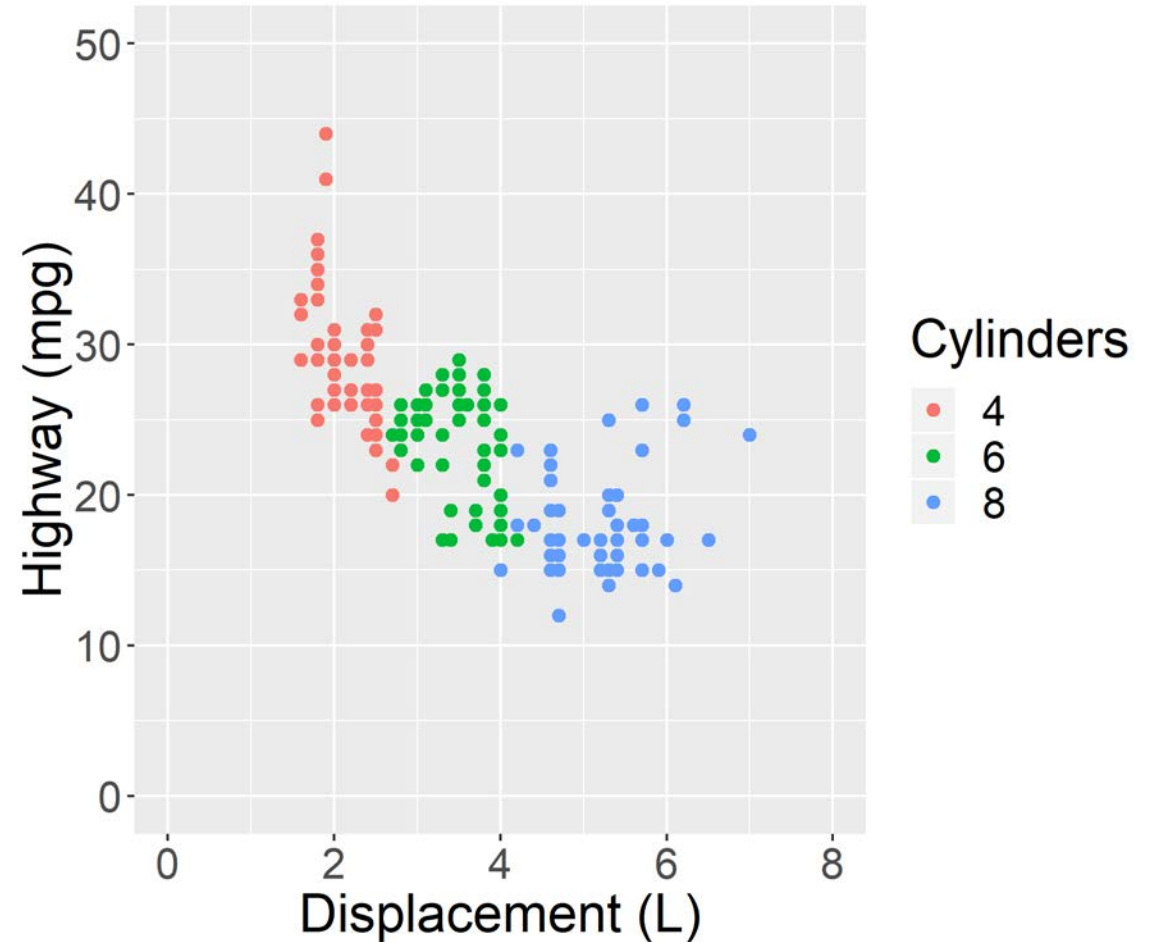
```
# Configure x, y, and color scales
```

```
layer2 <- layer1 +  
  scale_x_continuous(  
    name = "Displacement (L)",  
    limits = c(0, 8)  
  ) +  
  scale_y_continuous(  
    name = "Highway (mpg)",  
    limits = c(0, 50)  
  ) +  
  scale_color_discrete(  
    name = "Cylinders"  
  )
```



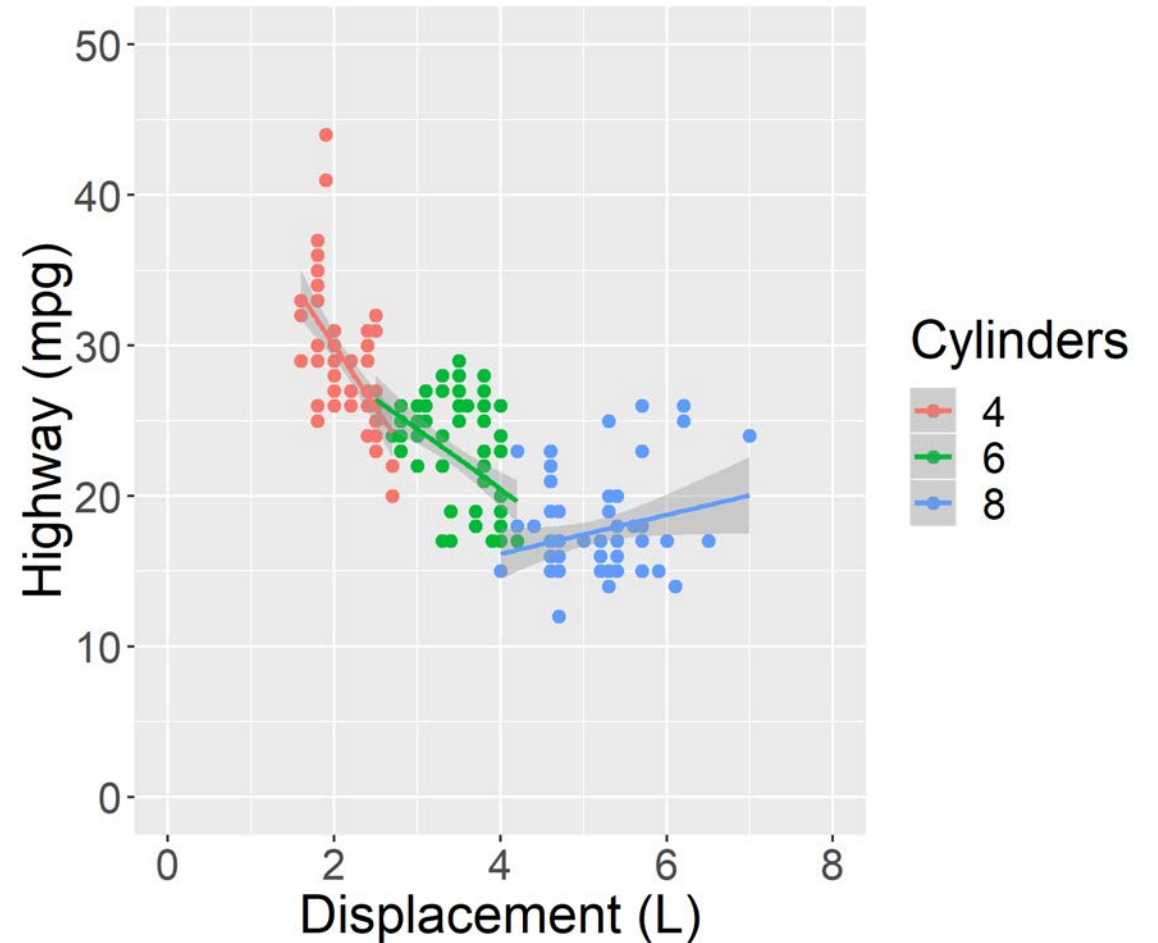
Introduction to ggplot2

```
# Plot data as points based on (x, y)
layer3 <- layer2 +
  geom_point(
    shape = "circle",
    size = 2
  )
```



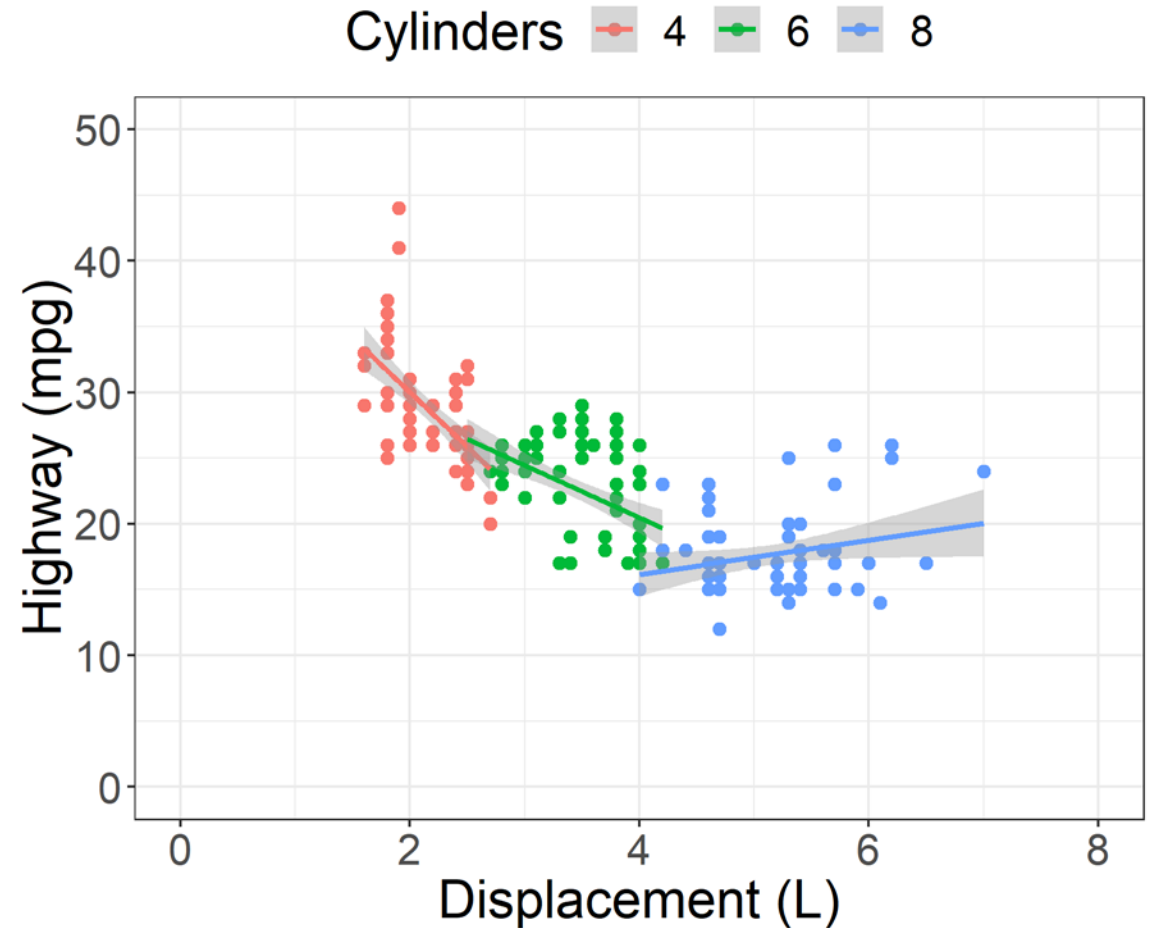
Introduction to ggplot2

```
# Add linear model plots  
layer4 <- layer3 +  
  geom_smooth(  
    method = "lm",  
    se = TRUE  
  )
```



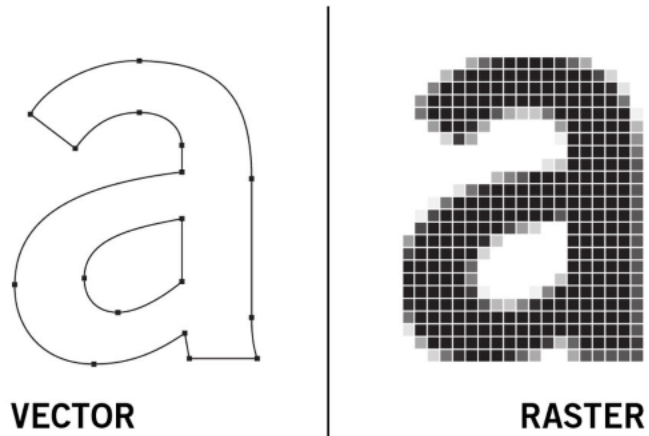
Introduction to ggplot2

```
# Configure theme for printing  
layer5 <- layer4 +  
  theme_bw() +  
  theme(legend.position = "top")
```



Exporting Graphics from ggplot2

- ggplots are created and saved as vectors
- They can be exported in various formats
 - Vectors: SVG, PDF, EPS, etc.
 - Rasters: JPEG, PNG, TIFF, etc.



```
# Save plot as PNG for PowerPoint
# Set to 6x5" at 300 dots per inch
ggsave(
  file = "fig1.png",
  plot = layer5,
  width = 6,
  height = 5,
  units = "in",
  dpi = 300
)
```

Visualization Workflow

Define before you design!

1. Define your **purpose**
2. Define your **audience**
3. Define your **data**
4. Define your **message**
5. Define your **values**

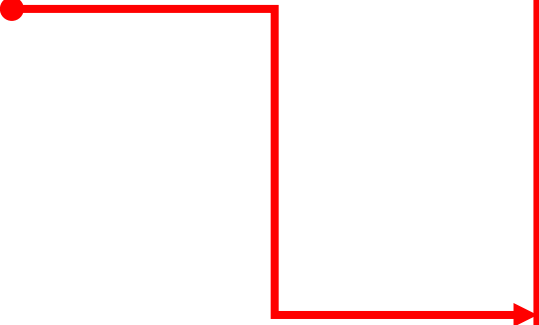
Design is a process!

6. Consider the **grammar**
7. Create some **prototypes**
8. Solicit **feedback**
9. **Update** your designs
10. **Iterate** until satisfied

Visualization Workflow

Define before you design!

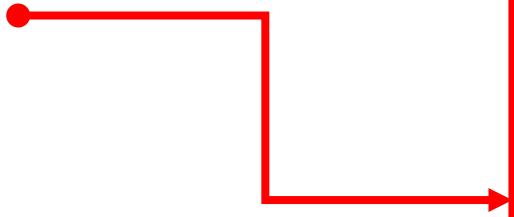
1. Define your **purpose**
2. Define your **audience**
3. Define your **data**
4. Define your **message**
5. Define your **values**

- 
- Why am I creating this graphic?
 - What are my goals for this graphic?
 - What level of "polish" is needed?
 - What format will it be displayed in?
 - What constraints on design exist?

Visualization Workflow

Define before you design!


1. Define your **purpose**
2. Define your **audience**
3. Define your **data**
4. Define your **message**
5. Define your **values**

- 
- Who is the graphic intended for?
 - What do they already know?
 - What do they need to know?
 - What will they understand?
 - What will they be expecting?

Visualization Workflow

Define before you design!


1. Define your **purpose**
2. Define your **audience**
3. Define your **data** ●
4. Define your **message**
5. Define your **values**

- 
- What data will be included?
 - Which observations to include?
 - Which variables to include?
 - Which groupings to enforce?
 - Will it use raw or summary scores?

Visualization Workflow

Define before you design!

1. Define your **purpose**
2. Define your **audience**
3. Define your **data**
4. Define your **message**
5. Define your **values**

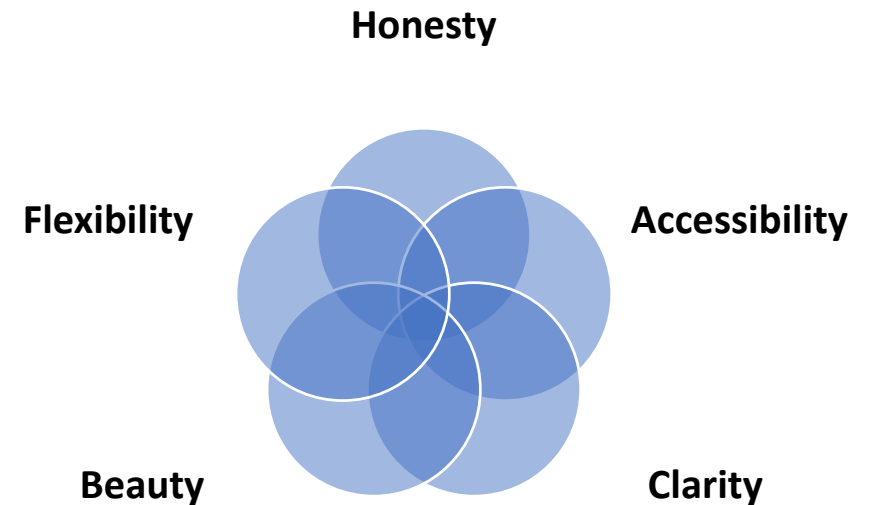
- 
- What is the main take-away?
 - What should viewers conclude?
 - How confident should they be?
 - What emotions should they feel?
 - What questions are they left with?

Visualization Workflow

Define before you design!

1. Define your **purpose**
2. Define your **audience**
3. Define your **data**
4. Define your **message**
5. Define your **values**

- What values are to be emphasized?
- What techniques can achieve them?



Visualization Workflow

- How many graphics are needed?
- Which mappings make sense?
- Which geoms and stats to use?
- Where are the eyes drawn?
- Is the message getting across?
- Was the purpose achieved?

← • **Design is a process!**

6. Consider the **grammar**
7. Create some **prototypes**
8. Solicit **feedback**
9. **Update** your designs
10. **Iterate** until satisfied

Visualization Resources

Books

- *R for Data Science*
- *ggplot2: Elegant Graphics for Data Analysis*
- *The Truthful Art: Data, Charts, and Maps...*
- *This Visual Display of Quantitative Information*

Websites

- ggplot2.org
- rstudio.com/resources/cheatsheets/
- stackoverflow.com
- color.adobe.com

ggplot2 Extensions

- *Stats and Geoms*: `ggrepel`, `ggforce`
- *Coordinate Systems*: `ggtern`, `circumplex`
- *Scales*: `scales`, `colorbrewer`, `viridis`
- More information at [ggplot2-exts.org](https://ggplot2-extensions.org)

Alternatives

- *Other R packages*: `ggvis`, `shiny`, `r2d3`
- *Other languages*: D3, `matplotlib`, `plotly`
- *Paid Software*: Tableau, Adobe Illustrator
- *Other Software*: Excel, SAS, STATA, SPSS, etc.