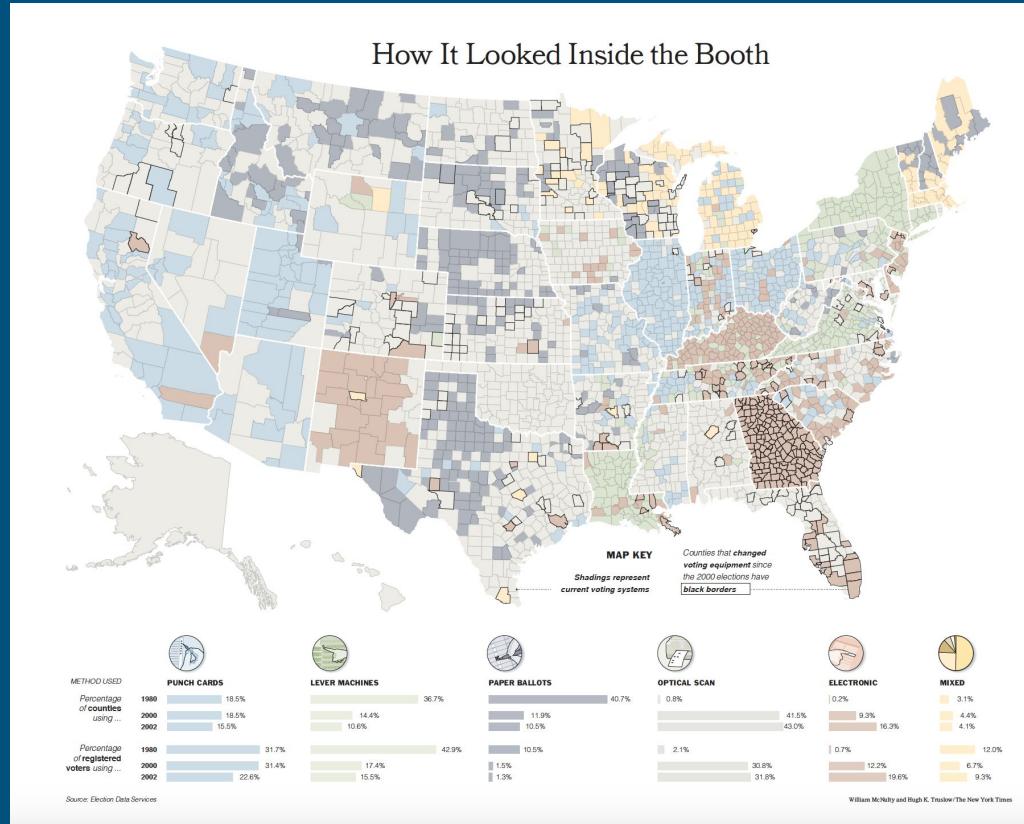


# Effective Data Visualization: Principles and Practices

---

Hugh Truslow  
Maps, Media, Data + Government Info.  
DataFest, Jan. 18, 2018 | CGIS S-030





William McNulty and Hugh K. Truslow. "How It Looked Inside the Booth." *New York Times*, Nov. 6, 2002. Page B9.

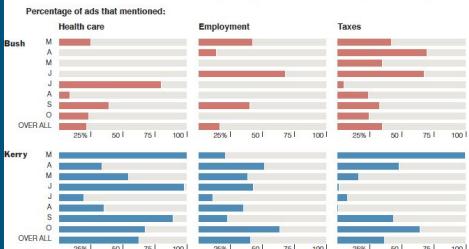
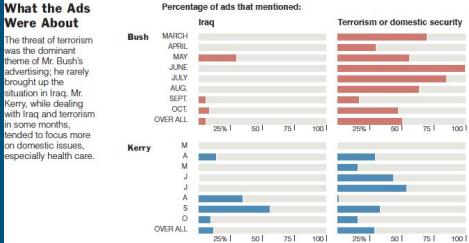
Matthew Ericson and  
Hugh K. Truslow. "The  
Great Ad Wars of 2004."  
*New York Times*, Nov. 1,  
2004. Page A19.

## The Great Ad Wars of 2004

Since March, the campaigns of President Bush and Senator John Kerry have bombarded living rooms across the country with over 630,000 commercials. A look at how the advertising war progressed.

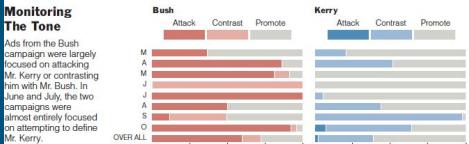
### What the Ads Were About

The threat of terrorism was the dominant theme of Mr. Bush's advertising; he rarely brought up the issue against Mr. Kerry, while dealing with Iraq and terrorism in some months, the two focus more on domestic issues, especially health care.



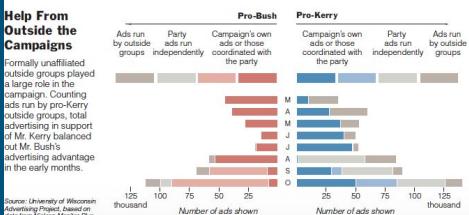
### Monitoring The Tone

Ads from the Bush campaign were largely focused on attacking Mr. Kerry, contrasting him with Mr. Bush. In June and July, the two campaigns were mainly focused on attempting to define Mr. Kerry.



### Help From Outside the Campaigns

Formerly unaffiliated groups played a large role in the campaign. Counting ads run by pro-Kerry outside groups, the total number of ads run by Mr. Kerry balanced out Mr. Bush's advertising advantage in the early months.



Sources: University of Minnesota Advertising Project, based on data from Nielsen Monitor-Plus

### The Shifting Advertising Map

The size of each circle represents the number of ads run by both campaigns in a given market.

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

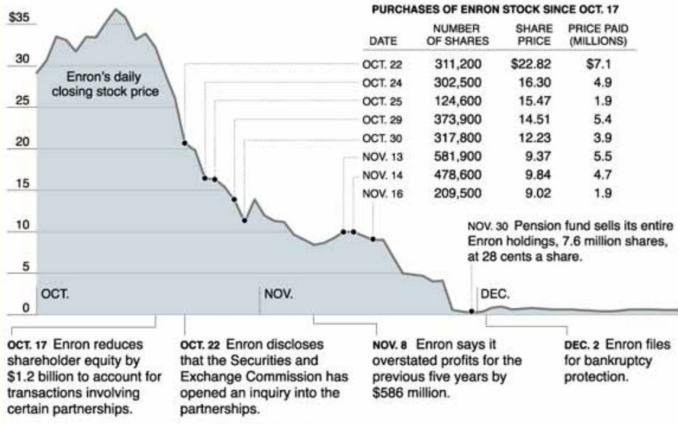
1,000, 2,500, 5,000

1,000, 2,500, 5,000

1,000, 2,500, 5,000

## Buying as the Ship Went Down

On the advice of Alliance Capital Management, one of its investment managers, the Florida state pension fund bought Enron stock even as the company's troubles became known. A former Alliance executive, Frank Savage, is also a member of Enron's board.



Sources: Dow Jones Interactive (stock price); Office of Senator Bill Nelson

### GRAPHIC OF THE DAY

The New York Times, February 6, 2002

This is a superb narrative of a stock price (graph), a narrative of the collapse of Enron (words annotating a time scale), and a narrative of lousy investments by the Florida state pension fund (table). The graph, table, and words are linked together very nicely. The caption at the top suggests a possible cause of the lousy investments. (The Florida state pension fund behaved like someone who left a jacket on the airplane; they were trying to get back on while everyone else was getting out.) Data sources are also indicated.

The major defect is that no designer is named. Someone did this good work and they should get credit for it. The Times gives the names of reporters and photographers: this graphic is a substantial piece of journalism, as valuable as a photograph or news story. Who did it?

[https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg\\_id=0003OT](https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0003OT)

# Definitions?

---

- > Data visualization
- > Information visualization
- > Scientific visualization
- > Information graphics, data graphics
- > Charts, graphs, plots, and figures
- > “Infographics”



“Amplify cognition”

“Facilitate  
understanding”

“Help people carry  
out tasks more  
effectively”

# There are some basics

---

- Label clearly

# There are some basics

---

- Label clearly
- Give a source for the data

# There are some basics

---

- Label clearly
- Give a source for the data
- Emphasize what's important

# There are some basics

---

- Label clearly
- Give a source for the data
- Emphasize what's important
- Watch out for distortions

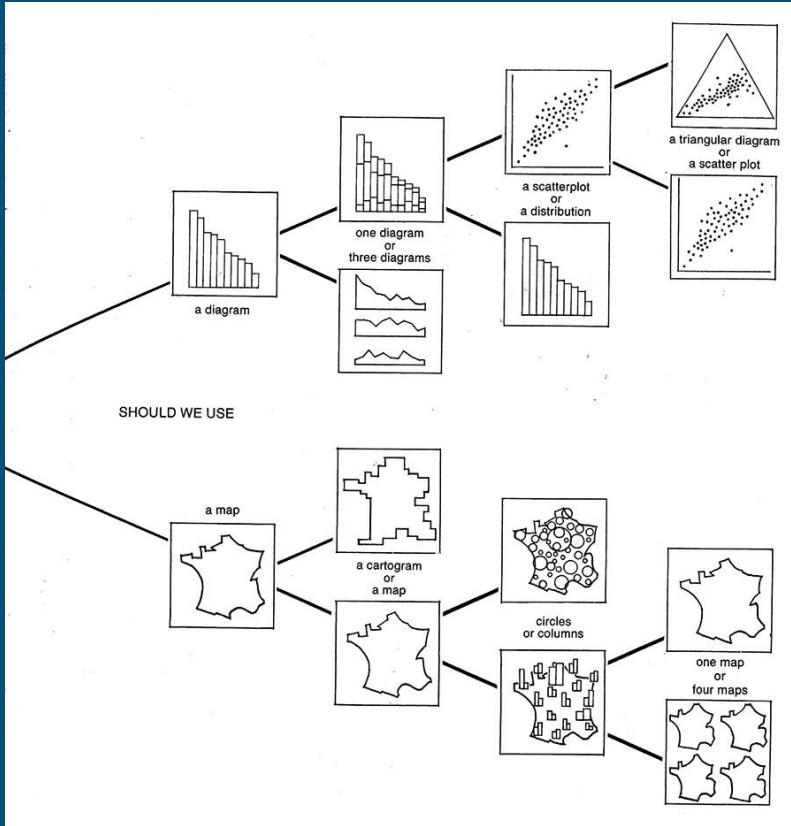
# There are some basics

---

- Label clearly
- Give a source for the data
- Emphasize what's important
- Watch out for distortions
- Eschew gimmickry

# Visual form

---



Bertin, Jacques. *Semiology of Graphics: Diagrams, Networks, Maps*. Trans. William J. Berg. Redlands, Calif.: ESRI Press, 2011. Page 101.

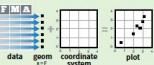
# Data Visualization with ggplot2

Cheat Sheet

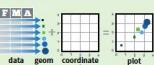


## Basics

ggplot2 is based on the grammar of graphics, the idea that you can build every graph from the same components: a **data set**, a **coordinate system**, and **geoms**—visual marks that represent data points.



To display values, map variables in the data to visual properties of the geom (**aesthetics**) like `size`, `color`, and `x` and `y` locations.



Complete the template below to build a graph.

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(
    mapping = aes(<MAPPINGS>),
    stat = <STAT>,
    position = <POSITION>
  ) +
  <COORDINATE_FUNCTION> +
  <FACET_FUNCTION> +
  <SCALE_FUNCTION> +
  <THEME_FUNCTION>
```

Required  
— Not required, sensible defaults supplied

```
ggplot(data = mpg, aes(x = cty, y = hwy))
  # Begins a plot that you finish by adding layers to.
  # Add one geom function per layer.
  # aesthetic mappings   data   geom
  # plot(x = cty, y = hwy, data = mpg, geom = <geom>)
  # Creates a complete plot with given data, geom, and
  # mappings. Supplies many useful defaults.
  last_plot()
  # Returns the last plot
  ggsave("plot.png", width = 5, height = 5)
  # Saves last plot as 5' x 5' file named "plot.png" in
  # working directory. Matches file type to file extension.
```

RStudio® is a trademark of RStudio, Inc. • CC BY RStudio • info@rstudio.com • 844-448-1212 • rstudio.com

**Geoms** - Use a geom function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

### Graphical Primitives

```
a <- ggplot(economics, aes(date, unemploy))
b <- ggplot(seals, aes(x = long, y = lat))
a + geom_blank() # (Useful for expanding limits)
```

```
b + geom_curve(aes(yend = lat + 1,
  xend = long + 1, curvature = 1), x, yend, alpha,
  angle, color, curvature, linetype, size)
```

```
a + geom_polygon(aes(group = group))
```

```
b + geom_rect(aes(ymin = long, ymax = lat,
  xmin = long + 1, ymin = lat + 1), xmax, xmin,
  ymax, ymin, alpha, color, fill, linetype, size)
```

```
a + geom_ribbon(aes(ymin = unemploy - 900,
  ymax = unemploy + 900))
```

```
b + geom_spoke(aes(angle = 1:155, xend = long + 1))
```

### Line Segments

common aesthetics: `x`, `y`, `alpha`, `color`, `linetype`, `size`

```
b + geom_abline(aes(intercept=0, slope=1))
```

```
b + geom_vline(aes(intercept = long))
```

```
b + geom_segment(aes(yend = lat + 1), xend = long + 1)
```

```
b + geom_spoke(aes(angle = 1:155, radius = 1))
```

### One Variable

#### Continuous

```
c <- ggplot(mpg, aes(hwy)); c2 <- ggplot(mpg)
```

```
c + geom_area(stat = "bin")
```

```
c + geom_density(kernel = "gaussian")
```

```
c + geom_dotplot()
```

```
c + geom_freqpoly()
```

```
c + geom_histogram(binwidth = 5)
```

```
c + geom_qq(aes(sample = hwy))
```

#### Discrete

```
d <- ggplot(mpg, aes(fct))
```

```
d + geom_bar()
```

```
d + geom_point()
```

— Not required, sensible defaults supplied

### Two Variables

#### Continuous X, Continuous Y

```
e + geom_label(aes(label = cty), nudge_x = 1,
  nudge_y = 1, check_overlap = TRUE)
```

```
e + geom_jitter(height = 2, width = 2)
```

```
e + geom_point()
```

```
e + geom_quantile()
```

```
e + geom_rug(sides = "bl")
```

```
e + geom_smooth(method = lm)
```

```
e + geom_text(aes(label = cty), nudge_x = 1,
  nudge_y = 1, check_overlap = TRUE)
```

— Not required, sensible defaults supplied

#### Discrete X, Continuous Y

```
f <- ggplot(mpg, aes(class, hwy))
```

```
f + geom_col()
```

```
f + geom_boxplot()
```

```
f + geom_hex()
```

```
f + geom_dotplot(binaxis = "y",
  stackdir = "center")
```

```
f + geom_pointrange()
```

```
f + geom_violin(scale = "area")
```

— Not required, sensible defaults supplied

#### Discrete X, Discrete Y

```
g <- ggplot(diamonds, aes(cut, color))
```

```
g + geom_count()
```

— Not required, sensible defaults supplied

#### Three Variables

```
seals$z <- with(seals, sqrt(delta_long^2 + delta_lat^2))
```

```
i <- ggplot(seals, aes(long, lat))
```

```
i + geom_contour(aes(z = z))
```

```
i + geom_raster(aes(fill = z), hijst = 0.5,
  vjst = 0.5, interpolate = FALSE)
```

```
i + geom_tile(aes(fill = z))
```

#### Continuous Bivariate Distribution

```
h <- ggplot(diamonds, aes(carat, price))
```

```
h + geom_bin2d(binwidth = c(0.25, 500))
```

```
h + geom_hex()
```

```
h + geom_density2d()
```

— Not required, sensible defaults supplied

#### Continuous Function

```
i <- ggplot(economics, aes(date, unemploy))
```

```
i + geom_area()
```

```
i + geom_line()
```

```
i + geom_step(direction = "hv")
```

— Not required, sensible defaults supplied

#### Visualizing error

```
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
```

```
j <- ggplot(df, aes(grp, fit, ymin = fit - se, ymax = fit + se))
```

```
j + geom_crossbar(fatten = 2)
```

```
j + geom_errorbar()
```

```
j + geom_linerange()
```

```
j + geom_pointrange()
```

— Not required, sensible defaults supplied

#### Maps

```
data <- data.frame(state = USArrests$Murder,
  state = tolower(rownames(USArrests)))
```

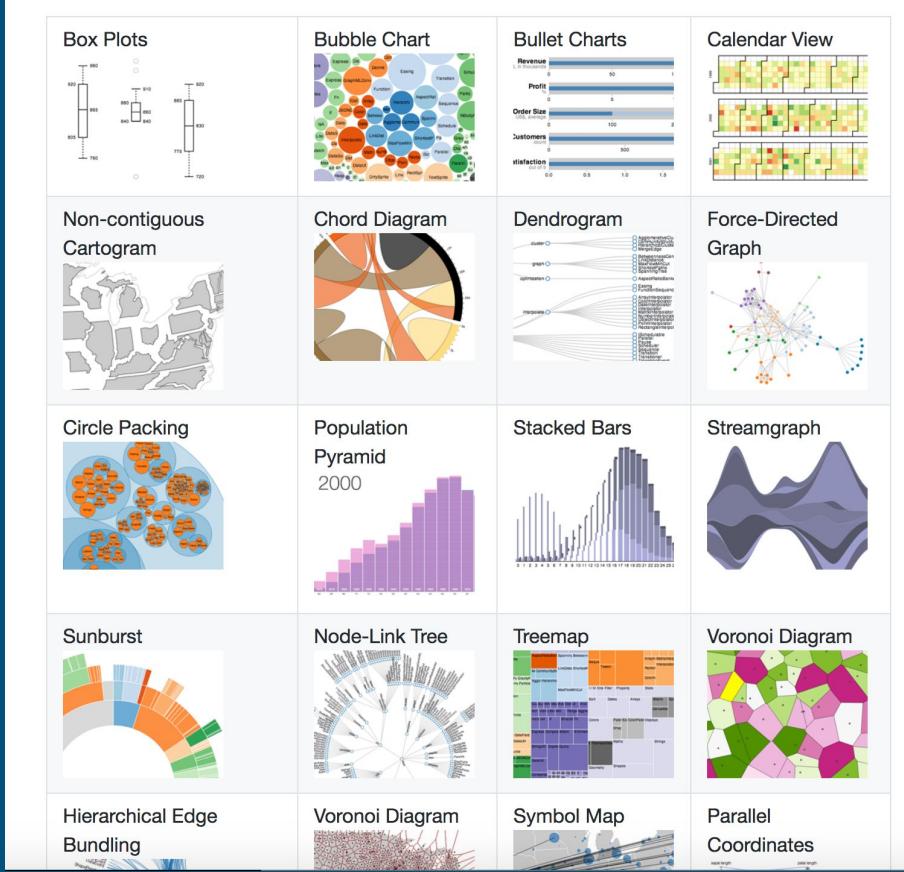
```
map <- map_data("state")
```

```
k <- ggplot(data, aes(fill = murder))
```

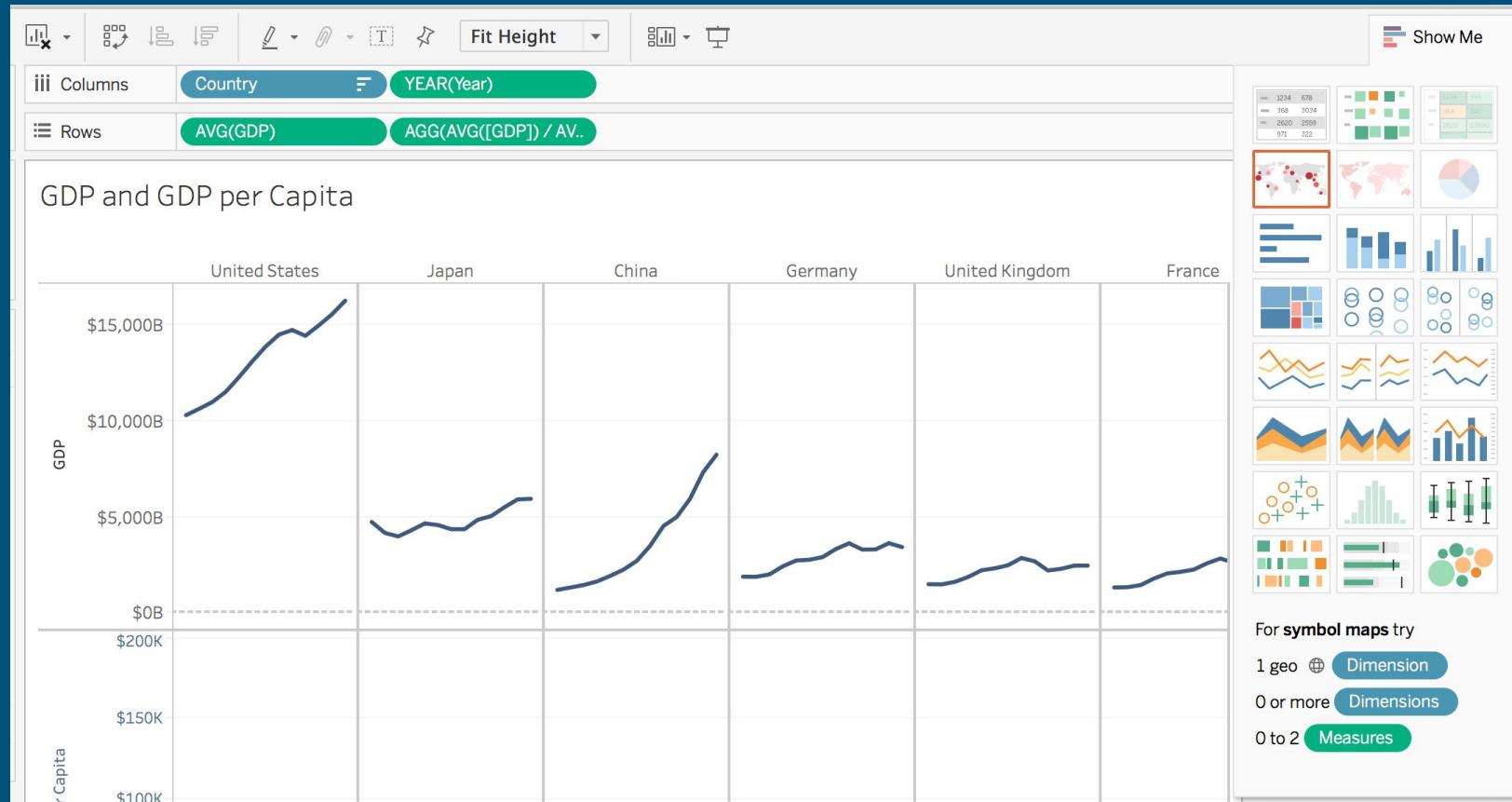
```
k + geom_map(aes(map_id = state), map = map) +
  expand_limits(x = map$long, y = map$lat)
```

```
map_id, alpha, color, fill, linetype, size
```

Learn more at [docs.ggplot2.org](http://docs.ggplot2.org) and [www.ggplot2-exts.org](http://www.ggplot2-exts.org) • ggplot2 2.1.0 • Updated: 11/16



[github.com/d3/d3/wiki/gallery](https://github.com/d3/d3/wiki/gallery)



## Tableau “Show Me”

# Tools? Don't get hung up

---

+ DATAVISUALIZATION.CH **SELECTED TOOLS**

All Maps Charts Data Color

Code?  Search...

**Arbor.js**  
A library of force-directed layout algorithms plus abstractions for graph organization and refresh handling.

**CartoDB**  
A web service for mapping, analyzing and building applications with data.

**Chroma.js**  
Interactive color space explorer that allows to preview a set of linear interpolated equidistant colors.

**Circos**  
A software package for visualizing data in a circular layout.

**Cola.js**  
A library for arranging networks using constraint-based optimization techniques.

**ColorBrewer**  
A web tool for selecting colors for maps.

**Cubism.js**  
A library for creating interactive time series and horizon graphs based on D3.js

**Cytoscape**  
An application for visualizing complex networks and integrating these with any type of attribute data.

**D3.js**  
An small, flexible and efficient library to create and manipulate interactive documents based on data.

**Dance.js**  
A simple data-driven visualization framework based on Data.js and Underscore.js

**Data.js**  
A data representation framework providing a uniform interface to domain data.

**Transform Script**  
Import Ex  
Split data repeatedly on newline into rows  
Split repeatedly on ;  
Promote row 0 to header  
Text Columns Rows Table Clean

**DataWrangler**  
An interactive web application for data cleaning and transformation.

**Degrada**  
A powerful declarative graphics framework for rich user interfaces, data visualizations and mapping.

**Envision.js**  
A library for creating fast, dynamic and interactive time series visualizations.

**Flare**  
A set of software tools for creating rich interactive data visualizations in ActionScript.

**GeoCommons**  
A public community and set of tools to access, visualize and analyze data with compelling map visualizations.

**Gephi**  
A visualization and exploration platform for networks with dynamic and hierarchical graphs.

**Google Chart Tools**  
A collection of simple to use, customizable and free to use interactive charts and data tools.

**Google Fusion Tables**  
A web application that makes it easy to host, manage, collaborate on, visualize, and publish data tables.

**I Want Hue**  
A web application to generate and refine palettes of optimally distinct colors.

**JavaScript InfoVis Toolkit**  
A JavaScript library that provides tools for creating interactive data visualizations for the web.

**Kartograph**  
A simple and lightweight framework for creating beautiful, interactive vector maps.

**Leaflet**  
A lightweight JavaScript library for making tile-based interactive maps for desktop and mobile browsers.  
<http://leaflet.cloudmade.com>

**Many Eyes**  
A web application to build, share and discuss graphic representation of user uploaded data.

**Output as** **JSON – Row Arrays**  

```
[{"Alan",12,"blue","Sep. 25, 2010"}, {"Shan",13,"green\blue","Sep. 25, 2010"}, {"John",45,"orange","Sep. 25, 2010"}]
```

**Update Output**  

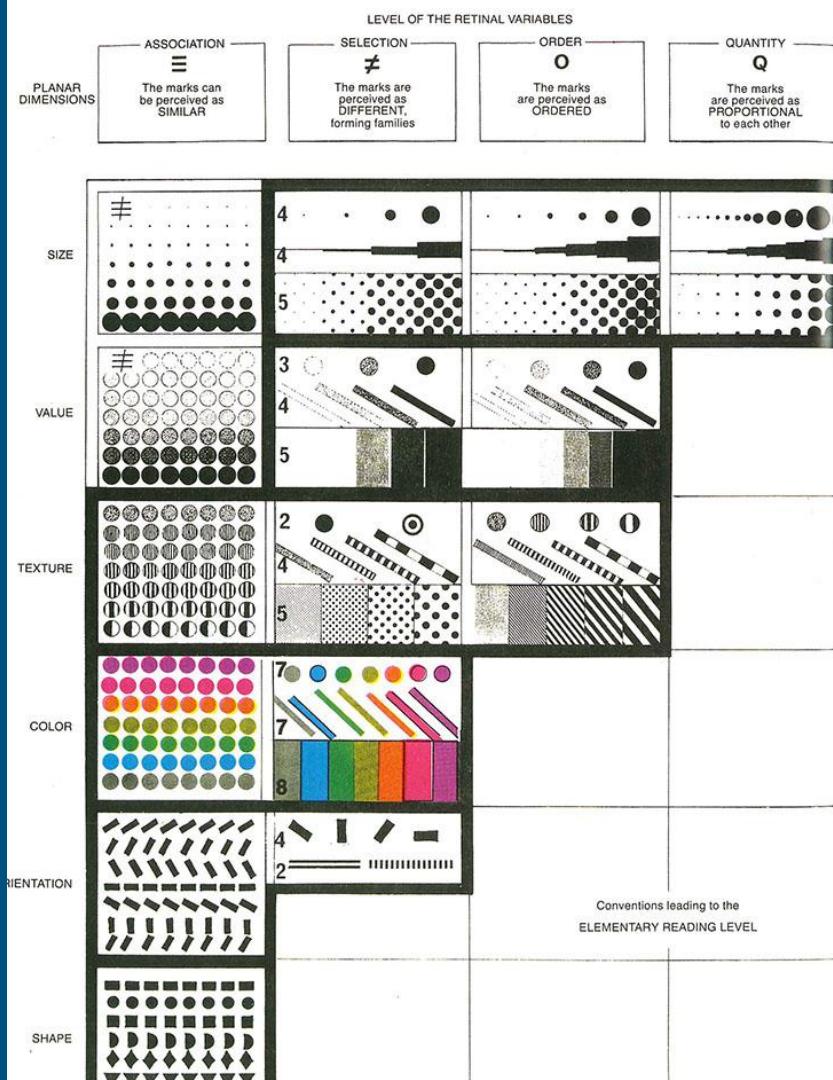
```
{"2010": [ {"fips": "6001", "totalpop": "1510271", "name": "New York City, NY"}, {"fips": "4601", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "1601", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "3601", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "5001", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "4801", "totalpop": "1499999", "name": "Philadelphia, PA"}, {"fips": "2401", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "4401", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "3201", "totalpop": "1499999", "name": "Dallas, TX"}, {"fips": "2601", "totalpop": "1499999", "name": "Austin, TX"}, {"fips": "4201", "totalpop": "1499999", "name": "San Jose, CA"}, {"fips": "1801", "totalpop": "1499999", "name": "Jacksonville, FL"}, {"fips": "2201", "totalpop": "1499999", "name": "Columbus, OH"}, {"fips": "3801", "totalpop": "1499999", "name": "Milwaukee, WI"}, {"fips": "4001", "totalpop": "1499999", "name": "St. Louis, MO"}, {"fips": "2801", "totalpop": "1499999", "name": "Portland, OR"}, {"fips": "3401", "totalpop": "1499999", "name": "Seattle, WA"}, {"fips": "1201", "totalpop": "1499999", "name": "Honolulu, HI"}, {"fips": "3001", "totalpop": "1499999", "name": "Baltimore, MD"}, {"fips": "1401", "totalpop": "1499999", "name": "Washington, DC"}, {"fips": "2001", "totalpop": "1499999", "name": "Tampa, FL"}, {"fips": "3801", "totalpop": "1499999", "name": "Orlando, FL"}, {"fips": "1001", "totalpop": "1499999", "name": "Nashville, TN"}, {"fips": "2801", "totalpop": "1499999", "name": "Memphis, TN"}, {"fips": "3201", "totalpop": "1499999", "name": "Baton Rouge, LA"}, {"fips": "3401", "totalpop": "1499999", "name": "New Orleans, LA"}, {"fips": "2201", "totalpop": "1499999", "name": "Austin, TX"}, {"fips": "3601", "totalpop": "1499999", "name": "Dallas, TX"}, {"fips": "4001", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "4401", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "4801", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "5001", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "5201", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "5401", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "5601", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "5801", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "6001", "totalpop": "1499999", "name": "Philadelphia, PA"}, {"fips": "6201", "totalpop": "1499999", "name": "Milwaukee, WI"}, {"fips": "6401", "totalpop": "1499999", "name": "Jacksonville, FL"}, {"fips": "6601", "totalpop": "1499999", "name": "St. Louis, MO"}, {"fips": "6801", "totalpop": "1499999", "name": "Portland, OR"}, {"fips": "7001", "totalpop": "1499999", "name": "Seattle, WA"}, {"fips": "7201", "totalpop": "1499999", "name": "San Jose, CA"}, {"fips": "7401", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "7601", "totalpop": "1499999", "name": "Austin, TX"}, {"fips": "7801", "totalpop": "1499999", "name": "Dallas, TX"}, {"fips": "8001", "totalpop": "1499999", "name": "Orlando, FL"}, {"fips": "8201", "totalpop": "1499999", "name": "Tampa, FL"}, {"fips": "8401", "totalpop": "1499999", "name": "Baltimore, MD"}, {"fips": "8601", "totalpop": "1499999", "name": "Washington, DC"}, {"fips": "8801", "totalpop": "1499999", "name": "Honolulu, HI"}, {"fips": "9001", "totalpop": "1499999", "name": "Milwaukee, WI"}, {"fips": "9201", "totalpop": "1499999", "name": "New Orleans, LA"}, {"fips": "9401", "totalpop": "1499999", "name": "Baton Rouge, LA"}, {"fips": "9601", "totalpop": "1499999", "name": "Austin, TX"}, {"fips": "9801", "totalpop": "1499999", "name": "Dallas, TX"}, {"fips": "0001", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "0201", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "0401", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "0601", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "0801", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "1001", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "1201", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "1401", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "1601", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "1801", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "2001", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "2201", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "2401", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "2601", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "2801", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "3001", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "3201", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "3401", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "3601", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "3801", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "4001", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "4201", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "4401", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "4601", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "4801", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "5001", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "5201", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "5401", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "5601", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "5801", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "6001", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "6201", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "6401", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "6601", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "6801", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "7001", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "7201", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "7401", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "7601", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "7801", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "8001", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "8201", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "8401", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "8601", "totalpop": "1499999", "name": "Phoenix, AZ"}, {"fips": "8801", "totalpop": "1499999", "name": "San Antonio, TX"}, {"fips": "9001", "totalpop": "1499999", "name": "San Diego, CA"}, {"fips": "9201", "totalpop": "1499999", "name": "Los Angeles, CA"}, {"fips": "9401", "totalpop": "1499999", "name": "Chicago, IL"}, {"fips": "9601", "totalpop": "1499999", "name": "Houston, TX"}, {"fips": "9801", "totalpop": "1499999", "name": "Phoenix, AZ"}]
```

**Output as** **JSON – Row Arrays**  

```
[{"Alan",12,"blue","Sep. 25, 2010"}, {"Shan",13,"green\blue","Sep. 25, 2010"}, {"John",45,"orange","Sep. 25, 2010"}]
```

# Marks, values, encodings

---



Bertin, *Semiology of Graphics*. Page 96.

MARK	EXAMPLE	DESCRIPTION
Point		The <i>point</i> mark has no variation ('constant') in the spatial dimension. It is largely a placeholder commonly used to represent a quantity through position on a scale, forming the basis of, for example, scatter plots.
Line		The <i>line</i> mark has one ('linear') spatial dimension. It is commonly used to represent quantitative value through variation in size, forming the basis of, for example, the bar chart.
Area		The <i>area</i> mark has two ('quadratic') spatial dimensions. It is commonly used to represent quantitative values through variation in size and position, forming the basis of, for example, bubble plots.
Form		The <i>form</i> mark has three ('cubic') spatial dimensions. It might be used to represent quantitative values through variation in size (specifically, through volume), forming the basis of, for example, a 3D proportional shape chart.

Figure 6.2 List of Mark Encodings

ATTRIBUTE	EXAMPLE	DESCRIPTION
QUANTITATIVE ATTRIBUTES		
Position		Position along a scale is used to indicate a quantitative value.
Size		Size (length, area, volume) is used to represent quantitative values based on proportional scales where the larger the size of the mark, the larger the quantity.
Angle/Slope		Variation in the size of angles forms the basis of pie chart sectors representing parts-of-a-whole quantitative values; the larger the angle, the larger the proportion. The slope of an incline formed by angle variation can also be used to encode values.
Quantity		The quantity of a repeated set of point marks can be used to represent a one-to-one or a one-to-many unit count.
Colour: Saturation		Colour saturation can be used (often in conjunction with other colour properties) to represent quantitative scales; typically, the greater the saturation, the higher the quantity.

ATTRIBUTE	EXAMPLE	DESCRIPTION
Colour: Lightness		Colour lightness can be used (often in conjunction with other colour properties) to represent quantitative scales; typically, the darker the colour, the higher the quantity.
Pattern		Variation in pattern density or difference in pattern texture can be used to represent quantitative scales or distinguish between categorical ordinal states.
Motion		Motion is more rarely seen but it could be used as a binary indicator to draw focus (motion vs no motion) or by incorporating movement through speed and direction to represent a quantitative scale ramp.
CATEGORICAL ATTRIBUTES		
Symbol/shape		Symbols or shapes are generally used with point markers to indicate categorical association.
Colour: Hue		Colour hue is typically used for distinguishing different categorical data values but can also be used in conjunction with other colour properties to represent certain quantitative scales.
RELATIONAL ATTRIBUTES		
Connection/Edge		A connection or edge indicates a relationship between two nodes. Sometimes arrows may be added to indicate direction of relationship, but largely it is just about the presence or absence of a connection.
Containment		Containment is a way of indicating a grouping relationship between categories that belong to a related hierarchical 'parent' category.

Figure 6.3 List of Attribute Encodings

Kirk, Andy. *Data Visualization*. Los Angeles, London: SAGE, 2016. Pages 153-54.

# Exploratory

---

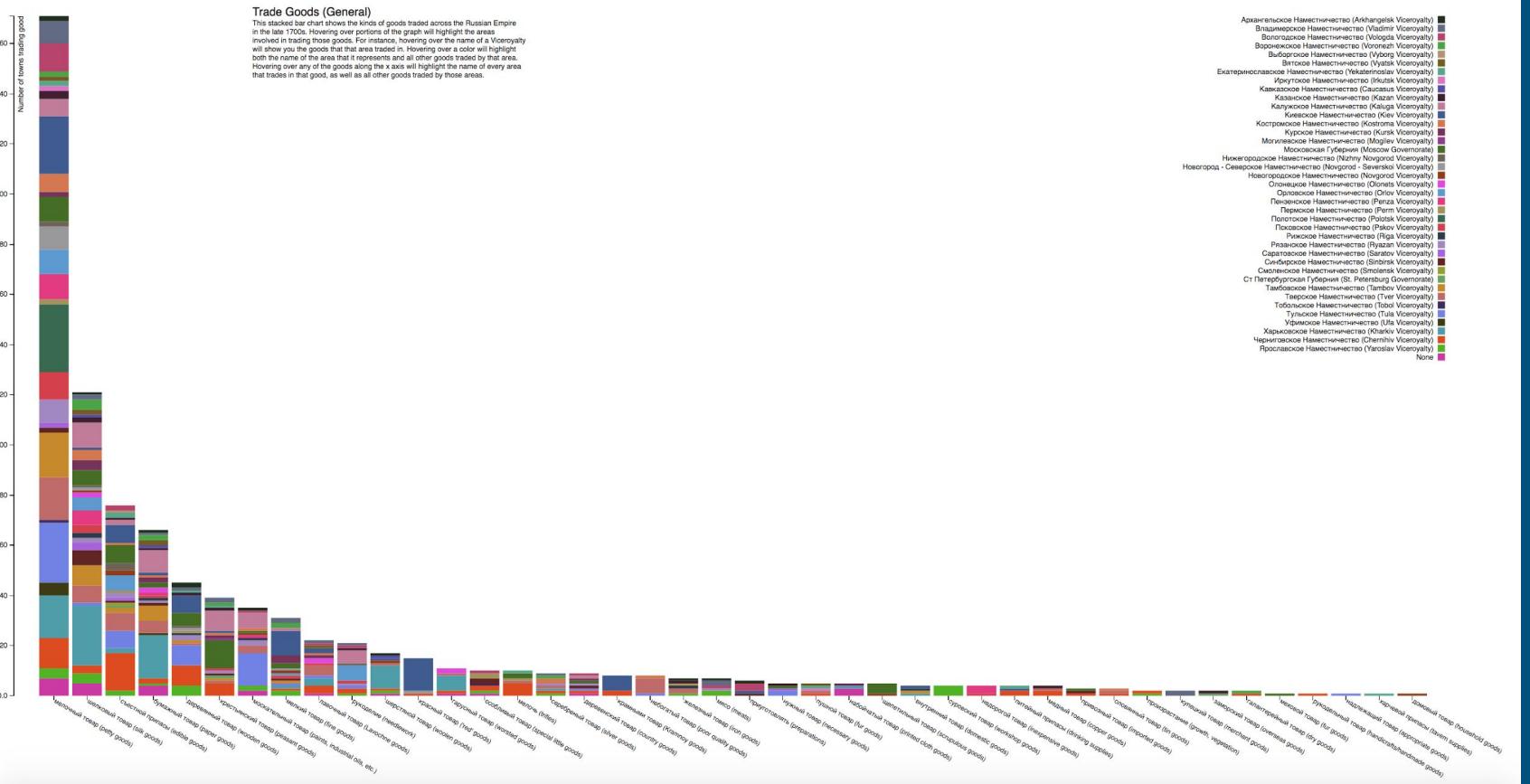
John W. Tukey

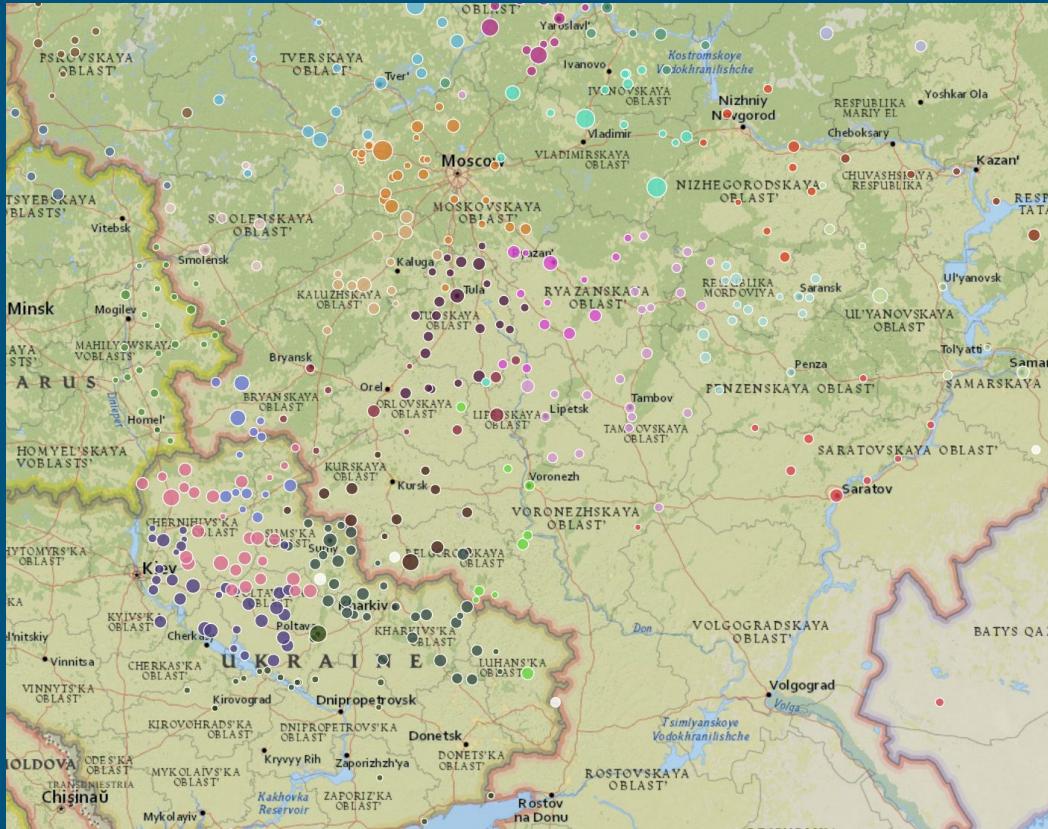
## EXPLORATORY DATA ANALYSIS



*"This book is about exploratory data analysis, about looking at data to see what it seems to say. It concentrates on simple arithmetic and easy-to-draw pictures. It regards whatever appearances we have recognized as partial descriptions, and tries to look beneath them for new insights. Its concern is with appearance, not with confirmation."*

Tukey, John W. *Exploratory Data Analysis*. Reading,  
Mass.: Addison-Wesley, 1977





<http://fungdavis.github.io/fungHGR/viz/leaflet/>

# wind map

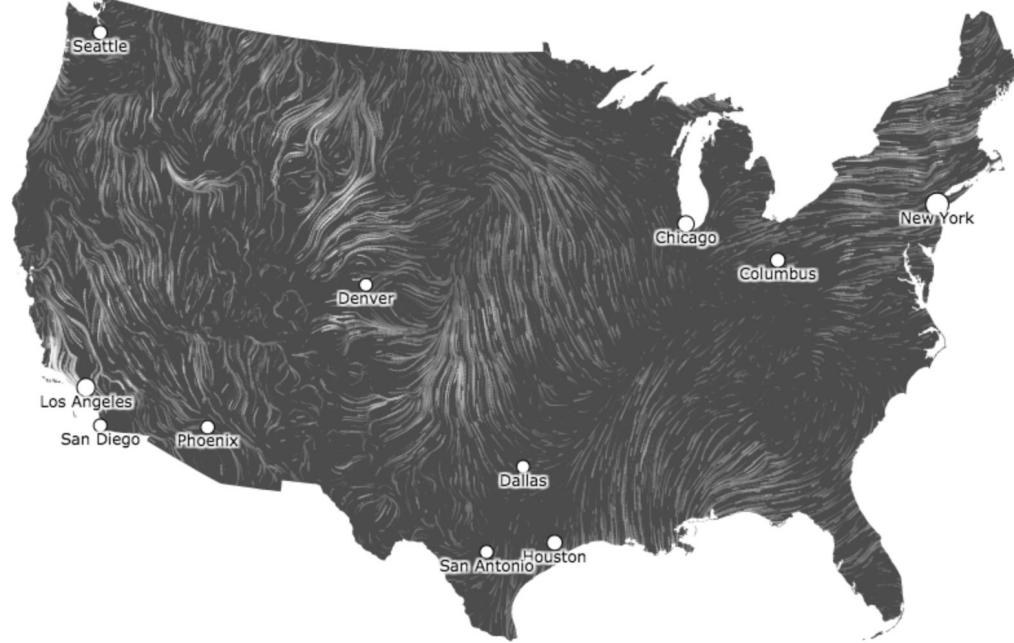
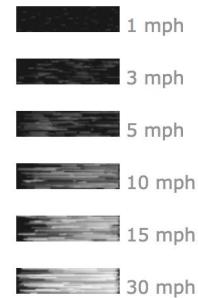
January 9, 2018

6:36 am EST

(time of forecast download)

top speed: **32.9 mph**

average: **8.2 mph**



[hint.fm/wind/](http://hint.fm/wind/)

# Visual hierarchy

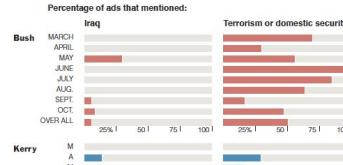
---

# The Great Ad Wars of 2004

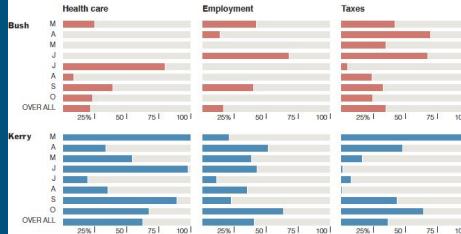
Since March, the campaigns of President Bush and Senator John Kerry have bombarded living rooms across the country with over 630,000 commercials. A look at how the advertising war progressed.

## What the Ads Were About

The threat of terrorism was the dominant theme of Mr. Bush's advertising; he rarely brought up the situation in Iraq. Mr. Kerry, while dealing with Iraq and terrorism in some months, tended to focus more on domestic issues, especially health care.

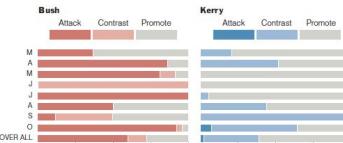


## Percentage of ads that mentioned:



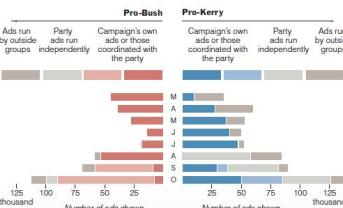
## Monitoring The Tone

Ads from the Bush campaign were largely focused on attacking Mr. Kerry or contrasting him with Mr. Bush. In June and July, the two campaigns were almost entirely focused on attempting to define Mr. Kerry.



## Help From Outside the Campaigns

Formerly unaffiliated outside groups played a large role in the campaign, running ads run by pro-Kerry outside groups, total advertising in support of Mr. Bush that helped put Mr. Bush's advertising advantage in the early months.



Source: University of Wisconsin Advertising Project, based on data from Nielsen Monitor Plus

## The Shifting Advertising Map

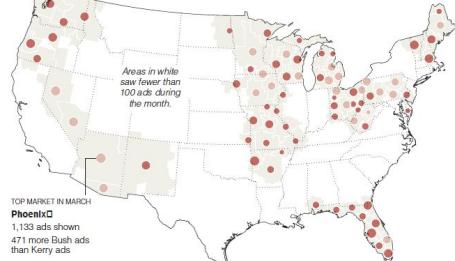
The size of each circle represents the number of ads run by both campaigns in a given market

The color of the circle represents the relative advantage in the number of ads shown

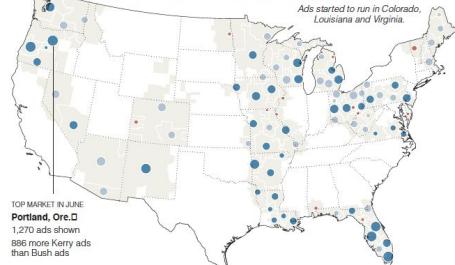
1,000, 2,500, 5,000

Large Bush, Slight Bush, the same, Slight Kerry, Large Kerry

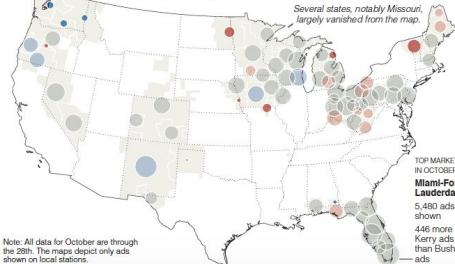
**March: Opening Salvoes** After Mr. Kerry became the presumed Democratic nominee, the two campaigns largely determined the 11 or so states they would focus on.



**June: The Playing Field Widens** While the volume of ads dropped somewhat, the number of states where the campaigns were running ads expanded to about 20.



**October: The Final Barrage** The map of targeted states contracted to the dozen or so most highly contested, where there was a huge increase in the volume of ads.



Matthew Ericson and Hugh E. Trudell/The New York Times

## The Shifting Advertising Map

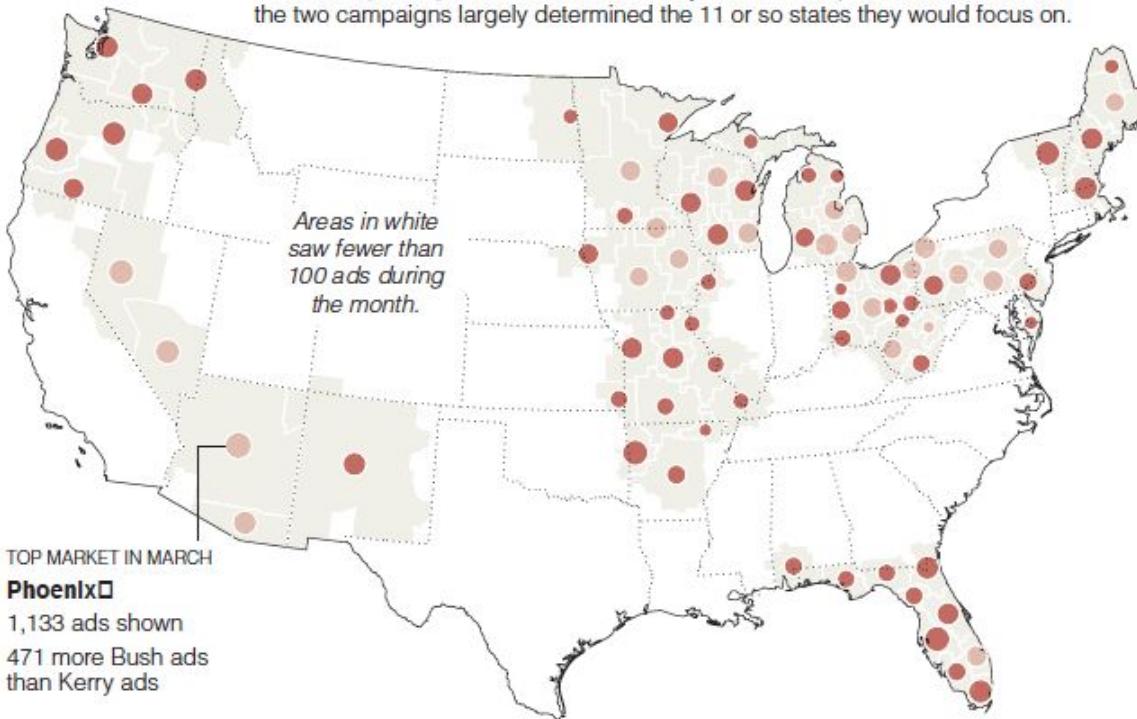
The size of each circle represents the number of ads run by both campaigns in a given market



The color of the circle represents the relative advantage in the number of ads shown



**March: Opening Salvo** After Mr. Kerry became the presumed Democratic nominee, the two campaigns largely determined the 11 or so states they would focus on.



# VISUAL ELOQUENCE

A HANDSON WORKSHOP FOR  
CREATING EFFECTIVE  
DATA VISUALIZATIONS

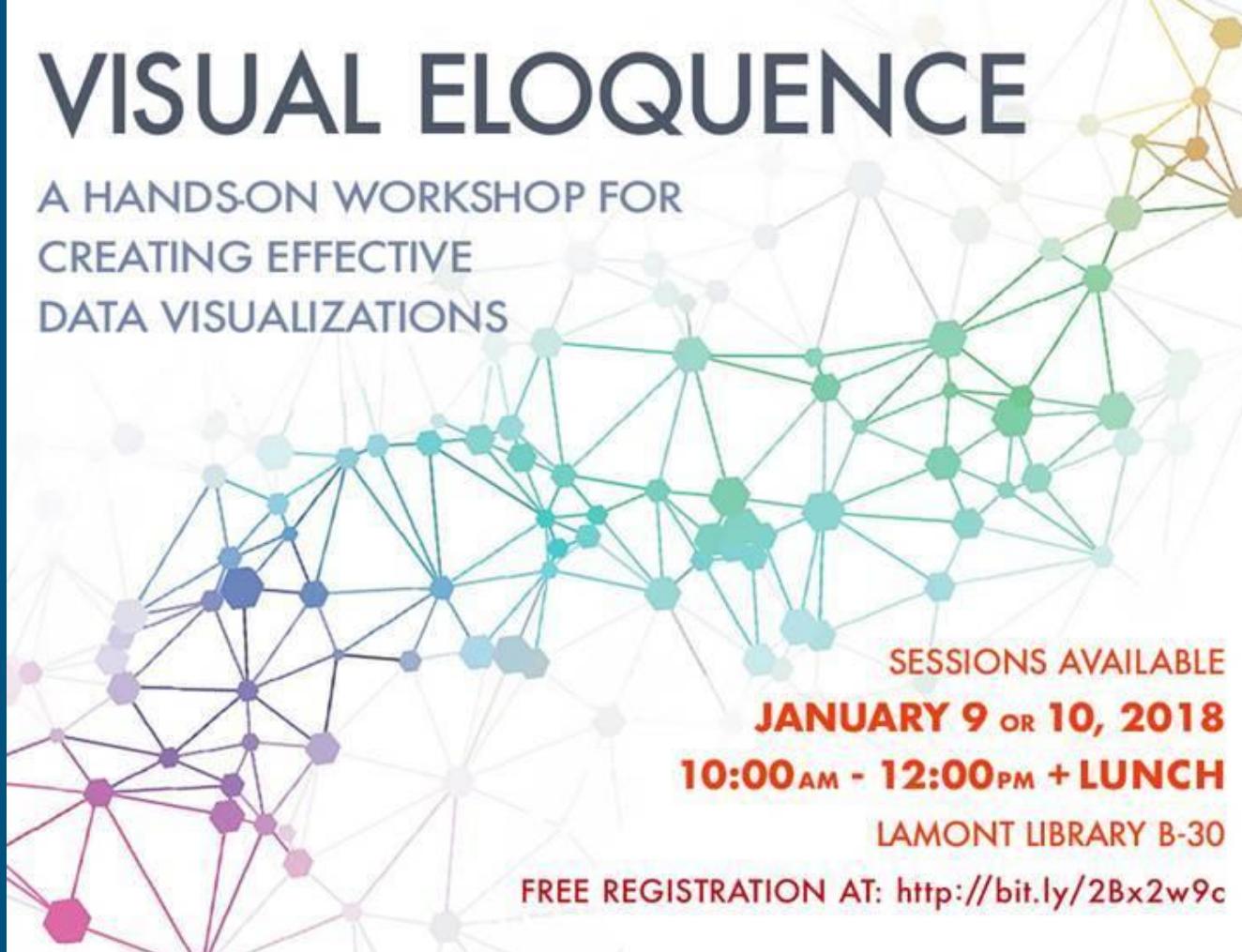
SESSIONS AVAILABLE

JANUARY 9 OR 10, 2018

10:00 AM - 12:00 PM + LUNCH

LAMONT LIBRARY B-30

FREE REGISTRATION AT: <http://bit.ly/2Bx2w9c>



# Thank you

---

[truslow@fas.harvard.edu](mailto:truslow@fas.harvard.edu)

