Designing with a Purpose

When you explore your own data, you don't need to do much in terms of storytelling. You are, after all, the storyteller. However, the moment you use your graphic to present information—whether it's to one person, several thousand, or millions—a standalone chart is no longer good enough.

Sure, you want others to interpret results and perhaps form their own stories, but it's hard for readers to know what questions to ask when they don't know anything about the data in front of them. It's your job and responsibility to set the stage. How you design your graphics affects how readers interpret the underlying data.
Prepare Yourself

You need to know your source material to tell good stories with data. This is an often overlooked part of designing data graphics. When you start, it’s easy to get excited about your end result. You want something amazing, beautiful, and interesting to look at, and this is great; but you can’t do any of that if you have no idea what you’re visualizing. You’ll just end up with something like Figure 9-1. How can you explain interesting points in a dataset when you don’t know the data?

Learn about the numbers and metrics. Figure out where they came from and how they were estimated, and see if they even make sense. This early data gathering process is what makes graphics in The New York Times so good. You see the end results in the paper and on the web, but you miss all the work that goes into the graphics before a single shape is drawn. A lot of the time, it takes longer to get all the data in order than it does to design a graphic.

So the next time you have a dataset in front of you, try not to jump right into design. That’s the lazy person’s way out, and it always shows in the end. Take the time to get to know your data and learn the context of the numbers.

Punch some numbers into R, read any accompanying documentation so that you know what each metric represents, and see if there’s anything that looks weird. If there is something that looks weird, and you can’t figure out why, you can always contact the source. People are usually happy to hear that someone is making use of the data they published and are eager to fix mistakes if there are any.

After you learn all you can about your data, you are ready to design your graphics. Think of it like this. Remember that part in The Karate Kid when Daniel is just starting to learn martial arts? Mister Miyagi tells him to wax a bunch of cars, sand a wooden floor, and refinish a fence, and then Daniel is frustrated because he feels like these are useless tasks. Then of course, it turns out that blocking and punching all of a sudden come natural to him because he’s been working on all the right motions. It’s the same thing with data. Learn all you can about the data, and the visual storytelling will come natural. If you haven’t seen the movie, just nod your head in agreement. And then go add The Karate Kid to your Netflix queue.
Big Graphic Blueprint

**THINGS**

<table>
<thead>
<tr>
<th>Cyan thing</th>
<th>Magenta thing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PERCENTAGE**

10%

**RANDOM STATISTIC**

44

**ANOTHER STATISTIC**

901

**ONE MORE STATISTIC**

101

**ONE MORE STATISTIC**

101

**GRAPHS THAT SHOW STUFF**

#1 #2 #3 #4 #5 #6 #7

#8 #9 #10 #11 #12 #13 #14

**REALLY BIG SHOCKING NUMBER**

1,121,455,492

**FIGURE 9-1** Big graphic blueprint. Go big or go home.
Prepare Your Readers

Your job as a data designer is to communicate what you know to your audience. They most likely didn’t look at the data, so they might not see the same thing that you see if there’s no explanation or setup. My rule of thumb is to assume that people are showing up to my graphics blindly, and with sharing via Facebook and Twitter and links from other blogs, that’s not all that far off.

For example, Figure 9-2 shows a screenshot of an animated map I made. If you haven’t seen this graphic before, you probably have no clue what you’re looking at. Given the examples in Chapter 8, “Visualizing Spatial Relationships,” your best guess might be openings for some store.

The map actually shows geotagged tweets that were posted around the world during the inauguration of President Barack Obama on Tuesday, January 20, 2009, at noon Eastern Standard Time. The animation starts early Monday morning, and as the day moves on, more people wake and tweet at a steady rate. The number of tweets per hour increases as the event nears, and Europe gets in on some of the action as the United States sleeps. Then Tuesday morning starts, and then boom—there’s huge excitement as the event actually happens. You can easily see this progression in Figure 9-3. Had I provided this context for Figure 9-2, it probably would’ve made a lot more sense.
INAUGURATION DAY ON TWITTER
A look at tweets around the world during the historic event.

FIGURE 9-3 Tweets during the inauguration of President Barack Obama

You don't have to write an essay to accompany every graphic, but a title and a little bit of explanation via a lead-in are always helpful. It's often good to include a link somewhere on your graphic so that people can still find your words even if the graphic is shared on another site. Otherwise, it can quickly become like a game of Telephone, and before you know it, the graphic you carefully designed is explained with the opposite meaning you intended. The web is weird like that.

As another example, the graphic in Figure 9-4 is a simple timeline that shows the top ten data breaches at the time.

It's basic with only ten data points, but when I posted it on FlowingData, I brought up how the breaches grow higher in frequency as you move from 2000 to 2008. The graphic ended up getting shared quite a bit, with a variant even ending up in Forbes magazine. Almost everyone brought up that last bit. I don't think people would've given the graphic much thought had I not provided that simple observation.
FIGURE 9-4 Major data breaches since 2000

The lesson: Don’t assume your readers know everything or that they can spot features in your graphic. This is especially true with the web because people are used to clicking to the next thing.

That’s not to say that people won’t spend time looking at data. As you might have seen, the OkCupid blog has been writing relatively long posts presenting results from thorough analyses of its online dating dataset. Titles include “The Best Questions for a First Date” and “The Mathematics of Beauty.”

Posts on the blog have been viewed millions of times, and people love what the OkCupid folks have to say. In addition to the tons of context in the actual post, people also come to the blog with a bit of context of their own. Because it is data and findings about dating and the opposite sex, people can easily relate with their own experiences. Figure 9-5, for example, is a graphic that shows what Asian guys typically like, which is from an OkCupid post on what people like, categorized by race and gender. Hey, I’m Asian and a guy. Instant connection.
On the other hand, when your graphic's topic is pollution levels or global debt, it can be a tough sell to a general audience if you don't do a good job of explaining.

Sometimes, no matter how much you explain, people simply don't like to read online, and they'll just skim. For example, I posted a map by FloatingSheep that compares number of bars to number of grocery stores in the United States, as shown in Figure 9-6. Red indicates areas where there are more bars than grocery stores, and orange indicates vice versa. The FloatingSheep guys called it the "beer belly of America."

Toward the end of the post, I wondered about the accuracy of the map and then finished up with, "Anyone who lives in the area care to confirm? I expect your comment to be filled with typos and make very little sense. And maybe smell like garbage." The lesson? Dry humor and sarcasm doesn't translate very well online, especially when people aren't used to reading your writing. I didn't actually expect comments to smell like garbage. Most
people got the joke, but there were also a good number of insulted Wisconsinites. Like I said, the web is an interesting place (in a good way).

**Visual Cues**

In Chapter 1, “Telling Stories with Data,” you saw how encodings work. Basically, you have data, and that data is encoded by geometry, color, or animation. Readers then decode those shapes, shades, and movement, mapping them back to numbers. This is the foundation of visualization. Encoding is a visual translation. Decoding helps you see data from a different angle and find patterns that you otherwise would not have seen if you looked only at the data in a table or a spreadsheet.

These encodings are usually straightforward because they are based on mathematical rules. Longer bars represent higher values, and smaller circles represent smaller values. Although your computer makes a lot of decisions during this process, it’s still up to you to pick encodings appropriate for the dataset at hand.

Through all the examples in previous chapters, you’ve seen how good design not only lends to aesthetics, but also makes graphics easier to read and can change how readers actually feel about the data or the story you tell. Graphics with default settings from R or Excel feel raw and mechanical. This isn’t necessarily a bad thing. Maybe that’s all you want to show for an academic report. Or if your graphic is just a supplement to a more important body of writing, it could be better to not detract from what you want people to focus on. Figure 9-7 shows a generic bar plot that is about as plain as plain can be.

If, however, you do want to display your graphic prominently, a quick color change can make all the difference. Figure 9-8 is just Figure 9-7 with different background and foreground colors.

A darker color scheme might be used for a somber topic, whereas a brighter color scheme can feel more happy-go-lucky (Figure 9-9).

Of course, you don’t always need a theme. You can use a neutral color palette if you like, as shown in Figure 9-10.
FIGURE 9-7 Plain bar plot

FIGURE 9-8 Default graph with dark color scheme
FIGURE 9-9 Default graph with light color scheme

FIGURE 9-10 Default graph with neutral color scheme
The main point is that color choice can play a major role in data graphics. It can evoke emotions (or not) and help provide context. It's your responsibility to choose colors that represent an accurate message. Your colors should match the story you are trying to tell. As shown in Figure 9-11, a simple color change can change the meaning of your data completely.

The graphic by designer David McCandless and design duo Always With Honor, explores the meaning of colors in different cultures. For example, black and white are often used to represent death; however, blue and green are more commonly used in Muslim and South American cultures, respectively.

Similarly, you can change geometry for a different look, feel, and meaning. For example, Figure 9-12 shows a randomly generated stacked bar chart with visualization researcher Mike Bostock's Data-Driven Documents. It has straight edges and distinct points, along with peaks and valleys.

**FIGURE 9-11** Colours In Culture by David McCandless and Always With Honor
Check out Lee Byron and Martin Wattenberg’s paper, “Stacked Graphs—Geometry and Aesthetics” for more information on streamgraphs. Several packages are also available, such as Protovis and D3, that enable you to design your own.

**FIGURE 9-12** Randomly generated stacked bar chart

If instead you used a streamgraph to show similar data, as shown in Figure 9-13, you clearly get a different feel. It’s more free-flowing and continuous, and instead of peaks and valleys, you have tightening and swelling. At the same time though, the geometry between the two chart types is similar. The streamgraph is basically a smoothed stacked bar chart with the horizontal axis in the center instead of on the bottom.

**FIGURE 9-13** Randomly generated streamgraph

Sometimes context can simply come from how you organize shapes and colors. Figure 9-14 shows a graphic that I made for fun to celebrate the holidays. The top part shows the ingredients that go into brining your turkey, and on the bottom is what goes into the turkey when you roast it in the oven.
CHRISTMAS TURKEY

Combine ingredients:
- 1 gallon vegetable stock
- 1 cup kosher salt
- 1/2 cup light brown sugar
- 1 tablespoon black peppercorns
- 1 1/2 teaspoons allspice berries
- 1 1/2 teaspoons chopped candied ginger

1 gallon heavily iced water

Soak turkey in brine 8 to 16 hours.

Put ingredients in turkey:
- 1 (14 to 16 pound) young turkey
- canola oil
- 1 red apple
- 1/2 onion
- 4 sprigs rosemary
- 6 leaves sage
- 1 cinnamon stick

Roast at 500 degrees F for 30 minutes.
Reduce to 350 degrees F and roast about 2 hours.

FIGURE 9-14 Recipe for Christmas turkey
The bottom line: At its most basic level, visualization is turning data, which can be numbers, text, categories, or any variety of things, into visual elements. Some visual cues work better than others, but applicability also varies by dataset. A method that’s completely wrong for one dataset could fit perfectly for another. With practice, you can quickly decide what fits your purpose best.

**Good Visualization**

Although people have been charting and graphing data for centuries, only in the past few decades have researchers been studying what works and what doesn’t. In that respect, visualization is a relatively new field. There still isn’t a consensus on what visualization actually is. Is visualization something that has been generated by a computer following a set of rules? If a person has a hand in the design process, does that make it not a visualization? Are information graphics visualization, or do they belong in their own category?

Look online, and you can find lots of threads discussing differences and similarities between information graphics and visualization or essays that try to define what visualization is. It always leads to a never-ending back and forth without resolution. These opposing opinions lead to varied criteria for what makes a data graphic good or bad.

Statisticians and analysts, for example, generally think of visualization as traditional statistical graphics that they can use in their analyses. If a graphic or interactive doesn’t help in analysis, then it’s not useful. It’s a failure. On the other hand, if you talk to graphic designers about the same graphic, they might think the work is a success because it displays the data of interest fairly and presents the data in an engaging way.

What you need to do is smush them all together, or at least get them in the same room together more often. The analytically minded can learn a lot from designers about making data more relatable and understandable, whereas design types can learn to dig deeper into data from their analytic counterparts.
I don't try to define what visualization is because the definition doesn't affect how I work. I consider the audience, the data in front of me, and ask myself whether the final graphic makes sense. Does it tell me what I want to know? If yes, then great. If no, I go back to the drawing board and figure out what would make the graphic better so that it answers the questions I have about the data. Ultimately, it's all about your goals for the graphic, what story you want to tell, and who you tell it to. Take all of the above into account—and you're golden.

Wrapping Up

A lot of data people see design as just a way to make your graphics look pretty. That's certainly part of it, but design is also about making your graphics readable, understandable, and usable. You can help people understand your data better than if they were to look at a default graph.

You can clear clutter, highlight important points in your data, or even evoke an emotional response. Data graphics can be entertaining, fun, and informative. Sometimes it'll just be the former, depending on your goal, but no matter what you try to design—visualization, information graphic, or data art—let the data guide your work.

When you have a big dataset, and you don't know where to begin, the best place to start is with a question. What do you want to know? Are you looking for seasonal patterns? Relationships between multiple variables? Outliers? Spatial relationships? Then look back to your data to see if you can answer your question. If you don't have the data you need, then look for more.

When you have your data, you can use the skills you learned from the examples in this book to tell an interesting story. Don't stop here, though. Think of the material you worked through as a foundation. At the core of all your favorite data graphics is a data type and a visualization method that you now know how to work with. You can build on these for more advanced and complex graphics. Add interactions, combine plots, or complement your graphics with photographs and words to add more context.
Remember: Data is simply a representation of real life. When you visualize data, you visualize what’s going on around you and in the world. You can see what’s going on at a micro-level with individuals or on a much larger scale spanning the universe. Learn data, and you can tell stories that most people don’t even know about yet but are eager to hear. There’s more data to play with than ever before, and people want to know what it all means. Now you can tell them. Have fun.
Symbols

$ (dollar sign), columns, 97
% (percent sign), Protovis tooltips, 154
? (question mark), R, 102, 279

A

Able2Extract, 27
ActionScript, 65–67, 167
colors, 316
maps, animation, 315
Modest Maps, 83–84
addChild(), 315
Adobe Flex Builder
interactive stacked area charts, 167–170
maps, animation, 311–313
Adobe Illustrator, 76, 77–78
Align window, 106
Area Graph Tool, 163
bar graphs, 102–107
Color window, 116–117, 127, 140, 236
Direct Selection Tool, 103, 104, 106, 139, 164
fonts, 104
Graph Tool, 78
heatmaps, 236
histogram matrix, 218–219
Layer window, 127
Line Graph Tool, 120–121
Mark Data Points, 122
MDS, 264
Object menu, 104, 105
Pen Tool, 140
pie charts, 137–141
Pie Graph Tool, 137
Rectangle Tool, 127
scatterplots, 113–117
Selection Tool, 103, 116
spreadsheets, 138–139
stacked area charts, 163–166
stacked bar charts, 111
step charts, 124, 127
tick marks, 122
time series charts, 120–123
Tools window, 103, 121, 137
Treemap Tool, 158
Type menu, 104
Type Tool, 107, 117
URL, 121
Window menu, 103
AggData, 24
aging, 8, 161–166
Align window, 106
Always With Honor, 337
Amazon Public Data Sets, 24
angles, pie charts, 17–18
animation, maps
ActionScript, 315
Adobe Flex Builder, 311–313
colors, 316–317
dots, 324
Flash, 315
latitude, 315, 320
listeners, 324
longitude, 315, 320
markers, 320–325
Modest Maps, 310, 313–314
Openings(), 314, 319
space, 309–325
Target, 309
time, 309–325
Walmart, 309, 325
zoom, 318
API. See Application Programming Interface
apipack, 239
Application Programming Interface [API], 24, 273, 274
ArcGIS, 83
Area Graph Tool, 163
arrays
plot(), 114
stacked bar charts, 152
art, 4-6
Asendorf, Kim, 5
axis labels
bar graphs, 98-99
design, 16-17
FlowingData, 16-17
plot(), 120
scale, 16
scatterplots, 116
step charts, 125
time series charts, 122
Type Tool, 117

B
bar graphs
Adobe Illustrator, 78, 102-107
axis labels, 98-99
borders, 98
colors, 98-99, 101, 335-336
creating, 95-102
histogram matrix, 226
lines, 106
Nathan's Hot Dog Eating Contest, 95-107
rectangles, 17
storytelling, 103
Tableau Software, 60
tick marks, 105-106
time axis, 94
time series, 94-107
titles, 107
value axis, 95
barplot(), 97-98, 101-102
col, 100
colors, 99
stacked bar charts, 110-111
Basketball Reference, 25
Beautiful Soup, 29-38
HTML, 49
region maps, 290-292
XML, 42
Berkeley Data Lab, 24
Bing, 82
birth rates
density plots, 209-213
histograms, 204-207
histogram matrix, 214-219
stem-and-leaf plots, 201-203
<body>, 145
borders
bar graphs, 98
heatmaps, 236
scatterplot matrix, 191
treemaps, 160
Bostock, Mike, 237, 337
bottom(), 147
boxplot(), 268
boxplots
bubble charts, 198
circles, 193
colors, 197
correlation, 192-200
creating, 194-200
crime, 194-200
labels, 198-199, 200
maps, 283-285
read.csv(), 195
rectangles, 198
relationships, 192-200
sizing, 18-19, 194
squares, 198, 199
stars, 198
symbols(), 195-198
text(), 198-199
thermometers, 198
value axis, 193
Brewer, Cynthia, 235, 286
bubbles, Polymaps, 85
bubble charts
boxplots, 198
circles, 193
colors, 197
correlation, 192-200
creating, 194-200
crime, 194-200
labels, 198-199, 200
maps, 283-285
read.csv(), 195
rectangles, 198
relationships, 192-200
sizing, 18-19, 194
squares, 198, 199
stars, 198
symbols(), 195-198
text(), 198-199
thermometers, 198
value axis, 193
business reports, design, 20
Byron, Lee, 338

c( ), 100, 256
Carter, Shan, 44
categories
interactive stacked area charts, 170
treemaps, 160
category axis
labels, 156
stacked bar charts, 148
causation, 11, 180-181
Census Bureau, 26, 170
crime, 183
chacter development, 7-8
Chernoff Faces, 238-243
Chestnut, Joey, 107
choropleths, 85
country maps, 301
region maps, 286, 294
unemployment, 294
circles
bubble charts, 193
maps, 277-279
Cleveland, William, 13, 127, 128
clipping masks, 103-104
close( ), 37
clusters, 263-264
cm.colors( ), 231
cmscale( ), 260
CMYK. See Cyan, Magenta, Yellow, and key
code
data formats, 46-51, 62-75
trade-offs, 74-75
JSON, 50
col, 98
barplot( ), 100
fill_colors, 100
scatterplots, 116
colors, 334-337
ActionScript, 316
bar graphs, 98-99, 101, 335-336
barplot( ), 99
bubble charts, 197
density plots, 210
differences, 228-229
Direct Selection Tool, 106
donut charts, 146
heatmap(), 233
heatmaps, 231-236
context, 234
scale, 233
histogram matrix, 225
interactive stacked area charts, 174
maps, 278-279
animation, 316-317
MDS, 261-263
multiple variables, 228-229
paral1el( ), 255
parallel coordinates plots, 253-256
pie charts, 139-140
Protoolis, 146
region maps, 286-302
scatterplot matrix, 191
stacked area charts, 164-165
stacked bar charts, 111, 152-153
step charts, 127
treemaps, 159
Color window, 116-117, 127, 140, 236
ColorBrewer, 153, 235
interactive stacked area charts, 175-176
Python, 292
region maps, 286-289
Colours In Culture, 337
columns
$ [dollarsign], 97
country maps, 298
heatmaps, 231
histograms, 203
latitude, 280
longitude, 280
parallel coordinates plots, 252
region maps, 288-289
scatterplots, 114
stacked bar charts, 109
star charts, 245
comma-delimited files, 40, 97
comma-separated values [CSV], 40, 44, 51, 96
country maps, 299
heatmaps, 231
histogram matrix, 215
data, 21-52
  breaches, 332
  checking, 12
  context, 22
  direct queries, 23
  finding sources, 22-26
  formats, 38-51
    code, 46-51, 62-75
    delimited text, 40
    JSON, 40
    spreadsheets, 45
    switching, 47-50
    Tool, 42-46
    XML, 41-42
  frames, 231
  general applications, 24
  geography, 25
  government, 26
  matrix, 231
  politics, 26
  provided by others, 22
  scatterplot matrix, 189
  scraping, 27-38
  search engines, 23
  sports, 25
  spreadsheets, 22
  topical, 24-26
  universities, 23-24
  world, 25-26
Data and Story Library (DASL), 23
Data-Driven Documents, 337
data.gov, 26
DataSF, 26
death probability, 3
decoding, 334
delimited files, 40, 48, 97, 210
density(), 209
density plots, 208-213
  birth rates, 209-213
  colors, 210
  creating, 209-213
  histograms, 212
LOESS, 208
plot(), 210
R, 209

D3, 338
DASL. See Data and Story Library

Cyan, Magenta, Yellow, and key (CMYK), 117

CSV. See comma-separated values
csv.reader(), 48, 291

CSS, 67-71
  country maps, 299
  donut charts, 145
  rows, 300-301
  stacked bar charts, 152, 157

Corel Draw, 79
corelation, 11
  bubble charts, 192-200
  causation, 180-181
  relationships, 180-200
  scatterplots, 181-192

Costco, 276-280, 281
country maps, 298-302
  choropleths, 301
  columns, 298
  Python dictionary, 299
  SVG, 299

crummy.com, 30

crime
  bubble charts, 194-200
  faces, 242-243
  Nightingale charts, 249
  outliers, 265
  scatterplots, 183-187
  star charts, 245-250

comparing
  histogram matrix, 214-219
  relationships, 213-226
compelling, 6-8
context, 19, 22, 107
  heatmap colors, 234
  maps, 330
continuous data
  stacked area charts, 162-166
  time series, 118-132

Cronby, 391

data, 21-52
  breaches, 332
  checking, 12
  context, 22
  direct queries, 23
  finding sources, 22-26
  formats, 38-51
    code, 46-51, 62-75
    delimited text, 40
    JSON, 40
    spreadsheets, 45
    switching, 47-50
  Tool, 42-46
  XML, 41-42
frames, 231
general applications, 24
geography, 25
government, 26
matrix, 231
politicalis, 26
provided by others, 22
scatterplot matrix, 189
scrapping, 27-38
search engines, 23
sports, 25
spreadsheets, 22
topical, 24-26
universities, 23-24
world, 25-26

Data and Story Library (DASL), 23
data-Driven Documents, 337
data.gov, 26
DataSF, 26
death probability, 3
decoding, 334
delimited files, 40, 48, 97, 210
density(), 209
density plots, 208-213
  birth rates, 209-213
  colors, 210
  creating, 209-213
  histograms, 212
LOESS, 208
plot(), 210
R, 209

D3, 338
data, 21-52

CSV. See comma-separated values
csv.reader(), 48, 291

CSS, 67-71
  country maps, 299
donut charts, 145
  rows, 300-301
  stacked bar charts, 152, 157

Cyan, Magenta, Yellow, and key (CMYK), 117
relationships, 208-213
tab-delimited files, 210
design, 13-20
  audience, 20
  axis labels, 16-17
  encodings, 13-16
  geometric shapes, 17-19
  labels, 13-16
  legends, 13-16
  sources, 19
Devlin, Susan, 127
diameter, bubble charts, 18
dictionary, Python, 291
country maps, 299
differences, 227-269
  colors, 228-229
  faces, 238-243
  heatmaps, 229-237
  maps, 305-308
  MDS, 258-264
  multiple variables, 228-258
  outliers, 265-269
  parallel coordinates plots, 251-258
  star charts, 244-250
Direct Selection Tool, 103, 104, 164
  colors, 106
  pie charts, 139
  stacked area charts, 164
  Tools window, 139
dist(), 260
distribution
  comparison, 213-226
  density plots, 208-213
  height, 5-6
  histograms, 203-208
  proportions, 136
  relationships, 200-213
<svg>, 142
width, 146
dots, 17
  maps, 276-280
  animation, 324
drawNavigation(), 319
dropout rates
  MDS, 260-264
  parallel coordinates plots, 257
E
education
  MDS, 260-263
  parallel coordinates plots, 252-258
emotions, 4, 5
encodings, 13-16, 334
Enos, Jeff, 158
entertainment, 5-6
Ericson, Matthew, 45
event(), 154
Excel, 39, 54-56
  conversion, 43-44
  CSV, 44
  Tableau Software, 60
time series charts, 121
Exploratory Data Analysis (Tukey), 201
Extensible Markup Language (XML), 41-42
Beautiful Soup, 42
CSV, 47-50
iterations, 49
Mr. Data Converter, 44
onLoadLocations(), 320
open(), 49
Python, 38, 42, 290
SVG, 289-290, 299
F
Facebook, 5
faces
  crime, 242-243
differences, 238-243
multiple variables, 238-243
NBA, 238-243
read.csv(), 239
faces(), 239, 240
Fakesville, 280–283
fertility rates, 283–285
File menu, 102
fill1, 290
fill1_colors, 100
fill1Style(), 154
filter(), 174
filtering, 185
findAll(), 32, 49, 292
Firefox, 13–14, 70
Fischer, Andreas Nicolas, 5
Flare, 66, 167
Flash, 65–67, 167
  Google Maps, 82
  maps, animation, 315
  Modest Maps, 83–84
Flash Builder. See Adobe Flex Builder
Flex Builder. See Adobe Flex Builder
Flex Navigator, 168
FlickrShapefiles, 25
FloatingSheep, 333
FlowingData, 9–10, 70, 77. See also your
  flowingdata
    axis labels, 16–17
    histogram matrix, 219, 225
    outliers, 267
    pie charts, 138
    PolyMaps, 85
    R, 76–77
    scatterplots, 114
    time maps, 305
    treemaps, 158, 161
Follow the Money, 26
fonts
  scatterplots, 116
  Selection Tool, 104, 116
  stacked bar charts, 111
  treemaps, 159
for, 51
Freebase, 24, 42

G

Gadget, 56
Gapminder Foundation, 6–7, 192
generate_css.py, 300
gecoder.us, 274
good coding, 273
GeoCommons, 25, 87
geography data, 25
geometric shapes, 17–19. See also specific shapes
Geopy, 273–274
GGobi, 252
Global Health Facts, 25
Good Maps Latitude Longitude Popup [Gorissen], 273
Google Chrome, 70
Google Finance, 57
Google Maps, 81, 82–83, 273, 274
Google Refine, 42–43
Google Spreadsheets, 56–57
Gore, AL, 6
Gorissen, Pierre, 273
government data, 26
grohs. See also specific graph types
  Excel, 55
  Python, 63
Graph Tool, 78
Graphical Perception and Graphical Methods for Analyzing Data [Cleveland and McGill], 13
gird lines, 118, 191
Gridworks, 42

H

happiest day of year, 5–6
Harr, Jonathan, 4–5
<head>, 145
header, 97
heatmap(), 229, 231, 233
heatmaps
  Adobe Illustrator, 236
  bars, 236
  colors, 231–236
  context, 234
  scale, 233
columns, 231
CSV, 231
data frames, 231
data matrix, 231
differences, 229-237
multiple variables, 229-237
NBA, 230-237
Protovis, 237
R, 229-237
YFD, 69-70
height
distribution, 5-6
donut charts, 146
histograms, 204
stacked bar charts, 152-153
hist(), 204-205, 221
histogram(), 212, 215-216
histograms, 203-208
birth rates, 204-207
columns, 203
density plots, 212
height, 204
maximum, 208
median, 208
minimum, 208
outliers, 267
rectangles, 203
relationships, 203-208
stem-and-leaf plots, 201, 202
value axis, 204
width, 204
histogram matrix
Adobe Illustrator, 218-219
bar graphs, 226
birth rates, 214-219
cells, 216
colors, 225
comparison, 214-226
CSV, 215
FlowingData, 219, 225
histogram(), 215-216
labels, 217
R, 214
relationships, 214-226
Rotten Tomatoes, 222-226
rows, 214-215, 217
television sizes, 219-222
horizontal axis
interactive stacked area charts, 170
scatterplots, 184
stacked bar charts, 152-153, 156-157
Horizontal Distribute Center, 106
HTML, 32-33, 67-71, 96, 154
Beautiful Soup, 49
donut charts, 144-145
interactive stacked bar charts, 151
Python, 38, 290
stacked bar charts, 157
<i>, 145
IBM Visual Communication Lab, 58
id, 159
if-else, 51
Illustrator. See Adobe Illustrator
IMDB. See Internet Movie Database
<img>, 32
Import window, 168-169
An Inconvenient Truth (Gore), 6
Infochimps, 24
Inkscape, 79
install.packages(), 276
interactive stacked area charts, 166-167
Adobe Flex Builder, 167-170
ColorBrewer, 175-176
colors, 174
horizontal axis, 170
JavaScript, 167
Protovis, 167
interactive stacked bar charts, 149-157, 167
HTML, 151
Obama, 151
Internet Explorer, 70, 85
Internet Movie Database (IMDB), 225
is.na(), 209
<i><i>, 41
iterations, 35, 49
data scraping, 37-38
loops, 51
Java applets, 60, 65
JavaScript, 67-71
donut charts, 145
Google Maps, 82
interactive stacked area charts, 167
Microsoft, 82
PolyMaps, 85
Processing, 65
Protovis, 151, 167
JavaScript Object Notation (JSON), 40
code, 50
CSV, 51
Mr. Data Converter, 44
JobVoyager, 170-173
journalism, 2-4
"Journalism in the Age of Data" [McGhee], 4
jQuery, 68-69
JSON. See JavaScript Object Notation

K
Kamvar, Sep, 4-5
Kane, David, 158
The Karate Kid (film), 328
keywords, 291

L
labels. See also axis labels
bubble charts, 198-199, 200
category axis, 156
design, 13-16
histogram matrix, 217
pie charts, 140-141
scatterplot matrix, 191
stacked area charts, 165
stacked bar charts, 154-156
star charts, 247
vertical axis, 156
labels, 240
latitude, 273-275
columns, 280
maps, animation, 315, 320
Layer window, 127
layers(), 154
left(), 147
legends, 13-16
length(), 100
lines
bar graphs, 106
maps, 280-283
parallel coordinates plots, 251
star charts, 244
line graphs. See also time series charts
Many Eyes, 58
proportions, 176-177
Line Graph Tool, 120-121
lineColor, 174
lineform, 79
lines, 212, 289
listeners, 324
lists, 100
locally weighted scatterplot smoothing
[LOESS], 127-132
curves, 128-132
density plots, 208
panel.smooth(), 190
plot(), 129
scatterplots, 185-186
scatterplot matrix, 192
unemployment, 128-132
LOESS. See locally weighted scatterplot
smoothing
longitude, 273-275
columns, 280
maps, animation, 315, 320
loops, 35, 51, 100

M
makeButton(), 318
Many Eyes, 58-60
maps, 87
scatterplot, 59
word tree, 58
map(), 279
MapEvent, 324
MapExtent(), 315
mapHeight, 315
map.market(), 159
maps, 80-88, 271-326. See also animation;
specific map types
API, 273
ArcGIS, 83
bubble charts, 283-285
circles, 277-279
colors, 278-279
context, 330
Costco locations, 276-280, 281
countries, 298-302
differences, 305-308
dots, 276-280
Fakesville, 280-283
fertility rates, 283-285
GeoCommons, 87
Google Maps, 82-83
latitude, 273-275
lines, 280-283
Tlines(), 289
longitude, 273-275
Many Eyes, 87
Microsoft, 82-83, 316
Modest Maps, 65, 83-84
multiple, 303-305
online-based solutions, 87
Package Installer, 276
points, 275-283
Polymaps, 84-85
R, 86-87
regions, 285-302
scaled points, 283-285
space, 302-325
sqrt(), 284
summary(), 284
symbols(), 277, 284
Tableau Software, 61
time, 302-325
title, 330
trade-offs, 87-88
Yahoo!, 82-83, 316
maps, 276
mapWidth, 315
Mark Data Points, 122
markers
maps, animation, 320-325
Modest Maps, 321
MarkersClip(), 320-322
maximum
histograms, 208
proportions, 136
McCandless, David, 337
McGhee, Geoff, 4
McGill, Robert, 13
mclust, 263-264
MDS. See multidimensional scaling
mean, 200, 207
median, 200, 207, 208
Microsoft, 82-83, 273, 316
minimum
histograms, 208
proportions, 136
Miscellaneous menu, 158
mode, 200
Modest Maps, 65, 83-84
maps, animation, 310, 313-314
markers, 321
MooTools, 68-69
mouseout events, 154
mouseover events, 147
Mr. Data Converter, 43-45
Mr. People, 45-46
multidimensional scaling (MDS)
Adobe Illustrator, 264
clusters, 263-264
cmdscale(), 260
colors, 261-263
differences, 258-264
dist(), 260
dropout rates, 260-264
education, 260-263
mclust, 263-264
plot(), 260
SAT scores, 260-262
text(), 261
multiple maps, 303-305
multiple variables
colors, 228-229
differences, 228-258
faces, 238-243
heatmaps, 229-237
parallel coordinates plots, 251-258
star charts, 244-250
MySQL, 64

N

names.org, 98
NameVoyager, 167
Nathan's Hot Dog Eating Contest
bar graphs, 95-107
R, 96-102
stacked bar charts, 108-111
Wikipedia, 96

NBA
faces, 238-243
heatmaps, 230-237
parallel coordinates plots, 251-252

Needlebase, 27
negative correlation, 181-182
New York Times, 2-4

Nightingale charts, 245-250
crime, 249
no correlation, 181-182
nobr, 33, 36
Numberary, 24
NYC DataMine, 26

O
Obama, Barack, 92, 330-331
interactive stacked bar charts, 151
pie charts, 150
stacked bar charts, 148-153

Object menu, 104, 105
<observation>, 49
OECD Statistics, 26
OkCupid, 5-6, 332
onLoadLocations, 320
onMapStartZooming, 324
onNextYear, 323
opacity, 139
open, 35, 48, 49
Openings, 314, 319
OpenSecrets, 26
OpenStreetMap, 25
OpenZoom Viewer, 305
order, 230
Oscar nominations, 14-15
outerRadius, 147

outliers
boxplots, 268
crime, 265
differences, 265-269
FlowingData, 267
histograms, 267
Weather Underground, 265
out-of-the-box, 54-62
Excel, 54-56
Google Spreadsheets, 56-57
Many Eyes, 58-60
Protovis, 67-68
Tableau Software, 60-61
trade-offs, 61-62
YFD, 61

P
Package Installer, 235, 276
pairs, 190
panel.smooth, 190
par, 221
parallel, 252, 255
parallel coordinates plots
c, 256
colors, 253-256
columns, 252
creating, 252-258
differences, 251-258
dropout rates, 257
education, 252-258
lines, 251
multiple variables, 251-258
NBA, 251-252
Protovis, 252
R, 252-255
read.csv, 252
reading_colors, 256
SAT scores, 252-256, 258
scale, 251
summary, 255
<path>, 290
patterns, 8-10
data scraping, 37-38
pch, 116

Pen Tool, 140
scatterplots, 117
tick marks, 117
Tools window, 105-106

PHP, 64, 219
pie charts
Adobe Illustrator, 78, 137-141
angles, 17-18
Color window, 140
colors, 139–140
creating, 137–141
Direct Selection Tool, 139
FlowingData, 138
labels, 140–141
Obama, 150
opacity, 139
Pen Tool, 140
proportions, 136–141
Tableau Software, 60–61
Pie Graph Tool, 137
play(), 323
Playfair, William, 136
playNextStore(), 323
plot(), 184
arrays, 114
axis labels, 120
density plots, 210
LOESS, 129
MDS, 260
R, 113, 129
scatterplots, 113
scatterplot matrix, 189
step charts, 125
time series charts, 119, 120
points
maps, 275–283
scale, 283–285
time, 93–118
points(), 116
polar area diagrams. See Nightingale charts
politics data, 26
polygon(), 210
Polymaps, 84–85
Portfolio
map.mark(), 159
R, 71–72, 159
treemaps, 158
positive correlation, 181–182, 184
postage rates, 124–127
posters, 20
poverty, 6
prettify(), 293
Processing, 64–65, 219
programming. See code
proportions, 135–178
donut charts, 141–148
line graphs, 176–177
pie charts, 136–141
stacked bar charts, 148–157
time series, 161–176
treemaps, 157–161
Protovis, 67
colors, 146
donut charts, 142–148
heatmaps, 237
histogram matrix, 219
interactive stacked area charts, 167
JavaScript, 151, 167
parallel coordinates plots, 252
Stack, 153
steamgraphs, 338
title(), 154
Python, 29–38, 63–64
ColorBrewer, 292
CSV, 51
dictionary, 291
country maps, 299
Geopy, 273–274
HTML, 38, 290
keywords, 291
Modest Maps, 83–84
region maps, 288–297
SVG, 293
XML, 38, 42, 290

Q
quartiles
boxplots, 268
definition, 268
region maps, 296
summary(), 207
unemployment, 296
questionable data, 12

R
R, 71–74
? (question mark), 102, 279
apilpack, 239
boxplots, 268
density plots, 209
File menu, 102
FlowData, 76–77
heatmaps, 229–237
histogram matrix, 214
id, 159
lists, 100
loops, 100
maps, 86–87
maps, 276
Miscellaneous menu, 158
Nathan’s Hot Dog Eating Contest, 96–102
parallel coordinates plots, 252–255
plot(), 113, 129
Portfolio, 71–72, 159
read.csv(), 183, 195, 215, 252
rectangles, 159
scatterplots, 71, 119
scatterplot matrix, 189, 191
stacked bar charts, 108–111
stars(), 245
statistics, 71
stem-and-leaf plot, 201
step charts, 124–132
storytelling, 74
summary(), 284
time series, 71
treemaps, 71–72, 77, 158–161
URL, 121, 183, 215
vectors, 100
radius, bubble charts, 18
Raven, 79
RColorBrewer, 235
read.csv(), 96, 97, 114
  bubble charts, 195
  faces, 239
  parallel coordinates plots, 252
  R, 183, 195, 215, 252
time series charts, 119
treemaps, 158–159
URL, 239
reading_colors, 256
Really Simple Syndication (RSS), 41
rectangles
  bar graphs, 17
  bubble charts, 198
  histograms, 203
  R, 159
treemaps, 19, 157–161
Rectangle Tool, 127
region maps, 285–302
  Beautiful Soup, 290–292
  choropleths, 286, 294
  ColorBrewer, 286–288
colors, 292–295
columns, 288–289
csv.reader(), 291
findall(), 292
<path>, 290
Python, 288–297
quartiles, 296
SVG, 288–297
unemployment, 288–297
relationships, 11–12, 179–226
  bubble charts, 192–200
  comparison, 213–226
correlation, 180–200
density plots, 208–213
distribution, 200–213
histograms, 203–208
histogram matrix, 214–226
scatterplots, 181–192
scatterplot matrix, 188–192
“Robust Locally Weighted and Smoothing scatterplots” (Cleveland), 128
Rosling, Hans, 6–7, 192
Rotten Tomatoes, 222–226
rows
  CSS, 300–301
  histogram matrix, 214–215, 217
  scatterplots, 114
time series charts, 120
treemaps, 160
  white space, 249
RSS. See Really Simple Syndication

S
SAS, 71
SAT scores
  MDS, 260–262
  parallel coordinates plots, 252–256, 258
SaturationEncoder(), 174
Scalable Vector Graphics (SVG)
  Beautiful Soup, 292
country maps, 299
CSV, 290-291

- donut charts, 142
- Polymaps, 85
- Python, 293
- region maps, 288-287
- XML, 289-290, 299

**scale.** See also multidimensional scaling
- axis labels, 16
- heatmap colors, 233
- maps, 283-285
- parallel coordinates plots, 251
- points, 283-285
- stacked bar charts, 152-153
- treemaps, 160

**scatterplots**
- Adobe Illustrator, 113-117
- axis labels, 116
- col, 116
- columns, 114
- correlation, 181-192
- creating, 113-118, 183-187
- crime, 183-187
- filtering, 185
- FlowingData, 114
- fonts, 116
- horizontal axis, 184
- LOESS, 185-186
- Many Eyes, 59
- pch, 116
- Pen Tool, 117
- plot(), 113
- R, 71, 119
- relationships, 181-192
- rows, 114
- scatter.smooth(), 185-186
- time axis, 112
- time series, 112-118
- time series charts, 119
- value axis, 112, 182

**scatterplot matrix**
- borders, 191
- colors, 191
- creating, 189-192
- data, 189
- grid lines, 191
- labels, 191
- LOESS, 192

**plot().** 189
- R, 189, 191
- relationships, 188-192
- scatter.smooth(), 130, 185-186
- Schneiderman, Ben, 157
- <script>, 145
- search engines, 23
- Selection Tool, 103, 104, 116
- sep, 97
- setMarkers(), 322-323
- setPoints(), 322
- setwd(), 97
- Silverlight, 82
- sizing
  - bubble diagrams, 18-19
  - two-dimensional shapes, 18
- skew, 201
- slide presentations, 20
- slippy maps, 80
- sorting, 230
- sources
  - context, 19, 107
  - design, 19
  - stacked bar charts, 111
- space maps, 302-325
- animation, 309-325
- spacing, 101-102
- <span>, 32, 33
- Sparkline plugin, 69
- SpatialKey, 87
- S-plus, 71
- spokes, 244
- sports data, 25
- spreadsheets. See also Excel
  - Adobe Illustrator, 78, 138-139
  - data, 22
  - formats, 45
  - Google, 56-57
  - Line Graph Tool, 121
- sqrt(), 284
- squares, 198, 199
- Stack, 153
- stacked area charts. See also interactive stacked area charts
  - Adobe Illustrator, 163-166
  - colors, 164-165
  - continuous data, 162-166
Direct Selection Tool, 164
interactive, 166–176
labels, 165
Protopis, 68
tick marks, 164
time axis, 162
typos, 164
value axis, 162
stacked bar charts. See also interactive
stacked bar charts
Adobe Illustrator, 111
arrays, 152
barplot(), 110–111
category axis, 148
colors, 111, 152–153
columns, 109
creating, 108–111
CSS, 152, 157
<div>, 152
fonts, 111
height, 152–153
horizontal axis, 152–153, 156–157
HTML, 157
labels, 154–156
Nathan’s Hot Dog Eating Contest, 108–111
Obama, 148–153
proportions, 148–157
R, 108–111
randomly generated, 338
scale, 152–153
sources, 111
tick marks, 156
time series, 108–111
Type Tool, 111
value axis, 148
vertical axis, 111, 153
width, 152–153
“Stacked Graphs—Geometry and Aesthetics”
(Byron and Wattenberg), 338
Stanford Visualization Group, 67
stars, bubble charts, 198
star charts
columns, 245
creating, 245–250
crime, 245–250
differences, 244–250
labels, 247
lines, 244
multiple variables, 244–250
spokes, 244
white space, 249
stars(), 245
startAnimation(), 323
statistics, 71
steam graphs, 68, 338
stem(), 202
stem-and-leaf plots, 201–203
step charts
Adobe Illustrator, 124, 127
axis labels, 125
colors, 127
creating, 124–127
plot(), 125
postage rates, 124–127
R, 124–132
smoothing and estimation, 127–132
time axis, 123
time series, 123–132
titles, 125
trends, 127
value axis, 123, 127
storytelling, 7
bar graphs, 103
R, 74
strings, 109
Stroke window, 106
strokeStyle(), 147
style.css, 300–301
SumedicinA (Asendorf), 5
summary(), 207, 255, 284
SVG. See Scalable Vector Graphics
symbols(), 195–198, 277, 284
T
\t, 48
tab-delimited files, 40, 48, 210
Tableau Software, 60–61
Takeru Kobayashi, 95, 107
Target, 309
television sizes, 219, 107
text(), 198–199, 261
text boxes, 107
textAlign(), 155
thermometers, 198

**tick marks**
- Adobe Illustrator, 122
- bar graphs, 105-106
- Horizontal Distribute Center, 106
- Pen Tool, 105-106, 117
- stacked area charts, 164
- stacked bar charts, 156
- time series charts, 122

**TIGER, 25**

**time axis**
- bar graphs, 94
- continuous data, 119
- scatterplots, 112
- stacked area charts, 162
- step charts, 123

**time maps, 302-325**
- animation, 309-325
- FlowingData, 305
- unemployment, 304-305

**time series, 9-10, 91-133**
- Adobe Illustrator, 78
- bar graphs, 94-107
- continuous data, 118-132
- discrete points of time, 93-118
- position, 17
- proportions, 161-176
- R, 71
- scatterplots, 112-118
- stacked bar charts, 108-111
- step charts, 123-132
- Tableau Software, 60
- trends, 92, 118

**time series charts**
- Adobe Illustrator, 120-123
- axis labels, 122
- creating, 119-123
- Excel, 121
- plot(), 119, 120
- read.csv(), 119
- rows, 120
- scatterplots, 119
- tick marks, 122

**timestamps, 36**

**title(), 147, 154**

**titles**
- bar graphs, 107
quartiles, 296
region maps, 288–297
time maps, 304–305
unique id, 291
universities, 23–24
URL
Adobe Illustrator, 121
R, 121, 183, 215
read.csv(), 239
url1lib2, 32, 35, 36

V
value axis
bar graphs, 95
bubble charts, 193
continuous data, 119
density plot, 208
histograms, 204
scatterplots, 112, 182
stacked area charts, 162
stacked bar charts, 148
step charts, 123, 127
variables. See multiple variables
vectors. See also Scalable Vector Graphics
R, 100
vertical axis
density plots, 212
labels, 156
stacked bar charts, 111, 153
vis.render(), 154

W
Walmart, 309, 325

Wattenberg, Martin, 167, 338
We Feel Fine [Harris and Kamvar], 4–5
Weather Underground, 27–38, 265
white space, 241, 249
Wickham, Hadley, 215
width
donut charts, 146
histograms, 204
stacked bar charts, 152–153
Wikipedia, 24, 96
Window menu, 103
wolfalpha.com, 23
word tree, 58
World Bank, 26, 298, 299
world data, 25–26
World Health Organization, 26
World Progress Report, 9
write(), 36
write.table(), 210

X
XML. See Extensible Markup Language

Y
Yahoo!, 82–83, 316
YFD. See your.flowingdata
your.flowingdata [YFD], 61, 69–70
y.ticks(), 156
yyyymmdd, 36

Z
zoom, 318