### Data Visualization

#### with ggplot2

**Cheatsheet**

**Basics**

- *ggplot2* is based on the grammar of graphics, the idea that you can build every graph from the same few components: a data set, a set of geoms—visual marks that represent data points, and a coordinate system.
- To display data values, map variables in the data set to aesthetic properties of the geom like color, size, and x and y locations.

**Graphical Primitives**

- Add a new layer to a plot with a geom_*() or stat_*() function. Each provides a geom, a set of aesthetic mappings, and a default stat and position adjustment.
- *aesthetics mappings* + *data* + *geom* + *add layers, elements with + layer = geom + default stat + layer specific mappings

**Examples**

```
# ggplot(data = mpg, aes(x = cty, y = hwy))
# Begins a plot that you finish by adding layers to. No defaults, but provides more control than qplot().

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

```
# ggplot(mpg, aes(hwy, cty)) +
# geom_point(aes(color = cyl)) +
map ~ map_data("state")
```

```
# ggplot(mpg, aes(class, hwy))
# ggplot(Seals, aes(x = long, y = lat))
```

**Build a graph with ggplot() or qplot()**

#### One Variable

- **Continuous**
  - `g <- ggplot(mpg, aes(hwy))`
  - `a + geom_area(stat = "bin")`
  - `x, y, alpha, color, fill, linetype, size`
  - `b + geom_density(alpha = .5)`
  - `x, y, alpha, color, fill, linetype, size, weight`

- **Discrete**
  - `b <- ggplot(mpg, aes(fli))`
  - `x, alpha, color, fill, linetype, size, weight`

**Two Variables**

- **Continuous X, Continuous Y**
  - `f <- ggplot(mpg, aes(cty, hwy))`
  - `f + geom_blank()`
  - `x, y, alpha, color, fill, shape, size`

- **Discrete X, Continuous Y**
  - `g <- ggplot(mpg, aes(class, hwy))`
  - `g + geom_bar(stat = "identity")`
  - `x, y, alpha, color, fill, linetype, size`

- **Continuous Bivariate Distribution**
  - `i <- ggplot(movies, aes(year, rating))`
  - `i + geom_bin2d(binwidth = c(5, 0.5))`
  - `xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size, weight`

**Three Variables**

- `m + geom_raster(aes(fill = z), hjust=0.5, vjust=0.5, interpolate=FALSE)`
  - `x, y, alpha, fill (fast)`
- `m + geom_tile(aes(fill = z))`
  - `x, y, alpha, color, fill, linetype, size (slow)`

**Visualizing error**

```
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
k <- ggplot(df, aes(grp, fit, ymin = fit-se, ymax = fit+se))
k + geom_errorbar() + geom_errorbarh()
k + geom_crossbar(fatten = 2)
k + geom_pointrange()
k + geom_errorbar()
k + geom_hline()
k + geom_vline()
k + geom_percentage()
k + geom_boxplot()
```

**Maps**

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

```
m + ggplot(data, aes(fill = murder))
```

```
l + geom_map(data = map, map_id = state)
```

```
l + geom_polygon(aes(map_id = state), map = map)
```

```
l + geom_hex(sides = "bl")
```

```
l + geom_rug trước = "hv"
```

```
l + geom_text(aes(label = cty))
```

**Learn more at docs.ggplot2.org • ggplot2 1.0.0 • Updated: 4/15**
Stats - An alternative way to build a layer

Some plots visualize a transformation of the original data set. Use a stat to choose a common transformation to visualize, e.g.

\[ g + \text{geom_bar()} \]

Each stat creates additional variables to map aesthetics to. These variables use a common name, syntax, stat functions and geom functions both combine a stat with a geom to make a layer, i.e. stat(geom="bar") does the same as geom_bar(stat="bin")

General Purpose scales

Use with any aesthetic: alpha, color, fill, linetype, shape, size

scale_*_continuous() - map cont values to visual values
scale_*_discrete() - map discrete values to visual values
scale_*_identity() - identity function

X and Y location scales

Use with x or y aesthetics (x shown here)

scale_x_*() - use x aesthetics (x shown here)

Color and fill scales

Discrete

- Stack elements on top of one another, normal height
- Arrange elements side by side

Continuous

- Map values linearly
- Include midpoints
- Value mapped to visual values

Shape scales

Manual shape values

Size scales

- Value mapped to visual values

Coordinate Systems

\[ r + \text{coord_cartesian(xlim=c(0, 5), ylim=c(0, 10))} \]

Faceting

- Grid facets based on one or more discrete variables
- Face into boxes based on one or more continuous variables
- Facet into rows and columns
- Wrap facets into a rectangular layout

Labels

- Change the label on the X axis
- Change the label on the Y axis
- Change the X axis limits
- Change the Y axis limits

Legends

- Set legend title and labels with a scale function
- Set legend type for each aesthetic: colorbar, legend, none (no legend)
- Set legend for each aesthetic: colorbar, legend, none (no legend)

Themes

- Change the theme of the plot
- Change the background
- Change the font size

Zooming

- Without clipping (preferred)
- With clipping (removes unseen data points)

Use scale functions to update legend labels

Learn more at docs.ggplot2.org • ggplot2 1.0.0 • Updated: 4/15