Topics in Data Visualization

Scales

Jul 16 2014

Charlotte Wickham

vis.cwick.co.nz

Assignments

#1 graded

#2 up, due next Wednesday, critique either:

one of your plots from Assn #1

or any plot you find and can replicate

#3 Preview: Using any data you want, produce a graphic (1-3 plots):

draft by Monday week 6 for group critique

final ready by Friday week 6 for class presentation

Your turn

```
iah <- read.csv("http://vis.cwick.co.nz/data/iah-summary.csv")
ggplot(iah, aes(DepHour, DayOfWeek)) +
  geom_tile(aes(fill = prop_over_15))
```

A subset of the data in the hflights package

Just George Bush Intercontinental Airport (IAH)

2011 flights, summary of departure delays, by day of week and hour of departure

see: vis.cwick.co.nz/lectures/10-get-iah-summary.R for code

```
scale_xxxx_yyyyyyy
scale_fill_gradient
```

aesthetic scale type

Every aesthetic has a scale. Scales control how:

the data is mapped to the aesthetic property

`trans`formations

`limits`

mapping (scale specific)

the mapping is communicated in the plot

`name`

`breaks, minor_breaks`

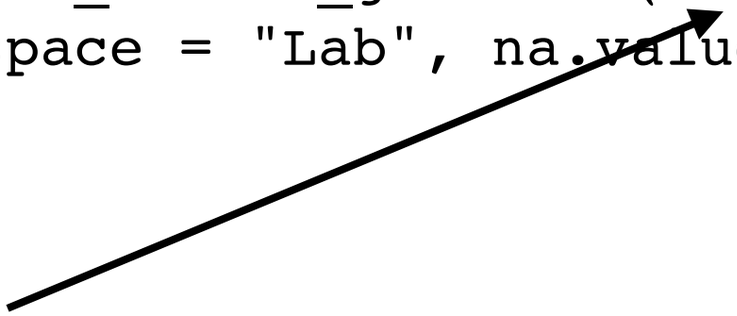
`labels`

`guide`

`args common to all scale functions`

scale_fill_gradient

```
scale_colour_gradient(..., low = "#132B43", high = "#56B1F7",
  space = "Lab", na.value = "grey50", guide = "colourbar")
```



Any of the arguments:

name,
limits,
trans,
labels,
breaks,
guide

Your turn

1. Explicity add: `scale_fill_gradient()`
2. Experiment using different named colors (`colors()`) for high and low.

http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf

3. Try a Brewer based palette instead

```
library(RColorBrewer)
display.brewer.all() # to find the name
+ scale_fill_distiller(palette = "YlOrRd")
```

Other color scales

continuous

`gradient`	Sequential (2) color gradient
`gradient2`	Diverging (3) color gradient
`gradientn`	n color gradient
`distiller`	Gradient through brewer palette
`grey`	Grey gradient

discrete

`brewer`	Brewer palette
`hue`	Hues with constant chroma and luminance
`manual`	Specify hex/named colors

Your turn

Read the help on one of the other continuous color scales and try it out.

trans and limits

limits - useful for zooming in, or making scales comparable across plots.

trans - equivalent to transforming the variable, except the untransformed values can be read off the scale.

Beware! These are applied before statistics. (Binning, averaging, box plots etc. will use the limited and transformed data).

To occur after statistics use:

```
coord_cartesian(xlim = , ylim = )
coord_trans(xtrans = , ytrans = )
```

Your turn

```
ggplot(iah, aes(DepHour, DayOfWeek)) +
  geom_tile(aes(fill = avg_delay_delayed))
```

Try limiting the color scale to 0 to 120.

Instead of limits try using a log transform, `trans = "log"`.

Compare to instead setting the aesthetic in geom_tile, `fill = log(avg_delay_delayed)`

Remind Charlotte to talk about: `scale_y_continuous`

Guides

Control the appearance of the legend.

Explore on your own:

http://docs.ggplot2.org/0.9.3/guide_colourbar.html

http://docs.ggplot2.org/0.9.3/guide_legend.html

Elements of a polished plot

Element	**Unpolished example**	**Polished example**	
Axis titles	avg_temp	Monthly average temperature (°C)	`+ xlab() + ylab()` or first argument to `scale_xxx_xxx()`
Aesthetic legend labels (axis tick labels, grouping levels)	0.2 0.4 0.6 1000 1500 2000 1e2 1e3 1e4 100000 200000 trt ctrl	20% 40% 60% $1000 $1500 $2000 100 1000 10000 100,000 200,000 treatment control	`labels` argument in `scale_xxx_xxx()`

All text should be unobscured and readable

(make sure it is big enough!)

All legends should be fully viewable and readable

Grid (major & minor breaks) should be meaningful and not distracting

Notes on polishing

Polishing the plot in a graphics editing program (e.g. Illustrator) is fine, but generally avoided because it's hard to reproduce.

You can add non-data text and geoms with `annotate`:

```
+ annotate(geom, x = NULL, y = NULL, xmin = NULL, xmax = NULL,
  ymin = NULL, ymax = NULL, ...)
```

I generally add annotation that is based on the data in R (like lines at important thresholds, or labels for particular data points).

Annotation that is more expository, captions, and titles, I tend to add where the plot is being presented (Word, Latex, Illustrator, Powerpoint).

Direct labeling - it's easier for viewers if they don't have to translate over to a legend to figure out groups. Consider labeling lines etc. directly.