Topics in Data Visualization

Jul 21 2014

Charlotte Wickham

vis.cwick.co.nz

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.

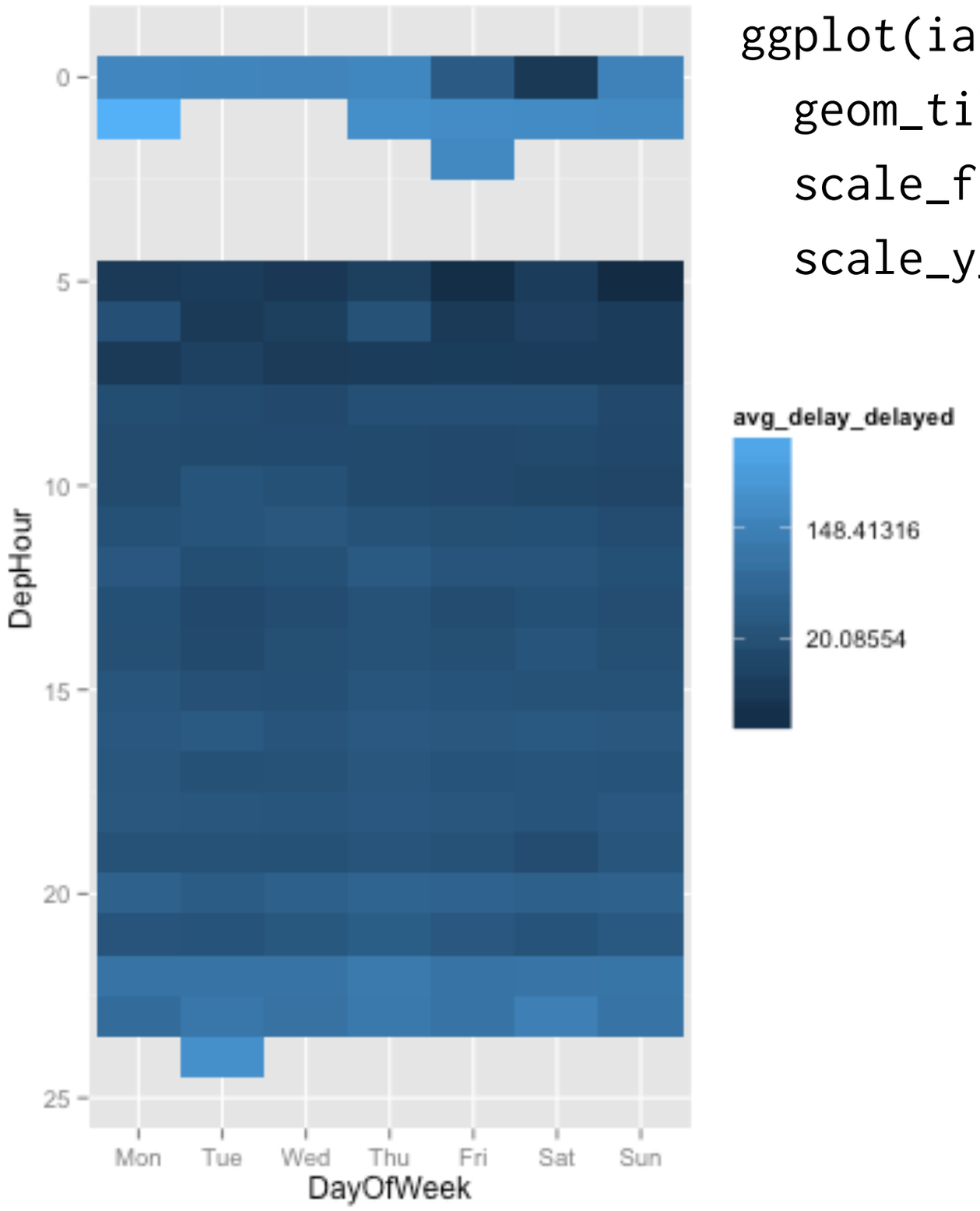
You can add non-data text and geoms with `annotate`:

```
+ 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).

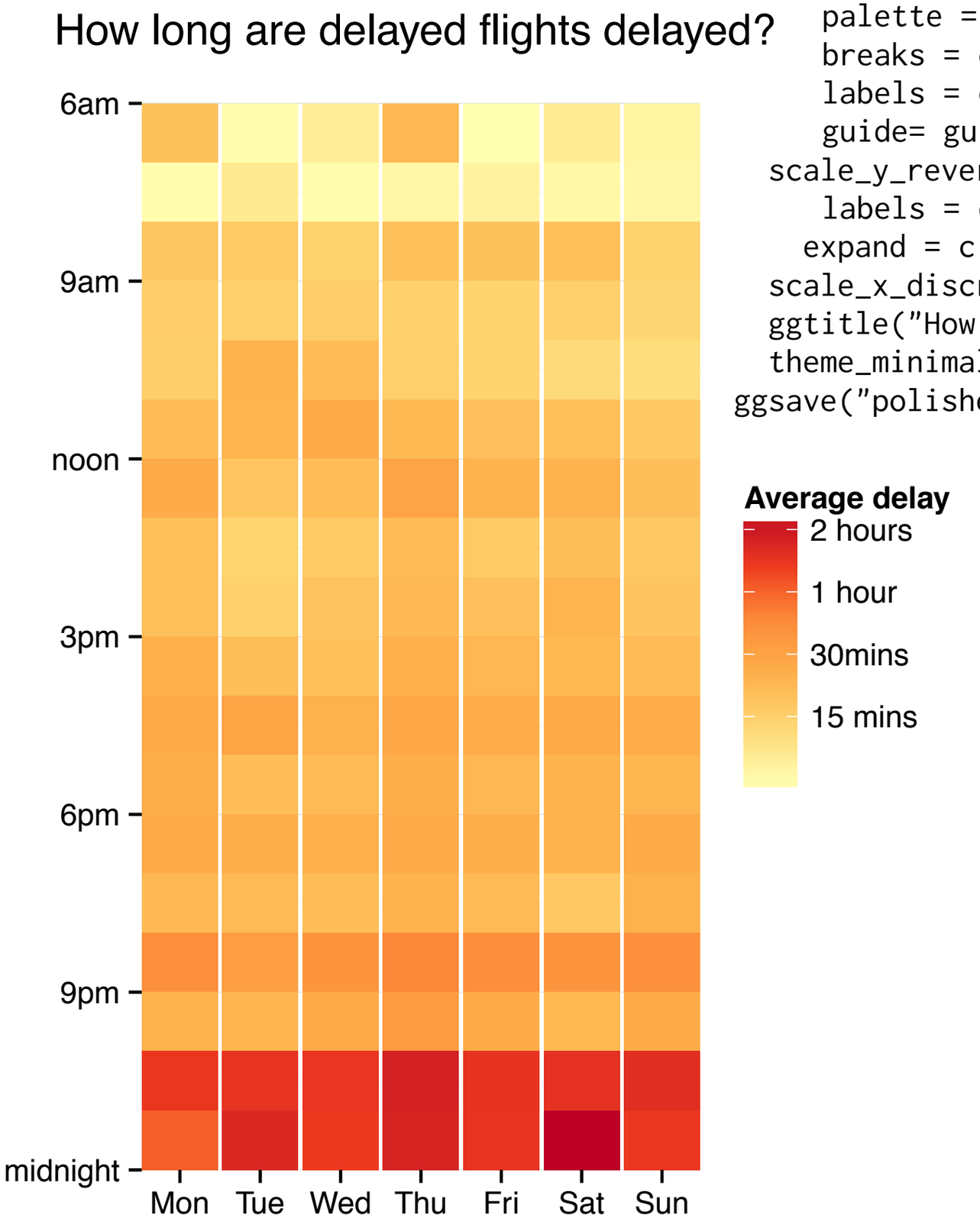
Annotation that is more expository, captions, and titles, I tend to add where the plot is being presented (Word, Latex, Illustrator, Powerpoint).

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.




```
ggplot(iah, aes(DayOfWeek, DepHour)) +
  geom_tile(aes(fill avg_delay_delayed)) +
  scale_fill_gradient(trans = "log") +
  scale_y_reverse()
```

```
ggplot(subset(iah, DepHour > 5 & DepHour < 24),
       aes(DayOfWeek, DepHour +0.5)) +
  geom_tile(aes(fill = avg_delay_delayed), width = 0.95)  +
  scale_fill_distiller("Average delay", trans = “log",
     palette = "YlOrRd",
     breaks = c(15, 30, 60, 120, 240),
     labels = c("15 mins", "30mins", "1 hour", "2 hours", "4 hours"),
     guide= guide_colorbar()) +
  scale_y_reverse("", breaks = seq(6, 24, 3),
     labels = c("6am", "9am", "noon", "3pm", "6pm", "9pm", "midnight"),
    expand = c(0,0)) +
  scale_x_discrete("", expand = c(0,0)) +
  ggtitle("How long are delayed flights delayed? \n") +
  theme_minimal()
ggsave("polished.pdf", height = 6, width = 5)
```



In groups of 4 or 5

Task: Recreate/Create a graphic using the delay data

Grab code from website
Same data as Friday with the addition of another airport (HOU)

Define your own purpose:

"Don't fly in the evening!"

"Flight time scheduler"

"Comparing patterns at IAH and HOU"

Evaluate your plots relative to your purpose.

Ideas from Friday posted online.

You are welcome to create different summaries if you want.

Deliverable

In a new thread under the *Graphics Critique* forum:

- Post your **polished** plot.
- Defend your plot choice based on your purpose.
- Comment on anything you think works particularly well or needs improvement.
- Attach R script that creates and polishes plot.

preferably during class, but at latest before **5pm today**