Topics in Data Visualization

Variability + Interaction/Animation

Jul 25 2014

Charlotte Wickham



Monday critique

Is there anyone who can't make it, but still wants to participate?

You can have your own "online" group.

Email me by Sunday night.

When is it appropriate to display averages?

When the average has a natural interpretation.

When the underlying distribution of data is well described by the average.

How could we communicate the variability in delay times?



One Monday at IAH

Uncertainty

Statistical uncertainty arises because the summary you display is based on data that is generated by a random process.

Is there uncertainty in our average of delays on Mondays at 10am?

Visually-Weighted Regression Solomon M. Hsiang

http://gspp.berkeley.edu/assets/uploads/research/pdf/SSRN-id2265501.pdf



Figure 4: (A) Simultaneous visual-weighting of a confidence interval and regression line using the visual-weighting function in Equation 3. The coloration of the confidence interval is "ink-preserving" because the quantity of ink used to color the graph, integrated along a vertical line, is constant across all values of X. The regression line is visually-weighted because its contrast with the confidence interval increases in regions where the confidence interval is narrower. (B) "Spaghetti" plot of regression results obtained by resampling the data. (C) Applying a mesh over the resampled estimates in B allows us to smooth over the spaghetti in the Y-dimension. (D) Smoothed results for continuous visual-weighting of the confidence interval in both the X- and Y-dimension, as described in Equation 4. This "watercolor regression" uses visual-weighting to depict the certainty of the regression line (via contrast with the confidence interval) as well as certainty within the confidence interval itself (via color saturation).

Gelman, Andrew. "Exploratory data analysis for complex models." Journal of Computational and Graphical Statistics 13.4 (2004).



Figure 9: Top row: (a) Histogram of 90 patient parameters θ_j as estimated from an expanded psychometric model, with the fitted mixture prior distribution overlain. (b) The suspended rootogram shows the simulations to fit the model reasonably well. Compare to Figure 8.

Bottom row: the same plots, but showing (jittered) histograms and suspended rootograms of 50 random draws of the vector θ from its posterior distribution.



Always be very specific about what your error bars represent (SDs, SEMs, X% Cls).

"Error" bars always have some assumptions. If those assumptions are inappropriate, so are your error bars.

What tools does Hans Rosling have that we don't have in ggplot2?

http://www.ted.com/talks/

hans rosling shows the best stats you ve ever seen

Beware animation

https://www.youtube.com/watch? v=1nL5ulsWMYc

Animation can be a useful tool for communication, if you prime people to help them see the changes you want them to.

Do not expect animation to be a good way to reveal interesting surprises.

Interaction

All our interaction with our ggplot2 plots is through code. Fine if you code!

- Interactivity has been available in R for awhile but it's generally been a bit cumbersome.
- shiny + ggvis is probably the way forward

shiny - "Turn your analyses into interactive web
applications"

ggvis - ggplot2-like web-based with interactivity

shiny http://shiny.rstudio.com/gallery/kmeans-example.html

ggvis

library(ggvis)

```
mtc <- mtcars
mtc$id <- 1:nrow(mtc) # Add an id column to use ask the key
all_values <- function(x) {
    if(is.null(x)) return(NULL)
    row <- mtc[mtc$id == x$id, ]
    paste0(rownames(row))
}
mtc %>% ggvis(x = ~wt, y = ~mpg, key := ~id) %>%
    layer_points(size.hover := 50, size.update := 25,
        opacity := input_slider(0, 1)) %>%
    add_tooltip(all_values, "hover")
```

shiny + ggvis

http://shiny.rstudio.com/gallery/movie-explorer.html

Next Week

Monday - group critique

Wednesday -?

high dimensional data (tours, parallel coordinate plots, heatmaps?)

other special topic (maps?, networks?,?)

show and tell (anything you want to share)

work session

Maps or High dimensional data

Friday - poster session