**Wk 5 workshop: effective data viz**

OVERPLOTTING

1. Section on dealing with overplotting, which will help introduce some key elements below in a more interesting way

replicate <https://www.data-to-viz.com/caveat/overplotting.html> (not the Marginal Distribution section, we won’t cover that) – note that you can click ‘CODE’ to see and copy the exact code for the figs.

HOW TO MANIPULATE PLOT FEATURES

1. What is aes()
2. Glossary of manipulations

The emphasis here is getting them to implement the graphical elements, not necessarily make the right choices. Hopefully the lecture will help them make the right choices, the impetus of the workshop is for them to see how each thing is implemented for different plots (and hopefully see that it’s pretty much agnostic to the kind of plot selected in ggplot, because that’s the benefit of using it)

Lean heavily on the plots from last week – some plots will have specific aesthetic things associated with it (e.g. horizontal vs vertical is really only for bar plots and histograms), but lots of elements only need to be introduced once and then can be used throughout all the rest of the sheet.

OR, if easier, lots can be pulled directly from here:

Boxplots:

<http://www.sthda.com/english/wiki/be-awesome-in-ggplot2-a-practical-guide-to-be-highly-effective-r-software-and-data-visualization#graphical-parameters>

Scatterplots:

<http://zevross.com/blog/2014/08/04/beautiful-plotting-in-r-a-ggplot2-cheatsheet-3/>

Note that I have dropped a few example tutorials that cover some of this already into the workshop folder – please borrow as much as possible from those!

Aesthetic adjustments that should be introduced for appropriate plots:

* The plot structure
  + Truncating (for scatterplots/time series)
  + Ordering (I know this is just a reiteration from last week, but good to have them see a direct comparison)
  + Horizontal vs vertical (barcharts)
* Labels & annotations
  + Axis labels (and angles where appropriate, and things like controlling how frequent the points are and whether they are decimals or not)
  + Titles
  + Legend (including position and removing the box around it)
  + Annotating plots with labels (e.g. labelling different time series lines)
  + Insert vertical lines in time series (and if possible label them?)
* Themes
  + Guide lines/grids
* Dots
  + Sizes of points and bars
  + Different shaped points
  + Transparency – including making a subset of the data more transparent than other subsets in order to highlight it (e.g. in the overfitting example above, or making one point in a scatterplot full colour and the rest more transparent to draw the eye)
* Colour
  + where appropriate for each plot, it would be good for them to do it with a bad choice of colour (e.g. too much colour variation where only one would suffice, or a divergent colour palette) and a good choice of colour (e.g. all one colour instead, or instead of using multiple colours, using different intensities of the same colour)
  + Using colour to highlight one thing, e.g. <https://www.reddit.com/r/RStudio/comments/b4dzh4/change_one_bar_to_different_color_in_ggplot/>
  + Palettes (n.b. my favourite is Vridis, let’s indoctrinate the students with that)
* Other
  + Two plots side by side (e.g. the “Position adjustments” section of <http://www.sthda.com/english/wiki/be-awesome-in-ggplot2-a-practical-guide-to-be-highly-effective-r-software-and-data-visualization#graphical-parameters>)
* WHAT ELSE IS MISSING?

1. I would love to find a tutorial we can replicate (or ask the students to replicate) for a plot remake, but I can’t find anything using ggplot2. If you come across something, can you shoot it through to me please? If you don’t, this part isn’t entirely necessary (i.e. don’t spend time explicitly hunting for this). Is there something here? <https://www.quora.com/What-is-the-most-elegant-plot-you-have-made-using-ggplot2-in-R-Embed-the-code-if-possible>

Or here

<https://www.reddit.com/r/rstats/comments/6hz3g3/help_me_improve_this_ggplot2_plot_for_practice/>

ASSESSMENT

TBC -

‘Go back to your plots from the assessment last week and re-design them using the theory from lectures and techniques from the workshop. In the .Rmd file for submission, for each question include your original image at half size with the code hidden (see below how to do this) and the re-designed one with the code printed. For each question give 1 sentence justification for the type of plot you chose (last week’s material) and 1 sentence justification for the aesthetics you chose (this week’s material). ‘