

Conceptual introduction to ggplot

Author: Nicholas G Reich

Made available under the Creative Commons Attribution-ShareAlike 3.0 Unported License: <http://creativecommons.org/licenses/by-sa/3.0/deed.en-US>

Choices for R graphics

You have three central choices for making graphics in R:

- ▶ “Base R graphics”
- ▶ `ggplot2`
- ▶ `lattice`

I use `ggplot` because:

1. it is integrated with the `tidyverse`
2. it is actively developed/maintained
3. there are a ton of extensions (see more later)

Understanding the “grammar” of ggplot2

The grammar ...

- ▶ layer (a ‘geom’, a ‘stat’, an ‘annotation’)
- ▶ aesthetics (‘aes’)
- ▶ scales
- ▶ facets
- ▶ data
- ▶ ... and more here: <http://ggplot2.tidyverse.org/reference/>

What is a layer?

Layers define the basic structure of the elements on the plot

- ▶ **Geoms**: point, line, tile, boxplot, ribbon, ...
- ▶ **Stats**: histogram, smooth, density, ...
- ▶ **Annotation**: hline, vline, text, ...

For more info check out the documentation:

<http://ggplot2.tidyverse.org/reference/>.

What are “aesthetics”?

Aesthetics define a mapping between **tidy data** and the information required to create a specific graphic¹

length	width	depth	trt
2	3	4	a
1	2	1	a
4	5	15	b
9	10	80	b



x	y	colour
2	3	a
1	2	a
4	5	b
9	10	b

¹ Figure credits: Hadley Wickham

geom_point

Each geom has a different set of aesthetics.

What information do we need to draw a scatterplot?

Or, asked another way, what aesthetics do we need for `geom_point`?

geom_point

Each geom has a different set of aesthetics.

What information do we need to draw a scatterplot?

Or, asked another way, what aesthetics do we need for `geom_point`?

- ▶ x (required)
- ▶ y (required)
- ▶ alpha
- ▶ color
- ▶ fill
- ▶ shape
- ▶ size

geom_line

What information do we need to draw a line of connected points?
Or, asked another way, what aesthetics do we need for `geom_line`?

geom_line

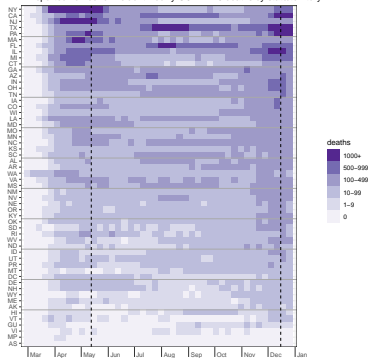
What information do we need to draw a line of connected points?
Or, asked another way, what aesthetics do we need for `geom_line`?

- ▶ x (required)
- ▶ y (required)
- ▶ alpha
- ▶ color
- ▶ linetype
- ▶ size

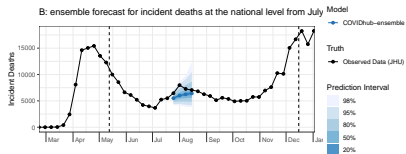
ggplot extensions that I used in [a recent paper](#)

gridExtra or cowplot for multi-plot alignment

A: reported number of incident weekly COVID-19 deaths by state/territory



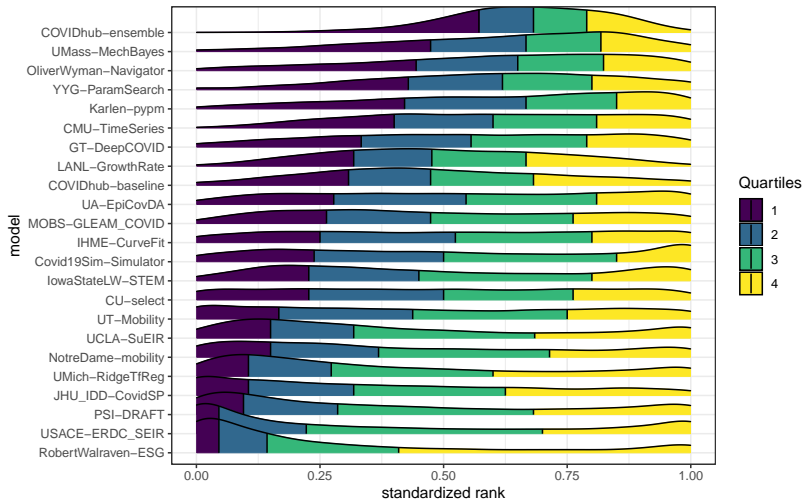
B: ensemble forecast for incident deaths at the national level from July



C: number of models submitting forecasts of incident deaths

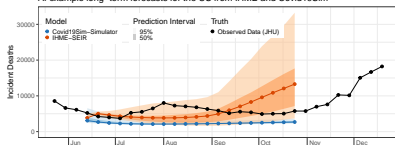


ggrides for ridgeplots

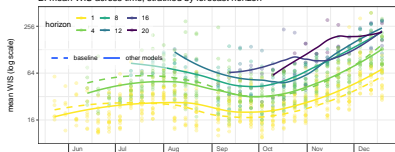


RColorBrewer and viridis for colors

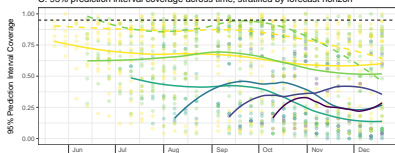
A: example long-term forecasts for the US from IHME and Covid19Sim



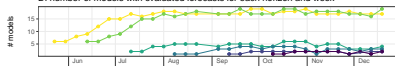
B: mean WIS across time, stratified by forecast horizon



C: 95% prediction interval coverage across time, stratified by forecast horizon



D: number of models with evaluated forecasts for each horizon and week



Breakout rooms

We have made the data from the CO2 emissions figure available in Google Drive. As a group, you will be asked to complete the following tasks:

1. Recreate the figure as close as possible to the original.
2. Improve the figure. Make some changes that you think make the figure more clear.

At the end of class, the class will vote on which figure is (1) closest to the original and (2) the best improvement.