**Mini-Assignment #1**

**Blogs**

One visualization that caught my eye during this week’s readings was the [COVID-19 dashboard](https://coronavirus.jhu.edu/map.html), produced by the Center for Systems Science and Engineering (CSSE; See Figure 1). Specifically, I appreciate the interactive plot for its easy interpretability, navigability, and customizability. After playing around with it, it has a lot of features that make the viewing experience as pleasant as possible: it contains widgets that zoom in or out or focus on particular panes, has options to alter what data is being plotted (e.g., incidence rates, case-fatality ratios, global vaccinations), as well as to acquire relevant percentages for particular countries/US States. Some ways that I might improve the dashboard: 1) Allow user to view the change in the data over time (e.g., with an option to “play” the animation for the data changing over specified time interval), 2) provide user with options to omit particular panes for more minimalistic viewing experience, and 3) provide user with search functions and options to filter the data by particular demographics (e.g., see how data differ by particular subpopulations, ethnicities, genders, etc…).

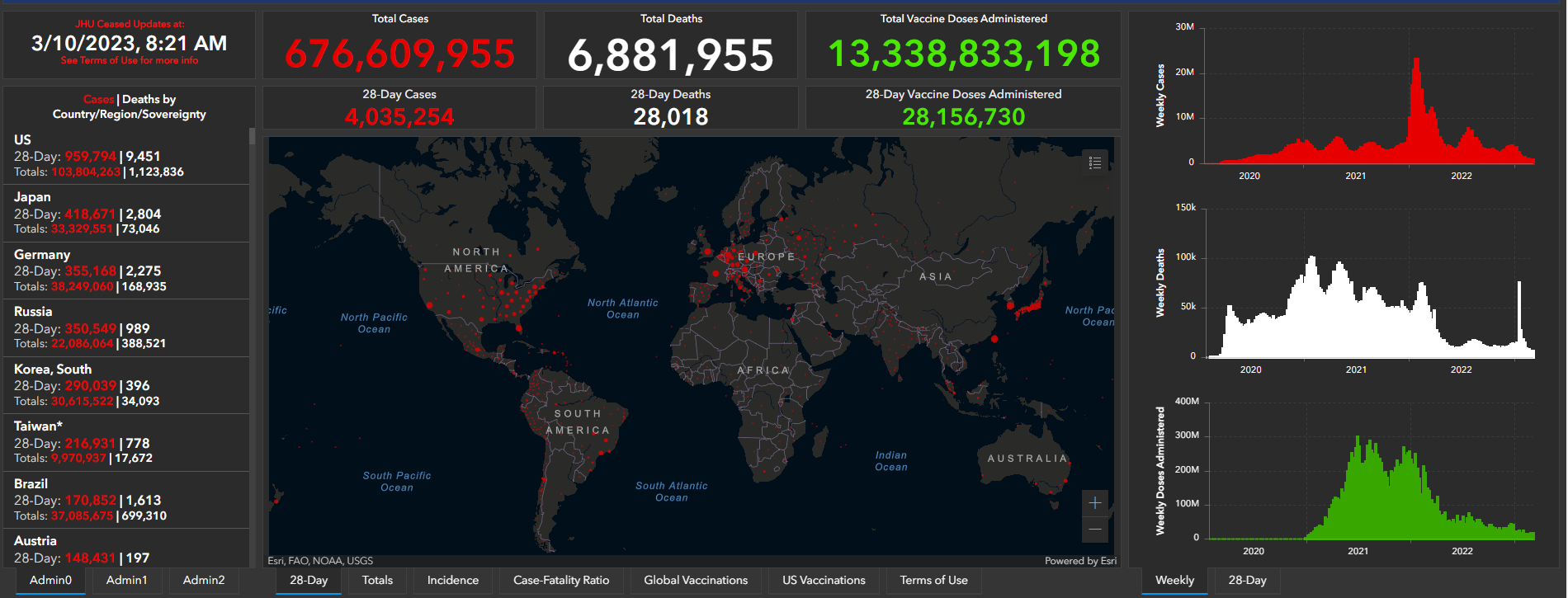
**Good/bad graphs**

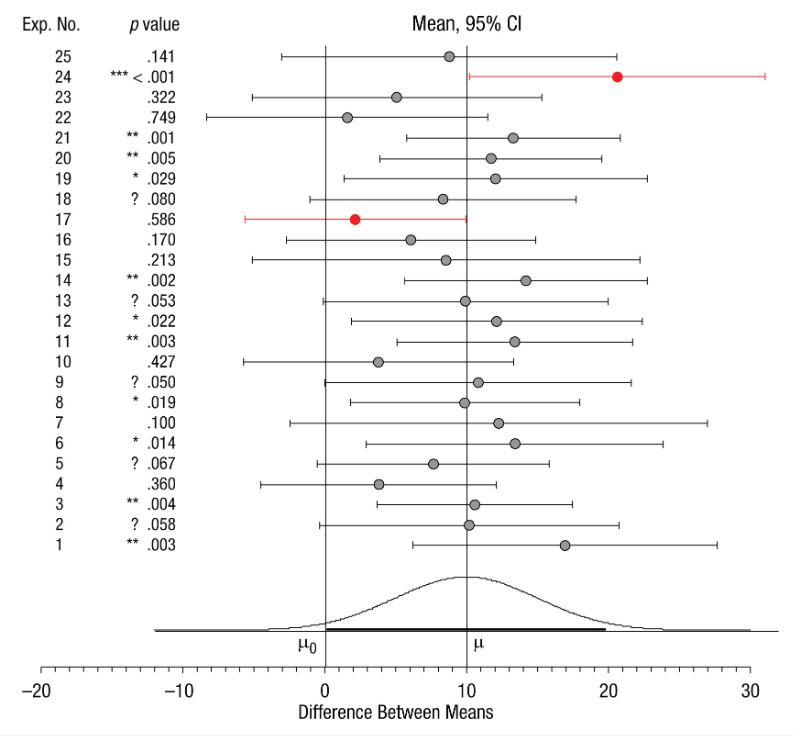
[Cumming (2013)](https://journals.sagepub.com/doi/10.1177/0956797613504966), in his *The New Statistics* paper, makes the argument that 95% CIs should be used instead of *p*-values. He does so by showing a clear and insightful plot: it depicts the *p*-values, means, and 95% CIs obtained from 20 simulated experimental studies (see Figure 1). Each study is based on a simple research scenario common in psychology: an independent samples *t*-test comparing two independent groups on a variable of interest[[1]](#footnote-0). In its minimalistic presentation, the plot allows the viewer to quickly gather the idea that, while the 95% CIs largely capture the true population mean differences (only 2/25 missed it), the *p*-values greatly differed! This plot complements Cumming’s explanation really well, and together they help convince the reader of Cumming’s position.

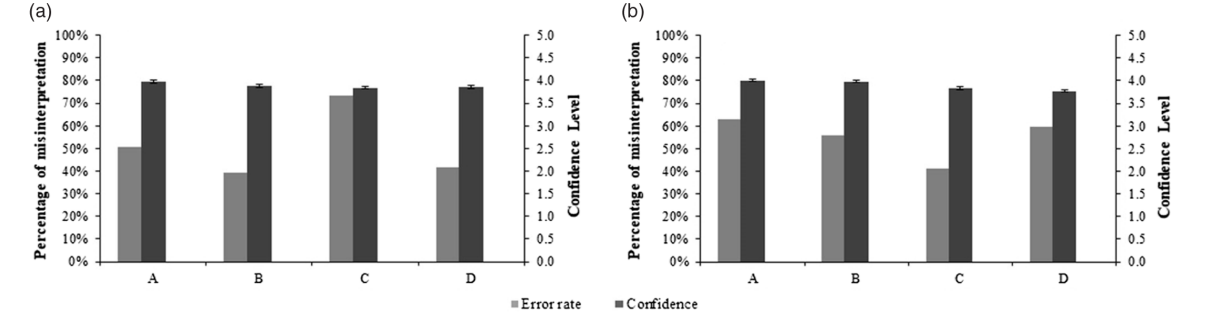
The “bad” graphic I found is from [Lyu et al. (2020)](https://journals.sagepub.com/doi/pdf/10.1017/prp.2019.28) (see Figure 2). The plot is meant to depict the % of misinterpretations (left y-axis) and the reported confidence levels (right y-axis) for both *p*-values (in plot a) and CIs (plot b) for their sample across different items they asked (items labeled A-D on the x-axis[[2]](#footnote-1)). While it is clear that participants tended to be more overconfident compared to their performance, the graphic says very little else. Really, the graphic is comparing the mean confidence score (with 95% CI) to the % of misinterpretations across the entire sample, but in doing so, the viewer has no idea of how the sample varied across these two continuous variables. Thus, the plot is misleading; it’s not as informative as it appears. Additionally, it is unknown what each color of bar represents without having to constantly reference the legend, the use of dual y-axes is evil, and the error bars on the confidence level means are tiny with high contrast with the bar itself.

**PLOTS**

*Covid-19 Dashboard Screenshot*



*Figure 1 From Cumming (2013) The New Statistics: Why and How*

*Figure 2 from Lyu et al. (2020) Beyond Psychology: Prevalence Of p Value And Confidence Interval Misinterpretation Across Different Fields*.

1. As Cumming describes, each group was normally distributed with σ = 20 and *n* = 32. The population means differ by 10 points (standardized ES δ = 0.5). [↑](#footnote-ref-0)
2. A-D actually represent different items for *p*-values vs. CIs! (e.g., A-D (a) = Items #1-4; A-D (b) = Items #5-8) [↑](#footnote-ref-1)