Practical Advice for Producing Better Graphs

Jonathan P. Kastellec*
Department of Politics
Princeton University
jkastell@princeton.edu

March 4, 2025

Abstract

In this short paper I present a few practical tips for producing better published graphs. These include: making labels big enough to read; avoiding legends and labeling lines directly; using small multiple plots; and using different line types and shapes to draw distinctions. I illustrate these suggestions by improving a few example published graphs. Finally, I provide replication code to implement these suggestions using *ggplot*.

*I thank Andrew Gelman and Eric Lawrence for helpful comments and suggestions. I also thank commenters at Gelman's "Statistical Modeling, Causal Inference, and Social Science" blog for helpful comments on this post: `https://tinyurl.com/3ucm9jbs` . The data and code to recreate all the graphs presented in the paper can be found at `https://tinyurl.com/3s7hkadj` .

Introduction

Compared to earlier eras, it is undoubtedly true that political scientists today are more likely to present statistical graphics in their published research, both in addition to and instead of using tables.[1] This shift can be attributed to several factors, including the increased popularity of *R* and the *ggplot2* package, as well as the increased ability of journals and book publishers to publish high-quality graphics, including graphs with colors.

At the same time, many published graphs leave much to be desired, in that they make the reader work harder to understand a graph than is either desirable or necessary. In this paper I present a few practical tips for producing better graphs. To be perfectly clear, none of these tips are original; they can be found in some or all of the excellent "how-to" books on producing graphs.[2] Rather, I focus on a few very basic suggestions that apply, in some cases, to all graphs, and in others, to many types of graphs that social scientists tend to produce.[3] To be clear, these are suggestions designed for better *published graphs*; they are not generally applicable for scholars' own visualizations in the course of their research workflow. Indeed, some of the basic problems arise because what works well as a default during the data exploration phase does not translate well to published graphs.

The paper is structured around four main suggestions. For each suggestion, I present the original version of a graph that could be improved and a revised version that incorporates a given suggestion or suggestions.[4] (Each revision usually involves two or more changes; I

[1]For advice on how to use graphs instead of tables, see Gelman, Pasarica and Dodhia (2002) and Kastellec and Leoni (2007).

[2]To list a few such books: Tukey (1977), Cleveland (1993), Tufte (2001), Chang (2018), Wickham (2016), Healy (2018), Murrell (2019), and Schwabish (2021); see also Franconeri et al. (2021) for a helpful review of recent advancements in the science of data visualizations. For an nice example of a political science book that applies these suggestions effectively, see Gelman (2009). Because I focus on a few basic suggestions, I omit detailed discussions of several more advanced issues surrounding the quality of data visualization, such as color choice and accessibility in visualizations. Where relevant, however, I point interested readers to suggested further reading in footnotes.

[3]At risk of de-anonymizing my reviews, these are the suggestions I tend to make most often when reviewing papers.

[4]The point of this exercise is certainly not to single out the authors of these graphs. The issues I note are quite general, and for each suggestion there are numerous candidate graphs I could focus on. These graphs

make these changes for each new graph, but the text for a given suggestion usually focuses on the relevant specific suggestion.) Code for each of the revised graphs, along with the necessary data to create them, can be found at `https://tinyurl.com/3s7hkadj` . Given the popularity of *ggplot* among political scientists, all of the code is written using that package.[5]

Make labels big enough to read easily

This mistake is both the most innocuous and the easiest to remedy. The defaults on statistical packages, including *ggplot*, are often designed to facilitate the visualization of the graph by the researcher him or herself. This includes defaults on text and numeric sizing, including titles, axis labels, and legends, among others. The problem is that what looks fine on one's computer screen can be too small once a graph is rendered for a final publication. In particular, the production of high-resolution images by journals can lead to graphs that are relatively quite small in terms of area on a page. While such figures are normally quite sharp in resolution, the reduction process can lead to labels that are too small to read easily.

As an example, consider the left graph in Figure 1, which reproduces Figure 1 from Hankinson and Magazinnik (2023). The point of this graph is to summarize the supply of housing units in cities, comparing them across treatment and control units, where the treatment is a switch from at-large to district elections for city council; the x-axis shows the years to the first district election (i.e. the "treatment" is at zero). While the figure is relatively clear, the axis labels, the tick mark labels, and the legend are difficult to read as rendered, because the text is quite small.

The right graph in Figure 1 reproduces this figure, but using larger labels. (Note that I reproduced the left graph in *R* using the same figure dimensions as the right graph so as to

met the simple criteria that a) they happened to be in articles I have read; and b) the authors made their code available, for which I thank them.

[5]To be clear, the suggestions are general and can applied to any statistical program and/or package, including Stata. In some cases, the defaults in *ggplot* actually make it easier to make some of the mistakes I document; for example, producing legends instead of labeling lines directly.

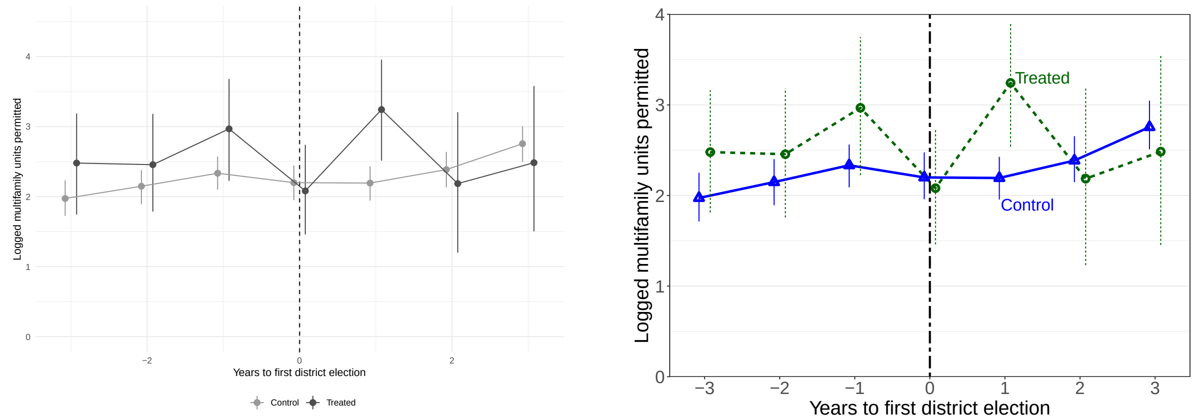


Figure 1: The left graph reproduces Figure 1 from Hankinson and Magazinnik (2023). The right graph uses larger text, making the labels easier to read.

make the sizing comparison an apples-to-apples one.) This simple change makes the graph much easier to read, particularly the axis labels.[6]

The revised graph also incorporates several other changes; some of these fall into the recommendations discussed later, but I briefly note them here. First, I label the lines directly instead of using a legend. Second, I use different line types for the treatment and control lines, as well as different point types, and add color to both, making them easier to distinguish. Third, I add labels for every tick mark on the x-axis; I also reduce the whitespace at the top and bottom of the graphs, and add a solid line around the perimeter of the plot region, which helps clarify the minima and maxima of the confidence intervals.[7] Finally, while this is mostly a matter of taste, I remove the vertical grid lines; this helps render more clearly both the vertical confidence intervals and the dashed line at zero.

Avoid legends and label lines directly

The next suggestion involves the visualization of grouped data, which is common in political science. The default method to distinguish groups in many statistical graphics

[6]The default text size in *ggplot* is 11. While the optimal size will be depend on the height and width of a given graph, my general rule of thumb is to set the axis text size (i.e. for labeling tick marks) to 14 and to set the axis label size to 18.

[7]I usually use `theme_bw` in *ggpplot*, which places a box around the plot region by default.

programs, including *ggplot*, is to use a legend, usually to the right of or under the main plot region. While legends are necessary in some cases, they have the distinct disadvantage of forcing a reader to map a particular group from the legend to the distinguishing plot type (e.g. lines, bars or points). While legends are sometimes unavoidable for space reasons, in most cases it is far superior to label lines, bars or points directly using text.[8]

As an example, consider the top graph in Figure 2, which reproduces Figure 3 from Grumbach (2022). The point of this graph is to emphasize a dramatic decline in democratic performance in North Carolina (based on a democracy index created by the author) in the early 2010s. To visualize this shift, Grumbach compares North Carolina to the other 49 states; Texas and Washington are singled out for specific comparison, while the light gray lines show the scores for the remaining states. While overall this is a nice graph, the use of a legend creates work for the reader in figuring out which line goes with which state. In addition, the legend takes up space that could be better used for displaying the data. (One nice feature of the graph is the states are depicted using different line types, which is a suggestion I turn to shortly.)

The bottom graph in Figure 2 reproduces this figure, but instead directly labels the state lines. With this simple change, the correspondence between states and lines is now immediate.[9] The saved horizontal whitespace from removing the legend also allows for larger axis tick and title labels.[10] In addition, while it not strictly necessary, I add color to further help distinguish the lines, and remove the "Year" label from the x-axis, since it is obvious in the context of the article and graph. Finally, I add minor tick marks (i.e. at every

[8] On the general importance of using text to improve data visualizations, see Borkin et al. (2015) and Bromley and Setlur (2023).

[9] I usually add labels manually in *ggplot* using the "annotate" function. There are some packages, however, that automate the process. See e.g. `https://r-graph-gallery.com/web-line-chart-with-labels-at-end-of-line.html` and `github.com/AllanCameron/geomtextpath` .

[10] Even if one uses a legend, it can often be placed in the plot region itself if one wants to waste space that could be used for the graph itself. In this case, for example, the legend could be placed horizontally in the lower left corner of the graph.

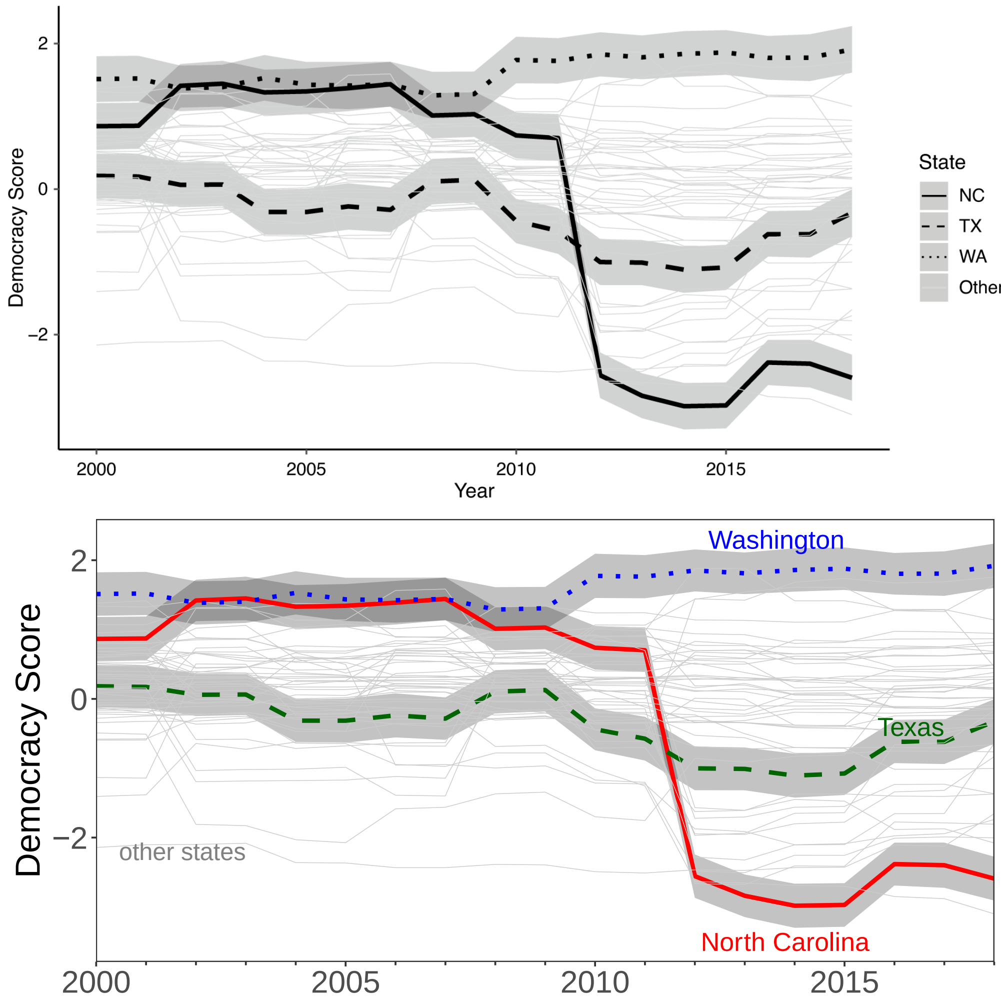


Figure 2: The top graph reproduces Figure 3 from Grumbach (2022). The bottom graph eschews the legend and labels the key lines directly.

year in between the major five-year intervals); this allows the reader to see more clearly the years (following redistricting in 2010) in which the estimated North Carolina decrease in democratic performance occurred.

Embrace Small Multiples

This suggestion also involves grouped data. While there is no set rule for when this occurs, at some point the number of groups becomes too large to effectively display in a single plot. When that happens, a better strategy is to use "small multiple" plots (or simply "small multiples"), a term coined by Tufte (2001), in which a single repeated graphical structure is used for every group. As Schwabish (2021, 42) argues, "the small multiples approach has at least three advantages: First, one the reader understands how to read one [plot], they know how to read all the [plots]. Second, you can display lots of information without confusing your reader. Third, small multiples let readers make comparisons across multiple variables."

As an example of the advantages of the small multiple approach, consider the top graphs in Figure 3, which reproduce Figure 3 from a 2019 note from the editors of the *American Political Science Review*. The graphs summarize data on both the duration of reviews ("days") and the number of words written by reviewers ("tokens"), broken down by subfield. The graph can be thought of a "spaghetti plot," withs lots of lines going all over the place.[11] While spaghetti plots can be useful in some contexts—for example, they are often good at visualizing the spread of hurricane forecasts (Sanyal et al. 2010)—in many cases they make it difficult to discern any structure in the data. In the APSR case, it is both very difficult to tell which lines go with which subfield and to discern the trends in the data, if there are any. For example, in discussing these figures, the editors (p. vii) state that "it is important to note that both indicators have been astonishingly stable across years on average, thus not revealing any signs of change in reviewer fatigue," but this conclusion is not readily apparent by looking at the respective graphs.

By contrast, the bottom plots in Figure 3 present small multiples version of the two APSR plots, broken down by subfield. Instead of being jammed together, each subfield gets its own plot, making it much easier to discern the trends for each. For example, while many of the subfields do exhibit stability across the time period, the left plot shows a steep decline in the average review time for methodology papers; the right plot shows a less severe but still noticeable decline in average length per review. To be sure, the small multiples approach makes it more difficult to make comparisons *across* subfields, since this now requires

[11] As best as I can tell, the term 'spaghetti plot" was coined by Allen (2010, 128).

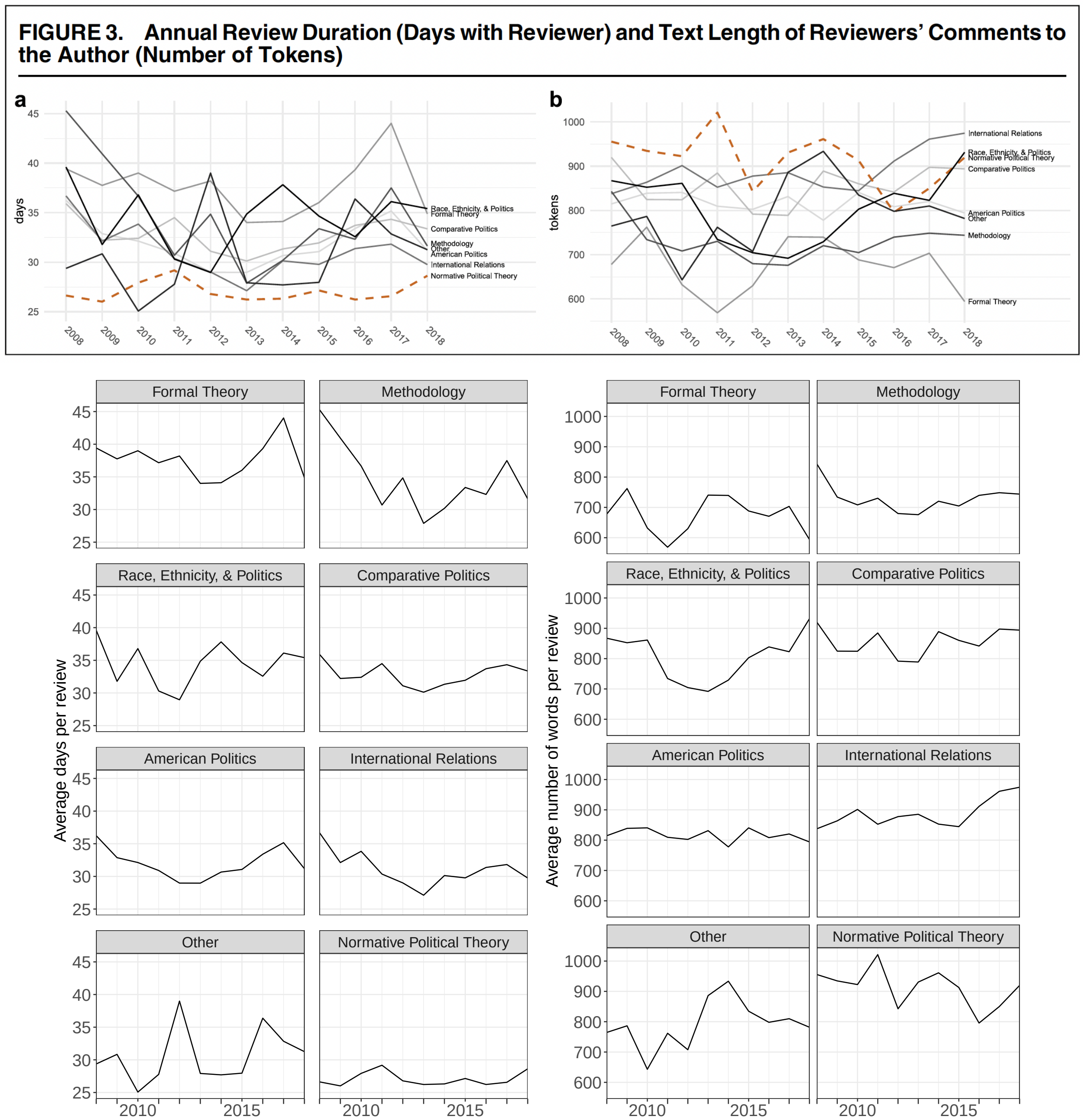


Figure 3: The top graph reproduces Figure 3 from the APSR's editors' note. The bottom graphs use small multiples instead of "spaghetti plots."

comparison across plots. But this tradeoff is worth it, given that such comparisons are effectively impossible in the spaghetti plot version.

A couple of points about the construction of the small multiple plots are worth noting. First, the default in *ggplot* is order the panels (or "facets") alphabetically. In most instances, this default should be avoided, as it results in what Wainer (2005, 72) calls the "Alabama first" error—the idea is that alphabetical order is essentially arbitrary and provides no useful information to the reader in making comparisons across panels.[12] Instead, a better choice is to order the panels based on some interesting type of ordering in the group-level data. The APSR groups be ordered in any number of ways. Here, I choose to sort them from the highest average days per review (pooling across all years) to the lowest average days per review; this results in formal theory going first and political theory going last. I then maintain this ordering in the right graph. (I also use informative y-axis labels and avoid the jargon of "tokens," which has no intuitive meaning. In addition, in this case it is not necessary to label every individual year on the x-axis; that creates unnecessary clutter.)

Second, as a general principle, the panels in small multiple plots should all have the same x- and y-axis limits, in order to make it easier to make comparisons across the plots. The vastly differing scales of the days per review and words per review variable makes that effectively impossible for the y-axis (though the y-axis scales are fixed *within* the left graph and right graph). To avoid this issue, one could simply choose to present each graph in separate figures. Alternatively, the two outcomes variables could be normalized, such that their scales would be directly comparable.[13]

Use different line types and shapes to draw distinctions

This suggestion can be viewed as a cousin of the suggestion to label line types directly. When presenting data that involves two or more distinctions, whether it be groups or different types of suggestions, it is good practice to construct a graph that maximizes the ability of

[12]The name comes from the common default practice of presenting data at the levels of U.S. states in alphabetical order—Alabama is the first state. The "Alabama first" principle applies to any ordering of categorial data in a plot—e.g. an ordered dotplot—and not just to small multiples.

[13]The problem of comparisons being difficult to make with spaghetti plots also frequently arises with stacked bar plots, especially when more than two variables are stacked together. In general, either side-by-side barplots (like Figure 4 below) or small multiple plots are preferable to stacked barplots, since it very difficult to compare across the interior bar categories, as they do not share a common baseline (Cleveland and McGill 1984).

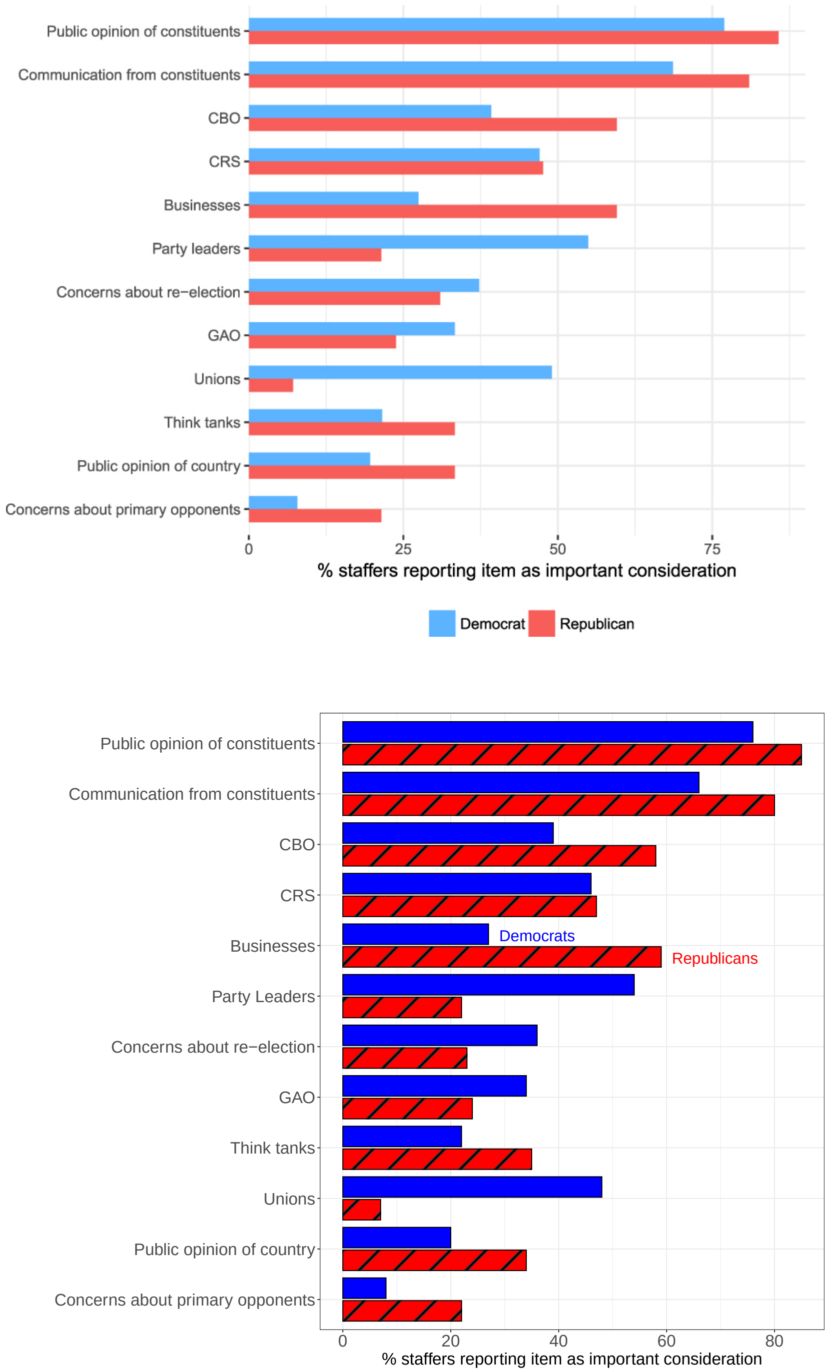


Figure 4: Using different barplot patterns to distinguish groups. The top graph reproduces Figure 2 from Hertel-Fernandez, Mildenberger and Stokes (2019). The bottom graph adds cross-hatches for the Republican bars (and labels the bars directly).

the reader to be able to discern the author's desired distinctions, no matter how the reader is viewing the graph (e.g. on a computer or on a black-and-white paper printout).

I present two examples in this section to illustrate this principle. The top graph in Figure 4 reproduces Figure 2 from Hertel-Fernandez, Mildenberger and Stokes (2019); it presents the results of a survey question in which senior congressional staff in the U.S. Congress are asked the following: "Think about the policy proposals you have worked on during your time on the Hill. What shaped your thinking on whether your Member should support or oppose these policies? Indicate how important each of the following considerations was in shaping your advice to your Member on various policy proposals." The graph shows the percentage of staffers reporting each item as an important consideration, broken down by party.

The graph uses a side-by-side barplot to show the results. If you are reading this paper on a computer or a paper copy with color printing, the graph is quite effective. The blue and red bars correspond to the usual party color connotations, and the categories are sorted from most important to least important (i.e. the authors avoid the "Alabama first" error). In addition, it is clear where priorities differ by party; unsurprisingly, for example, Republican staffers give more weight to businesses and Democratic staffers give more weight to unions.

If, by contrast, you are reading this in black-and-white, the differences between the party are basically unintelligible. This is because the use of a same pattern (i.e. fill) for the bars, combined with the use of a legend, effectively makes the bars indistinguishable from one another if viewed in grayscale. While some journals now feature color printing, not all do (even if they use color on the web-based versions of articles); indeed, in the print issue of the *American Political Science Review* in which the graph appears, it is presented in grayscale, meaning that readers of the actual journal issue cannot tell the difference between the bars.

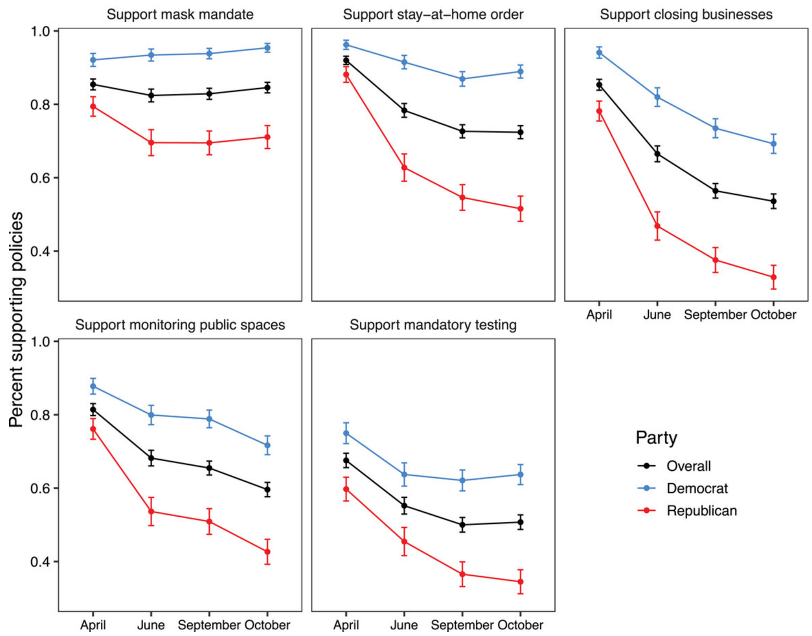
Given this possibility, it is good practice to construct graphs such that they are fully readable even if printed in grayscale. The bottom graph in Figure 4 accomplishes this goal by adding "cross-hatching" for the Republican bars. I maintain the color in the bars, such that the graph is still more effective (naturally) in color. In addition, the new version removes the legends and labels two bars directly (the choice of where to place these labels is somewhat arbitrary—the first two bars are the most natural candidate but the labels do not fit in the plot region when the bars go all the way across at the top). The combination of the cross-hatching and labeling the bars directly helps the reader in distinguishing the responses by party.[14]

The second example comes from Mehlhaff et al. (2024), who examine partisan differences in responses to COVID-related mitigation policies, across a multi-wave panel survey conducted in 2020. The top graph in Figure 5 reproduces Figure 1 from their paper; it plots average support for five mitigation policies across four months of the survey, broken down by opinion among Democratic and Republican identifiers, as well as overall opinion (that is, including all respondents). The vertical error-bars depict 95% confidence intervals for each estimate.

While the results are intuitive–Democrats are always more supportive of each policy than Republicans—the fact that the line types and point types are the same for each group makes it more difficult to distinguish them than is necessary (and in greyscale, they are indistinguishable). The legend is somewhat helpful, but in addition to the general problem with using legends (discussed above), the ordering in the legend being different than the ordering in the plots further complicates the mapping between groups and lines.

The bottom graph in Figure 5 provides two simple remedies. First, I use separate line types for the three groups, which makes it easier to distinguish them. Second, instead of using a legend, I label the lines directly in the first panel. Because the ordering of the groups in the same in each panel, just labeling the first panel seems sufficient; if the ordering

[14]Another issue with relying solely on color to distinguish line types and shapes is that, depending on the colors used, readers with color-blindness may not be able to distinguish between the colors. For general advice on how to use colors, see Setlur and Stone (2015), Ferreira (2020) and Nelli (2024); the website `colorbrewer2.org` also provides specific suggestions for workable color choices. While it is always good practice to choose colors that can be distinguished by color-blind readers, using different line types is a effective tool to also aid in distinguishing. (A related issue is that some readers may other types of vision impairments that inhibit proper procession of data visualizations; for overviews of improving accessibility when making graphs, see Elavsky, Bennett and Moritz (2022) and Schwabish and Feng (2022).)



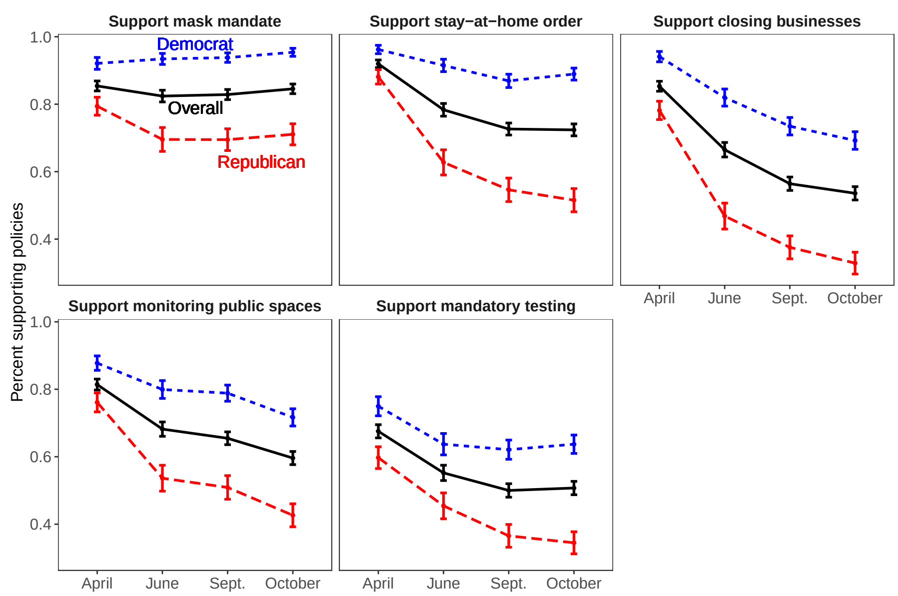


Figure 5: Using different line types to distinguish groups. The top graph reproduces Figure 1 from Mehlhaff et al. (2024). The bottom graph uses different line types to distinguish the partisan groups.

changed across policies, or if the groups were closer together on some of the issues, labeling many or all of the panels might be desirable. Finally, I bold the panel titles in order to make it clearer that they are not x-axis labels; I also abbreviate "September" to allow the axis labels to be larger.

Conclusion: Keep it Simple and Avoid Confusion

One way to unify all the above suggestions is with the slogan, "keep it simple and avoid confusion." It has never been easier to make effective data visualizations, but it is still incumbent on producers of graphs to keep their readers—and how they will perceive a graph—in mind. I hope these suggestions will help producers of statistical graphics achieve this goal.

References

Allen, Theodore T. 2010. *Introduction to Engineering Statistics and Lean Sigma: Statistical Quality Control and Design of Experiments and Systems*. Springer Science & Business Media.

Borkin, Michelle A, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister and Aude Oliva. 2015. "Beyond Memorability: Visualization Recognition and Recall." *IEEE Transactions on Visualization and Computer Graphics* 22(1):519–528.

Bromley, Dennis and Vidya Setlur. 2023. What is the Difference Between a Mountain and a Molehill? Quantifying Semantic Labeling of Visual Features in Line Charts. In *2023 IEEE Visualization and Visual Analytics (VIS)*. IEEE pp. 161–165.

Chang, Winston. 2018. *R Graphics Cookbook*. Sebastopol, CA: O'Reilly Media.

Cleveland, William S. 1993. *Visualizing Data*. Hobart Press.

Cleveland, William S and Robert McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79(387):531–554.

Elavsky, Frank, Cynthia Bennett and Dominik Moritz. 2022. How Accessible is my Visualization? Evaluating Visualization Accessibility with Chartability. In *Computer Graphics Forum*. Vol. 41 Wiley Online Library pp. 57–70.

Ferreira, CR. 2020. "Two Simple Steps to Create Colorblind-Friendly Data Visualizations." Available at `https://tinyurl.com/y8zm6w3j`.

Franconeri, Steven L, Lace M Padilla, Priti Shah, Jeffrey M Zacks and Jessica Hullman. 2021. "The Science of Visual Data Communication: What Works." *Psychological Science in the public interest* 22(3):110–161.

Gelman, Andrew. 2009. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton: Princeton University Press.

Gelman, Andrew, Cristian Pasarica and Rahul Dodhia. 2002. "Let's Practice What We Preach: Turning Tables into Graphs." *The American Statistician* 56(2):121–130.

Grumbach, Jacob M. 2022. "Laboratories of Democratic Backsliding." *American Political Science Review* pp. 1–18.

Hankinson, Michael and Asya Magazinnik. 2023. "The Supply-Equity Trade-off: The Effect of Spatial Representation on the Local Housing Supply." *The Journal of Politics* 85(3):1033–47.

Healy, Kieran. 2018. *Data Visualization: A Practical Introduction.* Princeton: Princeton University Press.

Hertel-Fernandez, Alexander, Matto Mildenberger and Leah C Stokes. 2019. "Legislative Staff and Representation in Congress." *American Political Science Review* 113(1):1–18.

Kastellec, Jonathan P and Eduardo L Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics* 5(4):755–771.

Mehlhaff, Isaac D, Timothy J Ryan, Marc J Hetherington and Michael B MacKuen. 2024. "Where Motivated Reasoning Withers and Looms Large: Fear and Partisan Reactions to the COVID-19 Pandemic." *American Journal of Political Science* 68(1):5–23.

Murrell, Paul. 2019. *R Graphics, Third Edition.* Boca Raton, FL: Taylor & Francis.

Nelli, Luca. 2024. "Color Quest: An Interactive Tool for Exploring Color Palettes and Enhancing Accessibility in Data Visualization." *Plos one* 19(3):e0290923.

Notes from the Editors. 2019. *American Political Science Review* 113(2):v–vii.

Sanyal, Jibonananda, Song Zhang, Jamie Dyer, Andrew Mercer, Philip Amburn and Robert Moorhead. 2010. "Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Ucertainty." *IEEE Transactions on Visualization and Computer Graphics* 16(6):1421–1430.

Schwabish, Jonathan. 2021. *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks.* New York: Columbia University Press.

Schwabish, Jonathan, Sue Popkin and Alice Feng. 2022. "Do No Harm Guide: Centering Accessibility in Data Visualization." Urban Institute. Available at `https://tinyurl.com/ms76fezm`.

Setlur, Vidya and Maureen C Stone. 2015. "A Linguistic Approach to Categorical Color Assignment for Data Visualization." *IEEE transactions on visualization and computer graphics* 22(1):698–707.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information, Second Edition.* Cheshire, CT: Graphics Press.

Tukey, John W. 1977. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley Publishing Company.

Wainer, Howard. 2005. *Graphic Discovery: A Trout in the Milk and Other Visual Adventures.* Princeton: Princeton University Press.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis, 2nd Edition.* Springer.