Uprooting sub-standard visualisation practices for decision-making in operational management

Emi Tanaka

Department of Econometrics and Business Statistics, Monash University, Melbourne, VIC 3800, emi.tanaka@monash.edu

Jessica Wai Yin Leung

Department of Econometrics and Business Statistics, Monash University, Melbourne, VIC 3800, jessica.leung@monash.edu

Dianne Cook

Department of Econometrics and Business Statistics, Monash University, Melbourne, VIC 3800, dicook@monash.edu

1. Introduction

Basole et al. (2021) express visualisation can aid or mislead depending on its use – we agree that visualisation is indeed a double-edged sword that can uncover unexpected information and communicate information effectively but poor practices can lead to misinterpretation or over-emphasis of a feature in the plot. Basole et al. (2021) apply the Gestalt principles of perceptual organization to demonstrate misrepresentations by inclusion and omission on a graph. We concur that appropriate visual encoding is vital for enhancing the visual decoding accuracy of a graph, however there is more discussion needed on the visual encoding process and the incorporation of statistical thinking when working with visual displays of data. These two points form the focus of this commentary.

2. Visualising encoding choices

Encoding refers to the process and choices made for displaying the data in a graph. Decoding describes the process of interpreting or perceiving the information from the graph. Different encoding approaches lead to differences in decoding information. In a visual encoding process, retinal variables, such as position, size, shape, color hue, color intensity, orientation, texture, and motion,

are used to represent different information in the data (Bertin 1983, Munzner 2014). Christodoulou (2018) provides a website with convenient explanation with a compact diagram to illustrate such choices. The encoding process was formalised into a grammar by Wilkinson 2005 that specifies the mapping of the data variables to retinal variables. The grammar was extended and implemented in the `ggplot2` R-package (Wickham 2016). The connection with statistics is tightened because the mapping requires that data is organised as "tidy data" (Wickham 2014), where variables are in columns and observations are in rows. For example, the data used in Figure 3 from Basole et al. (2021) contains the variables: state, coverage, readmission rate and cancer screening percentage, and additionally two spatial variables (x, y) specifying state tile location. These variables are mapped as follows to produce the visual display: x to the x-axis (like longitude), y to the y-axis (like latitude), size to coverage, using a circle geometry, and color to readmission rate (left plot), or cancer screening percentage (right plot).

Basole et al. (2021) use Figure 3 to assert that "In this depiction, one can discern that states with highly integrated care also tend to have (i.e. are predictive of) higher cancer screening rates and lower 30-day hospital readmissions." – we interpret integrated care to mean coverage here. *We strongly disagree with this statement, and feel that the authors have misinterpreted the plot.* A better visual encoding for examining association between two variables is the humble scatter plot, where coverage is mapped to the x-axis and either the readmission rate or cancer screening percentage is mapped to the y-axis, as shown here in Figure 1. Additionally, a local regression fit is added and text annotation displaying the sample correlation coefficient. What can be seen is that there is *NO strong associations* between coverage and either of the other two variables, disagreeing with the conclusion of Basole et al. (2021).

Making assertions from a single plot is flawed. And this is where statistical thinking needs to be employed. Making a conclusion from a single plot, is analogous to making a conclusion from a single numerical summary statistic without carefully calibrating by the standard deviation. Plots are (graphical summary) statistics and thus statistical inference machinery ought to be used to assess the significance of structure seen in the plot (Wickham et al. 2010).

3.   Making inferences from a single plot is prone to over-reading it

A single plot is commonly used to make inferences or assertions, however, this may result in over-reading a plot. A (data) plot can be treated as a statistic with its variation calibrated so as not to wrongfully over-emphasise an observed feature in the plot when the feature is a common artefact of random sampling. Buja et al. (2009) introduced the lineup protocol to calibrate a data plot by embedding it randomly among a set of null plots (a null plot has the same mapping and geometrical rendering as the data plot, using data generated under the null hypothesis of no pattern existing).
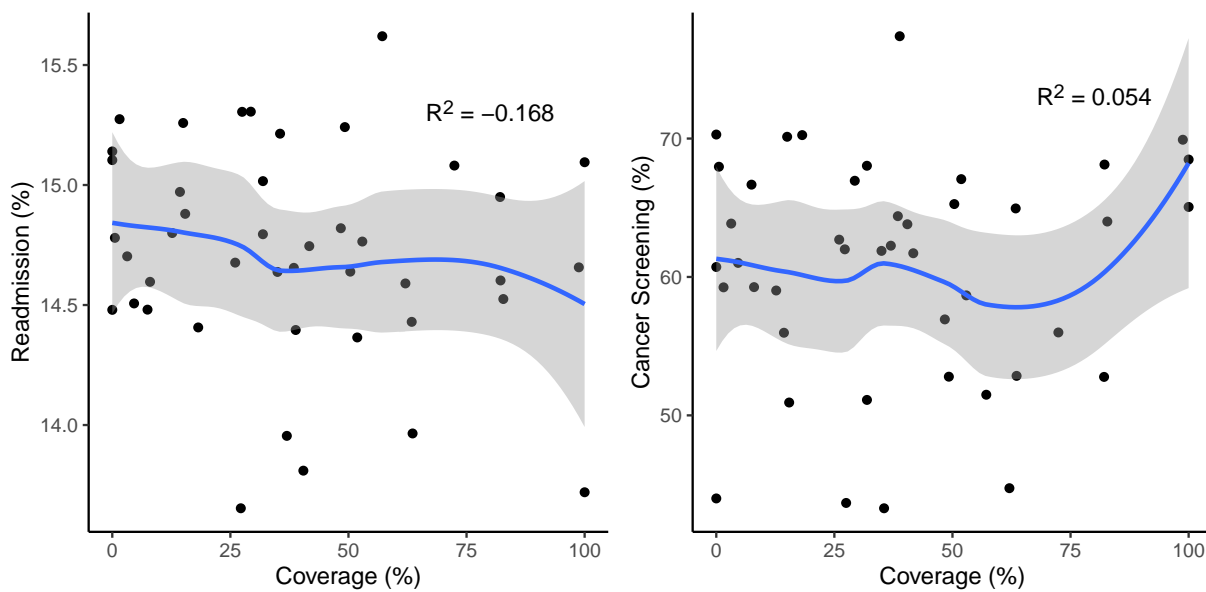
Figure 1 **The above figure show an alternative plot design to display the information in Figure 3 of Basole et al. (2021). The plot shows a scatter plot of percentage of readmission and coverage on the left and a scatter plot of percentage of cancer screening and coverage on the right. Both plots are superimposed by a local polynomial regression (displayed as a blue line) as implemented by the `loess` function in R with confidence interval for the line (displayed in gray) as implemented by the default `geom_smooth` function in the `ggplot2` package. See Figure S2 in the Supplementary Material for the code.**

The lineup protocol also adheres to a hypothesis testing framework and can be used to test for statistically significant patterns in plots. It would be recommended to remove any labels that may bias the reader because of pre-conceived notions about a problem. Figure 2 provides an example lineup as might be used in place of the right plot in Figure 3 from Basole et al. (2021). The null plots in Figure 2 are generated by permuting cancer screening percentage against the other variables, thus breaking any association with coverage. If the data plot exhibits a strong association between cancer screening percentage and coverage then most observers will detect it from among the null plots. On the converse, if most observers cannot detect the data plot, then it is not significantly different to null plots, and one must conclude that there is no association between coverage and cancer screening percentage. A lineup is presented in Figure 2 so you may use this to test if you can find a plot that appears strikingly different to others (the position of the data plot is only revealed in the Supplementary Material).

The lineup protocol is a rigorous method to compare the power of different visual encodings (see examples in Hofmann et al. 2012). The same lineup can be constructed using different visual encodings, and shown to multiple observers. The most powerful visual encoding will correspond to the data plot being detected the lineup more often. In the previous section, we advised that the
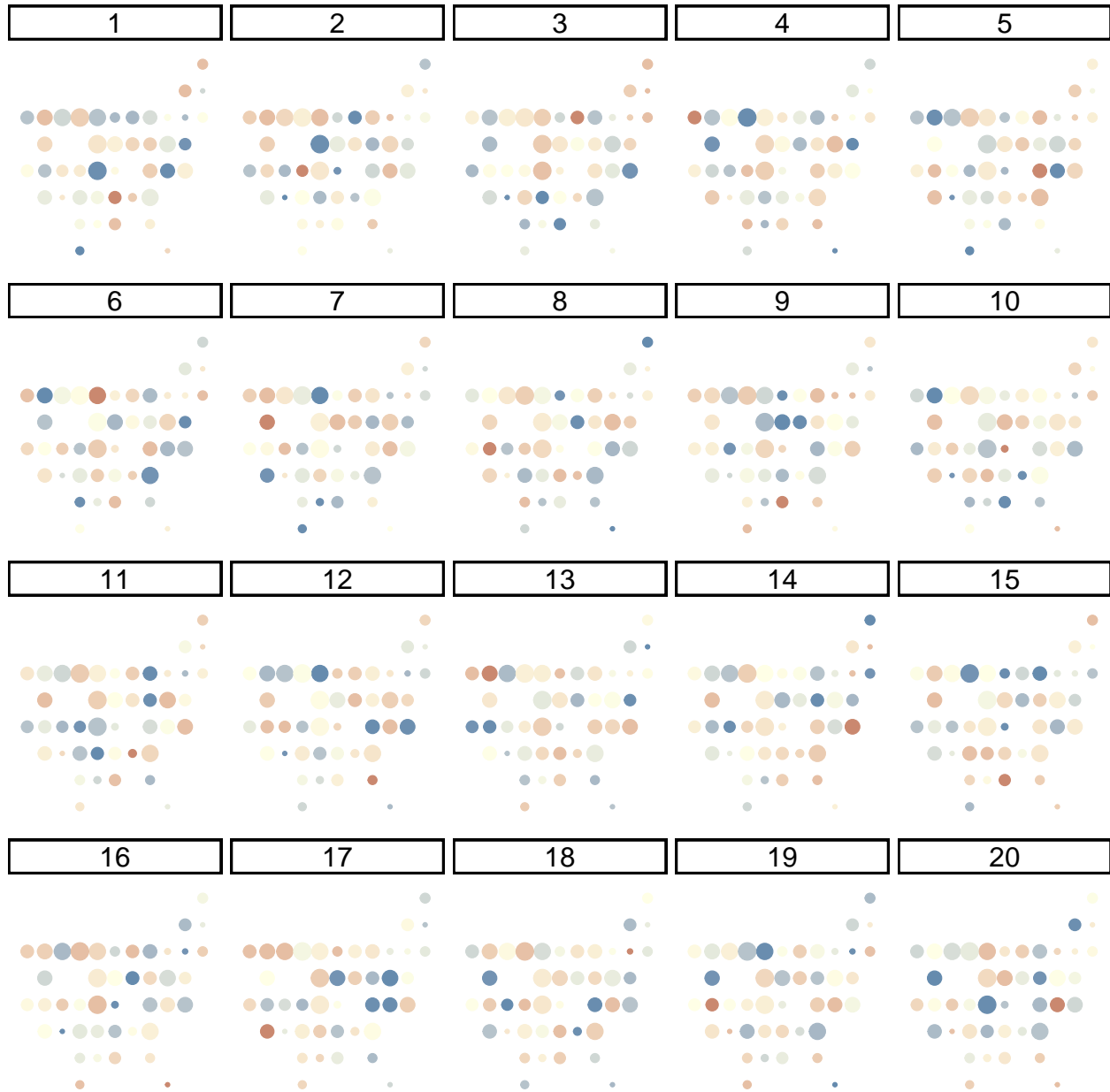
Figure 2 **A tile grid was used in Figure 3 of Basole et al. (2021) to display the cancer screening (by color) and the coverage (by size) with the tile order matching a rough spatial layout of USA. The above figure shows a lineup for this tile grid plot where one of the plots is made using the data, and the other nineteen plots are constructed after first permuting the percentage of cancer screening across different states. (Missing value structure is preserved.) Text and legends have been removed to minimise the bias in reading plots due to the reader being aware of the context. Your task is to identify any plot that is visually different from the others. See the Supplementary Material to see if you have selected the data plot as the most different, and also for the code to produce the lineup.**

scatterplot is a more powerful visual encoding, than the tile grid, for assessing association between two variables. The Supplementary Materials has additional lineups that support this.

4.   Final thoughts

The advancements of computational subsystems, such as `ggplot2` R-package (Wickham 2016), `matplotlib` python library (Hunter 2007) and `D3` javascript (Bostock et al. 2011), have proliferated the production of graphs in a reproducible manner. In particular, we can now generate motion graphs and, by extension, interactive graphs and dashboards very efficiently. This ease of graph production however does not prevent sub-standard visualisation practices.

Sub-standard visualisation practice burgeons as a result of informal treatment of plots, including over-emphasis on particular features of plots for decision-making processes. Considering plots as a statistic, treating it with the a similar rigour used to treat numerical statistics in inferences, can help in weeding out poor practices. We refer readers interested in learning more about visual inference to VanderPlas et al. (2020).

In addition, interactive graphs and dashboards have gained significant popularity in numerous business and e-commerce applications. An interactive visualization enables us to explore the "what-if" scenarios under different assumptions and to uncover unexpected patterns under different conditions. With live data updates, dashboards are often used for visual reports of key performance indicators, enabling interactive monitoring and real-time analysis. With the abundant and complex data available, it is evermore important that visual encoding choices are chosen to communicate the intended message to the audience efficiently, but also understood without misrepresentation. We believe future research benefits from extending existing visual inference foundations to interactive graphics (Cook et al. 2021) and look forward to cultivation of statistical rigour in data visualisation.

Supplementary Materials

All figures in this paper are produced by using the R language (R Core Team 2020) and the `ggplot2` package (Wickham 2016) are supplied at `https://emitanaka.org/supp-visOM/`.

Acknowledgments

References

Basole RC, Bendoly E, Chandrasekaran A, Linderman K (2021) Visualization in operations management research. *INFORMS Journal on Data Science* (to appear).

Bertin J (1983) *Semiology of graphics: diagrams, networks, maps* (ESRI Press), ISBN 978-1589482616, URL `http://dx.doi.org/10.1080/00690805.1987.10438353`.

Bostock M, Ogievetsky V, Heer J (2011) D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17(12):2301–2309, ISSN 1077-2626, URL `http://dx.doi.org/10.1109/TVCG.2011.185`.

6

Author: *Uprooting sub-standard visualisation practices for decision-making in operational management*
Article submitted to *INFORMS Journal on Data Science*; manuscript no. IJDS-0001-1922.65

Buja A, Cook D, Hofmann H, Lawrence M, Lee EK, Swayne DF, Wickham H (2009) Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367(1906):4361–4383, URL http://dx.doi.org/10.1098/rsta.2009.0120.

Christodoulou D (2018) Graph workflow model. https://graphworkflow.com/model/.

Cook D, Reid N, Tanaka E (2021) The foundation is available for thinking about data visualization inferentially. *Harvard Data Science Review* URL http://dx.doi.org/10.1162/99608f92.8453435d, https://hdsr.mitpress.mit.edu/pub/mpdasaqt.

Hofmann H, Follett L, Majumder M, Cook D (2012) Graphical tests for power comparison of competing designs. *IEEE Transactions on Visualization and Computer Graphics* 18(12):2441–2448, URL http://dx.doi.org/10.1109/TVCG.2012.230.

Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9(3):90–95, URL http://dx.doi.org/10.1109/MCSE.2007.55.

Munzner T (2014) *Visualization analysis and design* (CRC press), URL http://dx.doi.org/10.1201/b17511.

R Core Team (2020) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

VanderPlas S, Cook D, Hofmann H (2020) Testing statistical charts: what makes a good graph? *Annual Review of Statistics and Its Application* 7:61–88, URL http://dx.doi.org/10.1146/annurev-statistics-031219-041252.

Wickham H (2014) Tidy data. *Journal of Statistical Software, Articles* 59(10):1–23, ISSN 1548-7660, URL http://dx.doi.org/10.18637/jss.v059.i10.

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York), ISBN 978-3-319-24277-4, URL http://dx.doi.org/10.1007/978-0-387-98141-3.

Wickham H, Cook D, Hofmann H, Buja A (2010) Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics* 16(6):973–979, URL http://dx.doi.org/10.1109/TVCG.2010.161.

Wilkinson L (2005) *The Grammar of Graphics* (Springer), ISBN 9780387245447, URL http://dx.doi.org/10.1007/0-387-28695-0.