Table Dressing

Strategies for improving tabular displays of Monte Carlo simulation results

Matthew Sigal
York University

March 13th, 2017

Introduction

My dissertation research pertains to investigating and implementing methods for improving the presentation of Monte Carlo simulation study results.

Broadly speaking, such approaches can be categorized into three primary areas:

1. Tabular displays
2. Graphical displays
3. Interactive displays

Today's presentation will focus on tabular displays.

Introduction

Primary Goals

- Why Monte Carlo studies?
- Motivating example
- Specific issues with tabular displays
- Table tidiness
- Present recommendations
  - Typographic improvements
  - Structural improvements
  - Display elements
- `R` implementation via `SimDisplay`

Monte Carlo Designs...

Monte Carlo Simulation Studies (MCSS) provide a pivotal foundation for research in Quantitative Psychology, and in applied statistics at large.

What are Monte Carlo Simulation Studies?

MCSS are **experiments** in which certain parameters, which are known and fixed by the researcher, are used to generate random data and then estimate the behavior of other statistics. This is repeated over many *iterations*, and across many *conditions*.

Monte Carlo Designs. . .

For example, MCSS are often used to:

- Evaluate the accuracy of a new statistic
- Investigate the performance of a statistic under various assumption violations
- Determine Power or Type I error rates for various designs
- See how well parameters are recovered under varying conditions
- Simulate 'realistic' data to address hard to study phenomena

See Sigal and Chalmers (2016) to learn about running simulation studies in `R`!

MCSS Results

Conditions for a MCSS can be varied, but generally resolve around aspects like:

- Overall sample size
- Number of groups
- Sample size per group
- Magnitude of effect size
- Degree of within-group variability
- Generating distributions
- Method of estimation

MCSS Results

Further, researchers might be interested in a variety of results from each design cell. For instance, they might keep track of:

- Type I error rates
- Power rates
- Degree of bias/size of the root mean square error

Results necessarily take the form of a **multi-dimensional table**, with dimensions pertaining to the results for an outcome measure for a particular combination of design elements.

Prototypical results

Even results from fairly simple MCSS produce a voluminous amount of output, and it is customary for authors to simply omit conditions from published reports. Even so, this leads to some *very* long tables.

For example, we can look at Brown and Forsythe (1974). This classic simulation pertains to tests for heterogeneity of variance in analysis of variance (ANOVA) designs. It has 3 design variables:

- Generating distribution:
  - Gaussian;
  - χ^2 with 4 *df* (positively skewed);
  - Student *t* with four *df* (symmetric but long-tailed); and,
  - Cauchy (similar to t, but with longer tails)
- Group sizes: 40/40, 20/40, 10/10, 10/20.
- Population variance ratios: 1:1, 1:2, 1:4, 2:1, and 4:1.

Prototypical results

Truncated results from Brown and Forsythe (1974). Note full table does not include any results from the Cauchy distribution.

Empirical Size and Power ($\alpha = 5\%$)

n_1, n_2	$\sigma_1^2:\sigma_2^2$	*F*	*Jackknife*	*Layard* χ^2	*Levene* (W_0)	W_{10}	W_{50}
			A. Gaussian distribution				
40, 40	1:1	6.3	5.8	6.5	6.4	6.1	5.1
	2:1	57.5	54.2	56.3	51.1	50.8	48.4
	4:1	98.2	98.1	98.2	97.1	96.9	96.7
10, 10	1:1	5.4	4.6	6.3	5.5	4.9	2.9
	2:1	16.7	13.9	19.5	15.8	14.7	9.8
	4:1	51.3	42.1	51.1	44.3	41.4	31.7
20, 40	1:1	5.8	5.7	6.2	5.8	5.2	4.5
	2:1	43.4	41.3	37.8	38.8	37.5	32.9
	4:1	92.0	89.4	88.4	88.5	88.0	85.7
	1:2	36.6	38.0	42.9	33.2	32.9	31.1
	1:4	92.0	90.5	93.3	86.3	85.3	83.9
10, 20	1:1	4.8	5.2	6.6	5.7	5.2	4.0
	2:1	24.3	19.7	17.7	21.5	20.4	15.7
	4:1	71.7	63.0	59.8	62.2	59.8	52.7
	1:2	16.1	18.4	24.3	15.8	14.8	12.1
	1:4	57.0	57.4	66.0	49.9	48.6	43.0
			B. Student's t on 4 df				
40, 40	1:1	24.1	6.3	4.8	5.0	4.8	4.4
	2:1		33.7	31.2	36.4	34.9	33.2
	4:1		74.3	79.3	87.2	86.4	85.8
10, 10	1:1	15.9	6.8	8.4	5.9	5.3	3.5
	2:1		13.2		12.2	10.3	7.1
	4:1		31.9		32.1	28.6	22.4
20, 40	1:1	21.9	8.2	6.0	5.0	4.7	4.4
	2:1			19.4	27.5	25.7	22.9
	4:1			60.9	75.9	74.2	71.3

Issues with tabular displays

> “The graphic method has considerable superiority for the exposition of statistical facts over the tabular. A heavy bank of figures is grievously wearisome to the eye, and the popular mind is as incapable of drawing any useful lessons from it as of extracting sunbeams from cucumbers."
>
> Farquhar and Farquhar (1891, p. 55)

Issues with tabular displays

- Results nearly unreadable, except for looking up particular combinations of factors
- Many comparisons get hidden from view, especially for complex simulation designs with many factors
- Patterns generally difficult to discern, especially at a glance

How can this situation be improved?

Tidy Theory

To help conceptualize MCSS results, it is useful to think of them within the *tidy* framework.

- *Tidy Data*: a standard for structuring data, with each variable as a column, each observation as a row, and each type of observational unit a table (Wickham, 2014).
- *Tidy Theory*: a standard for structuring statistical results to make them easily amenable to visualization and summarization (Robinson, 2014), divides components into:
  1. statistical components (e.g., slope/intercepts),
  2. observations (e.g., fitted values and standard errors for particular data points),
  3. model level information (e.g., fit statistics/R^2 values).

Can we place MCSS results within this framework?

Tidy Theory

MCSS tables presented in journals are *messy*.

- Condition variables often are used in both columns *and* rows
- This is done to conserve space, which is at a premium
- However, MCSS **data** should be tidy!
  - Results should not span multiple files
  - Columns should pertain to design factors and output, with appropriate naming conventions
  - Rows should pertain to singular combinations of design factors
  - Often requires collapsing individual replications into an appropriate summarization
  - Thankfully, a properly structured simulation study should do this for you!

Brown & Forsythe, Tidied

	var_ratio	sample_size	groups_equal	distribution	F	Jacknife	Layard	Levene	W10	W50
1	4	80	TRUE	Gaussian	0.990	0.989	0.987	0.979	0.976	0.972
2	2	80	TRUE	Gaussian	0.572	0.533	0.555	0.511	0.508	0.474
3	1	80	TRUE	Gaussian	0.048	0.052	0.055	0.057	0.058	0.049
4	4	20	TRUE	Gaussian	0.504	0.420	0.508	0.429	0.423	0.339
5	2	20	TRUE	Gaussian	0.153	0.135	0.174	0.148	0.137	0.103
6	1	20	TRUE	Gaussian	0.047	0.047	0.069	0.058	0.054	0.035
7	4	60	FALSE	Gaussian	0.935	0.899	0.894	0.889	0.887	0.870
8	2	60	FALSE	Gaussian	0.458	0.408	0.381	0.405	0.402	0.368
9	1	60	FALSE	Gaussian	0.054	0.056	0.065	0.054	0.052	0.045
10	0.5	60	FALSE	Gaussian	0.345	0.362	0.415	0.320	0.314	0.295
11	0.25	60	FALSE	Gaussian	0.924	0.895	0.929	0.867	0.864	0.852
12	4	30	FALSE	Gaussian	0.686	0.602	0.566	0.601	0.579	0.511
13	2	30	FALSE	Gaussian	0.240	0.176	0.189	0.209	0.197	0.160
14	1	30	FALSE	Gaussian	0.048	0.055	0.064	0.053	0.047	0.033
15	0.5	30	FALSE	Gaussian	0.201	0.222	0.288	0.196	0.200	0.164

- Each row pertains to a unique combination of factor variables
- Results given own variable, in this case summarized by the empirical detection rate
- Data structure provides standardized basis for further visualization and exploration

Strategies to encourage recall

To promote recall and understanding, the most memorable visualizations have the following key characteristics (Borkin et al., 2016):

- Should be striking "at-a-glance"
- Have meaningul titles and text
- Can include pictograms
- Should not shy away from some redundancy

While these suggestions skew toward improving infographics, they are useful to keep in mind here.

Typographic improvements

- Numerical output is often not in an optimal human readable format
- e.g., `R` often prints numbers using scientific notation
- This makes comparisons tricky

For example:

Consider the following vector of coefficients from a fitted model:

```
##                     effect
## (Intercept)  1.135000e+02
## A           -1.350000e+01
## B            4.500000e+00
## C            2.450000e+01
## C1           6.927792e-14
## C2          -1.750000e+00
## D            1.650000e+01
```

What is the largest coefficient?

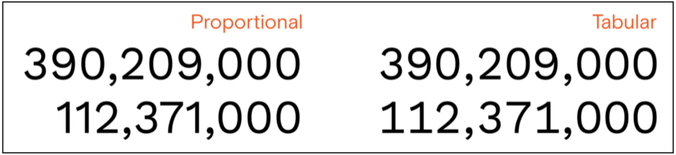
Typographic improvements

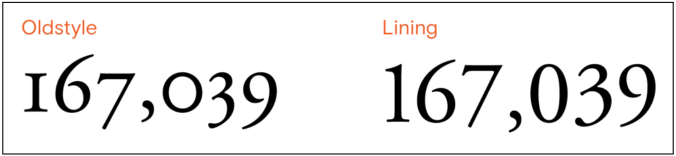
- **Lucid printing** (Wright, 2016)
  - `lucid` converts numeric output to character vectors, with appropriate white-space
  - Maintains column/digit consistency
  - MCSS results are often presented as decimals or percentages, so this can help if there is substantial variation among methods

```
##              effect
## (Intercept) 114
## A           -13.5
## B             4.5
## C            24.5
## C1            0
## C2           -1.75
## D            16.5
```

Typographic improvements

Ström (2016), a graphic designer for `Planetary.io`, wrote a great post about improving typographic elements for data tables.





Tables should always use fonts with a tabular-lining structure!

Typographic improvements

Further, alignment can provide easier comparison.

1. Right align numerical data (lines up digits for easy comparison)
    - Try to use a consistent number of significant digits
2. Left align textual data (read left-to-right, lines up first letter for comparison)
3. Headers should be aligned with their data (provides consistency and context)
4. Don't use center alignment (prevents "raggedness")

For MCSS, we often want to highlight particular entries. In such cases, use emphasis to signify important results (bold, italics)

Structural improvements

- Present rows and columns in a meaningful way
- Use effects ordering and/or rotation (e.g., manipulate rows using clustering algorithms)

Organizing principles for tabular displays

Miller (2007) presented a straight-forward framework for organizing tables for ease of use.

Organizing data to accompany a prose description

1. Empirical ordering (based on value or frequency)
2. Theoretical grouping (conceptually related sets, via panels)
3. Multiple organizing criteria
    - group theoretically then arrange within groups via an empirical consideration
    - arrange alphabetically within conceptual/empirical groupings
4. Supply a narrative to accompany table or chart, clearly indicating organizing principle

Organizing principles for tabular displays

Organizing data for reference use

- Little to no prose description
- Use familiar convention (e.g., alphabetical or ordering from a standard document)

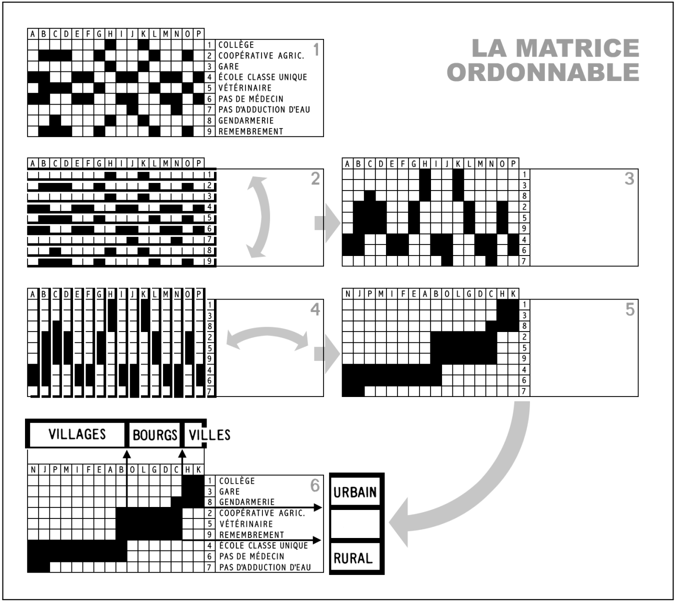
In both cases, the organizing criterion to use depends largely on the type of variables in question. However, a general guideline is to organize your data in the order you will describe them when reporting.

The reorderable matrix

A more complex approach that was popularized in 1967 by Bertin (1981, English translation) is the concept of the *reorderable matrix* (Siirtola and Mäkinen, 2005).

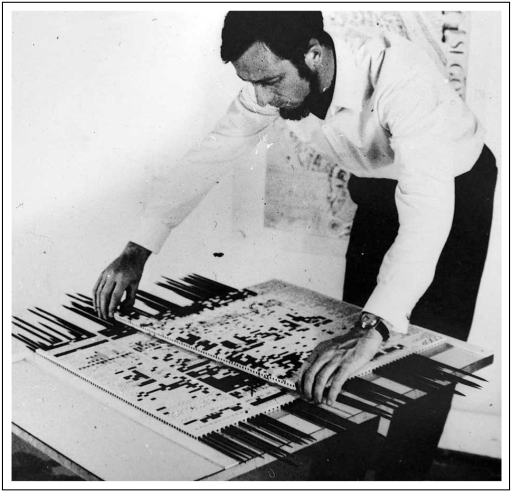
- transforms a multidimensional data set into a two-dimensional (2D) interactive graphic
- mimics table as graphic maintains row and column structure
- **Construction**: data values are replaced with symbols (e.g., circles/rectangles), with size relative to actual data value
- **Reconstruction**: rows and columns are permuted, one at a time, allowing different views of the data set

The reorderable matrix



The reorderable matrix

Bertin's original method relied heavily on visual judgment (Perin et al., 2014), which is very impractical for large datasets.



Serge Bonin with a physical reorderable matrix.

The reorderable matrix

Thankfully, multiple methods are now available to determine row and column placement algorithmically. For instance:

- Cluster analysis/hierarchical clustering
- Principal components analysis
- Multidimensional scaling
- Bond energy algorithm (BEA)

The Bond Energy Algorithm

McCormick et al. (1972) defined a measure of effectiveness (ME) for an *n* by *m* matrix $\mathbf{X} = (x_{ij})$ as:

$$\boxed{M(\mathbf{X}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} [x_{i,j+1} + x_{i,j-1} + x_{i+1,j} + x_{i-1,j}]}$$

This value is maximized if each element is as closely related numerically to its four neighboring elements as possible.

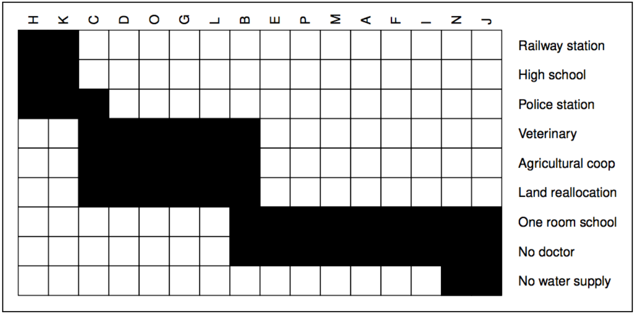
The Bond Energy Algorithm

The BEA uses the ME to reconstruct the given matrix (Hahsler et al., 2008).

1. Place one randomly chosen column.
2. Try to place each remaining column at each possible position left, right, and between the already placed columns.
3. For each placement, calculate the change in ME.
4. Choose the column and position which gives the largest increase in ME and place that column.
5. Repeat until all columns are placed.
6. Repeat for the rows.

The Bond Energy Algorithm

For example, applying the BEA to Bertin's townships data with the `seriate` package in `R` produces (Hahsler et al., 2008):



Display elements

Use colour?

- Map colors for redundant reinforcement (make cells 'pop')
- Colors could also map onto alternate value
  - e.g., table entries for power but shading per Type I error rates
  - This would highlight statistics that may have better power, but at the cost of being more liberal.
- Issues:
  - Use a colorblind-friendly palette: `colorbrewer2.org`
  - Ensure that scale is reasonable

R implementation via SimDisplay

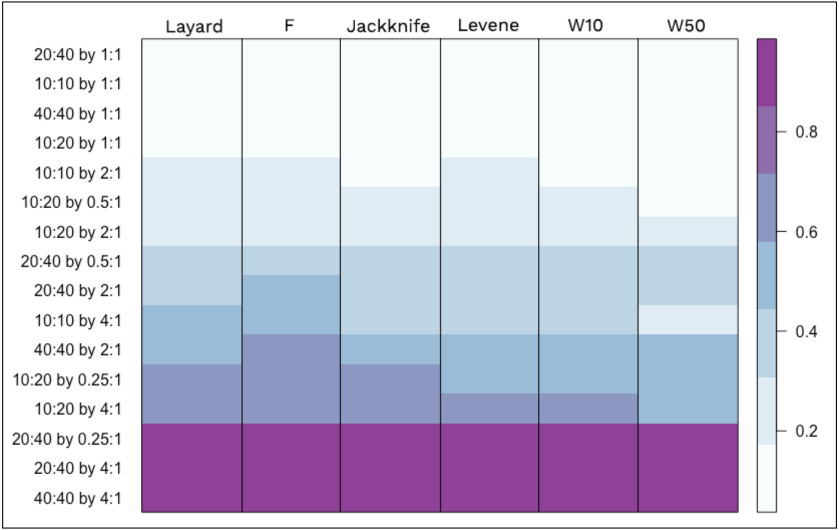
- Work is currently ongoing on package to aid in this endeavour
- Major functions for working with static tables and graphics, as well as provide interactive instances
- Will provide generics functions for:
  - `summary()` for `SimDesign` objects that give useful results, convenient for display, and easy to manipulate
  - `plot()` method(s) – will generate a standard set of multi-panel plots
  - `ftable()`/`xtable()` methods – formatting numerical tables on screen or in LaTeX
  - `tableplot()` methods – to provide semi-graphic display of results
- Is designed to work with `SimDesign` objects, but can also be applied to generic dataframes

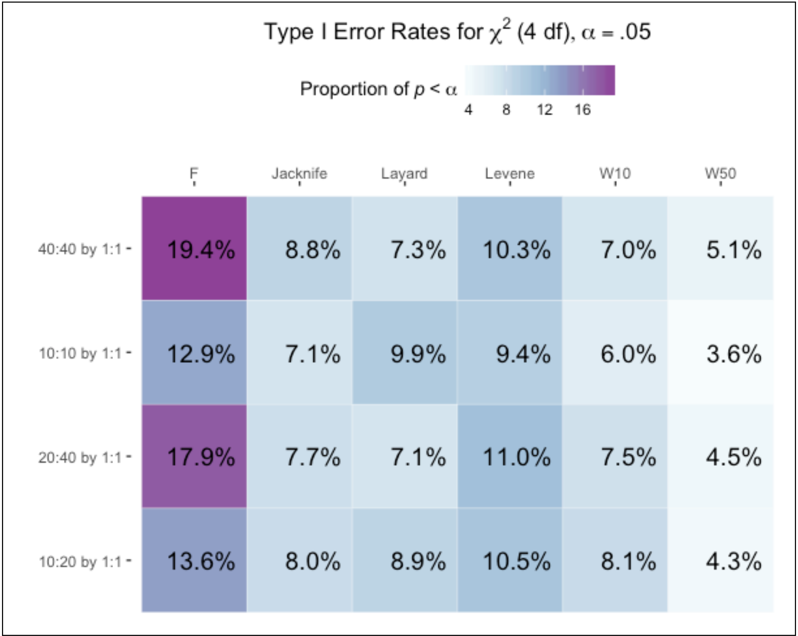
Type I Error Rates and Power for Gaussian, $\alpha = .05$

Proportion of $p < \alpha$: 25, 50, 75

	F	Jacknife	Layard	Levene	W10	W50
40:40 by 1:1	5.8%	5.9%	6.4%	5.8%	5.8%	5.2%
40:40 by 2:1	56.7%	55.3%	56.3%	50.8%	50.9%	49.0%
40:40 by 4:1	98.6%	98.0%	97.9%	97.0%	96.9%	96.4%
10:10 by 1:1	5.0%	5.0%	6.5%	6.5%	5.6%	3.7%
10:10 by 2:1	14.7%	13.1%	17.2%	15.0%	13.3%	9.8%
10:10 by 4:1	48.7%	39.3%	48.6%	39.9%	36.8%	28.1%
20:40 by 1:1	5.2%	5.2%	5.8%	5.5%	5.5%	4.4%
20:40 by 2:1	45.4%	40.5%	39.2%	41.3%	40.2%	35.2%
20:40 by 4:1	94.0%	92.0%	90.0%	88.3%	87.9%	86.5%
20:40 by 0.5:1	34.7%	36.0%	41.5%	32.7%	32.3%	30.1%
20:40 by 0.25:1	92.8%	91.4%	93.3%	87.5%	87.0%	85.5%
10:20 by 1:1	4.7%	6.1%	7.5%	6.3%	6.0%	5.2%
10:20 by 2:1	22.8%	17.1%	16.9%	18.9%	18.0%	14.5%
10:20 by 4:1	69.2%	60.2%	58.3%	61.5%	58.9%	52.0%
10:20 by 0.5:1	17.8%	19.7%	25.4%	16.8%	15.8%	12.5%
10:20 by 0.25:1	60.2%	57.7%	67.6%	49.5%	49.2%	43.5%

Using PCA Clustering





Type I Error Rates for χ2 (4 df), α = .05
Proportion of p < α
4
8
12
16
F
Jacknife
Layard
Levene
W10
W50
40:40 by 1:1
19.4%
8.8%
7.3%
10.3%
7.0%
5.1%
10:10 by 1:1
12.9%
7.1%
9.9%
9.4%
6.0%
3.6%
20:40 by 1:1
17.9%
7.7%
7.1%
11.0%
7.5%
4.5%
10:20 by 1:1
13.6%
8.0%
8.9%
10.5%
8.1%
4.3%

Overall

Static tables can convey substantial amounts of information and should still be recognized as a suitable method for presenting MCSS results, if thoughtfully constructed.

As tables become more complex, the added-value of proper sorting of factor levels, well-chosen color scheme and cut-values, and appropriate typography, over a purely numerical table *increases*.

These ideas and the accompanying `R` package should make it substantially easier to transfer these recommendations into practice.

Thank you!

Slides available at: www.matthewsigal.com/teaching

References I

Bertin, J. (1981). *Graphics and Graphic Information Processing*. Gruyter & Co., Berlin.

Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yeh, C. S., Borkin, D., Pfister, H., and Oliva, A. (2016). Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):519–528.

Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367.

Farquhar, A. B. and Farquhar, H. (1891). *Economic and industrial delusions*. G. P. Putnam's Sons, New York, NY.

Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting things in order: An Introduction to the R package seriation. *Journal of Statistical Software*, 25(3).

McCormick, W. T., Schweitzer, P. J., and White, T. W. (1972). Problem decomposition and data reorganization by a clustering technique. *Operations Research*, 20(5):993–1009.

Miller, J. E. (2007). Organizing data in tables and charts: Different criteria for different tasks. *Teaching Statistics*, 29(3):98–101.

References II

Perin, C., Dragicevic, P., and Fekete, J.-D. (2014). Revisiting Bertin matrices. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2082–2091.

Robinson, D. (2014). broom: An R package for converting statistical analysis objects into tidy data frames. *arXiv preprint*.

Sigal, M. J. and Chalmers, R. P. (2016). Play it again: Teaching statistics with Monte Carlo simulation. *Journal of Statistics Education*, 24(3):136–156.

Siirtola, H. and Mäkinen, E. (2005). Constructing and reconstructing the reorderable matrix. *Information Visualization*, 4(1):32–48.

Ström, M. (2016). Design better data tables. https://planetary.io/blog/design-better-data-tables.

Wickham, H. (2014). *Advanced R*. Chapman and Hall/CRC, Boca Raton, FL.

Wright, K. (2016). *lucid: Printing Floating Point Numbers in a Human-Friendly Format*. R package version 1.4.