29th ANNUAL AUSTRALIAN POULTRY SCIENCE SYMPOSIUM

SYDNEY, NEW SOUTH WALES

4TH -7TH FEBRUARY 2018

Organised by

THE POULTRY RESEARCH FOUNDATION
(University of Sydney)

and

THE WORLD'S POULTRY SCIENCE ASSOCIATION
(Australian Branch)

ENHANCING RESEARCH COMMUNICATION THROUGH INFORMATION DESIGN AND VISUAL STORYTELLING: REFLECTIONS ON 10 YEARS OF APSS PROCEEDINGS FIGURES

M. KRZYWINSKI[1]

Summary

Figures have the potential to illustrate complex concepts and patterns that would otherwise be difficult to express concisely in words or notice quickly from a table. Figures that are clear, concise and attractive help to form a strong connection with the audience, communicate with immediacy and accelerate understanding and progress. This can be achieved by employing principles of graphic design, which are based on our understanding of how we perceive, interpret and organize visual information. This article presents practical guidelines for visually communicating data and experimental protocols to assist poultry researchers in presenting their research to colleagues and those outside of the field. Several redesign examples using figures taken from past APSS Proceedings are presented to concretely illustrate the application of information design concepts such as data encoding, use of color, visual conventions and metaphors. The examples demonstrate how to more clearly show essential aspects of the data, combine different modalities and correlate patterns across observations and treatments. To strengthen reach to industry, the public and policy makers, an unadorned and field-agnostic visual style is used to improve consistency and clarity of how information is presented.

I. INTRODUCTION

All of us are schooled in 'written design' (grammar) and most have had some experience with 'verbal design' (public speaking). However, relatively few have had training in 'visual design' (information design and visualization). Thus, when we need to present complex information visually, we may find ourselves at a 'loss for words', graphically speaking.

Just as text must be grammatically and semantically sound, figures should embody equivalent visual qualities. These qualities can be realized by using principles of graphic and information design that are based on our understanding of how we perceive, interpret and organize visual information. This information is underpinned by conclusions from studies in visual perception and awareness (Cleveland and McGill, 1984, 1985; Yantis, 2005; Fecteau and Munoz, 2006). In addition to informed choice of data encoding (e.g. bar, line, scatter plot, etc.), the use of color, placement of labels and annotation, visual weight of navigation elements and flow of information across panels all influence how well the figure can communicate concepts, proportions and patterns. These elements must be in balance to match salience to relevance - drawing the eye to important patterns and subtly discouraging irrelevant or misleading interpretations that can arise when it is not clear where to look (Wong, 2011).

To make the design advice presented here practical, I have applied it to examples that are representative of the current poultry research literature. I have reviewed and tabulated all figures in the past 10 years (2008–2017) of the Proceedings of the Annual Australian Poultry Science Symposium (APSS), categorizing them by type, use of color, use of 3-D and texture (Figure 1). I counted 481 distinct figure panels, of which the majority (366, 76%) grouped into three basic types: 162 (34%) bar charts, 118 (25%) line charts and 86 (18%) scatter plots (Figure 1A). More than 50% of all figures were black-and-white and 34% of the top three

[1] Canada's Michael Smith Genome Sciences Center, Vancouver, Canada; martink@bcgsc.ca

types used one or more colors (Figure 1B). 23% of bar charts used textures and 51% of line plots used dotted, dashed, or otherwise stippled lines to distinguish traces (Figure 1C).

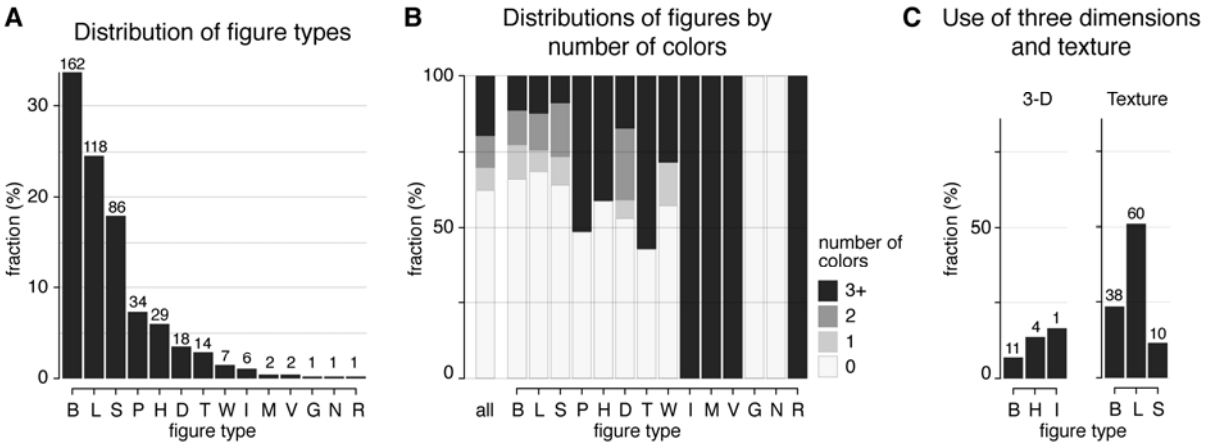

Figure 1 - The distribution of plot types and format properties of 481 figures that appeared in 10 years of Proceedings of the APSS during 2008–2017. Values above bars show the absolute count. (A) Fraction of each figure type. (B) Breakdown of all figures (first column) and figures by type (remaining columns) by the number of colors used (0 indicates black-and-white or greyscale). (C) The fraction of figures that appeared in three-dimensions and employed the use of textured patterns. Texture for line and scatter plots indicates use of dashed, dotted, or otherwise stippled lines. Plot type codes are: B (bar), L (line), S (scatter), P (photo), H (heat map), D (diagram), T (ternary), W (box plot), I (pie chart), M (geographical map), V (Venn diagram), G (gel), N (network layout), R (rose plot).

II. BAR CHARTS

The bar chart is the most common figure type in APSS Proceedings. This example will focus on presenting raw data together with output of statistical testing, selecting grey levels and managing labels in the plot and in the legend.

The original figure (Figure 2A) (Hughes, et al., 2016) shows the effect of a feed enzyme additive on the apparent metabolisable energy (AME) of wheat. In crowded charts (this one has 58 bars), white bars can be difficult to distinguish from the space between bars—an effect described by the figure-ground Gestalt principle (Wong, 2010a, b). Bars with a solid fill stand apart from the background more effectively (Figure 2B). Angled axis labels make it difficult to include labels for all 29 wheats in the original figure - only 15/29 labels are shown and only for 5/10 wheats with a significant effect. Angled labels also match poorly with their bars because they extend beyond neighbouring bars. If labels are long (e.g. complex experimental conditions) but need to be easily read, a horizontally formatted figure is better. When labels are short sample IDs, as here, a vertical label orientation is a good compromise between legibility and compactness. The axis tick marks appear between bar groups in Figure 2A, which is inconsistent with how ticks work in other data encodings. The role of the tick mark is as a callout line to its label and not a group separator - groups can be separated by space.

To visually assist hypothesis testing, figures should include the output of statistical tests, such as significance. The observation that 7/10 wheats with untreated AME < 13 showed an enzyme effect would be made more salient if the '*' significance symbol were placed closer to the untreated AME bar. However, this is difficult to achieve when bars are so close together. The significance stars are more optimally placed near the wheat labels (Figure 2B). This has the effect of easily connecting the treatment effect to the label as well as facilitating counting the stars (because they are now aligned) for a given AME cutoff.

Although reporting statistical significance using a *P* value cutoff (so-called bright line testing) is widely used, the *P* value alone does not tell the full story (Krzywinski and Altman, 2014b; Altman and Krzywinski, 2016, 2017). Ideally, both the effect size and 95% confidence interval (CI) should be reported (Krzywinski and Altman, 2013a) and both have been added for each wheat in Figure 2B. The effect size relates to the magnitude of the change that is observed, which may be statistically significant but not necessarily biologically relevant. The CI tells us the range of AME changes that are statistically compatible (do not test as significant) with the observed change. Thus, when the 95% CI includes zero, we conclude that no effect is present. Showing CIs enhances our interpretation of significance - we can now tell how close we are to the bright line cutoff (impossible from a statement like '$P < 0.05$') and whether increasing sample size might make the observation significant. Note the two wheats in Figure 2B that are annotated with an arrow - the left shows an effect (just barely) and the right does not (again, just barely). If the experiment was repeated, we would have good reason to expect that the significance status of these wheats changes. Always keep in mind that, because samples are variable (Krzywinski and Altman, 2013b), so are their 95% CIs.

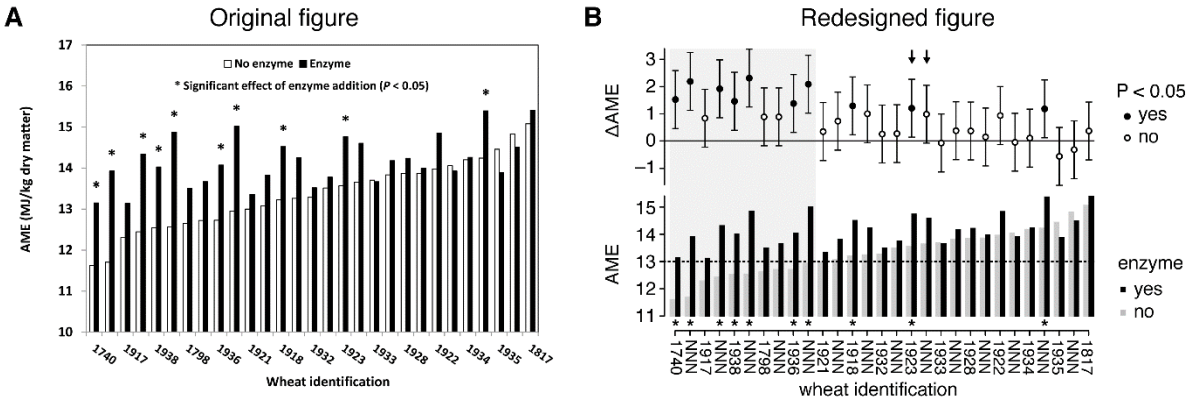

Figure 2 - A grouped bar chart can show the effect of a treatment but a large number of observations can reduce legibility of values and labels. (A) Original figure from (Hughes, *et al.*, 2016). (B) The redesigned figure. Statistical significance is communicated both with the conventional '*', conveniently next to the wheat label, as well as a scatter plot of AME change and 95% confidence interval (CI) (arrows point to one pair of near-identical cases of which only the left is significant, barely). Focus on wheats with AME < 13 is achieved with a horizontal dashed line and grey highlight. The label 'NNNN' is used for wheat whose identifier was not available in (A). The 95% CIs drawn are visual placeholders for actual intervals that would be drawn—ones shown are estimated and made all identical.

The panels in Figure 2B share the horizontal axis and both vertical axes for AME and ΔAME have the same scale to facilitate comparison of vertical distances between panels. Vertical axes are clipped to the range of the data to allocate more of the figure's space to where data are changing. In the original figure the AME axis range of 10–17 is too generous - there are no values for AME < 11 or AME > 16. Where possible, the legend should not encroach on space that would be better used by data and should be set as a table with as little duplication in text as possible. There is also usually no need for a top and right axis, which unnecessarily contain the plot and add clutter.

From the scatter plot of ΔAME in Figure 2B, we can glean that all the significant observations have ΔAME > 1, a reflection of the power of the experiment. Power should always be reported - when it is low, only large effects can be detected, and negative results (such as the sample called out with the right arrow in Figure 2B) cannot be reliably interpreted (Krzywinski and Altman, 2013c). Since a pair of bars showing untreated (AME–) and treated (AME+) AME can just as easily be shown as a scatter plot, I've adopted this approach in Figure 3A. I have also adopted the notation in which untreated and treated

conditions are denoted as a suffix (– vs +). Note that the treatment suffix is not a dash '-' but an n-dash '–', which is longer and visually more compatible with '+'. The n-dash is the appropriate character for intervals (e.g., 10–17) and negative quantities (e.g., $x = –2$).

The scatter plot of ΔAME vs AME– clearly shows that all differences ΔAME > 1 are significant and makes any outliers evident. It is more useful than the plot of AME+ vs AME– , since the experiment is more cogently explained in terms of AME– and ΔAME because it is the change in AME, not its final value, that is more relevant.

One essential observation from the experiment is the extent to which untreated AME affects the efficacy of enzyme treatment. The point that 7/10 wheats with AME < 13 showed an effect (Figure 2B) has already been made anecdotally. By fitting the fraction of wheat in a range of AME values (e.g. 11–12, 12–13, and so on) as a function of AME– we can quickly see that the higher the AME–, the less likely we are to observe an enzyme effect (Figure 3B). Furthermore, from the slope of the fitted line, we can say that each unit of AME– increase in the range 11–16 leads to an absolute decrease in the fraction of wheat with an effect by 25%.

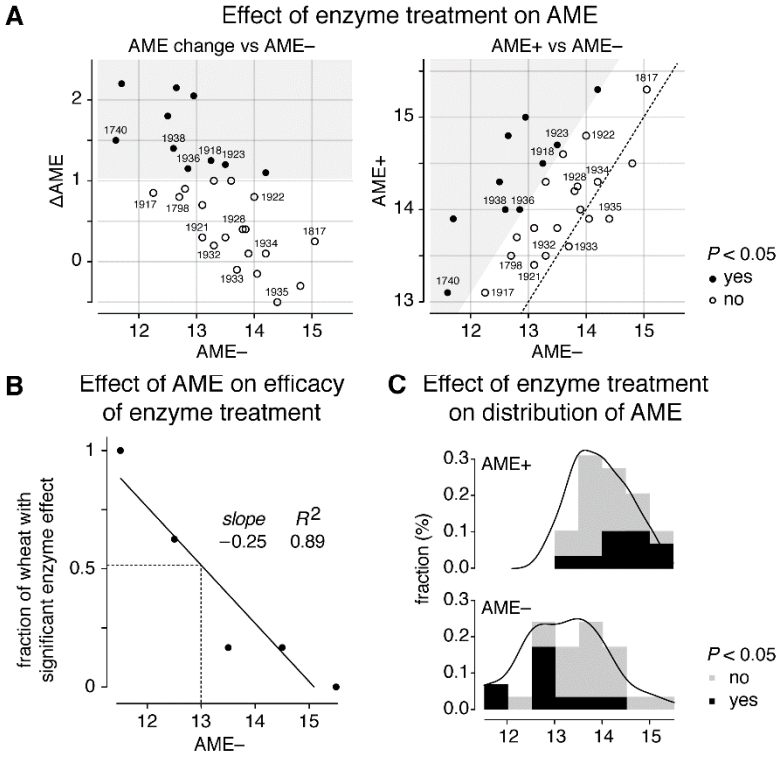

Figure 3 - Alternative ways to a bar plot for displaying a 2-level single factor experiment. (A) The change in AME in the presence of enzyme treatment (ΔAME, left) and treated AME (AME+, right) shown as a function of untreated AME (AME–). The first clearly shows the ΔAME > 1 significance cutoff, which is harder to identify quantitatively in the second panel because it is a vertical distance between two oblique lines. Point labels are wheat identifiers. (B) A fit to the fundamental relationship addressed by the experiment: the impact of enzyme efficacy on AME increase as a function of AME–. Only the slope and R^2 are reported for the line, since the intercept has no meaningful interpretation. (C) Distribution of AME– and AME+ is useful to understand the bounds of the measurement.

Finally, let's use a frequency plot to explore the distribution of untreated and treated AME values (Figure 3C). Although these distributions do not distinguish how AME changed for any specific wheat, they clearly indicate that AME reaches a maximum of about 15, regardless of treatment, and that treated AME has a smaller spread than untreated AME, presumably because the enzyme treatment is more effective for smaller AME values.

Frequency plots and box plots (Krzywinski and Altman, 2014a) powerfully communicate features of populations and samples. Unfortunately, both are underrepresented in the APSS Proceedings. Although they may be at first hand unfamiliar, they are simple to grasp and have great utility. Figure 1 in (Greenhalgh, et al., 2017) shows the variation in feed conversion ratio (FCR) of 140 birds using jittered points (Figure 4A), making it difficult to judge the spread and extent of outliers. These data are better shown as a histogram of bird count for each FCR interval (e.g. 0.1), which makes it is easier to see both the spread and the top and bottom 15% of birds, which were monitored in the experiment (Figure 4B). Importantly, the histogram visually justifies the 15% cutoff for classification as low- and high-efficiency, something that can't be identified from the original in Figure 4A.

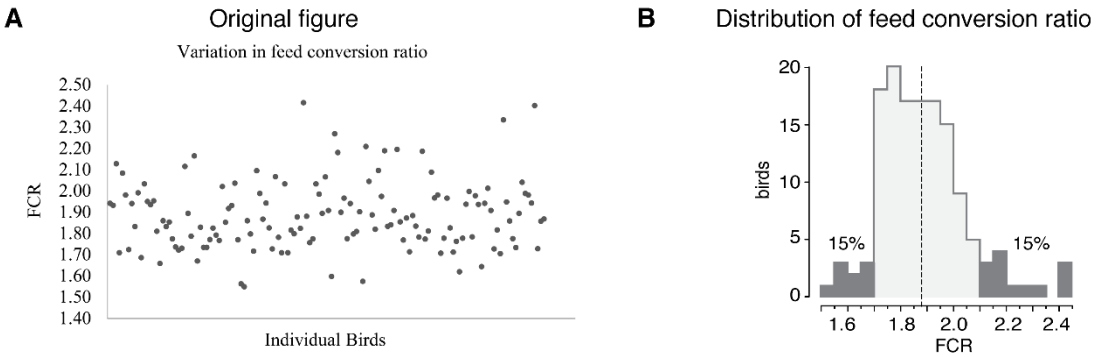

Figure 4 - A histogram clearly communicates the location and spread of a distribution. (A) Original figure from (Greenhalgh, *et al.*, 2017). (B) The same data shown as a histogram with bins of 0.1. The distribution average is shown as a vertical dashed line (1.88) and the bottom and top 15% percentile highlighted.

III. LINE PLOTS

This example is of an experiment studying the effect of 4 levels (1, 2, 3 and 4%) of Ca supplement in the diet. The original figure (Figure 5A) is taken from (Bradbury, et al., 2014) and shows the supplement's effect on daily feed and weekly separate source Ca intake.

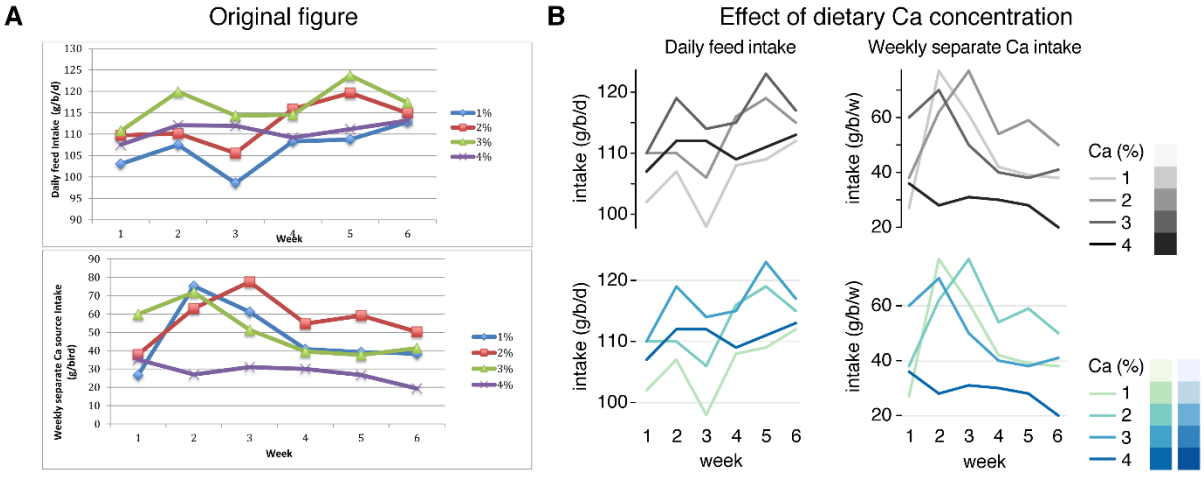

Figure 5 - Effect of Ca supplement level (1–4%) on daily feed intake and weekly separate Ca intake. (A) Original figure from (Bradbury, *et al.*, 2014). (B) Data shown in black-and-white (top row) or color (bottom row) with simplified formatting to emphasize trends in the line profiles. Colors 2–5 from the 5-color grey and blue-green Brewer palettes are used for the lines. Also shown is the 5-color blues Brewer palette (bottom row) as a perceptually quantitative option with constant hue.

The line plot - the second most common APSS Proceedings figure type - is a good choice here. The line emphasizes trends, especially when the number of observations is low

(6 points per trace). The choice of color (blue, red, green, purple), however, is inappropriate for the factor - these colors are more suitable for a categorical variable (one for which there is no inherent order). The Ca level is a quantitative variable and the colors chosen do not reflect this. Purple does not communicate that its level is 4× the level encoded by blue.

The figure can be reformatted to black-and-white (Figure 5B, top panels) without loss of clarity. The grey scales for each trace can be selected from the grey sequential Brewer palette (Harrower and Brewer, 2003), which has steps of grey that are approximately perceptually uniform (the difference in perceived brightness between adjacent tones is similar). Alternatively, a colored sequential Brewer palette (e.g. blue-green) can be used (Figure 5B, bottom). Using color makes comparing the plots easier—hue visually groups objects more powerfully than grey tone by the Gestalt principle of grouping (Wong, 2010a, b).

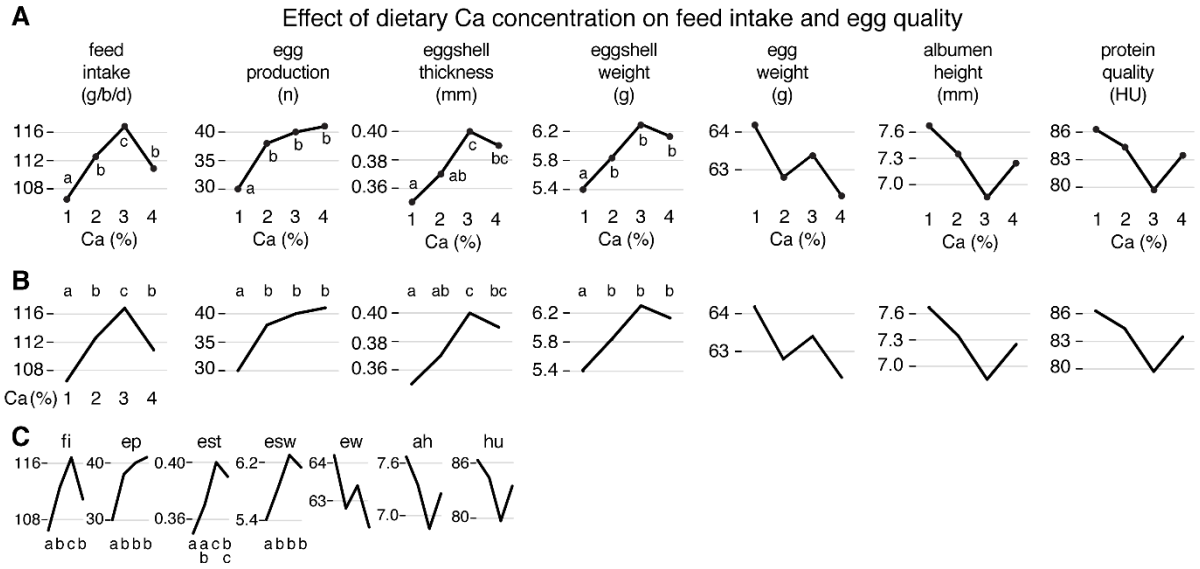

Figure 6 - Change in feed intake and egg quality as a function of Ca supplement level, shown at various level of design detail. (A) Line plots shown with data point glyphs and post-hoc pair-wise significance labels next to the points. The horizontal axis is labeled for each plot. (B) Data point glyphs have been removed and only the first plot has a horizontal axis label. Significance labels aligned on top of the panels to maintain emphasis on data. (C) A tight formatting of the plots, with fewer grid lines and plot titles shown as variable acronyms—a practice that can save space but requires explanation.

Not every plot element in Figure 5B is needed. The horizontal axis for the week number is not necessary—axis lines for categorical, ordinal or otherwise discrete variables can mislead the reader that intermediate values are possible. For example, in Figure 2B the *x*-axis line can be removed—the bottoms of the bars provide a sufficient visual anchor. In Figure 5B the vertical axis can be removed—now any line with a vertical component represents data.

Figures with subtle formatting are ideal when there are a lot of data or insufficient space (or both). Navigational elements (axes, grids, ticks) should have less visual weight than data to keep the data-to-ink ratio high (Krzywinski, 2013a). For example, a 20% opacity is recommended for grid lines (Stone and Bartram, 2009). Sometimes removing elements from a plot altogether helps add salience to data patterns. If you're not familiar with this practice, you may be initially uncomfortable with a plot missing elements that are traditionally present. However, many of these elements are actually superfluous (e.g. plot or legend border, frequent ticks or grids) and should be included only if they enhance your ability to understand the data, not because they are the default settings in your plotting.

Figure 5B shows only part of the picture. The experiment explored the effect of Ca supplement on 7 variables, all of which are tabulated (Table 1 in (Hughes, *et al.*, 2016)). Although tables are excellent for looking up precise values, they do not quickly reveal trends. If we plot the change in the 7 variables as function of Ca supplement level (Figure 6A), we can see that trends in the variables fall into groups. For example, feed intake, eggshell thickness and eggshell weight all from 1-3% then decrease. Similarly, albumen height and protein quality have almost the same profile shape, not surprising since the former informs the latter.

Not all trends in Figure 6A, however, are significant. Including letter subscripts to indicate significant pair-wise comparison (post-hoc Tukey test) helps focus on potentially meaningful changes. The letters should be carefully placed and aligned to avoid disorder and the inevitable confusion that follows. Letters chosen are arbitrary, though traditionally from the beginning of the alphabet in order of appearance. This is not optimal however, since some letters have similar shapes, such as 'a', 'c' and 'e'. It is better to select letters that have different shapes, such as 'h', 'q', 'x' (Krzywinski and Wong, 2013), especially for large plots in which many letters are used, such as Figure 4 from (Dersjant-Li and Kwakernaak, 2017) that uses 8 letters a–h with both positioning and color inconsistently applied.

When dealing with a large number of variable profiles, simplifying the figure by removing elements adds focus on the data (Figure 6B, C). For example, moving the pair-wise significance letters away from data points helps with comparison (because the letters are now aligned) and relieves clutter in the data panel. Other strategies to avoid crowding and emphasizing data and not formatting (Krzywinski, 2013b) include cropping the vertical scale to allow as much room for data as possible (McInerny and Krzywinski, 2015) and limiting the use of large glyphs, drop shadows or gradients (Figure 4A has all three).

IV. EXPERIMENTAL PROTOCOLS

The final example visualizes both the experimental protocol and data simultaneously. The experiment depicted in the original table and figure (Figure 7) explores the effect of two differently changing lighting schedules on egg laying times and rate (Cronin, et al., 2010).

A Original table

Table 1. Clock times (h) when the room lights were switched on and off each day over progressive weeks of the experiment, and the median times of egg laying per week for birds in the Gradual and Abrupt treatments.

	Gradual treatment				Abrupt treatment				Median egg laying times		
Hen age (wks)	On	Off	On	Off	On	Off	On	Off	Gradual	Abrupt	SED
17	0600	1900			0600	1900			0740	0740	0.553
18†	0600	1900	0230	0300	0600	1930			0620	0730	0.601
19	0600	1900	0200	0300	0600	2000			0545 [a]	0740 [b]	0.473
20	0600	1900	0130	0300	0600	2030			0630 [a]	0730 [b]	0.316
21	0600	1900	0100	0300	0600	2100			0540 [a]	0720 [b]	0.298
22	0600	1900	0030	0300	0600	2130			0650 [a]	0830 [b]	0.511
23‡	0600	1900	0000	0300	0600	1900	0300	0600	0620 [a]	0805 [b]	0.620
24	0600	1900	0000	0300	0600	1900	0300	0600	0650	0725	0.659
25	0600	1900	0000	0300	0600	1900	0300	0600	0720	0715	0.554
26	0600	1900	0000	0300	0600	1900	0300	0600	0730	0735	0.625

Within hen age pairs, values with different superscripts differ a,b: P<0.05
SED: standard error of difference between the means
Start of Gradual and Abrupt Light Introduction Treatments indicated by † and ‡, respectively

B Original figure

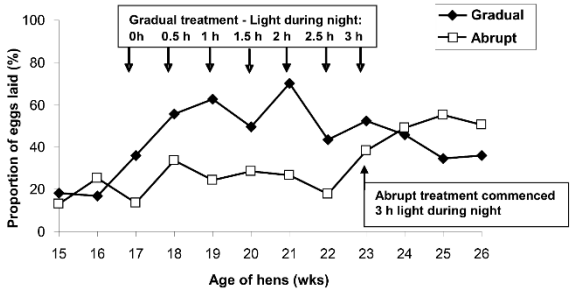

Figure 7 - The effect of two light-dark protocols on egg laying times and rate. From (Cronin, *et al.*, 2010) (A) The light schedule and median egg laying times for gradual (interrupted night) and abrupt (longer day) treatments. (B) The effect of the treatments in (A) on egg laying rate.

The light schedule table (Figure 7A) is easy to understand but requires time to parse. Many values are actually repeated and these would be better shown as a dot to indicate repetition of value in the row above. For example, it is not immediately clear that the amount of additional light each week is the same (30 minutes) for both protocols. This fact is more easily understood from the text and, at that point, the table is no longer required. Similarly, trends in the egg laying times are also difficult to spot.

It is reasonable to try to use a clock to explain the two lighting schedules (Figure 8). This format is familiar and it is easy to spot the difference between the schedules: interrupted night (gradual) and longer day (abrupt). However, it takes longer to identify the weekly constant 30 minute light increase.

For all its familiarity, the real trouble with Figure 8 is that it is very difficult to overlay egg laying times and rates in a way that is simple and quantitative—there is no single axis on which egg laying time can fall. Placing these times on the clocks would make identifying trends across the clocks difficult. Using the clock metaphor has helped one aspect of communication but hindered another. Although we could draw the observations in a separate figure, as done in the original publication (Figure 7B), let's see how both the schedule and observations can be combined the same panel—I present two options.

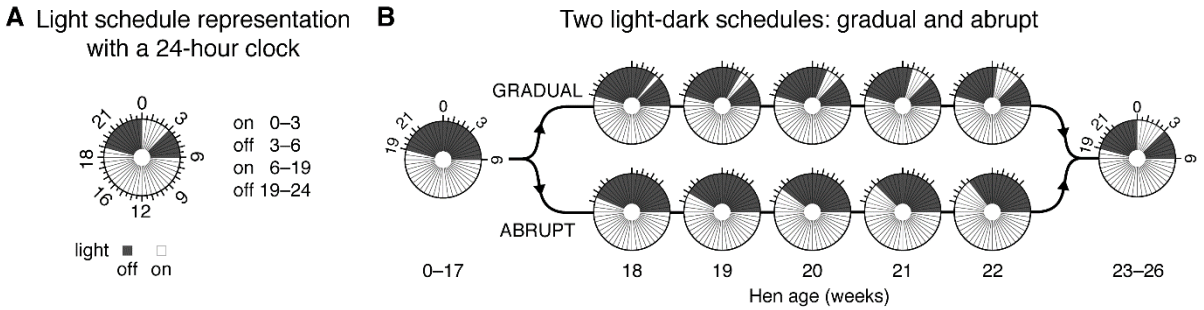

Figure 8 - Representing light schedules using a 24-clock. (A) The face of a 24-hour clock is colored based on light (white) or dark (black) conditions. Here light is applied between midnight–3am (0–3) and between 6am–7pm (6–19). (B) Two light-dark schedules for transitioning hens from a full night of darkness to 3 hours of light. The gradual schedule acclimatizes the hens to 3 hours of night light by subjecting them to increasingly more light each week. The abrupt schedule extends light during the day by 30 minutes each week but does not change night light conditions. For simplicity, only times at which light schedule is varying are labeled on the clock.

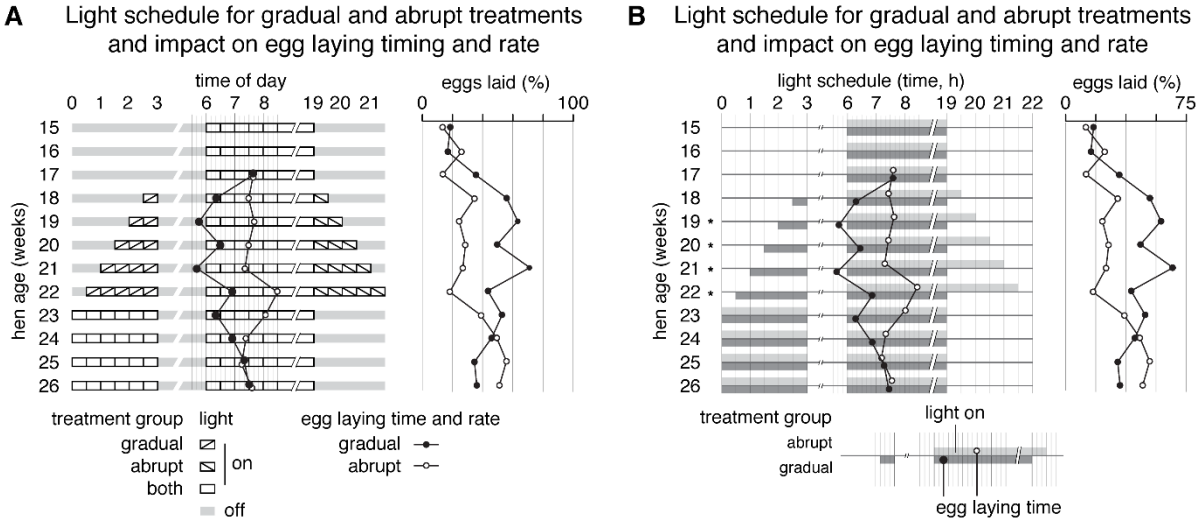

Figure 9 - Two alternatives for combining light-dark schedules and experiment results in a single panel. (A) Light/dark metaphor informs the color of boxes that indicate absence and presence of light. (B) Ink is used to show when light is on and darkness is inferred from where bars that indicate light are absent.

In the first option (Figure 9A) adopts the dark/light metaphor whereby white boxes are used to indicate light condition and grey strip as dark condition. Light during the gradual and abrupt treatments is distinguished by the direction of the line through the box. The day line acts as a natural time axis for the time laying times. By overlaying experimental

observations on top of light schedule, we can correlate patterns better. For example, we can quickly see something that is very hard to spot from the table in Figure 7A: there are two weeks at which hens are laying eggs in the dark (week 19 and 21 in the gradual treatment). We can also see that the highest egg laying rate happens during precisely these two weeks.

In the alternative presentation (Figure 9B), the light/dark metaphor is not used. Instead, horizontal bars show the duration of light for each group. Using ink in the figure to encode when light is present and no ink for when it's absent connects more closely to the theme of the experiment. Note that the legend of Figure 9B is formatted similarly to the figure itself—a practice that helps explain the position and identity of all the elements in a complex display.

V. PARTING THOUGHTS

An experiment that took months (or years) to complete deserves a figure that took several hours (or days) to prepare. Begin designing a figure by identifying the core message, which should be salient so that it can be grasped quickly. Avoid distracting and unnecessary elements to ensure that the visual communication is immediate. Default software settings often pollute output with unnecessary garnish—once done with a figure, check for elements that may be removed without impacting the message. Follow best practices (Tufte, 1992) and resist the urge to depart from effective visual paradigms unless your data set absolutely requires it.

REFERENCES

Altman N & Krzywinski M (2017) *Nature Methods* **14:** 213.
Altman N & Krzywinski M (2016) *Nature Methods* **14:** 3.
Bradbury EJ, Wilkinson SJ, Cronin GM, Bedford MR & Cowieson AJ (2014) *Proceedings of the Australian Poultry Science Symposium* **25:** 139-142.
Cleveland WS & McGill R (1985) *Science* **229:** 828-833.
Cleveland WS & McGill R (1984) *Journal of the American Statistical Association* **79:** 531-554.
Cronin GM, Borg SS, Storey TH, Downing JA & Barnett JL (2010) *Proceedings of the Australian Poultry Science Symposium* **21:** 130-133.
Dersjant-Li Y & Kwakernaak C (2017) *Proceedings of the Australian Poultry Science Symposium* **28:** 135-138.
Fecteau JH & Munoz DP (2006) *Trends in Cognitive Sciences* **10:** 382-390.
Greenhalgh S, Akter Y, Nolan B & O'Shea CJ (2017) *Proceedings of the Australian Poultry Science Symposium* **28:** 97-100.
Harrower M & Brewer CA (2003) *The Cartographic Journal* **40:** 27-37.
Hughes RJ, Black JL, Flinn PC, Tredrea AM & Diffey S (2016) *Proceedings of the Australian Poultry Science Symposium* **27:** 235-238.
Krzywinski M (2013a) *Nature Methods* **10:** 183.
Krzywinski M (2013b) *Nature Methods* **10:** 371.
Krzywinski M & Altman N (2013a) *Nature Methods* **10:** 921-922.
Krzywinski M & Altman N (2013b) *Nature Methods* **10:** 809-810.
Krzywinski M & Altman N (2013c) *Nature Methods* **10:** 1139-1140.
Krzywinski M & Altman N (2014a) *Nature Methods* **11:** 119-120.
Krzywinski M & Altman N (2014b) *Nature Methods* **11:** 597-598.

Krzywinski M & Wong B (2013) *Nature Methods* **10:** 451.
McInerny G & Krzywinski M (2015) *Nature Methods* **12:** 591.
Stone M & Bartram L (2009) *Proceedings of the 16th Color and Imaging Conference Final Program and Proceedings* **5:** 355-359.
Tufte E (1992) *The Visual Display of Quantitative Informationm,* Graphics Press USA, Cheshire, CT, USA.
Wong B (2010a) *Nature Methods* **7:** 863.
Wong B (2010b) *Nature Methods* **7:** 941.
Wong B (2011) *Nature Methods* **8:** 889.
Yantis S (2005) *Nature Neuroscience* **8:** 975-977.