

Graphs in phylogenetic comparative analysis: Anscombe's quartet revisited

Liam J. Revell^{1,2}  | Klaus Schliep¹  | Eugenio Valderrama³  | James E. Richardson^{2,4}

¹Department of Biology, University of Massachusetts Boston, Boston, Massachusetts

²Programa de Biología, Universidad del Rosario, Bogotá, Colombia

³Facultad de Psicología and Programa de Biología, Universidad el Bosque, Bogotá, Colombia

⁴Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, Scotland

Correspondence

Liam J. Revell

Email: liam.revell@phytools.org

Funding information

United States National Science Foundation, Grant/Award Number: DEB 1350474 and DBI 1759940

Handling Editor: Simon Blomberg

Abstract

1. In 1973, the statistician Francis Anscombe used a clever set of bivariate datasets (now known as Anscombe's quartet) to illustrate the importance of graphing data as a component of statistical analyses. In his example, each of the four datasets yielded identical regression coefficients and model fits, and yet when visualized revealed strikingly different patterns of covariation between x and y .
2. Phylogenetic comparative methods (the set of methodologies that use phylogenies, often combined with phenotypic trait data, to make inferences about evolution) are statistical methods too; yet visualizing the data and phylogeny in a sensible way that would permit us to detect unexpected patterns or unanticipated deviations from model assumptions is not a routine component of phylogenetic comparative analyses.
3. Here, we use a quartet of phylogenetic datasets to illustrate that the same estimated parameters and model fits can be obtained from data that were generated using markedly different procedures—including pure Brownian motion evolution and randomly selected data uncorrelated with the tree. Just as in the case of Anscombe's quartet, when graphed the differences between the four datasets are quickly revealed.
4. The intent of this article is to help build the general case that phylogenetic comparative methods are statistical methods and consequently that graphing or visualization should invariably be included as an essential step in our standard data analytical pipelines.
5. Phylogenies are complex data structures and thus visualizing data on trees in a meaningful and useful way is a challenging endeavour. We recommend that the development of graphical methods for simultaneously visualizing data and tree should continue to be an important goal in phylogenetic comparative biology.

KEYWORDS

comparative methods, macroevolution, phylogeny, plotting, visualization

1 | INTRODUCTION

Phylogenetic comparative biology is the discipline that involves using phylogenies, often combined with trait data for species, to draw inferences about evolutionary change through time (Felsenstein,

1985). Phylogenetic comparative methods now comprise an incredibly vast array of approaches (O'Meara, 2012). For instance, comparative methods have been developed to model diversification (speciation and extinction), trait evolution for single discrete or continuously valued characters, correlated evolution of multiple

traits, or to reconstruct ancestral values for one or more characters at internal nodes of the tree (Blomberg, Garland, & Ives, 2003; Butler & King, 2004; Garland, Harvey, & Ives, 1992; Hansen, 1997; Ives, Midford, & Garland, 2007; Lewis, 2001; Maddison, Midford, & Otto, 2007; Nee, May, & Harvey, 1994; O'Meara, Ané, Sanderson, & Wainwright, 2006; Pagel, 1994; Schluter, Price, Mooers, & Ludwig, 1997). Phylogenetic methods have been proposed to account for competition among species, diversity-dependent ecological interactions, host and parasite co-evolution, and a range of other phenomena of ecological or evolutionary relevance (e.g., Emerson & Gillespie, 2008; Etienne et al., 2011; Legendre, Desdevises, & Bazin, 2002; Mahler, Revell, Glor, & Losos, 2010). By any measure, the importance of phylogenetic comparative analyses in macroevolutionary research is already high and continues to grow each year. For instance (according to Google Scholar), in 2017 alone there were more than 1,400 published articles containing the phrase “phylogenetic comparative.” (In 2016, that number was around 1,200.)

Since phylogenies are complex data structures and thus (nearly) all phylogenetic comparative methods are specialized to the discipline, it is easy to overlook the fact that phylogenetic comparative analyses are statistical methods too. Usually these methods rely on explicit, sometimes simplistic, underlying models—just as do virtually all parametric statistical approaches designed for the analysis of nonphylogenetic information. For instance, a wide range of phylogenetic comparative methods assume a model called Brownian motion for the evolution of the traits under study or for the correlation structure in the residual error of a fitted linear model (or both).

Brownian motion is a continuous-time stochastic diffusion process in which the expected value is constant and the variance among lineages increases linearly with time (Felsenstein, 1985; O'Meara et al., 2006; Revell, 2008; Revell, Harmon, & Collar, 2008). It is used to model continuous character evolution on phylogenies because it is mathematically tractable, because it is a reasonable approximation of genetic drift under some assumptions or of natural selection under others, and because Brownian evolution results in the fairly sensible outcome that closely related taxa tend to exhibit similar trait values (O'Meara et al., 2006).

In 1973, the English statistician Francis (“Frank”) Anscombe published the analysis of a quartet of fictitious bivariate datasets in which he demonstrated that each was identical in many summary statistics, including fitted regression coefficients, sum-of-squares, model fit, and so on (Anscombe, 1973). Nonetheless, when visualized (Figure 1) these data revealed remarkably different patterns of covariation between the x and y dimensions of each plot. For instance, in Figure 1a we see a pattern that more or less matches what we might presume for a linear regression model fitted to evolutionary or ecological data. By contrast, Figure 1b shows an inarguably parabolic relationship between the two variables; the data of Figure 1c exhibit a perfect linear pattern with one striking outlier; finally, the plot of Figure 1d shows a pattern in which all but one value of y share a common value of x . Anscombe reports presenting this result to confront several commonly (at the time) indoctrinated notions in statistical analysis, including that “numerical calculations are exact, but graphs are rough” and that “performing intricate

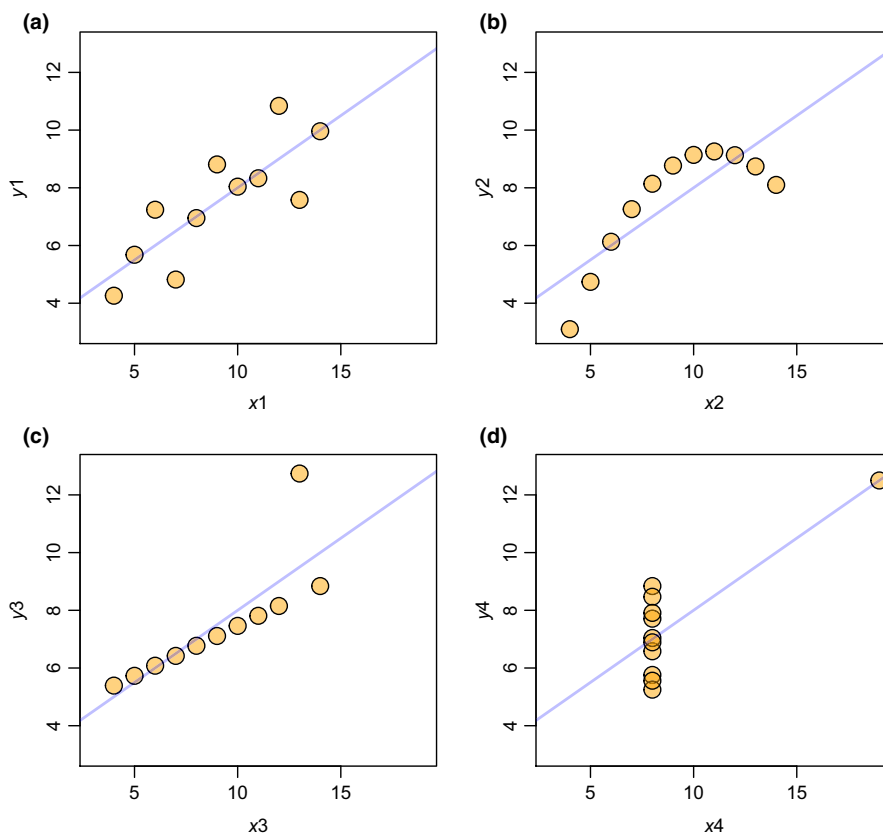


FIGURE 1 Panels (a–d) give Anscombe’s (1973) famous quartet of datasets. Each yield identical summary statistics, including fitted regression coefficients and model fits, yet graphing reveals markedly different patterns of covariation between x and y in each of the four datasets. The blue line in each subplot gives the fitted bivariate regression line for the data of the plot, which is identical in each panel

calculations is virtuous, whereas actually looking at the data is cheating" (Anscombe, 1973).

We argue that the same issue that Anscombe was attempting to address also afflicts phylogenetic comparative biology, and furthermore that the problem might actually be worse in our field than it may have been in statistics at the time Anscombe published his work. This is because phylogenies are complex, and properly visualizing phylogenetic comparative data often involves simultaneously representing our data and the relationships among taxa (but see Garland et al., 1992 for useful visualizations of phylogenetic data in a transformed space). In phylogenetic studies (unlike in Anscombe's example), an "outlier" (be it a genuine evolutionary anomaly that is driving our result, or an error in our data) might not be apparent when our raw data are plotted apart from the tree.

Herein, we present a novel quartet of fictitious datasets generated via totally different procedures that (much like the original Anscombe quartet) yield precisely the same fitted parameters of the Brownian motion process and model fits. We then visualize the data using three readily available plotting methods. We emphasize that neither the quartet of datasets nor the triplet of visualization techniques we have used in this study is intended to survey the range of ways in which phylogenetic data could violate model assumptions nor the scope of visualization methods that could be leveraged to identify this deviation. To the contrary, we have intentionally selected relatively simple data and visualization techniques in a mere effort to reiterate the point of Anscombe (1973) and to argue that it applies equally well to statistical analysis using phylogenetic comparative methods as it does to ordinary parametric statistics.

2 | METHODS AND RESULTS

2.1 | Preamble on Brownian motion

In this article, we have chosen to focus on the relatively simple model of Brownian motion evolution. This too follows in the spirit of Anscombe (1973) who elected to focus on bivariate data and the relatively simple statistical procedure of linear regression, rather than on more complicated data or methods. We chose Brownian motion because a pattern of evolution by this process, in which the expected value is constant and variance accrues between diverging lineages in direct proportion to time, produces a reasonably visually discernible pattern when graphed. This advantages us when we are most interested in the fit-to or deviation from Brownian motion of our own data; however, the fact that other, more complicated processes of evolutionary change for continuous character traits or other types of comparative data do not necessarily result in the same reasonably predictable pattern when plotted does not (in our opinion) negate the value of graphing. In the Discussion, we will highlight some plotting tactics for more complex evolutionary scenarios than those explored by this article.

The Brownian model has only two parameters. The first is the evolutionary rate—so called because it represents the instantaneous

rate of diffusion under the process, or (similarly) the instantaneous rate of accumulation of variance between a lineage and its ancestor. This parameter is normally denoted σ^2 , as it will be in this article (O'Meara et al., 2006). The second parameter of the Brownian process is the starting value at the root of the tree. This parameter is usually of less interest to macroevolutionary biologists, except inasmuch as it represents the estimated ancestral state value at the root. When referred to symbolically, this parameter will be denoted herein as a . To fit the model of Brownian motion to our data, we will use the statistical estimation procedure of maximum likelihood, though it is also possible to use other estimation procedures such as phylogenetically independent contrasts (Felsenstein, 1985) or least squares (Rohlf, 2001). The results of this article would be largely the same in either case.

2.2 | Four fictitious phylogenetic datasets

In this study, we first simulated a twenty-taxon pure-birth ("Yule") phylogeny (Figure 2). Next, we generated data for this tree using four different procedures. Following directly in the spirit of Anscombe (1973), the goal of these four procedures was to produce data vectors by very different methods but for which the fitted maximum likelihood estimates (MLEs) of the parameters of the Brownian process would be $\sigma^2 = 0.5$ and $a = 3.0$.

Note that our intent is not at all to comprehensively survey the many and varied evolutionary scenarios under which biological data could arise (e.g., Revell et al., 2008). To the contrary, our aim was merely (once again, following Anscombe, 1973) to create four fictitious datasets by inarguably widely different procedures that

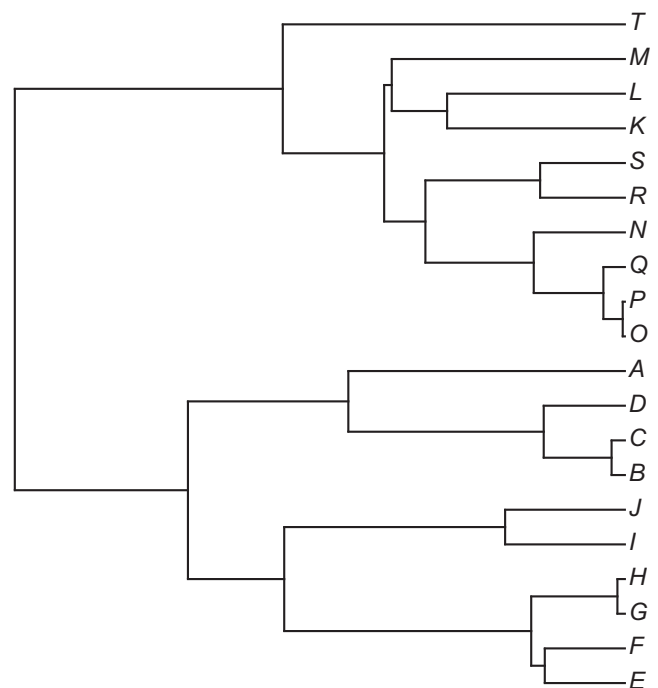


FIGURE 2 The simulated pure-birth ("Yule") phylogeny used in this study

nonetheless resulted in the same model fit and parameter estimates. We then intend to see if the differences between these datasets become more apparent when graphed.

First, for data vector y_1 , we began by simulating data via Brownian motion. Then, with the intent of exaggerating the phylogenetic structure of our data relative to Brownian evolution, we rearranged the data by the “cladewise” order of the terminal taxa of the tree (i.e., top-to-bottom order when the tree is plotted, or *vice versa*; Revell, 2014). Finally, we used a numerical optimization routine to rescale and recentre the data, so that it would have an estimated Brownian motion rate parameter, σ^2 equal to 0.5, and an ancestral value at the root node of 3.0. This procedure will result in a data structure similar to Brownian motion, but with an amplified tendency of related species to possess similar phenotypes. Second, to obtain data vector y_2 , we merely simulated data via Brownian motion and then rescaled and recentred it in its original order via numerical optimization as above. Third, to generate data vector y_3 , we simply drew values randomly from a uniform distribution on the interval [0, 6]. With these data too, we used numerical optimization to identify rescaling and recentring factors to ensure that the fitted value of σ^2 would be equal to 0.5, and the fitted value of the ancestral state at the root node, a , would be equal to 3.0. Finally, fourth, we obtained the final data vector, y_4 , by first simulating Brownian motion evolution on the tree but with a much lower rate of evolution ($\sigma^2=0.1$), and then optimizing the position of the ultimate (left-most) tip of the tree to obtain the same ancestral state and fitted rates as in the previous three cases.

The data from this exercise (rounded to two decimal places) are given in Table 1. Simulation code that can be used by the reader to generate his or her own “phylogenetic Anscombe datasets” is available as a supplementary appendix to this article. The fitted model parameter values, log-likelihoods, and so on are given in Table 2. (The reader can confirm that they are all identical to at least four significant digits.)

2.3 | Three useful visualizations

Though the fictitious datasets for this study were all obtained in radically different ways, they all yield precisely the same estimated parameters and model fits, as per our design. This kind of result could easily lead a naïve user to conclude that the four data vectors had evolved by a similar process. Nonetheless, relatively simple visualization tools can quickly tell us that this was probably not the case. Here, we have chosen to focus on three graphical methods—although other helpful plotting techniques for this type of data are readily imaginable.

The first method is a projection of the tree into phenotype space called a “traitgram” (Evans, Smith, Flynn, & Donoghue, 2009; Revell, 2013, 2014). In this class of plotting method, the projection involves first reconstructing values at all the internal nodes of the tree and then merely drawing lines to connect all parent nodes to their daughters, in which the horizontal position of the node is determined based on its height above the root of the tree (Revell, 2014).

TABLE 1 The phylogenetic Anscombe datasets used in this study

Taxon ID	y_1	y_2	y_3	y_4
A	3.12	2.53	2.74	3.30
B	2.54	3.53	3.25	2.86
C	2.75	3.34	3.26	2.93
D	3.01	2.45	3.05	2.83
E	1.28	3.12	3.25	2.69
F	1.41	2.65	2.89	2.57
G	1.75	2.50	3.24	2.84
H	1.97	2.52	3.16	2.84
I	2.23	2.14	3.07	2.74
J	2.39	2.27	2.69	2.92
K	3.78	4.05	2.84	2.26
L	3.78	3.41	3.37	2.32
M	3.94	3.56	3.08	2.27
N	3.49	3.38	2.82	2.23
O	3.15	3.25	3.43	2.03
P	3.19	3.26	3.33	2.05
Q	3.29	3.38	2.73	2.15
R	3.50	3.90	2.94	2.45
S	3.61	3.34	3.13	2.53
T	3.95	3.29	2.95	4.94

TABLE 2 Fitted parameter estimates, log-likelihood, AIC, and number of estimated parameters for each dataset. They are all numerically identical to four significant digits. Note that due to rounding the values from Table 1 will yield estimated parameters and likelihoods *slightly* different to those reported here

Dataset	σ^2	a	log(L)	AIC	k
y_1	0.5000	3.000	-8.108	20.22	2
y_2	0.5000	3.000	-8.108	20.22	2
y_3	0.5000	3.000	-8.108	20.22	2
y_4	0.5000	3.000	-8.108	20.22	2

Via this graphical method, we can easily see that the pattern of divergence among taxa is wildly different in the quartet of cases in this study (Figure 3). For instance, it is immediately apparent that for trait vector y_4 the overwhelming majority of lineages have diverged little from their ancestor, while one lineage is displaced greatly from the rest (an evolutionary anomaly, so to speak, or a statistical outlier). Fine details, however, can be difficult to distinguish in this type of plot. For instance, because their relative positions are so highly clumped, it is hard to discern the presence or absence of phylogenetic structure in the data for Figure 3c.

The second method uses a projection of the data onto the tree, rather than of the tree into trait space. This can be done via multiple methods, but in this case we have elected to use the visually appealing *contMap* (continuous character mapping) method of

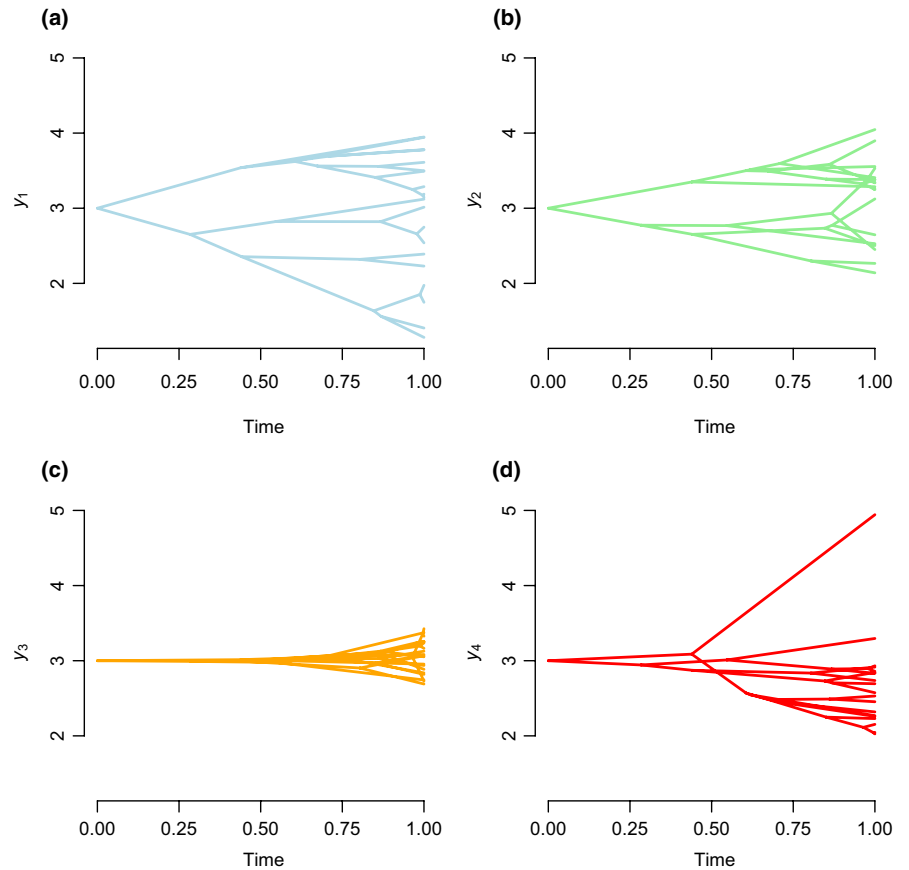


FIGURE 3 Traitgrams (projections of the tree into a space defined by observed or reconstructed trait value on the vertical axis and time since the root on the horizontal) for each of the four datasets used in this study. Panels (a–d) show projections of the tree of Figure 2 into the trait spaces defined by each of the trait vectors (columns) of Table 1. As noted in the article text and shown in Table 2, all four result in the same maximum likelihood estimates of the Brownian rate of evolution (σ^2) and the reconstructed ancestral state at the global root node of the tree (a), as well as identical log-likelihoods of the fitted models

Revell (2013; Figure 4). In this method, a colour gradient is used to map observed and reconstructed trait values onto the edges of the phylogeny. Simpler methods in which trait data are mapped to the tips of the tree serve a similar purpose and are also available. This method clearly illustrates some features of character variation that were less apparent in the traitgram plots of Figure 3. For instance, in Figure 4a,b the phylogenetic structure of the data is readily apparent, as is its absence in Figure 4c.

Finally, the third graphical diagnostic we elected to apply was an exclude-one sensitivity plot. In this case, our tactic was to sequentially exclude every taxon, one by one, and each time reestimate the Brownian rate parameter, σ^2 . The intent of this type of visualization is to measure the relative importance of each measured tip value in determining the fitted parameter value under the model. It is straightforward to overlay the global estimate of the rate as we have also done in this case using the vertical red line of each panel. (It is invariably $\sigma^2 = 0.5$ as per our design.) This analysis shows us that some tips carry much more influence in our analysis than do others and might be a handy way to identify outliers (evolutionary anomalies or measurement errors) in our data (Figure 5). Particularly revealing is Figure 5d in which we can easily identify the disproportionate influence of the left-most tip of the tree.

Most importantly, all three graphical methods suggest data that have arisen by quite different processes (perhaps barring y_1 and y_2) in spite of their aforementioned identical fitted parameter values and model fits. No single method, however, is a panacea in the sense

that each plot reveals complementary information about the deviation of each dataset from the assumptions of the fitted model.

2.4 | Notes on implementation

All analyses of this article were conducted in the R statistical computing environment (R Core Team, 2017). We performed simulation of the phylogeny and trait data using PHYTOOLS (Revell, 2012), which in turn depends on the important R phylogenetics packages APE (Paradis, Claude, & Strimmer, 2004) and PHANGORN (Schliep, 2011), as well as on a number of other R packages (Azzalini & Genz, 2016; Becker, Wilks, Brownrigg, Minka, & Deckmyn, 2016; Chasalow, 2012; Gilbert & Varadhan, 2016; Harmon, Weir, Brock, Glor, & Challenger, 2008; Jackson, 2011; Lemon, 2006; Ligges & Mächler, 2003; Pinheiro, Bates, DebRoy, & Sarkar, 2017; Plummer, Best, Cowles, & Vines, 2006; Qiu & Joe, 2015; Venables & Ripley, 2002; Xie, 2013). We performed model-fitting using maximum likelihood in the R package GEIGER (Harmon et al., 2008), which too depends on various contributed R libraries (Eastman, 2013; Eddelbuettel, 2013; Eddelbuettel & Francois, 2011; Eddelbuettel et al., 2017; Genz & Bretz, 2009; Genz et al., 2017; Ihaka, Murrell, Hornik, Fisher, & Zeileis, 2016; King, 2017; Paradis et al., 2004; Plummer et al., 2006; Soetaert, Petzoldt, & Setzer, 2010; Venables & Ripley, 2002). Figure 1 was generated using the DATASETS package (R Core Team, 2017). All remaining plots were produced exactly as represented in R using PHYTOOLS (Revell, 2012).

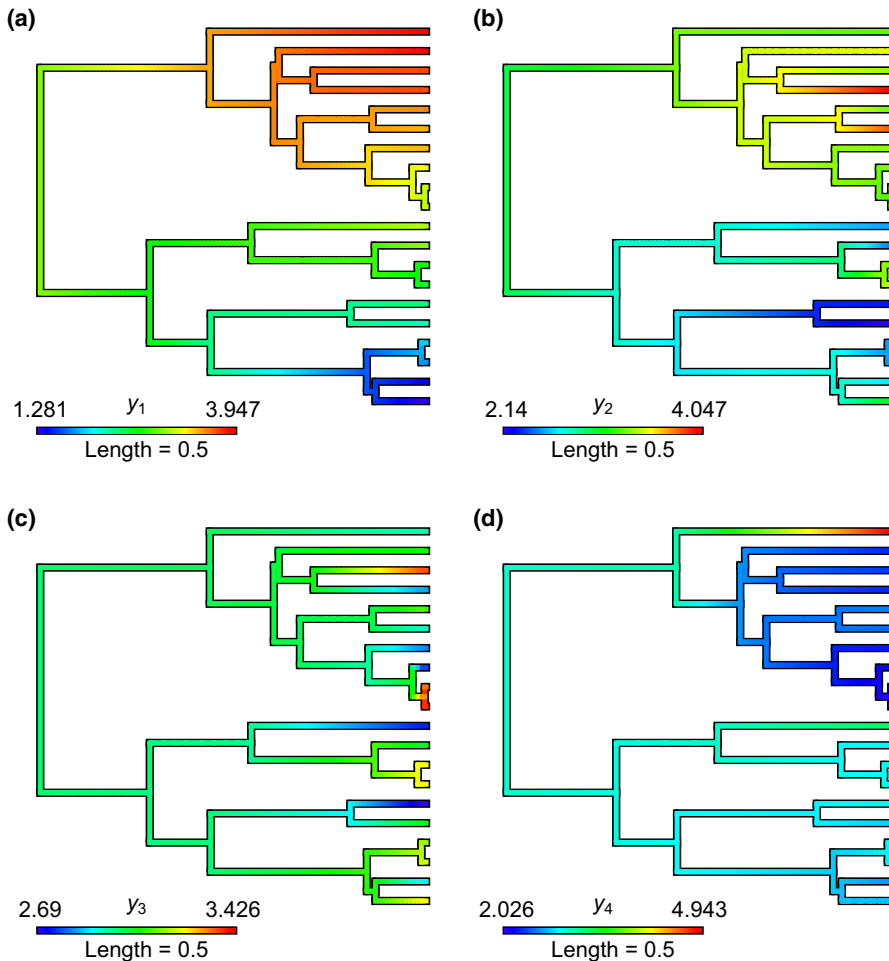


FIGURE 4 Continuous character mapping (*contMap*) plots for each of the four datasets in Table 1 and Figure 3. Observed and reconstructed values of each of the four columns of Table 1 have been projected onto the tree of Figure 2 in panels (a–d). Note the conspicuous absence of phylogenetic structure in the mapped character values of panel (c) of the plot

3 | DISCUSSION

In recent years, phylogenetic comparative methods have expanded dramatically in their use and importance in evolutionary biology, as well as invaded other disciplines such as linguistics, anthropology, and genomics (e.g., Brunet et al., 2006; Nunn, 2011). Due to the unusual structure of the phylogeny, the overwhelming majority of phylogenetic comparative methods are specialized to the discipline which makes it easy to ignore the fact that they are statistical methods too and thus equally, if not more, vulnerable to deviations from model assumptions and errors in the data.

Nearly 45 years ago, statistician Francis Anscombe published what would become one of the emblematic papers of his career. In his article, he cleverly identified a quartet of bivariate datasets that yielded identical summary statistics, regression coefficients, and model fits (Anscombe, 1973). Plotting the quartet, however, revealed strikingly different patterns of covariation between x and y in each of the four datasets (Figure 1). The point of this exercise was to reiterate the value of data visualization in an era in which fitting sophisticated models to data had taken primacy over what one can see with one's own eyes. Subsequently, Tukey (1977) helped popularize exploratory data analysis, including via the introduction of a number of new visualization methods that continue to be popular today.

Since the days of Anscombe and Tukey it would be hard to argue that data visualization has remained out of favour. To the contrary, computational tools such as the \mathbf{R} project (R Core Team, 2017), with its fancy suite of plotting methods, has helped put graphing back into vogue. In fact, even Anscombe's original example has seen something of a renaissance. For instance, the phrase "Anscombe's quartet" was mentioned in 125 articles in 2016 (according to Google Scholar); and Anscombe (1973) continues to inspire computer scientists to publish methods (e.g., based on simulated annealing) to generate datasets that graph in a peculiar way, yet nonetheless yield identical summary statistics (e.g., Matejka & Fitzmaurice, 2017).

However, phylogenetic comparative data are inherently difficult to plot because we must identify ways to simultaneously graph our comparative data together with the phylogeny in a meaningful way. This is because outliers (be they genuinely anomalous evolutionary outliers, or errors in the data or tree) may not be apparent when the data are graphed apart from the tree. Consequently, it is not a routine procedure to simultaneously visualize data and tree in phylogenetic comparative studies. In this article, we present a quartet of *phylogenetic* Anscombe datasets—that is, datasets obtained via wholly different procedures but that nonetheless match identically in fitted model parameters and model fits. The point of

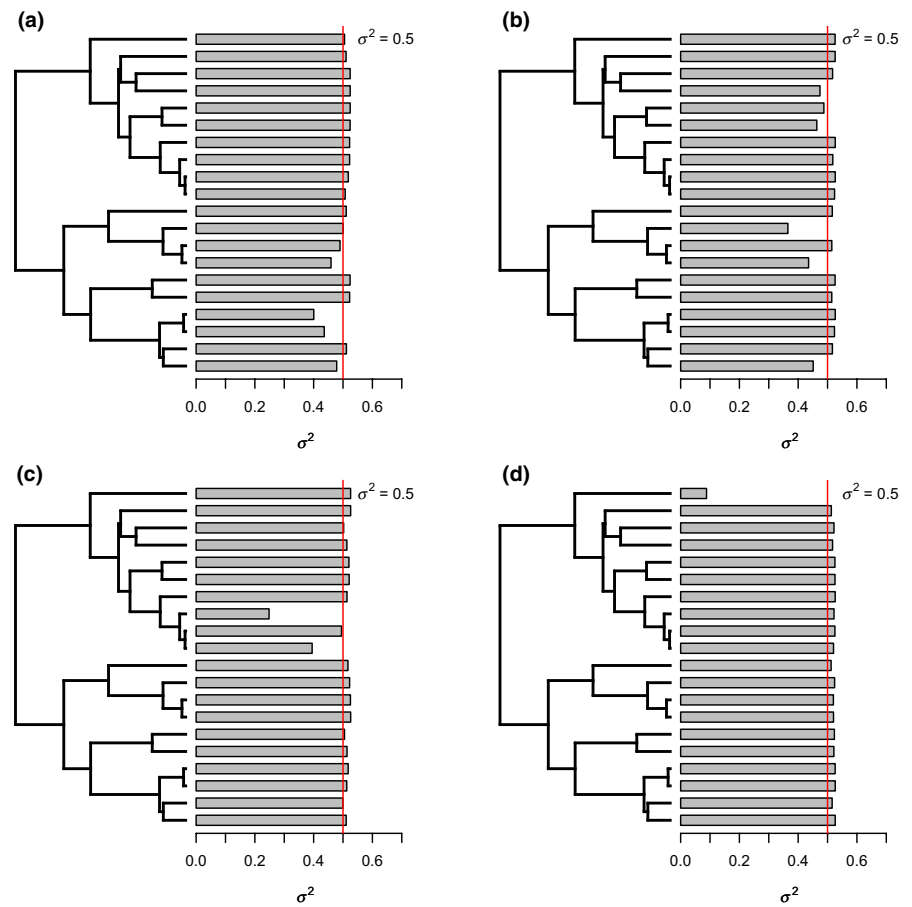


FIGURE 5 Remove-one sensitivity analysis of the estimated Brownian rate (σ^2) for the four datasets of Table 1 in panels (a–d). The bars to the right of each subplot show the estimated rate when each of the corresponding tips is dropped from the analysis

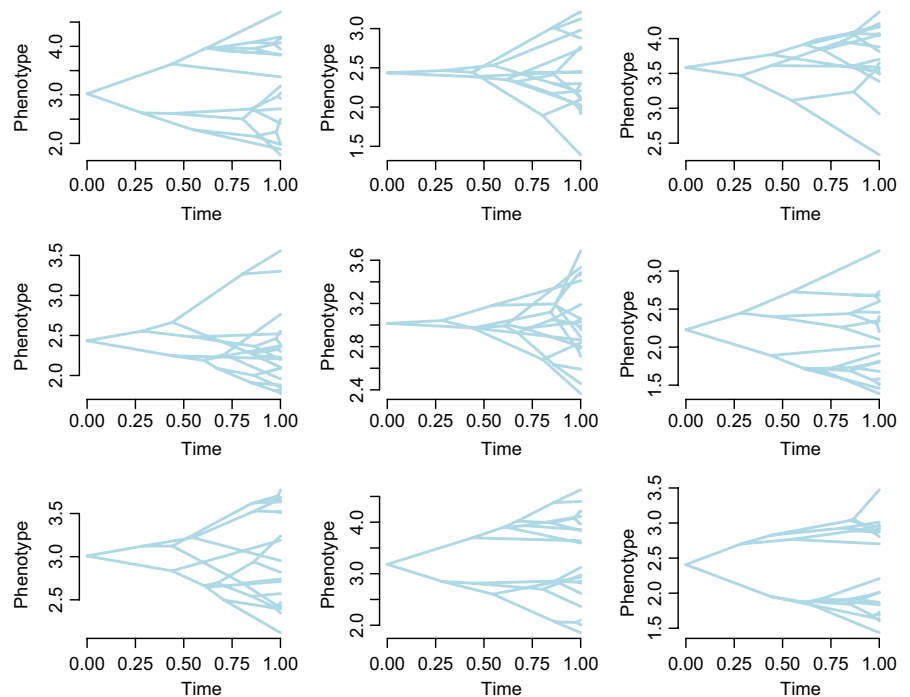


FIGURE 6 Traitgrams for nine simulated datasets obtained by simulating under the maximum likelihood estimated parameters ($\sigma^2 = 0.5$ and $\alpha = 3.0$) obtained from any of the quartet of datasets of this study

so doing is not to disillusion the reader, but merely to point out that the differences between these datasets are readily apparent when graphed!

Note that Garland et al. (1992) introduced a number of relatively simple diagnostic visualization methods for phylogenetic comparative data that employ a transformed space based on independent

contrasts (Felsenstein, 1985). The methods we selected for this article should not be seen as a replacement for the approaches proposed by Garland et al. (1992), but perhaps merely as a complement to them. In fact, we suspect that deviation from Brownian motion in our Anscombe phylogenetic datasets would also be quite apparent using some of the graphing methods of Garland et al. (1992)! Unfortunately, though these methods were once widely used, along with the evolution of the software employed for conducting phylogenetic comparative analyses the visual diagnostic procedures of Garland et al. (1992) have fallen somewhat out of favour. Nor do we suppose that we have identified the graphing procedures most appropriate for detecting deviations in our data from model assumptions. We merely selected these methods because they comprise a range of approaches and serve to emphasize the importance of exploratory plotting, in the spirit of Anscombe (1973).

Earlier, we noted that the focus on the relatively simple case of Brownian motion evolution was intentional—because this process produces a fairly predictable pattern of similarity and dissimilarity among species. However, we argue that visualization is important for all types of datasets and analyses. For instance, in the case of a bivariate or multivariate regression model fit using phylogenetic generalized least squares, one might use a traitgram plot (such as those of Figure 3) to project tree into the residual error spaces, or a *contMap* style visualization (such as those of Figure 4) to project the correlation structure of the residual error onto the tree. For more complex models, one can always conduct sensitivity analyses (such as those represented in Figure 5) to examine the relative weight of different terminal taxa in the final fitted model. Finally, if all else fails, it is frequently quite straightforward to simulate under the fitted model and visualize the results. For instance, Figure 6 shows nine simulated datasets generated using the fitted model parameters of all four phylogenetic Anscombe datasets of this study. By casual inspection, it seems quite obvious that none resemble panels (c) or (d) of Figure 3 in this article which might help us conclude that neither data vector had genuinely evolved by a biological process that is well-approximated by Brownian evolution.

Finally, we feel that it is important to emphasize that although we believe that graphical methods should be used in phylogenetic comparative analyses—they should not be abused. For instance, it would be relatively facile for us to examine Figure 3d and conclude that the uppermost lineage of the tree had evolved via a different process than the rest, fit a model in which the process for that lineage was permitted to vary, and obtain a significant result. This process of testing a hypothesis suggested by the data under conditions in which the data are limited has been labelled “uncomfortable science” by Tukey (Hoaglin, Mosteller, & Tukey, 1985). We should not be deceived, however, for statistical analyses generally assume that we have devised our hypotheses to be tested a priori—that is, independent of our observations. It is widely appreciated that this is seldom genuinely the case in empirical studies in which the specific data to be collected are often motivated by anecdotal (but heretofore untested) observations, themselves a form of data. It is even less true when our hypotheses are formulated from a visualization of the very data to be used to test our hypothesis!

4 | CONCLUSION

Phylogenetic comparative methods are those in which a phylogeny is often combined with trait data for the species in the tree to test hypotheses about macroevolution. Anscombe (1973) published a quartet of datasets to demonstrate the critical importance of graphing in statistical analyses. Phylogenetic comparative methods are statistical methods too and here we employ our own fictional quartet of phylogenetic Anscombe datasets to help illustrate the importance of data visualization in phylogenetic comparative studies.

ACKNOWLEDGEMENTS

The authors thank the United States National Science Foundation (DEB 1350474 and DBI 1759940 to L.J.R.) for the funding that permitted this work to take place. D. Bapst and D. Wright participated in a helpful social media discussion that led to significant improvements in this article. Associate editor S. Blomberg and two anonymous reviewers provided a number of useful comments that helped improve the article over earlier versions. Finally, L. Harmon coined the phrase “phylogenetic comparative methods are statistical methods too” (or some variant thereof) which we have reused on more than one occasion in this article.

AUTHORS' CONTRIBUTIONS

All authors conceived the project and contributed significantly to writing. L.J.R. and K.S. developed the visualization methods and/or implemented them in software. L.J.R. wrote the first draft of the manuscript.

DATA ACCESSIBILITY

Data have been deposited in the Dryad Digital Repository <https://doi.org/10.5061/dryad.11340t4> (Revell, Schliep, Valderrama, & Richardson, 2018).

ORCID

Liam J. Revell  <http://orcid.org/0000-0003-0767-4713>

Klaus Schliep  <http://orcid.org/0000-0003-2941-0161>

Eugenio Valderrama  <http://orcid.org/0000-0002-6320-7792>

REFERENCES

- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, 27, 17–21.
- Azzalini, A., & Genz, A. (2016). The R package “MNORMT”: The multivariate normal and “t” distributions (version 1.5-5). Retrieved from <https://CRAN.R-project.org/package=mnormt>
- Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P., & Deckmyn, A. (2016). *MAPS: Draw geographical maps*. R package version 3.1.1. <https://CRAN.R-project.org/package=maps>.

- Blomberg, S. P., Garland Jr, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*, 717–745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- Brunet, F. G., Crollius, H. R., Paris, M., Aury, J.-M., Gilbert, P., Jaillon, O., ... Robinson-Rechavi, M. (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Molecular Biology and Evolution*, *23*, 1808–1816. <https://doi.org/10.1093/molbev/msl049>
- Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *American Naturalist*, *164*, 683–695. <https://doi.org/10.1086/426002>
- Chasalow, S. (2012). COMBINAT: Combinatorics utilities. R package version 0.0-8. <https://CRAN.R-project.org/package=combinat>.
- Eastman, J. (2013). NCBIT: Retrieve and build NCBI taxonomic data. R package version 2013.03.29. <https://CRAN.R-project.org/package=ncbit>.
- Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. New York: Springer. <https://doi.org/10.1007/978-1-4614-6868-4>
- Eddelbuettel, D., & Francois, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*, 1–18.
- Eddelbuettel, D., Lucas, A., Tuszynski, J., Bengtsson, H., Urbanek, S., Frasca, M., ... Simko, V. (2017). DIGEST: Create compact hash digests of R objects. R package version 0.6.12. <https://CRAN.R-project.org/package=digest>.
- Emerson, B. C., & Gillespie, R. G. (2008). Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology and Evolution*, *23*, 619–630. <https://doi.org/10.1016/j.tree.2008.07.005>
- Etienne, R. S., Haegeman, B., Stadler, T., Aze, T., Pearson, P. N., Purvis, A., & Phillimore, A. B. (2011). Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society Series B*, *279*, 1300–1309.
- Evans, M. E. K., Smith, S. A., Flynn, R. S., & Donoghue, M. J. (2009). Climate, niche evolution, and diversification of the “bird-cage” evening primroses (Oenothera, sections Anogra and Kleinia). *American Naturalist*, *173*, 225–240. <https://doi.org/10.1086/595757>
- Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, *125*, 1–15. <https://doi.org/10.1086/284325>
- Garland Jr, T., Harvey, P. H., & Ives, A. R. (1992). Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology*, *41*, 18–32. <https://doi.org/10.1093/sysbio/41.1.18>
- Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*, Vol. 195, *Lecture Notes in Statistics*. Heidelberg: Springer-Verlag. <https://doi.org/10.1007/978-3-642-01689-9>
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2017). MVTNORM: Multivariate normal and t distributions. R package version 1.0-6. <http://CRAN.R-project.org/package=mvtnorm>.
- Gilbert, P., & Varadhan, R. (2016). NUMDERIV: Accurate numerical derivatives. R package version 2016.8-1. <https://CRAN.R-project.org/package=numDeriv>.
- Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*, *51*, 1341–1351. <https://doi.org/10.1111/j.1558-5646.1997.tb01457.x>
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008). GEIGER: Investigating evolutionary radiations. *Bioinformatics*, *24*, 129–131. <https://doi.org/10.1093/bioinformatics/btm538>
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1985). *Exploring data tables, trends, and shapes*. New York: John Wiley & Sons.
- Ihaka, R., Murrell, P., Hornik, K., Fisher, J. C., & Zeileis, A. (2016). COLORSPACE: Color space manipulation. R package version 1.3-2. <https://CRAN.R-project.org/package=colorspace>.
- Ives, A. R., Midford, P. E., & Garland Jr, T. (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology*, *56*, 252–270. <https://doi.org/10.1080/10635150701313830>
- Jackson, C. H. (2011). Multi-state models for panel data: The MSM package for R. *Journal of Statistical Software*, *38*, 1–29.
- King, A. A. (2017). SUBPLEX: Unconstrained optimization using the subplex algorithm. R package version 1.4-1. <https://CRAN.R-project.org/package=subplex>.
- Legendre, P., Desdevises, Y., & Bazin, E. (2002). A statistical test for host-parasite coevolution. *Systematic Biology*, *51*, 217–234. <https://doi.org/10.1080/10635150252899734>
- Lemon, J. (2006). PLOTRIX: A package in the red light district of R. *R-News*, *6*, 8–12.
- Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*, 913–925. <https://doi.org/10.1080/106351501753462876>
- Ligges, U., & Mächler, M. (2003). SCATTERPLOT3D – An R package for visualizing multivariate data. *Journal of Statistical Software*, *8*, 1–20.
- Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, *56*, 701–710. <https://doi.org/10.1080/10635150701607033>
- Mahler, D. L., Revell, L. J., Glor, R. E., & Losos, J. B. (2010). Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution*, *64*, 2731–2745. <https://doi.org/10.1111/j.1558-5646.2010.01026.x>
- Matejka, J., & Fitzmaurice, G. (2017). Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. In *ACM Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, (pp. 1290–1294). New York, NY: ACM.
- Nee, S., May, R. M., & Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society London B*, *344*, 305–311. <https://doi.org/10.1098/rstb.1994.0068>
- Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology*. Chicago, IL: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226090009.001.0001>
- O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: A review of methods. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 267–285. <https://doi.org/10.1146/annurev-ecolsys-110411-160331>
- O'Meara, B. C., Ané, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, *60*, 922–933. <https://doi.org/10.1111/j.0014-3820.2006.tb01171.x>
- Pagel, M. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society London B*, *255*, 37–45. <https://doi.org/10.1098/rspb.1994.0006>
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290. <https://doi.org/10.1093/bioinformatics/btg412>
- Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D., & R Core Team. (2017). NLME: Linear and nonlinear mixed effects models. R package version 3.1-131. <https://CRAN.R-project.org/package=nlme>
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R-News*, *6*, 7–11.
- Qiu, W., & Joe, H. (2015). CLUSTERGENERATION: Random cluster generation (with specified degree of separation). R package version 1.3.4. <https://CRAN.R-project.org/package=clusterGeneration>.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Revell, L. J. (2008). On the analysis of evolutionary change along single branches in a phylogeny. *American Naturalist*, *172*, 140–147. <https://doi.org/10.1086/588078>
- Revell, L. J. (2012). PHYTOOLS: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*, 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>

- Revell, L. J. (2013). Two new graphical methods for mapping trait evolution on phylogenies. *Methods in Ecology and Evolution*, 4, 754–759. <https://doi.org/10.1111/2041-210X.12066>
- Revell, L. J. (2014). Graphical methods for visualizing comparative data on phylogenies. In L. Z. Garamszegi (Ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice* (pp. 77–103). Berlin, Heidelberg: Springer-Verlag.
- Revell, L. J., Harmon, L. J., & Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, 57, 591–601. <https://doi.org/10.1080/10635150802302427>
- Revell, L. J., Schliep, K., Valderrama, E., & Richardson, J. E. (2018). Data from: Graphs in phylogenetic comparative analysis: Anscombe's quartet revisited. *Dryad Digital Repository*, <https://doi.org/10.5061/dryad.11340t4>
- Rohlf, F. J. (2001). Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution*, 55, 2143–2160.
- Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27, 592–593. <https://doi.org/10.1093/bioinformatics/btq706>
- Schluter, D., Price, T., Mooers, A. Ø., & Ludwig, D. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution*, 51, 1699–1711. <https://doi.org/10.1111/j.1558-5646.1997.tb05095.x>
- Soetaert, K., Petzoldt, T., & Setzer, R. W. (2010). Solving differential equations in R: Package `DEsolve`. *Journal of Statistical Software*, 33, 1–25.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer. <https://doi.org/10.1007/978-0-387-21706-2>
- Xie, Y. (2013). ANIMATION: An R package for creating animations and demonstrating statistical methods. *Journal of Statistical Software*, 53, 1–27.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Revell LJ, Schliep K, Valderrama E, Richardson JE. Graphs in phylogenetic comparative analysis: Anscombe's quartet revisited. *Methods Ecol Evol*. 2018;9:2145–2154. <https://doi.org/10.1111/2041-210X.13067>