

Visualizations of Circular Data

A Supplement to “*Working with Circular Data: A Tutorial for Cognitive and Behavioral Research*”

Stella Wernicke^{1*}, Julie de Falco¹, Zahara Gironés^{1,2,3}, Chris Jungerius¹,
Ivan Tomić⁴, Adam Triabhall¹, Xiaolu Wang^{1,5}, and Paul M Bays¹

¹Department of Psychology, University of Cambridge, United Kingdom

²Department of Engineering, University of Cambridge, United Kingdom

³Department of Clinical Neuroscience, University of Cambridge, United Kingdom

⁴Department of Psychology, University of Zagreb, Croatia

⁵Department of Psychology, Sun Yat-sen University, China

*Corresponding author: saiw2@cam.ac.uk

Visualizing circular data requires approaches that respect its inherent periodicity and directional nature. A key decision to make is whether data is unwrapped onto a line and represented in a Euclidean coordinate system or represented in a two-dimensional polar coordinate system – where angle is mapped to direction (θ) and magnitude to radius (r). In some cases, linear and polar visualizations convey similar information, while in others, one format provides a much clearer or more intuitive representation of the data. This supplemental material will outline common approaches and potential pitfalls in the visualization of circular data, concluding with a discussion of methods for visualizing summary statistics. We also provide practical guidance on producing these visualizations in commonly used programming languages.

Linear Representations of Circular Data

One common way to visualize circular data is by unwrapping it into a *linear histogram* along the circular dimension, using binned counts (absolute or relative frequencies) or continuous density estimates. This requires two key decisions: selecting the interval and – as for Euclidean data – defining the bins. Both the standard interval $[0, 2\pi)$ and the signed interval $[-\pi, \pi)$, are common choices for linear visualizations. The latter is especially useful if there is a meaningful zero point, such as for errors relative to a reference angle. The choice of the interval determines the wrap-around point, which critically affects how clearly the structure of the data is displayed. When data are unimodal and highly concentrated, a linear histogram can closely resemble a standard histogram of non-circular data and often provides a simple, interpretable summary. In such cases, the interval should be chosen so that the region of highest density lies in the center of the x -axis. For multimodal distributions, boundaries should, where possible, be positioned in regions of low density to ensure that all modes are clearly represented. For instance, in Figure 5A, the shape of the distribution around the dominant mode is

clearly presented, while the secondary mode is positioned too close to the boundary, leading to density spilling over to the opposite side of the plot.

As for Euclidean data, the choice of bin size further shapes the appearance of the histogram: wider bins “smooth” over noise but may conceal fine structure in the data, whereas narrower bins reveal more detail but are susceptible to spurious fluctuations (Figure 5B). Common rules for selecting the size and number of bins in Euclidean space (e.g., Freedman & Diaconis, 1981; Scott, 1979, 2010; Sturges, 1926) are sometimes applied to circular data by treating angular values as lying on the real number line. However, heuristics based on dispersion measures may perform poorly in circular space, and even yield pathological results, such as vanishing bin widths for near-uniform data. Currently, no universally accepted bin-selection rule exists for circular histograms. In practice, researchers often use a fixed number of bins. Crucially, bin widths and locations should be selected so that edges coincide with the boundaries of the presentation interval. Another consequence of mapping circular data to a linear axis is that subgroup visualization no longer requires any circular-specific treatment; the same methods used for Euclidean data apply directly, including grouped, overlaid, or stacked histograms (e.g., Figure 5I).

A related method is to plot a *linear continuous density curve*, which provides a smoothed representation of the distribution. For sufficiently dense data, continuous density curves intuitively summarize the overall structure of the distribution without being tied to arbitrary bin edges. However, as for bin size in histograms, the choice of smoothing parameter strongly influences the result: too much smoothing can mask structure, while too little smoothing exaggerates noise (Figure 5C). Additionally, a smoothing algorithm that takes into account the connectedness of the two ends of the interval must be chosen. As for the Euclidean case, circular continuous density plots are not appropriate for discrete or heavily rounded data or for sparse data, where density es-

timates can be unstable or misleading. For a short review of nonparametric methods for smooth density estimation of circular data, see Chaubey (2022).

The *empirical cumulative distribution function* (eCDF) summarizes the data without requiring a choice of bin size or smoothing parameter (Figure 5D). It shows the cumulative proportion of observations along the circular axis and can help highlight changes in density while avoiding some of the visual noise of histograms. Like other linear representations, however, eCDFs still require choosing an interval, and selecting the interval so that the wrap-around point lies opposite the region of highest concentration helps to center the distribution.

For producing linear plots of circular data, readers can use the same software and functions familiar from visualizing Euclidean data. For this, it is essential that the data are correctly wrapped (i.e., mapped to their representative values within the interval selected for visualization), and if applicable, that density estimation is conducted using methods accounting for the periodicity of the data. Once the circular data is appropriately preprocessed, linear plots can be generated by using the same software for (binning and) plotting as for Euclidean data.

Polar Representations of Circular Data

We saw that linear visualizations such as histograms and eCDFs are intuitive and effective for concentrated, unimodal data. However, they may obscure structure in more dispersed or multimodal distributions, especially near wrap-around points. In such cases, polar visualizations often provide a clearer representation.

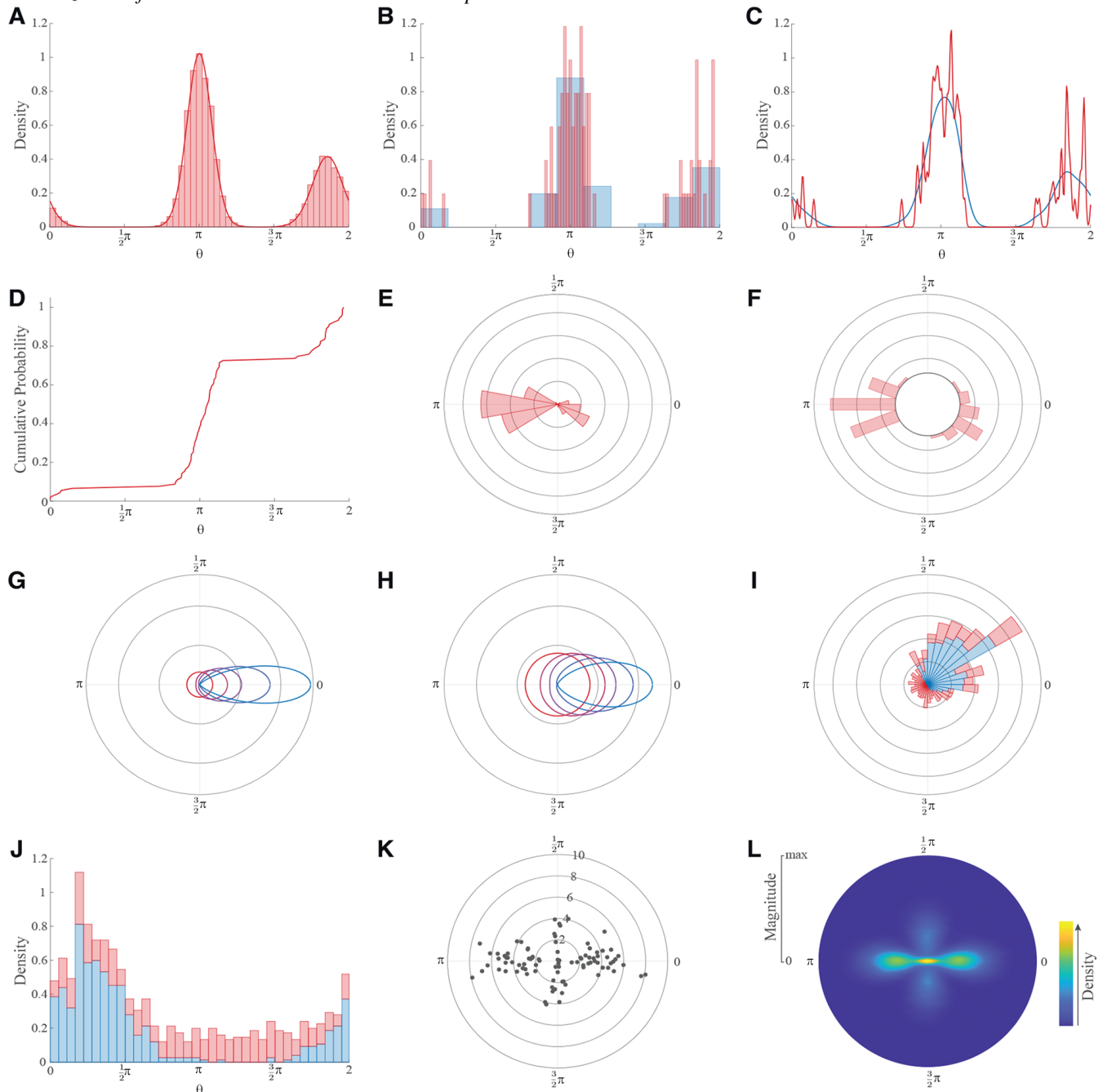
Polar histograms are widely used for polar representation of circular data (Figures 5E and 5F). The *rose plot* is a type of polar histogram that represents frequency data as radial wedges of equal angular width radiating from the center of a circle (Figure 5E). The height of each wedge indicates the number or proportion of observations in that angular bin. Unlike linear histograms, this approach preserves the continuity of the circular scale and avoids edge artifacts introduced by unwrapping. It also makes features such as the central tendency and directional biases easier to see. However, the representation is still influenced by the choice of bin size and the placement of the bin edges (for an illustration, see Figure 1 in Wells, 2000).

Analogous to its linear equivalent, a *polar continuous density curve* maps the smoothed density around the circle (Figures 5G and 5H). This representation conveys the same information as the linear density plot but emphasizes the cyclical structure of the data and does not require choosing a wrap-around point. As before, the choice of smoothing parameter strongly influences the result, and continuous density plots are also unsuitable for sparse data and data rounded to a coarse resolution, where density estimates can become unstable. A

key consideration in polar representations is how magnitude is mapped to radial distance. Because the area of a polar region (i.e., a wedge or the area under a smooth curve) grows with the square of its radius, mapping frequency or density directly to radial height causes area to scale quadratically (Figure 5G). This distorts the apparent magnitude because small differences in height translate into large differences in area. As a result, dense regions are spatially overrepresented, while sparse areas are understated and difficult to distinguish. The problem arises for both binned counts and smooth densities. To correct this, radial height can be mapped to the square root of frequency or density (Mardia & Jupp, 2009). This ensures that the area of each sector is linearly proportional to the actual quantity, making sector areas directly comparable across angles. However, this comes at the cost of interpretability of the radial axis: because radial distance now represents the square root of frequency or density, values can no longer be read directly from the radius. Figure 5H shows continuous density plots of the same data as Figure 5G, using an area-proportional representation. For a more extensive discussion on this topic, see Xu and Wang (2020).

A version of polar histograms that avoids a distorted representation of area is *wrapped bar histograms*, which use rectangular bars arranged around a reference circle instead of wedges (Figure 5F). The base of each bar is aligned with the reference circle's circumference, and the radius of the reference circle is set arbitrarily to ensure that the bars fit around its perimeter. This method effectively wraps a linear histogram around a circle, preserving the height of each bar relative to its linear counterpart. This approach avoids the distortions of apparent magnitude typical of wedge-based rose plots, because both height and area are proportional to frequency. However, the gaps between adjacent bars complicate overlaying smooth density curves: to remain consistent with bar heights, density must be mapped to height rather than area; as a result, area under the density curve can misrepresent magnitude, with a disproportionately larger area plotted around longer bars compared to shorter bars. For a more detailed discussion of polar histograms and their design considerations, see Sanderson and Peacock (2020).

Stacked polar histograms (Figure 5I) can be used to show subgroup compositions within each angular bin. Annular sectors corresponding to different groups are layered radially one on top of another. However, this design inherits the same distortion mentioned previously: subgroups farther from the center occupy disproportionately more area, even when counts are equal. This occurs because area increases with the square of the radial distance, so visually comparing sector sizes can give a misleading impression of the underlying proportions. *Stacked linear histograms* show subgroup magnitudes in a clearer way (Figure 5J), even though the stacked representation still makes it difficult to compare the same subgroup across bins or relative subgroup contri-

Figure 5*Visualization of Circular Data in Linear and Polar Spaces.*

Note. (A) Linear histogram of a bimodal circular distribution. Bars show the empirical density estimated from a random sample ($N = 9000$), and the line shows the corresponding theoretical distribution. (B–F) All plots show the same sample data ($N = 90$) drawn from the theoretical distribution shown as the line in (A). (B) Effect of bin widths on the representation of circular data. Both histograms show the same data, but with coarse (blue, 11 bins) and fine (red, 111 bins) binning. (C) Effect of smoothing strength on circular density estimation. Both curves represent the same data with stronger (blue) and weaker (red) smoothing. (D) Empirical cumulative distribution function. (E, F) Polar histograms showing density as a function of angle: polar histogram with wedges (E) and wrapped bar histogram with rectangular bars (F). (G) Polar plot of von Mises distributions (circular normal distributions, see the [Probability Theory and Models for Circular Variables](#) section) with probability density plotted in the radial dimension. Note that dense regions appear exaggerated while sparse regions are understated because area scales with the square of the radius. (H) Polar plot of the same von Mises distributions as in (G) with the square

root of probability density on the radial axis. This area-proportional mapping ensures that enclosed areas are linearly proportional to density. (I, J) Stacked histograms of the same data with two subgroups (blue and red) layered one on top of another in polar (I) and linear circular space (J). Note that the red subgroup is approximately uniformly distributed. However, as area grows with the square of the radius in rose plots, bins of the red subgroup appear larger for bins where density in the blue subgroup is higher in the polar representation (I). (K, L) Representations of cylindrical data with one circular and one Euclidean dimension. (K) Polar scatter plot showing cylindrical data with the circular and Euclidean dimensions represented by angle and distance from the center, respectively. (L) Polar heat map showing a large number of samples drawn from the same distribution as data in (K). Density is represented smoothly by a change in hue across both angular and radial dimensions. Note that this figure is a reproduction of Figure 5 from the [Visualizations of Circular Data](#) section in the main paper.

butions within a bin. For these comparisons, alternatives such as grouped polar bars, unstacked linear histograms, or proportion-normalized plots provide more reliable summaries.

To produce *polar plots*, MATLAB provides the built-in function `polarplot`. In R, a linear plot generated with `ggplot2` (Wickham, 2016) can be transformed into a polar representation by adding `coord_polar()` as a coordinate layer; alternatively, polar plots can be directly created using the `polar.plot` function from the package `plotrix` (Lemon, 2006). In Python, the visualization library `matplotlib` (Hunter, 2007) supports polar plotting via the polar projection in `matplotlib.pyplot.subplot`. For special considerations when creating polar plots of time-series data, refer to Lee Rodgers et al. (2014).

Visualizations of Cylindrical Data

Data consisting of paired circular and linear variables is referred to as cylindrical. Examples include the pairing of movement direction and amplitude, and pairs of visual features like edge orientation and spatial frequency. One way of visualizing cylindrical data is a *polar scatter plot*, where each point is positioned according to its angle around the circle and its distance from the center (Figure 5K). For large datasets or continuous variables, scatter plots can become cluttered and difficult to interpret. In these cases, *polar heatmaps* offer a useful alternative (Figure 5L). Such plots summarize data density or intensity smoothly across both the circular and radial dimensions, making it easier to identify patterns in the data. Cylindrical data can also be represented using two-dimensional Cartesian plots, where one axis corresponds to one period of the circular data displayed in linear form.

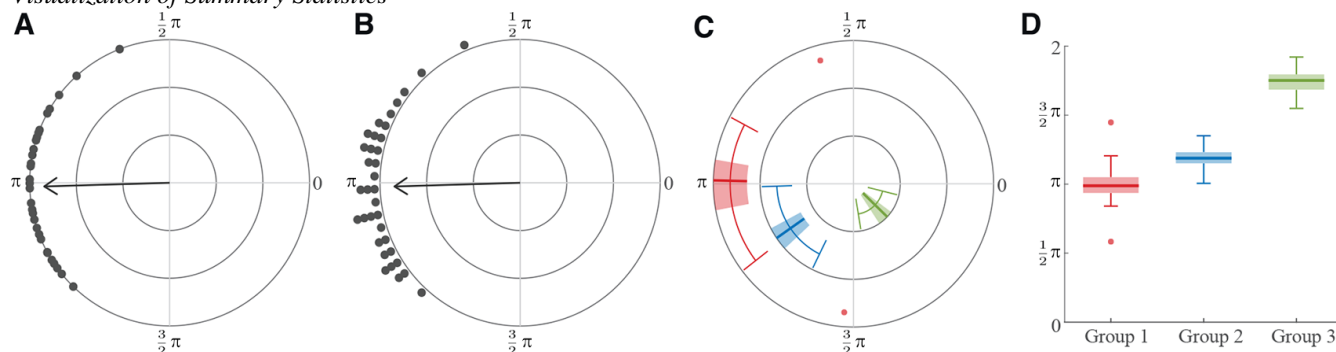
Polar scatter plots can be generated using MATLAB's built-in function `polarscatter`, the polar projection with `scatter` in `matplotlib` (Hunter, 2007) in Python, and `geom_point()` combined with `coord_polar()` in `ggplot2` (Wickham, 2016) or `radial.plot` from `plotrix` (Lemon, 2006) in R. Polar heatmaps can be implemented by first converting values to Cartesian coordinates using the transformations defined in the [Representations of Circular Data](#) section and then applying standard two-dimensional binning and rendering functions familiar from generating heatmaps of Euclidean data.

Visualizing Summary Statistics

While histograms and density plots illustrate the overall distribution of data, visualizations of summary statistics convey key properties, such as the dominant direction in the data or how strongly clustered the observations are. Some of the concepts that we will introduce in this subsection (e.g., mean vector and circular median) are explained in greater detail in the [Descriptive Statistics of Circular Data](#) section. For now, the focus is on how these statistical summaries can be effectively represented in visual form, rather than on details of their calculation.

Rayleigh plots (Figure 6A and 6B) provide a simple and informative way to visualize both individual data points and summary statistics in a single plot. They are related to circular histograms but – in addition to individual observations – also visually summarize central tendency and spread. Instead of grouping observations into bins, Rayleigh plots represent each data point as a mark positioned at its corresponding angle on the circumference (Figure 6A). While typically used for ungrouped data, Rayleigh plots can be adapted for grouped data. In such cases, or when multiple observations share the same angle, instead of being overlaid, points can be stacked radially to show frequency (Figure 6B). Stacking points makes individual observations easier to distinguish compared to the classic Rayleigh plot, where overlapping points can obscure one another. A distinctive feature of the Rayleigh plot is the mean vector, drawn as an arrow from the center of the circle. The angle of this arrow indicates the direction of central tendency, while its length reflects the spread of the data. Thus, it provides an intuitive summary of both the main direction and the consistency of observations around it. For a more detailed mathematical definition of the mean vector, see the [Circular Mean](#) section.

For Euclidean data, a *box plot* typically displays lower quartile (Q1), median, upper quartile (Q3), with additional elements indicating the spread beyond the quartiles and potential outliers. Circular box plots (Figure 6C and 6D) aim to represent similar features but face unique challenges due to the periodic nature of angular data. A critical challenge is that, in circular space, the median and quartiles are not always uniquely defined, particularly in multimodal or uniform distributions (see the [Circular Median](#) section). If these order statistics are not uniquely defined, a circular box plot cannot be uniquely defined either. Despite this limitation, var-

Figure 6*Visualization of Summary Statistics*

Note. (A, B) Rayleigh plots. The arrows indicate the mean resultant vector (Circular Mean section). (A) Rayleigh plot showing 35 individual observations (dark gray circles). Note that it is not possible to see if multiple observations share the same value. (B) Stacked Rayleigh plot of the same data shown in (A), but now binned into 5° intervals, with observations stacked radially within each bin. (C, D) Box plots summarizing three groups of circular observations with circular median (bold line), interquartile range from the 25th to the 75th percentile (opaque box), range of data not including outliers (whiskers), and outlier values (dots). Note that data dispersion is similar for Group 1 (red) and 3 (green), which is apparent in the linear representation (D). In the polar representation (C), the radial distance between the two groups gives the impression that Group 1 has a higher dispersion. Note that this figure is a reproduction of Figure 6 from the [Visualizations of Circular Data](#) section in the main paper.

ious implementations attempt to visualize the angular spread and central range using arcs or radial sectors to represent interquartile ranges, with additional markers for extreme values when identifiable ([supplemental materials on outlier classification and exclusion](#)). Like other visualizations of circular data, circular box plots can be displayed either on a circular (Figure 6C) or linear scale (Figure 6D). The circular format can be particularly useful for comparing multiple circular variables or groups, where box plots are positioned at increasing radii. However, comparing such box plots can be misleading: box plots plotted farther from the origin will be physically longer than those closer in, even if their angular spreads are identical. This can distort the perceived variability and exaggerate or diminish differences between groups. In the example shown in Figure 6C and 6D, the three groups appear more different in terms of spread when displayed on a circular plot compared to a linear one. To support interpretation, circular box plots can include radial grid lines and concentric circles, which serve as visual reference points for comparing angular spread and position across groups plotted at different radii.

The implementation of Rayleigh plots follows the same approach as described for polar scatter plots, with the radius of each point fixed to a constant value for ungrouped data or incremented radially for each observation sharing the same angle for grouped data. The mean vector can be added as an arrow using `quiver` in MATLAB, `annotate with arrowprops` in `matplotlib` in Python, and `geom_segment()` in `ggplot2` in R. For a more detailed discussion of possible ways to implement circular box plots,

refer to Abuzaid et al. (2012) and Buttarazzi et al. (2018).

References

- Abuzaid, A. H., Mohamed, I. B., & Hussin, A. G. (2012). Boxplot for circular variables. *Comput. Stat.*, 27(3), 381–392. <https://doi.org/10.1007/s00180-011-0261-5>
- Buttarazzi, D., Pandolfo, G., & Porzio, G. C. (2018). A Boxplot for Circular Data. *Biometrics*, 74(4), 1492–1501. <https://doi.org/10.1111/biom.12889>
- Chaubey, Y. P. (2022). On nonparametric density estimation for circular data: An overview [Series Title: Forum for Interdisciplinary Mathematics]. In A. SenGupta & B. C. Arnold (Eds.), *Directional Statistics for Innovative Applications* (pp. 351–378). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-1044-9_19
- Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator: L 2 theory [Publisher: Springer Science and Business Media LLC]. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4), 453–476. <https://doi.org/10.1007/bf01025868>
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Lee Rodgers, J., Beasley, W. H., & Schuelke, M. (2014). Graphical data analysis on the circle: Wrap-around time series plots for (interrupted) time series designs. *Multivariate Behavioral Research*, 49(6),

- 571–580. <https://doi.org/10.1080/00273171.2014.946589>
- Lemon, J. (2006). plotrix: A package in the red light district of R. *R-News*, 6(4), 8–12.
- Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics*. John Wiley & Sons.
- Sanderson, D. J., & Peacock, D. C. P. (2020). Making rose diagrams fit-for-purpose. *Earth-Science Reviews*, 201, 103055. <https://doi.org/10.1016/j.earscirev.2019.103055>
- Scott, D. W. (1979). On optimal and data-based histograms [Publisher: Oxford University Press (OUP)]. *Biometrika*, 66(3), 605–610. <https://doi.org/10.1093/biomet/66.3.605>
- Scott, D. W. (2010). Scott's rule [Publisher: Wiley]. *WIREs Computational Statistics*, 2(4), 497–502. <https://doi.org/10.1002/wics.103>
- Sturges, H. A. (1926). The Choice of a Class Interval [Publisher: Informa UK Limited]. *Journal of the American Statistical Association*, 21(153), 65–66. <https://doi.org/10.1080/01621459.1926.10502161>
- Wells, N. A. (2000). Are there better alternatives to standard rose diagrams? *Journal of Sedimentary Research*, 70(1), 37–46. <https://doi.org/10.1306/2DC408FC-0E47-11D7-8643000102C1865D>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Xu, D., & Wang, Y. (2020). Area-proportional visualization for circular data. *Journal of Computational and Graphical Statistics*, 29(2), 351–357. <https://doi.org/10.1080/10618600.2019.1654881>