It is all about the circle:
An inventory of circular data

Bachelorthesis Onderwijskunde (200600207)

Department of Methodology and Statistics

Utrecht University

Jolien Ketelaar

3921514

Irene Klugkist

25/06/2015

Utrecht

Abstract

Circular data are data measured in angles or directions on the unit circle. Because circular data have a periodical nature, the calculation of descriptive and inferential statistics works different than in linear statistics. This study aims at investigating how circular data are described and analyzed in scientific articles and which characteristics circular data have. Therefore, seventeen articles were reviewed and additional analyses were performed on six datasets. In the literature review was found that there is no unity in the descriptions reported and analyses used for circular data. The additional analyses resulted in a lot of variation in the variances and more striking, almost no distributions that fitted the von Mises distribution. This is very interesting because the von Mises distribution is the most commonly used model in circular statistics. Despite the limited number of papers and datasets investigated in this study, the results can be used as guidelines for researchers analyzing circular data. For example, an illustration was included to explain the steps that can be performed when using a test to compare groups on their mean direction. It is proposed that researchers that work with circular data at least report a sample size, a measure of location and a measure of spread and investigate the distribution if researchers want to use inferential statistics.

Keywords: circular data, circular statistics, directional statistics, von Mises distribution

Introduction

Circular data are data measured in angles, directions or orientations (Fisher, 1995; Mardia, & Jupp, 1999). For circular data the unit circle is used as support, as opposed to the real line in linear statistics (Pewsey, Neuhäuser, & Ruxton, 2013). Circular data occur in all sorts of disciplines, such as neuroscience, criminology, social sciences and geoscience. A more specific example of circular data measured in angles is the angle of the head of hearing respondents with impaired hearing on one ear relative to a sound (Brimijoin, McShefferty and Akeroyd, 2012). For example, Figure 1 shows data with the directions, defined with θ, 10°, 30°, 330° and 350°. In this example, the direction 0° is the point where the sound comes from and the four data points, also called vectors, are measured relative to this point. When analyzing circular data it is important to take into account their periodical nature. The linear mean of these four measurements would be 180°, which clearly is incorrect. It is not even near to the observed points, but in the opposite direction as can be seen in Figure 1. The true mean direction $\bar{\theta}$ is 0°, which makes more sense because it is closer to the observation points.

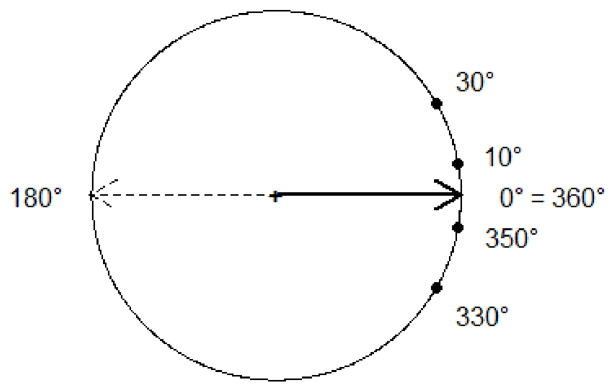
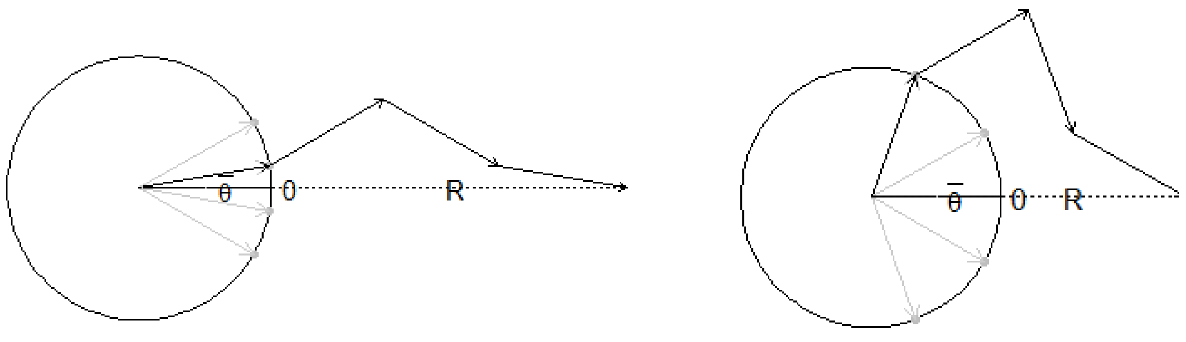


Figure 1. Example of circular data. The points on the circumference of the circle are the observations measured in degrees, the dashed arrow is the linear mean, the solid arrow is the circular mean.

Fisher (1995) described a relatively simple method to obtain the circular mean direction, which is demonstrated in Figure 2A. When the arrows on the circle are added, the line from the middle of the circle to the endpoint of the arrows is called the resultant vector and points out the mean direction. The length of this line is called the resultant length R. The resultant length is related to the variance of the data and the sample size (Baayen et al., 2012). This is visualized in Figure 2B, where four data points are more spread on the circle providing a smaller resultant length.



A. B.

Figure 2. Visualizations of the mean direction and resultant length direction for different variances. The directions of Figure 1 were used for (A), namely 10°, 30°, 330°, 350°. The directions of (B) are 30°, 70°, 290°, 330°. The mean direction of zero is given by $\bar{\theta}$ and the resultant length is given by R.

This study aims to investigate what empirical circular data look like. By taking into account the characteristics of circular data with the development of new circular statistical models and analyses, methods can be developed that fit empirical circular data better. Additionally, the aim is to present the possibilities and preferred methods of working with circular data and to provide guidelines for applied researchers. The first research question of this study is: How is empirical circular data described and analyzed in articles? This question is investigated by assessing the purpose of using circular data and which visualizations, descriptive statistics, distribution checks and inferential statistics are reported in articles. The second research question is: Which characteristics do empirical circular data have? This question is about which characteristics and distributions occur in circular data.

The following section is a theoretical background on circular data. A method section follows and subsequently the results of the literature review, the additional analyses and the illustration of this study are reported. The paper ends with a discussion.

Theoretical background

Different representations

Circular data are described in angles that can be measured in degrees or radians. A radian is “an angle measure related to the arc length of a circle” (Akkoc, 2008, p. 858). One radian is the arc length that is equal to the radius of the circle. Figure 3 shows the relation between the radius and the radian.

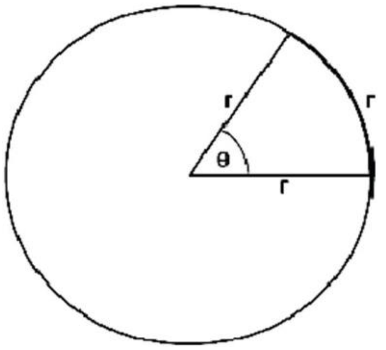
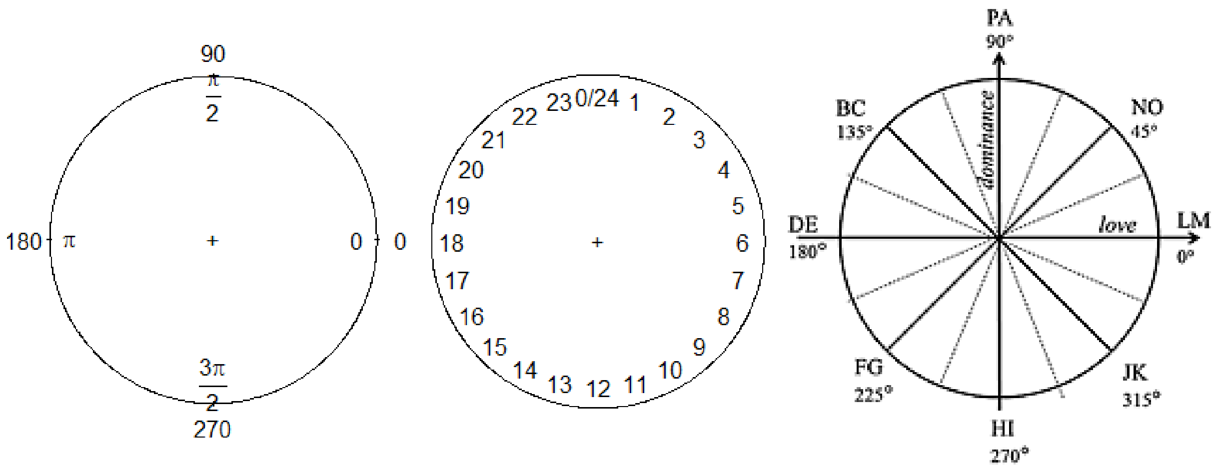


Figure 3. Visualization of one radian. The radius and the radian are indicated with r and have the same length. The angle is described with θ. Retrieved from “Pre-service mathematics teachers’ concept images of radian.” by H. Akkoc, 2008, *International Journal of Mathematical Education in Science and Technology, 39*, p. 858.

In Figure 4A is shown how data can be measured on the circle in radians and in degrees. The different measures are related to each other. One radian is equal to 180/π degrees. The circumference of a circle is 2π radians or 360 degrees. It is also possible to have a negative angle. This depends on the measured direction, clockwise or counterclockwise. A negative angle can be transformed to a positive angle by adding 2π to the concerning angle (Gill, & Hangartner, 2010). Fisher (1995) describes that from a mathematical point of view angles are mostly measured counterclockwise from the starting point, for example (1, 0), on the unit circle.

Besides directions and orientations, time can also be measured on a circular scale. The circle can be divided in, for example, 24 hours as can be seen in Figure 4B. The periodical nature of time can be taken into account when analyzing it circular. When time is used, data are mostly presented clockwise.

A circumplex can also be analyzed circularly. A circumplex model is organized in a two-dimensional space defined by two orthogonal axes or dimensions (Horowitz et al., 2006; König, Onnen, Karl, Rosner, & Butollo, 2015). An example of a circumplex, shown in Figure 4C, is the interpersonal circumplex model, that consists of the axes love and dominance (König et al., 2015). Another interpersonal circumplex is that of Gurtman (2009) where the axes consist of agency and communion. The interpersonal circumplex can be used to make up a circular profile that presents the score on the eight octants. Gurtman (2009) describes how circular data can be analyzed with a circumplex.



A. B. C.

Figure 4. Different representations of circular data in degrees and radians (A), in time (B) and a circumplex (C). (C) retrieved from "Interpersonal Subtypes and Therapy Response in Patients Treated for Posttraumatic Stress Disorder." by J. König, M. Onnen, R. Karl, R. Rosner, and W. Butollo, 2015, *Clinical Psychology & Psychotherapy*, p. 2.

Another distinction that is relevant for circular data is the difference between axial and vectorial data. Vectorial data are data with a specific direction, such as the direction of flying birds (Fisher, 1995). Axial data are data with an undirected line. This means that each direction is equivalent to the opposite direction (Mardia, & Jupp, 1999). An example of axial data are fractures in a rock. Axial data can be transformed to vectorial data, by doubling them and reducing them modulo 360 degrees (Fisher, 1995). For axial data the angles θ and $\theta + \pi$ radians are the same, which can be seen in Figure 5 (Pewsey et al., 2013).

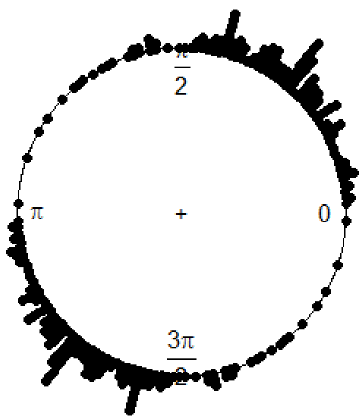


Figure 5. Example of axial data.

Summary statistics

To describe a particular direction θ the sine and cosine can be used (Fisher, 1995). The value of the cosine (θ) can be represented on the x-axis and the value of the sine (θ) can be represented on the y-axis, which can be seen in Figure 6 where an angle of 60° is shown. The coordinates are x = cos(60°) = 0.50 and y = sin(60°) = 0.87. The values of the sine and cosine always lie between -1 and 1.

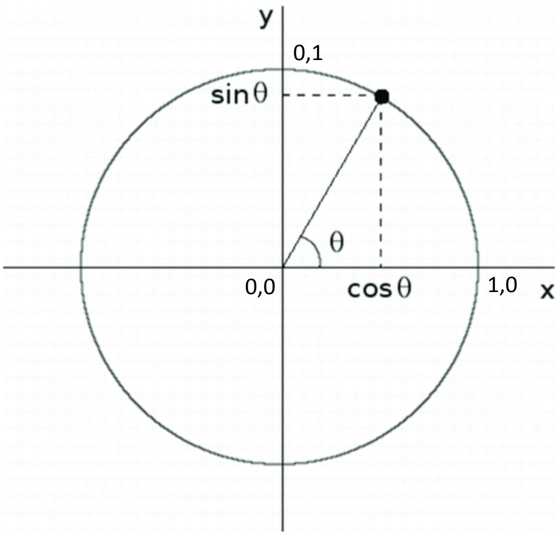


Figure 6. An angle θ represented with corresponding x coordinate cos (θ) and y coordinate sin (θ). Adapted from "A Test of Order-Constrained Hypotheses for Circular Data With Applications to Human Movement Science." by C. Baayen, I. Klugkist, and F. Mechsner, 2012, *Journal of motor behavior, 44,* p. 355.

The sine and cosine coordinates of the data points on the circle can be used for calculation of the mean direction. The mean direction can be computed with the following formula, given a set of angular observations $\theta_1, \dots, \theta_n$. Calculate

$$C = \sum_{i=1}^{n} \cos \theta_i , \qquad S = \sum_{i=1}^{n} \sin \theta_i, \qquad R^2 = C^2 + S^2 \ \ (R \geq 0)$$

(Fisher, 1995, p. 31). The coordinates of the mean direction $\bar{\theta}$ can be found with

$$\cos \bar{\theta} = C/R, \qquad \sin \bar{\theta} = S/R.$$

With the above formula you also find the resultant length, R. The resultant length can be used to calculate the mean resultant length, $\bar{R}$, which is defined by

$$\bar{R} = R/n$$

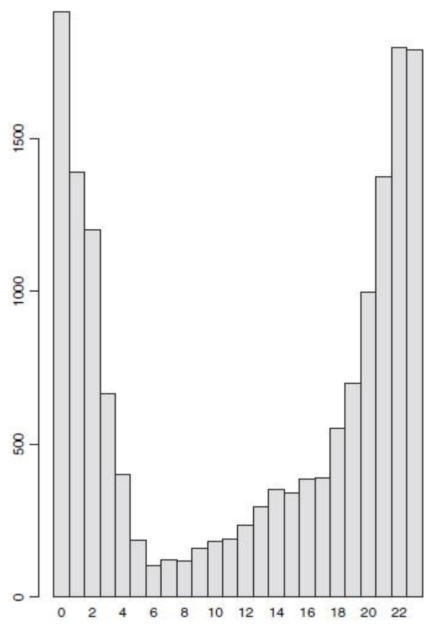
(Fisher, 1995, p. 32). The mean resultant length can be used to find the circular variance, which is calculated with

$$V = 1 - \bar{R}.$$

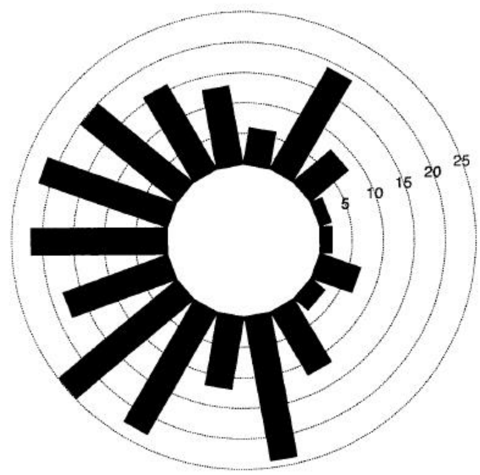
The values of the mean resultant length and variance both lie within the range of 0 to 1 (Fisher, 1995). A variance of 0 and a resultant length of 1 means all vectors are at the same angle. A variance of 1 and a resultant length of 0 does not necessarily mean the data are uniformly distributed around the circle (Fisher, 1995). For example, a bimodal distribution with peaks in opposite directions can result in a variance around 1. For more information about descriptives, such as circular standard devation and skewness, is referred to Fisher (1995). For the calculations of the descriptives in the statistical program R, R Core Team (2014), is referred to Pewsey et al. (2013).

Visualizations

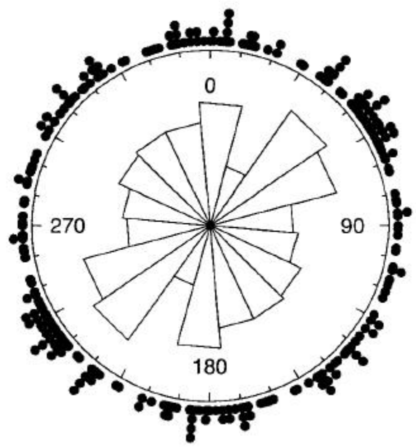
Different visualizations can be used to summarize or show circular data. A common visualization is the raw circular data plot, which is used in Figure 1 - 2 and Figure 4 - 5. In Figure 7A a linear histogram is shown, which is a familiar representation. The disadvantage of a linear display is that the periodical nature cannot be visualized (Pewsey et al., 2013). When a reader does not know the data are periodical, it seems the data in Figure 7A have two peaks. Other circular visualizations exist, such as the angular histogram that is displayed in Figure 7B. In Figure 7C a rose diagram can be seen, in combination with a raw circular data plot. Disadvantageous of the angular and rose diagram is that the bandwidth and starting point have to be chosen arbitrarily and this choice affects the representation of the data (Fisher, 1995). The advantage of these methods is that the data can be viewed easily. The added value of the display in figure 7C is that a raw circular data plot is represented among the rose diagram. Figure 7D shows a density estimate, which gives a more smoothened view of the data (Fisher, 1995). By smoothening the data can easily be viewed how the data are distributed. The type and bandwidth of the density estimate also have to be adjusted, which can also affect the visualization (Pewsey et al., 2013). It depends on the data which plot is most informative at what time.



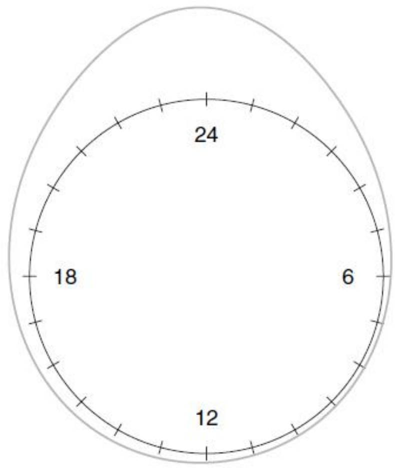
A.



B.



C.

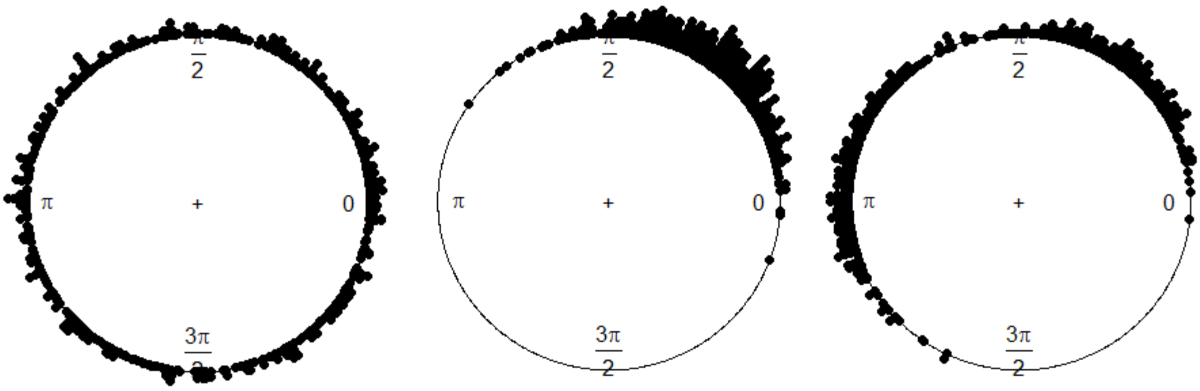


D.

Figure 7. Different visualizations of circular data. (A) Linear histogram, retrieved from "Circular data in political science and how to handle it." by J. Gill and D. Hangartner, 2010, *Political Analysis*, *18*, p. 3. (B) Angular histogram, retrieved from *Statistical analysis of circular data* (p. 19), by N.I. Fisher, 1995, New York: Cambridge University Press. (C) Rose diagram with a raw circular data plot, retrieved from *Statistical analysis of circular data* (p. 20), by N.I. Fisher, 1995, New York: Cambridge University Press. (D) Density estimate, retrieved from "Circular data in political science and how to handle it." by J. Gill and D. Hangartner, 2010, *Political Analysis*, *18*, p. 4.

Distributions and tests for circular data

Circular data can be differently distributed on the circle. For instance, the data can be uniform, unimodal and multimodal. Circular data are called uniform when all directions between 0 and 360 degrees are equally likely, which is also named isotropic (Fisher, 1995). A uniform distribution is visualized in Figure 8. Circular data are unimodal when there is a single cluster of data points found in the data. A common unimodal distribution is the von Mises distribution, which can be considered a circular analogue of the normal distribution. The von Mises distribution is a function of the mean direction (μ) and the dispersion (κ), a measure for the centeredness of the data (Fisher, 1995). A multimodal distribution means there exist two or more clusters in the data, see Figure 8 (Fisher, 1995). Data with two clusters are also called bimodal. To analyze multimodal data complicated techniques are needed, which are not further elaborated on.



Uniform Unimodal Multimodal

Figure 8. Different distributions of circular data.

By testing the uniformity of a dataset, it can be seen if the data are random or have a preference for one or more specific directions. For further testing, it is important to know what the distribution is. To test the uniformity of the dataset, a Rayleigh test can be used (Fisher, 1995). This test uses the mean resultant length to compare randomness to a undefined unimodal distribution. For other tests of uniformity and the application of those tests in the statistical program R, R Core Team (2014), is referred to Pewsey et al. (2013). For other tests and distribution fitting is referred to Fisher (1995) and Mardia and Jupp (1999).

Different modelling approaches

A distribution defined on the circle, such as the von Mises distribution, is called the intrinsic approach. Two different approaches, the wrapped and the embedded approach, for analyzing circular data exist.

The wrapped approach wraps an existing distribution on the circle modulo 2π. For example, with the wrapped normal circular distribution the normal distribution is wrapped on the circle (Pewsey et al., 2013). This means there are more layers of the normal distribution at a specific point of the circular distribution. The height of the density of a specific angle is derived from the sum of the different wraps. For a theoretical explanation of the wrapped approach is referred to Ravindran and Ghosh (2011). In Pewsey et al. (2013) a summary of different intrinsic and wrapped circular models and possible applications can be found.

The embedded approach assumes that the circular variable has a projected bivariate linear distribution. This model treats the given data points on the circle as projections of unobserved, underlying data points in a two-dimensional space. To use this model the underlying bivariate data have to be estimated, while using a latent variable representing the distance between the underlying bivariate data and the origin. For further explanations of the embedded approach is referred to Nuñez-Antonio, & Gutiérrez-Peña (2005).

Methods

Search for articles and descriptions of datasets

For the inventory of circular data, seventeen articles were explored. The criterion for including articles was that they contained a description of an empirical circular dataset Different approaches were followed to find articles. A search was conducted in the search engine Google Scholar, where the keywords ‘circular’ and ‘data’ were used in combination. Also the word ‘social’ was included as a keyword to find more articles about topics of the social sciences. When articles were found, a snowballing technique was used, that is, articles that were referred to were also examined. To find more applied papers, citation information was used, which means that articles that cited books and articles about circular statistics were included. Papers that were already accessible to the researcher were included likewise.

Literature review

When articles were found, a search was conducted for descriptions of empirical circular data. The reported characteristics of the described datasets were analyzed using open coding. The reviewed articles gave a varied image of circular data. Although the review is illustrative and not systematic, it is interesting to summarize the used descriptions of circular data. At first the articles are described by the discipline it was found in and by the purpose of the article. After that, the results are summarized by which plots, graphs and other visualizations are used in the article and which descriptive statistics are used. The results section of the literature review ends with which tests and modelling assumptions are performed and described in the papers.

Additional statistical analyses

To answer the second research question, additional statistical analyses were performed on six datasets. Some of the analyzed datasets were already available to the researcher, some of them were found in statistical packages and others were retrieved. To retrieve datasets the first author of an article was sent an e-mail to ask for the used dataset. Descriptions of two additional datasets were retrieved from the reviewed articles and also used. More information about the analyzed datasets can be found in Appendix B.

The datasets were analyzed in the statistical program R to gain further insight in the descriptives and the distribution of the data (R Core Team, 2014). A function was developed to calculate multiple descriptives at once for multiple datasets and to convert the data to radians. Descriptives that were investigated were the sample size, mean direction, mean resultant length, variance and the circular standard deviation. The Rayleigh test was performed for the datasets, as well as Watson test for the von Mises distribution. From the two datasets retrieved from the reviewed articles, some characteristics had to be calculated, transferred to radians or rounded off to two decimals. All results are reported in positive angles.

Illustration

The use of circular statics is illustrated by demonstrating the steps that can be taken to compare independent groups on their mean direction, for which an example of music listening behavior was used**.** The Music dataset was used by Herrera et al. (2010) to investigate music listening behavior of music tracker users. The data were collected and placed online by Celma (2010). Three users were randomly selected from the raw dataset, namely user 6, 35 and 44. In Appendix B more information can be found about the dataset and the users.

At first, the distributions of the dataset were examined by visualizing them. Density estimates were plotted for all three users and the spread was further examined by comparing it with a von Mises distribution using a QQ-plot. To compare the distribution of a user with a von Mises distribution, first the von Mises distribution had to be fitted to the data to find a useful μ and κ. This was done with maximum likelihood estimation (MLE). To make the QQ-plots, the empirical data were plotted against the estimated von Mises distributions. For the MLE and the QQ-plot the R-script from Pewsey et al. (2013, p. 102-105) was used. The three users from the Music dataset were compared on their mean direction with the Watson's large-sample nonparametric test, because the data did not fit the von Mises distribution (Pewsey et al., 2013, p. 134-135).

Results

To report the results, at first the literature review is reported. It is described in which disciplines circular data were found and the purpose of the articles is explained. After that, the results are summarized by the used visualizations, the reported descriptive statistics, the distribution assumptions and checks and finally the inferential statistics that were used. In Appendix A, a table is shown where it can be seen which articles reported which descriptions of circular data. After the results of the literature review the outcomes of the additional analyses are explained and the results of the illustration are reported at the end.

Literature review

Purpose and occurred disciplines. The investigated articles differed in purpose. Seven of the seventeen found articles were methodology papers, in those articles circular data were used as illustration of the used or developed methods. The other ten articles were applied papers, where existing methods were applied on research with circular data.

Circular data were found in a lot of different disciplines. The methodology articles used illustrations from neuroscience, criminology, psychology, computer science, physics and political science. The applied papers were from zoology, geography, neuroscience, psychology, hearing science and musicology. All three articles about neuroscience worked with the orientation of neurons (Berens, 2009; Tolias et al., 2007; Maldonado, Gödecke, Gray, & Bonhoeffer, 1997). For instance, Maldonado et al. (1997) compared the orientations of neurons in two different domains in the brain. Three articles studied movement directions, namely the movement of animals in different conditions and the commute of people to their work. (Cochran, Mouritsen, & Wikelski (2004); Bulbert, Page and Bernal, 2015; Corcoran, Chhetri, & Stimson, 2009) Three of the articles, all from different disciplines, analyzed time patterns. Brunsdon and Corcoran (2006) and Herrera, Resa and Sordo (2010) analyzed patterns per day and per week and used examples from criminology and musicology, respectively. Time patterns were also analyzed by Kubiak and Jonas (2007). They introduced circular statistics to detect patterns in time with a procedure called the maximization of mean vector length and contrasted this method with time-series analyses in psychology. Four other articles came from psychology, from which two of them used a circumplex (Mechsner, Stenneken, Cole, Aschersleben, & Prinz, 2007; Kirschner, & Tomasello, 2009; König et al., 2015; Mainhard, Brekelmans, Den Brok, & Wubbels, 2011). As an example, König et al. (2015) used a circumplex model with four quadrants to investigate therapy responses in patients with posttraumatic stress disorder, where patients in the four quadrants were compared on their responses. Brimijoin et al. (2012) did research on the head movements of listeners with hearing impairment on one ear. They compared the movement of the head in these groups and in conditions with noise from different directions. A more unusual article was that of Hanbury (2003) about computer science. He put the color spectrum on the circle and used circular statistics to summarize color images, where a correction for the saturation of the image was made. Finally, two other methodological papers used illustrations from physics and political science (Van Doorn, Dhruva, Sreenivasan, & Cassella, 2000; Gill, & Hangartner, 2010).

Visualizations of observed data. Almost all articles used some sort of visualization to display circular data, which resulted in a lot of variation. In the theoretical background was spoken about certain circular representations, such as the raw circular data plot, and about a linear display, the linear histogram. Besides the visualizations discussed in the theoretical background, other displays were found. In Table 1 the most used representations are stated with corresponding frequencies of articles that used them.

About a third of the articles used a combination of linear and circular visualizations and the rest only used one of the two displays. Interesting was that solely circular images were used most in the applied papers, while the methodological studies mostly used linear visualizations or a combination.

Of the circular displays the raw circular data plot was used most, namely in about a third of the articles. All of those articles included the mean resultant length or median resultant length in their circular plot. The angular histogram, rose diagram and density estimate were also found in the papers, in two articles each. Besides the visualizations introduced in the theoretical background, one other circular-like display was found, namely the map. A map is a representation where the directions are visualized with arrows or colors, this sort of display was used in about one fifth of the articles. One of those used a choropleth map to display circular mean directions, dispersions and other descriptives (Tolias et al., 2007). A choropleth map is a map where areas are colored in proportion to the measurement.

From the linear visualizations linear histograms were used most often, namely in around half of the articles. Unexpected visualizations that were found, were the linear plot, linear graph and boxplot. Approximately one fifth of the articles used linear graphs. It was mostly used to represent changes in angles over time or to show estimated von Mises functions.

Table 1. Most often found visualizations in frequencies of the seventeen found articles that used the display.

	Linear histogram	**Raw circular data plot**	**Linear graph**	**Linear plot**	**Map**
Reported	9	6	5	4	3

Visualizations were used not solely to summarize the data, but also to display descriptives of the data. For example, Kirschner and Tomasello (2009) used boxplots compare mean deviations and resultant lengths per condition. Likewise, Mechsner et al. (2007) used a linear histogram to visualize the circular standard deviation for the different conditions.

Some of the articles used other visualizations for information besides relatively simple descriptives of circular data. An example is a density estimate of Brunsdon and Corcoran (2006) to view $\varphi(\theta)$, which is here the probability of something happening at a specific location at a certain time. In none of the articles a clear explanation was given why they used a specific visualization.

Descriptive statistics. A lot of variation is found in the descriptives that are reported in the articles. The most often reported descriptives are the sample size, measures of location and measures of spread. In Table 2 a short summary is shown of the most used descriptives.

Approximately three-quarters of the articles described the sample size in their article. It was not clear in all of the cases how much data points were exactly on the circle. To interpret the results of a study, it is important to know what the sample size is, defined as data points on the circle.

More than half of the reviewed articles reported a measure of location. Most of them reported the mean direction. A median direction was reported in two papers (Bulbert et al., 2015; Brimijoin et al., 2012). Bulbert et al. (2015) explained the use of the median direction by stating that their data were not normally distributed, which makes the median a better measure of location. The two articles that used the median were not from the same discipline.

Furthermore, slightly less than half of the articles reported a circular measure of spread. About a quarter of the articles reported the mean resultant length. Related to the mean resultant length is the circular variance, which was described in only two articles. The circular standard deviation was found in three articles. Maldonado et al. (1997) described a mean orientation scatter and a mean orientation range with corresponding deviations as a measure of spread. An orientation scatter is a measure related to the standard deviation, but is independent of sample size. Maldonado et al. (1997) probably chose to use this because they have variations in sample sizes.

Besides circular measures of location and spread, linear measures were found. Both of the articles that reported linear measures worked with a circumplex, which can explain the use of linear measures, because a circumplex consists of two linear axes. Mainhard et al. (2011) described their data solely linear and used a linear mean and standard deviation per axis. König et al. (2015) described a linear variance and furthermore they included a structural summary as descriptive. This structural summary describes a circumplex with a sinusoid as base (Wright, Pincus, Conroy, & Hilsenroth, 2009). The structural summary consists of the amplitude, elevation, displacement and the goodness-of-fit statistic R^2 of the data. For more information about this method is referred to Wright et al. (2009).

A measure that was reported in about one fifth of the articles was the 95 percent confidence interval. In one of these articles, also the 99 percent confidence interval was reported (Cochran et al., 2004). In none of those articles an explanation was given why they reported a confidence interval and all articles were from different disciplines.

Across the reviewed articles no straight forward steps or standard procedures for describing the data were found. A lot of variation occurred in reported descriptives, from which some of them were only described in one paper. Examples of these descriptives were the upper and lower quartiles, modus angle, skewness and kurtosis. In Appendix A the complete picture of descriptives can be found.

Table 2. Most often found descriptives in frequencies of the seventeen found articles that used the measure

	Sample size	Mean direction	Mean resultant length	Circular standard deviation	Confidence interval 95%
Reported	12	8	5	3	3

Distribution assumptions and checks. In around two-third of the articles the distribution was investigated. The majority of the researchers tested if the distribution was uniform. In some of the articles it was not clear which test they used to test uniformity or they reported that a test was performed, but the results were not or only partly described. The article of Herrera et al. (2010) is one of those articles. The researchers only reported percentages with corresponding standard deviations of the subsamples that were non-uniformly distributed, as a result of the uniformity tests on their data. In general, distributions were most often examined with the Rayleigh test, namely in four of the articles. Other reported tests for uniformity, such as the Omnibus test, were only found in one article.

Other checks that were performed concerning the distribution of the data were aimed at investigating if the data fitted a von Mises distribution, for example with the Watson test. Other methods were aimed at fitting von Mises or Gaussian distributions using maximum likelihood estimation with least squares. Two of the three articles that fitted a distribution were methodological papers.

Some other tests related to the distribution occurred only once. The tests that were found compared distributions or were used to measure normality. Bulbert et al. (2015) used the most checks in their article. They used the Moore test to investigate if the distribution of the samples was the same, the Watson's two-sample test of homogeneity to compare distributions and a circular correlation as a correction.

More than ten different test were found that investigated or compared distributions. The majority of those tests only occurred in one article, which indicates a lot of variation. Most of the articles that inspected the distribution also performed inferential statistics. Sometimes inferential statistics were performed without reporting results of assumption checks.

Inferential statistics. About two-third of the articles used inferential statistics. Most of the tests were aimed at comparing groups or conditions. Less than half of the articles that performed inferential statistics used a statistical test specifically for circular data. The test that was most used was the Watson-Williams test, this test occurred in two articles (Berens, 2009; Cochran et al., 2004). Other circular tests that were found were, for example, a circular two-way ANOVA and a Bayesian regression. The circular tests that were not aimed at comparing groups or conditions, were only found in methodological papers.

More than a half of the articles that used inferential statistics, used tests for linear data. Most used were the ANOVA and the t-test. Mechsner et al. (2007) explicitly described that they performed a linear ANOVA because a circular ANOVA was not available for their complex repeated-measures design with multiple conditions. ANOVA's and t-tests were also used when circular data were not the direct outcome measure, but the mean resultant length, standard deviation or range. For example, Kirschner and Tomasello (2009) used the mean resultant length and performed a correction on it to be able to use a linear repeated-measures ANOVA, because there was no circular test available.

A few other linear tests occurred in the articles, such as the Wilcoxon rank sum test by Tolias et al. (2007). The other inferential tests that were found, occurred only once. From the studies that did not use inferential statistics, some performed analyses but did not use a formal test to make inferences.

Additional statistical analyses

Additional statistical analyses were performed to further examine the characteristics of the data, such as the variance and the distribution. In Appendix B more information can be found about the used datasets. From the analyzed datasets all characteristics are reported in radians and summarized in Table 3. It can be seen that the mean resultant length, variance and standard deviation are related measures. The variances that occurred in the additional analyses are diverse and range from very small to quite large. The datasets with the smallest variance and standard deviation are the Cake datasets. This probably has to do with the restricted range of these datasets from zero to 90 degrees. When the data are centered at only a small part of the circle, such as with the Cake datasets, it may not be necessary to use a circular test. Because the two most extreme values do not come closer to each other, which is the case if the data are more spread, a linear test will likely give the same results.

The most spread dataset is Circling Condition 17, which has a circular standard deviation of 2.47 radians. This means the data are spread across a large part of the circle. As stated in the theoretical background, a maximum variance does not necessarily have to mean that the data are uniformly distributed. The large variance of Circling Condition 17 can also be due to a bimodal distribution. A Rayleigh test was performed to further examine the datasets. As can be seen in Table 3 none of the data are uniformly distributed.

An interesting finding is that almost all of the datasets are not von Mises distributed, according to the Watsons test that was performed. The test statistics and p-values can be found in Table 3. This is an interesting finding, because the von Mises model the most commonly used model in circular statistics. Only the dataset Fisher 5 is von Mises distributed. This dataset contains axial data, but that is probably not an explanation for the distribution of the data. In the next section an illustration of circular statistics is described, where the distributions of the Music datasets are further explored with a QQ-plot.

Table 3. Descriptives and distribution checks of the investigated datasets

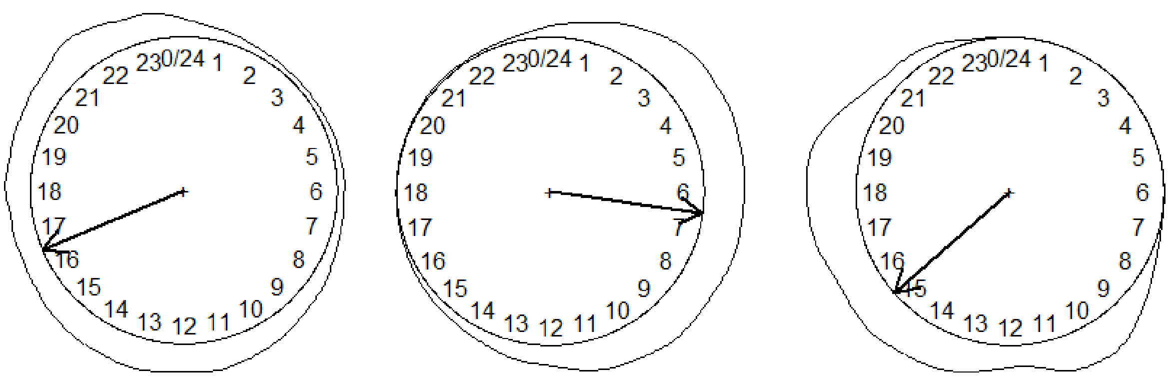
	Sample size	Mean direction	Mean resultant length	Variance	Standard deviation	Rayleigh test for uniformity	Watson test for von Mises
Cake							
Recipe A	90	0.58	0.99	0.01	0.13	0.99**	0.19**
Recipe B	90	0.55	0.99	0.01	0.15	0.99**	0.22**
Recipe C	90	0.55	0.99	0.01	0.15	0.99**	0.21**
Fisher 5[×]	60	0.35	0.24	0.76	1.69	0.24**	0.03
Fisher 7	100	3.20	0.61	0.39	0.99	0.61**	0.10**
Fisher 12	15	3.00	0.63	0.36	0.95	0.63**	0.08*
Music							
User 6	29021	4.31	0.27	0.73	1.62	0.27**	17.80**
User 35	9458	1.72	0.56	0.44	1.07	0.56**	6.22**
User 44	13169	3.99	0.59	0.41	1.03	0.59**	15.79**
Pittsburgh crimes	15831	5.92	0.55	0.45	1.10	0.55**	14.87**
Neuron orientation							
Neuron 1	643	2.36	0.26	0.74	1.65	**	
Neuron 2	353	5.38	0.73	0.27	0.79	**	
Neuron 3	141	5.45	0.81	0.19	0.66	**	
Circling							
Condition 14		0.19	0.92	0.08	0.40		
Condition 17		4.42	0.05	0.95	2.47		

Note. For the Neuron orientation and Circling datasets are the descriptions of the article used, the other datasets were available and analyzed. *$p < 0.05$ ** $p < 0.01$, significance means the sample is not distributed like the tested distribution [×]The Fisher 5 dataset is axial, therefore a transformation is performed following Fisher (1995).

Illustration

To illustrate working with circular data and using circular statistics, a possible procedure to compare three samples on their mean direction is explained. The example used, is the Music dataset, which contains data of music listening behavior of three music tracker users.

A logical first step is examining the data, which can be done by visualizing it. Because the Music dataset contains a lot of data, namely thousands to ten thousands of data points, a density estimate is chosen as visualization. With a density estimate the peaks and distribution can be viewed easily despite the large sample size. Different bandwidths were tried and a bandwidth of 1/100 of the sample size was chosen because big differences were still visualized, but the small differences not anymore. By choosing the same proportion of bandwidth per user, the displays can easily be compared. In Figure 9 the density estimates can be seen, including the means of the samples. It can be seen that the data, for all three datasets, are wide spread across the circle. User 6 has the most data on the opposite direction of the mean direction in comparison with the other two users. User 35 has a smooth density that seems a bit skewed, which can be seen in Figure 9. User 44 has no data points on the opposite direction of the mean direction and the density estimate has three small fluctuations. To represent the data other visualizations can be appropriate, but a circular display is preferred because the periodical nature of the data can be taken into account. To further explore the data some descriptives can be calculated, such as the mean or the median, and a measure of spread, such as the circular standard deviation.



User 6 User 35 User 44

Figure 9. Density estimates of the three users from the Music dataset, the arrow indicates the mean direction

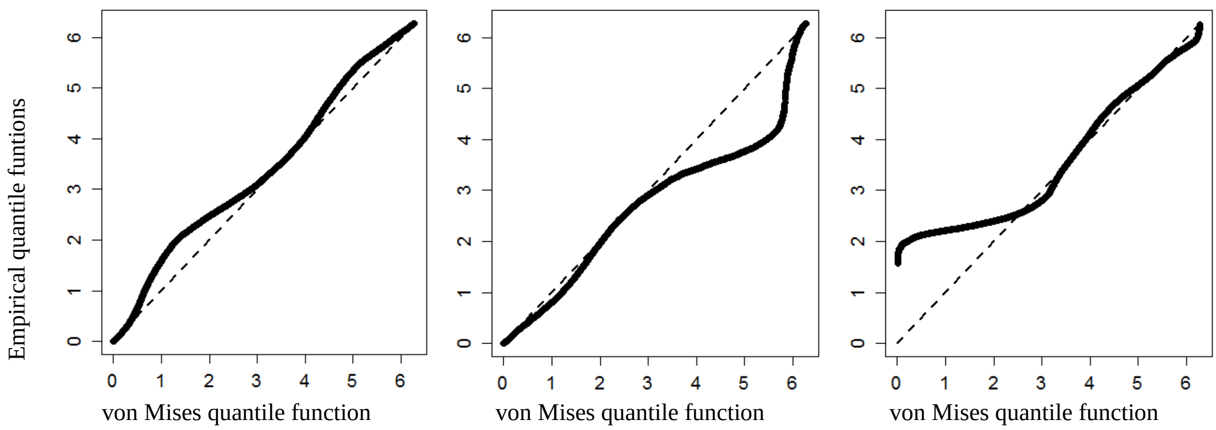
To further explore the distribution of the Music data some tests can be performed, such as the Rayleigh test for uniformity and the Watsons test for a von Mises distribution, see Table 3. If these tests both give significant results, a QQ-plot can be used to compare the empirical distribution of a user with a von Mises distribution, which was done here. To compare the distribution, first the von Mises distribution had to be fitted to the data to find a useful μ and κ. In Table 4 the estimates can be seen.

Table 4. Results of maximum likelihood estimation for fitting a von Mises distribution

	$\hat{\mu}$	**Standard error of $\hat{\mu}$**	**Bias-corrected κ**	**Standard error of κ**
Music 6	4.31	0.02	0.55	0.01
Music 35	1.72	0.01	1.37	0.02
Music 44	3.99	0.01	1.46	0.02

The distributions with the estimated μ and κ were used to compare with the empirical distribution of the concerning users in a QQ-plot, which can be seen in Figure 10. The distribution of the data from User 6 seems to fit the von Mises distribution best, because the line is almost linear, although it seems like the data have a bimodal tendency. It can be seen that the data of User 35 fit the von Mises distribution quite good in the beginning, but later, where the von Mises distribution is more spread out, the empirical data are more clustered together. Figure 10 shows that the distribution of User 44 data has less data on the small quantiles than a von Mises distribution with the corresponding parameters. As reported in Table 4 all estimated κ's are below two, which means it is not easy to establish whether the model fits the data, according to Fisher (1995). In the literature review κ's below two were also found. The researchers of the articles did not report anything about this low κ and did not act differently. The reason can be that there was no alternative available or known to the researchers.

Because the Watson tests gave significant results and because the QQ-plots were not linear, it is concluded that the three datasets are not von Mises distributed. Other model fitting procedures might be more appropriate for these datasets.



User 6 User 35 User 44

Figure 10. Q-Q plots for the von Mises distribution fitted to the data of the Music dataset.

Because it is concluded that the datasets are not von Mises distributed, a non-parametric test is used to compare the three users. The test that was performed is the Watson's large-sample nonparametric test. Results indicate that the groups have a significant different mean direction, $Y_g = 20749.64$, $p < 0.01$. This is not a surprising result, because the mean direction of user 35 deviates a lot from the mean direction of the other two users. To investigate which users differ exactly in their mean direction a post hoc test could probably be performed, but it was not known to the researcher if such a test existed. Because it is unknown how the data came about and how the hours of the different users relate to each other, no conclusions can be made with respect to the content of the Music data. The primary use of this example was illustrating circular statistics.

Discussion

The first purpose of this study was to investigate how empirical circular data are described and analyzed in articles. It has been found that circular data occurs in many different disciplines and is used in methodological and applied papers. It was found that many different visualizations were used to display circular data. The circular representation that occurred most was the raw circular data plot, where mostly a measure of location was included. Visualizations were used to summarize the data, but also to display descriptives of the data, such as the circular standard deviation. No unity was found in the reported descriptives, not per discipline or for methodological or applied papers. An interesting finding was that only about half the articles reported some measure of spread. This is striking, because when reporting linear descriptives, a standard deviation is reported most of the time. A possible explanation can be that the researchers are not familiar with circular statistics and do not know which descriptives are important to report. From this research is concluded that to give an image of how circular data look like, it is important to at least report a sample size, a measure of location and a measure of spread. Adding a visualization can also be useful to give a summary of the data, for example a raw circular data plot. A circular visualization of the data is preferred, because the periodical nature of the data can be displayed.

Most of the articles tested the distribution of their data, but the used tests differed a lot. In most of the cases it was tested if the distribution was uniform with a Rayleigh test. In comparison, in the additional analyses of this study the distributions were checked with a Rayleigh test and a Watson test for von Mises. A striking finding was that most of the datasets were not von Mises distributed, because this distribution is used a lot for modelling approaches and tests in circular statistics. In the illustration QQ-plots were made to further investigate the distribution of the datasets and relatively divergent distributions were observed. The question is how representative it is that the datasets did not fit the von Mises distribution, but it is still a fascinating finding because only one dataset of the additional analyses fitted.

Besides tests for the distribution, inferential statistics were used in the reviewed articles. The reported tests were sometimes circular, but also linear methods had been used because circular methods were not available for that type of research design. This sometimes resulted in a linear test on an outcome measures of circular data, for example on the mean resultant length. Further research need to be done to see if this approach is most obvious or best. Because circular analyses were not available for complex designs, further research can be about developing new circular tests. Because of the huge variation in used methods, it is important to formulate guidelines for working with circular data, such as first investigating the distribution and describe the tests used with clear corresponding results.

The papers that were looked into for circular data were a convenient selection. It is not known if the used articles are representative of all circular data. Despite the fact that the review is not systematic, it gives an image of the variation of the choices researchers make when it is about circular data. It would be interesting to do a systematic review on circular data to get a complete view of how circular data are described.

To conclude, there is major variation in used methods and descriptions when using circular data. Besides, variation is found in the characteristics of circular data, such as the variances and distributions. Guidelines for analyzing circular data are needed to bring more unity and clearness in the research. The illustration in this study hopefully adds to this.

Analyzing data circularly has an advantage because it takes the periodical nature of the data into account. If the data are more spread around the circle, for example distributed over more than half of the circle, it is important to take into account that the two most extreme values come closer to each other instead of further away. Circular statistics provide for including this phenomenon.

References

Agostinelli, C., & Lund, U. (2013). R package 'circular': Circular Statistics (version 0.4-7). Retrieved from http://cran.r-project.org/web/packages/circular/circular.pdf

Akkoc, H. (2008). Pre-service mathematics teachers' concept images of radian. *International Journal of Mathematical Education in Science and Technology, 39*(7), 857-878.

Baayen, C., Klugkist, I., & Mechsner, F. (2012). A Test of Order-Constrained Hypotheses for Circular Data With Applications to Human Movement Science. *Journal of motor behavior, 44*(5), 351-363.

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software*, *31*(10), 1-21.

Bolker, B., Bates, D., Maechler, M., Walker, S., Christensen, R. H. B., Singman, H., & Dai, B. (2014). The lme4 package (version 1.1-7). Retrieved from http://cran.r-project.org/web/packages/lme4/lme4.pdf

Brimijoin, W. O., McShefferty, D., & Akeroyd, M. A. (2012). Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. *Hearing research*, *283*(1), 162-168.

Brunsdon, C., & Corcoran, J. (2006). Using circular statistics to analyse time patterns in crime incidence. *Computers, Environment and Urban Systems*, *30*(3), 300-319.

Bulbert, M. W., Page, R. A., & Bernal, X. E. (2015). Danger Comes from All Fronts: Predator-Dependent Escape Tactics of Túngara Frogs. *PLoS ONE, 10*(4): e0120546. doi: 10.1371/journal.pone.0120546

Celma, Ò. (2010). *Last.fm Dataset - 1K users.* Retrieved from http://www.dtic.upf.edu/~ocelma/MusicRecommendationDataset/lastfm-1K.html

Cochran, W. W., Mouritsen, H., & Wikelski, M. (2004). Migrating songbirds recalibrate their magnetic compass daily from twilight cues. *Science*, *304*(5669), 405-408.

Corcoran, J., Chhetri, P., & Stimson, R. (2009). Using circular statistics to explore the geography of the journey to work*. *Papers in regional science*, *88*(1), 119-132.

Fisher, N. I. (1995). *Statistical analysis of circular data.* New York: Cambridge University Press.

Gill, J., & Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis*, *18*(3), 316-336.

Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass, 3*(4), 601-619.

Hanbury, A. (2003). Circular statistics applied to colour images. In *8th Computer Vision Winter Workshop*, 91(1-2), 53-71.

Herrera, P., Resa, Z., & Sordo, M. (2010). Rocking around the clock eight days a week: an exploration of temporal patterns of music listening. In *1st Workshop On Music Recommendation And Discovery (WOMRAD), ACM RecSys, 2010, Barcelona, Spain*.

Horowitz, L. M., Wilson, K. R., Turan, B., Zolotsev, P., Constantino, M. J., & Henderson, L. (2006). How interpersonal motives clarify the meaning of interpersonal behavior: A revised circumplex model. *Personality and Social Psychology Review, 10*(1), 67-86.

Kirschner, S., & Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *Journal of experimental child psychology, 102*(3), 299-314.

König, J., Onnen, M., Karl, R., Rosner, R., & Butollo, W. (2015). Interpersonal Subtypes and Therapy Response in Patients Treated for Posttraumatic Stress Disorder. *Clinical Psychology & Psychotherapy*. doi: 10.1002/cpp.1946

Kubiak, T., & Jonas, C. (2007). Applying circular statistics to the analysis of monitoring data: Patterns of social interactions and mood. *European Journal of Psychological Assessment*, *23*(4), 227.

Last.fm (2015). *Discover more music.* Retrieved from http://www.last.fm/

Mainhard, M. T., Brekelmans, M., den Brok, P., & Wubbels, T. (2011). The development of the classroom social climate during the first months of the school year. *Contemporary educational psychology*, *36*(3), 190-200.

Maldonado, P. E., Gödecke, I., Gray, C. M., & Bonhoeffer, T. (1997). Orientation selectivity in pinwheel centers in cat striate cortex. *Science*, *276*(5318), 1551-1555.

Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics*. John Wiley & Sons.

Mechsner, F., Stenneken, P., Cole, J., Aschersleben, G., & Prinz, W. (2007). Bimanual circling in deafferented patients: Evidence for a role of visual forward models. *Journal of neuropsychology, 1*(2), 259-282.

Nuñez-Antonio, G., & Gutiérrez-Peña, E. (2005). A Bayesian analysis of directional data using the projected normal distribution. *Journal of Applied Statistics, 32*(10), 995-1001.

Pewsey, A., Neuhäuser, M., & Ruxton, G. D. (2013). *Circular statistics in R.* New York: Oxford University Press.

Ravindran, P., & Ghosh, S. K. (2011). Bayesian analysis of circular data using wrapped distributions. *Journal of Statistical Theory and Practice, 5*(4), 547-561. doi: 10.1080/15598608.2011.10483731

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Tolias, A. S., Ecker, A. S., Siapas, A. G., Hoenselaar, A., Keliris, G. A., & Logothetis, N. K. (2007). Recording chronically from the same neurons in awake, behaving primates. *Journal of neurophysiology*, *98*(6), 3780-3790.

Van Doorn, E., Dhruva, B., Sreenivasan, K. R., & Cassella, V. (2000). Statistics of wind direction and its increments. *Physics of Fluids*, *12*(6), 1529-1534.

Wright, A. G., Pincus, A. L., Conroy, D. E., & Hilsenroth, M. J. (2009). Integrating methods to optimize circumplex description and comparison of groups. Journal of Personality Assessment, 91(4), 311-322.

Appendix A

Table A1. Summary of the seventeen found articles that described empirical circular data

Article	Purpose of the article	Visualizations of observed data	Descriptive statistics	Distribution assumptions and checks	Inferential statistics
Berens (2009)	Methodological paper Explaining circular data and methods with an example of the orientation tuning of neurons. This example comes from Tolias et al. (2007).	- Raw circular data plot with mean resultant length and a plotted area	- Sample size - Mean direction - Variance - Circular standard deviation - Angular deviation - Skewness - Kurtosis - Confidence interval 95%	- Rayleigh test - Omnibus test	- Watson-Williams test - Post hoc Watson-Williams test to make pairwise comparisons
Brimijoin et al. (2012)	Applied paper Studying the way listeners that are impaired on one ear move their head while listening to a sound in different noise conditions.	- Linear graph - Linear plot - Linear histogram	- Sample size - Median direction - Mode direction - Upper and lower quartile	- Shapiro-Wilks test of normality	- Circular two-way ANOVA
Brunsdon, & Corcoran (2006)	Methodological paper Analyzing time patterns in circular data with an example of criminal damage and public disorder.	- Angular histogram - Density estimate - Linear histogram	- Sample size - Mean direction - Dispersion parameter	- Von Mises distribution assumed, doubt stated because data seems multimodal	- Watson test - Non-parametric τ-test (developed and used) - Formula for estimating $\varphi(\theta)$ (developed)
Bulbert et al. (2015)	Applied paper Examining escape responses of frogs against bat and snake models.	- Raw circular data plot with median resultant length (including images)	- Sample size - Median direction - Mean resultant length* - Range - Upper and lower quartile	- Watson test for uniformity - Watson test for von Mises - Watson's two-sample test of homogeneity (to compare distributions) - Circular correlation - Moore's test (to test for repeatability)	- Binomial test (to compare directions)

Article	Purpose of the article	Visualizations of observed data	Descriptive statistics	Distribution assumptions and checks	Inferential statistics
Cochran et al. (2004)	Applied paper Investigating the source of migration behavior of birds, namely stars, the sun or a magnetic field.	- Raw circular data plot with mean resultant length - Map with directions	- Sample size - Mean direction - Mean resultant length** - Confidence interval 95% and 99%	- Test for uniformity**	- Watson-Williams test
Corcoran et al. (2009)	Applied paper Investigating the commutes of people using circular statistics.	- Map (line and choropleth)	- Mean direction*** - Mean resultant length***	- Rayleigh test***	
Gill, & Hangartner (2010)	Methodological paper Describing regression modeling framework and apply these methods on examples of gun crimes, domestic terrorism and Bundestag elections.	- Linear histogram - Angular histogram - Rose diagram - Density estimate	- Sample size - Mean direction - Variance - Circular standard deviation		- Circular maximum likelihood estimation (MLE) regression - Circular Bayesian regression (developed and used)
Hanbury (2003)	Methodological paper Demonstrating methods to summarize images by placing the color spectrum on the circle and developing a correction for the saturation of the colors.	- Linear histogram	- Unweighted hue mean**** - Saturation-weighted hue mean**** - Range		
Herrera et al. (2010)	Applied paper Investigating if listeners have a preference to listen to specific artists or genres at certain moments of the day or week and using this to predict new listening behavior.	- Raw circular data plot with mean resultant length - Rose diagram with mean resultant length	- Sample size	- Test for uniformity (only percentages of non-uniformity with corresponding standard deviation reported)	- Investigated the predictability of listening behavior with pre-determined accuracy measures, but not with a formal test

Article	Purpose of the article	Visualizations of observed data	Descriptive statistics	Distribution assumptions and checks	Inferential statistics
Kirschner, & Tomasello (2009)	Applied paper Comparing joint drumming behavior of children in different conditions.	- Raw circular data plot with mean resultant length - Boxplot (of median, median deviation and mean resultant length) - Raw circular data plot (of mean vectors R with grand mean vector)	- Sample size	- Rayleigh test	- Linear repeated measures ANOVA, 3 x 3 x 2 design (on mean resultant length)
König et al. (2013)	Applied paper Using a interpersonal circumplex model to see which methods work better for patients with post-traumatic stress disorder per quadrant on the circumplex.		- Sample size - Linear variance - Confidence interval 95% - Structural summary (amplitude, elevation, displacement, R^2)		- Linear one-factor ANOVA - Post hoc t-test - Linear ANOVA, 2 x 2 x 4 design - Correlation
Kubiak, & Jonas (2007)	Methodological paper Developing a method to detect patterns over time using the maximization of mean vector length (MMVL)	- Raw circular data plot with mean resultant length - Linear histogram - Linear plot (of MMVL results)	- Mean direction - Mean resultant length	- Rayleigh test - Rao test (to test for uniformity)	- Iterative procedure: MMVL
Mainhard et al. (2011)	Applied paper Studying the development of classroom social climate using a circumplex model with the axes control and affiliation.	- Linear graph (of two axes of circumplex) - Boxplot	- Sample size - Linear mean of both axes - Linear deviation of both axes		

Article	Purpose of the article	Visualizations of observed data	Descriptive statistics	Distribution assumptions and checks	Inferential statistics
Maldonado et al. (1997)	Applied paper Comparing neuronal orientations in pinwheel centers and iso-orientation domains.	- Angular map with colors - Polar map with colors - Linear histogram - Linear graph (of Gaussian functions)	- Sample size - Mean orientation scatter with corresponding deviations - Mean orientation range with corresponding deviations	- Compared Gaussians fit's χ^2 with fit of straight line (to test uniformity) - Fitted a Gaussian function to the distribution using MLE - Kolmogorov-Smirnov test (to test if the two samples came from the same distribution)	- T-test - Simple linear regression
Mechsner et al. (2007)	Applied paper Investigating the role of visual feedback using a platform that patients had to circle around bimanually in different conditions.	- Linear histogram (of circular standard deviation (cSD)) - Linear plot	- Mean direction - Circular standard deviation - Mean resultant length	- Hotellings test (to check if individual means can form a grand mean)	- Linear repeated measures ANOVA, 2 x 2 x 2 design (on cSD) - Post hoc t-test
Tolias et al. (2007)	Methodological paper Developing a statistical framework to investigate how the responses of individual and groups of neurons change over time.	- Linear graph (of fitted von Mises distributions) - Linear histogram (of normalized Euclidean distance)	- Sample size	- Rayleigh test (only reported $p < 0.5$ for all data) - Fitting von Mises distributions using least squares	- Normalized Euclidean distance (to compare orientations) - Wilcoxon rank sum test (to compare samples)
Van Doorn et al. (2000)	Methodological paper Developing measurements of changes in wind directions.	- Linear histogram - Linear graph - Linear plot (scatter plot)		- Compared the probability density function of θ with a Gaussian function	

Note. *Bulbert et al. (2015) reported a rho. Concerning the text it is assumed they calculated the mean resultant length. **Cochran et al. (2004) reported an R and a p-value. Concerning the text, it is assumed they calculated the mean resultant length and a test for uniformity, respectively. ***Corcoran et al. (2009) reported the mean direction, mean resultant length and the results of the Rayleigh test in a visualization, where these measures were categorized. Besides, they reported that the application of a von Mises model and a Watson test can permit establishing a comparison. ****A color can be specified with the luminance, hue and the saturation, where the hue "is an angular measure around the achromatic axis with respect to an origin at pure red" (Hanbury, 2003, p. 53).

Appendix B

Table B1. Explanation and origin of the datasets

Dataset	Explanation	Origin
Cake	Breaking angles of cakes from three different recipes and six different baking temperature. The range of this data is restricted, the possible degrees are between the zero and 90 degrees. In this study only the data of the three recipes were used.	Retrieved from package lme4 in R (Bates, Sarkar, Bates, & Matrix, 2007)
Fisher 5	Measurements of orientations of fractures in basalt (Fisher, 1995). This dataset contains axial data.	Randomly selected from twelve datasets in the package circular in R (Agostinelli, & Lund, 2013)
Fisher 7	Directions of ants that were put on a black target with the same illumination (Fisher, 1995).	Randomly selected from twelve datasets in the package circular in R (Agostinelli, & Lund, 2013)
Fisher 12	Movement directions of pigeons, released Northwest of their loft (Fisher, 1995)	Randomly selected from twelve datasets in the package circular in R (Agostinelli, & Lund, 2013)
Music	Listening behavior of users of the music discovery service Last.fm (2015). It contains the variables user, timestamp, artist and song of almost 1000 users and was collected using the user.getRecentTracks(). From these respondents only 49 were completely available for the researcher. Three music tracker users, namely user 6, 35 and 44, were randomly selected. In Table B2 information about the three users can be found.	Retrieved online from the webpage of Celma (2010), analyzed in Herrera et al. (2010)
Pittsburgh crimes	Time of reported crimes in Pittsburgh, Pennsylvania, from 1987 to 1998.	Sent by Gill, analyzed in Gill and Hangartner (2010)
Neuron orientation	Neuronal orientations of three neurons measured with tetrodes.	Reported descriptions used of Berens (2009)
Circling	Data of bimanual circling in two patients in different conditions. Two conditions were randomly selected from the dataset.	Reported descriptions used of Mechsner et al. (2007). This dataset was also sent by Mechsner, but only the descriptions from their article were used.

Table B2. Information of the users from the Music dataset

User	Gender	Age	Country	Registered
6		24	Russian Federation	May 18, 2006
35	m	22	Australia	Mar 7, 2005
44	f	21	Italy	Feb 27, 2007