Tutorial 3 - Plots and distributions

Introduction

It is important to investigate the way that data is distributed. Not only do many models assume that data has a certain distribution, even simple descriptives such as the median or mean may paint an incorrect picture of our data if we take them at face value.

When we analyse circular data it is just as important that we visualize the distribution of data points. In the previous tutorial we noted it was possible for mean directions to be undefined and for mean resultant lengths to indicate a very low concentration even when in our data this may not be the case. For example, take the dataset specified below.

```
example <- as.circular(c(350, 355, 0, 5, 10,
                         190, 185, 180, 175, 170), units = "degrees")
mean(example)
```

```
## Circular Data:
## Type = angles
## Units = degrees
## Template = none
## Modulo = asis
## Zero = 0
## Rotation = counter
## [1] NA
```

```
rho.circular(example)
```

```
## [1] 3.861386e-17
```

The mean direction returns `NA` indicating it is not defined. Additionally, the mean resultant length is practically equal to zero. This value would lead us to believe that the data is spread entirely across the circle. However, when inspecting the actual distribution we can see that there are two modal groups on opposite ends of the circle.
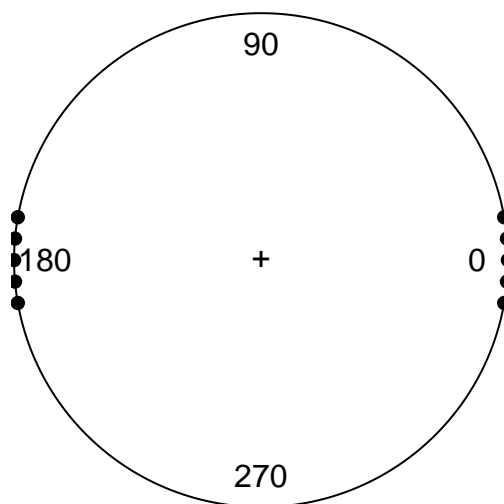

Dataset with two exact opposite modal groups.

In this example, exact opposites were used. Similar problems also arise when data is distributed in two largely opposite groups. We will find a mean resultant length close to zero even though the data is not evenly spread across the circle. The mean direction is thus not indicative for the actual location of the data and may also drastically shift due to small changes in observed values.

Distributions on the circle

Generally speaking there are three types of circular distributions: unimodal, multimodal and uniformly distributed data.

Unimodally distributed data is data that contains a single area on the circle in which most of the observations are found. Many statistical tests assume unimodally distributed data. It is however also possible to find multimodal data. This is data in which two or more modal groups are present. This may pose a problem if these modal groups all belong to the same set of observations. Uniformly distributed data is also problematic. In a uniform distribution, observations are spread out entirely over the entire circular range.
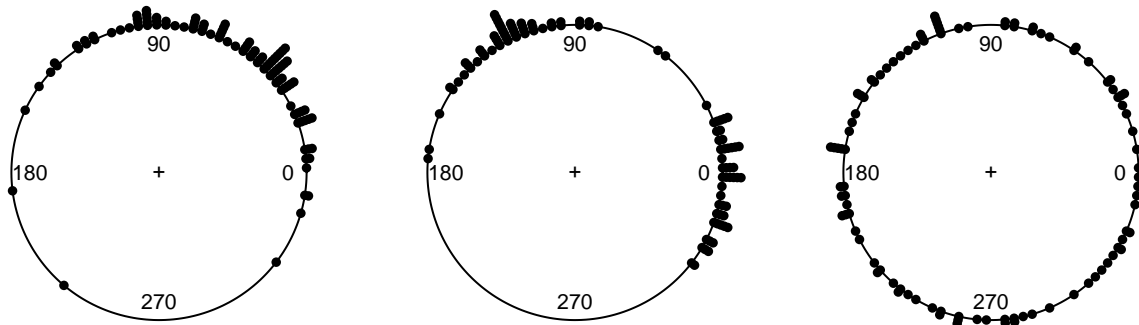
From left to right: A unimodal, bimodal and uniform circular distribution.

Models for circular data as well as simple descriptives make assumptions about the distribution of data points on the circle. It is thus always a good idea to plot the dataset before any analysis takes place. Additionally, uniformity of circular data can sometimes be hard to judge just by eye. Data points can be scattered across the circle but still have one or more modal groups. For this reason, it is common practice to test uniformity with a statistical test. A future tutorial will cover some of the possible ways to test uniformity.

Plotting circular data in R

The `circular` package contains the function `plot.circular` for this purpose. An example of this is given below, using a dataset on flight direction of pigeons. Note how the argument `stack = T` is passed onto the function. The default value for this is `FALSE`, which would result in similar values overlapping. A second useful argument is `sep`, which can be used to specify the amount of seperation between stacked data points. The command `cex` can be used to specify the desired size of the datapoints themselves. Increasing `tol` will prevent the figure from cutting off stacked values. The zero point and rotation can also be changed using the arguments `zero` and `rotation`. `zero` requires a value in radians and `rotation` requires the string "*clock*" or "*counter*".
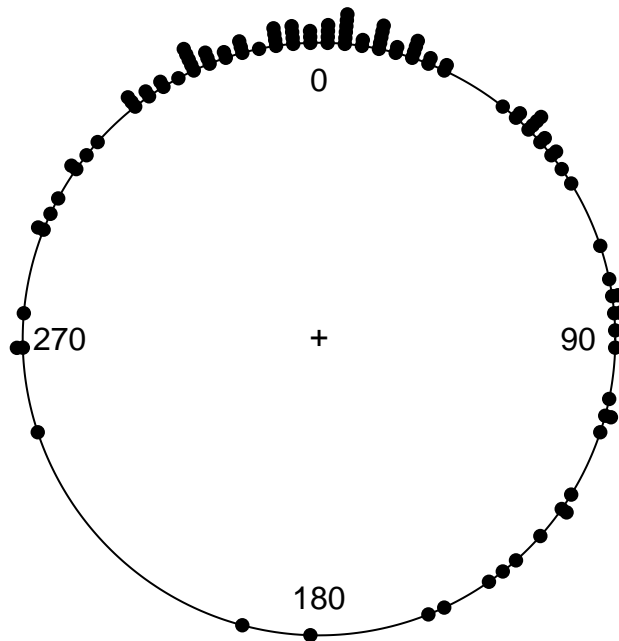
The example below uses a dataset on pigeon flight directions to demonstrate these arguments.

```r
data <- as.circular(pigeons$bearing, units = "degrees")

par(mar = c(1,1,1,1)) #sets the margins for an image in R.
plot(x = data,
```

```
        stack = T,
        sep = 0.02,
        tol = 0.2,
        cex = 1,
        zero = pi/2,
        rotation = "clock",
        main = "", shrink = .9)
```
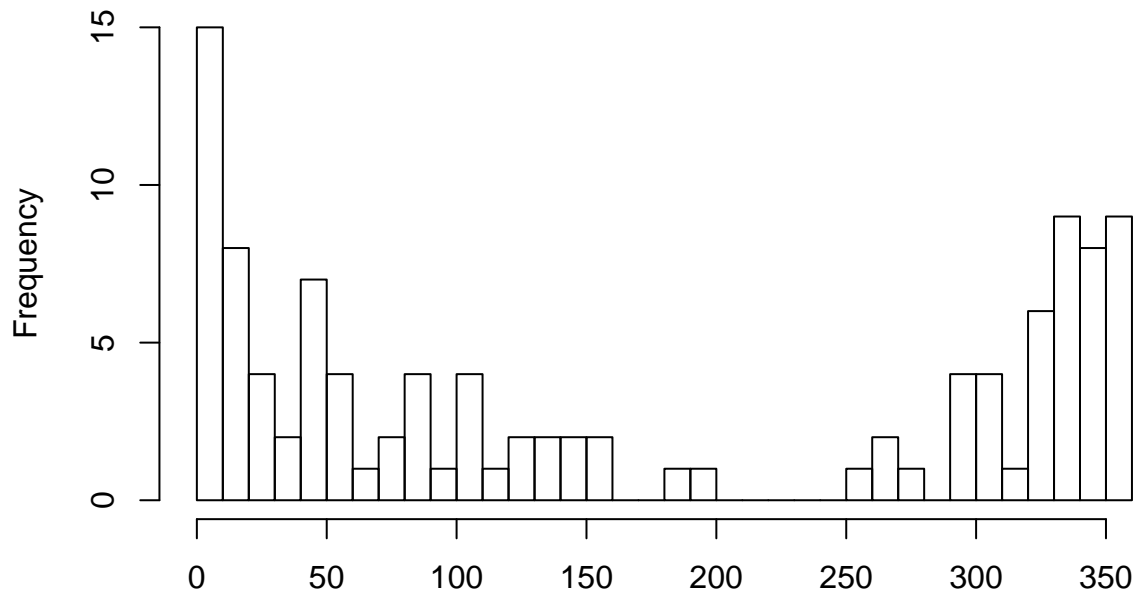

Pigeon flight directions plotted on the circle.

It can sometimes also be useful to inspect circular data on a linear scale. This can be done by transforming the data from circular to continuous using `as.numeric`. Keep in mind that values that are outside of the circular range will need to be calculated down to their remainder using `%%`. This will ensure all values are within either 360 degrees or 2π radians. For example a value of -60 degrees can be converted to 300 degrees using `-60 %% 360`.

```
hist(as.numeric(data%%360),
     breaks = 50,
     xlim = c(0,360),
     main = "",
     xlab = "",
     shrink = .9)
```

Pigeon flight direction plotted as linear data.

Be mindful that a unimodal dataset is not mistaken for a bimodal dataset. In the example above, there are two peaks. One around the value of 0 and one around the value of 360. Since we are dealing with circular data these are in fact the same modal group. One can choose to transform the data so that the modal group is at the center of the histogram.

Conclusion

The way that data is distributed on the circle is important to investigate. Under certain distributions descriptives may be inaccurate. Many circular models also make assumptions about the distribution of data that need to be met. In the following tutorials we will cover various types of circular models, such as the circular regression and circular analysis of variance. We will also cover how to deal with data that violates the assumption of unimodally distributed data.