Tutorial 3 - Plots and distributions

Arjan Huizing

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 a 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.

```r
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
```

```r
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.

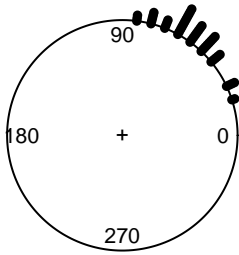# Dataset with two exact opposite modal groups
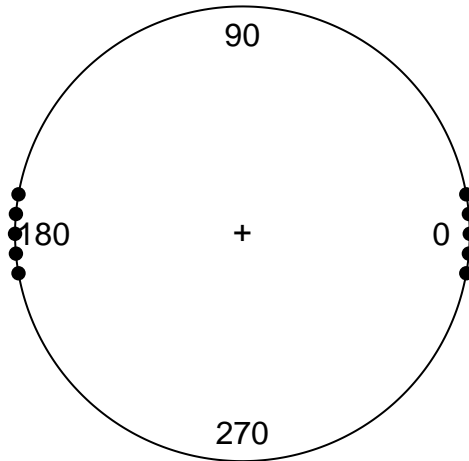
In this example, exact opposites were used. Similar problems also arise when data is distributed in two largely opposite groups. We will still find a mean resultant length close to zero even though the data is not evenly spread across the circle. We will obtain a mean direction, however the mean direction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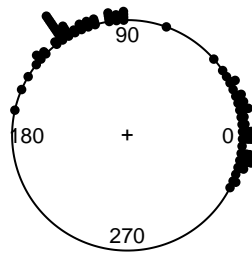
Distributional assumptions

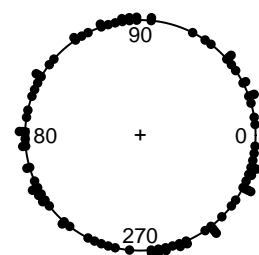Generally speaking there are three types of circular distributions

**Unimodally distributed data**   **Bimodally distributed data**   **Uniformly distributed data**

Plotting circular data in R

It is thus always a good idea to plot the dataset before any analysis takes place. 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.

Pigeon flight directions
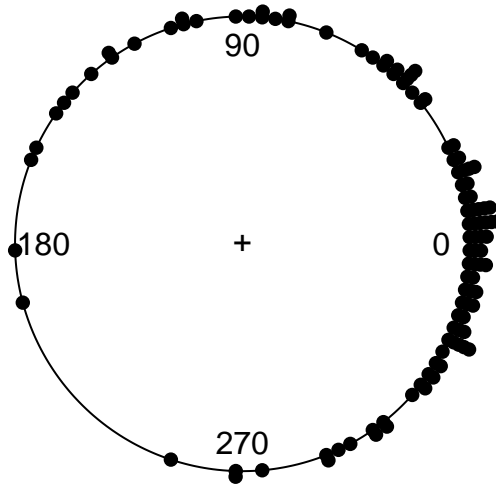
It can sometimes also be useful to inspect circular data on a continues scale. This can be done by transforming the data from circular to continues using `as.numeric`. Keep in mind that values that are outside of the circular range will need to be calculated down to their remainder using `%%`.

```r
hist(as.numeric(data),
     breaks = 360,
     xlim = c(0,360),
     main = "Pigeon flight directions")
```

Pigeon flight directions
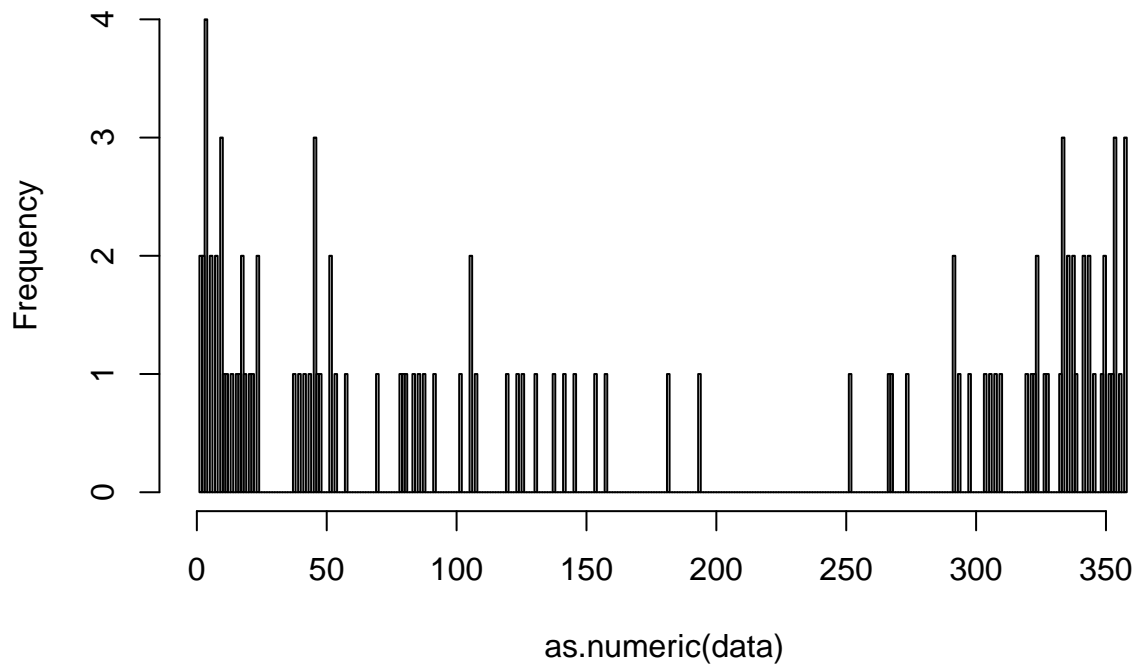

```
#Testing symmetry using a symmetry plot?
```

Uniformity

Uniformally distributed data indicates an even spread of observations across the entire circular range. This can sometimes be difficult to observe visually, which is why it is recommended to also assess uniformity using uniformity plots and statistical tests.

```
#test for uniformity.

#uniformity plot
#kuiper.test(data)
#rao.test(data)
```