Implementation of a radial bar chart geom in ggplot2

-Scientific Team Project KMD-

**Supervisor:** Uli Niemann, M.Sc. uli.niemann@ovgu.de

**Project Outline**

**ggplot2** [1] is the most popular R package for data visualizations. It is an implementation of the “grammar of graphics” [2], a tool to concisely describe the components of a graphic by decomposing it into multiple layers. For example, a scatterplot can be thought of as a combination of four layers: points, axes, a coordinate system, text annotations, e.g. the plot title. As most of these layers are plot type agnostic, **ggplot2** allows to easily combine multiple generic layers into a custom data visualization. As a result, **ggplot2** provides more flexibility compared to R’s inbuilt plot functions like `hist()` or `boxplot()`.

Various extensions\(^1\) have been implemented, including “geoms” (geometric objects) for Sankey diagrams, tree maps, mosaic charts, and radar charts. The goal of this project is to design and implement\(^2\) a geom layer for the radial barchart visualization proposed in [3], cf. Figure 1.

---

Figure 1: Radial barcharts visualizing two clusters of tinnitus patients. (From [3]).

The barchart provides a graphical representation of a subpopulation. In particular, the height of a bar depicts the subpopulation average for a feature. Each feature is z-score normalized. The radial spatial layout distributes the bars around a circle where each bar starts at the black 0 line which represents the feature average over the whole population. Due to feature scaling, bars inclined to the outside represent

\(^1\)https://exts.ggplot2.tidyverse.org/gallery/

\(^2\)https://ggplot2.tidyverse.org/articles/extending-ggplot2.html
feature averages above the overall population mean and bars inclined to the inside represent feature averages below the population mean. This interpretation can optionally be visually supported by color-coded bars. Feature names can optionally be shown on top of each bar. All values are depicted in terms of standard deviation away from the population mean. For example, a value of -1 indicates that the subpopulation average is 1 standard deviation smaller than the overall population average. The standard deviation within a subpopulation is represented as grey error lines facing the colored inner circle. To facilitate quick feature localization, a custom feature categorization can be provided (see inner circle), alongside the subpopulation title and the number of instances within that subpopulation.

Requirements

Interested students must have practical programming experience with R. Preferably, they have completed “Data Science with R”.

Team size: 2 students.

References

1 Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer 2016. doi:10.1007/978-3-319-24277-4