

# Learning to Create Proportional Venn Diagrams in R for Data Visualization

Authored by  
**Mohammed looti**

November 11, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Learning to Create Proportional Venn Diagrams in R for Data Visualization*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=17054>

The [Venn diagram](#) remains a cornerstone of **set theory** and descriptive statistics, using overlapping circles to graphically illustrate the logical relationships and shared elements between distinct groups. While standard Venn diagrams are highly effective for conceptual representation--showing *which* sets overlap--they inherently lack the capacity to convey the actual magnitude or frequency of the data involved. This limitation often renders them insufficient for rigorous quantitative analysis where scale is paramount.

To address this crucial gap, the **proportional Venn diagram** (often technically categorized as an [Euler diagram](#) when areas are precisely scaled) becomes an indispensable tool in the analyst's arsenal. A proportional diagram ensures that the area occupied by each circle, as well as the resultant overlap, is directly scaled according to the frequency or sample size of the respective groups. This critical geometric fidelity guarantees an accurate visual representation, significantly enhancing the interpretability of complex data relationships, especially when dealing with sets of vastly different sizes.

For users working within the statistical computing environment of [R](#), creating these geometrically accurate diagrams efficiently is best accomplished using the highly specialized [eulerr package](#). This robust package provides sophisticated functions, including `plot()` and `euler()`, which manage the complex numerical optimizations necessary to accurately map numerical counts to corresponding geometric areas. The subsequent sections will provide a step-by-step guide on harnessing this powerful computational tool for professional [data visualization](#).

## The Necessity of Area-Proportional Visualization

Effective [data visualization](#) demands techniques that communicate both the structure of the relationships and the true scale of the underlying data. Consider a simple scenario: a visualization indicates that 50 individuals belong to both Group A and Group B. Without proportional scaling, this diagram fails to show whether those 50 shared members constitute a negligible 1% or a overwhelming 90% of the total population. Proportional diagrams elegantly resolve this ambiguity by tying the visual area directly to the numerical frequency, thereby preventing misleading interpretations and fostering accurate conclusions.

The primary technical challenge in generating these proportional representations lies in **geometric optimization**. When representing interactions among three or more sets, it is often mathematically challenging, or even impossible, to draw perfect circles where the area of every segment precisely matches the required count. The [eulerr package](#) overcomes this geometric hurdle by implementing advanced numerical routines. These routines calculate the optimal arrangement and size of the circles, minimizing the error between the desired area (based on input counts) and the actual area rendered in the visualization. This rigorous approach ensures that the resulting visual output maintains the highest possible fidelity to the source data, establishing `eulerr` as an

essential resource for academic research, statistical modeling, and advanced business analytics.

It is important to clarify the often blurred distinction between **proportional Venn diagrams** and [Euler diagrams](#) in this context. While a traditional Venn diagram must display all possible intersections (even those with a zero count), an Euler diagram only requires the visual inclusion/exclusion logic to be met (e.g., disjoint sets must not overlap). When area scaling is the priority, as it is here, the resulting visualization is fundamentally an area-proportional Euler diagram. The [eulerr package](#) is expertly designed to create these visualizations, prioritizing accurate area mapping regardless of whether the final output strictly adheres to the Venn or Euler definition.

## Preparing Data for the eulerr Package

Before any visualization can be executed, users must ensure the necessary software infrastructure is in place. Specifically, the [eulerr package](#) must be installed and loaded into the active **R** environment. Installation is standard and utilizes R's native package management functions. Once loaded, the central task is structuring the input data correctly to define both the size of the individual sets and the counts of their various intersections.

The core function, `euler()`, requires input in the form of a **named vector**. The names assigned to this vector are crucial, as they define the specific set combinations (intersections or unique elements), and the corresponding numerical values represent the counts associated with those combinations. This format provides the function with the precise breakdown of the total population across all sets and subsets.

For instance, if we are analyzing a comparison involving two sets, designated A and B, we must explicitly provide three distinct numerical values: the count of elements unique to A (A only), the count of elements unique to B (B only), and the count of elements shared by both (A&B). Supplying these exact counts allows the `euler()` function to perform the intricate calculations required to determine the necessary radii, positions, and overall geometry needed for an accurate proportional representation. Failure to correctly specify all unique and overlapping counts will result in an inaccurate or incomplete diagram fit.

## Example 1: Constructing the Basic Proportional Diagram

Let us apply the principles discussed by examining a simple, two-group scenario. Imagine we are tracking membership in two hypothetical services, Group A and Group B. Our objective is to visually represent the significant difference in the scale of these two groups and the precise magnitude of their overlap.

We rely on the following hypothetical observed counts to demonstrate the effect of proportional

scaling:

**A (Members unique to Group A): 100**

**B (Members unique to Group B): 500**

**A & B Overlap (Shared Membership): 75**

The crucial task is to visually communicate that Group B (500 unique members) is substantially larger than Group A (100 unique members), while also placing the shared membership (75) into the correct context. We achieve this by utilizing the `euler()` function to calculate the fit and the `plot()` function to render the result in the R environment:

```
library(eulerr)
```

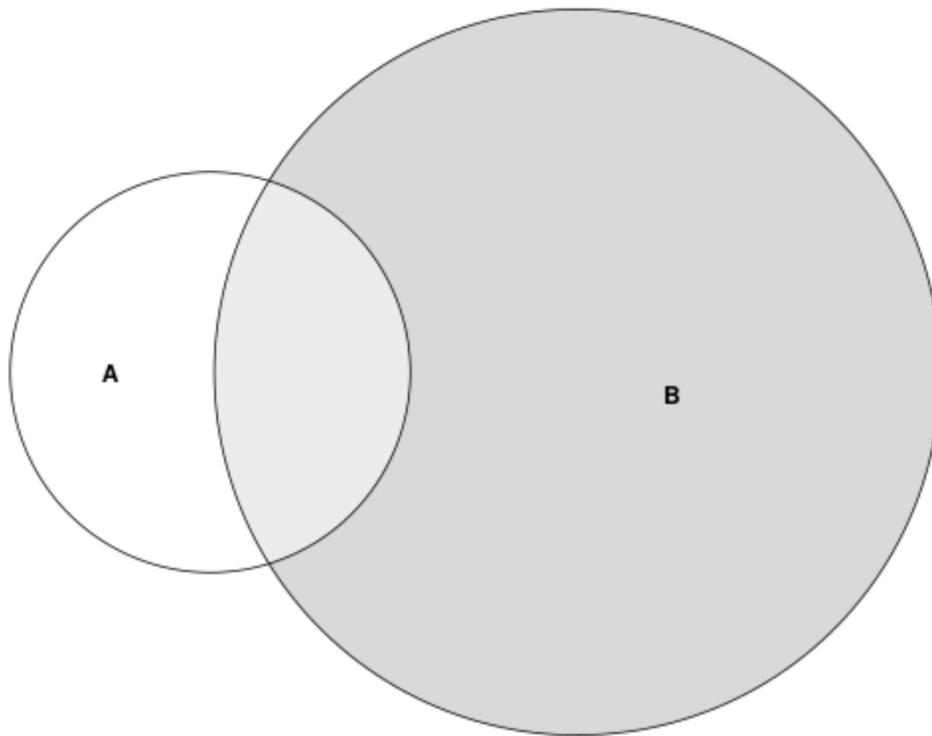
```
#specify values to use in venn diagram
```

```
fit <- euler(c('A' = 100, 'B' = 500, 'A&B' = 75))
```

```
#create venn diagram
```

```
plot(fit)
```

Execution of the preceding code block immediately generates the initial visualization, where the geometric size disparity between the two sets is the primary visual cue, reinforcing the data structure.



Analyzing this initial output confirms that the areas of the circles in the proportional diagram accurately reflect the specified input values. The circle representing Group B is significantly larger in area than the circle for Group A, visually affirming the numerical reality of their unique memberships. Furthermore, the overlap area, while clearly defined, is correctly scaled relative to the size of the larger set (B). This high level of visual fidelity between the input data and the resulting geometric representation underscores the significant analytical benefit provided by the **eulerr** package.

## Example 2: Enhancing Aesthetics Through Customization

While the default plots generated by the `eulerr` package are functionally clean and highly informative, professional reporting often necessitates customization to match specific branding, reporting standards, or to improve visual contrast and clarity. Fortunately, the `plot()` function provides extensive flexibility through various arguments for aesthetic control, allowing analysts to fine-tune the visualization output.

One of the most frequently required aesthetic adjustments involves modifying the color scheme of the sets. This is easily accomplished using the `fill` argument within the `plot()` function. The `fill` argument accepts a vector of color identifiers (such as standard color names or specific hexadecimal codes), which are sequentially applied to the sets defined in the input data (A, B, C, and so on).

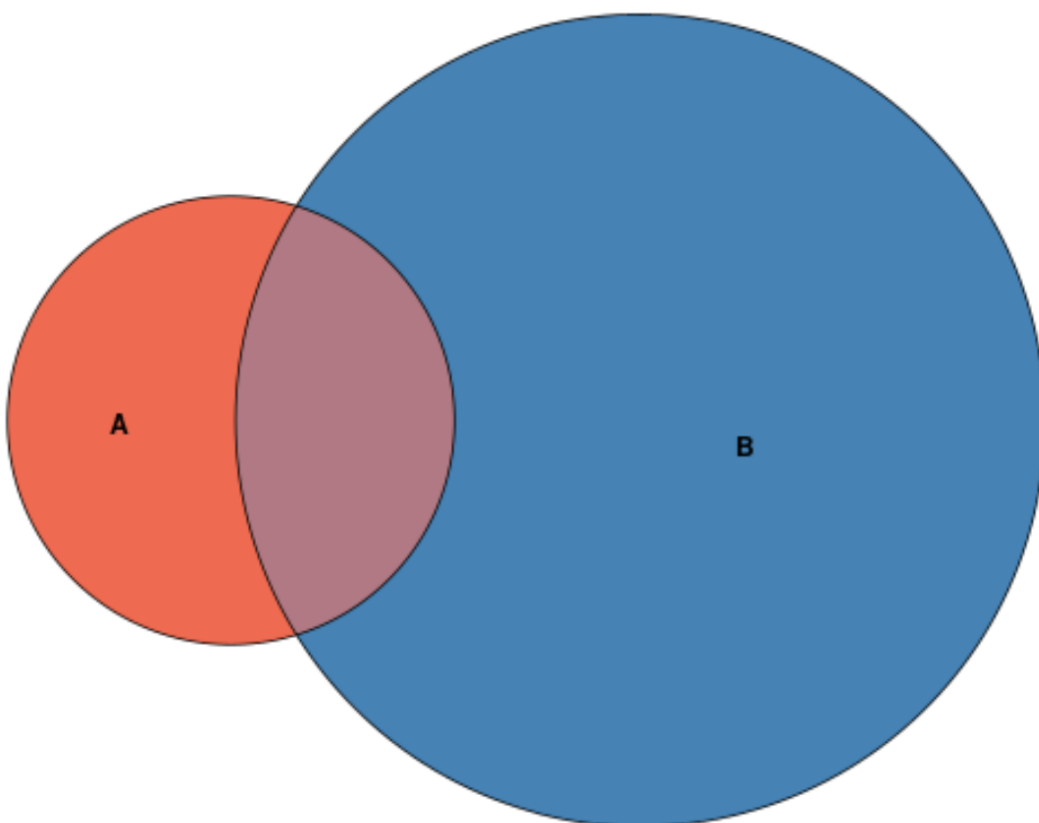
To demonstrate this capability, let's update our previous example to use specific, high-contrast colors--a warm tone for Set A and a cool tone for Set B. We incorporate this modification directly into the plotting command:

```
library(eulerr)
```

```
#specify values to use in venn diagram  
fit <- euler(c('A' = 100, 'B' = 500, 'A&B' = 75))
```

```
#create venn diagram with custom colors  
plot(fit, fill=c('coral2', 'steelblue'))
```

Executing this revised code generates a significantly customized diagram, showcasing the package's versatility in meeting diverse visualization requirements.



As seen in the resulting visualization, the A circle is now filled with the specified **coral2** hue, and the B circle utilizes **steelblue**, exactly matching the color vector provided to the `fill` argument. This powerful customization capability allows analysts to produce visually appealing and highly differentiated visualizations suitable for complex analytical reports. Beyond fill colors, advanced

users can adjust parameters such as border thickness, transparency (alpha levels), and font styling for labels, all contributing to a more refined and professional final graphic.

## Advanced Interpretation and Further Applications

The core value proposition of these proportional visualizations is the instant, intuitive comprehension of scale provided to the viewer. The area of the circle serves as the most immediate and impactful visual cue. In our running example, despite the overlap count of 75 being a substantial proportion of Group A's unique size (100), its visual influence is correctly minimized by the overwhelming scale of Group B (500). This proportional accuracy prevents the viewer from overestimating the importance of the overlap relative to the total size of the system, a common pitfall of non-proportional diagrams.

For users seeking to expand their analytical capabilities, the `eulerr` package is engineered to support diagrams involving up to five sets, although the visual complexity and potential for geometric fitting errors increase dramatically beyond three sets. Furthermore, the package offers crucial flexibility by allowing the plotting of either absolute counts or relative proportions, depending on the analytical requirement. It also provides essential diagnostics related to the **goodness-of-fit**-a vital feature when dealing with highly constrained data sets where achieving mathematically perfect area scaling is geometrically impossible. These diagnostics help users gauge the reliability of the visual representation.

Mastering the `eulerr` package equips **R** users with a robust and superior method for visualizing set relationships compared to rudimentary approaches. It transforms raw frequency data into highly accurate, geometrically faithful, and aesthetically flexible graphical summaries. For a comprehensive exploration of all advanced functionalities, including dealing with complex multi-set intersections and detailed fit diagnostics, users are strongly encouraged to consult the official documentation for the `eulerr` package.

## Additional Resources for R Data Visualization

If you are interested in further developing your [data visualization](#) expertise in **R**, particularly by leveraging the comprehensive capabilities of the [ggplot2](#) framework, the following curated resources offer valuable guidance:

[How to Create a Grouped Boxplot in R Using ggplot2](#)

[How to Create a Heatmap in R Using ggplot2](#)

[How to Create a Gantt Chart in R Using ggplot2](#)