How can I create a Pareto chart in R using step-by-step instructions?

Authored by

stats writer

April 26, 2024

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2024). *How can I create a Pareto chart in R using step-by-step instructions?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=139474

A Pareto chart is a type of data visualization tool that displays the frequency or relative importance of various categories in a dataset. It is a combination of a bar graph and a line graph, with the bars representing the individual categories and the line showing the cumulative percentage of the total. In R, creating a Pareto chart can be easily achieved by following these step-by-step instructions:

1. First, load the necessary packages for creating charts in R, such as "ggplot2" and "dplyr".

2. Next, import your dataset into R or create a new one with the relevant data.

3. Use the "dplyr" package to calculate the cumulative percentage for each category in your dataset.

4. Plot a bar graph of the categories using the "ggplot2" package.

5. Add a line graph to the bar graph using the "geom_line" function, with the cumulative percentage as the y-axis.

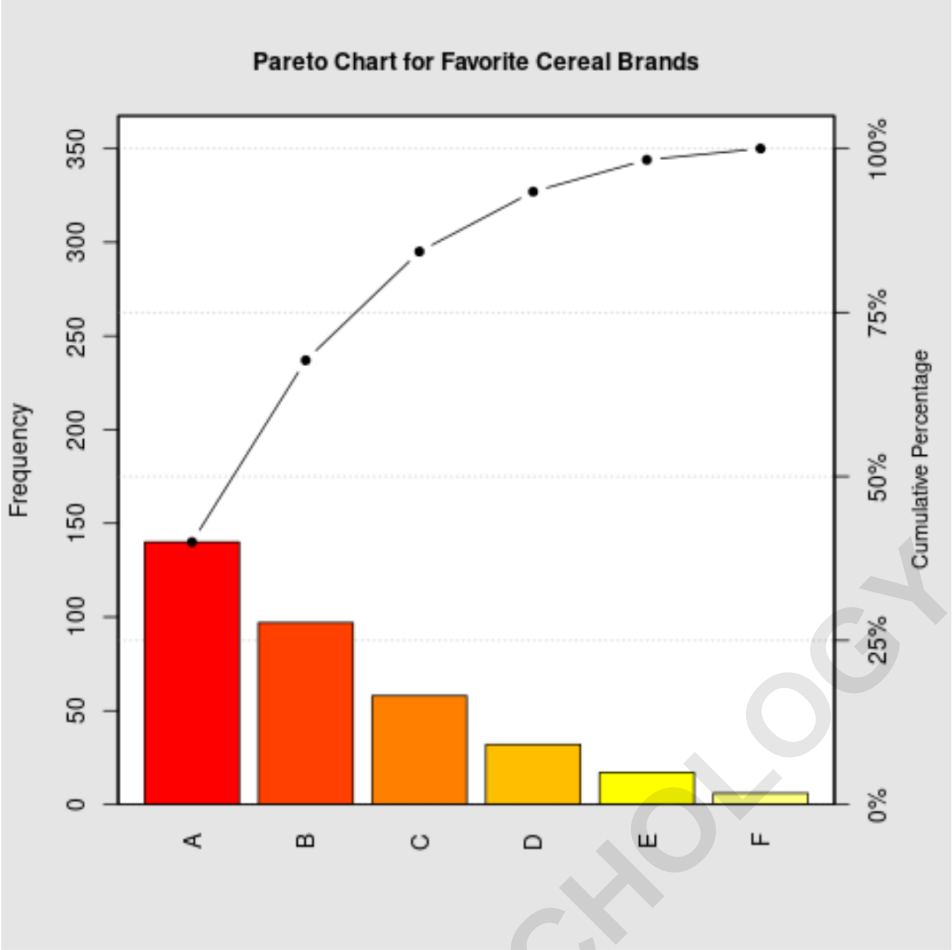
6. Customize the chart by adding labels, titles, and adjusting the colors and styles as desired.

7. Finally, save the chart as an image or display it directly in the R console. By following these steps, you can easily create a visually appealing and informative Pareto chart in R.

Create a Pareto Chart in R (Step-by-Step)

A Pareto chart is a type of chart that displays the frequencies of different categories along with the cumulative frequencies of categories.

ARABPSYCHOLOGY.COM



ARABPSYCHOLOGY.COM

This tutorial provides a step-by-step example of how to create a Pareto chart in R.

Step 1: Create the Data

Suppose we conduct a survey in which we ask 350 different people to identify their favorite cereal brand between brands A, B, C, D, and E.

The following dataset shows the total votes for each brand:

```
#create data
df <- data.frame(favorite=c('A', 'B', 'C', 'D', 'E', 'F'),
count=c(140, 97, 58, 32, 17, 6))

#view data
df

favorite count
1 A 140
2 B 97
3 C 58
4 D 32
5 E 17
6 F 6
```

Step 2: Create the Pareto Chart

To create a Pareto chart to visualize the results of this survey, we can use the pareto.chart() function from the qcc package:

```
library(qcc)

#create Pareto chart
pareto.chart(df$count)
```

ARABPSYCHOLOGY.COM

Pareto chart analysis for df$count

Frequency Cum.Freq. Percentage Cum.Percent.

A 140.000000 140.000000 40.000000 40.000000

B 97.000000 237.000000 27.714286 67.714286

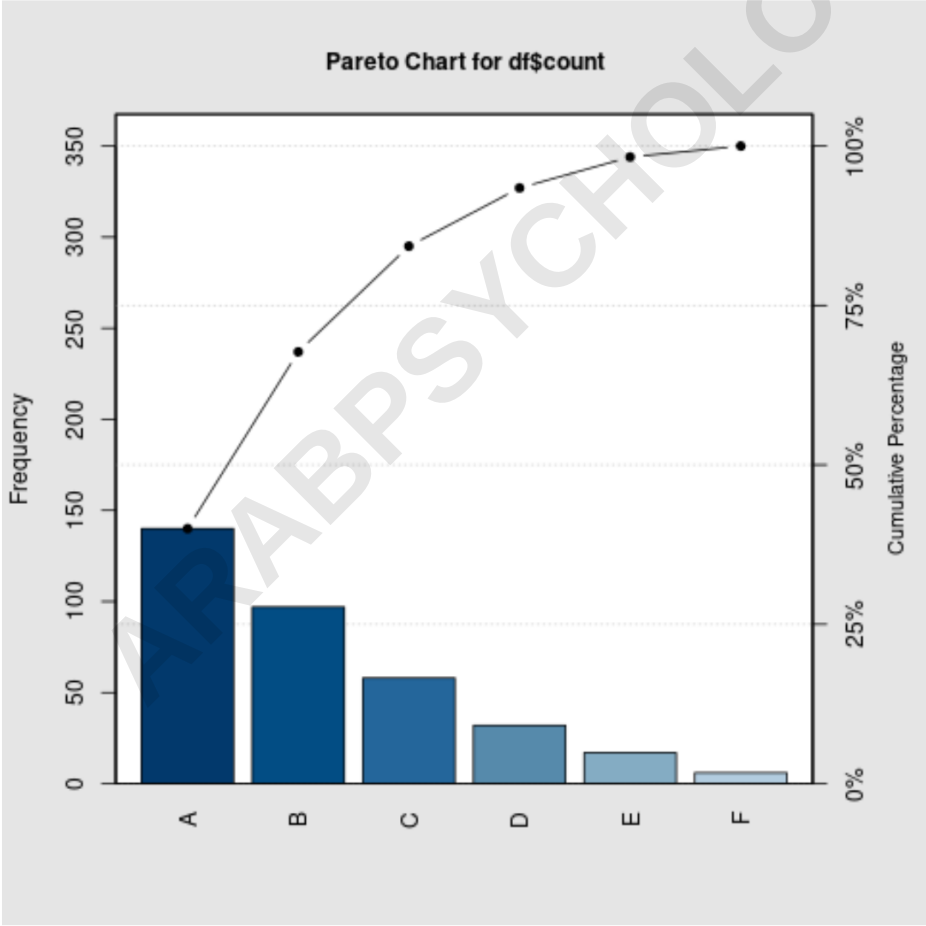
C 58.000000 295.000000 16.571429 84.285714

D 32.000000 327.000000 9.142857 93.428571

E 17.000000 344.000000 4.857143 98.285714

F 6.000000 350.000000 1.714286 100.000000

ARABPSYCHOLOGY.COM



The table in the output shows us the frequency and

cumulative frequency of each brand. For example:

Frequency of brand A: 140 | Cumulative frequency: 140Frequency of brand B: 97 | Cumulative frequency of A, B: 237Frequency of brand C: 58 | Cumulative frequency of A, B, C: 295

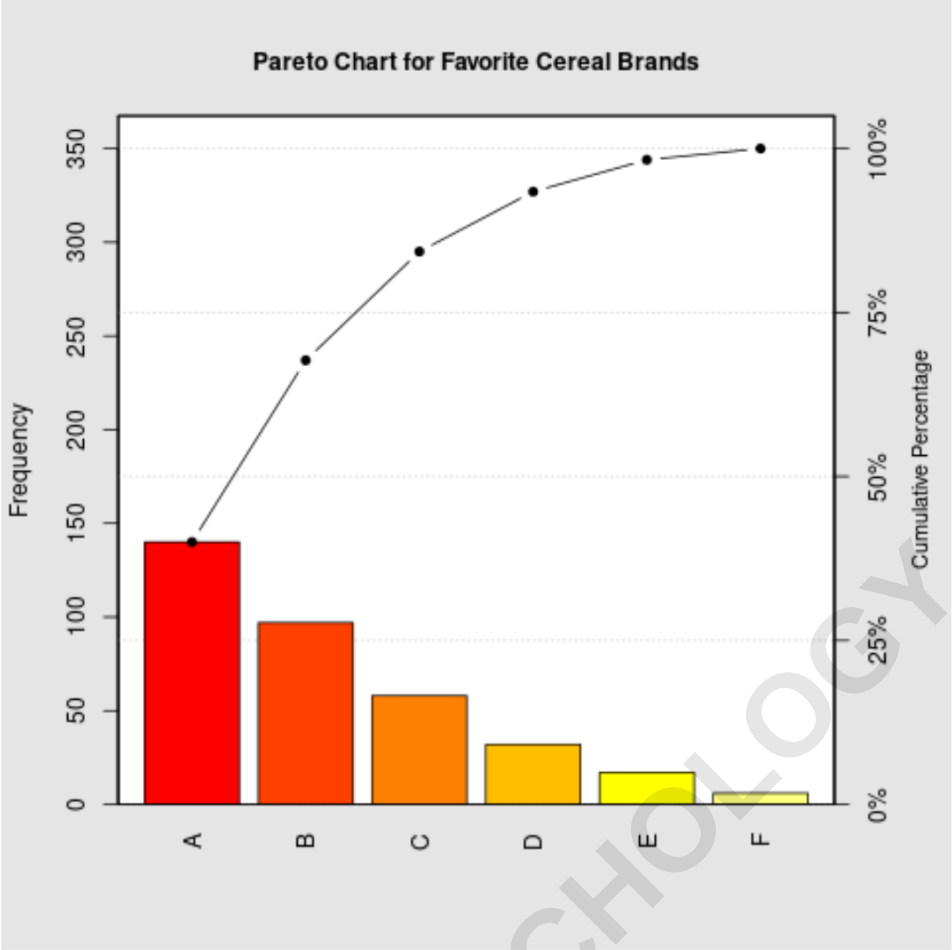
And so on.

Step 3: Modify the Pareto Chart (Optional)

The following code shows how to modify the title of the chart along with the color palette used:

```
pareto.chart(df$count,
main='Pareto Chart for Favorite Cereal Brands',
col=heat.colors(length(df$count)))
```

ARABPSYCHOLOGY.COM



You can find a complete list of color palettes available in .

ARABPSYCHOLOGY.COM