

# R Toolkit Digest



*...dose of hidden gems, packages,  
and practical R tricks...*

## Factors and Ordering Categories in ggplot2



# Why Factors Matter in R

Have you ever created a chart and noticed that the categories appear in the wrong order?

This is especially common with:

- age groups
- survey responses
- education levels
- Likert scales (Strongly Disagree → Strongly Agree)

By default, R often sorts categories alphabetically. While this may be technically correct, it is not always meaningful for communication.

In this presentation, we'll learn how factors help us control the order of categories and create more informative visualizations.

# Understanding Factors

## What is a Factor?

A factor is a categorical variable whose values can be arranged in a specific order.

### Examples:

- Low, Medium, High
- Male, Female
- Strongly Disagree → Strongly Agree

Unlike regular text values, factors allow us to tell R how categories should appear in tables, summaries, and visualizations.

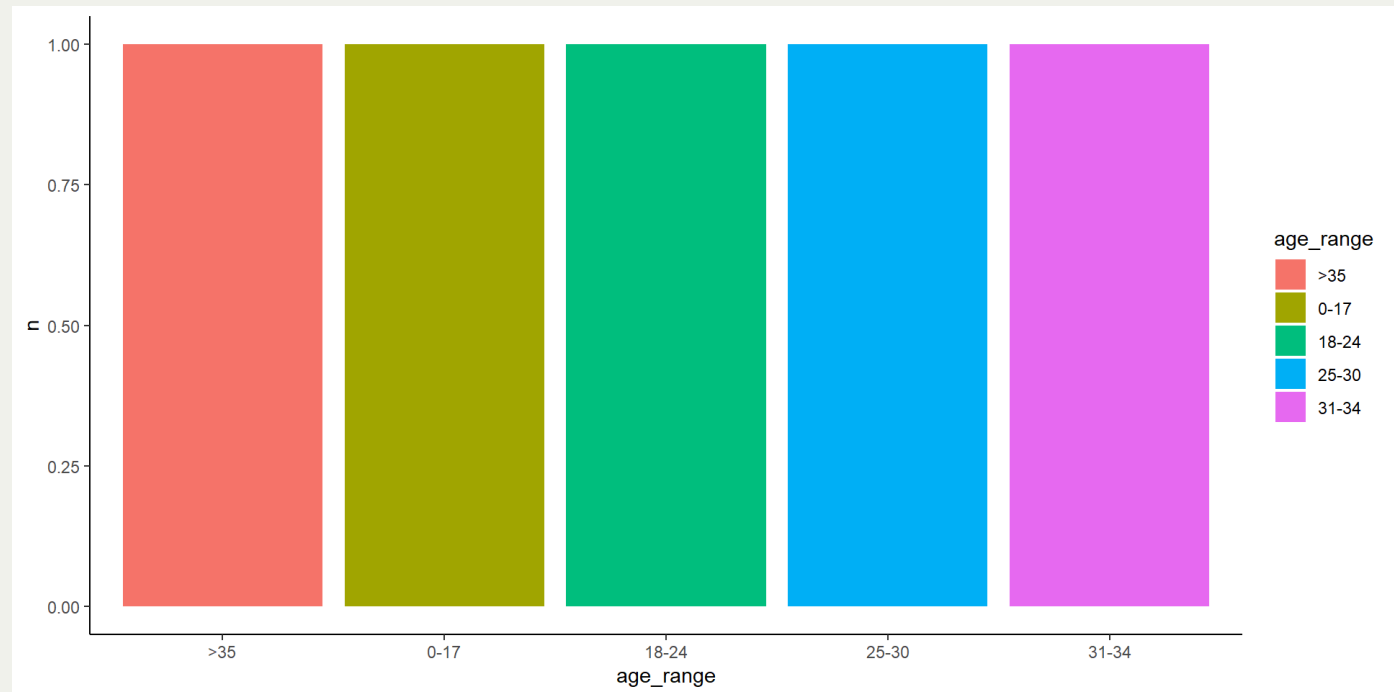
# Sample Data

Suppose a survey collected respondents' age groups.

These categories have a natural progression from younger to older respondents.

```
age_range
1      0-17
2     18-24
3     25-30
4     31-34
5     >35
```

Notice when visualized the categories appear in an ascending order.



Sometimes that is not the order we want because when categorical variables are stored as text, R may not display them in the order we expect.

# Using Factors to Control Order

For survey data, age groups, and reporting categories, we often need complete control over how categories appear.

# What Does `factor()` Do?

The `factor()` function allows us to define the exact order in which categories should appear.

The order is specified using the `levels` argument.

```
1 factor(  
2   variable,  
3   levels = c(...)  
4 )
```

Whatever order is provided in `levels` becomes the order used in tables and visualizations.

# Checking Factor Levels

You can inspect the order of categories using:

```
1 levels(data$response)
```

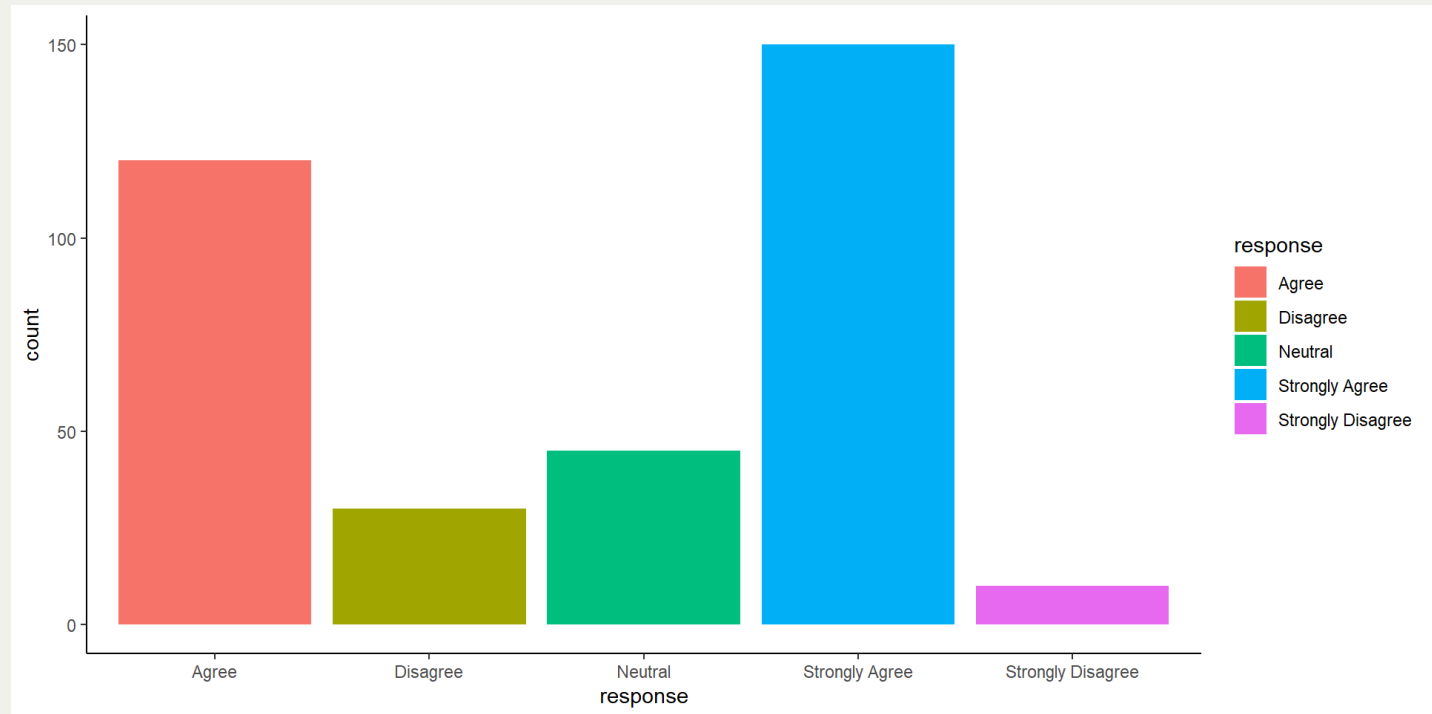
This is useful for confirming that categories will appear in the correct order before plotting.

# A Common Monitoring & Evaluation Example

Many survey tools collect responses such as the one below:

Without factors, charts may appear in alphabetical order, which trumps meaningful communication.

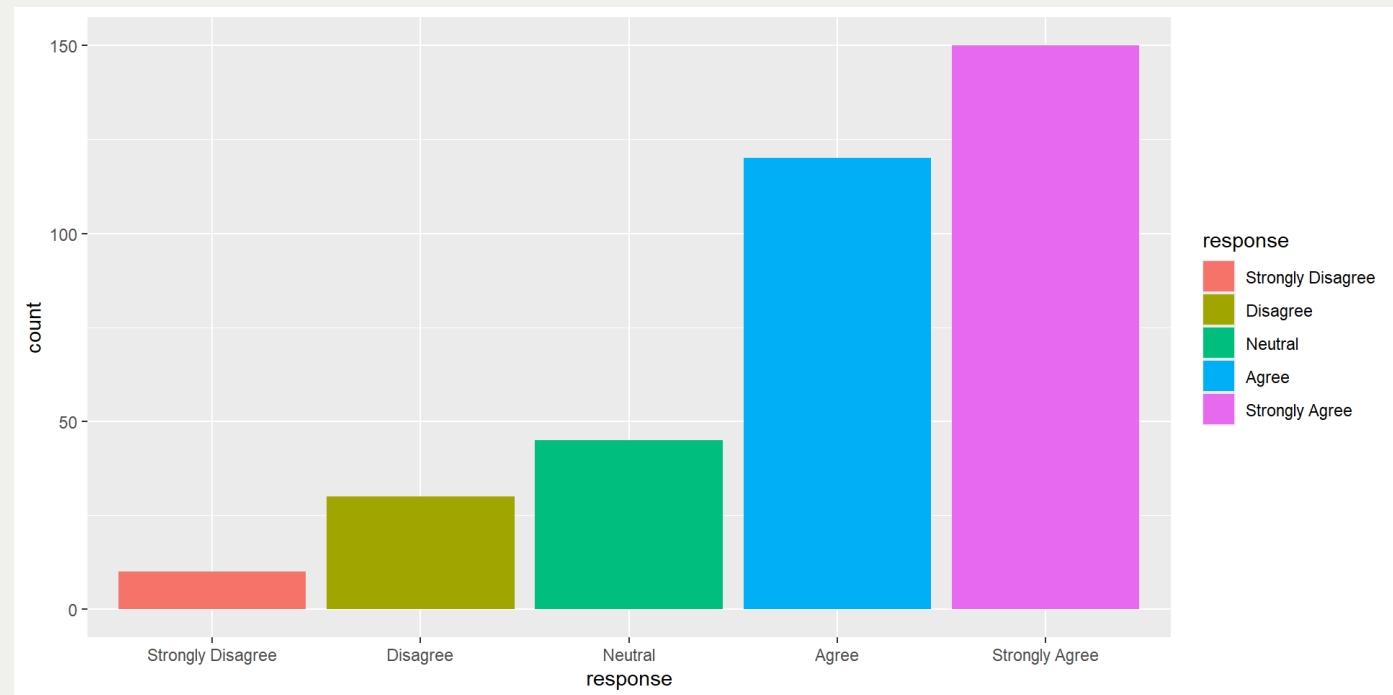
	response	count
1	Strongly Agree	150
2	Agree	120
3	Neutral	45
4	Disagree	30
5	Strongly Disagree	10



# Using Factors to Control Order

With factors, responses can be displayed according to the survey scale, making interpretation much easier.

```
1 data <- data %>%
2   mutate(
3     response = factor(
4       response,
5       levels = c(
6         "Strongly Disagree",
7         "Disagree",
8         "Neutral",
9         "Agree",
10        "Strongly Agree"
11      )
12    )
13  )
14
15 data %>%
16 ggplot(
17   aes(
18     x = response,
19     y = count,
20     fill = response
21   )
22 ) +
23 geom_col()
```



Now the categories follow a meaningful order.

# Key Takeaways

- Factors control how categories are ordered.
- ggplot2 respects factor levels when creating charts.
- Ordered categories improve interpretation.
- Factors are especially useful for survey responses, age groups, education levels, and Likert scales.

A small change in data structure can make a big difference in how results are communicated.



# Further Reading

- [forcats package documentation](#). Click here to read
- [R for Data Science \(Factors Chapter\)](#). Click here to read
- [ggplot2 Documentation](#). Click here to read
- [The Tidyverse Website](#). Click here to read

## Practice Datasets

- [Stack Overflow Developer Survey](#). Click here
- [Kaggle Survey Datasets](#). Click here



# Explore My Work

I share dashboards, R projects, Shiny apps, and data tools on my portfolio.

👉 Portfolio

## Includes:

- Interactive dashboards
- R packages in development
- Data visualization projects
- MEAL and monitoring tools