

R

Programming

tidyverse

a consistent way to organize your data in R, an organization called tidy data.
Getting your data into this format requires some up-front work, but that work pays off in the long term.

Making tidy data

There are three interrelated rules which make a dataset tidy:

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.

dplyr, ggplot2, and all the other packages in the tidyverse are designed to work with tidy data.

Tibble in R

Tibbles are data frames, but they tweak some older behaviors to make life a little easier.

Eg: `Table1`, `Table4a`

Functions to tidy up data

Gathering- When the column names are values of variable. Use `gather()`

Ex: table4a

Determine the name of the variable whose values form the column names. Call it as key value pair.

Spreading is the opposite of gathering.

You use it when an observation is scattered across multiple rows.

Separating and Pull

Separate pulls apart one column into multiple columns by splitting wherever a separator character appears.

By default it will separate according to non-numeric character.

Eg: table3

Unite will combine multiple columns into single column.

Eg: table5

Missing Values

The NA values are known as missing values.

We can fill the values with previous values or omit the rows.

Line Graphs

A line chart is a graph that connects a series of points by drawing line segments between them.

Line charts are usually used in identifying the trends in data.

`plot(v,type,col,xlab,ylab)`

- `v` is a vector containing the numeric values.
- `type` takes the value "p" to draw only the points, "l" to draw only the lines and "o" to draw both points and lines.
- `xlab` is the label for x axis.
- `ylab` is the label for y axis.
- `main` is the Title of the chart.
- `col` is used to give colors to both the points and lines.

Pie Charts

A pie-chart is a representation of values as slices of a circle with different colors. The slices are labeled and the numbers corresponding to each slice is also represented in the chart.

In R the pie chart is created using the **pie()** function

`pie(x, labels, radius, main, col, clockwise)`

- **x** is a vector containing the numeric values used in the pie chart.
- **labels** is used to give description to the slices.
- **radius** indicates the radius of the circle of the pie chart.(value between -1 and +1).
- **main** indicates the title of the chart.
- **col** indicates the color palette.
- **clockwise** is a logical value indicating if the slices are drawn clockwise or anti clockwise.

R - Bar Charts

A bar chart represents data in rectangular bars with length of the bar proportional to the value of the variable.

R uses the function **barplot()** to create bar charts.

```
barplot(H,xlab,ylab,main, names.arg,col)
```

- **H** is a vector or matrix containing numeric values used in bar chart.
- **xlab** is the label for x axis.
- **ylab** is the label for y axis.
- **main** is the title of the bar chart.
- **names.arg** is a vector of names appearing under each bar.
- **col** is used to give colors to the bars in the graph.

BoxPlot

Boxplots are a measure of how well distributed is the data in a data set.

It divides the data set into three quartiles.

This graph represents the minimum, maximum, median, first quartile and third quartile in the data set.

Boxplot

`boxplot(x, data, notch, varwidth, names, main)`

- **x** is a vector or a formula.
- **data** is the data frame.
- **notch** is a logical value. Set as TRUE to draw a notch.
- **varwidth** is a logical value. Set as true to draw width of the box proportionate to the sample size.
- **names** are the group labels which will be printed under each boxplot.
- **main** is used to give a title to the graph.

Scatterplot

Scatterplots show many points plotted in the Cartesian plane.

Each point represents the values of two variables.

One variable is chosen in the horizontal axis and another in the vertical axis.

scatterplots

`plot(x, y, main, xlab, ylab, xlim, ylim, axes)`

- **x** is the data set whose values are the horizontal coordinates.
- **y** is the data set whose values are the vertical coordinates.
- **main** is the title of the graph.
- **xlab** is the label in the horizontal axis.
- **ylab** is the label in the vertical axis.
- **xlim** is the limits of the values of x used for plotting.
- **ylim** is the limits of the values of y used for plotting.
- **axes** indicates whether both axes should be drawn on the plot.

R Markdown- Interactive Reports

For making interactive reports with our program we can use rmarkdown.

Install package-rmarkdown

For making pdf install tinytex- `tinytex::install_tinytex()`