

Additional Material for Plotting Bar Charts and Scatter Plot

Sudhakar Kumar

29 May 2019

Data frame for analysis

We will use a data frame (ÇETINKAYA-RUNDEL 2019) comprising information on **movies**. It contains 651 observations (rows), each representing a movie, and 32 variables (columns). Let us load this data frame by using `load` function.

```
load(url("https://stat.duke.edu/~mc301/data/movies.Rdata"))
paste("Information on", dim(movies)[1], "movies loaded.")
```

```
## [1] "Information on 651 movies loaded."
```

In **movies** data frame, there is one variable named `imdb_rating`. It contains the IMDB rating (on a scale of 0 to 10) for each of the 651 movies. For example, the movies titled **The Godfather, Part II** and **Disaster Movie** have lowest and highest IMDB rating, respectively. This can also be verified by using `range` function.

```
minMax <- range(movies$imdb_rating)
paste(movies$title[movies$imdb_rating==minMax[1]], "has lowest IMDB rating of", minMax[1])
```

```
## [1] "Disaster Movie has lowest IMDB rating of 1.9"
```

```
paste(movies$title[movies$imdb_rating==minMax[2]], "has highest IMDB rating of", minMax[2])
```

```
## [1] "The Godfather, Part II has highest IMDB rating of 9"
```

Plotting a bar chart

A **bar chart** represents data in rectangular bars with length of the bar proportional to the value of the variable (tutorialspoint 2018). **R** uses the function `barplot` to create **bar charts**. We can draw both vertical and horizontal bars in the **bar chart** using **R**. The basic syntax to create a **bar chart** in **R** is:

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

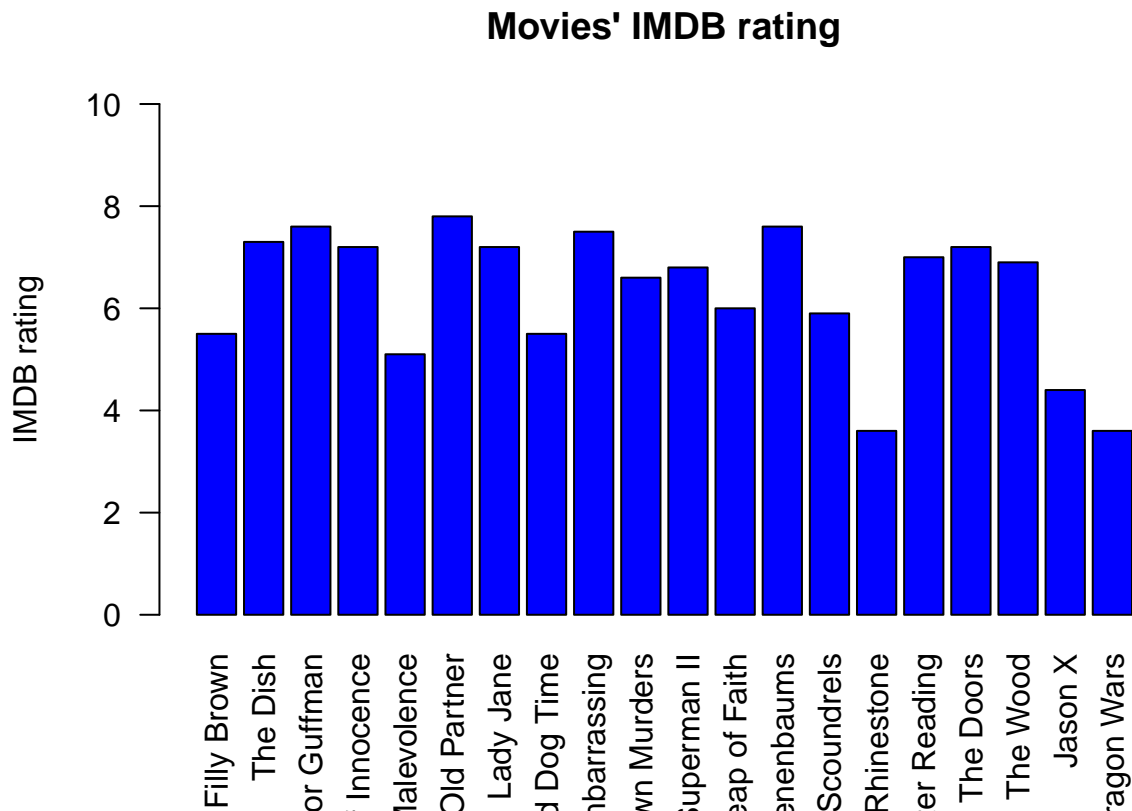
Following is the description of the parameters used:

- `data` is either a vector or matrix of values describing the bars which make up the plot.
- `xlab` and `ylab` are the labels for x and y axes, respectively.
- `main` is the title of the bar chart.
- `names.arg` is a vector of names to be plotted below each bar or group of bars.
- `col` is used to give colors to the bars in the graph.

We will create a **bar chart** from the variable `imdb_rating`. For the sake of simplicity, we consider only the first 20 observations of **movies** data frame.

```
moviesSub <- movies[1:20,]
barplot(moviesSub$imdb_rating,
        ylab = "IMDB rating",
        col = "blue",
        ylim = c(0,10),
```

```
main = "Movies' IMDB rating",
names.arg = moviesSub$title,
las = 2)
```

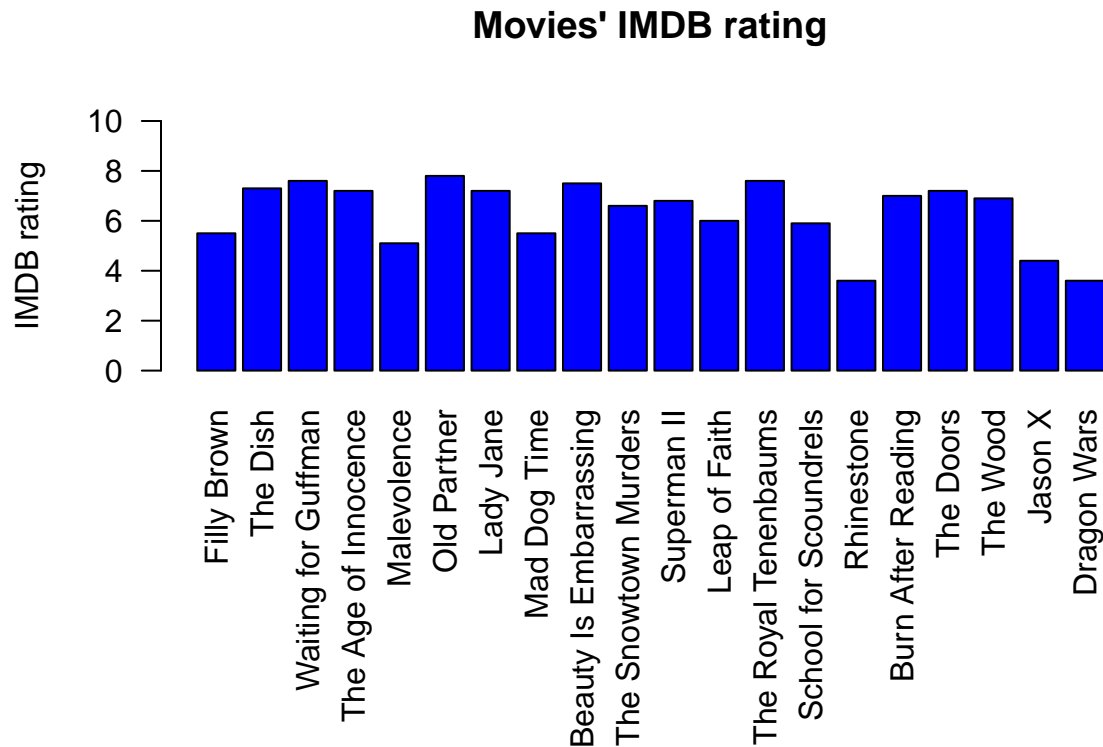


In this **bar chart**, we can see that the titles of movies are present on the X-axis. However, the longer titles are being truncated. This can be fixed either by defining the plotting area or by adding more functions followed by `barplot` function.

Adjusting labels of a bar chart by using `par` function

First, we will apply the former method by using `par` function. According to the documentation of **R** programming language, `par` can be used to set or query graphical parameters. Parameters can be set by specifying them as arguments to `par` in `tag = value` form, or by passing them as a list of tagged values. Here, we pass a `tag` named `mar`, which refers to margin. It is a numerical vector of the form `c(bottom, left, top, right)` which gives the number of lines of margin to be specified on the four sides of the plot.

```
par(mar=c(11,5,5,2)) # c(bottom, left, top, right)
barplot(moviesSub$imdb_rating,
        ylab = "IMDB rating",
        col = "blue",
        ylim = c(0,10),
        main = "Movies' IMDB rating",
        names.arg = moviesSub$title,
        las = 2)
```



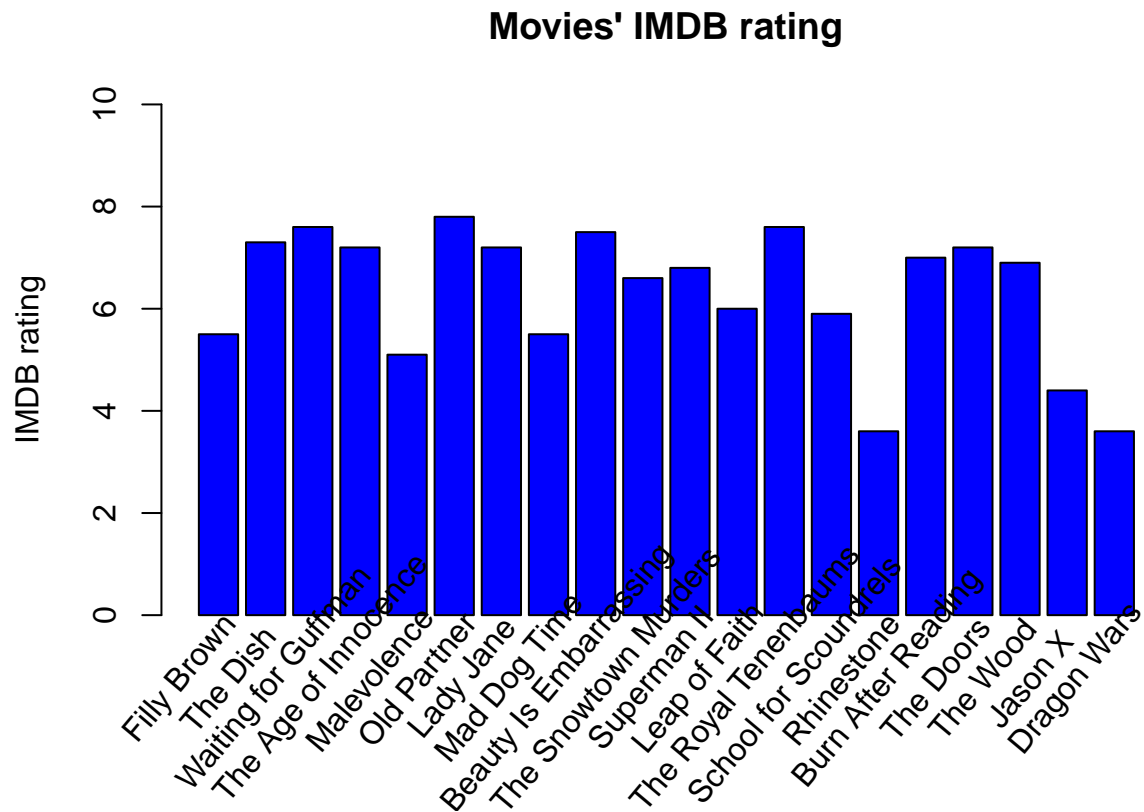
Now, the names are not being truncated, as it is evident from the graph plotted above.

Adjusting labels of a bar chart by using text function

Here, we will add one more function named `text` after the `barplot` function. According to the documentation of **R** programming language, `text` draws the strings given in the vector labels at the coordinates given by `x` and `y`. The required coordinates can be found out by trial and error method. We will pass the following arguments in this function:

- **coordinates** (`x` and `y`) – numeric vectors of coordinates where the text labels should be written.
- **labels** – a character vector or expression specifying the text to be written.
- **xpd** – used to enable things to be drawn outside the plot region.
- **srt** – used to rotate the text by an angle.

```
x <- barplot(moviesSub$imdb_rating,
             ylab = "IMDB rating",
             ylim = c(0,10),
             main = "Movies' IMDB rating",
             col = "blue")
labs <- moviesSub$title
text(x = x - 0.95,
     y = -1.25,
     labs,
     xpd = TRUE,
     srt = 50)
```



There are certainly other ways also to adjust the labels in plots.

References

- ÇETINKAYA-RUNDEL, MINE. 2019. “movies.RData – Mine Çetinkaya-Rundel.” <http://www2.stat.duke.edu/~mc301/data/movies.html>.
- tutorialspoint. 2018. “R - Bar Charts.” https://www.tutorialspoint.com/r/r_bar_charts.htm.