post01-wayne-li

Wayne Li

11/17/2017

Drawing Heatmaps in R

I. Introduction

By now, we have seen many useful graphs and plots. To name a few, line graphs, bar graphs, histogram and so on. However, there are still more powerful graphs not covered in class. This post will introduce heat map, a graphical representation of data where the individual values contained in a matrix are represented as colors. Drawing heatmaps can be accomplished with help from *Plotly*, *ggplot2* and now even the base package. I will focus on explaining how heat map can be constructed in R using *Plotly* in this post. Hopefully, readers will get a sense of how powerful heat map and *Plotly* is and take advantage of them in the future.

II. Example Time

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

For all the examples, I will use the data set that records information on nba players and their performance throughout the year. The data is from www.basketball-reference.com and a sample link for the data source of GS Warriors https://www.basketball-reference.com/teams/GSW/2017.html. Or it can be accessed via this link https://raw.githubusercontent.com/ucb-stat133/stat133-fall-2017/master/data/nba2017-stats.csv.

Data preparation

We will first sort and prepare the data

Sort

I sort the data by the minutes they played in descending order.

```
#sort data
nba_stats <- nba_stats[order(- nba_stats$minutes),]
```

Name Rows

We now name the rows by player names so we can get rid of the first column

```
#name rows
row.names(nba_stats) <- nba_stats$player
#delete first column
nba_stats <- nba_stats[,2:20]
```

We are coming to the final step of data preparation. To make a heatmap, we are required to put our data into a data matrix. However, nba_stats was loaded into data frame. To make that change, with all the work we did above, it will just be one line of code. #####Final step

```
#convert data frame into data matrix
nba_matrix <- data.matrix(nba_stats)
```

Plotting Time

Real Basic Plot

We have our basic heatmap but we can't read anything from it since there is no label on the x-axis nor y-axis.

```
#plot basic heatmap
p <- plot_ly(z = nba_matrix, type = "heatmap")
p
```

Plot With Labels

To make our heatmap more readable and presentable, we will add labels on both x-axis and y-axis

```r
#new plot with labels
p <- plot_ly(z = nba_matrix, type = "heatmap", x = colnames(nba_matrix), y = rownames(nba_matrix))
p
```

Color Selection

By this point, if we just want to see what our data looks like from the heatmap, we have done the job. Whereas, we can still make our gragh pretiier by changing the color scheme

```r
#costomize a color scheme
vals <- unique(scales::rescale(c(volcano)))
o <- order(vals, decreasing = FALSE)
cols <- scales::col_numeric("Blues", domain = NULL)(vals)
colz <- setNames(data.frame(vals[o], cols[o]), NULL)
#apply the colorscale
p <- plot_ly(z = nba_matrix, type = "heatmap", x = colnames(nba_matrix), y = rownames(nba_matrix), colorscale = co
lz)
p
```

As one can see, heatmap can tell us quite a lot about the data. For example, generally, more minutes players playered, they tend to have more field goal attemps.

VII. Summary

The key to this post is to present an elegant and useful graph as an alternate to data visualization. Different graphs have different advantages and disadvantages of presenting data. Depending on the nature of data sets and analysis needs, one should choose the appropriate graphs to visualize their data.

References

1. https://en.wikipedia.org/wiki/Heat_map
2. https://plot.ly/r/heatmaps/
3. http://flowingdata.com/2010/01/21/how-to-make-a-heatmap-a-quick-and-easy-solution/
4. https://www.rdocumentation.org/packages/stats/versions/3.4.1/topics/heatmap
5. https://www.youtube.com/watch?v=7bVCiMrAxe4
6. https://www.r-bloggers.com/heat-maps-using-r/
7. http://rstudio-pubs-static.s3.amazonaws.com/98032_acc2fdbb8d134d438b1beca8a42d4d62.html