R Notebook

Introduction

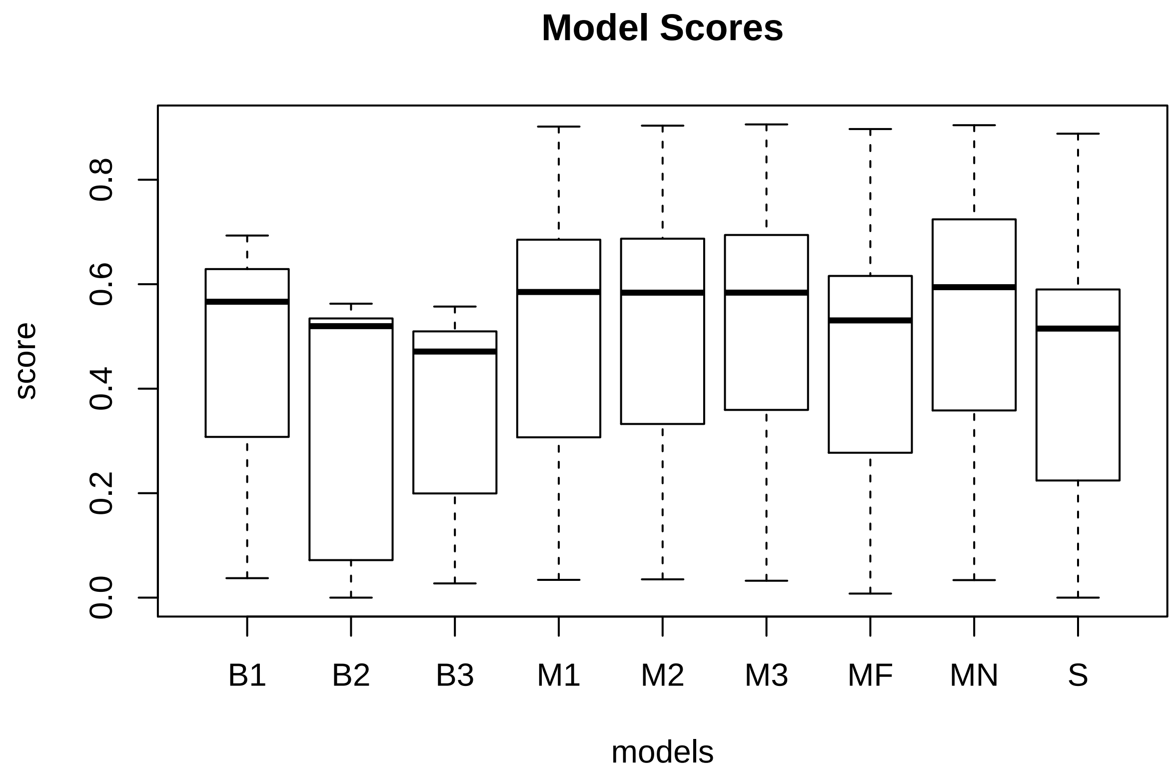
To read the data and understand the distribution of observations we did the following.

```r
library('ggplot2')
library('emmeans')
data = read.csv('../data/data.csv')
table(data$model, data$TeD)
```

```
##
##      TeD1 TeD2 TeD3 TeD4 TeD5 TeD6 TeD7
##   B1    1    1    1    1    1    1    1
##   B2    1    1    1    1    1    1    1
##   B3    1    1    1    1    1    1    1
##   M1   78   78   78   78   78   78   78
##   M2   77   77   77   77   77   77   77
##   M3   72   72   72   72   72   72   72
##   MF   80   80   80   80   80   80   80
##   MN   70   70   70   70   70   70   70
##   S    48   48   48   48   48   48   48
```

We see the Baselines are tested once against the testv sets. The transfer learning methods have been tested multiple times on the test sets to see the efeect of use of different training sets on the score. i.e. does the use of specific training sets help certain tasks (TeD).

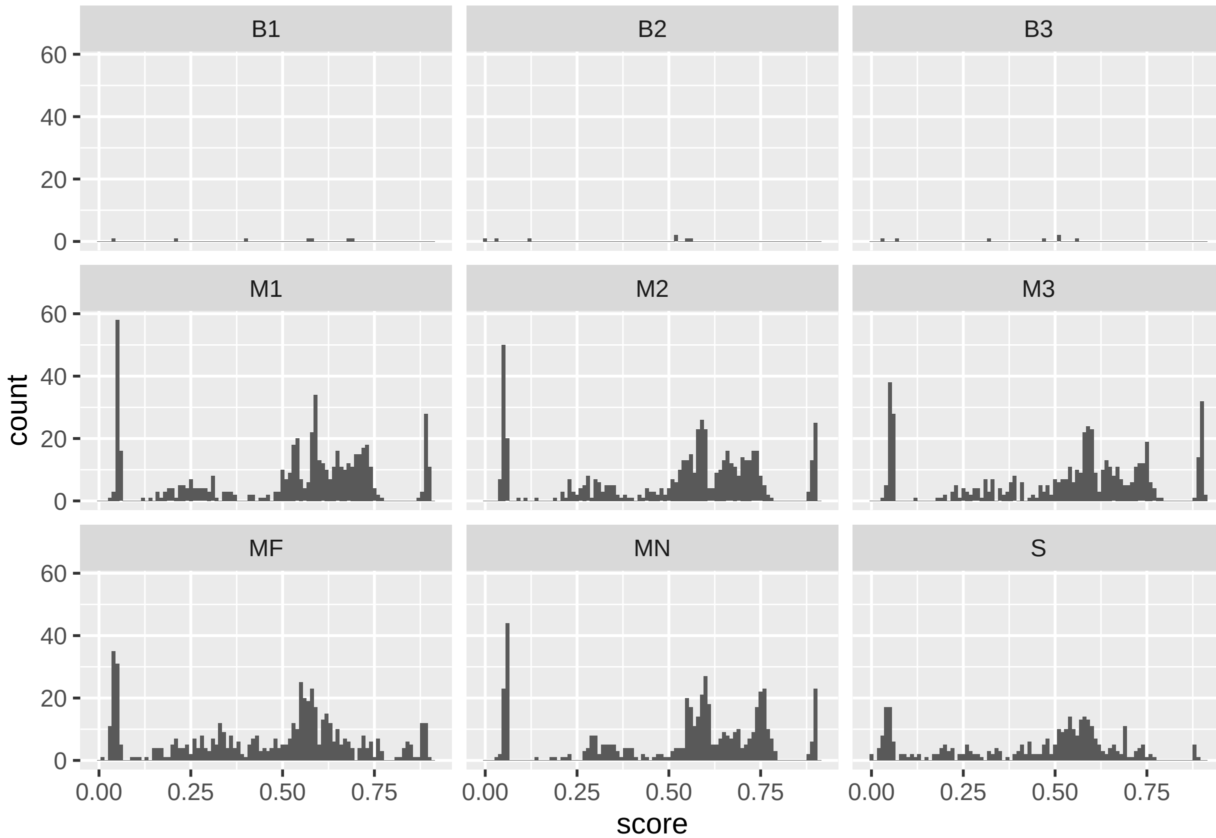
```r
plot(data$model, data$score, main="Model Scores", xlab='models', ylab='score')
```



If we naively observe the data, we see that models that are trained/refined partwise in transferlearning perform better on average than the baselines.

Lets see if this is valid.

```
ggplot(data, aes(score)) + geom_histogram(binwidth = 0.01) + facet_wrap(~model)
```



We see that the data is not normally distributed and we cannot naively derive conclusions from simple statistical analysis (like means) and linear models. We need to consider the following

- Effect of multiple IVs on the score.
- Interaction effect of different IVs.
- Type of task (TeD) we are testing on.
- Significance of the use of particular Method on test set.
- Effect of particular training set on the scoring of a test set by a model.