Analysis of marathon data

## Dmitri Timašjov, Ragnar Vent, Hindrek Teder

We tried to investigate pace groups - people who ride together throughout the race, with the person in front of the group setting the pace. In order to find the pace groups, we assumed that the pace group had to have a minimum of 2 people and the split times were within few seconds of each other. This job seemed like a fit for DBSCAN clustering method.
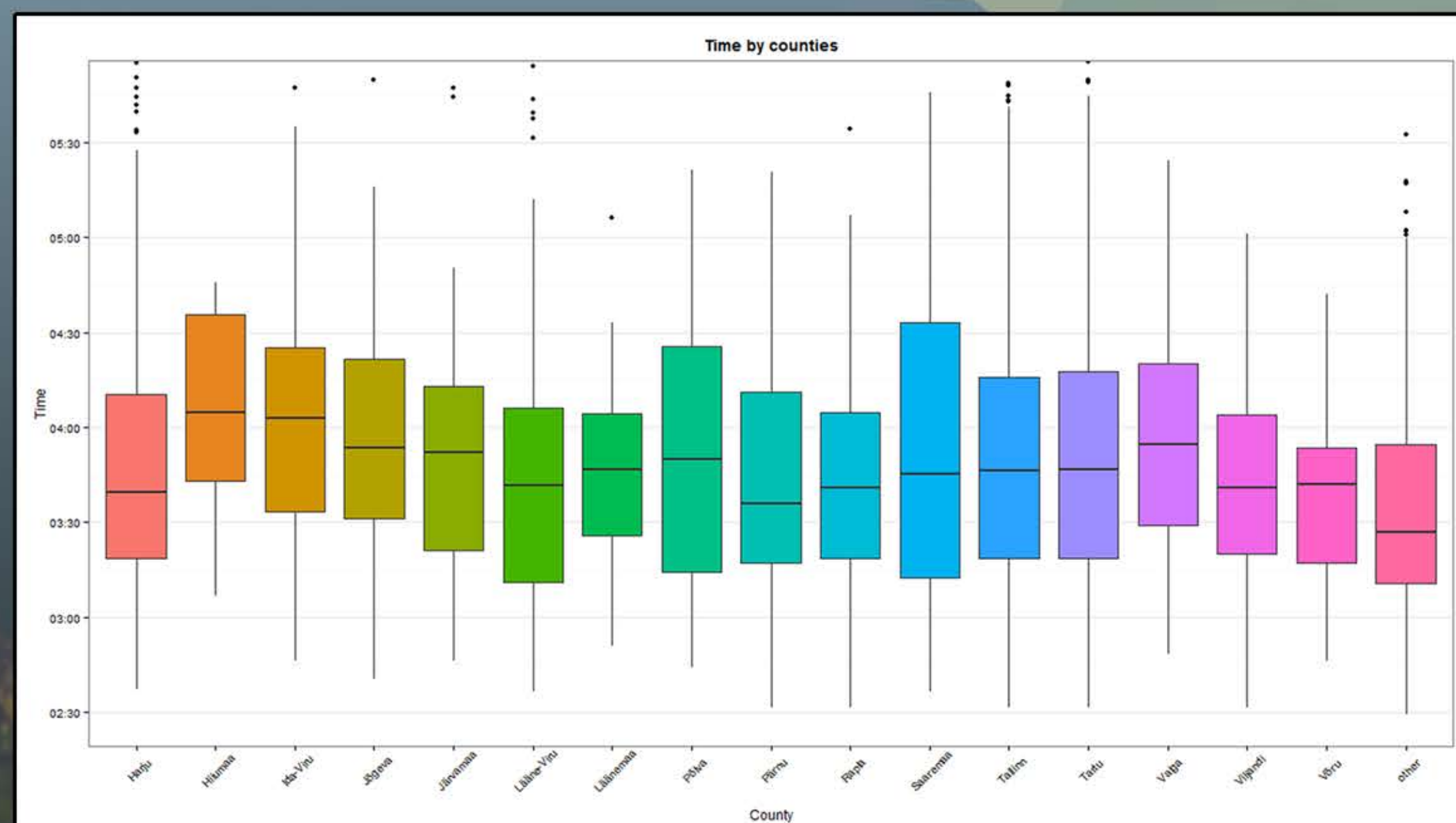
• 5% of people ride in pace group and their times are 15% better compared to solo competitors

• Average pace group size is 3

• 88% of people in the pace group have completed at least 1 marathon before

We tried to predict how well will someone do in a marathon, given person's age, nationality and how many times one had participated before. Based on the decision tree, we conclude that foreign participants between 17-21 will most likely have the best result.