Benford’s Project

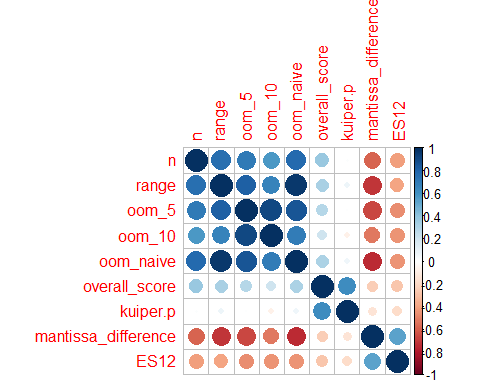
Figs 5,6,8,9,10: Summary Charts

Edward Stansfeld

January 2022

## Introduction

The India data contain 36 sets of regional data, each containing around 500 records, and the Scotland 14 sets of regional data. This paper focuses on the properties of the datasets for confirmed cases, deaths and recoveries. The distributions are plotted for each region and indicative statistics are calculated.



Conformant data was then encoded with a “1” and non-conformant with a “0” and a decision tree was created using all the data. This is helpful to see the relative importance of the various factors. The data was first split with 80% moved into the training set and 20% to the test set using the caTools package. The result shows the importance of the geographic area of the state, with the smaller states having higher conformancy at two levels of the tree. For the larger states, low literacy and an imbalanced sex ratio are associated with lower conformance.

