

Cricket Analytics with R

Samrit Pramanik

2023-11-27

Table of contents

Preface

This is a Quarto book on Cricket Analytics with application in R.

Like all other Indian boys, in my childhood, cricket was more than just a game, it is a profound emotion, a way of life, and an unrestrained source of happiness.

1 Warming Up

Data Collection

To know the game of cricket, you need to know the history, when and how it was started.

2 Data Collection and Sources

There are lot of data sources available online to get started our analysis.

Here, we will mainly focus on the ball-by-ball data so that we can achieve the insights from more granular level.

We will be using open source data from cricsheet.org.

2.1 Data Description

The raw data will be downloaded as a zip file. After unzipping we should have multiple .csv files that corresponds to each T20 International matches. There will be another file with same match ids but a suffix “__info” giving the information of match summary. Our first objective will be to collate all the ball by ball data in a single file or a data frame in R.

3 Powerplay

Data Cleaning and Wrangling

4 The Cricket Data

Cricket data is not so clean at when it was extracted from the database. Most of them are ball by ball on different formats such as T20I, ODI, Test, IPL, CPL etc.