

01DataWrangling

Rimjhim Saxena

2022-12-21

Getting data together

Use “cricketdata” r package for cricket data from ESPN

```
rm(list = ls())
library(cricketdata)
```

Get data for all test, odi and t20s

```
a = fetch_cricsheet(type = "bbb",gender = "male",competition = "tests")
b = fetch_cricsheet(type = "bbb",gender = "male",competition = "odis")
c = fetch_cricsheet(type = "bbb",gender = "male",competition = "t20is")
d = fetch_cricsheet(type = "bbb",gender = "male",competition = "t20is_unofficial")
f = fetch_cricsheet(type = "bbb",gender = "male",competition = "ipl")
g = rbind(c,d)
```

Writing the data

```
write.csv(a,"~/OneDrive - UCB-0365/CricketCC/Data/all-game/test.csv",row.names = F)
write.csv(b,"~/OneDrive - UCB-0365/CricketCC/Data/all-game/odi.csv",row.names = F)
write.csv(c,"~/OneDrive - UCB-0365/CricketCC/Data/all-game/t20.csv",row.names = F)
write.csv(d,"~/OneDrive - UCB-0365/CricketCC/Data/all-game/t20_unofficial.csv",row.names = F)
write.csv(f,"~/OneDrive - UCB-0365/CricketCC/Data/all-game/ipl.csv",row.names = F)
write.csv(g,"~/OneDrive - UCB-0365/CricketCC/Data/all-game/allt20.csv",row.names = F)
```

Subset data to recent seasons

```
a1 = a[which(a$season=="2022"|a$season=="2022/23"),]
b1 = b[which(b$season=="2022"|b$season=="2022/23"),]
c1 = c[which(c$season=="2022"|c$season=="2022/23"),]
d1 = d[which(d$season=="2022"|d$season=="2022/23"),]
f1 = f[which(f$season=="2022"|f$season=="2022/23"),]
g1 = g[which(g$season=="2022"|g$season=="2022/23"),]
```

Writing data for recent seasons

```
write.csv(a1,"~/OneDrive - UCB-0365/CricketCC/Data/recent-season/test.csv",row.names = F)
write.csv(b1,"~/OneDrive - UCB-0365/CricketCC/Data/recent-season/odi.csv",row.names = F)
write.csv(c1,"~/OneDrive - UCB-0365/CricketCC/Data/recent-season/t20.csv",row.names = F)
write.csv(d1,"~/OneDrive - UCB-0365/CricketCC/Data/recent-season/t20_unofficial.csv",row.names = F)
write.csv(f1,"~/OneDrive - UCB-0365/CricketCC/Data/recent-season/ipl.csv",row.names = F)
write.csv(g1,"~/OneDrive - UCB-0365/CricketCC/Data/recent-season/allt20.csv",row.names = F)
```

World cup games

```
c$start_date = as.Date(c$start_date)
wc = c[which(c$start_date>="2022-10-16" & c$start_date<="2022-11-13"),]
```

```
write.csv(wc,"~/OneDrive - UCB-0365/CricketCC/Data/worldcup-games/wcgames.csv",row.names = F)
```