*“He felt that his whole life was some kind of dream and he sometimes wondered whose it was and whether they were enjoying it.”*

*“The ships hung in the sky in much the same way that bricks don’t.”*

*“We demand rigidly defined areas of doubt and uncertainty!”*

*“For a moment, nothing happened. Then, after a second or so, nothing continued to happen.”*

*“The Answer to the Great Question… Of Life, the Universe and Everything… Is… Forty-two,’ said Deep Thought, with infinite majesty and calm.”*

The Hitchhiker's Guide to the Galaxy - Douglas Adams

**Introduction**

In this post, I introduce 2 new functions in my R package ‘cricketr’ (cricketr v0.22) see [Re-introducing cricketr! : An R package to analyze performances of cricketers](https://gigadom.in/2016/05/14/re-introducing-cricketr-an-r-package-to-analyze-performances-of-cricketers/) which enable granular analysis of batsmen and bowlers. They are

1. **Step 1**: getPlayerDataHA – This function is a wrapper around getPlayerData(), getPlayerDataOD() and getPlayerDataTT(), and adds an extra column ‘homeOrAway’ which says whether the match was played at home/away/neutral venues. A CSV file is created with this new column.
2. **Setp 2**: getPlayerDataOppnHA – This function allows you to slice & dice the data for batsmen and bowlers against specific oppositions, at home/away/neutral venues and between certain periods. This reduced subset of data can be used to perform analyses. A CSV file is created as an output based on the parameters of opposition, home or away and the interval of time

**Note** All the existing cricketr functions can be used on this smaller fine-grained data set for a closer analysis of players

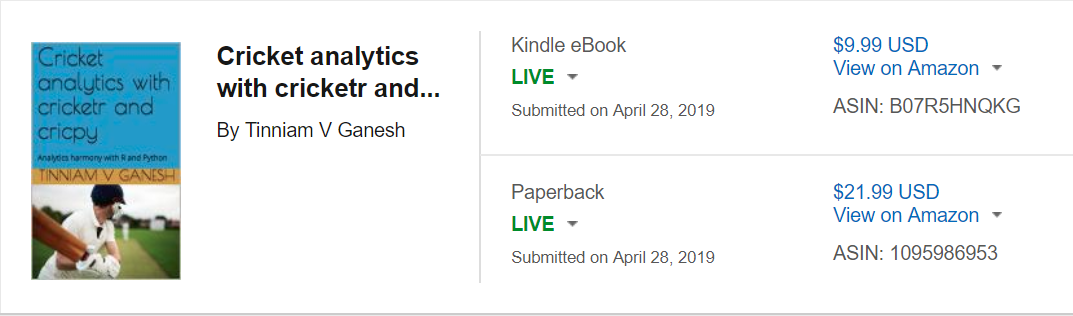
**Note 1**: You have to call the above functions only once. You can reuse the CSV files in other functions

**Important note**: Don’t go too fine-grained by choosing just one opposition, in one of home/away/neutral and for too short a period. Too small a dataset may defeat the purpose of the analysis!

This post has been published in Rpubs and can be accessed at [Cricketr learns new tricks](http://rpubs.com/tvganesh/512168)

You can download a PDF version of this post at [Cricketr learns new tricks](https://drive.google.com/file/d/1zsmHLd3vWL3V02msdviZoz50hXALqKDS/view?usp=sharing)

If you are passionate about cricket, and love analyzing cricket performances, then check out my racy book on cricket ‘Cricket analytics with cricketr and cricpy – Analytics harmony with R & Python’! This book discusses and shows how to use my R package ‘cricketr’ and my Python package ‘cricpy’ to analyze batsmen and bowlers in all formats of the game (Test, ODI and T20). The [paperback](https://www.amazon.com/dp/1095986953) is available on Amazon at $21.99 and  the [kindle](https://www.amazon.com/dp/B07R5HNQKG) version at $9.99/Rs 449/-. A must read for any cricket lover! Check it out!!



**1. Analyzing Tendulkar at 3 different stages of his career**

The following functions analyze Sachin Tendulkar during 3 different periods of his illustrious career. a) 1st Jan 2001-1st Jan 2002 b) 1st Jan 2005-1st Jan 2006 c) 1st Jan 2012-1st Jan 2013

# Get the homeOrAway dataset for Tendulkar in matches

#Note: I have commented the lines to getPlayerDataHA() as I already have

# CSV file

#df=getPlayerDataHA(35320,tfile="tendulkarTestHA.csv",matchType="Test")

# Get Tendulkar's data for 2001-02

df1=getPlayerDataOppnHA(infile="tendulkarHA.csv",outfile="tendulkarTest2001.csv",

startDate="2001-01-01",endDate="2002-01-01")

# Get Tendulkar's data for 2005-06

df2=getPlayerDataOppnHA(infile="tendulkarHA.csv",outfile="tendulkarTest2005.csv",

startDate="2005-01-01",endDate="2006-01-01")

# Get Tendulkar's data for 20012-13

#df3=getPlayerDataOppnHA(infile="tendulkarHA.csv",outfile="tendulkarTest2012.csv",

# startDate="2012-01-01",endDate="2013-01-01")

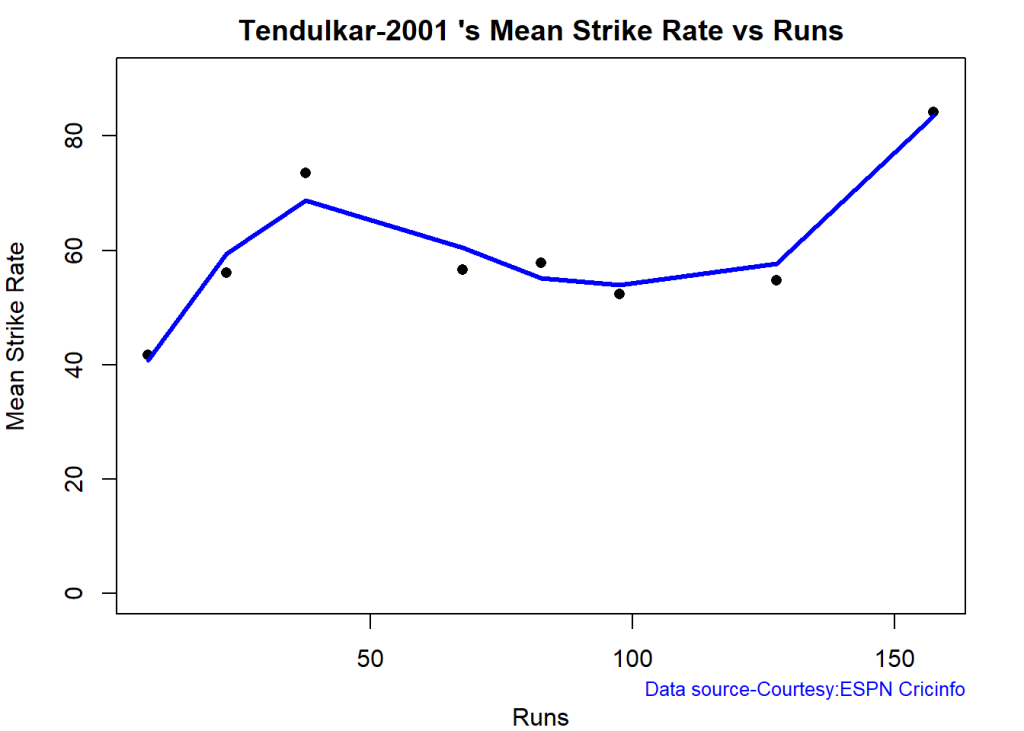
`

**1a Mean strike rate of Tendulkar in 2001,2005,2012**

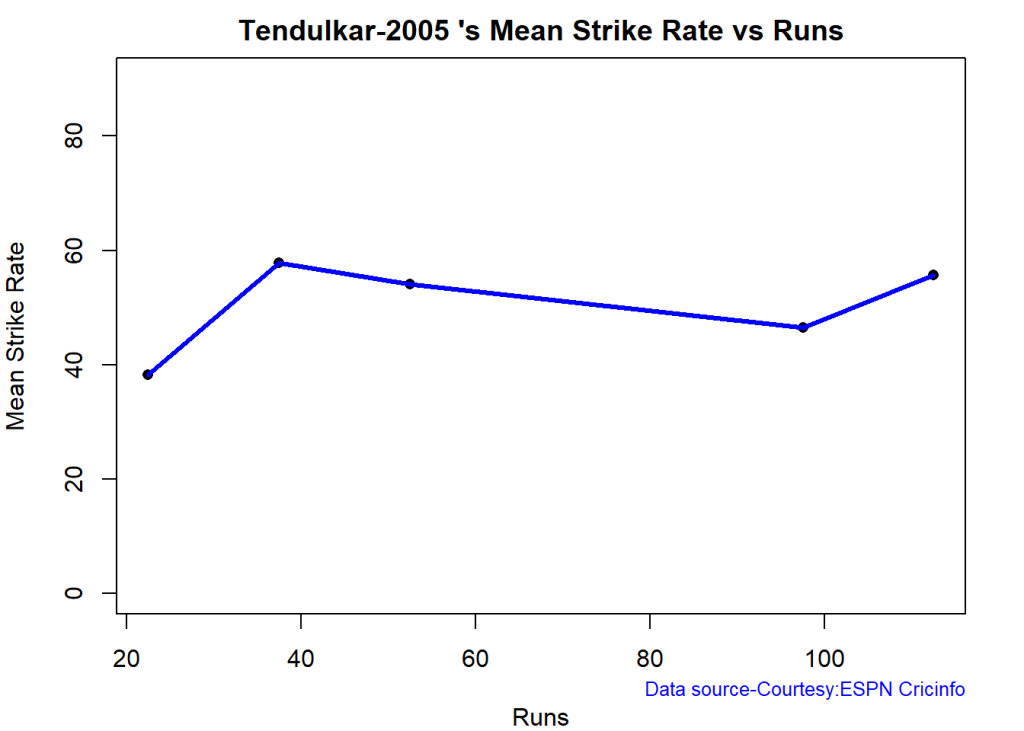
Note: Any of the cricketr R functions can be used on the fine-grained subset of data as below. The mean strike rate of Tendulkar is of the order of 60+ in 2001 which decreases to 50 and later to around 45

# Compute and plot mean strike rate of Tendulkar in the 3 periods

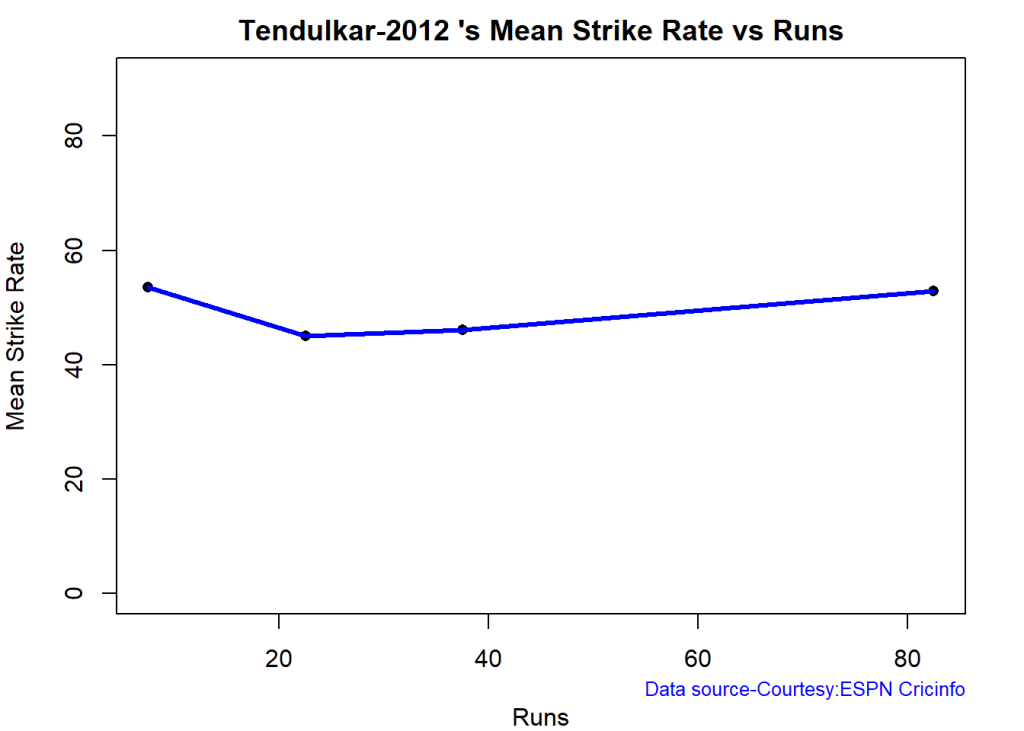
batsmanMeanStrikeRate ("./tendulkarTest2001.csv","Tendulkar-2001")



batsmanMeanStrikeRate ("./tendulkarTest2005.csv","Tendulkar-2005")



batsmanMeanStrikeRate ("./tendulkarTest2012.csv","Tendulkar-2012")



**1b. Plot the performance of Tendulkar at venues during 2001,2005,2012**

On an average Tendulkar score 60+ in 2001 and is really blazing. This performance decreases in 2005 and later in 2012

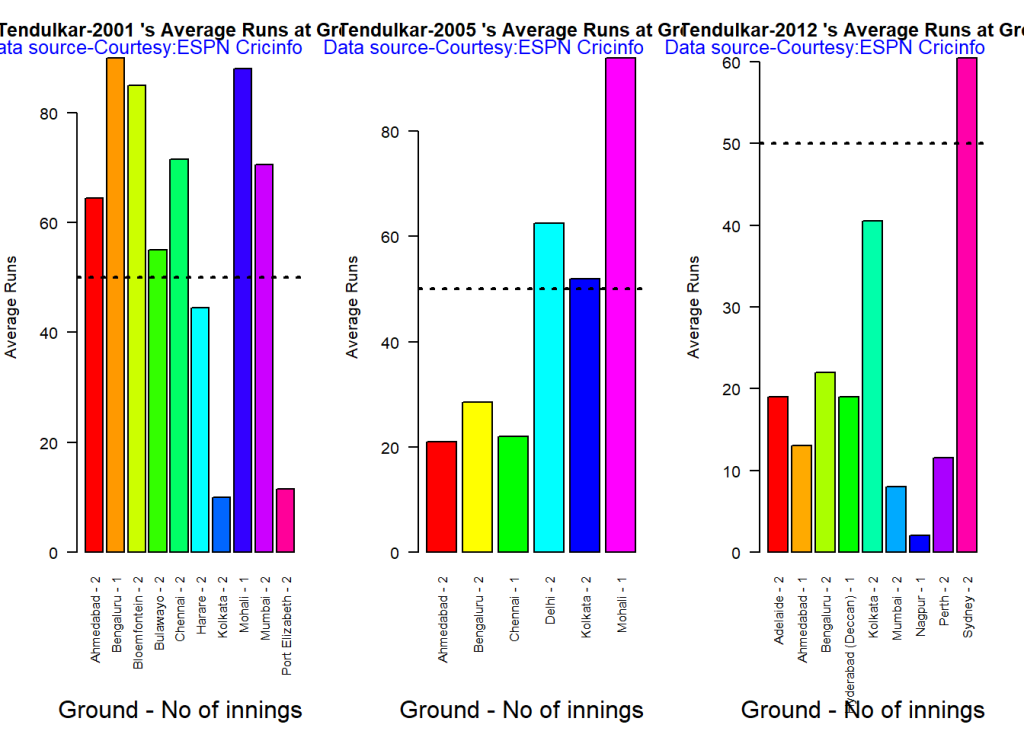
par(mfrow=c(1,3))

par(mar=c(4,4,2,2))

batsmanAvgRunsGround("tendulkarTest2001.csv","Tendulkar-2001")

batsmanAvgRunsGround("tendulkarTest2005.csv","Tendulkar-2005")

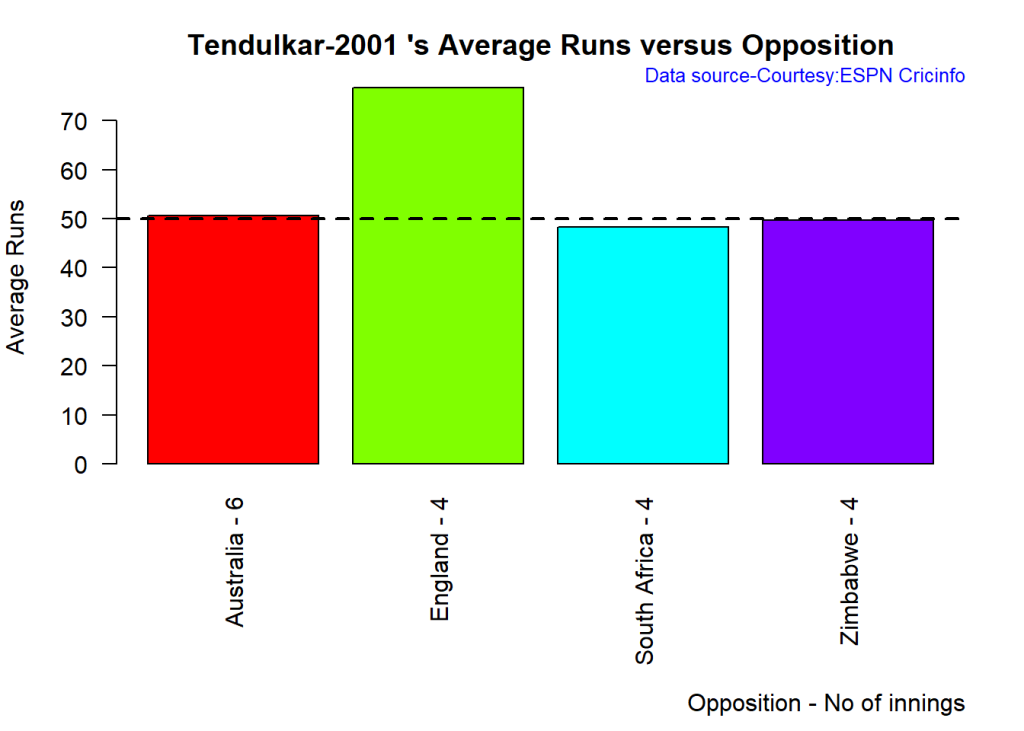
batsmanAvgRunsGround("tendulkarTest2012.csv","Tendulkar-2012")



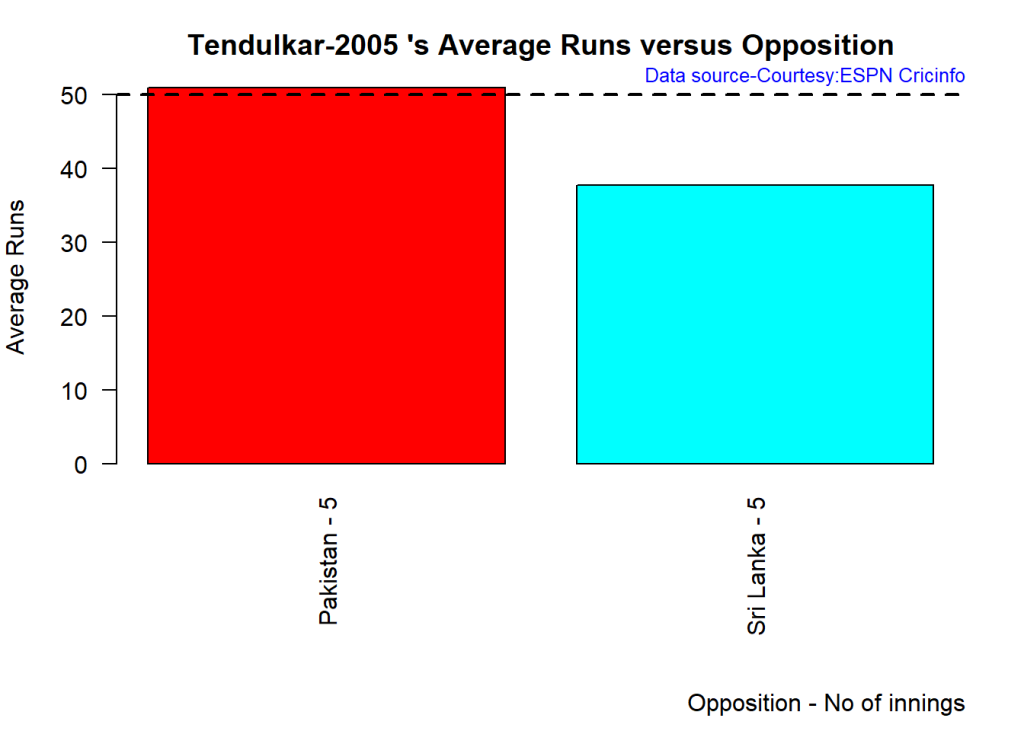
dev.off()

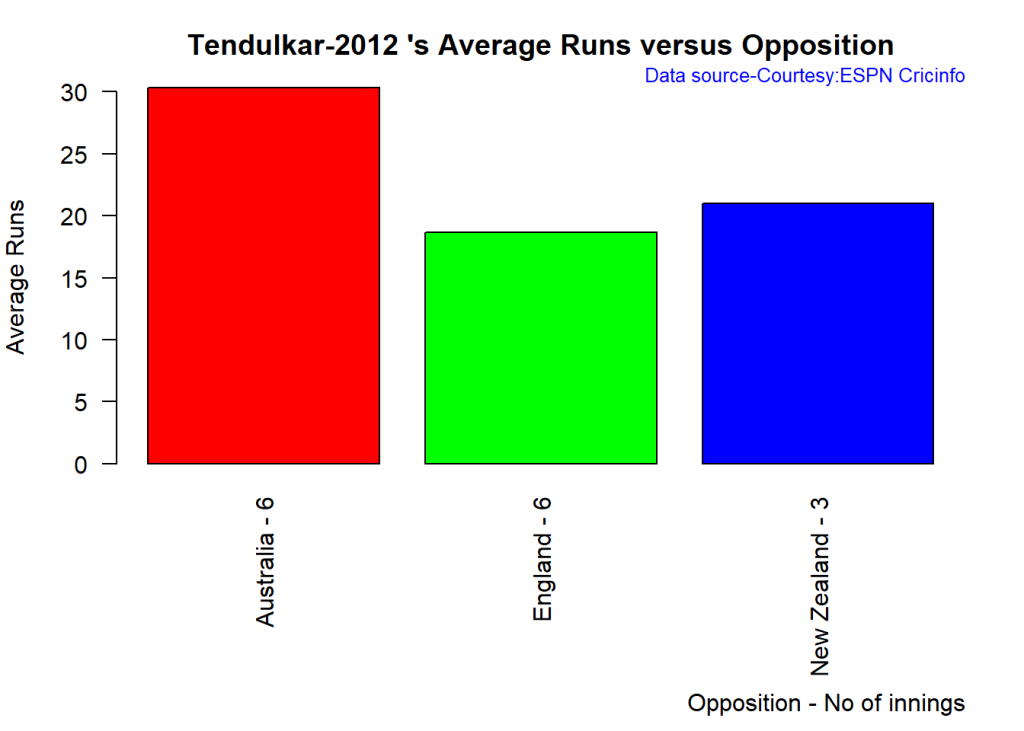
**1c. Plot the performance of Tendulkar against different oppositions during 2001,2005,2012**

Sachin uniformly scores 50+ against formidable oppositions in 2001. In 2005 this decreases to 40 in 2005 and in 2012 it is around 25

batsmanAvgRunsOpposition("tendulkarTest2001.csv","Tendulkar-2001")

batsmanAvgRunsOpposition("tendulkarTest2005.csv","Tendulkar-2005")



batsmanAvgRunsOpposition("tendulkarTest2012.csv","Tendulkar-2012")

**1d. Plot the relative cumulative average and relative strike rate of Tendulkar in 2001,2005,2012**

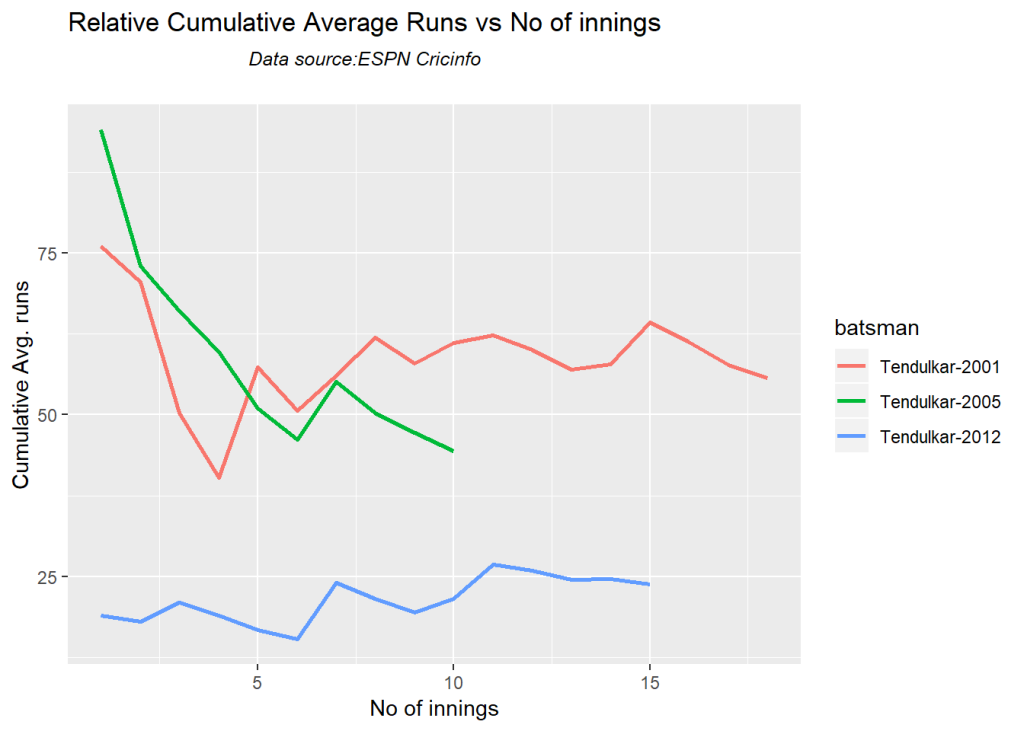
The plot below compares Tendulkar’s cumulative strike rate and cumulative average during 3 different stages of his career

1. The cumulative average runs of Tendulkar is in the high 60+ in 2001, which drops to ~50 in 2005 and later plummets to the low 25s in 2012
2. The strike rate in 2001 for Tendulkar is amazing 60+

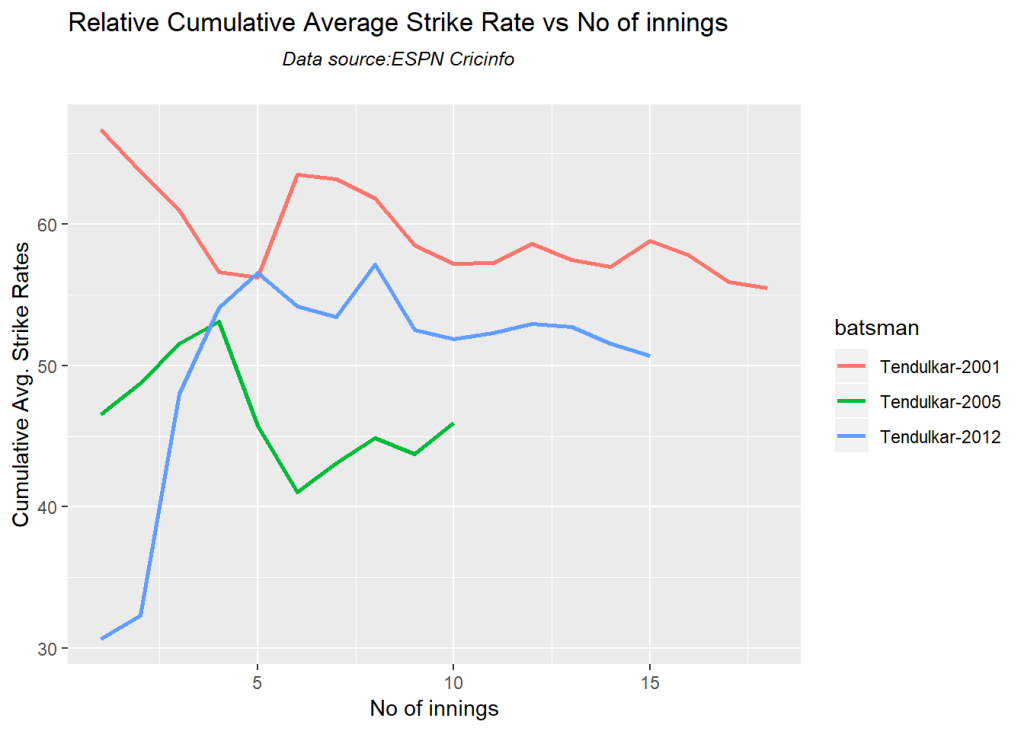
frames=list("tendulkarTest2001.csv","tendulkarTest2005.csv","tendulkarTest2012.csv")

names=list("Tendulkar-2001","Tendulkar-2005","Tendulkar-2012")

relativeBatsmanCumulativeAvgRuns(frames,names)



relativeBatsmanCumulativeStrikeRate(frames,names)



**2. Analyzing Virat Kohli’s performance against England in England in 2014 and 2018**

The analysis below looks at Kohli’s performance against England in ‘away’ venues (England) in 2014 and 2018

# Get the homeOrAway data for Kohli in Test matches

#df=getPlayerDataHA(253802,tfile="kohliTestHA.csv",type="batting",matchType="Test")

# Get the subset if data of Kohli's performance against England in England in 2014

df=getPlayerDataOppnHA(infile="kohliTestHA.csv",outfile="kohliTestEng2014.csv",

opposition=c("England"),homeOrAway=c("away"),startDate="2014-01-01",endDate="2015-01-01")

# Get the subset if data of Kohli's performance against England in England in 2018

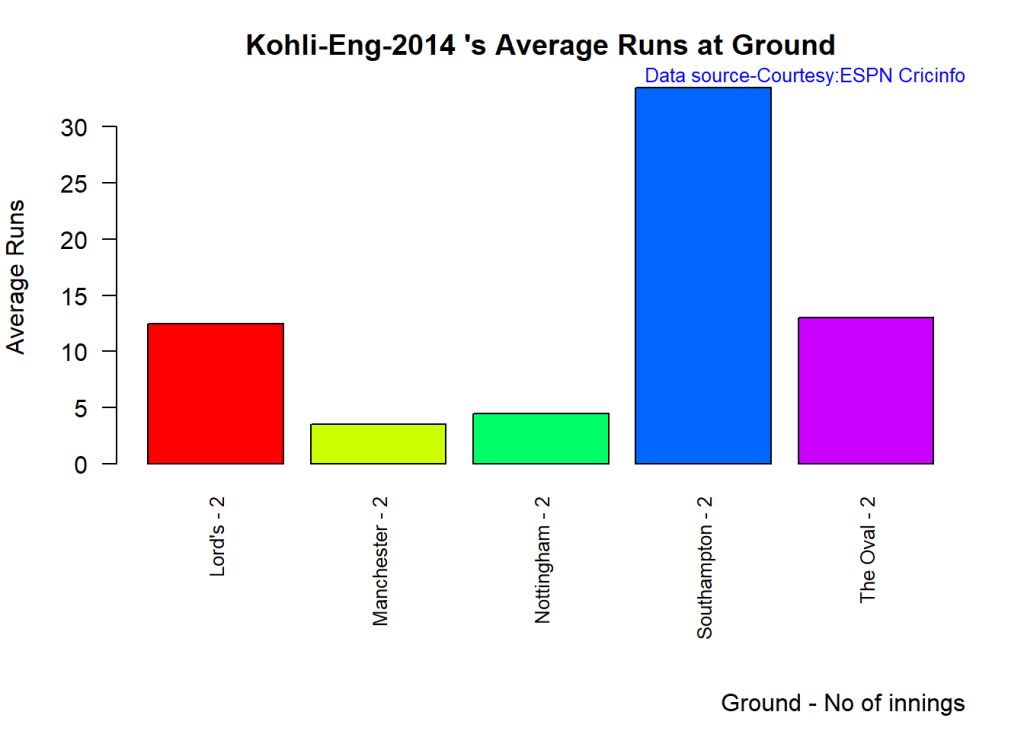
df1=getPlayerDataOppnHA(infile="kohliHA.csv",outfile="kohliTestEng2018.csv",

opposition=c("England"),homeOrAway=c("away"),startDate="2018-01-01",endDate="2019-01-01")

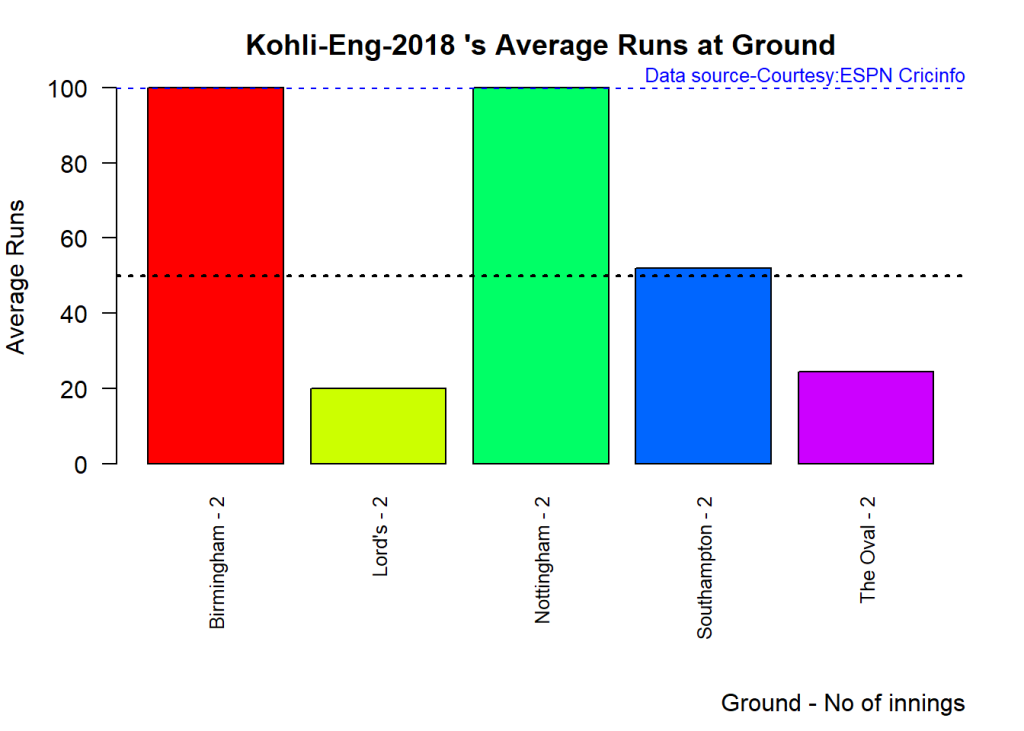
**2a. Kohli’s performance at England grounds in 2014 & 2018**

Kohli had a miserable outing to England in 2014 with a string of low scores. In 2018 Kohli pulls himself out of the morass

batsmanAvgRunsGround("kohliTestEng2014.csv","Kohli-Eng-2014")



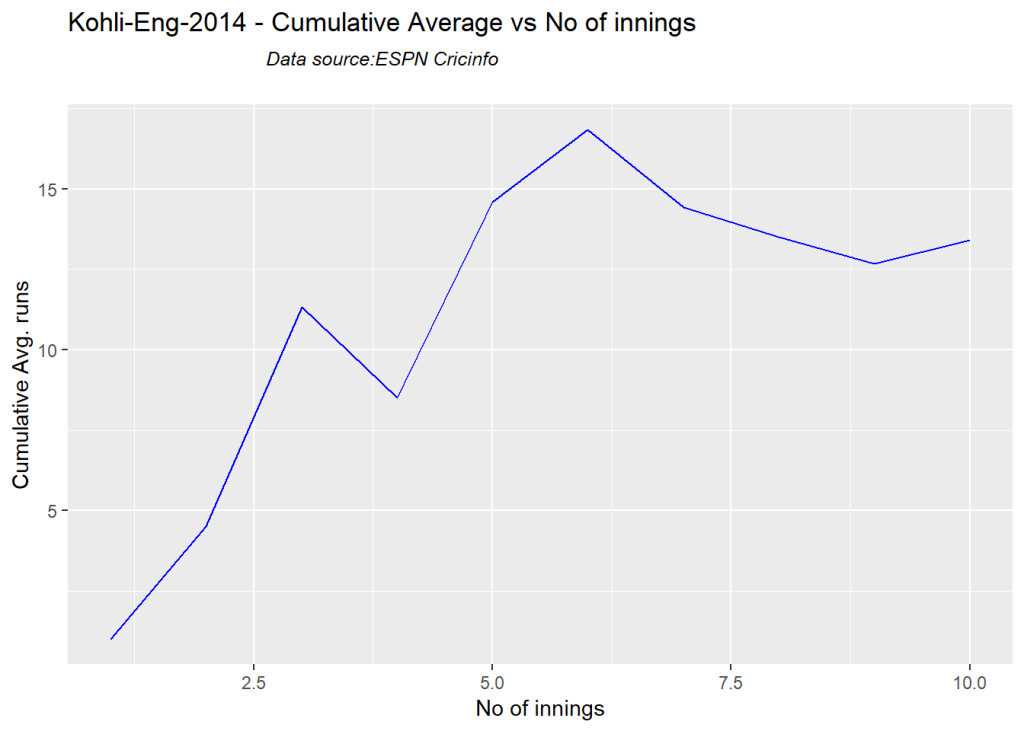
batsmanAvgRunsGround("kohliTestEng2018.csv","Kohli-Eng-2018")



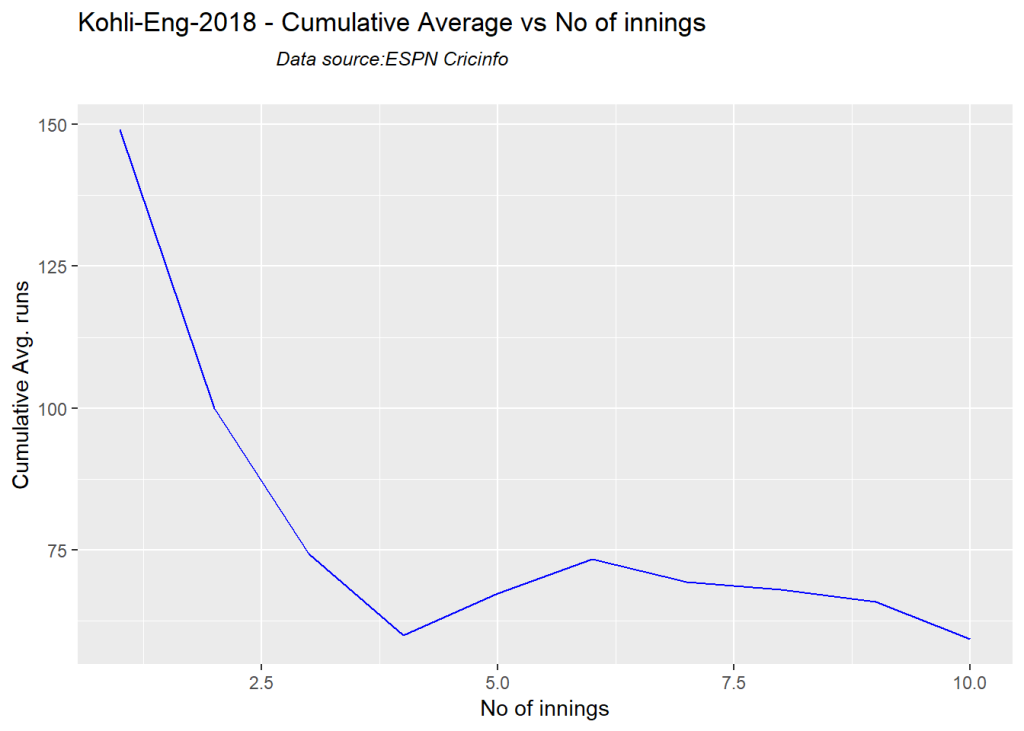
**2a. Kohli’s cumulative average runs in 2014 & 2018**

Kohli’s cumulative average runs in 2014 is in the low 15s, while in 2018 it is 70+. Kohli stamps his class back again and undoes the bad memories of 2014

batsmanCumulativeAverageRuns("kohliTestEng2014.csv", "Kohli-Eng-2014")



batsmanCumulativeAverageRuns("kohliTestEng2018.csv", "Kohli-Eng-2018")



**3a. Compare the performances of Ganguly, Dravid and VVS Laxman against opposition in ‘away’ matches in Tests**

The analyses below compares the performances of Sourav Ganguly, Rahul Dravid and VVS Laxman against Australia, South Africa, and England in ‘away’ venues between 01 Jan 2002 to 01 Jan 2008

#Get the HA data for Ganguly, Dravid and Laxman

#df=getPlayerDataHA(28779,tfile="gangulyTestHA.csv",type="batting",matchType="Test")

#df=getPlayerDataHA(28114,tfile="dravidTestHA.csv",type="batting",matchType="Test")

#df=getPlayerDataHA(30750,tfile="laxmanTestHA.csv",type="batting",matchType="Test")

# Slice the data

df=getPlayerDataOppnHA(infile="gangulyTestHA.csv",outfile="gangulyTestAES2002-08.csv"

,opposition=c("Australia", "England", "South Africa"),

homeOrAway=c("away"),startDate="2002-01-01",endDate="2008-01-01")

df=getPlayerDataOppnHA(infile="dravidTestHA.csv",outfile="dravidTestAES2002-08.csv"

,opposition=c("Australia", "England", "South Africa"),

homeOrAway=c("away"),startDate="2002-01-01",endDate="2008-01-01")

df=getPlayerDataOppnHA(infile="laxmanTestHA.csv",outfile="laxmanTestAES2002-08.csv"

,opposition=c("Australia", "England", "South Africa"),

homeOrAway=c("away"),startDate="2002-01-01",endDate="2008-01-01")

**3b Plot the relative cumulative average runs and relative cumative strike rate**

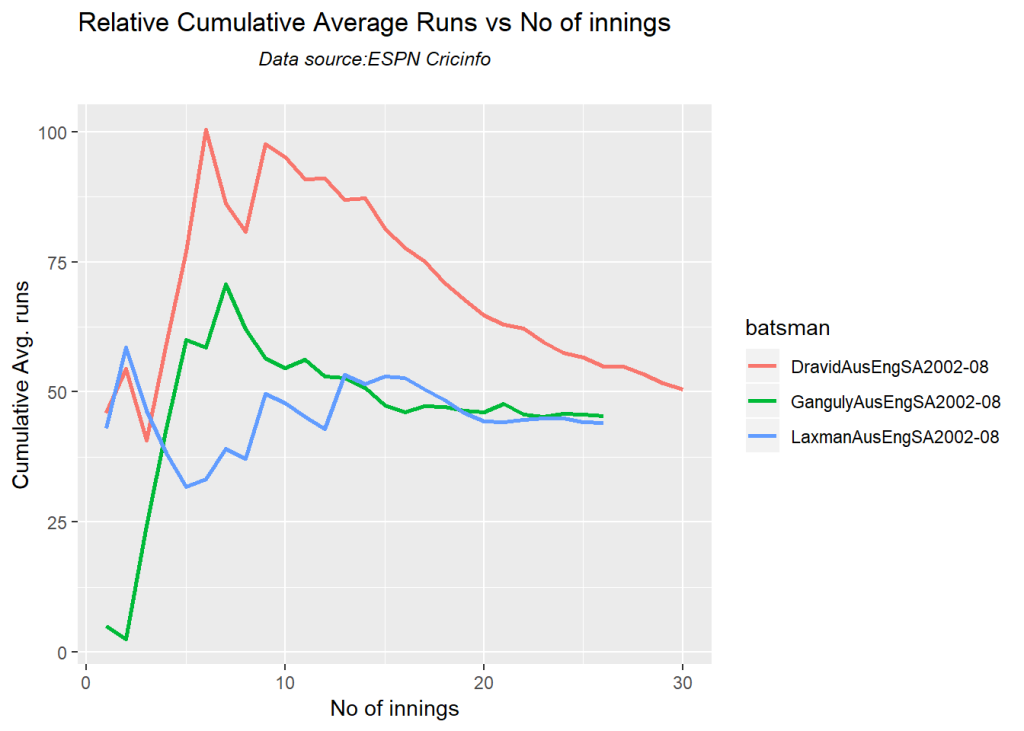
Plot the relative cumulative average runs and relative cumative strike rate of Ganguly, Dravid and Laxman

-Dravid towers over Laxman and Ganguly with respect to cumulative average runs. – Ganguly has a superior strike rate followed by Laxman and then Dravid

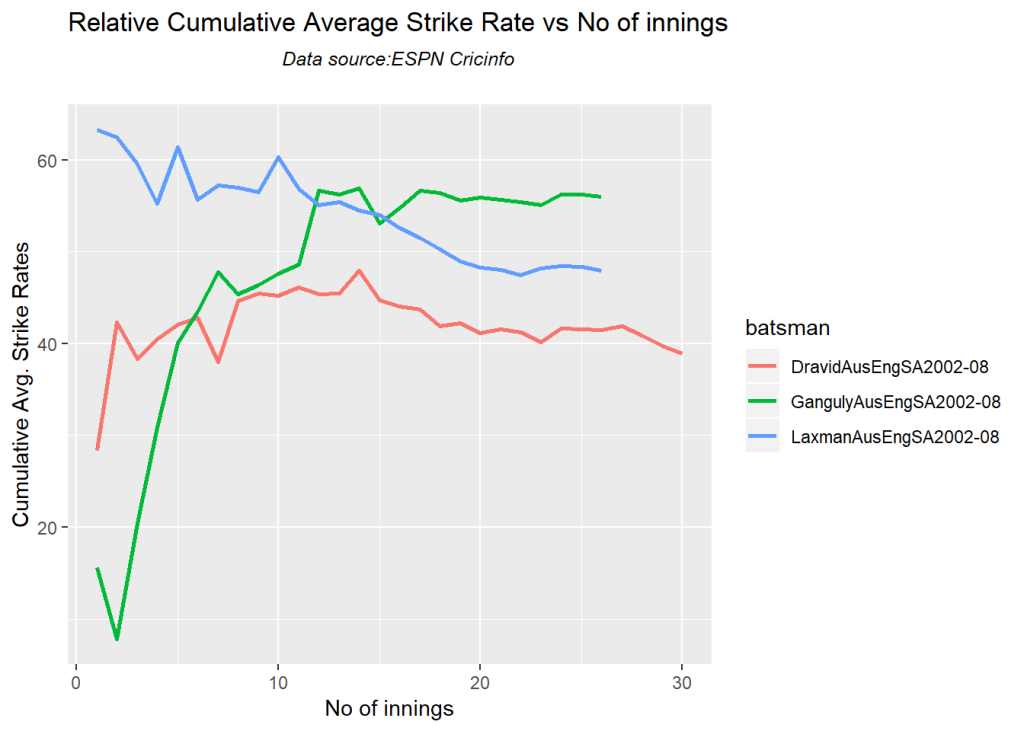
frames=list("gangulyTestAES2002-08.csv","dravidTestAES2002-08.csv","laxmanTestAES2002-08.csv")

names=list("GangulyAusEngSA2002-08","DravidAusEngSA2002-08","LaxmanAusEngSA2002-08")

relativeBatsmanCumulativeAvgRuns(frames,names)



relativeBatsmanCumulativeStrikeRate(frames,names)



**4. Compare the ODI performances of Rohit Sharma, Joe Root and Kane Williamson against opposition**

Compare the performances of Rohit Sharma, Joe Root and Kane williamson in away & neutral venues against Australia, West Indies and Soouth Africa

* Joe Root piles us the runs in about 15 matches. Rohit has played far more ODIs than the other two and averages a steady 35+

# Get the ODI HA data for Rohit, Root and Williamson

#df=getPlayerDataHA(34102,tfile="rohitODIHA.csv",type="batting",matchType="ODI")

#df=getPlayerDataHA(303669,tfile="joerootODIHA.csv",type="batting",matchType="ODI")

#df=getPlayerDataHA(277906,tfile="williamsonODIHA.csv",type="batting",matchType="ODI")

# Subset the data for specific opposition in away and neutral venues

df=getPlayerDataOppnHA(infile="rohitODIHA.csv",outfile="rohitODIAusWISA.csv"

,opposition=c("Australia", "West Indies", "South Africa"),

homeOrAway=c("away","neutral"))

df=getPlayerDataOppnHA(infile="joerootODIHA.csv",outfile="joerootODIAusWISA.csv"

,opposition=c("Australia", "West Indies", "South Africa"),

homeOrAway=c("away","neutral"))

df=getPlayerDataOppnHA(infile="williamsonODIHA.csv",outfile="williamsonODIAusWiSA.csv"

,opposition=c("Australia", "West Indies", "South Africa"),

homeOrAway=c("away","neutral"))

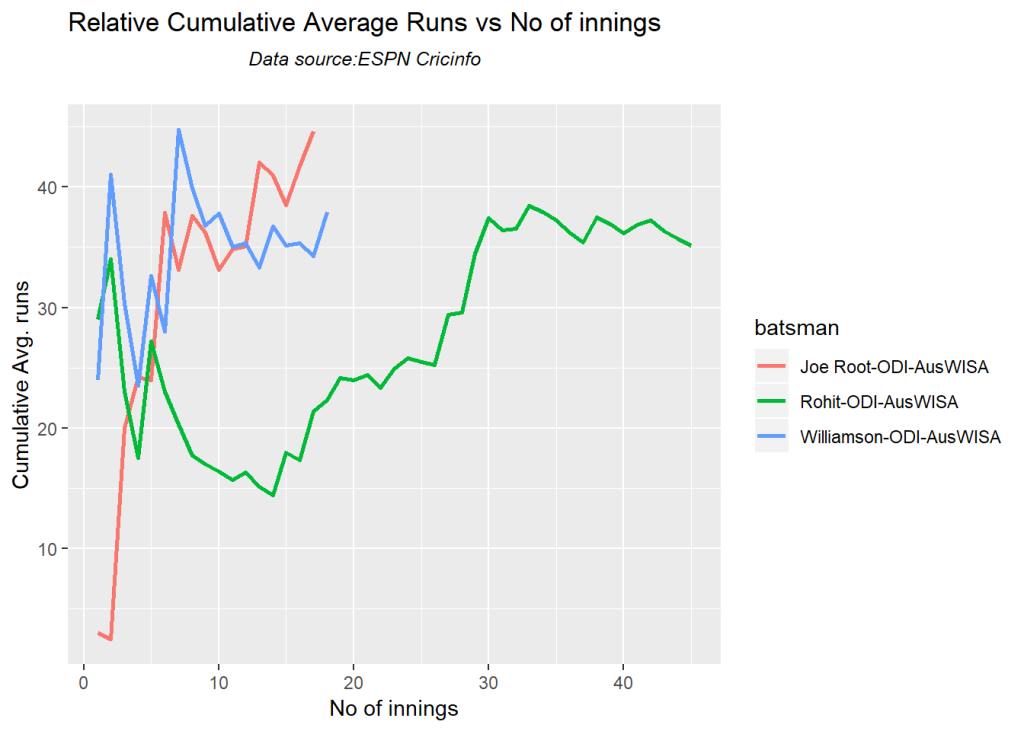
**4a. Compare cumulative strike rates and cumulative average runs of Rohit, Root and Williamson**

The relative cumulative strike rate of all 3 are comparable

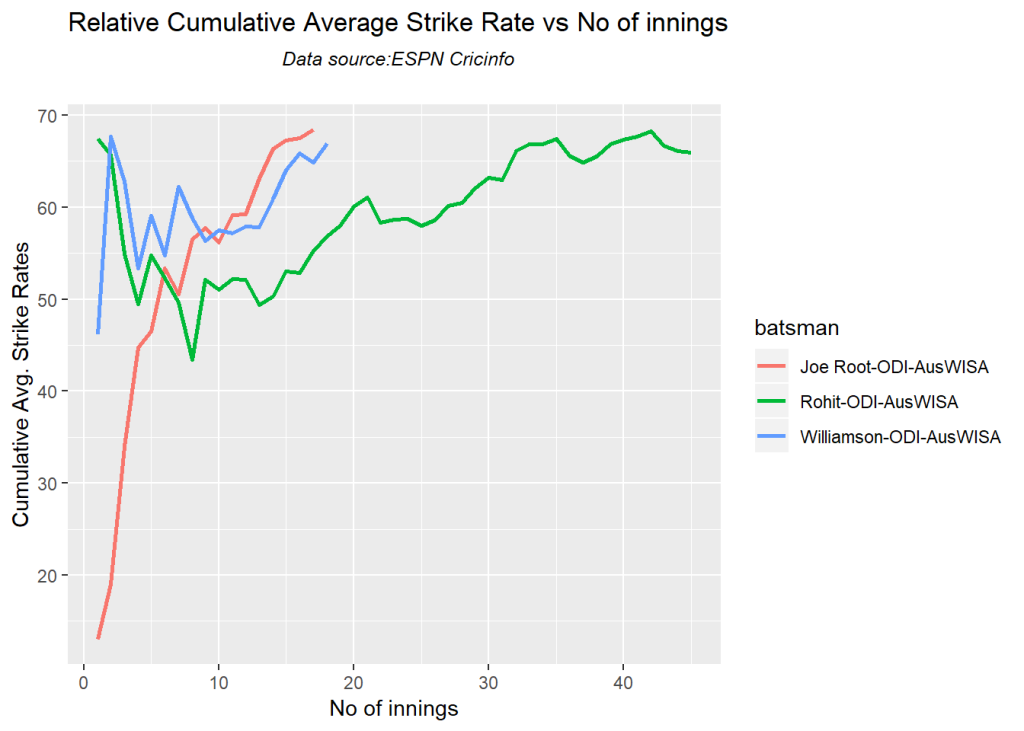
frames=list("rohitODIAusWISA.csv","joerootODIAusWISA.csv","williamsonODIAusWiSA.csv")

names=list("Rohit-ODI-AusWISA","Joe Root-ODI-AusWISA","Williamson-ODI-AusWISA")

relativeBatsmanCumulativeAvgRuns(frames,names)



relativeBatsmanCumulativeStrikeRate(frames,names)



**5. Plot the performance of Dhoni in T20s against specific opposition at all venues**

Plot the performances of Dhoni against Australia, West Indies, South Africa and England

# Get the HA T20 data for Dhoni

#df=getPlayerDataHA(28081,tfile="dhoniT20HA.csv",type="batting",matchType="T20")

#Subset the data

df=getPlayerDataOppnHA(infile="dhoniT20HA.csv",outfile="dhoniT20AusWISAEng.csv"

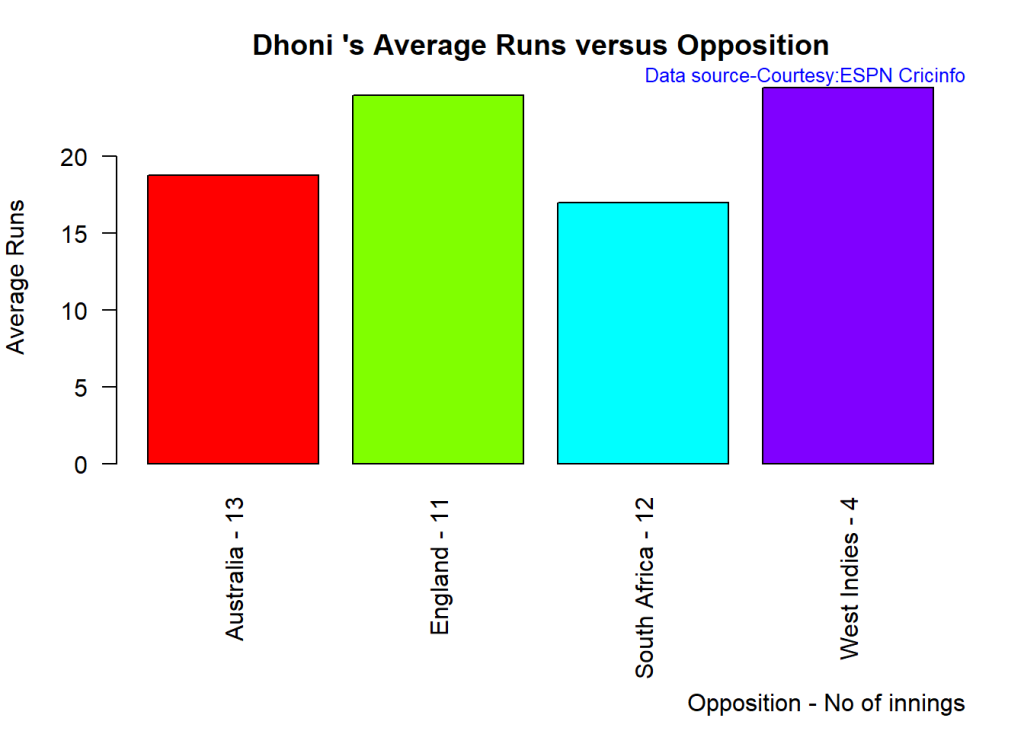
,opposition=c("Australia", "West Indies", "South Africa","England"),

homeOrAway=c("all"))

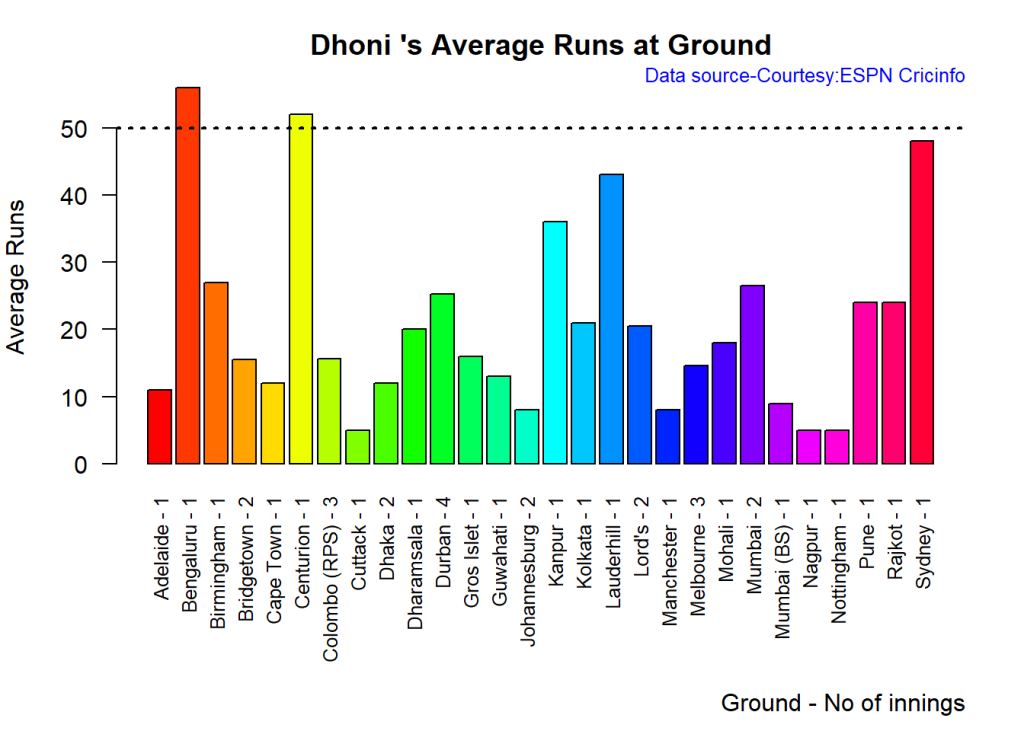
**5a. Plot Dhoni’s performances in T20**

**Note** You can use any of cricketr’s functions against the fine grained data

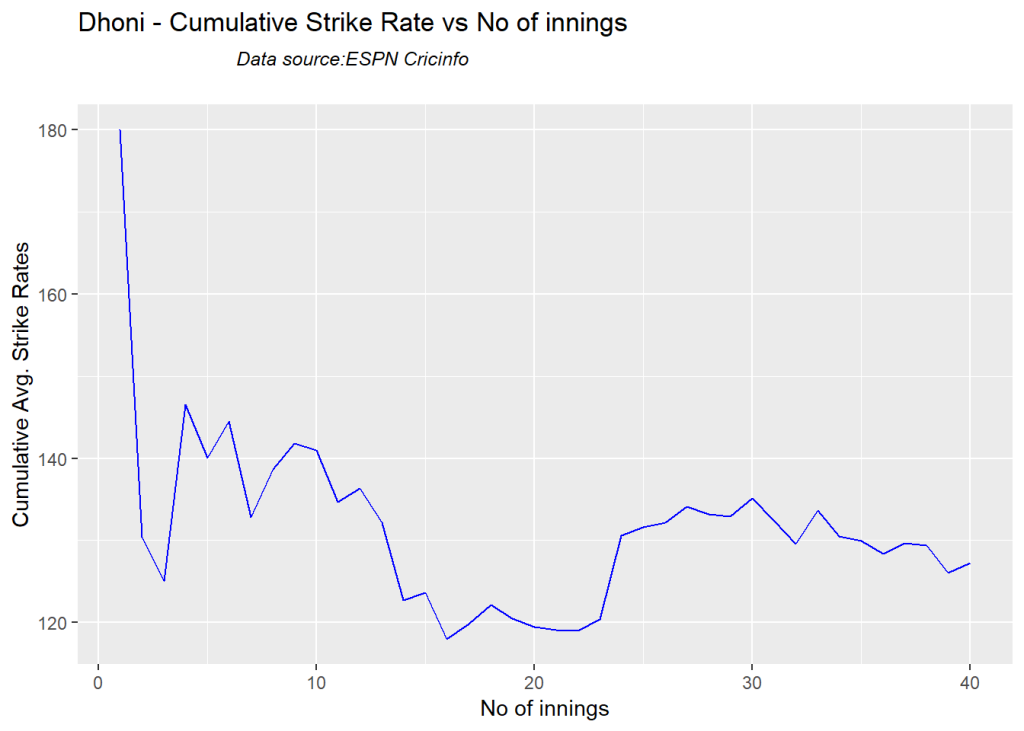
batsmanAvgRunsOpposition("dhoniT20AusWISAEng.csv","Dhoni")



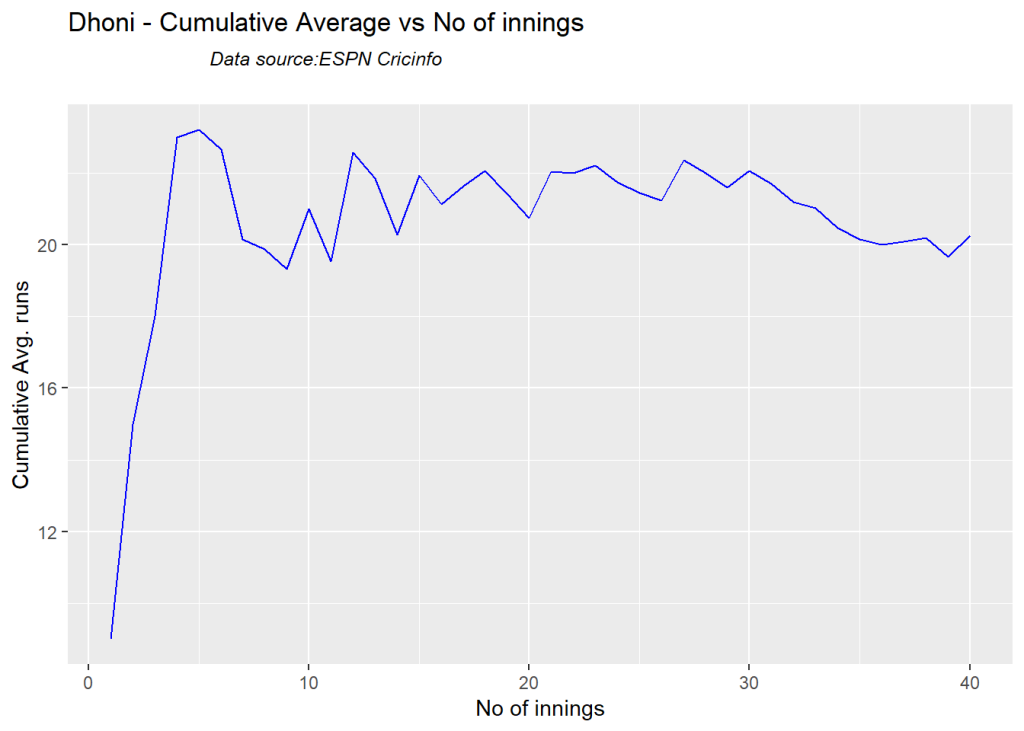
batsmanAvgRunsGround("dhoniT20AusWISAEng.csv","Dhoni")



batsmanCumulativeStrikeRate("dhoniT20AusWISAEng.csv","Dhoni")



batsmanCumulativeAverageRuns("dhoniT20AusWISAEng.csv","Dhoni")



**6. Compute and performances of Anil Kumble, Muralitharan and Warne in ‘away’ test matches**

Compute the performances of Kumble, Warne and Maralitharan against New Zealand, West Indies, South Africa and England in pitches that are not ‘home’ pithes

# Get the bowling data for Kumble, Warne and Muralitharan in Test matches

#df=getPlayerDataHA(30176,tfile="kumbleTestHA.csv",type="bowling",matchType="Test")

#df=getPlayerDataHA(8166,tfile="warneTestHA.csv",type="bowling",matchType="Test")

#df=getPlayerDataHA(49636,tfile="muraliTestHA.csv",type="bowling",matchType="Test")

# Subset the data

df=getPlayerDataOppnHA(infile="kumbleTestHA.csv",outfile="kumbleTest-NZWISAEng.csv"

,opposition=c("New Zealand", "West Indies", "South Africa","England"),

homeOrAway=c("away"))

df=getPlayerDataOppnHA(infile="warneTestHA.csv",outfile="warneTest-NZWISAEng.csv"

,opposition=c("New Zealand", "West Indies", "South Africa","England"),

homeOrAway=c("away"))

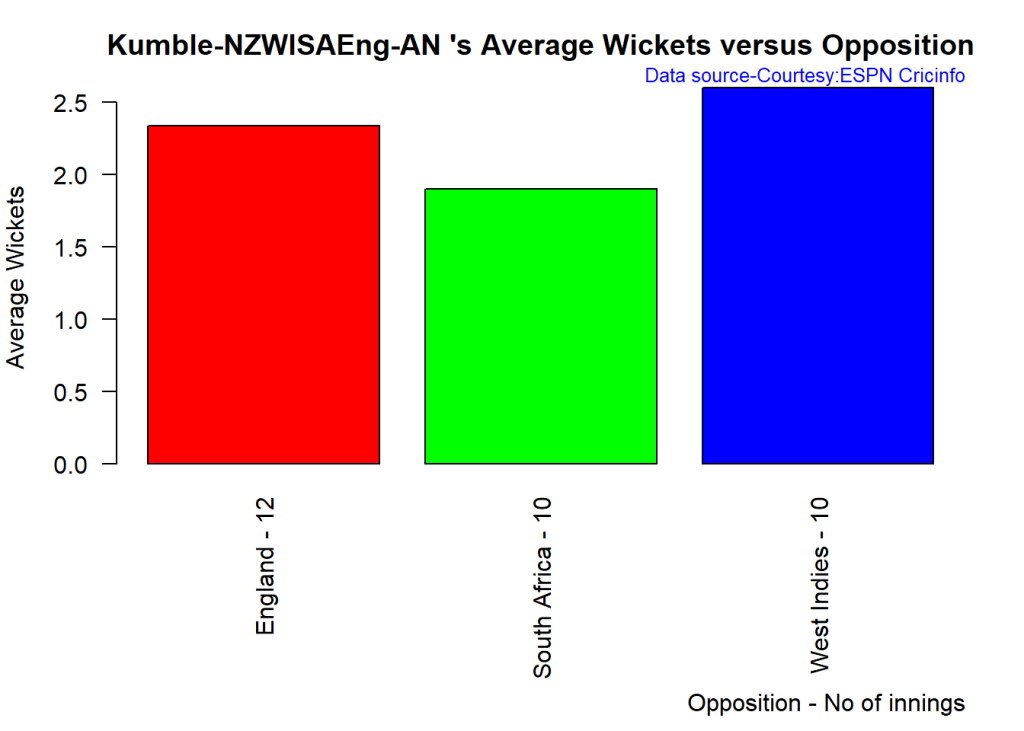
df=getPlayerDataOppnHA(infile="muraliTestHA.csv",outfile="muraliTest-NZWISAEng.csv"

,opposition=c("New Zealand", "West Indies", "South Africa","England"),

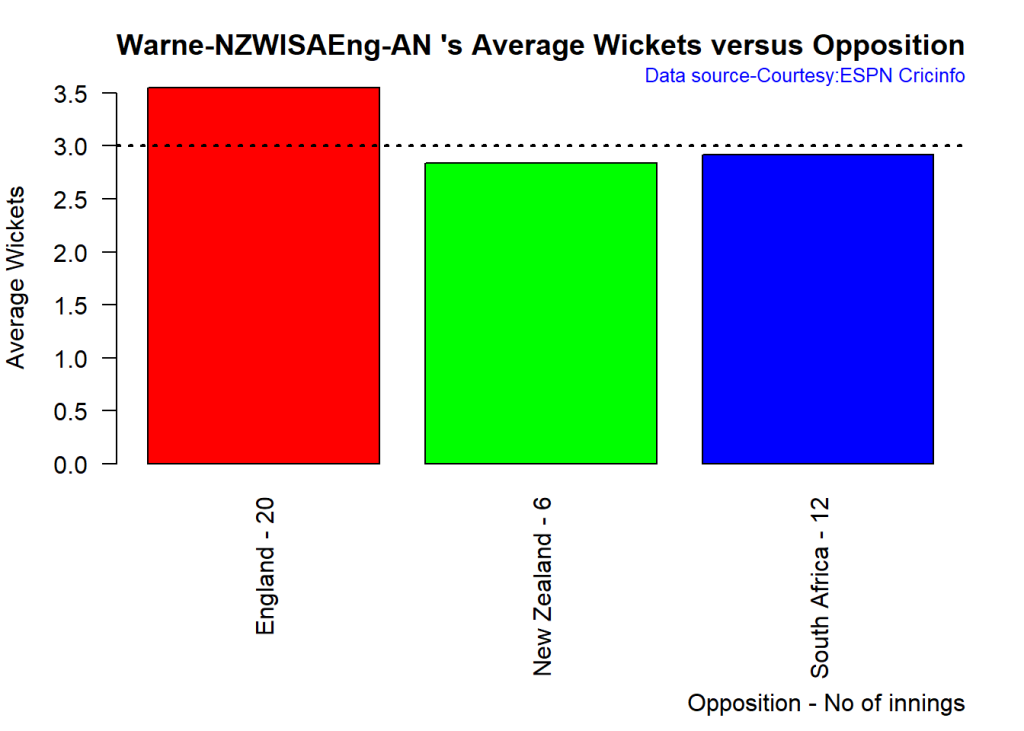
homeOrAway=c("away"))

**6a. Plot the average wickets of Kumble, Warne and Murali**

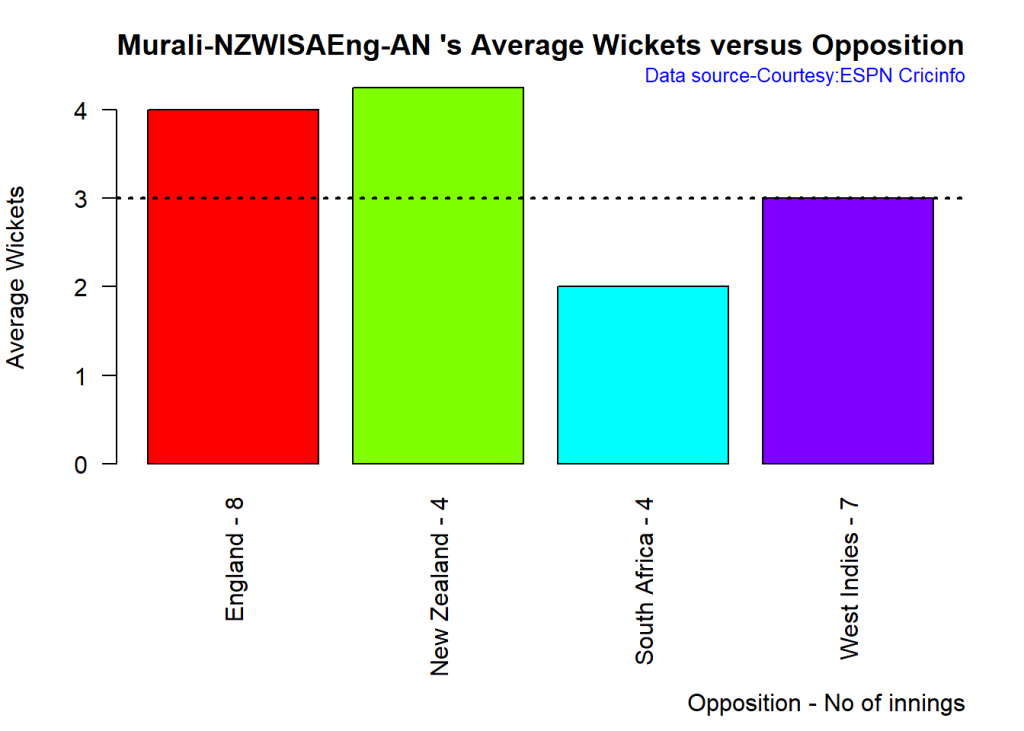
bowlerAvgWktsOpposition("kumbleTest-NZWISAEng.csv","Kumble-NZWISAEng-AN")



bowlerAvgWktsOpposition("warneTest-NZWISAEng.csv","Warne-NZWISAEng-AN")

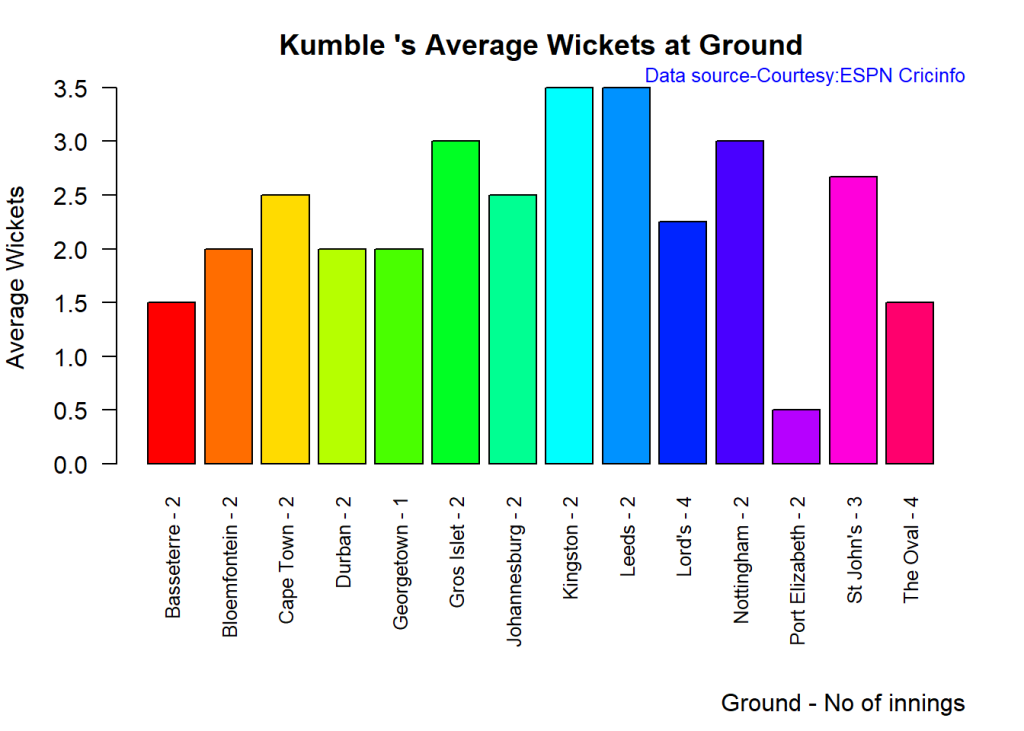


bowlerAvgWktsOpposition("muraliTest-NZWISAEng.csv","Murali-NZWISAEng-AN")

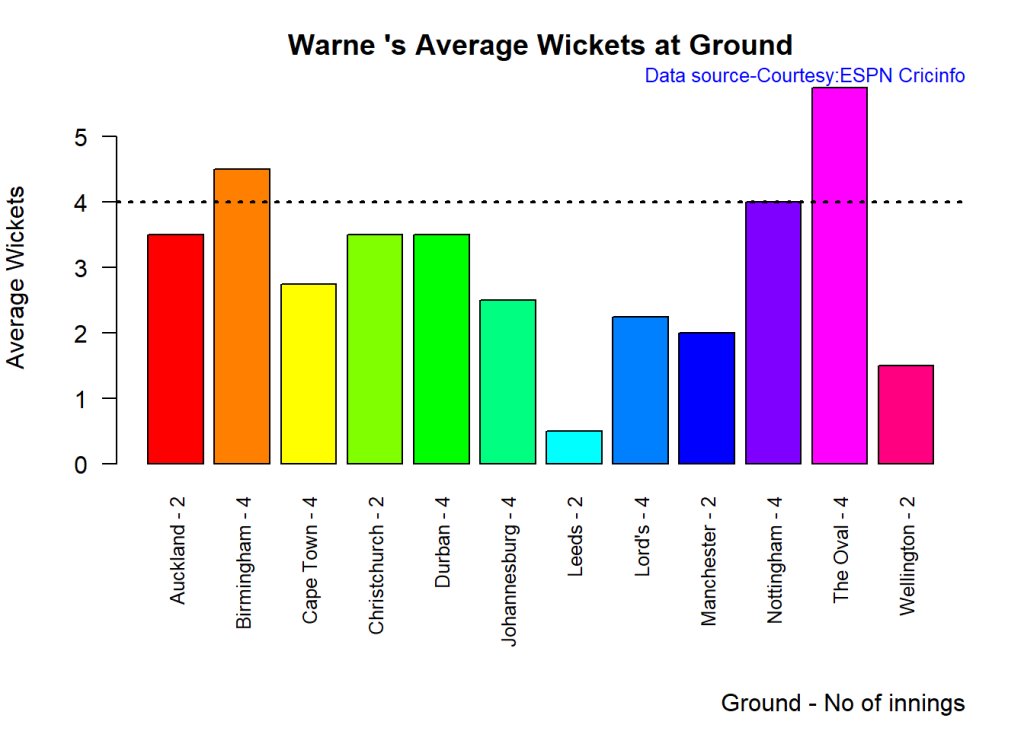


**6b. Plot the average wickets in different grounds of Kumble, Warne and Murali**

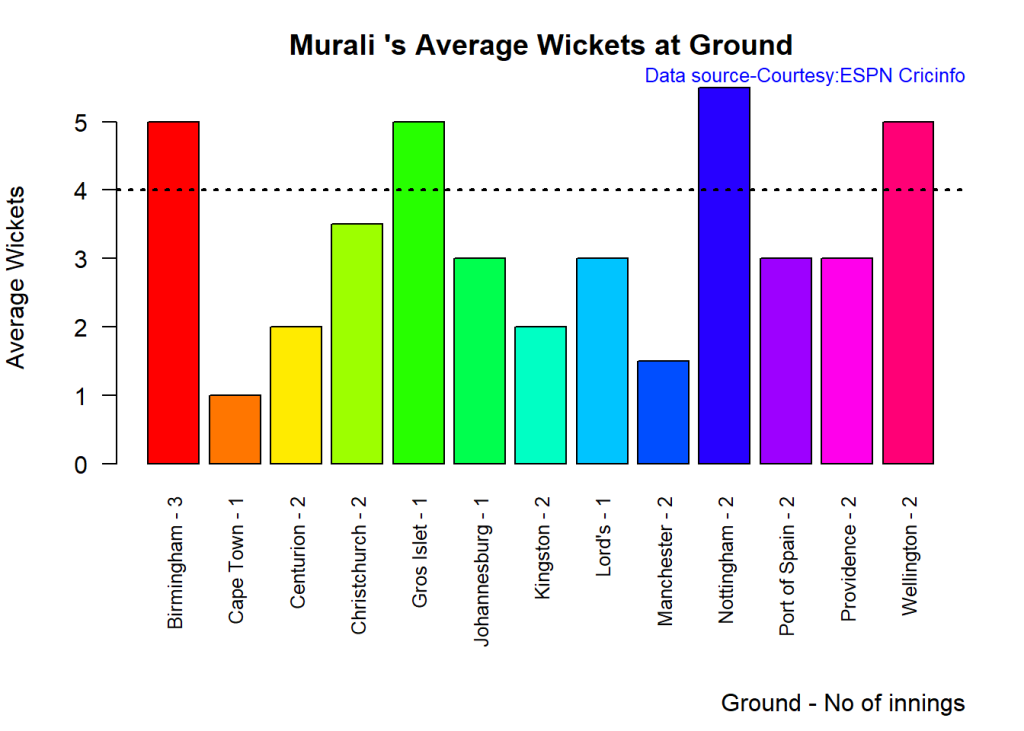
bowlerAvgWktsGround("kumbleTest-NZWISAEng.csv","Kumblew")



bowlerAvgWktsGround("warneTest-NZWISAEng.csv","Warne")



bowlerAvgWktsGround("muraliTest-NZWISAEng.csv","murali")



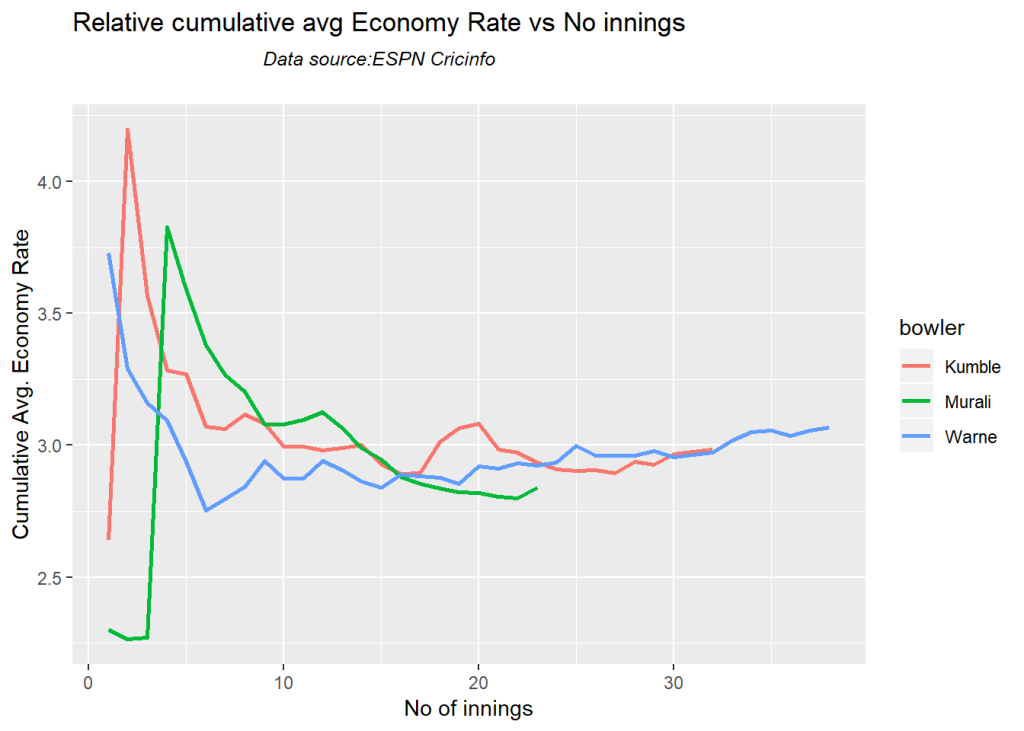
**6c. Plot the cumulative average wickets and cumulative economy rate of Kumble, Warne and Murali**

* Murali has the best economy rate followed by Kumble and then Warne
* Again Murali has the best cumulative average wickets followed by Warne and then Kumble

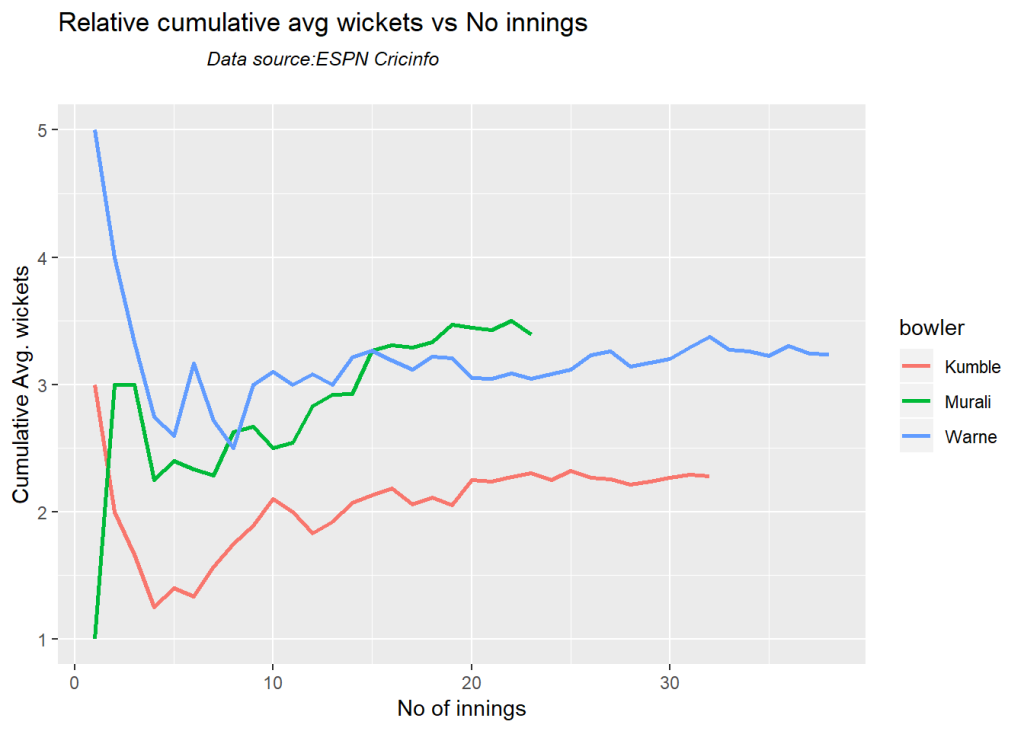
frames=list("kumbleTest-NZWISAEng.csv","warneTest-NZWISAEng.csv","muraliTest-NZWISAEng.csv")

names=list("Kumble","Warne","Murali")

relativeBowlerCumulativeAvgEconRate(frames,names)



relativeBowlerCumulativeAvgWickets(frames,names)



**7. Compute and plot the performances of Bumrah in 2016, 2017 and 2018 in ODIs**

# Get the HA data for Bumrah in ODI in bowling

df=getPlayerDataHA(625383,tfile="bumrahODIHA.csv",type="bowling",matchType="ODI")

## [1] "Working..."

# Slice the data for periods 2016, 2017 and 2018

df=getPlayerDataOppnHA(infile="bumrahODIHA.csv",outfile="bumrahODI2016.csv",

startDate="2016-01-01",endDate="2017-01-01")

df=getPlayerDataOppnHA(infile="bumrahODIHA.csv",outfile="bumrahODI2017.csv",

startDate="2017-01-01",endDate="2018-01-01")

df=getPlayerDataOppnHA(infile="bumrahODIHA.csv",outfile="bumrahODI2018.csv",

startDate="2018-01-01",endDate="2019-01-01")

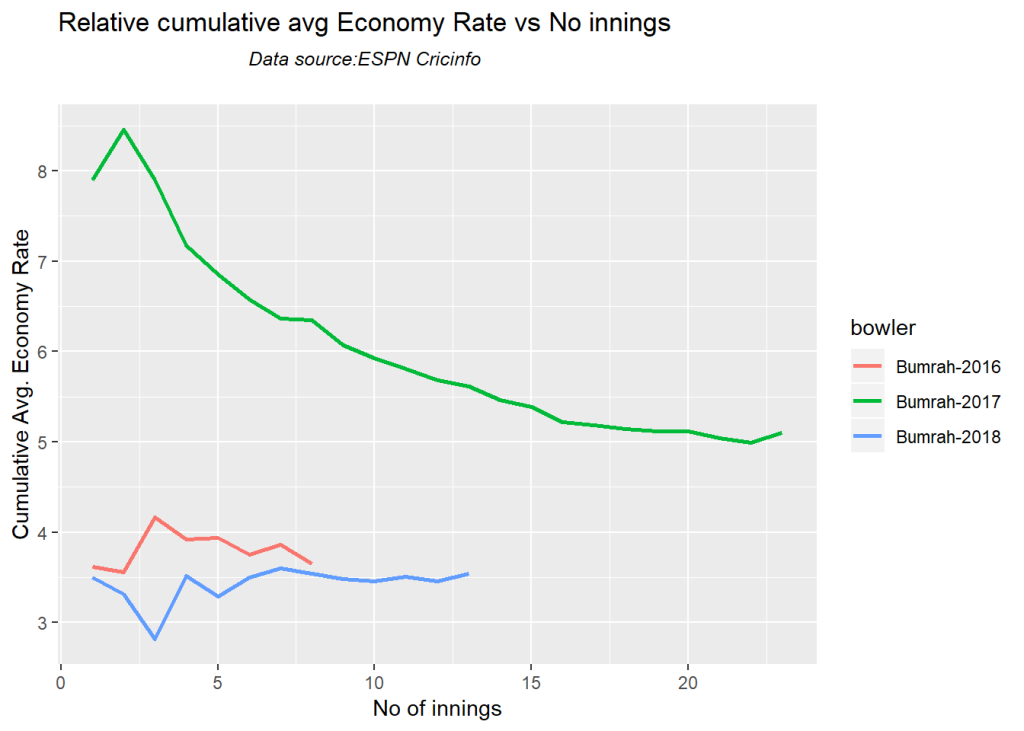
**7a. Compute the performances of Bumrah in 2016, 2017 and 2018**

* Very clearly Bumrah is getting better at his art. His economy rate in 2018 is the best!!!
* Bumrah has had a very prolific year in 2017. However all the years he seems to be quite effective

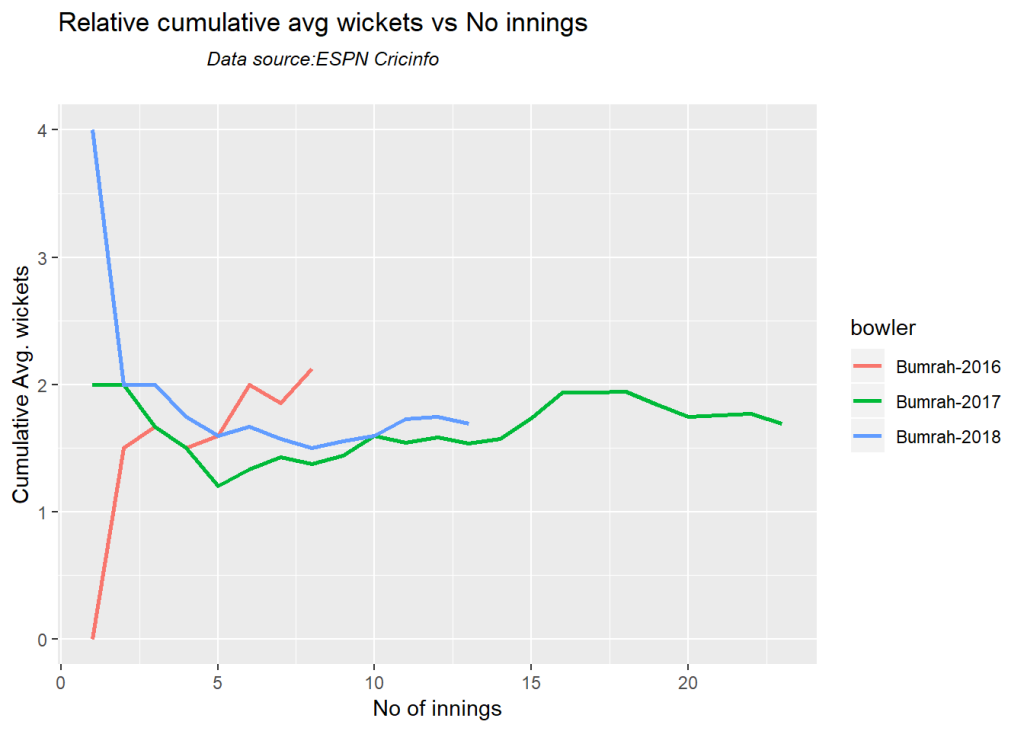
frames=list("bumrahODI2016.csv","bumrahODI2017.csv","bumrahODI2018.csv")

names=list("Bumrah-2016","Bumrah-2017","Bumrah-2018")

relativeBowlerCumulativeAvgEconRate(frames,names)



relativeBowlerCumulativeAvgWickets(frames,names)



**8. Compute and plot the performances of Shakib, Bumrah and Jadeja in T20 matches for bowling**

# Get the HA bowling data for Shakib, Bumrah and Jadeja

df=getPlayerDataHA(56143,tfile="shakibT20HA.csv",type="bowling",matchType="T20")

## [1] "Working..."

df=getPlayerDataHA(625383,tfile="bumrahT20HA.csv",type="bowling",matchType="T20")

## [1] "Working..."

df=getPlayerDataHA(234675,tfile="jadejaT20HA.csv",type="bowling",matchType="T20")

## [1] "Working..."

# Slice the data for performances against Sri Lanka, Australia, South Africa and England

df=getPlayerDataOppnHA(infile="shakibT20HA.csv",outfile="shakibT20-SLAusSAEng.csv"

,opposition=c("Sri Lanka","Australia", "South Africa","England"),

homeOrAway=c("all"))

df=getPlayerDataOppnHA(infile="bumrahT20HA.csv",outfile="bumrahT20-SLAusSAEng.csv"

,opposition=c("Sri Lanka","Australia", "South Africa","England"),

homeOrAway=c("all"))

df=getPlayerDataOppnHA(infile="jadejaT20HA.csv",outfile="jadejaT20-SLAusSAEng.csv"

,opposition=c("Sri Lanka","Australia", "South Africa","England"),

homeOrAway=c("all"))

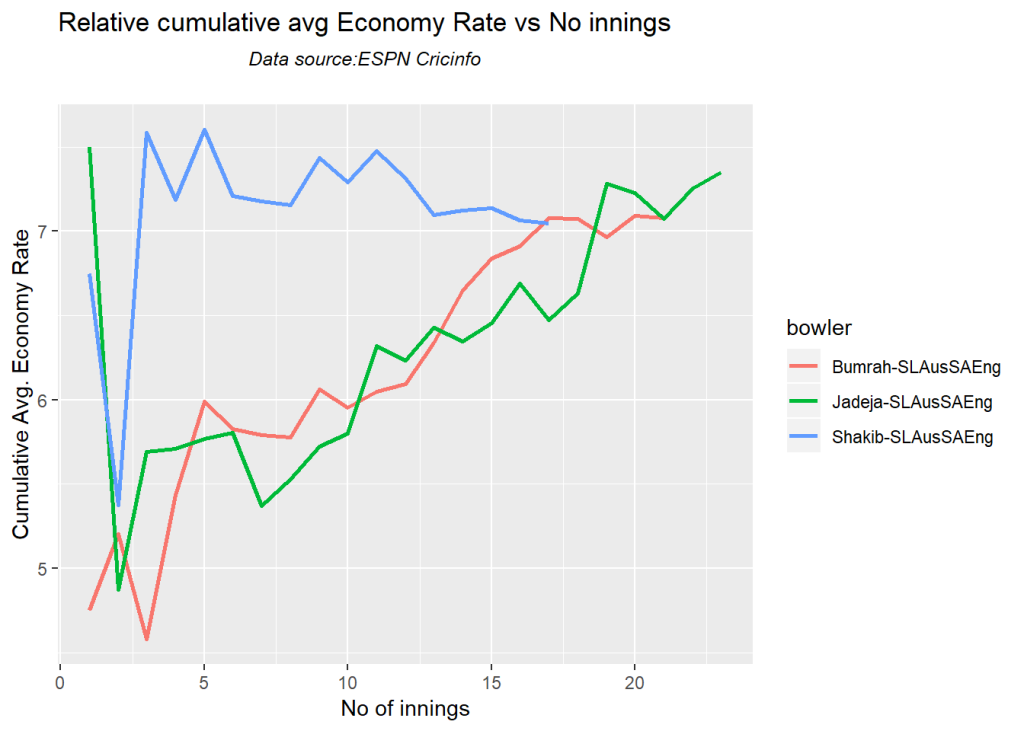
**8a. Compare the relative performances of Shakib, Bumrah and Jadeja**

* Jadeja and Bumrah have comparable economy rates. Shakib is more expensive
* Shakib pips Bumrah in number of cumulative wickets, though Bumrah is close behind

frames=list("shakibT20-SLAusSAEng.csv","bumrahT20-SLAusSAEng.csv","jadejaT20-SLAusSAEng.csv")

names=list("Shakib-SLAusSAEng","Bumrah-SLAusSAEng","Jadeja-SLAusSAEng")

relativeBowlerCumulativeAvgEconRate(frames,names)



relativeBowlerCumulativeAvgWickets(frames,names)

