Data Mining Lab Course
Cricket Group

Presentation of progress in week 3

Avinash Mishra, Matti Lorenzen AND Raveekiat Singhaphandu

Work in progress

- Parsing player profiles, crawl more information of the **player** from **ESPN Cricinfo** for further information such as, batting style, bowling style, batting and fielding averages, and bowling averages.
- Getting to know more about data mining tools
  - **Tableu(software)**
  - RapidMiner
  - R Language
  - **Microsoft Excel**
  - Waikato Environment for Knowledge Analysis (WEKA)

What we have done so far…(1)

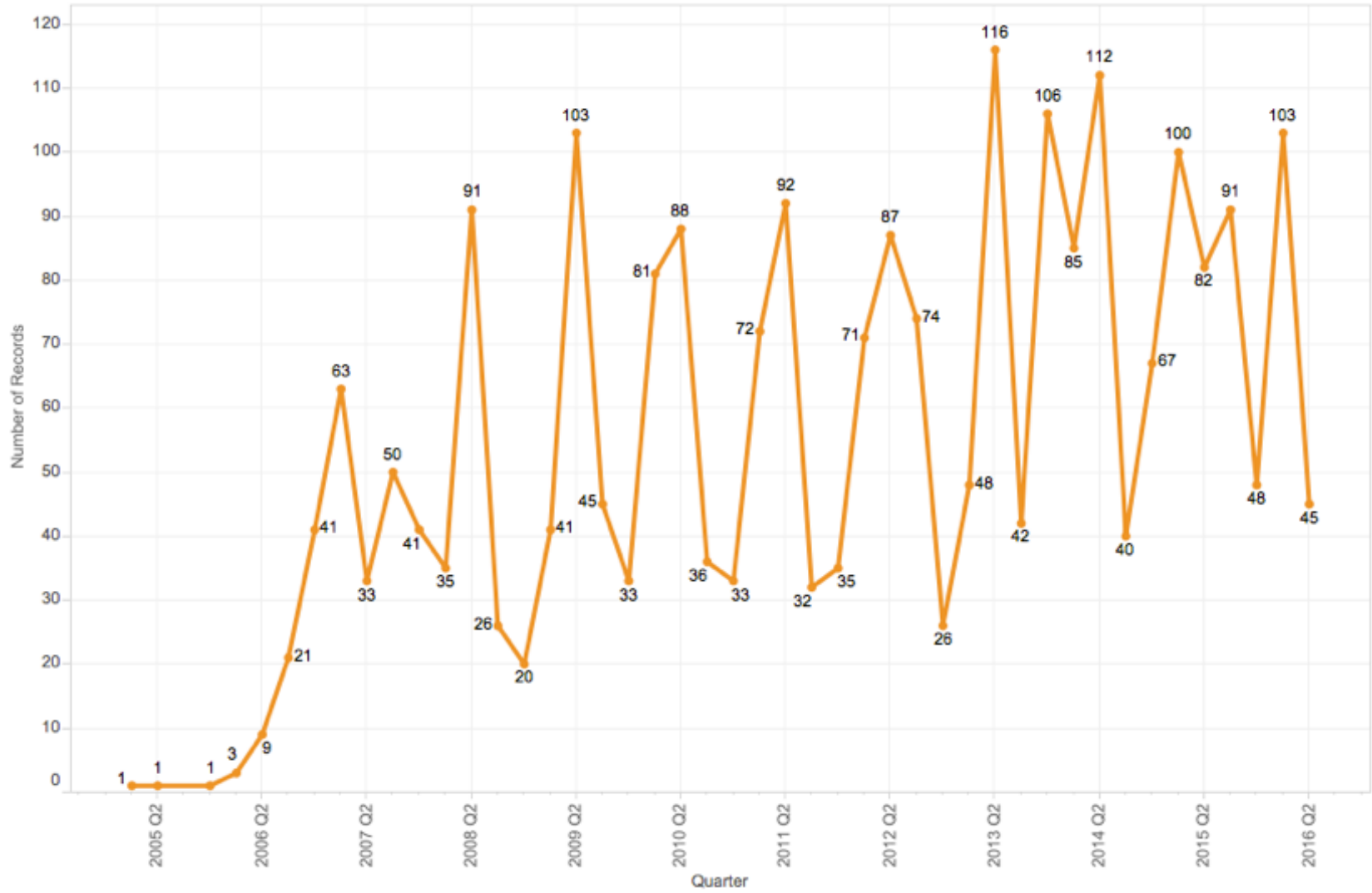
- Parsing all 2,470 matches including ball by ball data in to two csv files
  - Matches information and meta data
    - Such as dates, city, venue, match type, teams, winner, score, and etc.
  - Innings (ball-by-ball) information
    - 957,000 records of ball by ball information
    - Batsman, bowler, non striker, runs and wicket score

What we have done so far…(2)

- Using both Microsoft Excel and Tableau to visualize some data and information…
- First section: the information about the match
  - Date Occurrence
  - Venue and City
  - Match Type
  - Team appearance
  - **Histogram of wicket and run score**
  - **Number of match(es) win**
  - Toss Decision, match outcome of the team who win the toss, percentage of toss win

Figure 1

- Date occurrence
-



Skip

- Figure 2 to 5 and 9
- They are not that much informative
- If you are still interested, please take a look in our Wiki.

Figure 6

- Histogram of win by run scores
-

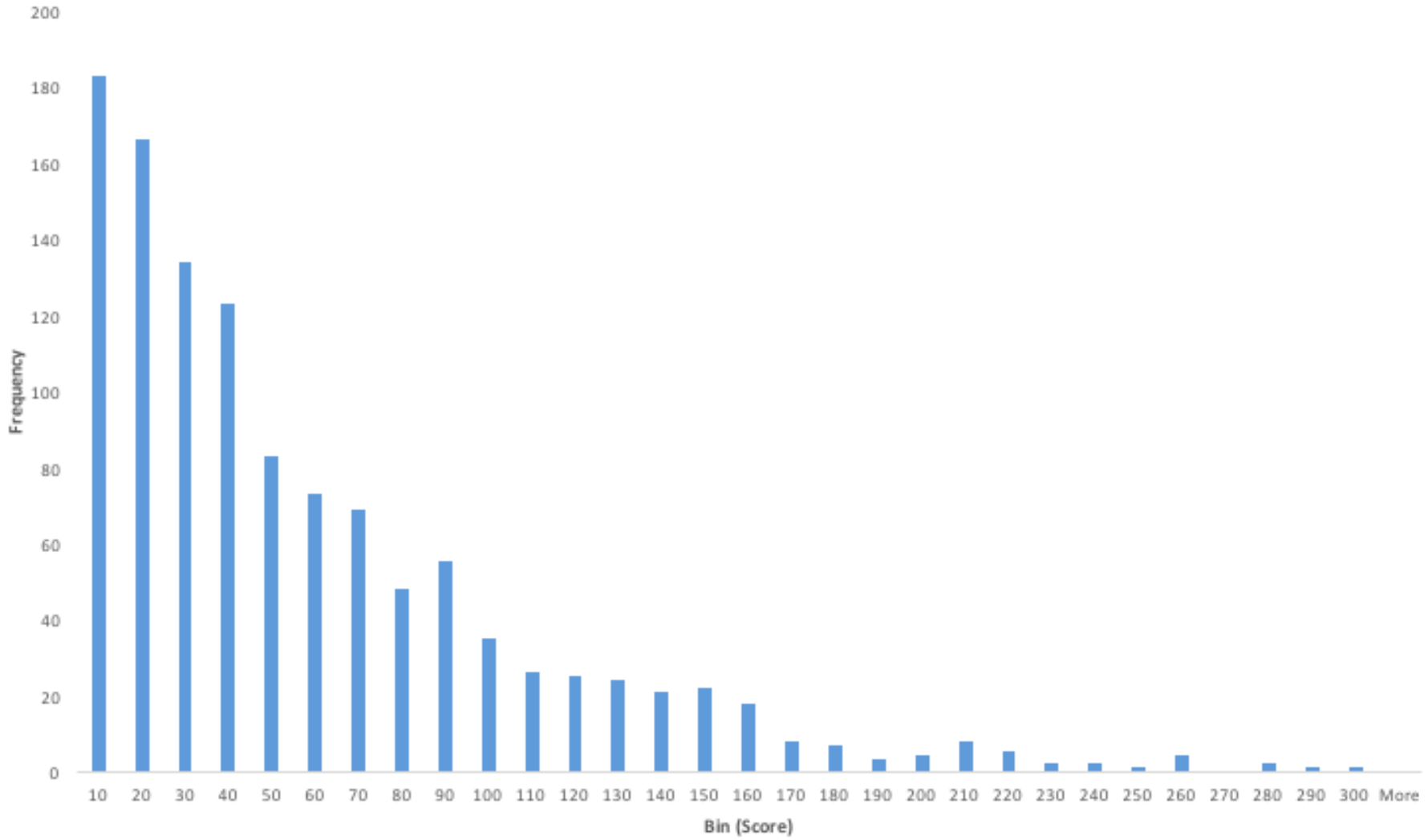


Figure 7

- Histogram of win by wicket scores
-

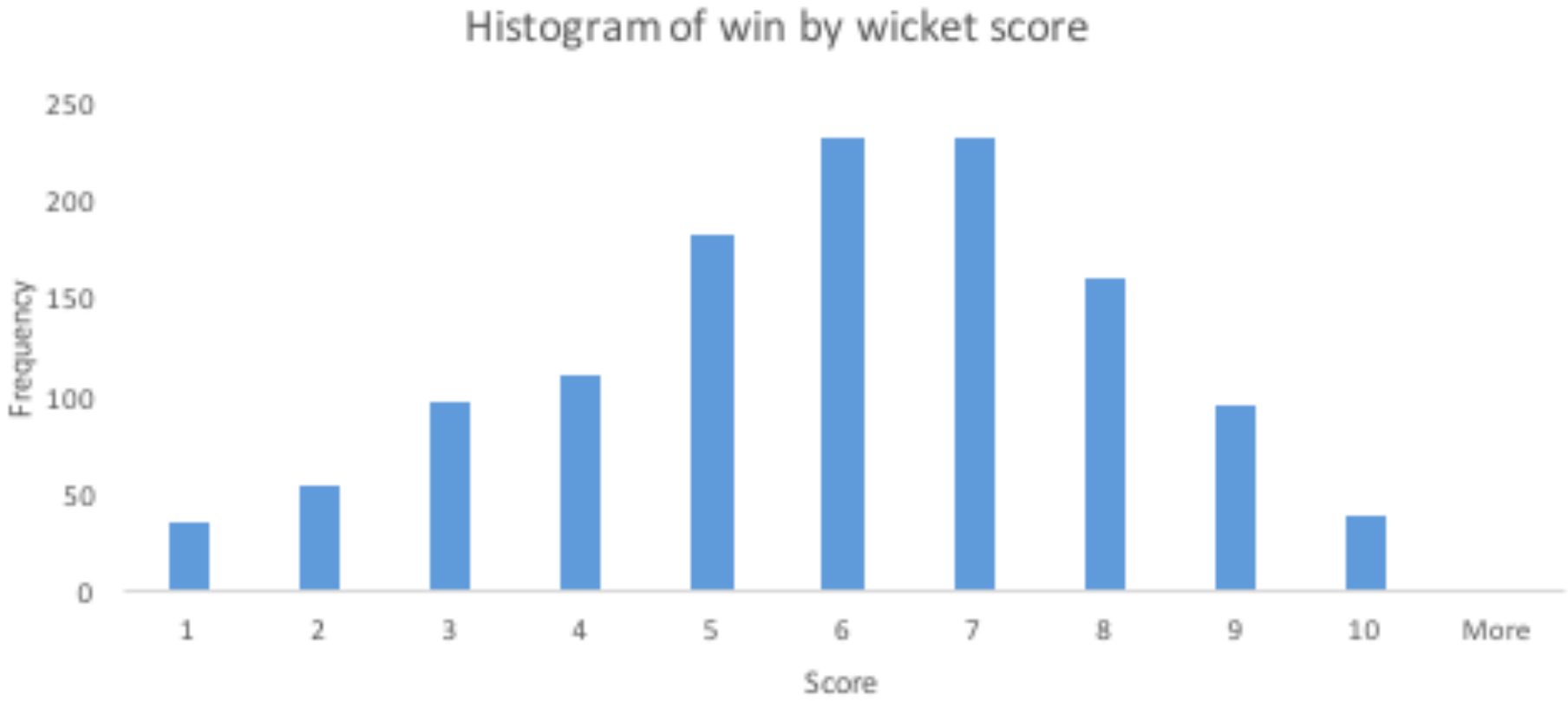
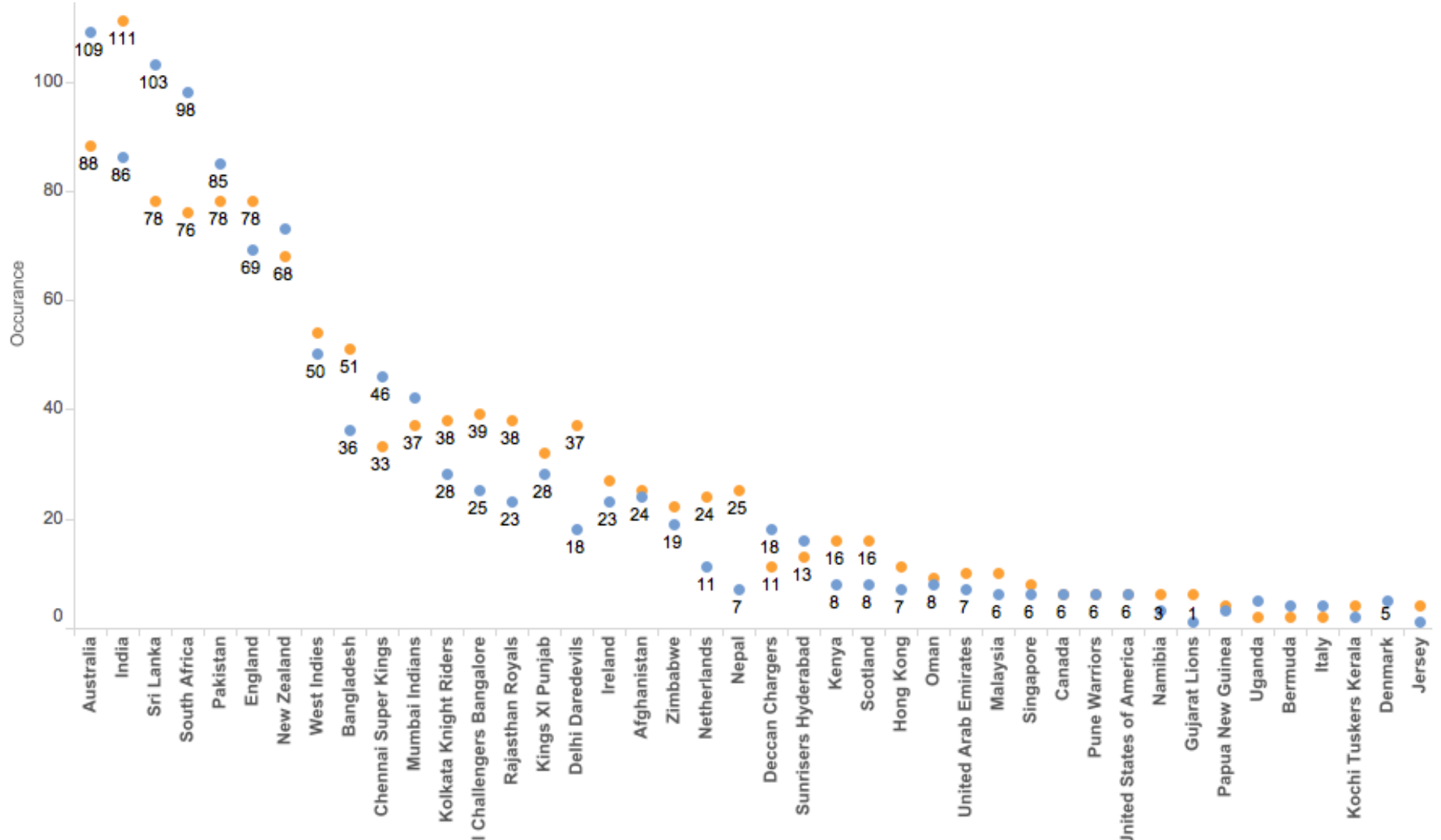


Figure 8

- Amount of win from each team (ORANGE: win by wickets, BLUE: win by runs)
-



Conclusion from figure 6-8b

- Now we know that…
  - They are two type of winning in the cricket matches
    - **By run** 1153 matches
    - **By wicket** 1128 matches
    - Only few of the matches they do not have any winner
  - **Each team has a different boldness (for example)**
    - Australia mostly win by runs
    - India mostly win by wickets
    - Most of the team win by wickets

What we have done so far…(3)

- Second section: by-ball-ball innings information
  - **Amounts of Face-offs between a specific bowler and batsman**
  - **Runs scored against a bowler by a specific batsman**
  - **Average amount of runs of a specific batsman against a specific bowler**

Figure 10

- Amounts of Face-offs between a specific bowler and batsman
-

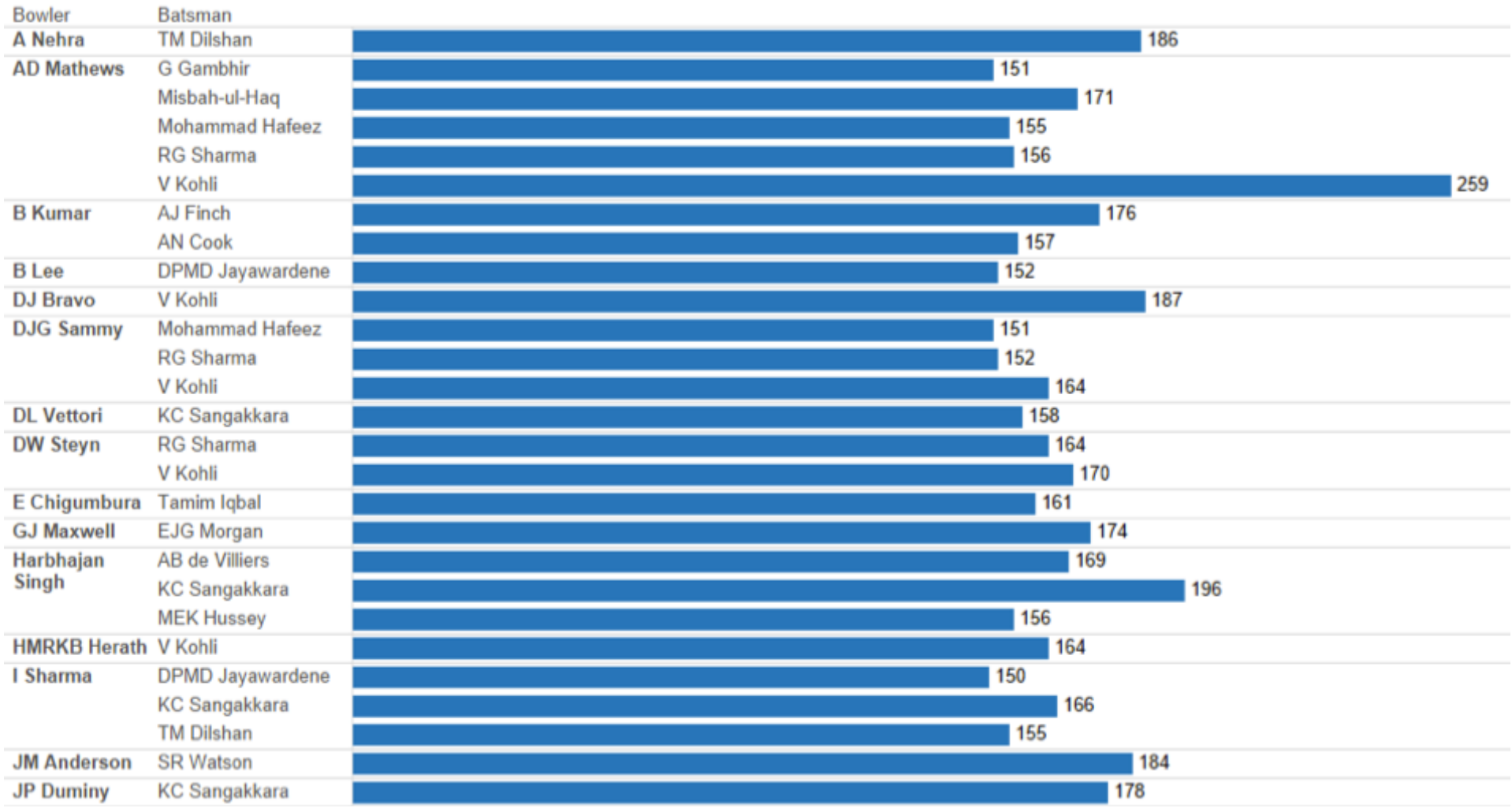


Figure 11

- Runs scored against a bowler by a specific batsman
-

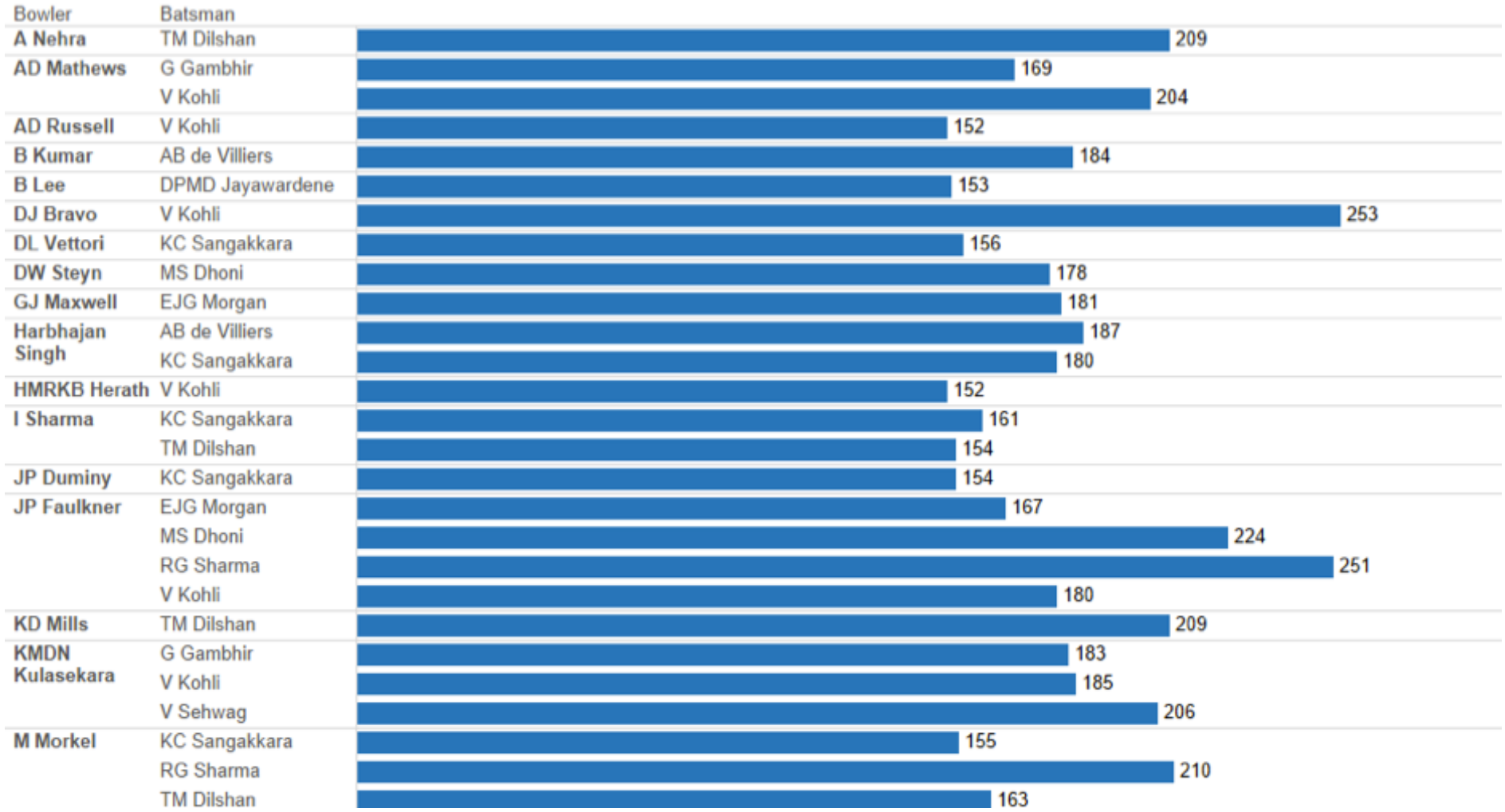


Figure 12

- Average amount of runs of a specific batsman against a specific b
-

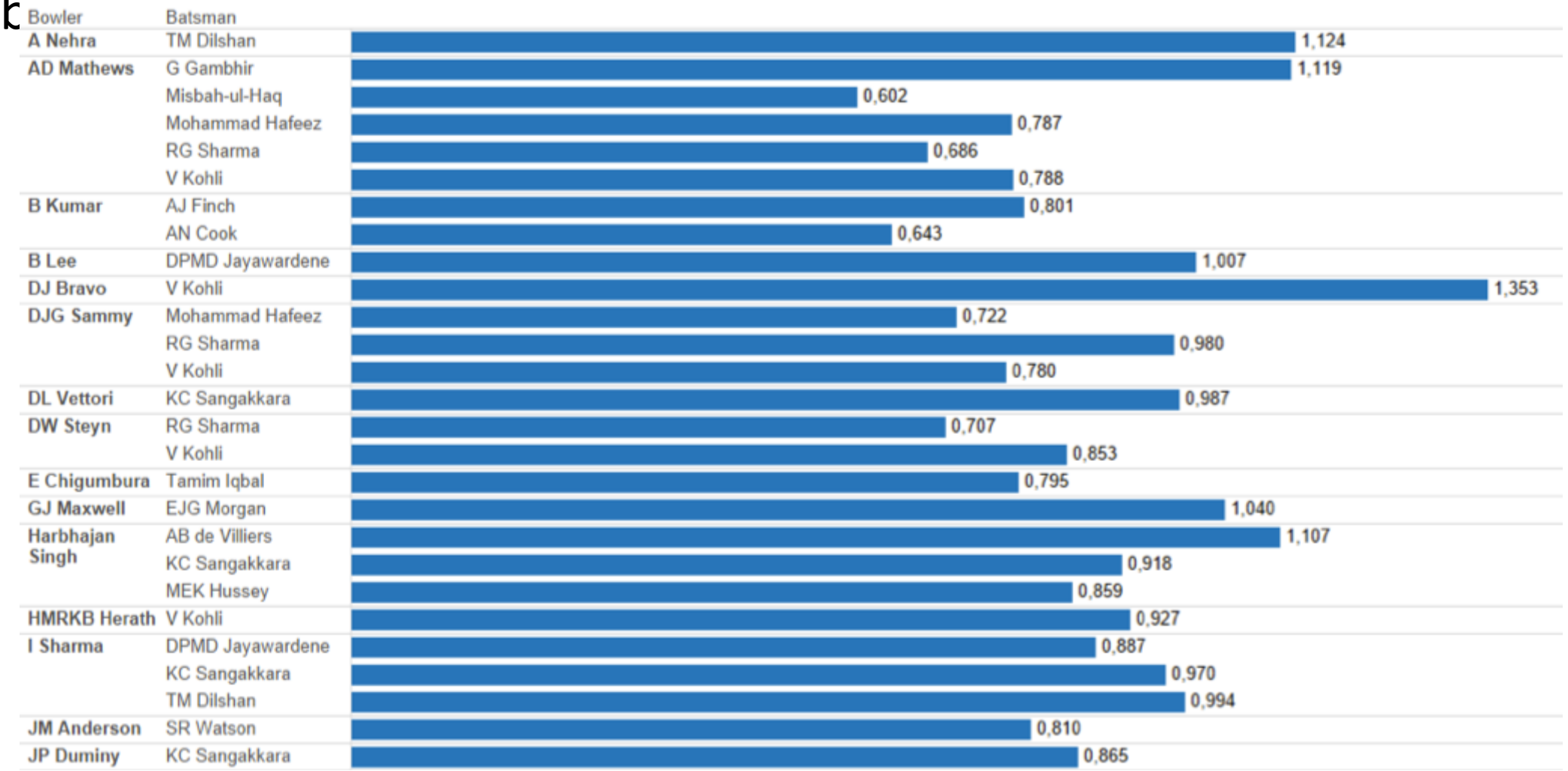
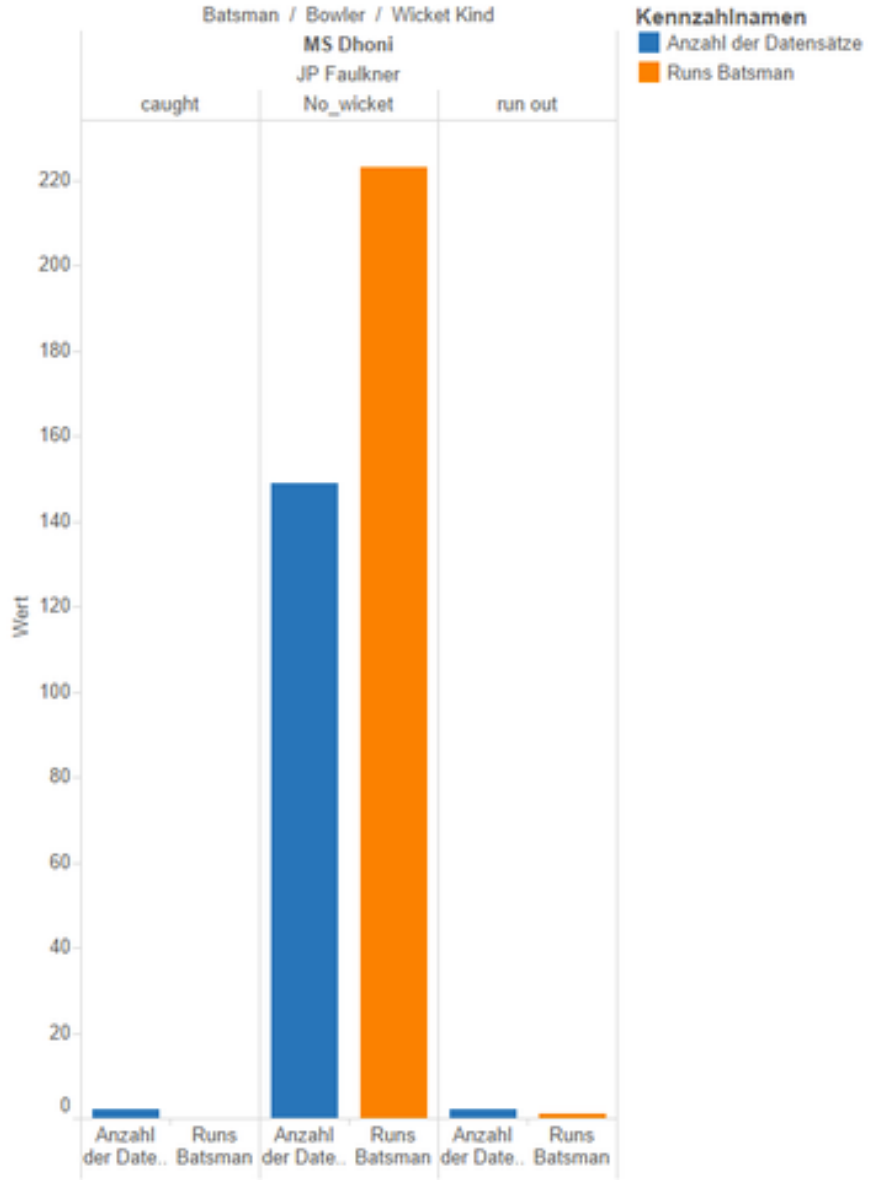
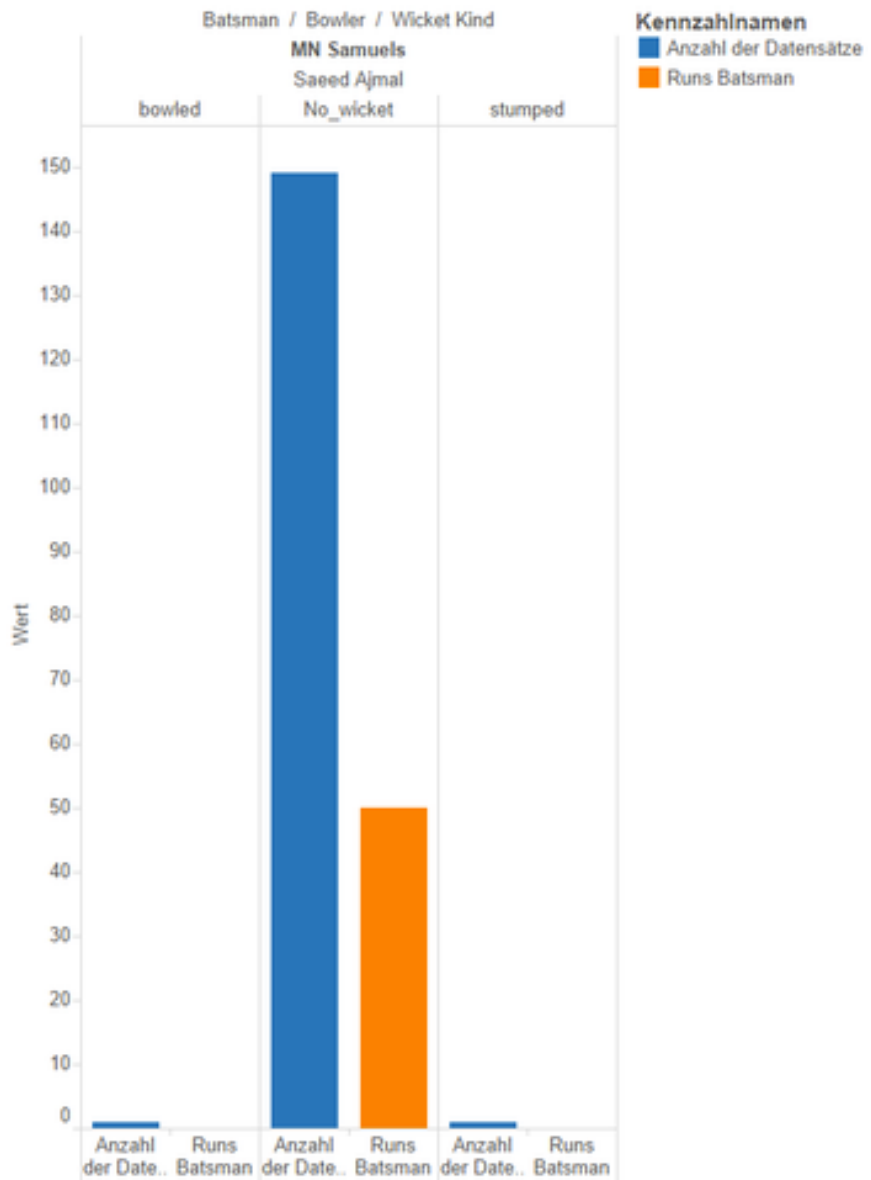


Figure 13 a&b


Batsman / Bowler / Wicket Kind
MS Dhoni
JP Faulkner
caught
No_wicket
run out
Kennzahlnamen
Anzahl der Datensätze
Runs Batsman
Wert
0
20
40
60
80
100
120
140
160
180
200
220
Anzahl der Date..
Runs Batsman
Anzahl der Date..
Runs Batsman
Anzahl der Date..
Runs Batsman


Batsman / Bowler / Wicket Kind
MN Samuels
Saeed Ajmal
bowled
No_wicket
stumped
Kennzahlnamen
Anzahl der Datensätze
Runs Batsman
Wert
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
Anzahl der Date..
Runs Batsman
Anzahl der Date..
Runs Batsman
Anzahl der Date..
Runs Batsman

Conclusion from figure 10-13b

- What we can…
  - Evaluate and visualize specific face-offs between a bowler and a batsman
  - Make statements about the performance of the players as we can sum up all runs ever scored by a batsman against each bowler
  - Compute an average of runs against each opponent
  - Have a closer look at the player pairings that have the highest and the lowest run average.
  - In fact, the pairing between MS Dhoni and JP Faulkner has been historical as there has been a new world record set up [link]
  - **Evaluations like this can be easily created for any pairing. The information we gather cannot yet be found in popular cricket statistic web sites.**

Feedback for Team 3 :Food Inspection Dataset

- Clear and detailed description of work.. Good Work, keep it up.

- **Year wise inspection Result**
  Normalized data for 2016, would have been good.

- **Brand wise Inspection Result Analysis:**
  Interesting Results bit little bit confused about the statement “Dunkin Donuts has
  more outlets but they has fewer failed inspections than Mc Donalds”
  because I can see from given figure that Dunkin Donuts has less no of inspections.

- **Visualizations of "Inspection Result" history based on location :**
  really love the idea to show all details based on location and make it more interactive. Love to know more description about how you did it.