

# Title: Peek into Indian Premier League

<https://github.com/bkommineni/DataVisualization-FinalProject>

Page to host the project: <https://bkommineni.github.io/DataVisualization-FinalProject/>

---

By: Anjani Bajaj ([aabajaj2@dons.usfca.edu](mailto:aabajaj2@dons.usfca.edu))

Bhargavi Kommineni ([bkommineni@dons.usfca.edu](mailto:bkommineni@dons.usfca.edu))

## 1. Background and Motivation

Cricket is an extremely popular sport in India, however it is under represented in the world. We come from India, where we grew up watching cricket. The Indian Premier League (IPL) is a professional Twenty20 cricket league in India contested during April and May of every year by teams representing Indian cities and some states. IPL started in 2008 and includes cricket players from all over the world. It is also the most-attended cricket league in the world and in 2014 ranked sixth by average attendance among all sports leagues.

Hence, we decided to explore and visualize IPL data from kaggle in this project. We plan to visualize the data effectively and hope to find some great insights throughout the process.

### Cricket Basics

- **Bat** : the wooden implement with which the batsman attempts to strike the ball
- **Ball** : the round object which the batsman attempts to strike with the bat. Also a delivery.
- **Batting** : the act and skill of defending one's wicket and scoring runs
- **Bowling** : the act of delivering the cricket ball to the batsman.
- **Bowling Average** : A bowler's bowling average is defined as the total number of runs conceded by the bowler divided by the number of wickets taken by the bowler.
- **Batting Average** : A batsman's batting average is defined as the total number of runs scored by the batsman divided by the number of times he has been dismissed
- **Catch** : To dismiss a batsman by a fielder catching the ball after the batsman has hit it with his bat but before it hits the ground
- **Caught** : Its a method of dismissing a batsman in the sport of cricket. Being caught out is the most common method of dismissal at higher levels of competition. A batsman is out caught if a fielder catches the ball fully within the field of play without it bouncing once the ball has touched the striker's bat, glove or only the leg of the batsman.
- **Caught and bowled** : When a player is dismissed by a catch taken by the bowler.
- **Caught behind** : A catch by the wicket-keeper.
- **Four** : A shot that reaches the boundary after touching the ground, which scores four runs to the batting side.
- **Six** : a shot which passes over or touches the boundary without having bounced or rolled, so called because it scores six runs to the batting side
- **Hit wicket** : A batsman getting out by dislodging the bails of the wicket behind him either with his bat or body as he tries to play the ball or set off for a run.

- **Inning** : One player's or one team's turn to bat (or bowl). Unlike in baseball, the cricket term "innings" is both singular and plural.
- **LBW** : Leg Before Wicket, It means when there is leg in front of wicket then the batsman is given out, but bowler have to make an appeal to umpire.
- **Leg bye** : Extras taken after a delivery hits any part of the body of the batsman other than the bat or the gloved hand that holds the bat. If the batsman makes no attempt to play the ball with the bat or evade the ball that hits him, leg byes may not be scored
- **Match referee** : an official whose role is to ensure that the spirit of the game is upheld.
- **No ball** : An illegal delivery; the batting side is awarded one extra, the bowler must deliver another ball in the over, and the batsman cannot be dismissed by the bowler on a no-ball. Most usually a front-foot no ball, in which the bowler oversteps the popping crease; other reasons include bowling a full toss above waist height (see beamer), throwing, having more than two fielders (excluding the wicketkeeper) behind square on the leg side, or breaking the return crease in the delivery stride
- **Non-striker** : The batsman standing at the bowling end
- **Obstructing the field** : It is one of the nine methods of dismissing a batsman in the sport of cricket. It dictates that either batsman can be given out if he wilfully attempts to obstruct or distract the fielding side by word or action.
- **Over** : The delivery of six consecutive legal balls by one bowler.
- **Retired out** : For a batsman to voluntarily leave the field during his innings, usually because of injury. A player who retired through injury/illness may return in the same innings at the fall of a wicket, and continue where he left off. A player who is uninjured may return only with the opposing captain's consent
- **Run out** : Dismissal by a member of the fielding side breaking the wicket while the batsman is outside his/her crease in the process of making a run.
- **Stump**
  - one of the three vertical posts making up the wicket ("off stump", "middle stump" and "leg stump");[2]
  - a way of dismissing a batsman in which the wicketkeeper breaks the batsman's wicket with the ball when the batsman is outside his crease but has not attempted a run; or
  - In a match lasting more than one day, "stumps" refers to the end of a day's play when the match is not complete (e.g. a progress score after the first day may be described as the score "at stumps on Day 1"). See also draw stumps

- **Toss** : In the sport of cricket, a coin is tossed to determine which team bats first. This is known as the toss.
- **Umpire** : one of the two (or three) enforcers of the laws[46] and adjudicators of play.
- **Wicket** :
  - a set of stumps and bails;
  - the pitch; or
  - the dismissal of a batsman.
- **Wide** : A delivery that passes illegally wide of the wicket, scoring an extra for the batting side. A wide does not count as one of the six valid deliveries that must be made in each over – an extra ball must be bowled for each wide
- **Duckworth-Lewis(DL)** : A mathematically based rule that derives a target score for the side batting second in a rain-affected one-day match
- **Extra** : A run not attributed to any batsman;
  - there are five types:
    - byes,
    - leg byes,
    - penalties,
    - wides and
    - no-balls.
  - The first three types are called 'fielding' extras (i.e. the fielders are determined to be at fault for their being conceded)
  - The last two are called 'bowling' extras (the bowler being considered to be at fault for their being conceded) which are included in the runs conceded by the bowler.

## Match Format

Twenty20 match format is a form of limited overs cricket in that it involves two teams, each with a single innings, the key feature being that each team bats for a maximum of 20 overs. In terms of visual format, the batting team members do not arrive from and depart to traditional dressing rooms, but come and go from a bench (typically a row of chairs) visible in the playing arena, analogous to association football technical area or a baseball dugout. <sup>[1]</sup>

## 2. Project Objectives

The goal of this project is to efficiently visualize the cricket data i.e. data about performance of the players and IPL matches for over 10 years (2008-2017) and identify if there exists some trends.

After the successful completion of this project we hope to:

- Visualize maps to analyze the location of the players, teams and matches
- Visually compare the performance of teams over the years
  - Logistics for performance -
    - Number of matches played
    - Wickets Taken
    - Number of runs scored, etc.
- Analysis of player performance across different teams and across different years, to find out correlations of their performance with team or their experience as a player in IPL
  - Visually compare the batting average, wickets taken and other attributes for different players over the years

### 3. Data

Source of the data is Kaggle - <https://www.kaggle.com/manasgarg/ipl>

The dataset contains 2 files: deliveries.csv and matches.csv.

- matches.csv contains details related to the match such as location, contesting teams, umpires, results, etc.
- deliveries.csv is the ball-by-ball data of all the IPL matches including data of the batting team, batsman, bowler, non-striker, runs scored, etc.

### 4. Data Processing

We plan to use python for data processing. The main task would be to generate the data of the players for each team for all the years. As we do not have all the details related to players grouped according to a team. We plan to derive it and use them for generate specific visualizations.

### 5. Visualization Design

All the visualizations will be implemented in d3.js with interactions. We brainstormed a bunch of different visualizations and finalized these-

- **BarChart:**
  - Filter: Year
  - X-axis: all the teams
  - Y-axis: Number of matches won
  - Drop down: Total number of runs scored and number of matches played
- **Stacked Bar Chart:**
  - Filter: Year
  - X-axis: All the teams
  - Y-axis: Number of dismissals

Type: stacked bar chart with different dismissal types legend with colors

- **ScatterPlot:**

Player level analysis:

Filter: Batsman

Each Datapoint :

X-axis: Total no of runs scored (in each match)

Y-axis: Total no of balls faced (in each match)

More interactions can show which Year and Match details

- **Line Trends:**

1. Player performance over the years

Filter: Player

X-axis: Year

Y-axis: Number of runs scored

2. Year-wise trends for all the teams

Filter: Team

X-axis: Year

Y-axis: Number of matches won

- **GeoMaps:**

1. Starting with India's map we can plot states corresponding to the team which will be colored by the team color.

Filter: Year

2. Plot a world map with the location of all the players

## Optional Features:

- **Time-Series Plots using brushing:**

- We can have the user select a time-frame for the plots instead of a filter for the line trends or scatter plots in the player level analysis.
- Example: The user can select like 2 or 3 years time frame to analyze the performance of the player.

<http://mlvl.github.io/timeseries/>

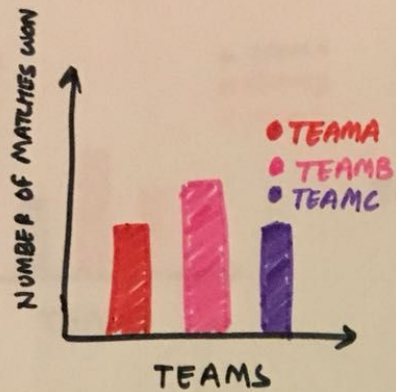
- **Chord Diagram:**

- Mouseover to focus on a team
- The tool tip can give you the overall matches for a team
- The thickness of links between countries can encode the relative frequency of matches between two teams: Thicker links represent more matches which result in more wins.
- Links will be directed: So, mouse over on a link/chord to see wins between the linked countries. Links will be colored by the team with more wins.

[Example chord diagram](#)

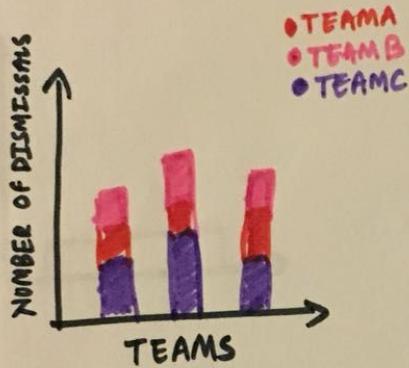
[Example cricket chord diagram](#)

# SHEET 1 : Brainstorm A Designs

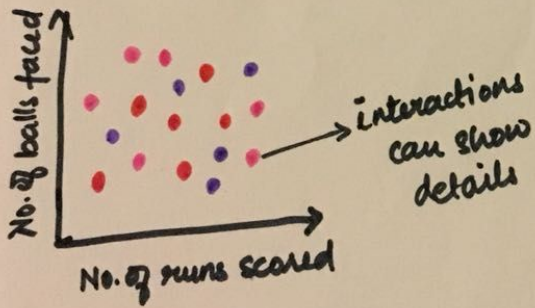


YEAR ← FILTER

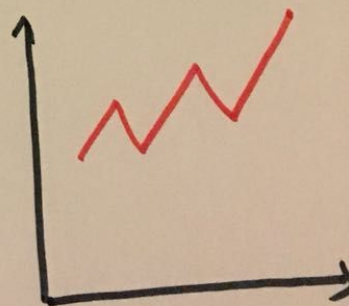
BAR CHART for  
Year v/s teams  
Number  
of matches  
won



STACKED BAR CHART  
Number of dismissals  
v/s  
teams

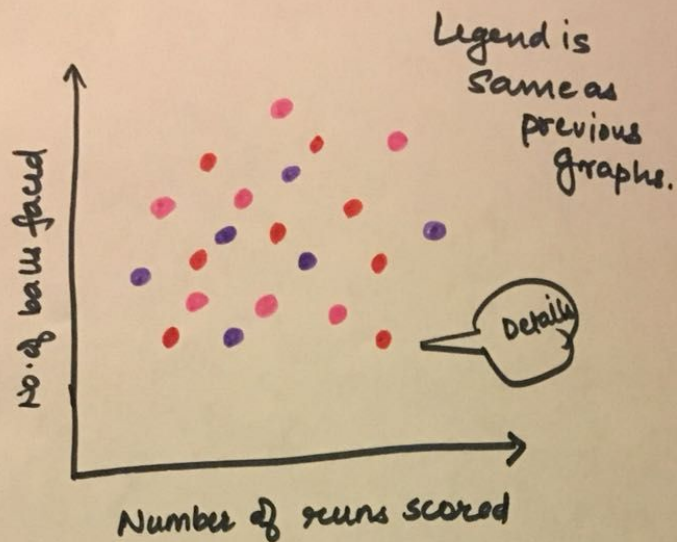


Some trends ?

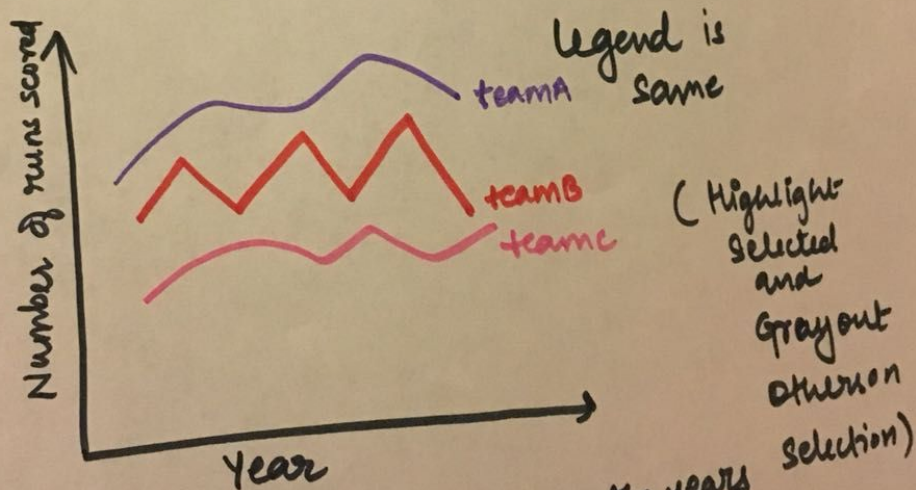




SHEET 3



TITLE: Scatter Plot for performance



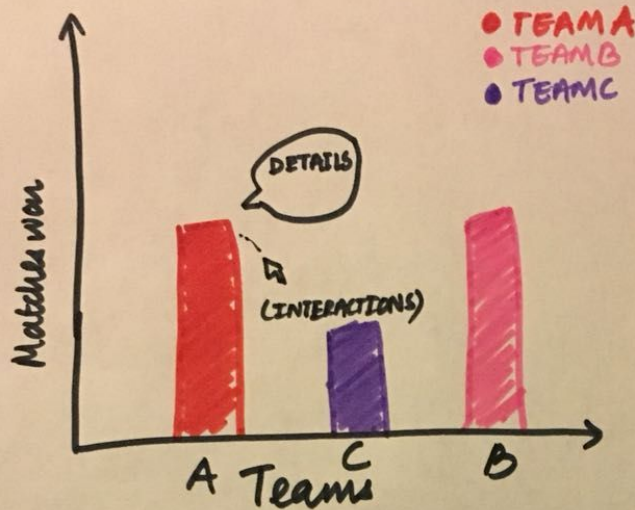
TITLE: Player performance over the years

SHEET 2 :

TITLE: Match comparison

Filter by Year

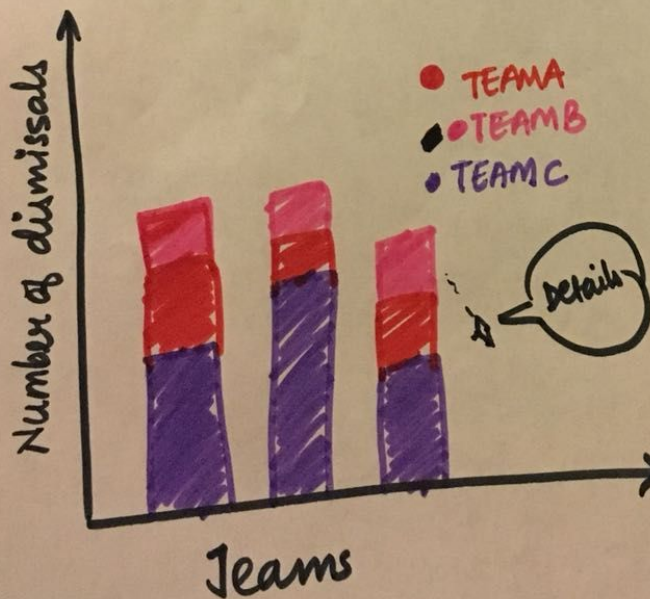
(A different bar chart for each other!) year



TITLE: Dismissal comparison

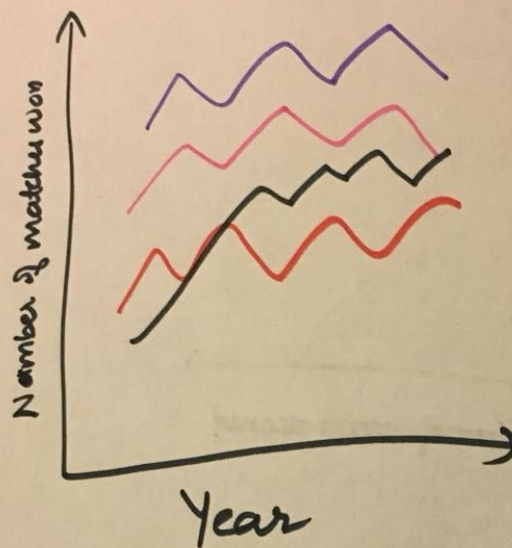
Filter by Year.

(A different stacked chart for each year)





Year wise trends for all teams



Legend is the same

(Filter can be team too)

(A dropdown maybe?)

Detail:

Data processing: Python

Visualizations: P3.js

## 6. Project Schedule

Task	Date
Project Proposal	4/6/18
Revised Proposal, Related Work, and Project Website	4/11/18
Alpha Release	4/16/18
Beta Release	4/25/18
Project Presentation	5/7/18 and 5/9/18
Presentation slides, Code, Data, Process Book, and README	5/16/18

## 7. References & Related Work

1. [Project on IPL data visualization using python packages](#)
2. [Line trends for players example](#)
3. [Example cricket chord diagram](#)
4. [Example on cricket visualizations](#)
5. <https://knoema.com/insights/cricket>