**FIT 5147**

**Assignment 2**

**Data Exploration and Visualisation Project Proposal**

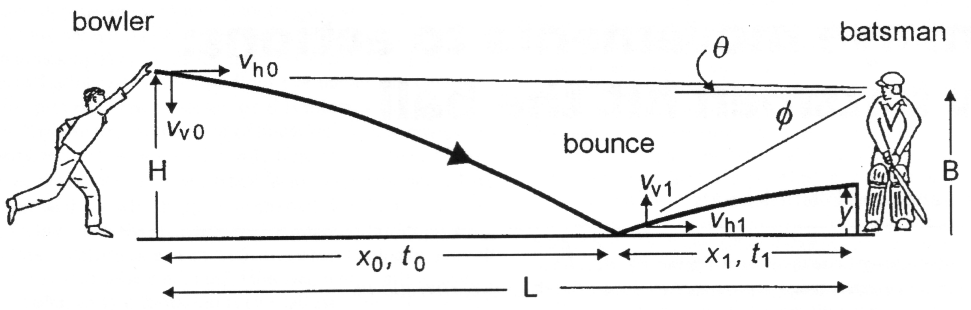
Linus McManamey

**Introduction**

This is a proposal for a web based Shiny application written in the R programming language. The application will attempt to display information from a dataset that recorded the XYZ coordinates of a cricket ball at three different positions on a cricket pitch, the positions are:

* The XYZ coordinates of the cricket ball when it was released by the bowler
* The XYZ coordinates of the cricket ball when it struck the ground
* The XYZ coordinates of the cricket ball when it arrived at the batsman's stumps

The other important data to be used by this application is the speed of the ball at each of the three location mentioned above and the angle at which the ball left the bowler's hand and the angle of the ball as it it bounced of the surface of the cricket pitch.



Using the data mentioned above this application will attempt to define physical attributes of a cricket pitch based on the differing trajectories and movement of the cricket ball.

**Problem description and motivation**

**The Problem**

The overall problem space defined in this project is to establish the display of certain areas of a cricket pitch that generate different results when it is struck by a cricket ball.

The display of this information will need to take into account the speed and angle of the ball when it was released by the bowler, the speed and angle of the ball as it left the pitch and the speed of the ball as it arrives at the far end of the cricket pitch. There has also been two recorded variables “Swing” and “Deviation” which track the ball's movement through the air. The **motivation** for displaying this data as interactive visual information would be to discover certain characteristic of a cricket pitch. The most useful characteristic would be :

* Areas on the cricket pitch that created high bounce
* Areas on the cricket pitch that created low bounce
* Areas on the cricket pitch that affected ball swing in the air
* Where ball speed was affected either in a positive or negative manner
* An overall effect on the above variables by time of day

**The Data Source Used and Data Cleaning** **Data Transformations**

The data source is a csv file of approximately 1750 records. The major data transformation task at this stage of the project was separating the data into two ends. The initial data seta has all the balls being bowled from one end. For the application to be of any use the correct bowling ends need to be established.

The data wrangling process, the data cleaning and data transformation process is attached in a python Notebook - file.ipynd. The ipython notebook will export a csv file to be then utilised in the Shiny application.

**Some Preliminary Data Exploration**

Based on the book “Physics of Cricket” by Mark Kidger certain values will be used to calculate the different reactions the the ball when it strikes the cricket pitch.

“The initial velocity of the ball is 160km/h, but **air resistance slows it by 12%...**” (Kidger, 2011). The global variable PRE\_BOUNCE\_DRAG = 0.88 will be used to calculate this physical effect of air drag on the ball.

“On bouncing, the ball loses kinetic energy and thus its speed will drop.  **It will rebound with about 55-60%** of its original speed. Air resistance will slow the ball by a further 7% before it reaches the batsman” (Kidger, 2011) The global variables REBOUND\_SPEED\_COEFF = 0.45 and AFTER\_BOUNCE\_DRAG = 0.93 will be be used to calculate the above mentioned effects on the ball.

“A **cricket ball** bounces to about one third of that height (0.67 m), in which case it rebounds at a speed of 3.61 m/s. The ratio of these two speeds is 3.61/6.26 = 0.58 and is called the **coefficient of restitution** (COR)”. (Physics of Cricket, 2005) The global variable C\_O\_R = 0.42 Will be used to calculate the height of the ball should bounce back to.

The model to demonstrate different reactions the cricket ball has on the cricket pitch will use the above mentioned coefficients on each ball. The projected outcome versus the real out come will be compared allowing the user to decided if there are any patterns that may useful or give insight into the makeup of the cricket pitch.

**Description of what message you want to convey in the project and what**

**the intended audience is.**

The applications proposed specification has only allowed for a static data source, however a dynamic data source implementation would allow the intended audience for this application - broadcast media the opportunity to use the application in a historic (previous day) or real time (previous ball) tool whilst live broadcasting.

The coaching staff of a cricket team would also be able to utlise this application for analysis of bowler and cricket pitch. It could also be used by the curators of a particular ground in identifying area of the pitch that need special attention to maintain a good playing surface.

**Where are you up to in the project**

This is as far as i have progressed with this project. From this point it is predominantly about developing the Shiny interface and the use of the plotly graphics package. I would also like to investigate the use of PCA analysis or a multivariate regression model to test the accuracy of the application.

**Youtube Video Presentation**

[**https://youtu.be/2oo0gf9qdQI**](https://youtu.be/2oo0gf9qdQI)

**References:**

Kidger, Mark. Physics of Cricket, edited by Mark Kidger, Nottingham University Press, 2011. ProQuest Ebook Central.

<http://www.physics.usyd.edu.au/~cross/cricket.html>.Physics of Cricket