

The 2015 Cricket World Cup Database

David Ebert

Tarleton State University

29 November 2016

GitHub: [dpebert7/2015cwc](https://github.com/dpebert7/2015cwc)



Introduction

Cricket Basics

Overview

Data
Collection

Database

Extensions

- Two teams of 11 players
- Toss determines who *bats* and *fields* first
- The fielding team *bowls* 50 *overs* of 6 *balls* each
- Two *batsmen* at a time from the batting team score as many runs as possible without losing their *wicket*
- After the batting team loses 10 wickets fall or 50 overs are completed, the teams switch and the fielding team *chases* the batting team's score
- After the second team has batted, the team with the highest score wins

Cricket Basics

Overview

Data
Collection

Database

Extensions

- Two teams of 11 players
- Toss determines who *bats* and *fields* first
- The fielding team *bowls* 50 *overs* of 6 *balls* each
- Two *batsmen* at a time from the batting team score as many runs as possible without losing their *wicket*
- After the batting team loses 10 wickets fall or 50 overs are completed, the teams switch and the fielding team *chases* the batting team's score
- After the second team has batted, the team with the highest score wins

E.g. "South Africa 227/7 (50 overs) defeated Pakistan 209/10 (48.5 overs) by 17 runs"

Tournament

Overview

Data
Collection

Database

Extensions

- February-March 2015 Cricket World Cup
- Hosted in Australia and New Zealand
- Tournament Structure
 - 2 Pools of 7 teams
 - Stage 1: Round-robin play within pools
 - Stage 2: Seeded 8-team single-elimination playoffs
- 49 matches

ESPNcricinfo

LIVE SCORES

Like 443

Tweet

ICC Cricket World Cup, 11th Match, Pool A: Australia v Bangladesh at Brisbane, Feb 21, 2015

[Scorecard](#) | [Commentary](#) | [Almanack report](#) | [Report](#) | [Articles \(7\)](#) | [Photos \(8\)](#) | [Videos \(2\)](#)

ICC Cricket World Cup - 11th match, Pool A

Australia v Bangladesh

Match abandoned without a ball bowled

ODI no. 3608a | 2014/15 season

Played at Brisbane Cricket Ground,
Woolloongabba, Brisbane

21 February 2015 - day/night match (50-
over match)



Andrew Ramsey ✓
@ARamseyCricket

Follow

Sad scene at Gabba turnstiles where water not patrons
flooding in. Rain heavier now than for past hour #CWC15

11:18 PM - 20 Feb 2015

Retweets 275 Likes 162



Alt Cricket
@AltCricket

[Follow](#)

AUSTRALIA WEATHER UPDATE: Cloudy with a chance of Sharknado.

12:35 AM - 21 Feb 2015

🔄 69 ❤️ 49

••••• Vodafone AU 4G 11:32 am 68% 🔋



underwater cricke

3



Scott Craig
@Scott_CraigEnig

[Follow](#)

The @CricketAus boys are getting some practice in at The Gabba before tonight's game #CWC15 #AUSvBAN

7:35 PM - 20 Feb 2015

🔄 15 ❤️ 11

Tournament

Overview

Data
Collection

Database

Extensions

- 2015 Cricket World Cup
- Hosted in Australia and New Zealand
- Tournament Structure
 - 2 Pools of 7 teams
 - Stage 1: Round-robin play within pools
 - Stage 2: Seeded 8-team single-elimination playoffs
- 49 **48** matches
- Data Collected from ESPNcricinfo website

Web Scraping - Match Page

Overview

Data
Collection

Database

Extensions

The screenshot shows a web browser window with two tabs. The active tab is titled 'Final: Australia v New Zealand' and shows the URL 'www.espnricinfo.com/icc-cricket-world-cup-2015/engine/current/match/656495.html'. The URL is circled in red. The page content includes the ESPN Cricinfo logo, 'LIVE SCORES', and social media buttons for 'Like 30K' and 'Tweet'. Below this is the match title 'ICC Cricket World Cup, Final: Australia v New Zealand at Melbourne, Mar 29, 2015'. A navigation bar contains links for 'Scorecard', 'Commentary', 'Wickets', 'Partnership table', 'Player v player table', 'Over comparison', 'Career averages', 'Almanack report', 'Report', 'Articles (30)', 'Photos (66)', 'Videos (20)', 'Hawk-Eye', 'Wagon wheel', 'Manhattan', 'Worm', 'Run rate graph', 'Player v player graph', 'Partnership graph', 'Scoring shots graph', 'Wickets pie', and 'Extras pie'. The match status is 'ICC Cricket World Cup - Final Australia v New Zealand' with the note 'Australia won by 7 wickets (with 101 balls remaining)'. Match details include 'ODI no. 3646 | 2014/15 season', 'Played at Melbourne Cricket Ground', and '29 March 2015 - day/night match (50-over match)'. A table titled 'New Zealand innings (50 overs maximum)' displays batting statistics for four players.

New Zealand innings (50 overs maximum)		R	M	B	4s	6s	SR
MJ Guptill	b Maxwell	15	48	34	1	1	44.11
BB McCullum*	b Starc	0	3	3	0	0	0.00
KS Williamson	c & b Johnson	12	52	33	1	0	36.36
LRPL Taylor	c †Haddin b Faulkner	40	104	72	2	0	55.55

Web Scraping - Match Page

Overview

Data
Collection

Database

Extensions

David

www.espnricinfo.com/icc-cricket-world-cup-2015/engine/current/match/656495.json

```
{
  "centre": {
    "batting": [
      {
        "balls_faced": 72,
        "batting_style": "rhb",
        "control_percentage": 92,
        "dismissal_name": "bowled",
        "dot_ball_percentage": 59,
        "known_as": "Michael Clarke",
        "live_current_name": "",
        "match_award": 0,
        "notout": 0,
        "player_id": 8876,
        "popular_name": "Cl",
        "preferred_shot": {
          "balls_faced": 13,
          "runs": 20,
          "runs_summary": [
            "5", "4", "0", "0", "4", "0", "0", "0"],
          "shot_name": "cut shot",
          "runs": 74,
          "runs_summary": [
            "43", "11", "4", "3", "10", "0", "1", "0"],
          "scoring_shots": 29,
          "strike_rate": 102.77,
          "wagon_zone": [
            {
              "runs": 1,
              "runs_summary": [0, 1, 0, 0, 0, 0, 0, 0],
              "scoring_shots": 1,
              {
                "runs": 3,
                "runs_summary": [0, 1, 1, 0, 0, 0, 0, 0],
                "scoring_shots": 2,
                {
                  "runs": 9,
                  "runs_summary": [0, 1, 2, 0, 1, 0, 0, 0],
                  "scoring_shots": 4,
                  {
                    "runs": 12,
                    "runs_summary": [0, 4, 0, 0, 2, 0, 0, 0],
                    "scoring_shots": 6,
                    {
                      "runs": 8,
                      "runs_summary": [0, 0, 1, 0, 0, 0, 1, 0],
                      "scoring_shots": 2,
                      {
                        "runs": 21,
                        "runs_summary": [0, 0, 0, 3, 3, 0, 0, 0],
                        "scoring_shots": 6,
                        {
                          "runs": 7,
                          "runs_summary": [0, 3, 0, 0, 1, 0, 0, 0],
                          "scoring_shots": 4,
                          {
                            "runs": 13,
                            "runs_summary": [0, 1, 0, 0, 3, 0, 0, 0],
                            "scoring_shots": 4
                        }
                      }
                    }
                  }
                }
              }
            }
          ]
        },
        "balls_faced": 71,
        "batting_style": "rhb",
        "control_percentage": 87,
        "dismissal_name": "not out",
        "dot_ball_percentage": 47,
        "known_as": "Steven Smith",
        "live_current_name": "striker",
        "match_award": 0,
        "notout": 1,
        "player_id": 50281,
        "popular_name": "Smith",
        "preferred_shot": {
          "balls_faced": 5,
          "runs": 11,
          "runs_summary": [
            "0", "3", "0", "0", "2", "0", "0", "0"],
          "shot_name": "pull",
          "runs": 56,
          "runs_summary": [
            "34", "26", "6", "2", "3", "0", "0", "0"],
          "scoring_shots": 37,
          "strike_rate": 78.87,
          "wagon_zone": [
            {
              "runs": 2,
              "runs_summary": [0, 0, 1, 0, 0, 0, 0, 0],
              "scoring_shots": 1,
              {
                "runs": 13,
                "runs_summary": [0, 7, 1, 0, 1, 0, 0, 0],
                "scoring_shots": 9,
                {
                  "runs": 12,
                  "runs_summary": [0, 6, 1, 0, 1, 0, 0, 0],
                  "scoring_shots": 8,
                  {
                    "runs": 4,
                    "runs_summary": [0, 4, 0, 0, 0, 0, 0, 0],
                    "scoring_shots": 4,
                    {
                      "runs": 3,
                      "runs_summary": [0, 3, 0, 0, 0, 0, 0, 0],
                      "scoring_shots": 3,
                      {
                        "runs": 9,
                        "runs_summary": [0, 2, 2, 1, 0, 0, 0, 0],
                        "scoring_shots": 5,
                        {
                          "runs": 6,
                          "runs_summary": [0, 1, 1, 1, 0, 0, 0, 0],
                          "scoring_shots": 3,
                          {
                            "runs": 7,
                            "runs_summary": [0, 3, 0, 0, 1, 0, 0, 0],
                            "scoring_shots": 4
                        }
                      }
                    }
                  }
                }
              }
            }
          ]
        },
        "bowling": [
          {
            "bowling_style": "rfm",
            "conceded": 46,
            "economy_rate": 5.01,
            "known_as": "Matt Henry",
            "live_current_name": "current bowler",
            "maidens": 0,
            "match_award": 0,
            "overall_lhb": {
              "balls": 11,
              "conceded": 9,
              "economy_rate": 4.90,
              "wickets": 1,
              "overall_rhb": {
                "balls": 146,
                "conceded": 37,
                "economy_rate": 5.04,
                "wickets": 1,
                "overs": 9.1,
                "pitch map_lhb":

```

Data Processing

- Match data from ESPNcricinfo website

Data Processing

- Match data from ESPNcricinfo website
- `python-espncricinfo` module from GitHub used to scrape data

Data Processing

- Match data from ESPNcricinfo website
- `python-espncricinfo` module from GitHub used to scrape data
- Python list comprehensions and `pandas` used to organize data and add missing columns

Data Processing

- Match data from ESPNcricinfo website
- `python-espncricinfo` module from GitHub used to scrape data
- Python list comprehensions and `pandas` used to organize data and add missing columns
- `sqlite3` module used for conversion to SQL

ER Diagram

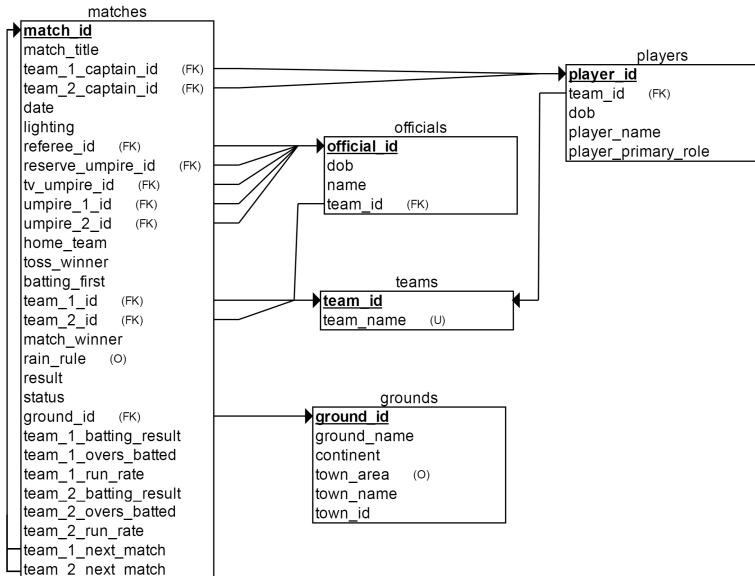
Overview

Data

Collection

Database

Extensions

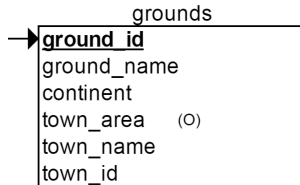


match_id	
match_title	
team_1_captain_id	(FK)
team_2_captain_id	(FK)
date	
lighting	
referee_id	(FK)
reserve_umpire_id	(FK)
tv_umpire_id	(FK)
umpire_1_id	(FK)
umpire_2_id	(FK)
home_team	
toss_winner	
batting_first	
team_1_id	(FK)
team_2_id	(FK)
match_winner	
rain_rule	(O)
result	
status	
ground_id	(FK)
team_1_batting_result	
team_1_innings	
team_1_overs_batted	
team_1_run_rate	
team_2_batting_result	
team_2_innings	
team_2_overs_batted	
team_2_run_rate	
team_1_next_match	
team_2_next_match	

Tables: Matches

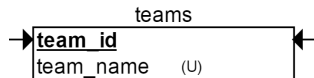
- match_id is primary key
- 2 team_id keys point to teams table
- match_id is primary key
- 5 official ids point to officials table
- 2 captain_id keys point to players table
- ground_id points to ground table
- 2 next_match keys point back to matches table
- all other columns provide details about the match time, conditions, match status and result

Tables: Grounds



- **ground_id** is primary key
- **town_area** is optional
- **town_id** is presumably a foreign key to another ESPNcricinfo table.

Tables: Teams



- **team_id** is primary key
- **team_name** is team abbreviation
- ESPNcricinfo's database includes additional attributes including team full name, logo, etc.

Tables: Players



- `player_id` is primary key
- `player_primary_role` is usually batsman, bowler, or all-rounder

Tables: Officials



- `official_id` is primary key
- `team_id` matches to `teams` table.
(Officials may not officiate their home country's match.)

Extensions

- Extend to pool A and B results
- Include more match details for players, including runs scored and wickets taken
- Scrape ball-by-ball detail
- Scrape ball tracking data
- Streamline data collection and organization