Sports Analytics

Basketball Data Analysis

The SportsAnalytics package (version 0.2) for R is used for this module. The version, along with its dependencies, can be installed from

http://cran.r-project.org/src/contrib/SportsAnalytics_0.2.tar.gz

The above package is still in its initial stages of development. The NBA player statistics can be downloaded using this package. The fetch_NBAPlayerStatistics function returns a data frame of the player statistics for the specified season. By default, the complete statistics (home and away games) is included. The raw data for this function is obtained from the website

http://dougstats.com

```
> library(SportsAnalytics)
>
> nba0708 <- fetch_NBAPlayerStatistics("07-08")
```

The data frame provides the player information using the following columns.

```
> names(nba0708)
 [1] "League"              "Name"
 [3] "Team"                "Position"
 [5] "GamesPlayed"         "TotalMinutesPlayed"
 [7] "FieldGoalsMade"      "FieldGoalsAttempted"
 [9] "ThreesMade"          "ThreesAttempted"
[11] "FreeThrowsMade"      "FreeThrowsAttempted"
[13] "OffensiveRebounds"   "TotalRebounds"
[15] "Assists"             "Steals"
[17] "Turnovers"           "Blocks"
[19] "PersonalFouls"       "Disqualifications"
[21] "TotalPoints"         "Technicals"
[23] "Ejections"           "FlagrantFouls"
[25] "GamesStarted"
```

The teams in the data frame can be viewed as follows.

```
> levels(nba0708$Team)
 [1] "ATL" "BOS" "CHA" "CHI" "CLE" "DAL" "DEN" "DET"
 [9] "GSW" "HOU" "IND" "LAC" "LAL" "MEM" "MIA" "MIL"
[17] "MIN" "NJN" "NOR" "NYK" "ORL" "PHI" "PHO" "POR"
[25] "SAC" "SAN" "SEA" "TOR" "UTA" "WAS"
```

The player statistics for the Boston Celtics for the 2007–2008 season can be filtered using the subset function as shown below. Some of the columns in the resulting data frame are shown using the Google Table chart function.

```
> nba0708.bos <- subset(nba0708, Team == 'BOS')
>
> chart1 <- gvisTable(nba0708.bos[,c(2,4:12)])
> plot(chart1)
```
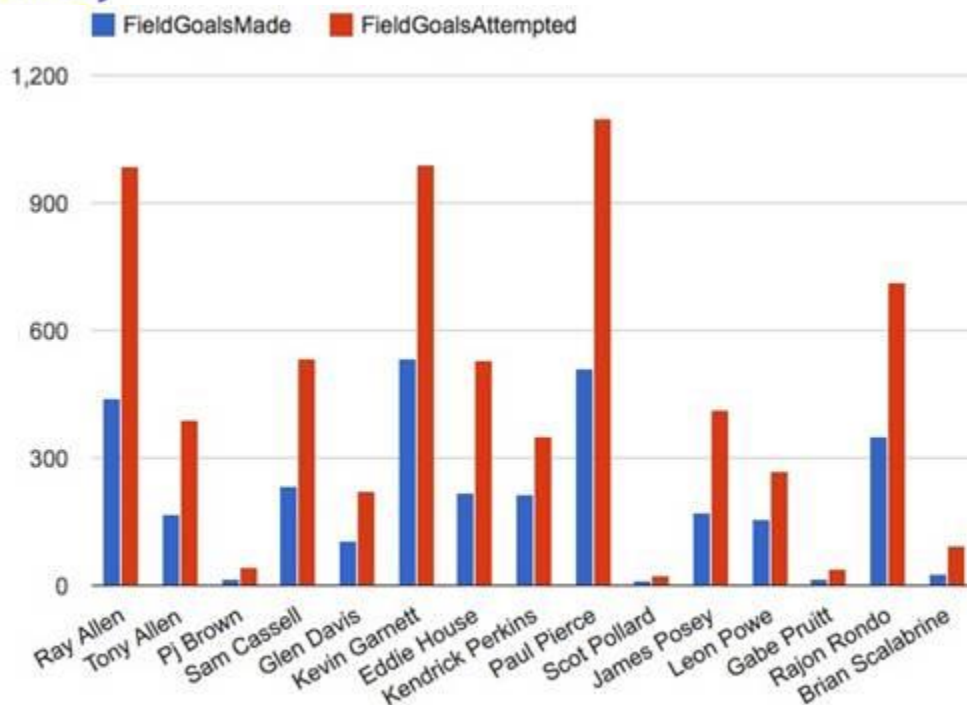
Name	Position	GamesPlayed	TotalMinutesPlayed	FieldGoalsMade	FieldGoalsAttempted	ThreesMade	ThreesAttempted	FreeThrowsMade	FreeThrowsAttempted
Ray Allen	SG	73	2621	439	987	180	452	215	237
Tony Allen	PG	75	1377	169	388	18	57	138	181
PJ Brown	PF	18	208	14	41	0	1	11	16
Sam Cassell	PG	55	1274	233	532	23	76	127	144
Glen Davis	SF	69	941	107	221	0	0	99	150
Kevin Garnett	PF	71	2329	534	990	0	11	269	336
Eddie House	SG	78	1481	217	530	117	298	33	36
Kendrick Perkins	C	78	1912	214	349	0	1	114	183
Paul Pierce	SF	80	2873	509	1098	143	365	409	485
Scot Pollard	C	22	173	12	23	0	0	15	22
James Posey	SF	74	1816	173	414	106	279	93	115
Leon Powe	SF	56	809	154	268	0	1	137	193
Gabe Pruitt	PG	15	96	14	39	3	12	1	2
Rajon Rondo	PG	77	2308	351	713	5	19	107	175
Brian Scalabrine	PF	48	512	29	94	15	46	15	20

Data: nba0708.bos[, c(2, 4:12)] • Chart ID: TableID72ee78f7aff44 • googleVis-0.5.8
R version 3.1.2 (2014-10-31) • Google Terms of Use • Documentation and Data Policy

The Google Column chart shown below compares the field goals made versus the field goals attempted for each player in the Boston team. Other analysis about players from a team or players across all the teams can similarly be carried out.

```
> chart2 <-
+    gvisColumnChart(
+        nba0708.bos,
+        xvar = "Name",
+        yvar = c("FieldGoalsMade", "FieldGoalsAttempted"),
+        options=list(
+          legend="top",
+          height=500, width=850))
> plot(chart2)
```


Data: data • Chart ID: ColumnChartID72ee58872de2 • googleVis-0.5.8
R version 3.1.2 (2014-10-31) • Google Terms of Use • Documentation and Data Policy

Basketball Championship Data
The website
http://www.landofbasketball.com/championships/year_by_year.htm provides year-by-year NBA championship information. More statistics about the games is also available from this website. The following example retrieves the data from the HTML table and provides the means to do analysis on the data.

```
> library(XML)
> library(stringr)
>
> webpage <-
+     paste0("http://www.landofbasketball.com/",
+             "championships/year_by_year.htm")
> data <-
+     readHTMLTable(webpage,
+                     which = 1,
+                     stringsAsFactors = FALSE)
> names(data) <- c("Input")
```

The table provided by the website is read as a single column in the resulting data frame. The data in the first row for the named Input column is shown below.

```
> data[1, "Input"]
[1] "2013-14\n          \n    \n  \n        San Antonio Spurs\n
\n       4-1\n        Miami Heat\n    \n  \n        \n      Fina
ls MVP:\n       \t\t  \n        Kawhi Leonard          (Spurs)\n
\n        \n      Season MVP:\n                \t\n            Kevin
Durant         (Thunder)"
```

For extracting the data into multiple columns, the above column is first split using the newline and tab characters as delimiters and inserted into the data frame.

```
> data$Split <-
+     strsplit(data$Input, split='(\n|\t)')
```

The data in the first row for the computed Split column is shown below.

```
> data[1, "Split"]
[[1]]
 [1] "2013-14"
 [2] "            "
 [3] "     "
 [4] "   "
 [5] "          San Antonio Spurs"
 [6] "      "
 [7] "       4-1"
 [8] "       Miami Heat"
 [9] "     "
[10] "    "
[11] "         "
[12] "       Finals MVP:"
[13] "        "
[14] ""
[15] "    "
[16] "       Kawhi Leonard         (Spurs)"
[17] "          "
[18] "        "
[19] "       Season MVP:"
[20] "           "
[21] ""
[22] "           Kevin Durant              (Thunder)"
```
Using the above parsed data, the year for the season, the winning team, the series score, and the opponent columns are computed.

```
> data$Year <-
+    sapply(data$Split,
+            FUN = function (x)
+                substr(x[[1]], 3, 7))
>
> data$Winner <-
+    sapply(data$Split,
+            FUN = function (x)
+                str_trim(x[[5]]))
>
> data$Series <-
+    sapply(data$Split,
+            FUN = function (x)
+                str_trim(x[[7]]))
>
> data$Opponent <-
+    sapply(data$Split,
+            FUN = function (x)
+                str_trim(x[[8]]))
```

If the Input and the Split columns are no longer needed, they can be removed from the data frame.

```
> data$Split <- NULL
> data$Input <- NULL
```

The resulting data frame can be plotted using the Google Table chart as shown below.

```
> chart3 <- gvisTable(data,
+                            options=list(
+                                 height=500))
> plot(chart3)
```

Year	Winner	Series	Opponent
13-14	San Antonio Spurs	4-1	Miami Heat
12-13	Miami Heat	4-3	San Antonio Spurs
11-12	Miami Heat	4-1	Oklahoma City Thunder
10-11	Dallas Mavericks	4-2	Miami Heat
09-10	Los Angeles Lakers	4-3	Boston Celtics
08-09	Los Angeles Lakers	4-1	Orlando Magic
07-08	Boston Celtics	4-2	Los Angeles Lakers
06-07	San Antonio Spurs	4-0	Cleveland Cavaliers
05-06	Miami Heat	4-2	Dallas Mavericks
04-05	San Antonio Spurs	4-3	Detroit Pistons
03-04	Detroit Pistons	4-1	Los Angeles Lakers
02-03	San Antonio Spurs	4-2	New Jersey Nets
01-02	Los Angeles Lakers	4-0	New Jersey Nets
00-01	Los Angeles Lakers	4-1	Philadelphia 76ers
99-00	Los Angeles Lakers	4-2	Indiana Pacers
98-99	San Antonio Spurs	4-1	New York Knicks
97-98	Chicago Bulls	4-2	Utah Jazz
96-97	Chicago Bulls	4-2	Utah Jazz
95-96	Chicago Bulls	4-2	Seattle Supersonics
94-95	Houston Rockets	4-0	Orlando Magic
93-94	Houston Rockets	4-3	New York Knicks
92-93	Chicago Bulls	4-2	Phoenix Suns
91-92	Chicago Bulls	4-2	Portland Trail Blazers

Data: data • Chart ID: TableID72ee3276844b • googleVis-0.5.8
R version 3.1.2 (2014-10-31) • Google Terms of Use • Documentation and Data Policy

References

Package 'SportsAnalytics': Infrastructure for Sports Analytics, February 19, 2015.
http://cran.r-project.org/web/packages/SportsAnalytics/SportsAnalytics.pdf

Doug's NBA & MLB Statistics Home Page.
http://dougstats.com

Land of Basketball.com: Teams, player profiles, awards, stats, records and championships.
http://www.landofbasketball.com/index.htm