Basketball Intelligence Software based on NBA data

Languages used: Python 2.7, R

Libraries used: Standard Python libraries, csv, math, reportlab, ggplot2, gridExtra, grid , png

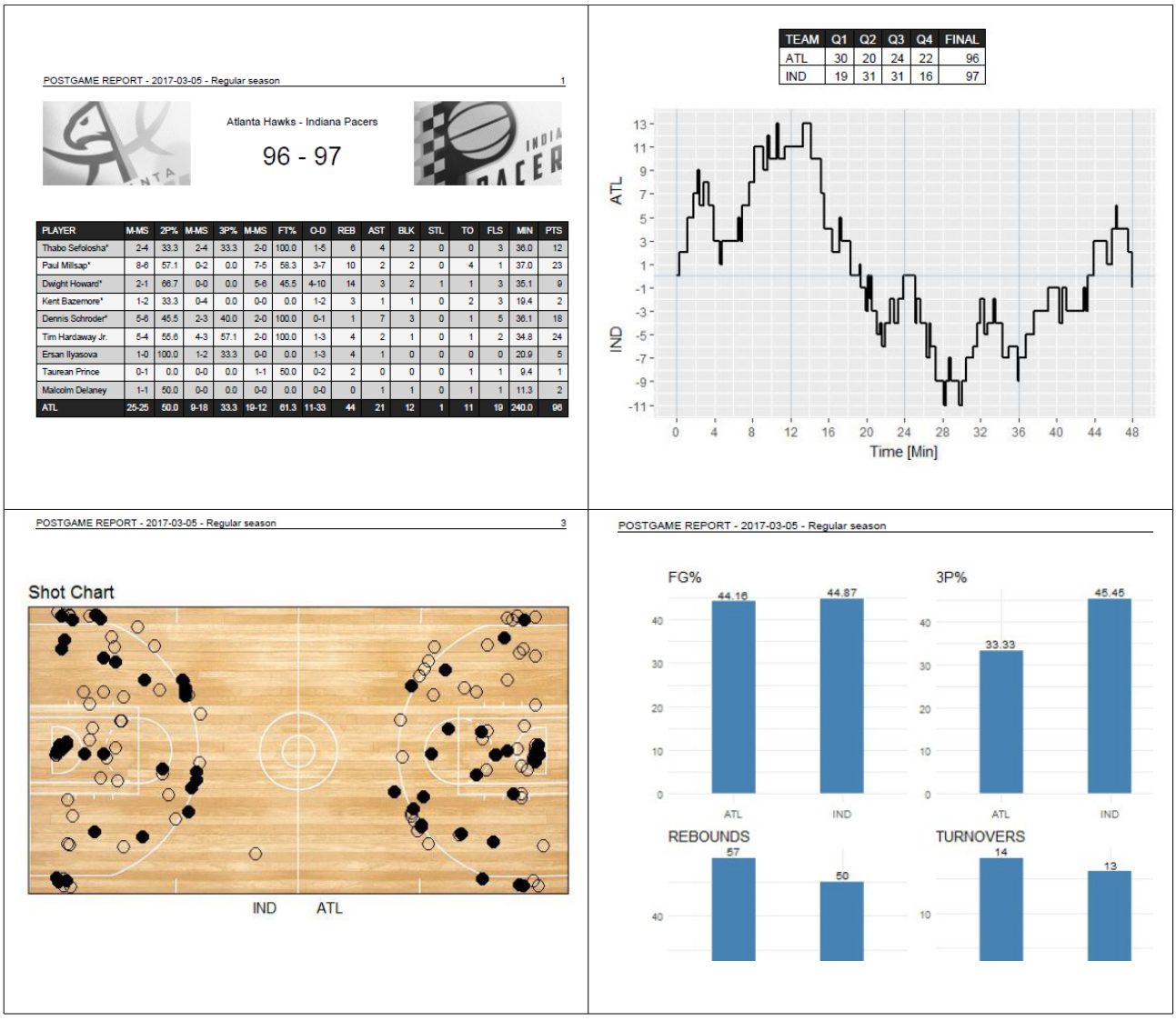
Description: The software is designed to generate well-known traditional (classic) and advanced basketball statistics. It can help you evaluate and monitor your team's and players' play (on-court achievements), offer comparison methods and introduce new ideas.

Input data: Play-by-play files.

Output data: Reports in the form of PDF and CSV files. There are about 43 different reports plus a number of postgame reports. PDF Reports are equipped with graphs and charts which visualize the results, thus ensure a better overview of the data.

Lines of code: +8,000

Processed data: 2 seasons; +2,500 games; +450,000 possessions; +1,000,000 plays



POSTGAME REPORT - 2017-03-05 - Regular season 1

Atlanta Hawks - Indiana Pacers

96 - 97

PLAYER	M-MS	2P%	M-MS	3P%	M-MS	FT%	O-D	REB	AST	BLK	STL	TO	FLS	MIN	PTS
Thabo Sefolosha*	2-4	33.3	2-4	33.3	2-0	100.0	1-5	6	4	2	0	0	3	36.0	12
Paul Millsap*	8-8	57.1	0-2	0.0	7-5	58.3	3-7	10	2	2	0	4	1	37.0	23
Dwight Howard*	2-1	66.7	0-0	0.0	5-6	45.5	4-10	14	3	2	1	1	3	35.1	9
Kent Bazemore*	1-2	33.3	0-4	0.0	0-0	0.0	1-2	3	1	1	0	2	3	19.4	2
Dennis Schroder*	5-6	45.5	2-3	40.0	2-0	100.0	0-1	1	7	3	0	1	5	36.1	18
Tim Hardaway Jr.	5-4	55.6	4-3	57.1	2-0	100.0	1-3	4	2	1	0	1	2	34.8	24
Ersan Ilyasova	1-0	100.0	1-2	33.3	0-0	0.0	1-3	4	1	0	0	0	0	20.9	5
Taurean Prince	0-1	0.0	0-0	0.0	1-1	50.0	0-2	2	0	0	0	1	1	9.4	1
Malcolm Delaney	1-1	50.0	0-0	0.0	0-0	0.0	0-0	0	1	1	0	1	1	11.3	2
ATL	25-25	50.0	9-18	33.3	19-12	61.3	11-33	44	21	12	1	11	19	240.0	96

TEAM	Q1	Q2	Q3	Q4	FINAL
ATL	30	20	24	22	96
IND	19	31	31	16	97

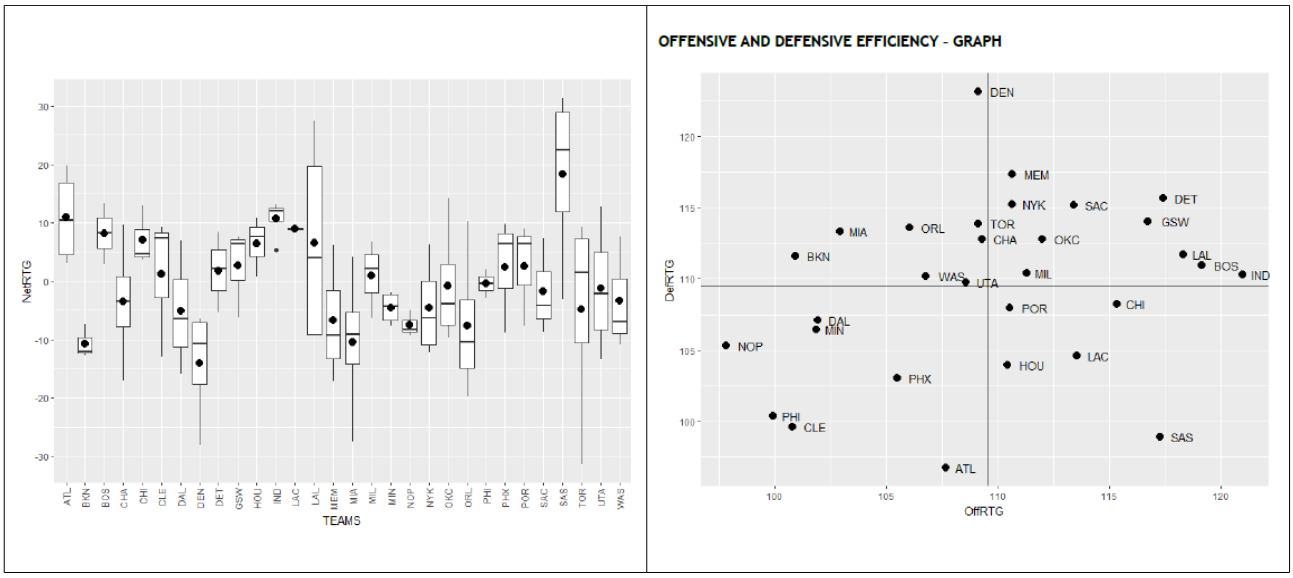
POSTGAME REPORT - 2017-03-05 - Regular season 3

POSTGAME REPORT - 2017-03-05 - Regular season

The software includes many validation functions to ensure data are correct. Some of them include:

- corrections of players' names,
- checking corrections of play, quarters and games time lengths,
- checking if there are exactly 2 teams competing,
- comparing scoring change and final score,
- isolating each team possessions and checking if teams alternate possessions,
- and many more.

The software is divided into two main parts. Python is used for preparing data (extracting, cleaning, transforming). R is used for statistical and visualization purposes.



Thanks to the software I was able to generate postgame reports, team benchmarks and power rankings (presented on YouTube channel).