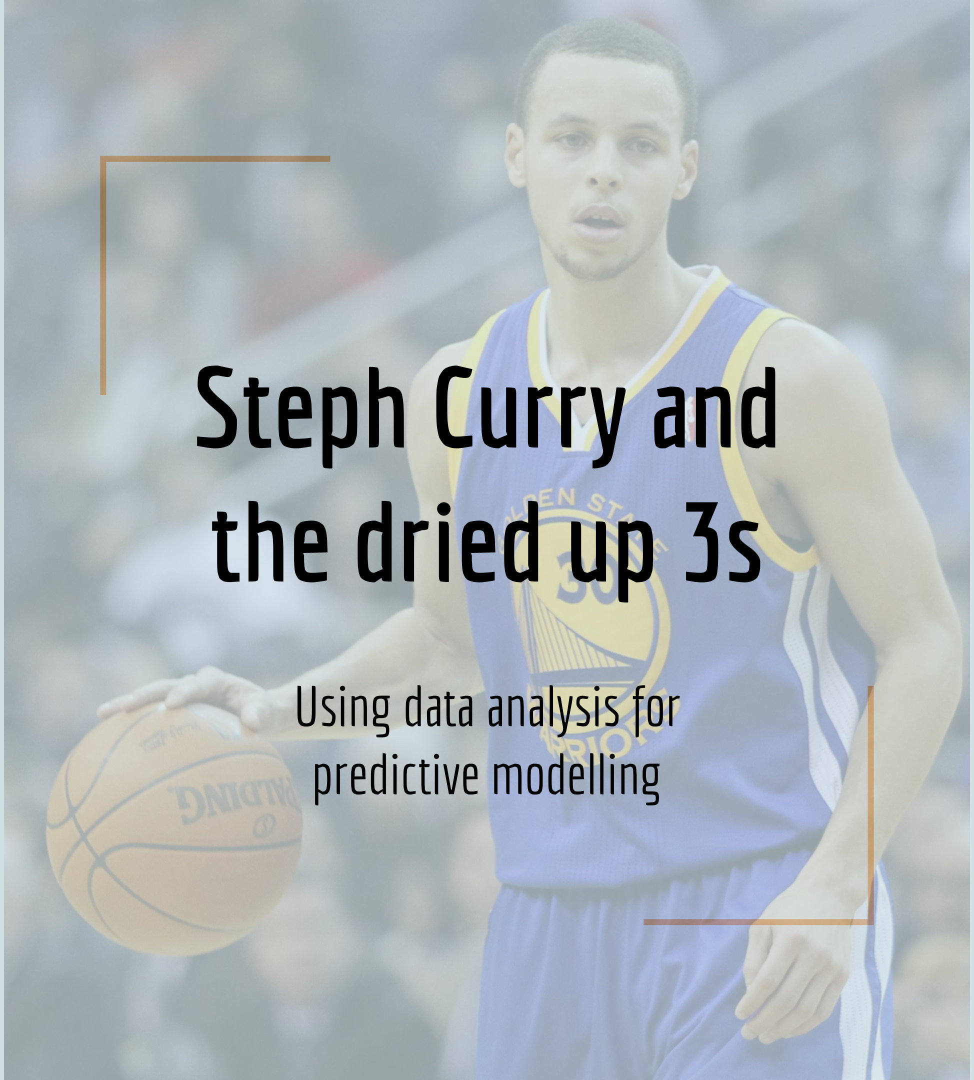


Steph Curry and the dried up 3s

Using data analysis for predictive modelling

So what is data science, anyway?

When you hear the term data scientist, what do you think of? If you're like most people, you might think of something really complex, with statistical terms and programming languages no one can understand. You might think that you might need to go to university to learn how to do data science.

Data science is just a 21st century spin on maths that people have been doing for centuries. Big data, data science and analytics are just fancy ways of saying we use information available to gain insight and improve a business.

Whether it's a small Excel spreadsheet or a 100 million records in a database, the goal is always the same: to find value.

Step 1: Have a question or something you’re curious about

In the 2015 NBA Playoffs, Steph Curry and the Golden State Warriors were down 2 games to 1 against the Memphis Grizzlies, and Curry’s 3-point shooting was down in the previous two games which the Warriors lost. Commentators were speculating; have the Grizzlies figured out Steph Curry? Can he bounce back and guide the Warriors to victory in the best of 7 series?

Grizzlies grind down MVP Stephen Curry, Warriors' offense in Game 3 win

Framing the hypothesis

Let’s now frame the hypothesis.

Curry’s poor 3 point percentage was shooting at below 0.20 for these two matches. But normally he shoots at 0.40. If he shoots over 0.40 then his team wins the game 89% of the time in 2014-15.

You can frame the hypothesis as a return to form or repeat poor performance

Golden State Warriors will return to form and win the series because Steph Curry should be able to average a 0.40 3 point percentage after his two poor shooting games against the Memphis Grizzlies in the 2014-15 NBA playoffs.

Golden State Warriors will lose and bow out of the series against the Memphis Grizzlies in the 2014-15 NBA playoffs because Steph Curry will average clearly less than a 0.40 3 point percentage.

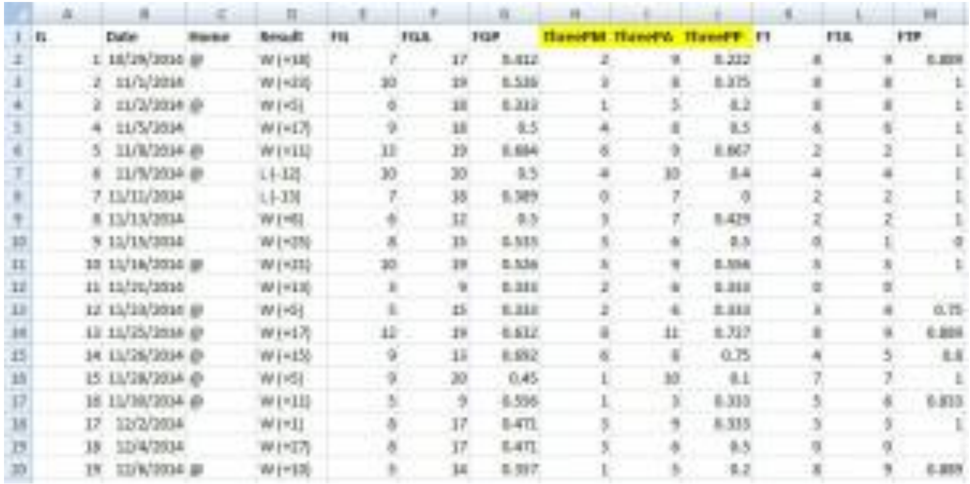
Deciding on the data required

- Average 3 Point Shooting percentage for all matches would be helpful to develop a baseline
- Average 3 Point Shooting percentage for all matches after he has had a poor shooting performance
- “For all matches” - just that season, or how long do you use data, is it less relevant if it’s for a previous year?
- Are there any other statistics that could be useful? Is there an issue with just playing that opponent?

Step 2: Gather data that exists for your area of interest

We can use easily available data from basketball-reference.com in this situation.

I simply took Steph Curry's game log for the 2014-15 regular season and created a .CSV file (uploaded here if you want to download it). Here's what the data looked like:

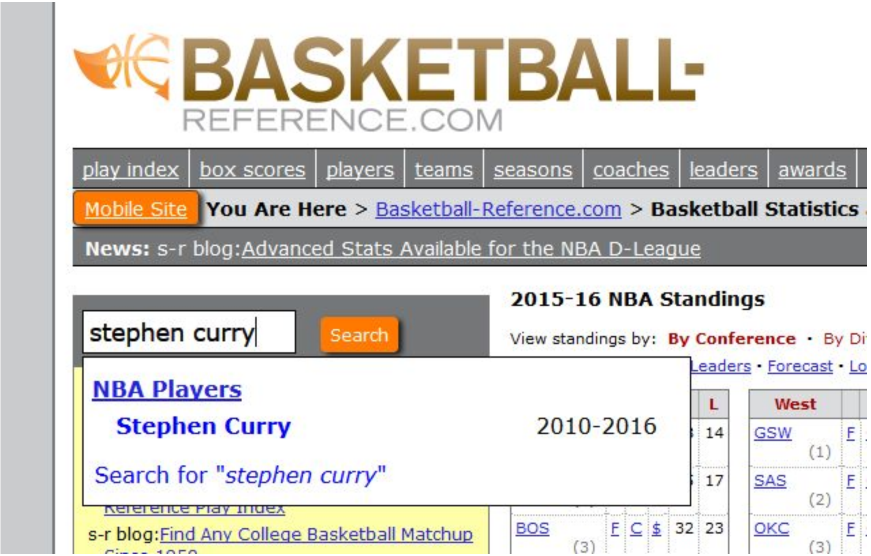


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	G	Date	Home	Result	FG	FGA	FGP	ThreePM	ThreePA	ThreePP	FT	FTA	FTP
2	1	10/29/2014	@	W (+18)	7	17	0.412	2	9	0.222	8	9	0.889
3	2	11/1/2014		W (+20)	10	19	0.526	3	8	0.375	8	8	1
4	3	11/2/2014	@	W (+5)	6	18	0.333	1	5	0.2	8	8	1
5	4	11/5/2014		W (+17)	9	18	0.5	4	8	0.5	6	6	1
6	5	11/8/2014	@	W (+11)	13	19	0.684	6	9	0.667	2	2	1
7	6	11/9/2014	@	L (-12)	10	20	0.5	4	10	0.4	4	4	1
8	7	11/11/2014		L (-13)	7	18	0.389	0	7	0	2	2	1
9	8	11/13/2014		W (+6)	6	12	0.5	3	7	0.429	2	2	1
10	9	11/15/2014		W (+25)	8	15	0.533	3	6	0.5	0	1	0
11	10	11/16/2014	@	W (+21)	10	19	0.526	5	9	0.556	3	3	1
12	11	11/21/2014		W (+13)	3	9	0.333	2	6	0.333	0	0	
13	12	11/23/2014	@	W (+5)	5	15	0.333	2	6	0.333	3	4	0.75
14	13	11/25/2014	@	W (+17)	12	19	0.632	8	11	0.727	8	9	0.889
15	14	11/26/2014	@	W (+15)	9	13	0.692	6	8	0.75	4	5	0.8
16	15	11/28/2014	@	W (+5)	9	20	0.45	1	10	0.1	7	7	1
17	16	11/30/2014	@	W (+11)	5	9	0.556	1	3	0.333	5	6	0.833
18	17	12/2/2014		W (+1)	8	17	0.471	3	9	0.333	3	3	1
19	18	12/4/2014		W (+27)	8	17	0.471	3	6	0.5	0	0	
20	19	12/6/2014	@	W (+13)	5	14	0.357	1	5	0.2	8	9	0.889

Process for gathering the data …

Navigate your web browser to basketball-reference.com

Type Stephen Curry in the search bar and click on the associated link:


BASKETBALL-
REFERENCE.COM
play index | box scores | players | teams | seasons | coaches | leaders | awards
Mobile Site
stephen curry
Search
View standings by: By Conference · By Di
Leaders · Forecast · Lo
L
West
GSW
SAS
OKC
BOS

You Are Here > Basketball-Reference.com > Basketball Statistics

News: s-r blog:Advanced Stats Available for the NBA D-League

2015-16 NBA Standings

NBA Players

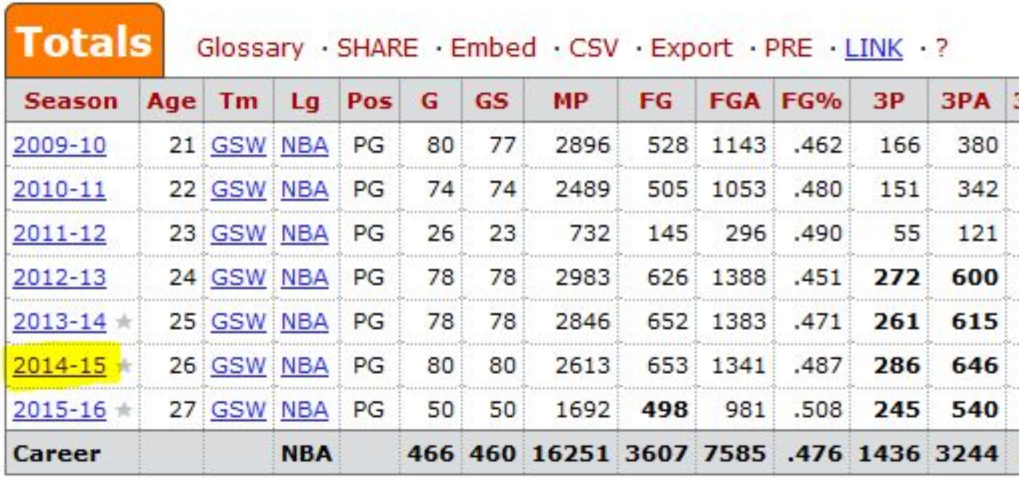
Stephen Curry 2010-2016

Search for "*stephen curry*"

s-r blog:Find Any College Basketball Matchup

Process for gathering the data …

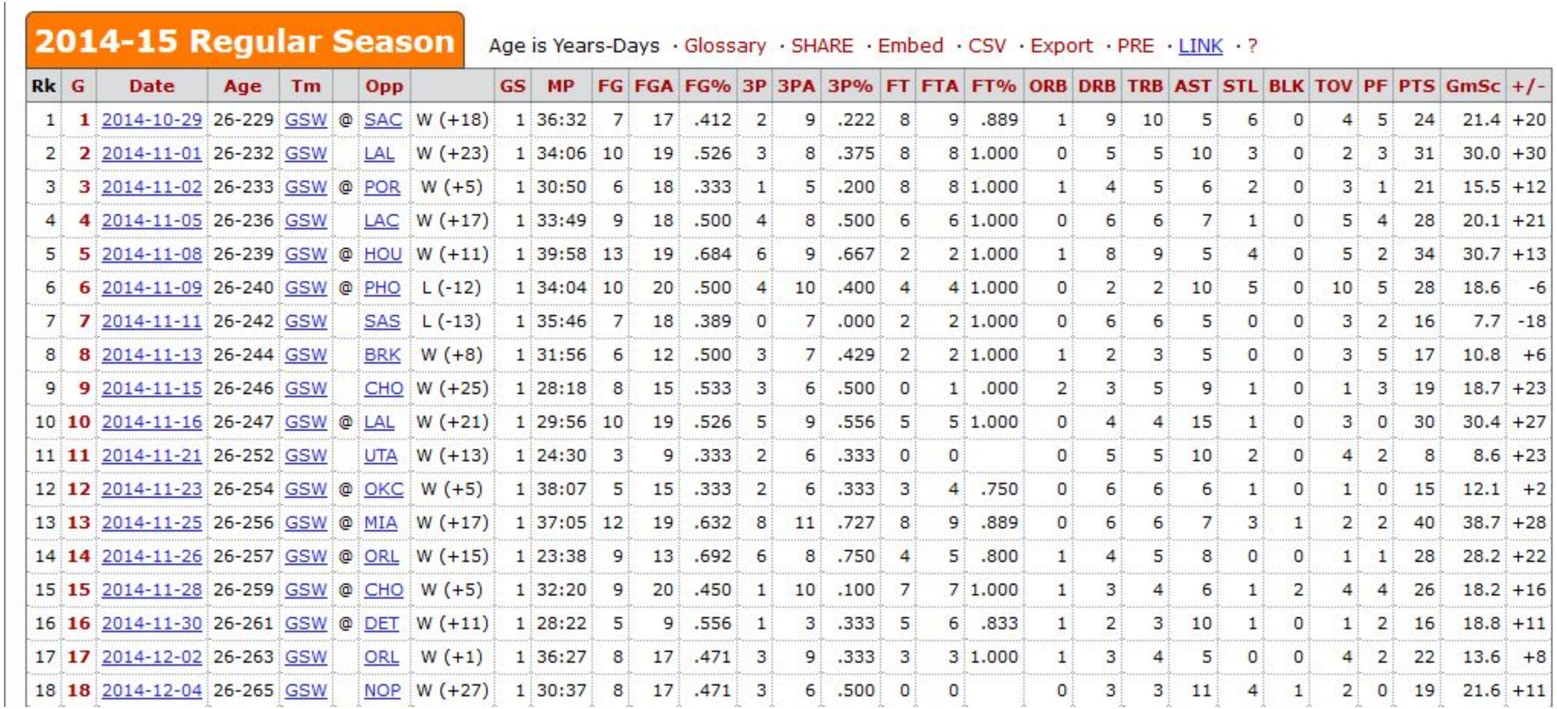
Scroll down and click on 2014-15 to display data from the appropriate season



Season	Age	Tm	Lg	Pos	G	GS	MP	FG	FGA	FG%	3P	3PA
2009-10	21	GSW	NBA	PG	80	77	2896	528	1143	.462	166	380
2010-11	22	GSW	NBA	PG	74	74	2489	505	1053	.480	151	342
2011-12	23	GSW	NBA	PG	26	23	732	145	296	.490	55	121
2012-13	24	GSW	NBA	PG	78	78	2983	626	1388	.451	**272**	**600**
2013-14 ★	25	GSW	NBA	PG	78	78	2846	652	1383	.471	**261**	**615**
2014-15 ★	26	GSW	NBA	PG	80	80	2613	653	1341	.487	**286**	**646**
2015-16 ★	27	GSW	NBA	PG	50	50	1692	**498**	981	.508	**245**	**540**
Career			**NBA**		**466**	**460**	**16251**	**3607**	**7585**	**.476**	**1436**	**3244**

Process for gathering the data ...

Scroll down so that you can see all of the games played by Curry in 2014-15



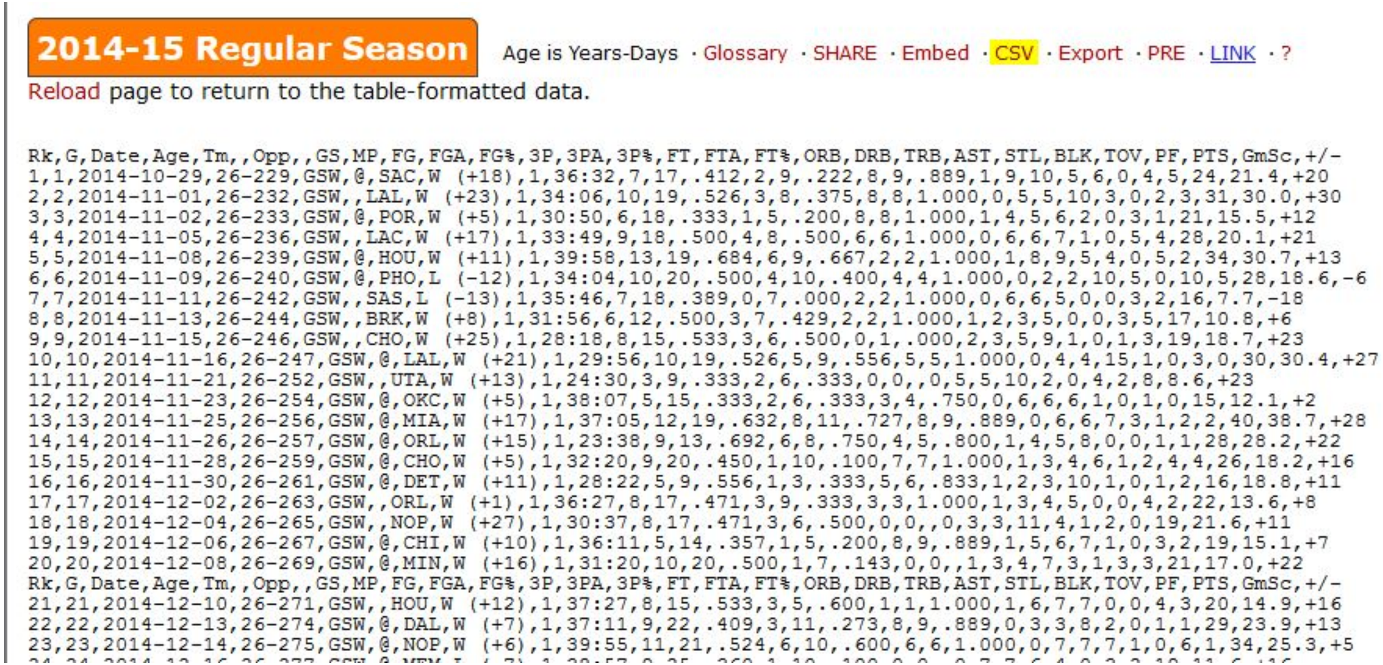
2014-15 Regular Season

Age is Years-Days · Glossary · SHARE · Embed · CSV · Export · PRE · LINK · ?

Rk	G	Date	Age	Tm		Opp		GS	MP	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	GmSc	+/-
1	1	2014-10-29	26-229	GSW	@	SAC	W (+18)	1	36:32	7	17	.412	2	9	.222	8	9	.889	1	9	10	5	6	0	4	5	24	21.4	+20
2	2	2014-11-01	26-232	GSW		LAL	W (+23)	1	34:06	10	19	.526	3	8	.375	8	8	1.000	0	5	5	10	3	0	2	3	31	30.0	+30
3	3	2014-11-02	26-233	GSW	@	POR	W (+5)	1	30:50	6	18	.333	1	5	.200	8	8	1.000	1	4	5	6	2	0	3	1	21	15.5	+12
4	4	2014-11-05	26-236	GSW		LAC	W (+17)	1	33:49	9	18	.500	4	8	.500	6	6	1.000	0	6	6	7	1	0	5	4	28	20.1	+21
5	5	2014-11-08	26-239	GSW	@	HOU	W (+11)	1	39:58	13	19	.684	6	9	.667	2	2	1.000	1	8	9	5	4	0	5	2	34	30.7	+13
6	6	2014-11-09	26-240	GSW	@	PHO	L (-12)	1	34:04	10	20	.500	4	10	.400	4	4	1.000	0	2	2	10	5	0	10	5	28	18.6	-6
7	7	2014-11-11	26-242	GSW		SAS	L (-13)	1	35:46	7	18	.389	0	7	.000	2	2	1.000	0	6	6	5	0	0	3	2	16	7.7	-18
8	8	2014-11-13	26-244	GSW		BRK	W (+8)	1	31:56	6	12	.500	3	7	.429	2	2	1.000	1	2	3	5	0	0	3	5	17	10.8	+6
9	9	2014-11-15	26-246	GSW		CHO	W (+25)	1	28:18	8	15	.533	3	6	.500	0	1	.000	2	3	5	9	1	0	1	3	19	18.7	+23
10	10	2014-11-16	26-247	GSW	@	LAL	W (+21)	1	29:56	10	19	.526	5	9	.556	5	5	1.000	0	4	4	15	1	0	3	0	30	30.4	+27
11	11	2014-11-21	26-252	GSW		UTA	W (+13)	1	24:30	3	9	.333	2	6	.333	0	0		0	5	5	10	2	0	4	2	8	8.6	+23
12	12	2014-11-23	26-254	GSW	@	OKC	W (+5)	1	38:07	5	15	.333	2	6	.333	3	4	.750	0	6	6	6	1	0	1	0	15	12.1	+2
13	13	2014-11-25	26-256	GSW	@	MIA	W (+17)	1	37:05	12	19	.632	8	11	.727	8	9	.889	0	6	6	7	3	1	2	2	40	38.7	+28
14	14	2014-11-26	26-257	GSW	@	ORL	W (+15)	1	23:38	9	13	.692	6	8	.750	4	5	.800	1	4	5	8	0	0	1	1	28	28.2	+22
15	15	2014-11-28	26-259	GSW	@	CHO	W (+5)	1	32:20	9	20	.450	1	10	.100	7	7	1.000	1	3	4	6	1	2	4	4	26	18.2	+16
16	16	2014-11-30	26-261	GSW	@	DET	W (+11)	1	28:22	5	9	.556	1	3	.333	5	6	.833	1	2	3	10	1	0	1	2	16	18.8	+11
17	17	2014-12-02	26-263	GSW		ORL	W (+1)	1	36:27	8	17	.471	3	9	.333	3	3	1.000	1	3	4	5	0	0	4	2	22	13.6	+8
18	18	2014-12-04	26-265	GSW		NOP	W (+27)	1	30:37	8	17	.471	3	6	.500	0	0		0	3	3	11	4	1	2	0	19	21.6	+11

Process for gathering the data …

Click on the CSV option and the data should transform as follows:

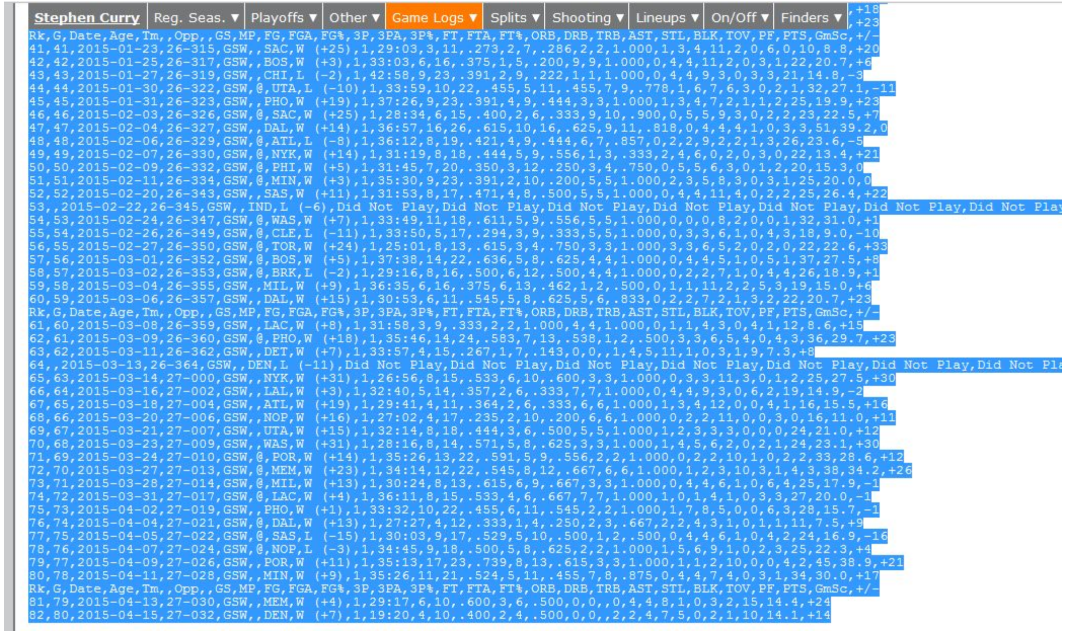

2014-15 Regular Season
Age is Years-Days · Glossary · SHARE · Embed · CSV · Export · PRE · LINK · ?

Reload page to return to the table-formatted data.

```
Rk,G,Date,Age,Tm,,Opp,,GS,MP,FG,FGA,FG%,3P,3PA,3P%,FT,FTA,FT%,ORB,DRB,TRB,AST,STL,BLK,TOV,PF,PTS,GmSc,+/-
1,1,2014-10-29,26-229,GSW,@,SAC,W (+18),1,36:32,7,17,.412,2,9,.222,8,9,.889,1,9,10,5,6,0,4,5,24,21.4,+20
2,2,2014-11-01,26-232,GSW,,LAL,W (+23),1,34:06,10,19,.526,3,8,.375,8,8,1.000,0,5,5,10,3,0,2,3,31,30.0,+30
3,3,2014-11-02,26-233,GSW,@,POR,W (+5),1,30:50,6,18,.333,1,5,.200,8,8,1.000,1,4,5,6,2,0,3,1,21,15.5,+12
4,4,2014-11-05,26-236,GSW,,LAC,W (+17),1,33:49,9,18,.500,4,8,.500,6,6,1.000,0,6,6,7,1,0,5,4,28,20.1,+21
5,5,2014-11-08,26-239,GSW,@,HOU,W (+11),1,39:58,13,19,.684,6,9,.667,2,2,1.000,1,8,9,5,4,0,5,2,34,30.7,+13
6,6,2014-11-09,26-240,GSW,@,PHO,L (-12),1,34:04,10,20,.500,4,10,.400,4,4,1.000,0,2,2,10,5,0,10,5,28,18.6,-6
7,7,2014-11-11,26-242,GSW,,SAS,L (-13),1,35:46,7,18,.389,0,7,.000,2,2,1.000,0,6,6,5,0,0,3,2,16,7.7,-18
8,8,2014-11-13,26-244,GSW,,BRK,W (+8),1,31:56,6,12,.500,3,7,.429,2,2,1.000,1,2,3,5,0,0,3,5,17,10.8,+6
9,9,2014-11-15,26-246,GSW,,CHO,W (+25),1,28:18,8,15,.533,3,6,.500,0,1,.000,2,3,5,9,1,0,1,3,19,18.7,+23
10,10,2014-11-16,26-247,GSW,@,LAL,W (+21),1,29:56,10,19,.526,5,9,.556,5,5,1.000,0,4,4,15,1,0,3,0,30,30.4,+27
11,11,2014-11-21,26-252,GSW,,UTA,W (+13),1,24:30,3,9,.333,2,6,.333,0,0,,0,5,5,10,2,0,4,2,8,8.6,+23
12,12,2014-11-23,26-254,GSW,@,OKC,W (+5),1,38:07,5,15,.333,2,6,.333,3,4,.750,0,6,6,6,1,0,1,0,15,12.1,+2
13,13,2014-11-25,26-256,GSW,@,MIA,W (+17),1,37:05,12,19,.632,8,11,.727,8,9,.889,0,6,6,7,3,1,2,2,40,38.7,+28
14,14,2014-11-26,26-257,GSW,@,ORL,W (+15),1,23:38,9,13,.692,6,8,.750,4,5,.800,1,4,5,8,0,0,1,1,28,28.2,+22
15,15,2014-11-28,26-259,GSW,@,CHO,W (+5),1,32:20,9,20,.450,1,10,.100,7,7,1.000,1,3,4,6,1,2,4,4,26,18.2,+16
16,16,2014-11-30,26-261,GSW,@,DET,W (+11),1,28:22,5,9,.556,1,3,.333,5,6,.833,1,2,3,10,1,0,1,2,16,18.8,+11
17,17,2014-12-02,26-263,GSW,,ORL,W (+1),1,36:27,8,17,.471,3,9,.333,3,3,1.000,1,3,4,5,0,0,4,2,22,13.6,+8
18,18,2014-12-04,26-265,GSW,,NOP,W (+27),1,30:37,8,17,.471,3,6,.500,0,0,,0,3,3,11,4,1,2,0,19,21.6,+11
19,19,2014-12-06,26-267,GSW,@,CHI,W (+10),1,36:11,5,14,.357,1,5,.200,8,9,.889,1,5,6,7,1,0,3,2,19,15.1,+7
20,20,2014-12-08,26-269,GSW,@,MIN,W (+16),1,31:20,10,20,.500,1,7,.143,0,0,,1,3,4,7,3,1,3,3,21,17.0,+22
Rk,G,Date,Age,Tm,,Opp,,GS,MP,FG,FGA,FG%,3P,3PA,3P%,FT,FTA,FT%,ORB,DRB,TRB,AST,STL,BLK,TOV,PF,PTS,GmSc,+/-
21,21,2014-12-10,26-271,GSW,,HOU,W (+12),1,37:27,8,15,.533,3,5,.600,1,1,1.000,1,6,7,7,0,0,4,3,20,14.9,+16
22,22,2014-12-13,26-274,GSW,@,DAL,W (+7),1,37:11,9,22,.409,3,11,.273,8,9,.889,0,3,3,8,2,0,1,1,29,23.9,+13
23,23,2014-12-14,26-275,GSW,@,NOP,W (+6),1,39:55,11,21,.524,6,10,.600,6,6,1.000,0,7,7,7,1,0,6,1,34,25.3,+5
```

Process for gathering the data ...

Highlight all of the comma separated values and copy (Ctrl-C) to clipboard



```
,+18
,+23
Rk,G,Date,Age,Tm,,Opp,,GS,MP,FG,FGA,FG%,3P,3PA,3P%,FT,FTA,FT%,ORB,DRB,TRB,AST,STL,BLK,TOV,PF,PTS,GmSc,+/-
41,41,2015-01-23,26-315,GSW,,SAC,W (+25),1,29:03,3,11,.273,2,7,.286,2,2,1.000,1,3,4,11,2,0,6,0,10,8.8,+20
42,42,2015-01-25,26-317,GSW,,BOS,W (+3),1,33:03,6,16,.375,1,5,.200,9,9,1.000,0,4,4,11,2,0,3,1,22,20.7,+6
43,43,2015-01-27,26-319,GSW,,CHI,L (-2),1,42:58,9,23,.391,2,9,.222,1,1,1.000,0,4,4,9,3,0,3,3,21,14.8,-3
44,44,2015-01-30,26-322,GSW,@,UTA,L (-10),1,33:59,10,22,.455,5,11,.455,7,9,.778,1,6,7,6,3,0,2,1,32,27.1,-11
45,45,2015-01-31,26-323,GSW,,PHO,W (+19),1,37:26,9,23,.391,4,9,.444,3,3,1.000,1,3,4,7,2,1,1,2,25,19.9,+23
46,46,2015-02-03,26-326,GSW,@,SAC,W (+25),1,28:34,6,15,.400,2,6,.333,9,10,.900,0,5,5,9,3,0,2,2,23,22.5,+7
47,47,2015-02-04,26-327,GSW,,DAL,W (+14),1,36:57,16,26,.615,10,16,.625,9,11,.818,0,4,4,4,1,0,3,3,51,39.2,0
48,48,2015-02-06,26-329,GSW,@,ATL,L (-8),1,36:12,8,19,.421,4,9,.444,6,7,.857,0,2,2,9,2,2,1,3,26,23.6,-5
49,49,2015-02-07,26-330,GSW,@,NYK,W (+14),1,31:19,8,18,.444,5,9,.556,1,3,.333,2,4,6,0,2,0,3,0,22,13.4,+21
50,50,2015-02-09,26-332,GSW,@,PHI,W (+5),1,31:45,7,20,.350,3,12,.250,3,4,.750,0,5,5,6,3,0,1,2,20,15.3,0
51,51,2015-02-11,26-334,GSW,@,MIN,W (+3),1,35:30,9,23,.391,2,10,.200,5,5,1.000,2,3,5,8,3,0,3,1,25,20.0,0
52,52,2015-02-20,26-343,GSW,,SAS,W (+11),1,31:53,8,17,.471,4,8,.500,5,5,1.000,0,4,4,11,4,0,2,2,25,26.4,+22
53,,2015-02-22,26-345,GSW,,IND,L (-6),Did Not Play,Did Not Play,Did Not Play,Did Not Play,Did Not Play,Did Not Play,Did Not Play
54,53,2015-02-24,26-347,GSW,@,WAS,W (+7),1,33:49,11,18,.611,5,9,.556,5,5,1.000,0,0,0,8,2,0,0,1,32,31.0,+1
55,54,2015-02-26,26-349,GSW,@,CLE,L (-11),1,33:50,5,17,.294,3,9,.333,5,5,1.000,0,3,3,6,1,0,4,3,18,9.0,-10
56,55,2015-02-27,26-350,GSW,@,TOR,W (+24),1,25:01,8,13,.615,3,4,.750,3,3,1.000,3,3,6,5,2,0,2,0,22,22.6,+33
57,56,2015-03-01,26-352,GSW,@,BOS,W (+5),1,37:38,14,22,.636,5,8,.625,4,4,1.000,0,4,4,5,1,0,5,1,37,27.5,+8
58,57,2015-03-02,26-353,GSW,@,BRK,L (-2),1,29:16,8,16,.500,6,12,.500,4,4,1.000,0,2,2,7,1,0,4,4,26,18.9,+1
59,58,2015-03-04,26-355,GSW,,MIL,W (+9),1,36:35,6,16,.375,6,13,.462,1,2,.500,0,1,1,11,2,2,5,3,19,15.0,+6
60,59,2015-03-06,26-357,GSW,,DAL,W (+15),1,30:53,6,11,.545,5,8,.625,5,6,.833,0,2,2,7,2,1,3,2,22,20.7,+23
Rk,G,Date,Age,Tm,,Opp,,GS,MP,FG,FGA,FG%,3P,3PA,3P%,FT,FTA,FT%,ORB,DRB,TRB,AST,STL,BLK,TOV,PF,PTS,GmSc,+/-
61,60,2015-03-08,26-359,GSW,,LAC,W (+8),1,31:58,3,9,.333,2,2,1.000,4,4,1.000,0,1,1,4,3,0,4,1,12,8.6,+15
62,61,2015-03-09,26-360,GSW,@,PHO,W (+18),1,35:46,14,24,.583,7,13,.538,1,2,.500,3,3,6,5,4,0,4,3,36,29.7,+23
63,62,2015-03-11,26-362,GSW,,DET,W (+7),1,33:57,4,15,.267,1,7,.143,0,0,,1,4,5,11,1,0,3,1,9,7.3,+8
64,,2015-03-13,26-364,GSW,,DEN,L (-11),Did Not Play,Did Not Play,Did Not Play,Did Not Play,Did Not Play,Did Not Play,Did Not Pla
65,63,2015-03-14,27-000,GSW,,NYK,W (+31),1,26:56,8,15,.533,6,10,.600,3,3,1.000,0,3,3,11,3,0,1,2,25,27.5,+30
66,64,2015-03-16,27-002,GSW,,LAL,W (+3),1,32:40,5,14,.357,2,6,.333,7,7,1.000,0,4,4,9,3,0,6,2,19,14.9,-2
67,65,2015-03-18,27-004,GSW,,ATL,W (+19),1,29:41,4,11,.364,2,6,.333,6,6,1.000,1,3,4,12,0,0,4,1,16,15.5,+16
68,66,2015-03-20,27-006,GSW,,NOP,W (+16),1,27:02,4,17,.235,2,10,.200,6,6,1.000,0,2,2,11,0,0,3,0,16,11.0,+11
69,67,2015-03-21,27-007,GSW,,UTA,W (+15),1,32:14,8,18,.444,3,6,.500,5,5,1.000,1,2,3,3,3,0,0,0,24,21.0,+12
70,68,2015-03-23,27-009,GSW,,WAS,W (+31),1,28:16,8,14,.571,5,8,.625,3,3,1.000,1,4,5,6,2,0,2,1,24,23.1,+30
71,69,2015-03-24,27-010,GSW,@,POR,W (+14),1,35:26,13,22,.591,5,9,.556,2,2,1.000,0,2,2,10,1,0,2,2,33,28.6,+12
72,70,2015-03-27,27-013,GSW,@,MEM,W (+23),1,34:14,12,22,.545,8,12,.667,6,6,1.000,1,2,3,10,3,1,4,3,38,34.2,+26
73,71,2015-03-28,27-014,GSW,@,MIL,W (+13),1,30:24,8,13,.615,6,9,.667,3,3,1.000,0,4,4,6,1,0,6,4,25,17.9,-1
74,72,2015-03-31,27-017,GSW,@,LAC,W (+4),1,36:11,8,15,.533,4,6,.667,7,7,1.000,1,0,1,4,1,0,3,3,27,20.0,-1
75,73,2015-04-02,27-019,GSW,,PHO,W (+1),1,33:32,10,22,.455,6,11,.545,2,2,1.000,1,7,8,5,0,0,6,3,28,15.7,-1
76,74,2015-04-04,27-021,GSW,@,DAL,W (+13),1,27:27,4,12,.333,1,4,.250,2,3,.667,2,2,4,3,1,0,1,1,11,7.5,+9
77,75,2015-04-05,27-022,GSW,@,SAS,L (-15),1,30:03,9,17,.529,5,10,.500,1,2,.500,0,4,4,6,1,0,4,2,24,16.9,-16
78,76,2015-04-07,27-024,GSW,@,NOP,L (-3),1,34:45,9,18,.500,5,8,.625,2,2,1.000,1,5,6,9,1,0,2,3,25,22.3,+4
79,77,2015-04-09,27-026,GSW,,POR,W (+11),1,35:13,17,23,.739,8,13,.615,3,3,1.000,1,1,2,10,0,0,4,2,45,38.9,+21
80,78,2015-04-11,27-028,GSW,,MIN,W (+9),1,35:26,11,21,.524,5,11,.455,7,8,.875,0,4,4,7,4,0,3,1,34,30.0,+17
Rk,G,Date,Age,Tm,,Opp,,GS,MP,FG,FGA,FG%,3P,3PA,3P%,FT,FTA,FT%,ORB,DRB,TRB,AST,STL,BLK,TOV,PF,PTS,GmSc,+/-
81,79,2015-04-13,27-030,GSW,,MEM,W (+4),1,29:17,6,10,.600,3,6,.500,0,0,,0,4,4,8,1,0,3,2,15,14.4,+24
82,80,2015-04-15,27-032,GSW,,DEN,W (+7),1,19:20,4,10,.400,2,4,.500,0,0,,2,2,4,7,5,0,2,1,10,14.1,+14
```

Process for gathering the data ...

Open up a spreadsheet application and paste the data into cell A1 (Ctrl-V)

In Google Sheets

- In Cell B1, type: =arrayformula(substitute(split(substitute(A1,",",",|"),","), "|",""))
- Fill down Cell B1
- Highlight B1 to AE87 and Copy to clipboard (Ctrl-C)
- Edit-Paste Special-Paste values only
- Delete Column A

In Excel

- Click Data-Text to Columns
- Choose Delimited, click Next
- Choose Comma, click Next
- Click Finish

Step 3 - Cleanse the data for use

Now we need to cleanse the data to make it usable by keeping only data relevant to our question.

- Adjust cell widths
- Ctrl-H find and replace Did Not Play with DNP
- Adjust cell widths again
- Remove repeated headers (where you see Rk/G every 20 rows)
- Name column F and H (Home, Result)
- Find which columns are needed for this comparison
  - Removed: Rk Age Tm Opp GS MP ORB DRB TRB AST STL BLK TOV PF PTS GmSc +/-

Process for cleansing the data ...

- Changed column header with % in it: FG%, 3P% and FT% to FGP, 3PP and FTP
- Changed column header with 3 in it: 3P 3PA 3PP to ThreePM ThreePA ThreePP
- Deleted DNP data rows from table
- Saved data as Curry3PP_1415.csv

Step 4 - Analyse the data

Using R and R Studio

R is a data analysis package that is powerful and R Studio is a good GUI that helps you use it.

Downloaded and installed R for windows

- https://cran.r-project.org/

Downloaded and installed RStudio for windows

- https://www.rstudio.com/products/rstudio/download/

Using Microsoft Excel / Google Sheets

Though it's at times harder to run multiple simulations of data in Excel than R, it's an easier tool to use and generally already installed.

All of the instructions (to the best of my knowledge) could be completed using either Excel or Sheets interchangeably.

Process for analysing data …

In R you can create a .r file that contains code that enables you to answer questions.

Here I want to load my CSV file into data and create an array that holds Curry’s average 3 Points Percentage in all games and his average when he’s had a bad game (<0.20)

```
data = read.csv("Curry3PP_1415.csv", header=TRUE)
attach(data)

steph3low_pct = array()
steph3low_pct = steph3low_pct[!is.na(steph3low_pct)]
steph3avg_pct = array()
steph3avg_pct = steph3avg_pct[!is.na(steph3avg_pct)]
```

I first wanted to know what Curry’s average 3 Points percentage was for every game he played during the season.

I can find this using an AVERAGE function on the entire 3 Points Percentage column

Process for analysing data ...

I then store values in these arrays using code - most notably a for loop, then calculate and display the average (mean) for that array.

```
steph3avg_pct = c(steph3avg_pct,data$ThreePP[1])
for (i in 2:80) {
  steph3avg_pct = c(steph3avg_pct,data$ThreePP[i])
    if (data$ThreePP[i-1] <= .20)
            {steph3low_pct =
c(steph3low_pct,data$ThreePP[i])}}

mean(steph3avg_pct)
mean(steph3low_pct)
```

I then wanted to know what Curry's average 3 Points percentage was for each game AFTER he'd had a bad one.

I used a conditional format to highlight games where there were <0.20 3 point percentage

I then used an IF statement in a new column to only display a percentage value if the previous game was <0.20 3 point percentage

I then could use an AVERAGE statement on values in that new column to find out what Curry scores the match after a 0.20 game

Process for analysing data ...

The process determines the average scores for Curry both for all games, and the game after he has a bad one:

All games: 0.435

After bad one: 0.424

The cells which contain the AVERAGE formulas should provide the average scores for Curry both for all games and the game after he has had a bad one:

All games: 0.435

After bad one: 0.424

Step 5 - Evaluating your hypothesis ...

Hypothesis - assume first one chosen:

Golden State Warriors will return to form and win the series because Steph Curry should be able to average a 0.40 3 point percentage after his two poor shooting games against the Memphis Grizzlies in the 2014-15 NBA playoffs.

Curry averages a 0.435 3 point percentage in his 2014-15 matches. So he is expected to return to his average and therefore proves the hypothesis true (or at least more likely)

Curry also averages a 0.424 3 point percentage after having one bad game using his 2014-15 matches. He is expected to return to this average and therefore proves the hypothesis true (or at least more likely)

Step 5 - Evaluating your hypothesis ...

Hypothesis - assume first one chosen:

Golden State Warriors will return to form and win the series because Steph Curry should be able to average a 0.40 3 point percentage after his two poor shooting games against the Memphis Grizzlies in the 2014-15 NBA playoffs.

Curry also averages a 0.422 3 point percentage in his 2014-15 regular season matches against Memphis. So he is expected to return to his average and therefore proves the hypothesis true (or at least more likely)*

*This analysis can be completed in Excel using an AVERAGEIF formula

Regression to the mean

The concept that Curry will bounce back and return to form suggests that over the long term his statistics will return to his average. This is known as “regression to the mean”.

A way to question the simplicity of this analysis is by:

- checking whether his 80 games in the season is a large enough sample
- is there more reasons behind his low shooting percentage in his last two playoff games (less minutes, more shots instead of passing, more assists instead of shooting)?
- does he have a historically (> 1 season) bad record against Memphis (struggles vs one team) or in playoffs in general (struggles under pressure)?

Step 6 - Present the Conclusion

Using data that we have analysed so far, within the scope of our investigation time period, we present the conclusion that:

Golden State Warriors will return to form and win the series because Steph Curry should be able to average a 0.40 3 point percentage after his two poor shooting games against the Memphis Grizzlies in the 2014-15 NBA playoffs.

In the next 3 playoff games against the Memphis Grizzlies, Curry shoots at 0.444, 0.462 and 0.615 for 3 point percentages and the Golden State Warriors win by more than 10 points in each game.

The rest is history

Golden State - NBA Champions
2014-15