Predicting Champions

Using season stats to predict sports champions!

Daniel Bolella, David Fonorow, Benjamin Rose & Brian Donohue

Can we predict the next world Champion?

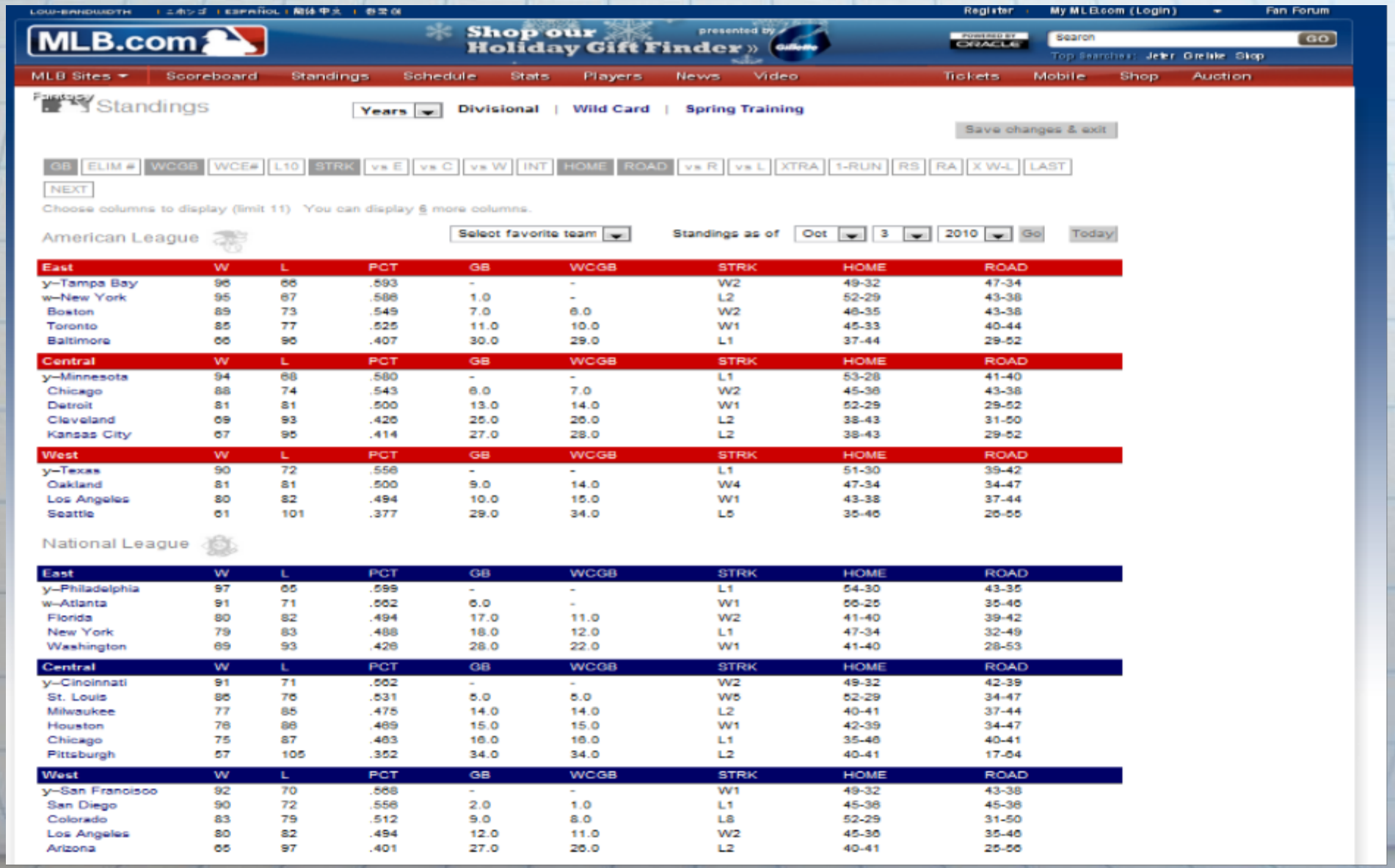
- This study was designed as a test of intermediate statistical methods and how they apply when attempting to define a seemingly chaotic system - professional sports.
- We gathered sports data for hockey, basketball, football, and baseball on a seasonal basis and attempted to identify any significant patterns within those data sets.
- Patterns identified, we expanded by attempting to predict the winning team in each sport for the current sports year.
- To assess the accuracy of our algorithms, we attempt to predict the winning team at 25%, 50%, 75%, and before the playoffs of each sport.

Our Approach

- Attempting to predict sporting event outcomes is not a new area of research
- Much money can be gained from doing so properly.
- Year after year in many sports, the same teams make it into the playoffs.
- Our goal of the study is to determine what regular season factors have influenced playoff performance in the few years prior.
- The interpreted results will be used to predict the winner of the playoffs in each sport this year.

Baseball- America's Favorite Pasttime

Gathered data from the official MLB website:



MLB.com — Shop our Holiday Gift Finder — MLB Sites | Scoreboard | Standings | Schedule | Stats | Players | News | Video | Tickets | Mobile | Shop | Auction

Standings — Years | Divisional | Wild Card | Spring Training

GB | ELIM # | WCGB | WCE# | L10 | STRK | vs E | vs C | vs W | INT | HOME | ROAD | vs R | vs L | XTRA | 1-RUN | RS | RA | X W-L | LAST | NEXT

Choose columns to display (limit 11) You can display 6 more columns.

American League — Select favorite team — Standings as of Oct 3 2010 Go Today

East	W	L	PCT	GB	WCGB	STRK	HOME	ROAD
y–Tampa Bay	96	66	.593	-	-	W2	49-32	47-34
w–New York	95	67	.586	1.0	-	L2	52-29	43-38
Boston	89	73	.549	7.0	6.0	W2	46-35	43-38
Toronto	85	77	.525	11.0	10.0	W1	45-33	40-44
Baltimore	66	96	.407	30.0	29.0	L1	37-44	29-52

Central	W	L	PCT	GB	WCGB	STRK	HOME	ROAD
y–Minnesota	94	68	.580	-	-	L1	53-28	41-40
Chicago	88	74	.543	6.0	7.0	W2	45-36	43-38
Detroit	81	81	.500	13.0	14.0	W1	52-29	29-52
Cleveland	69	93	.426	25.0	26.0	L2	38-43	31-50
Kansas City	67	95	.414	27.0	28.0	L2	38-43	29-52

West	W	L	PCT	GB	WCGB	STRK	HOME	ROAD
y–Texas	90	72	.556	-	-	L1	51-30	39-42
Oakland	81	81	.500	9.0	14.0	W4	47-34	34-47
Los Angeles	80	82	.494	10.0	15.0	W1	43-38	37-44
Seattle	61	101	.377	29.0	34.0	L5	35-46	26-55

National League

East	W	L	PCT	GB	WCGB	STRK	HOME	ROAD
y–Philadelphia	97	65	.599	-	-	L1	54-30	43-35
w–Atlanta	91	71	.562	6.0	-	W1	56-25	35-46
Florida	80	82	.494	17.0	11.0	W2	41-40	39-42
New York	79	83	.488	18.0	12.0	L1	47-34	32-49
Washington	69	93	.426	28.0	22.0	W1	41-40	28-53

Central	W	L	PCT	GB	WCGB	STRK	HOME	ROAD
y–Cincinnati	91	71	.562	-	-	W2	49-32	42-39
St. Louis	86	76	.531	5.0	5.0	W5	52-29	34-47
Milwaukee	77	85	.475	14.0	14.0	L2	40-41	37-44
Houston	76	86	.469	15.0	15.0	W1	42-39	34-47
Chicago	75	87	.463	16.0	16.0	L1	35-46	40-41
Pittsburgh	57	105	.352	34.0	34.0	L2	40-41	17-64

West	W	L	PCT	GB	WCGB	STRK	HOME	ROAD
y–San Francisco	92	70	.568	-	-	W1	49-32	43-38
San Diego	90	72	.556	2.0	1.0	L1	45-36	45-36
Colorado	83	79	.512	9.0	8.0	L8	52-29	31-50
Los Angeles	80	82	.494	12.0	11.0	W2	45-36	35-46
Arizona	65	97	.401	27.0	26.0	L2	40-41	25-56

Baseball- America's Favorite Past-time

Took the win-lose data for every quarter of each season of the past 5 years.

Year	Opening Day	1st Quarter	All-Star Break	3rd Quarter	End of Season
2006	04/02	05/22	07/11	08/22	10/02
2007	04/01	05/21	07/10	08/21	10/02
2008	03/30	05/23	07/15	08/23	09/30
2009	04/13	05/29	07/14	08/25	10/06
2010	04/05	05/25	07/13	08/24	10/05

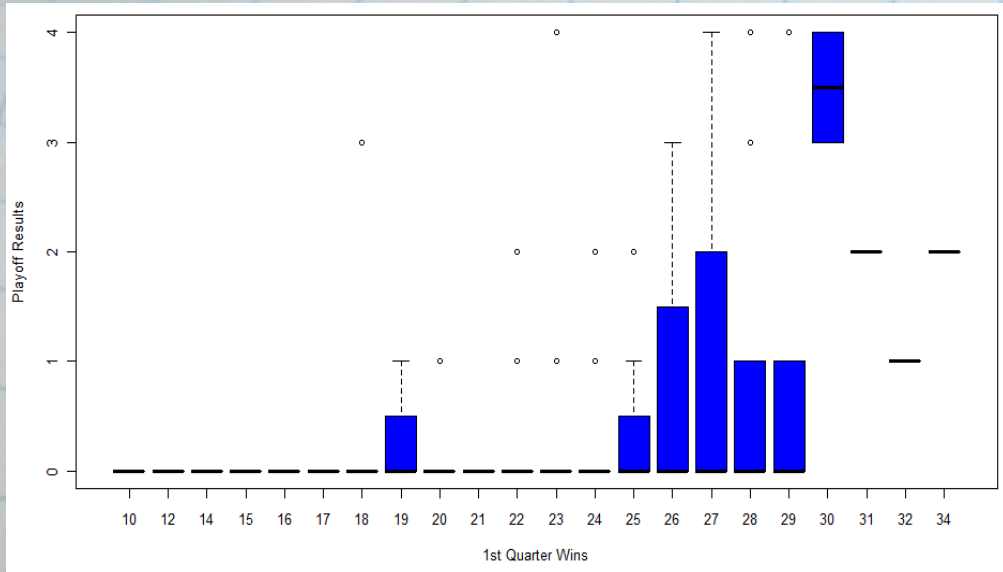
Baseball- America's Favorite Past-time

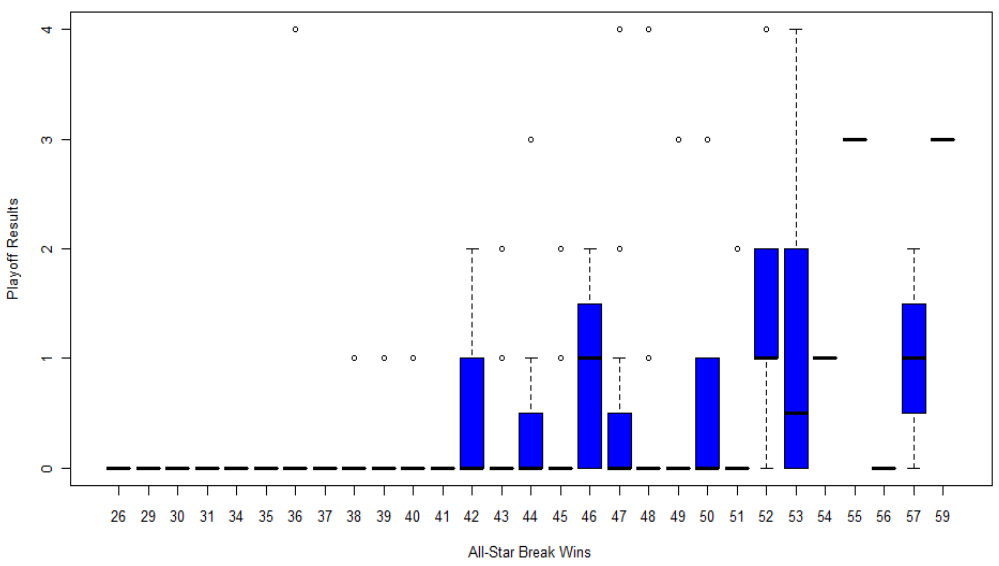
My Prediction:

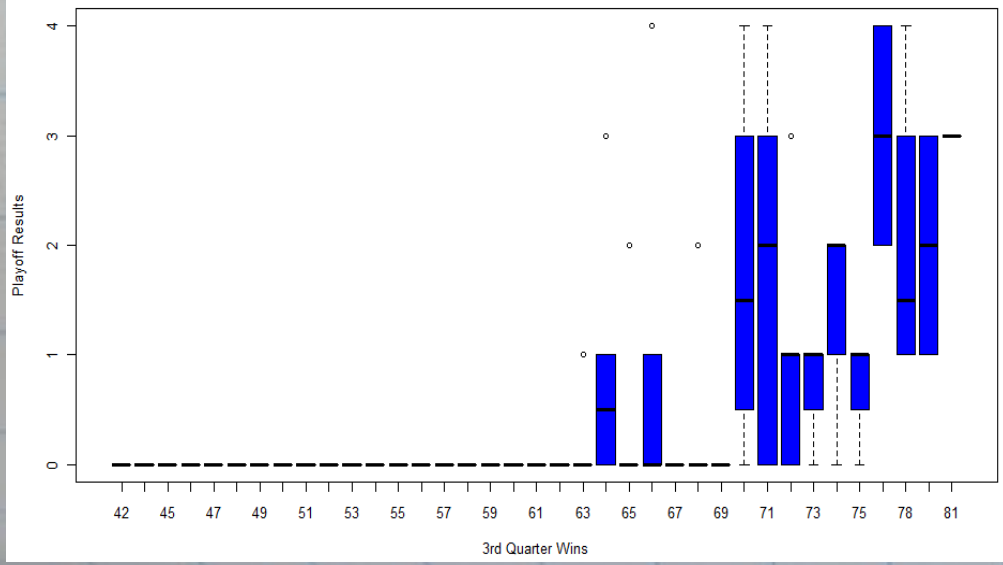
- There will be a very small relation between the standings at a certain point during the season and who will make it to the playoffs.
- Unless it is half-way through the season and the team is really far behind first place, then I feel comebacks and slumps are very possible and can change the outcome of divisional champs.

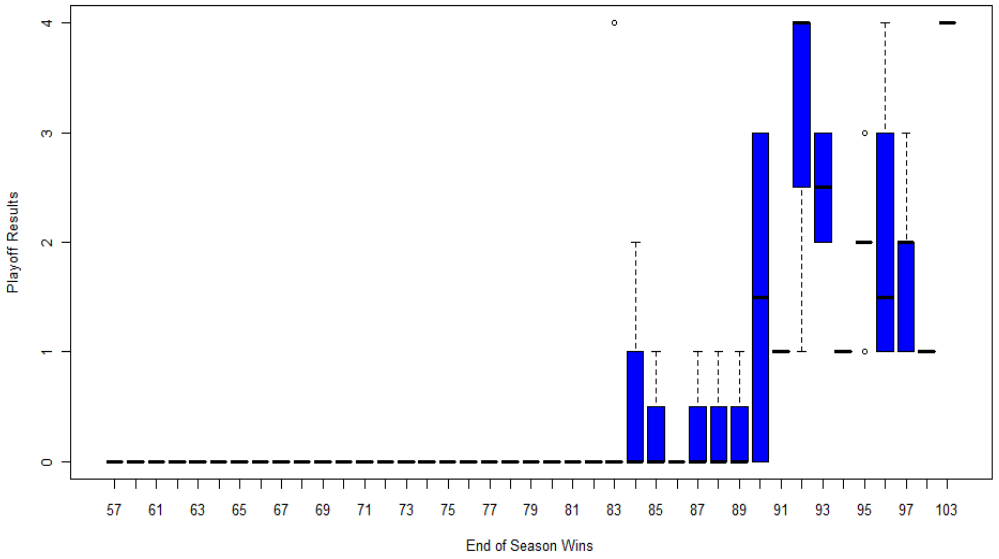
Baseball- America's Favorite Past-time

Comparison of Wins vs. Playoff Results









Baseball- America's Favorite Past-time

Results:
A Multi-Variable Anova revealed that the most significant factors were 1st Quarter, 3rd Quarter, and End Season results, as well as 1st & 3rd combined, and 3rd and EOS combined.

```
> anova(lm(PLAYOFF ~ W1 * W2 * W3 * W4))
Analysis of Variance Table

Response: PLAYOFF
             Df Sum Sq Mean Sq F value    Pr(>F)
W1            1 20.716 20.7155 37.0292 1.146e-08 ***
W2            1  2.945  2.9450  5.2642 0.0233239 *
W3            1 19.368 19.3681 34.6207 3.039e-08 ***
W4            1  8.208  8.2077 14.6713 0.0001958 ***
W1:W2         1  2.748  2.7483  4.9126 0.0283487 *
W1:W3         1  6.361  6.3615 11.3712 0.0009752 ***
W2:W3         1  0.732  0.7315  1.3076 0.2548614
W1:W4         1  1.433  1.4325  2.5607 0.1119066
W2:W4         1  0.079  0.0791  0.1413 0.7075792
W3:W4         1  7.313  7.3127 13.0715 0.0004233 ***
W1:W2:W3      1  0.017  0.0168  0.0300 0.8627069
W1:W2:W4      1  0.483  0.4832  0.8637 0.3543700
W1:W3:W4      1  1.223  1.2232  2.1866 0.1415671
W2:W3:W4      1  0.352  0.3521  0.6293 0.4290112
W1:W2:W3:W4   1  0.558  0.5584  0.9981 0.3195742
Residuals   134 74.965  0.5594
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Football (the real one)

Goal: By analyzing statistics about teams over the last ten years, try to predict the winner of the super bowl this season.

Prediction: Algorithm will predict the team with the highest record has the best chance of winning. Little to no influence based on other factors.

Football - Data Collection

Data was collected from www.pro-football-reference.com

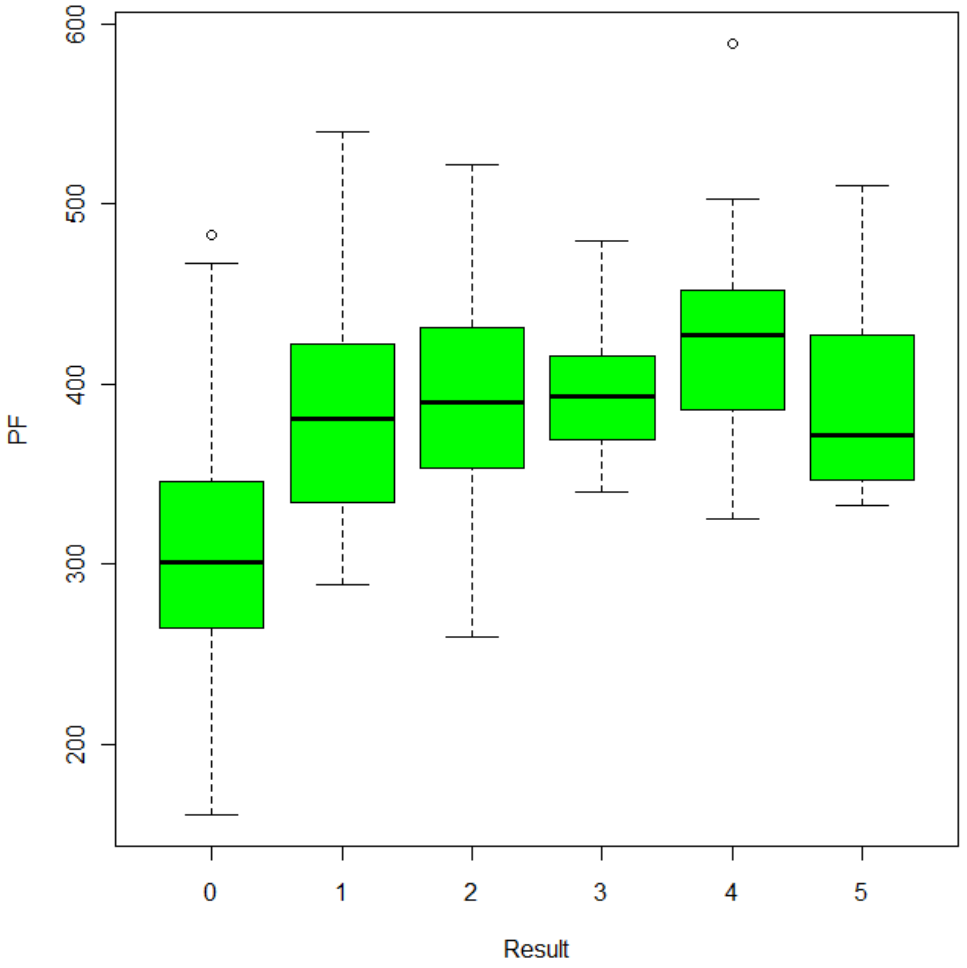
Team Games & Schedule

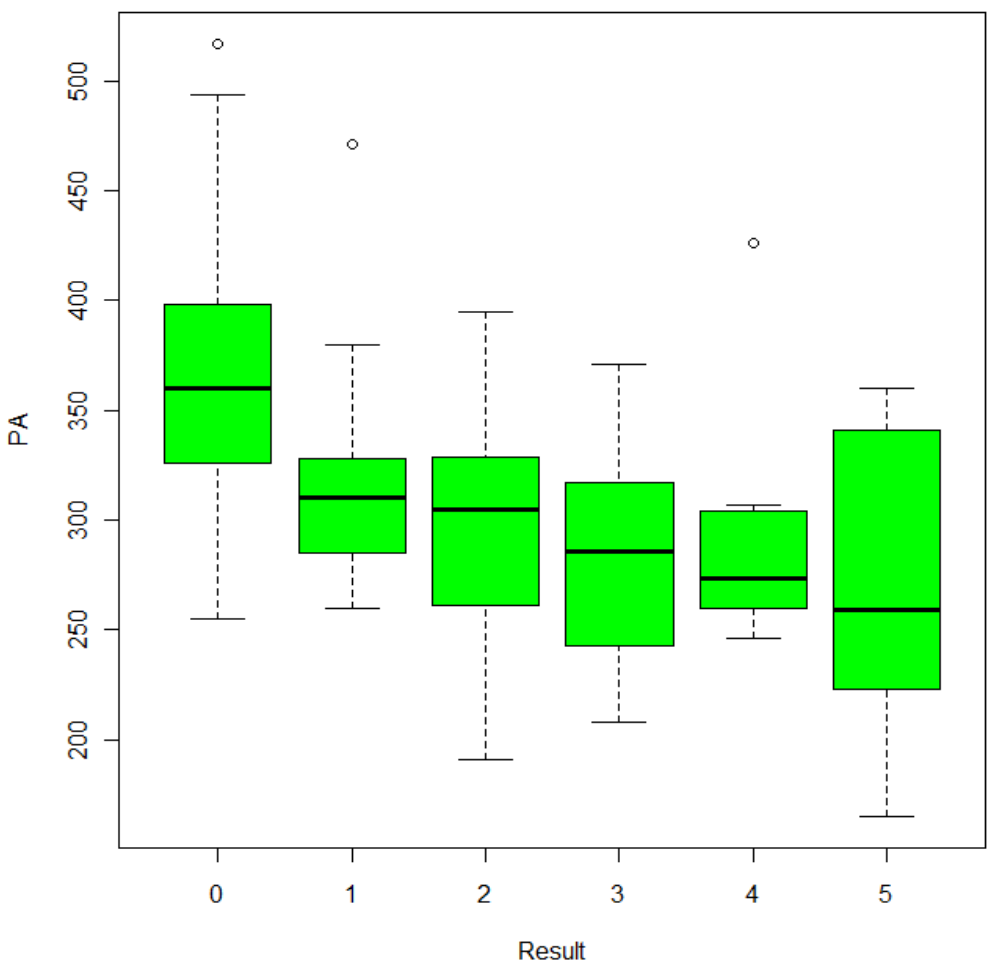
Glossary · SHARE · CSV · PRE · LINK · More Tools

								Score		Offense					Defense				
Week	Day	Date			OT	Rec		Opp	Tm	Opp	1stD	TotYd	PassY	RushY	TO	1stD	TotYd	PassY	RushY	TO
1	Sun	September 3	boxscore	L		0-1	@	New York Giants	16	21	19	355	312	43	4	20	395	172	223	2
2	Sun	September 10	boxscore	W		1-1		Dallas Cowboys	32	31	21	322	224	98	1	17	330	240	90	2
3								**Bye Week**												
4	Sun	September 24	boxscore	L		1-2		Green Bay Packers	3	29	12	209	181	28	4	23	455	279	176	1
5	Sun	October 1	boxscore	L		1-3	@	San Francisco 49ers	20	27	16	365	239	126	2	18	345	215	130	1
6	Sun	October 8	boxscore	W		2-3		Cleveland Browns	29	21	20	315	169	146	2	12	240	136	104	
7	Sun	October 15	boxscore	L		2-4		Philadelphia Eagles	14	33	17	305	207	98	3	28	391	219	172	
8	Sun	October 22	boxscore	L		2-5	@	Dallas Cowboys	7	48	15	276	167	109	3	23	347	147	200	
9	Sun	October 29	boxscore	L		2-6		New Orleans Saints	10	21	24	394	269	125	4	17	247	158	89	1
10	Sun	November 5	boxscore	W		3-6		Washington Redskins	16	15	11	178	133	45	2	27	422	258	164	3
11	Sun	November 12	boxscore	L		3-7	@	Minnesota Vikings	14	31	14	249	198	51	2	29	460	302	158	2
12	Sun	November 19	boxscore	L		3-8	@	Philadelphia Eagles	9	34	12	210	156	54	1	21	338	204	134	2
13	Sun	November 26	boxscore	L		3-9		New York Giants	7	31	14	267	179	88	4	25	371	225	146	1
14	Sun	December 3	boxscore	L		3-10	@	Cincinnati Bengals	13	24	15	340	268	72	2	27	398	106	292	1
15	Sun	December 10	boxscore	L		3-11	@	Jacksonville Jaguars	10	44	11	189	149	40	1	25	469	255	214	
16	Sun	December 17	boxscore	L		3-12		Baltimore Ravens	7	13	18	309	258	51	4	14	214	37	177	2
17	Sun	December 24	boxscore	L		3-13	@	Washington Redskins	3	20	14	245	141	104	5	19	315	175	140	2

Football - Points For/Against

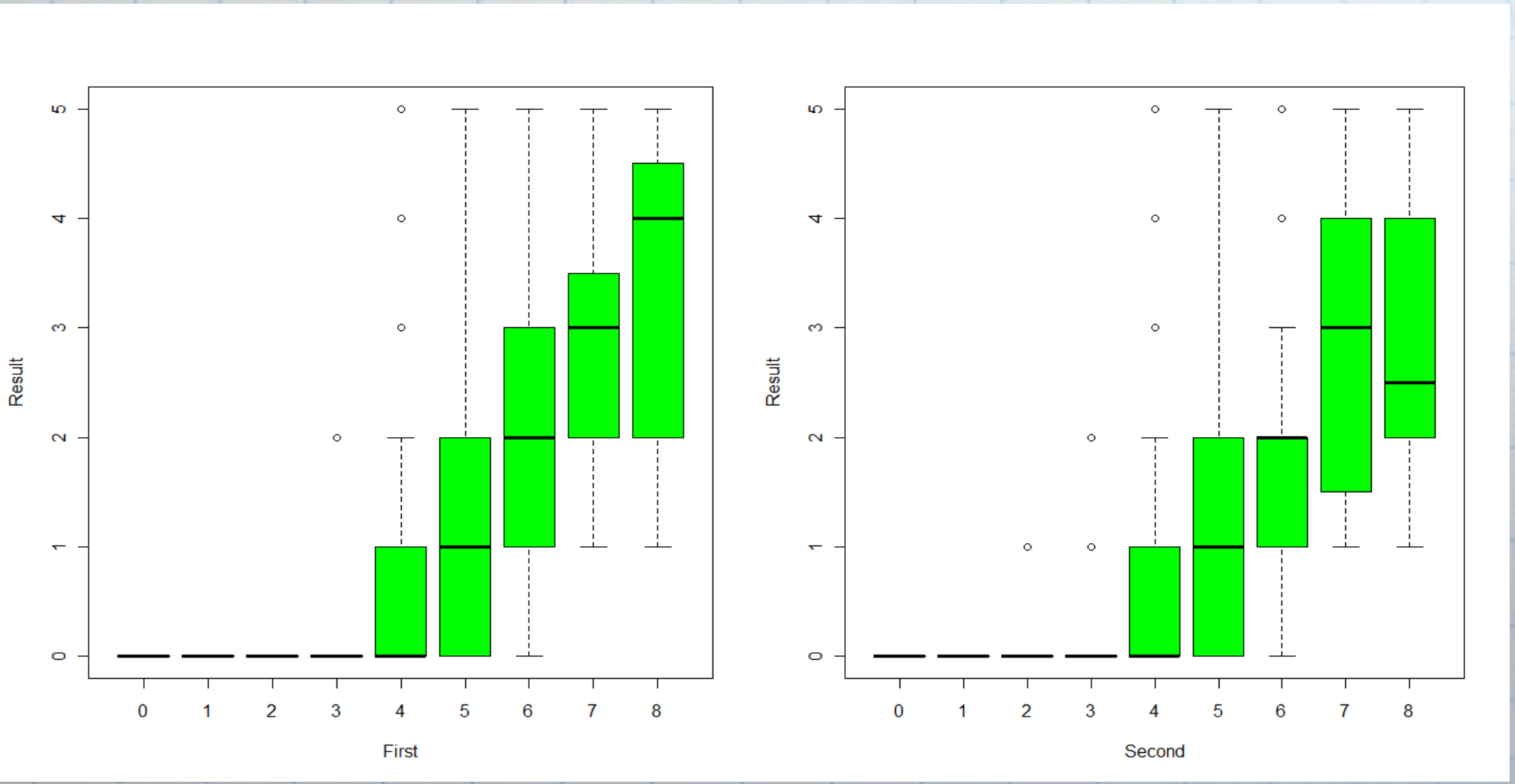
- PF and PA are significant when determining results





Football - First/Second Half Wins

- Still no surprises, more wins in either half means better results



Football - Correlation Table

Correlation(x,y)	Result
First half wins	0.61772
Second half wins	0.5945279
Total wins	0.7090561
Points for	0.5182356
Points against	-0.5287885
PF/PA difference	0.6650967

Football - multi ANOVA

- Some interesting results from multi ANOVA test

```
> anova(lm(Result ~ First * Second * PF * PA))
Analysis of Variance Table

Response: Result
                  Df  Sum Sq Mean Sq  F value    Pr(>F)
First              1 213.474 213.474 330.8517 < 2.2e-16 ***
Second             1  69.582  69.582 107.8416 < 2.2e-16 ***
PF                 1   0.227   0.227   0.3518  0.553547
PA                 1   1.440   1.440   2.2315  0.136269
First:Second       1  58.240  58.240  90.2628 < 2.2e-16 ***
First:PF           1   0.784   0.784   1.2158  0.271060
Second:PF          1   0.354   0.354   0.5481  0.459681
First:PA           1   0.499   0.499   0.7738  0.379732
Second:PA          1  10.140  10.140  15.7157 9.195e-05 ***
PF:PA              1   0.111   0.111   0.1716  0.678982
First:Second:PF    1   3.200   3.200   4.9593  0.026688 *
First:Second:PA    1   0.655   0.655   1.0159  0.314305
First:PF:PA        1   0.029   0.029   0.0447  0.832717
Second:PF:PA       1   0.526   0.526   0.8145  0.367508
First:Second:PF:PA 1   5.332   5.332   8.2631  0.004333 **
Residuals        302 194.858   0.645
```

Football - Predictions

After deriving a number of formulas, applying their significance and averaging them out, the following teams have the greatest chance of winning this year:

- New England
- Atlanta
- New Orleans
- Chicago
- Pittsburgh
- Green Bay
- New York Jets

Basketball

Goal

Using trends from the previous 10 years, predict the outcome of the championship finals in the next season by analyzing common trends at the 25%, 25%, 75%, and 100% points during the regular season.

Methodology

Statistics were collected, then loaded into the R software suite for analysis and plot creation. Further, in depth analysis was performed on this data and the plots rendered using statistical techniques discussed in this course.

Basketball

Statistics were collected from the National Basketball Association's official website: www.nba.com

Supplemental Information: dougstats.com

STATISTICS

Category: Offensive | Conference: NBA | Period: Season-to-date | Update

Team Offensive Statistics for 2010-2011

	TEAM	GAMES	PPG			FG%		3PT%		FT%		APG		
			HME	OPP	DIFF	HME	OPP	HME	OPP	HME	OPP	HME	OPP	DIFF
1	Toronto	6	109.67	101.83	+7.84	0.458	0.467	0.393	0.323	0.681	0.700	21.33	22.67	-1.34
2	Minnesota	6	106.33	99.17	+7.16	0.438	0.411	0.467	0.290	0.756	0.736	21.33	20.00	+1.33
3	Golden State	4	106.00	101.75	+4.25	0.471	0.430	0.366	0.407	0.829	0.738	22.00	21.75	+0.25
4	Portland	5	105.60	101.80	+3.80	0.477	0.486	0.475	0.414	0.786	0.688	21.80	20.60	+1.20
5	Denver	5	104.00	104.40	-0.40	0.435	0.438	0.345	0.331	0.765	0.796	20.40	25.80	-5.40
6	Utah	6	104.00	96.33	+7.67	0.492	0.450	0.356	0.362	0.752	0.747	22.17	18.67	+3.50
7	Orlando	5	103.20	79.00	+24.20	0.499	0.356	0.385	0.320	0.730	0.795	21.20	15.20	+6.00
8	Memphis	6	102.67	94.00	+8.67	0.492	0.413	0.313	0.373	0.730	0.692	17.00	20.83	-3.83
9	New York	5	100.40	102.00	-1.60	0.437	0.453	0.323	0.353	0.747	0.776	21.00	18.20	+2.80
10	Sacramento	5	98.80	101.00	-2.20	0.428	0.458	0.359	0.237	0.723	0.721	18.20	19.80	-1.60
11	Boston	7	98.57	92.00	+6.57	0.482	0.400	0.430	0.358	0.766	0.736	19.86	18.43	+1.43
12	L.A. Clippers	7	97.43	105.86	-8.43	0.439	0.459	0.331	0.410	0.723	0.762	21.86	21.57	+0.29
13	Houston	6	97.17	90.17	+7.00	0.450	0.425	0.340	0.295	0.773	0.772	19.00	18.50	+0.50
14	Oklahoma City	4	96.50	100.25	-3.75	0.438	0.500	0.444	0.344	0.761	0.719	19.00	21.25	-2.25
15	Detroit	6	96.00	99.50	-3.50	0.446	0.475	0.400	0.406	0.773	0.701	21.33	22.17	-0.84
16	Washington	6	95.83	94.00	+1.83	0.449	0.436	0.380	0.344	0.651	0.712	21.33	22.00	-0.67
17	Milwaukee	6	95.00	94.83	+0.17	0.415	0.453	0.350	0.329	0.836	0.702	17.33	19.67	-2.34
18	L.A. Lakers	5	94.80	98.40	-3.60	0.421	0.450	0.250	0.361	0.707	0.761	22.20	21.80	+0.40
19	Phoenix	6	94.67	110.33	-15.66	0.435	0.472	0.254	0.341	0.673	0.769	19.67	23.50	-3.83
20	Philadelphia	5	93.80	99.40	-5.60	0.384	0.466	0.247	0.424	0.809	0.712	19.20	20.00	-0.80
21	Dallas	6	93.50	94.00	-0.50	0.420	0.420	0.267	0.324	0.698	0.745	21.83	22.00	-0.17
22	New Jersey	6	93.17	89.50	+3.67	0.440	0.391	0.402	0.330	0.699	0.765	19.33	17.17	+2.16
23	Atlanta	4	93.00	97.50	-4.50	0.472	0.472	0.369	0.426	0.622	0.798	21.25	20.75	+0.50
24	San Antonio	5	93.00	90.60	+2.40	0.410	0.417	0.321	0.362	0.695	0.722	22.60	18.80	+3.80
25	Indiana	5	92.40	98.00	-5.60	0.401	0.426	0.313	0.317	0.718	0.688	17.80	18.00	-0.20
26	Chicago	6	92.33	96.67	-4.34	0.434	0.474	0.356	0.337	0.679	0.749	22.00	20.50	+1.50
27	Cleveland	6	91.17	85.17	+6.00	0.416	0.393	0.325	0.323	0.790	0.653	18.83	21.17	-2.34
28	Miami	5	90.60	90.00	+0.60	0.450	0.403	0.298	0.415	0.767	0.811	20.40	16.80	+3.60
29	New Orleans	5	86.60	98.60	-12.00	0.374	0.468	0.376	0.398	0.796	0.759	17.40	20.20	-2.80
30	Charlotte	5	82.60	90.00	-7.40	0.405	0.391	0.373	0.385	0.718	0.750	19.20	17.80	+1.40

***FG%:** Field Goal Percentage ***3PT%:** Three-Point FG Percentage ***FT%:** Free Throw Percentage
***PPG:** Points Per Game ***APG:** Assists Per Game

Basketball

Fields collected and considered

- team name
- games won
- games lost
- total minutes played
- field goals made
- field goals attempted
- threes made
- threes attempted
- free throws made
- free throws attempted
- offensive rebounds
- total rebounds
- assists
- steals
- turnovers
- blocks
- personal fouls
- technicals
- ejections
- flagrant fouls
- total points
- championship rank score*

* = Score rank was determined by downloading rank data from the NBA brackets, then matching simple text filters to find the teams logo HTML section on the webpage. Those winning the championship earned a '5' (their logo advanced all the way across the rendered bracket), those who did not qualify a '0' (their logo did not appear on the page).

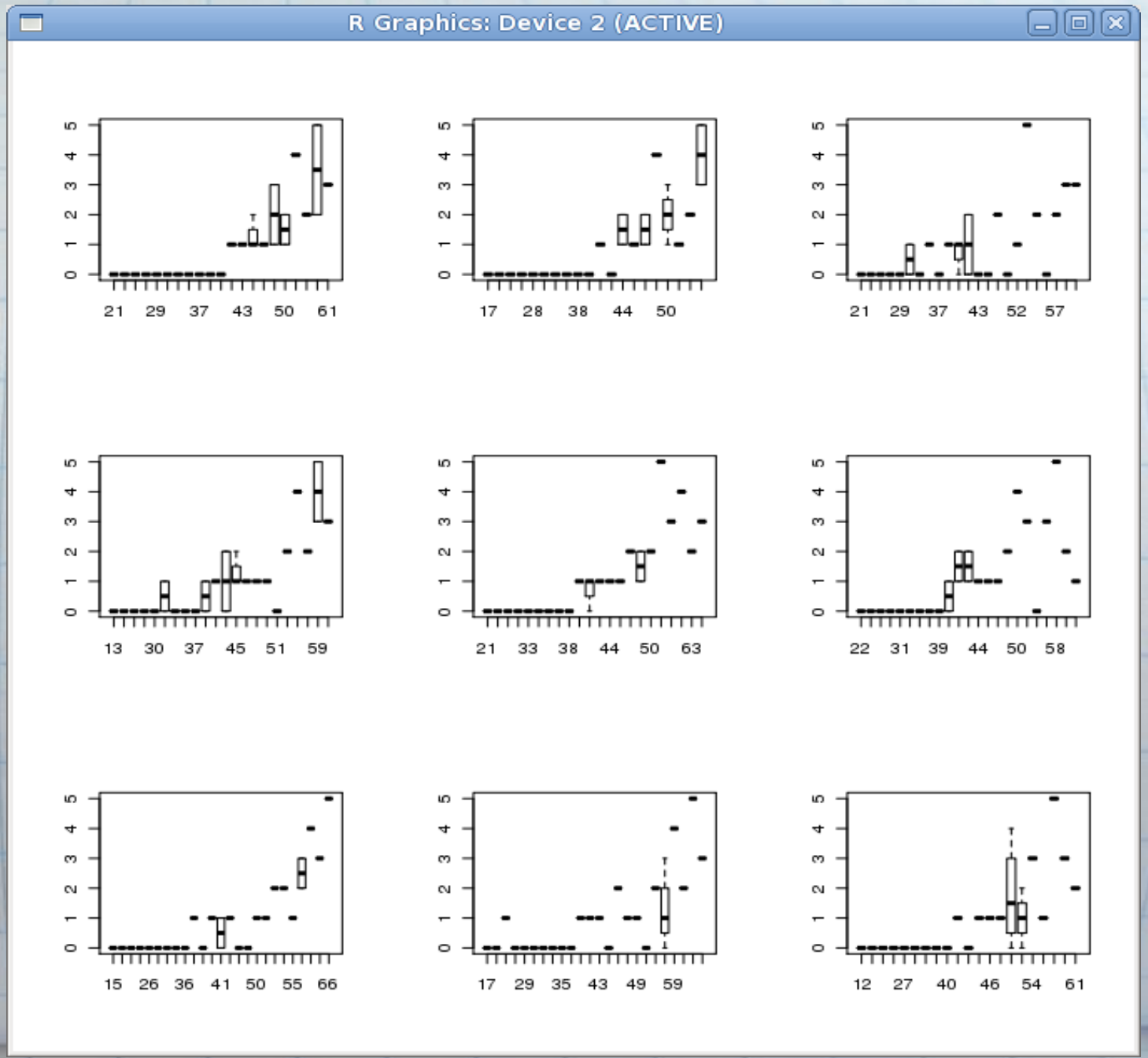
Basketball - Analysis of wins per season

Trends through NBA history:

Rarely does the team with the most wins also bring home the championship.

In fact, if you win the most games, you probably won't make it past the final four!

2001 omitted.



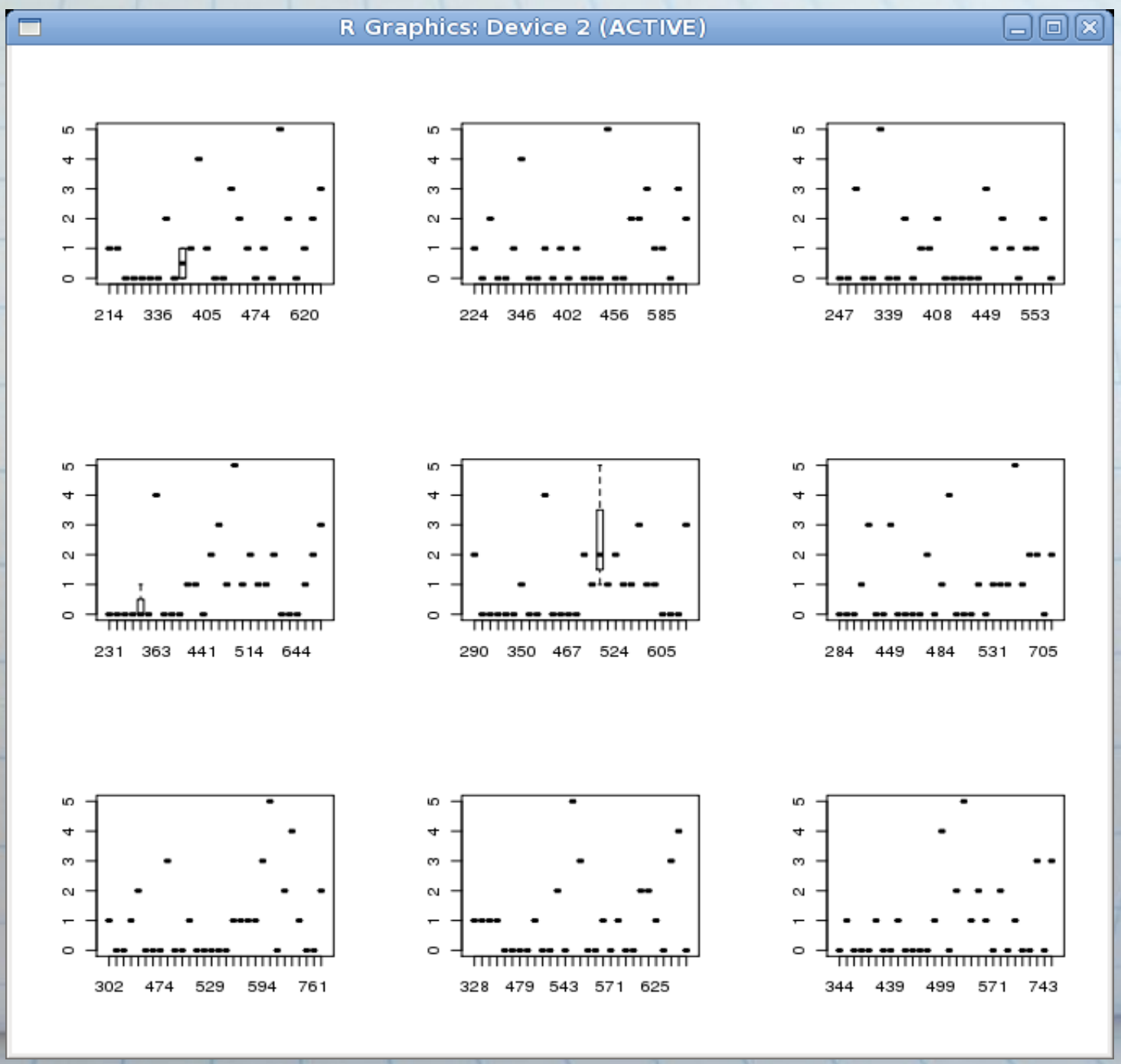
Basketball Skill - Analysis of 3-pointers!

Trends through NBA history:

If your team is constantly shooting field-goals, don't expect to make it into the final four.

If your team has an accumulation of field-goals in the 400-500 range, you're good! (except that one time)

2001 omitted.



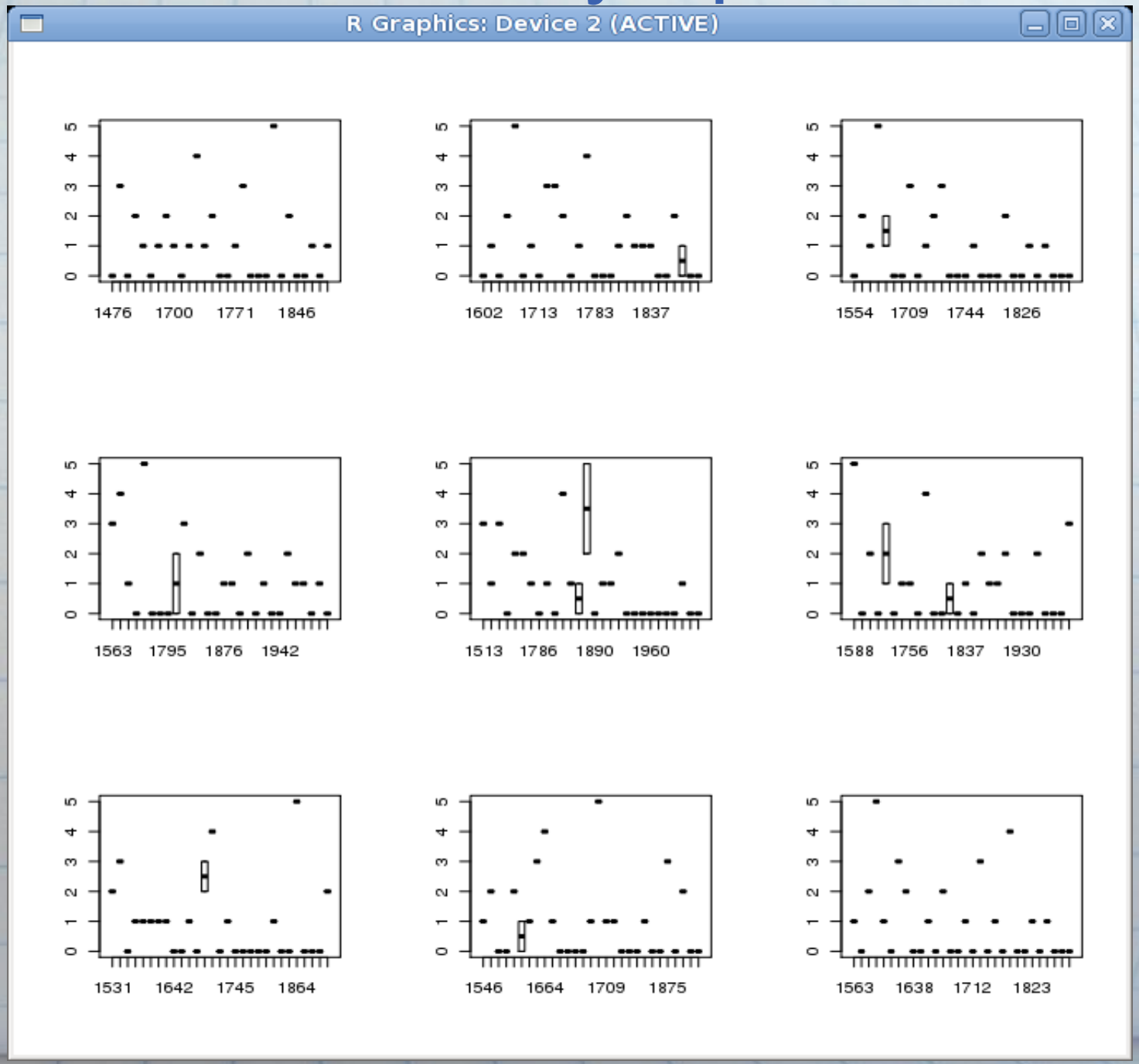
Hugatreeball - The Friendly Sport

Trends through NBA history:

Each year, a strikingly similar number of strikings occur.

Playing nicely with others is often rewarded, especially in 2006.

2001 omitted.



Basketball - Simple multivariate anova calculations

```
> anova(lm(season10$final ~ season10$won * season10$pf * season10$X3m))
Analysis of Variance Table

Response: season10$final
                                Df  Sum Sq Mean Sq F value    Pr(>F)
season10$won                     1 22.3425 22.3425 20.1772 0.0001814 ***
season10$pf                      1  0.9059  0.9059  0.8181 0.3755448
season10$X3m                     1  0.4188  0.4188  0.3782 0.5448787
season10$won:season10$pf         1  0.5130  0.5130  0.4633 0.5032056
season10$won:season10$X3m        1  2.6129  2.6129  2.3597 0.1387680
season10$pf:season10$X3m         1  0.3783  0.3783  0.3416 0.5648424
season10$won:season10$pf:season10$X3m  1  0.3345  0.3345  0.3021 0.5881213
Residuals                       22 24.3609  1.1073
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(lm(season09$final ~ season09$won * season09$pf * season09$X3m))
Analysis of Variance Table

Response: season09$final
                                Df  Sum Sq Mean Sq F value    Pr(>F)
season09$won                     1 26.8853 26.8853 30.5097 1.498e-05 ***
season09$pf                      1  0.0013  0.0013  0.0014   0.96998
season09$X3m                     1  0.0337  0.0337  0.0382   0.84681
season09$won:season09$pf         1  0.0537  0.0537  0.0609   0.80730
season09$won:season09$X3m        1  4.3525  4.3525  4.9393   0.03685 *
season09$pf:season09$X3m         1  0.0559  0.0559  0.0635   0.80342
season09$won:season09$pf:season09$X3m  1  0.1976  0.1976  0.2243   0.64047
Residuals                       22 19.3866  0.8812
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Basketball - Complex multivariate anova calculations

To analyze what factor - or combination of factors - influence the statistical likelihood of basketball teams winning, lengthy anova(lm()) calculations were performed on a quad-core Xeon server. The results for the most recent 4 seasons, 2007 - 2010, have returned. The next batch, 2003 - 2006, are currently processing. Trends can still be heavily analyzed from the initial findings, however, justifying the CPU time (if they all finish in time).

When the more complex calculations are complete, it will expose how each variable and set of variables relates to post-season performance, and hopefully, to each-other in a grand-scheme picture representative of trends evident in the last 10 years of NBA history.

Basketball - Complex multivariate anova calculations

2010

season10$won	1	22.3425	22.3425	14.0720	0.002148 **
season10$min	1	1.1557	1.1557	0.7279	0.407935
season10$fgm	1	0.2348	0.2348	0.1479	0.706329
season10$fga	1	0.4207	0.4207	0.2650	0.614737
season10$X3m	1	0.0652	0.0652	0.0411	0.842350
season10$X3a	1	0.1635	0.1635	0.1030	0.753035
season10$ftm	1	1.4186	1.4186	0.8935	0.360563
season10$fta	1	0.3556	0.3556	0.2240	0.643305
season10$or	1	0.1162	0.1162	0.0732	0.790698
season10$tr	1	0.2214	0.2214	0.1394	0.714435
season10$as	1	1.2391	1.2391	0.7804	0.391928
season10$st	1	0.3637	0.3637	0.2291	0.639597
season10$to	1	0.9983	0.9983	0.6287	0.441051
season10$bk	1	0.2099	0.2099	0.1322	0.721602
season10$pf	1	0.3332	0.3332	0.2099	0.653895
Residuals	14	22.2282	1.5877		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

2009

season09$won	1	26.8853	26.8853	44.7964	1.022e-05 ***
season09$min	1	1.4280	1.4280	2.3793	0.14525
season09$fgm	1	0.2686	0.2686	0.4476	0.51436
season09$fga	1	1.6576	1.6576	2.7619	0.11875
season09$X3m	1	0.7190	0.7190	1.1979	0.29221
season09$X3a	1	0.0333	0.0333	0.0555	0.81721
season09$ftm	1	1.3229	1.3229	2.2041	0.15980
season09$fta	1	4.2206	4.2206	7.0323	0.01896 *
season09$or	1	3.5762	3.5762	5.9586	0.02853 *
season09$tr	1	0.4635	0.4635	0.7723	0.39434
season09$as	1	0.0917	0.0917	0.1527	0.70183
season09$st	1	1.4345	1.4345	2.3902	0.14440
season09$to	1	0.3838	0.3838	0.6395	0.43725
season09$bk	1	0.0247	0.0247	0.0412	0.84201
season09$pf	1	0.0547	0.0547	0.0911	0.76717
Residuals	14	8.4023	0.6002		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

2008

season08$won	1	30.3211	30.3211	38.6891	2.237e-05
season08$min	1	1.0291	1.0291	1.3131	0.2710
season08$fgm	1	1.5642	1.5642	1.9958	0.1796
season08$fga	1	2.1697	2.1697	2.7685	0.1184
season08$X3m	1	0.0399	0.0399	0.0509	0.8247
season08$X3a	1	0.2740	0.2740	0.3496	0.5638
season08$ftm	1	0.0334	0.0334	0.0426	0.8395
season08$fta	1	0.0843	0.0843	0.1076	0.7477
season08$or	1	0.1238	0.1238	0.1580	0.6970
season08$tr	1	0.2567	0.2567	0.3275	0.5762
season08$as	1	0.2252	0.2252	0.2874	0.6003
season08$st	1	1.6636	1.6636	2.1228	0.1672
season08$to	1	0.3197	0.3197	0.4080	0.5333
season08$bk	1	0.4693	0.4693	0.5988	0.4519
season08$pf	1	1.4207	1.4207	1.8128	0.1996
Residuals	14	10.9720	0.7837		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

2007

season07$won	1	23.6528	23.6528	24.6652	0.0002070 ***
season07$min	1	0.0539	0.0539	0.0562	0.8161016
season07$fgm	1	0.1638	0.1638	0.1708	0.6856211
season07$fga	1	0.2475	0.2475	0.2581	0.6193559
season07$X3m	1	0.0402	0.0402	0.0420	0.8406507
season07$X3a	1	0.8352	0.8352	0.8710	0.3664973
season07$ftm	1	0.2931	0.2931	0.3057	0.5890738
season07$fta	1	4.2820	4.2820	4.4653	0.0530294 .
season07$or	1	0.3760	0.3760	0.3921	0.5412819
season07$tr	1	1.5450	1.5450	1.6112	0.2250259
season07$as	1	3.2189	3.2189	3.3566	0.0882894 .
season07$st	1	0.1310	0.1310	0.1367	0.7171595
season07$to	1	1.9147	1.9147	1.9967	0.1794885
season07$bk	1	0.4626	0.4626	0.4824	0.4987136
season07$pf	1	1.3579	1.3579	1.4160	0.2538520
Residuals	14	13.4254	0.9590		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Basketball - Predictions?!?

By analyzing previous trends of winners, we see that the following variables are highly important:

- You don't win the most, but rather around the 75% proficiency mark in that category.
- Your team doesn't consist of a bunch of super-stars.
- You play by the rules.

So far this season, the following teams have been the closest to matching those attributes: (but what does this *really* mean?)

- Chicago Bulls
- Indiana Pacers
- Denver Nuggets
- Phoenix Suns
- New Orleans Hornets

Hockey

Hypothesis:

- Using statistical analysis, an attempt will be made to determine which regular season data can indicate the winner of the Stanley Cup will occur on regular season data in an attempt to determine which regular season factors may indicate the winner of the Stanley Cup.
- This type of analysis is difficult to determine because the regular season performance of a team may not necessarily indicate the post season performance.
- Factors such as interaction between teams, aggressiveness of a team, and injuries may inhibit such an analysis

Hockey - Data Set

- The data set used comprises regular season data over the past five years for all teams of the NHL
- Regular season data includes: wins, goals scored, goals scored against, overtime won, overtime lost, penalties, penalties in minutes, power-play opportunities, and power-play goals
- In addition to the regular season data, the number of wins each team had in the post-season games is factored in.
- Because the number of wins for each team is slightly misleading (for instance it's possible a team makes the playoffs, but doesn't win a game), an additional factor is added that ranks the team by how far the team went in the playoffs (i.e. quarter-finals, finals, and Stanley Cup winner)

Hockey - (Naiive)Correlation in Data

Correlation(x,y)	Playoff Level	Playoff Wins
Wins Regular Season	0.6160447	0.5457638
Goals Scored	0.4544496	0.4272315
Goals Scored Against	-0.463293	-0.4061466
Overtimes Won	0.1638642	0.1059421
Overtimes Lost	-0.0872057	-0.06243047
Penalties	-0.1184132	-0.08396708
Penalties in Minutes	-0.1419841	-0.1042888
Power Play Opportunities	0.04501688	0.04478294
Power Play Goal	0.2774943	0.2793165

Hockey - Correlation in Data

Correlation(x,y)	Playoff Level	Playoff Wins
Wins Regular Season	0.1802523	0.1919077
Goals Scored	0.2874196	0.2748414
Goals Scored Against	-0.01905782	-0.03588322
Overtimes Won	-0.1248323	-0.1385355
Overtimes Lost	0.0624842	0.06598831
Penalties	0.01890814	0.02833453
Penalties in Minutes	0.03463016	0.03129029
Power Play Opportunities	0.09830155	0.0802194
Power Play Goal	0.2057550	0.2212174

Hockey - Multivariate Anova for Wins

Analysis of Variance Table

Response: Win_post

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Win_reg	1	61.11	61.11	2.8862	0.09391 .
Goals_for	1	69.34	69.34	3.2750	0.07476 .
Goals_against	1	44.65	44.65	2.1087	0.15106
OT_won	1	0.11	0.11	0.0051	0.94345
OT_lost	1	0.33	0.33	0.0156	0.90093
Penalties	1	4.83	4.83	0.2281	0.63450
PIM	1	10.99	10.99	0.5188	0.47381
PP_opp	1	11.21	11.21	0.5295	0.46933
PPG	1	16.98	16.98	0.8018	0.37371
Residuals	68	1439.76	21.17		

Hockey - Logistic Regression with Playoff Level as response

Deviance Residuals:

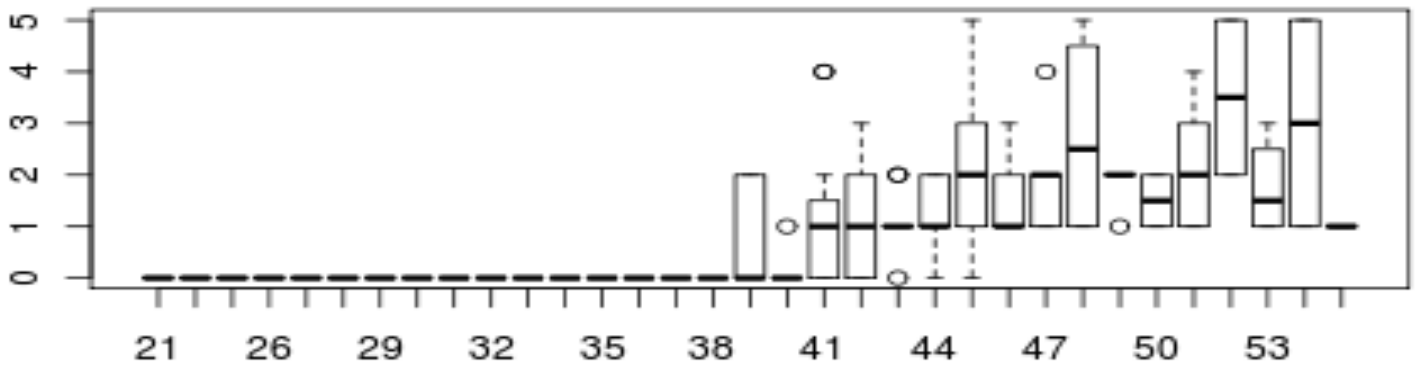
Min	1Q	Median	3Q	Max
-1.8513	-1.0022	0.4812	1.0114	1.7175

Coefficients:

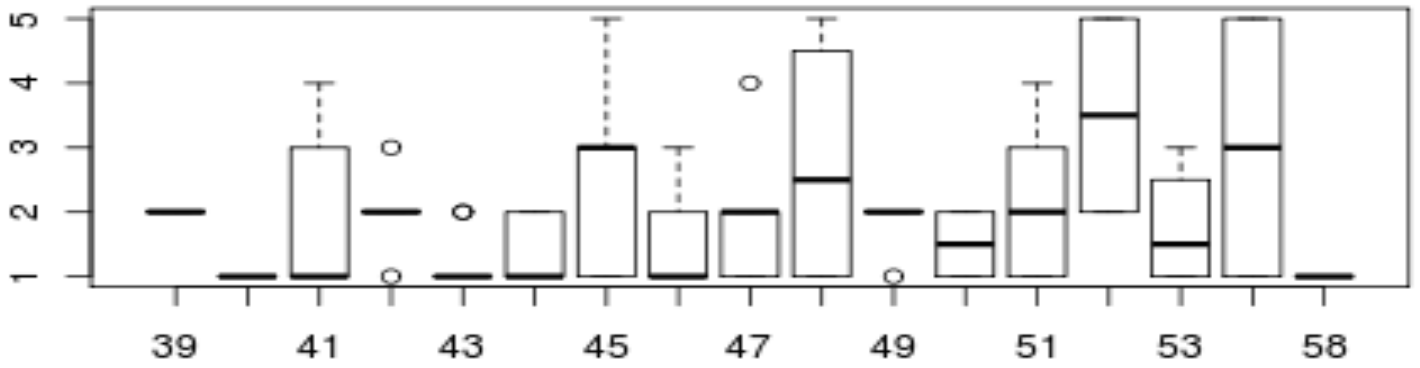
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	11.268326	7.486436	1.505	0.1323	
Win_reg	-0.257301	0.134528	-1.913	0.0558	.
Goals_for	0.045043	0.023356	1.929	0.0538	.
Goals_against	-0.055620	0.025542	-2.178	0.0294	*
OT_won	0.051313	0.090364	0.568	0.5701	
OT_lost	-0.046478	0.139037	-0.334	0.7382	
Penalties	-0.024543	0.014341	-1.711	0.0870	.
PIM	0.006732	0.004570	1.473	0.1407	
PP_opp	0.004378	0.008388	0.522	0.6017	
PPG	0.055042	0.035586	1.547	0.1219	

Hockey - Boxplots for Regular Season wins vs Playoff Levels

Full dataset (Naiive analysis):

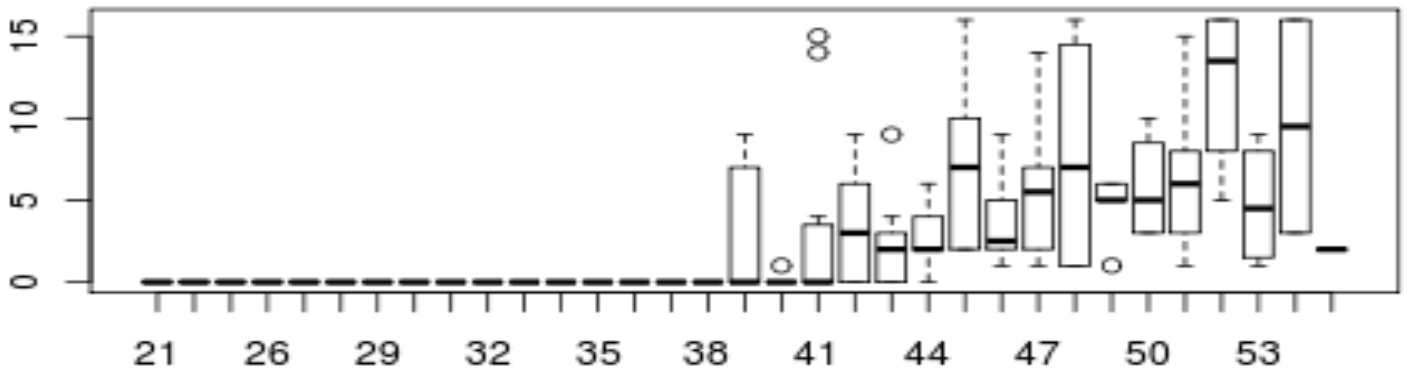


Playoff teams only:

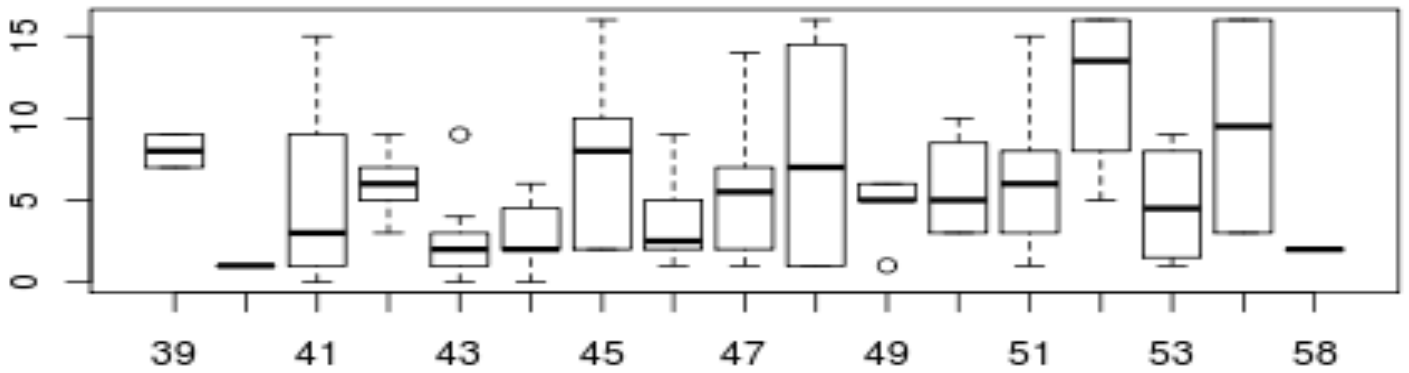


Hockey - Boxplot for Regular Season Wins vs. Postseason Wins

Full dataset (Naiive analysis):



Playoff teams only:



Hockey - Conclusions

- Analysis of the correlations, significance in Anova, and boxplot graphs indicate that given solely regular season data it is very difficult to develop a statistical model for determining the Stanley Cup winner.
- Does this assertion conform to experience?
- Last season Philadelphia Fliers were the 8th seed team in the playoffs (last place), and made it all the way to the Finals
- Also, last season Washington Capitals were first seed in the playoffs, and lost in the first round of the playoffs

Concluding Remarks

- Baseball
- Football
- Basketball
- Hockey

Any questions?