

NBA FREE AGENT CONTRACTS ANALYSIS

BY SEAN VOLPI



BACKGROUND

- NBA players are being paid more than ever.
- Will look at NBA players' current Free Agent contract's salaries, when they were initially signed and their season stats prior to the signing.
- I want to see what variables significantly impact how much a free agent gets paid when they sign their contract.



RESPONSE VARIABLE

- NBA players' current FA contract salaries.
- WILL NOT observe any other salary type (rookies, extensions, etc.)



45 EXPLANATORY VARIABLES

- The basics:
 - G, GS, MP, FG, FGA, FG%, 3P, 3PA, 3P%, 2P, 2PA, 2P%, eFG%, FT, FTA, FT%, ORB, DRB, TRB, AST, STL, BLK, TOV, PF, & PTS.
- Advanced statistics:
 - TS%, PER, 3PAr, FTr, ORB%, DRB%, TRB%, AST%, STL%, BLK%, TOV%, USG%, OWS, DWS, WS, WS/48, OBPM, DBPM, BPM, & VORP.

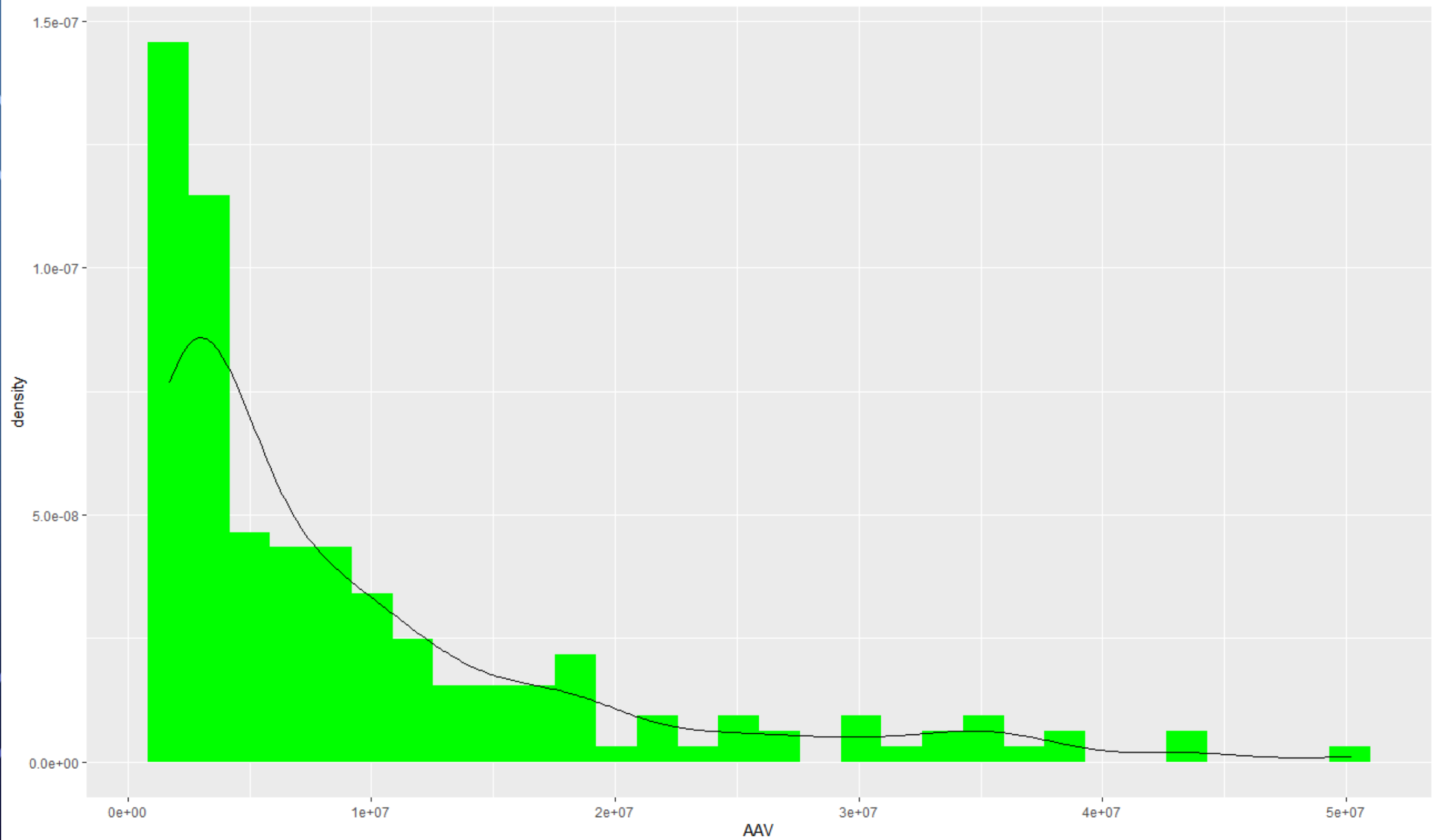


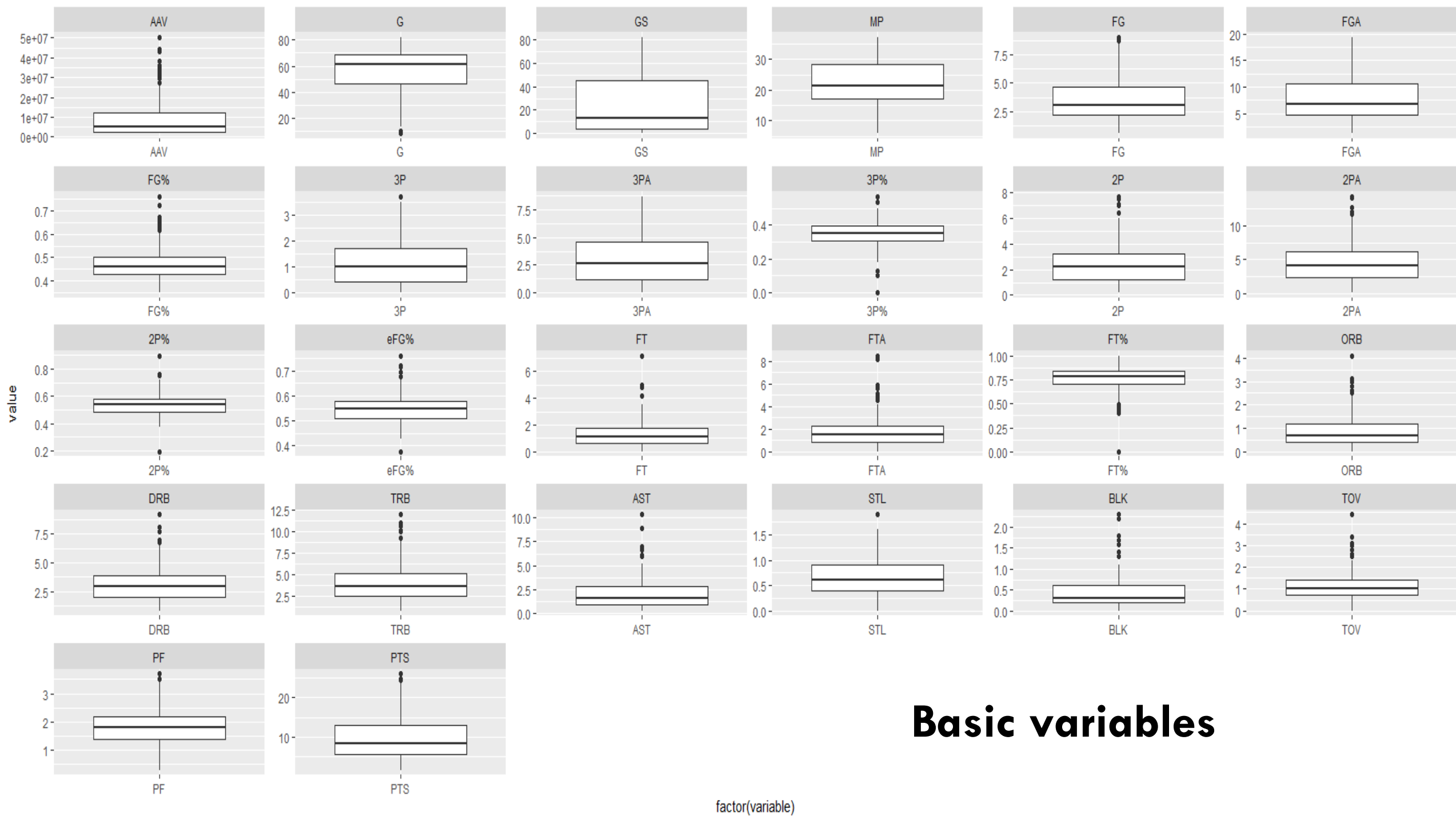
Player	Season	AAV (CONTRACT SIGNED AFTER SEASON)	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%
Nikola Jokic	17-18	29542010.00	75	73	32.6	6.7	13.5	0.499	1.5	3.7	0.396	5.2	9.7	0.538	0.554
Klay Thompson	18-19	37980720.00	78	78	34	8.4	18	0.467	3.1	7.7	0.402	5.3	10.3	0.516	0.553
Tobias Harris	18-19	36000000.00	82	82	34.7	7.5	15.3	0.487	1.9	4.8	0.397	5.5	10.5	0.528	0.549
Jimmy Butler	18-19	35197650.00	65	65	33.6	6.4	13.9	0.462	1	3	0.347	5.4	10.9	0.494	0.499
D'Angelo Russell	18-19	29331375.00	81	81	30.2	8.1	18.7	0.434	2.9	7.8	0.369	5.2	10.9	0.482	0.512
Khris Middleton	18-19	35500000.00	77	77	31.1	6.6	14.9	0.441	2.3	6.2	0.378	4.2	8.8	0.485	0.519
Kyrie Irving	18-19	34122650.00	67	67	33	9	18.5	0.487	2.6	6.5	0.401	6.4	12	0.533	0.557
Al Horford	18-19	27250000.00	68	68	29	5.7	10.6	0.535	1.1	3	0.36	4.6	7.6	0.604	0.586
Nikola Vucevic	18-19	25000000.00	80	80	31.4	8.8	16.9	0.518	1.1	2.9	0.364	7.7	14	0.549	0.549

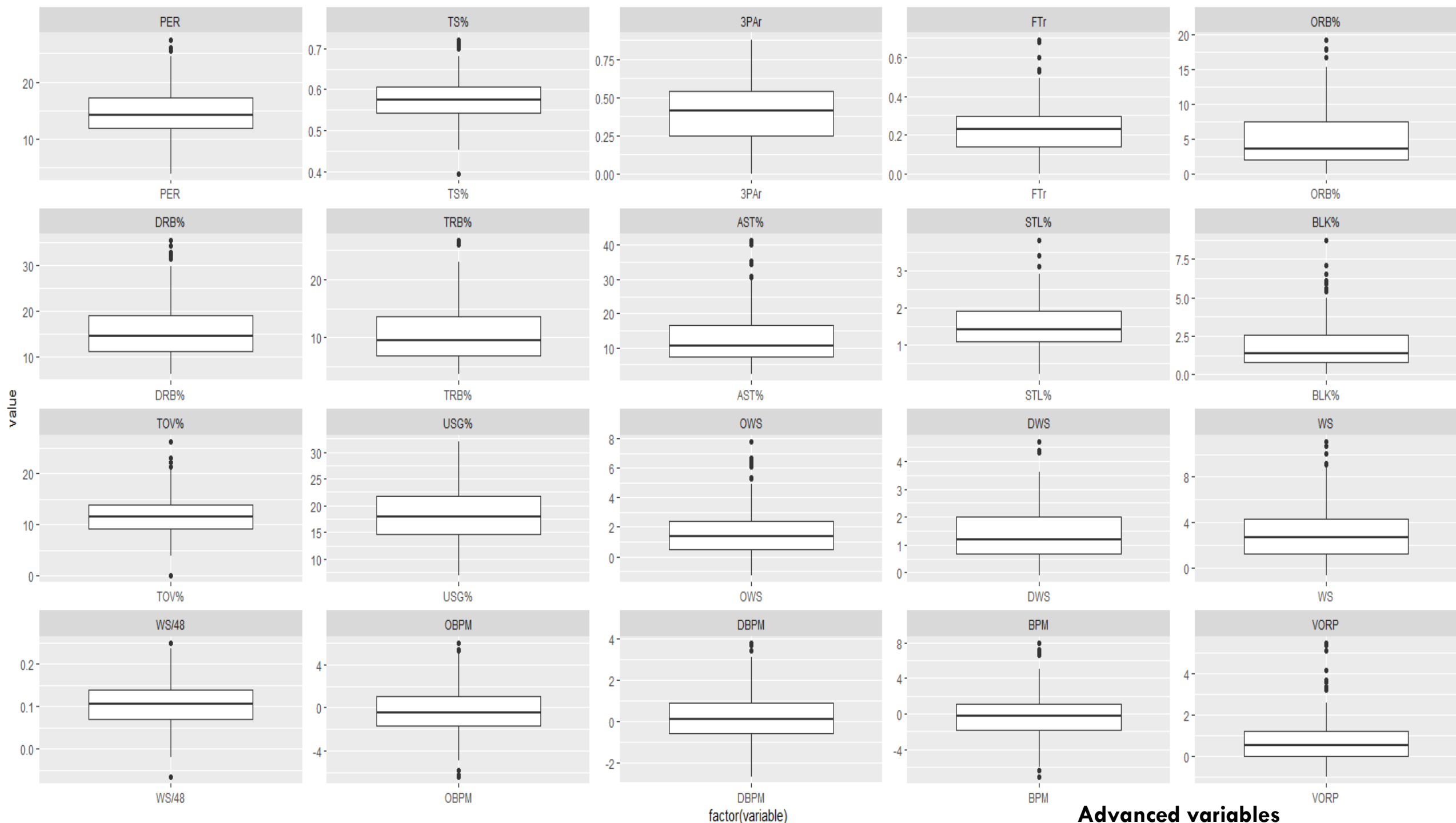
I observed every player that is currently on a FA contract whilst removing outliers: totaling to 194 NBA players in all.

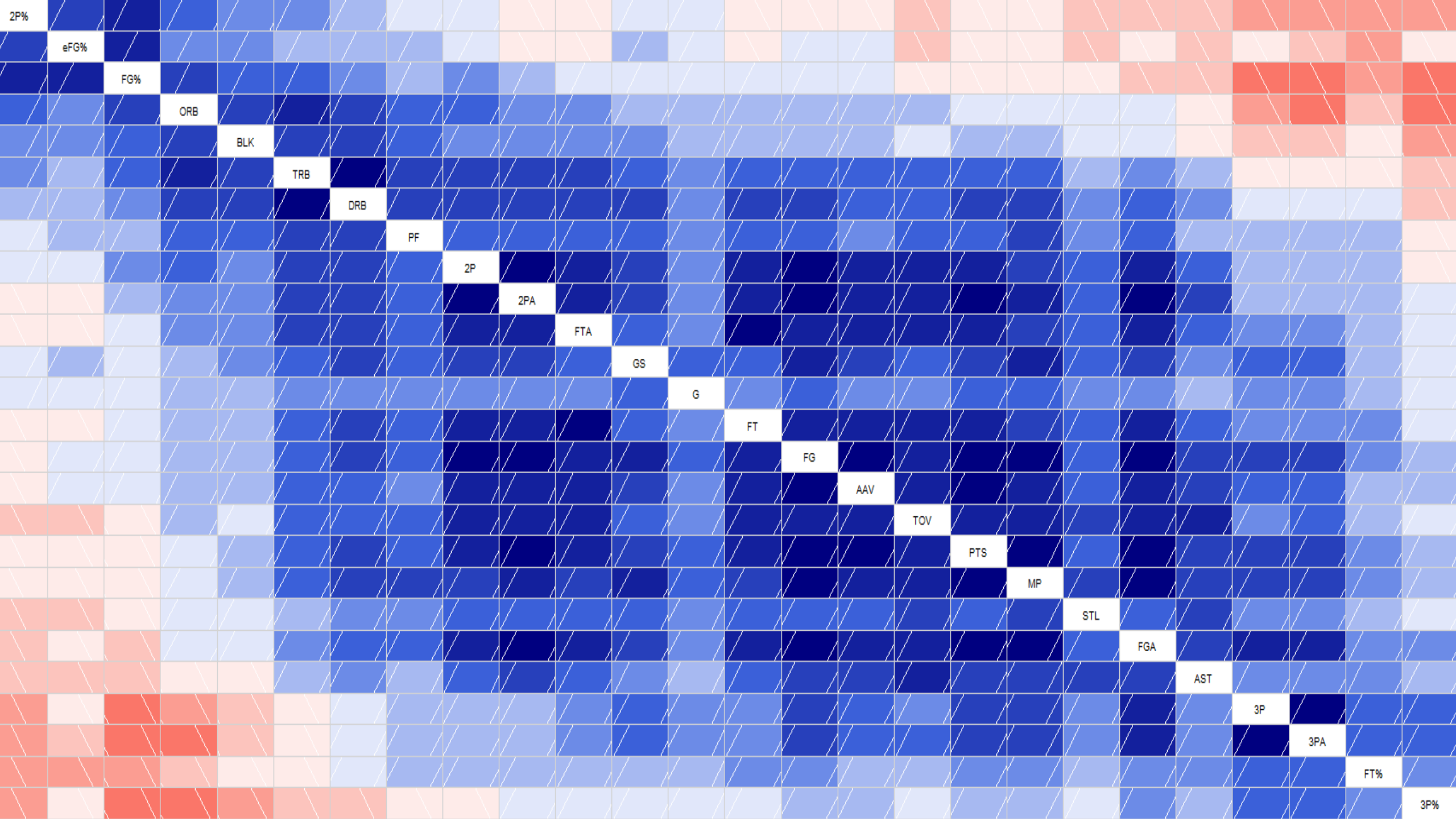
The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit boards or data paths, consisting of lines and small circles.

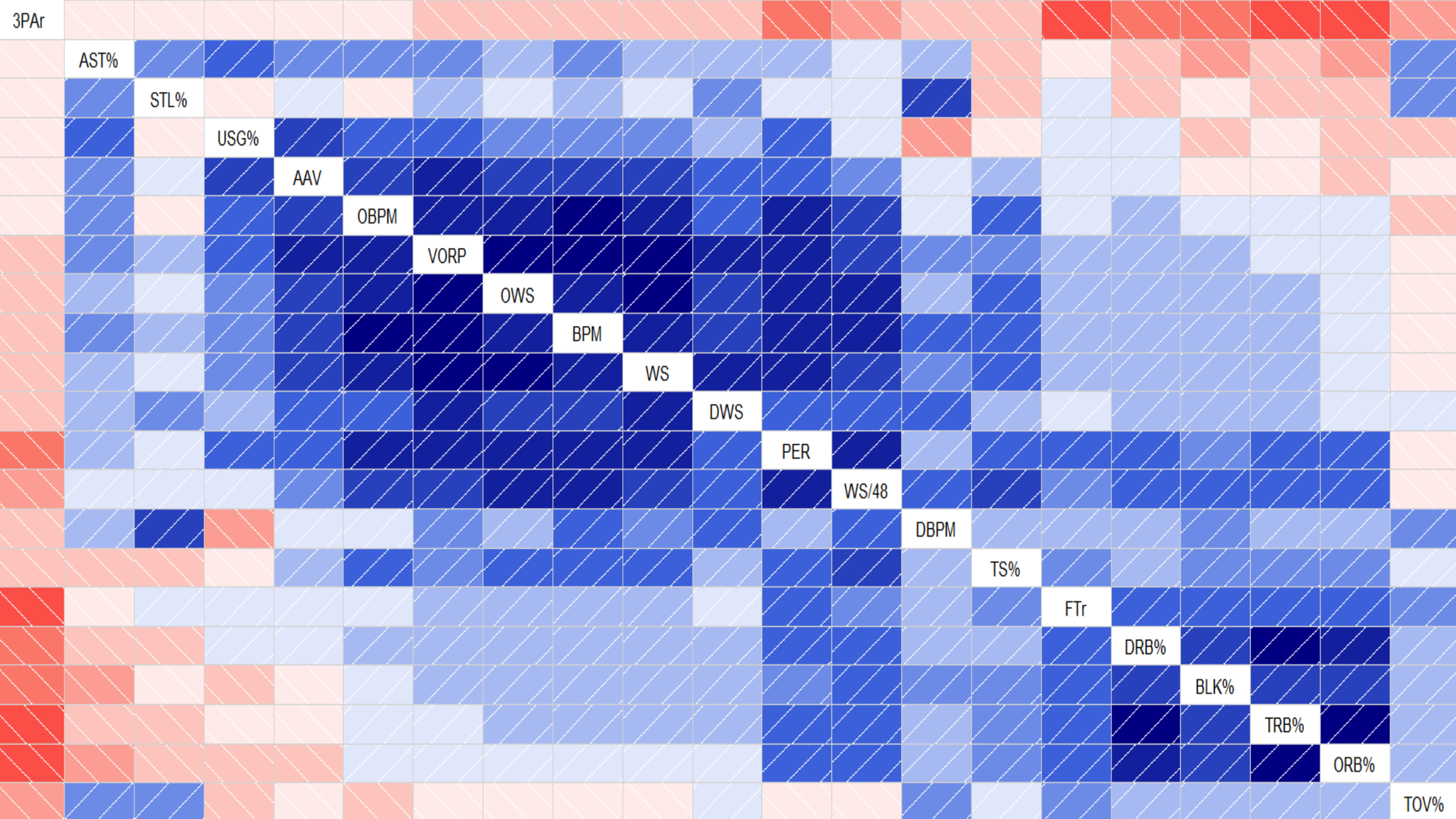
**Let's do some
exploratory data
analysis!**

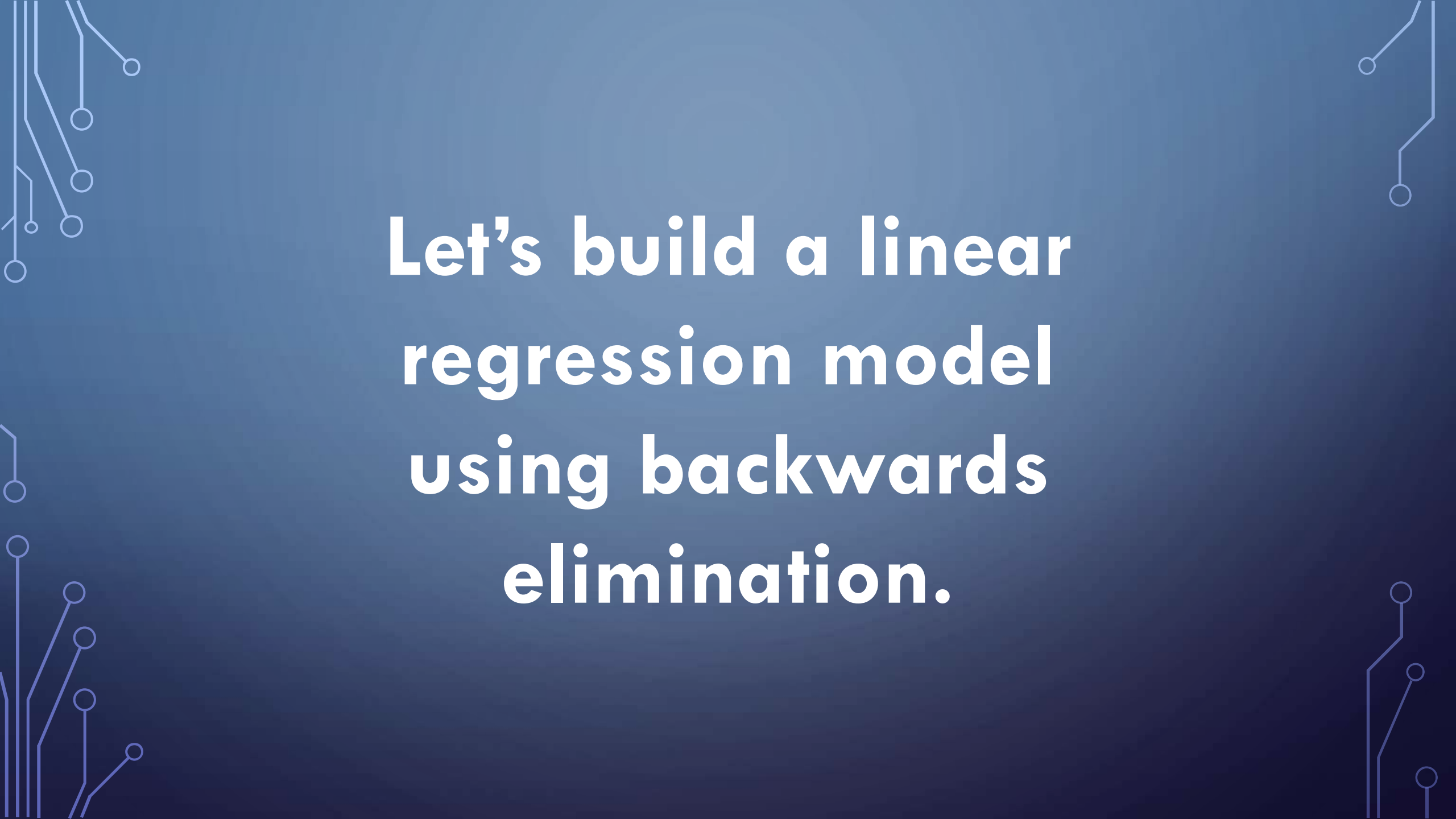










The background is a solid dark blue. It is decorated with white, stylized circuit board traces. These traces are located in the corners: top-left, top-right, bottom-left, and bottom-right. Each trace consists of thin lines that branch out and terminate in small white circles, resembling electronic components or connection points.

**Let's build a linear
regression model
using backwards
elimination.**

```

Residuals:
    Min       1Q   Median       3Q      Max
-11373982 -1833621    4010   1781320  14611917

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	22407498	18596196	1.205	0.2302
G	-34641	45570	-0.760	0.4484
GS	29530	25023	1.180	0.2399
MP	-649256	392975	-1.652	0.1006
FG	-7912328	10676421	-0.741	0.4598
FGA	-2191729	7521156	-0.291	0.7711
`FG%`	-58149832	107072479	-0.543	0.5879
`3P`	-10294794	9814341	-1.049	0.2959
`3PA`	2747516	7410795	0.371	0.7114
`3P%`	-1382878	5480730	-0.252	0.8011
`2P`	416112	7456137	0.056	0.9556
`2PA`	566356	7521656	0.075	0.9401
`2P%`	6312796	11662742	0.541	0.5891
`eFG%`	8244261	106727696	0.077	0.9385
FT	-7058872	5339126	-1.322	0.1882
FTA	1072018	3610352	0.297	0.7669
`FT%`	-477295	4551673	-0.105	0.9166
ORB	10682874	7173375	1.489	0.1386
DRB	11539444	7455008	1.548	0.1238
TRB	-11857809	7190959	-1.649	0.1013
AST	2056597	1420295	1.448	0.1497
STL	-3161209	4508616	-0.701	0.4843
BLK	-4509531	3843831	-1.173	0.2426
TOV	49775	3087439	0.016	0.9872
PF	-2703757	1197048	-2.259	0.0254 *
PTS	8785057	4585565	1.916	0.0573 .
PER	-1848560	1153912	-1.602	0.1113
`TS%`	43694857	61391717	0.712	0.4778
`3Par`	-12194566	21160737	-0.576	0.5653
FTR	-2234262	8126737	-0.275	0.7838
`ORB%`	-5282	1870984	-0.003	0.9978
`DRB%`	-494727	1664071	-0.297	0.7667
`TRB%`	1452320	3362021	0.432	0.6664
`AST%`	-51062	264407	-0.193	0.8471
`STL%`	2682648	2184413	1.228	0.2214
`BLK%`	1827430	1148998	1.590	0.1139
`TOV%`	-53506	232461	-0.230	0.8183
`USG%`	-660538	465330	-1.420	0.1579
OWS	-3580639	6302244	-0.568	0.5708
DWS	-3665202	6348950	-0.577	0.5646
WS	4577535	6137450	0.746	0.4570
`WS/48`	-2176444	48040920	-0.045	0.9639
OBPM	15511338	6888749	2.252	0.0258 *
DBPM	14882579	6885305	2.161	0.0323 *
BPM	-14612237	6882204	-2.123	0.0354 *
VORP	131408	1973793	0.067	0.9470

Signif. codes:

0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ', 1

Residual standard error: 4075000 on 147 degrees of freedom
Multiple R-squared: 0.8703, Adjusted R-squared: 0.8306
F-statistic: 21.92 on 45 and 147 DF, p-value: < 2.2e-16

Residuals:

Min	1Q	Median	3Q	Max
-11158877	-2136533	98359	1839048	19060705

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	11772451	2985861	3.943	0.000115	***
GS	53885	18735	2.876	0.004505	**
MP	-806142	151472	-5.322	2.99e-07	***
AST	1345801	276535	4.867	2.45e-06	***
PF	-2406166	732440	-3.285	0.001223	**
PTS	3288892	310396	10.596	< 2e-16	***
`TRB%`	241318	82884	2.912	0.004046	**
`USG%`	-1071106	176720	-6.061	7.61e-09	***
OBPM	13274438	5894260	2.252	0.025511	*
DBPM	14447590	5894302	2.451	0.015186	*
BPM	-13945789	5904655	-2.362	0.019242	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3970000 on 182 degrees of freedom

Multiple R-squared: 0.8476, Adjusted R-squared: 0.8392

F-statistic: 101.2 on 10 and 182 DF, p-value: < 2.2e-16


10 SIGNIFICANT EXPLANATORY VARIABLES

- The basics:
 - GS, MP, **AST**, PF, & **PTS**.
- Advanced statistics:
 - TRB%, **USG%**, **OBPM**, **DBPM**, & BPM.



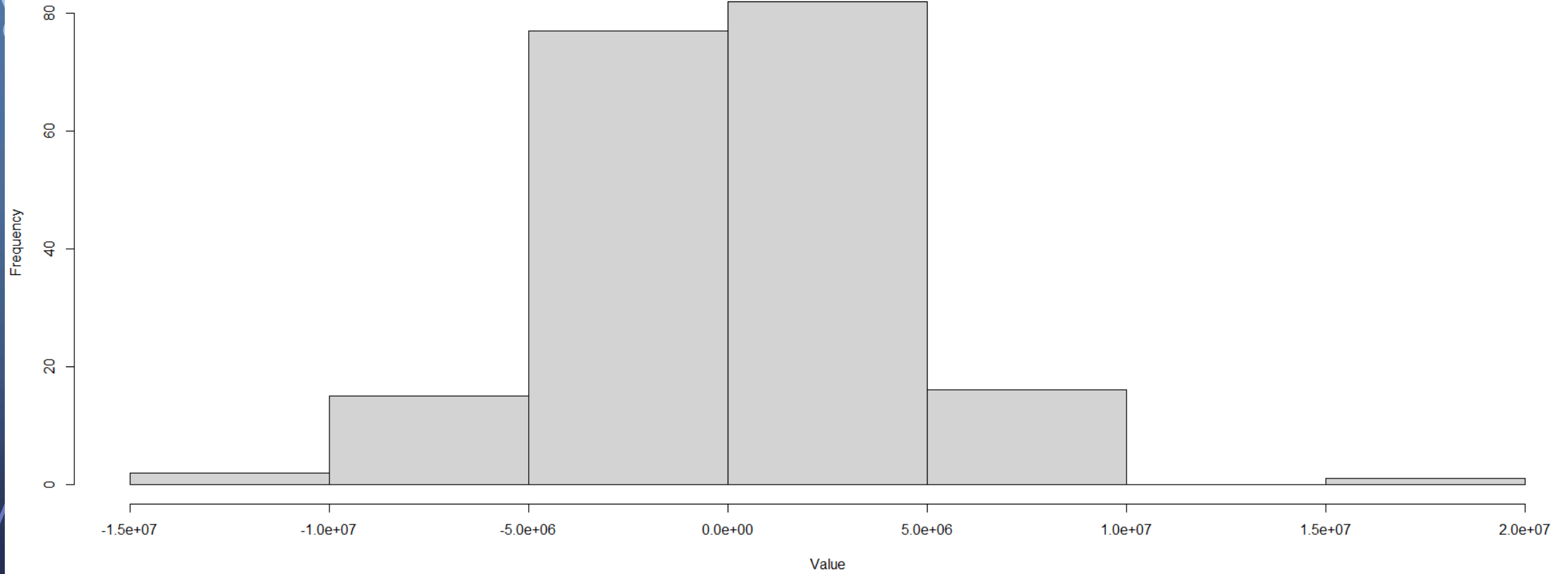
Red variables are offensive stats.

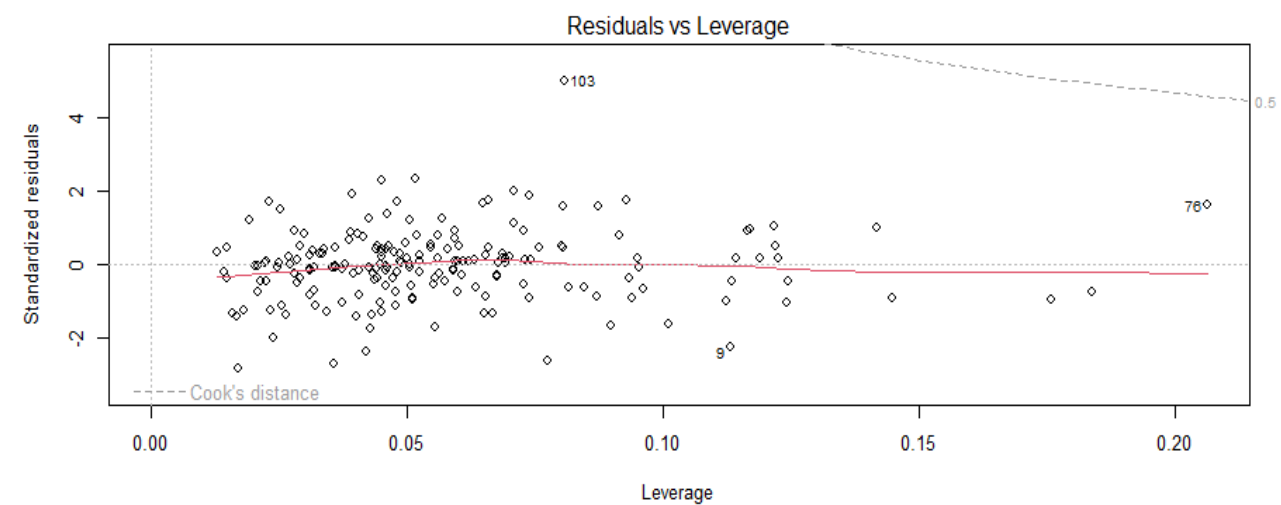
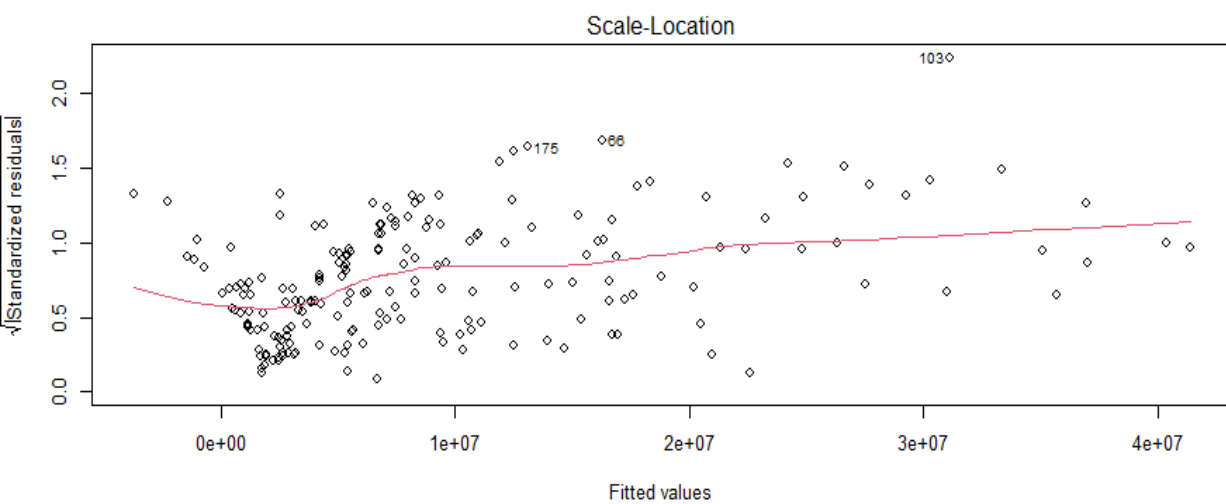
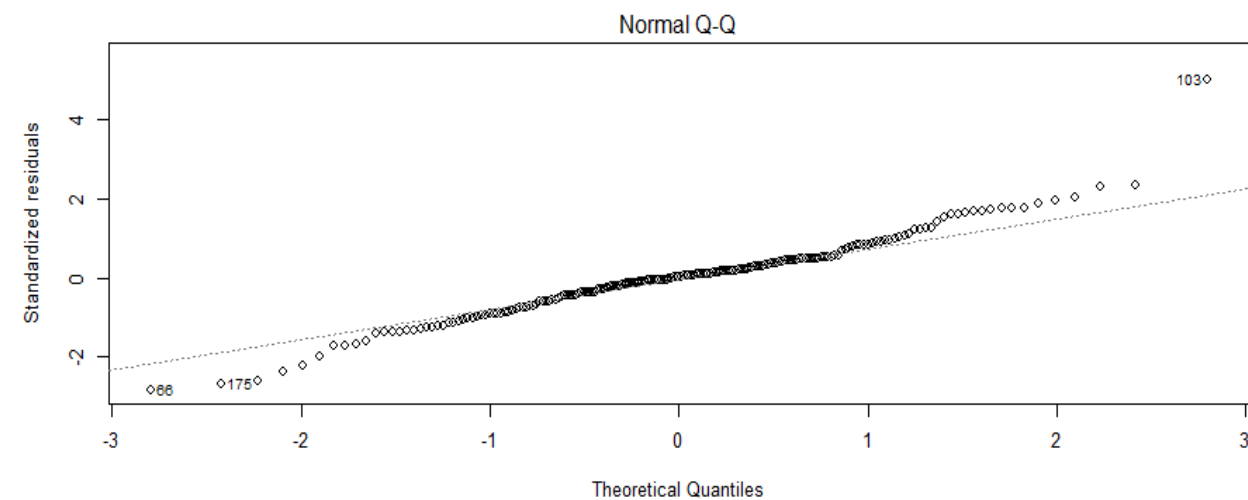
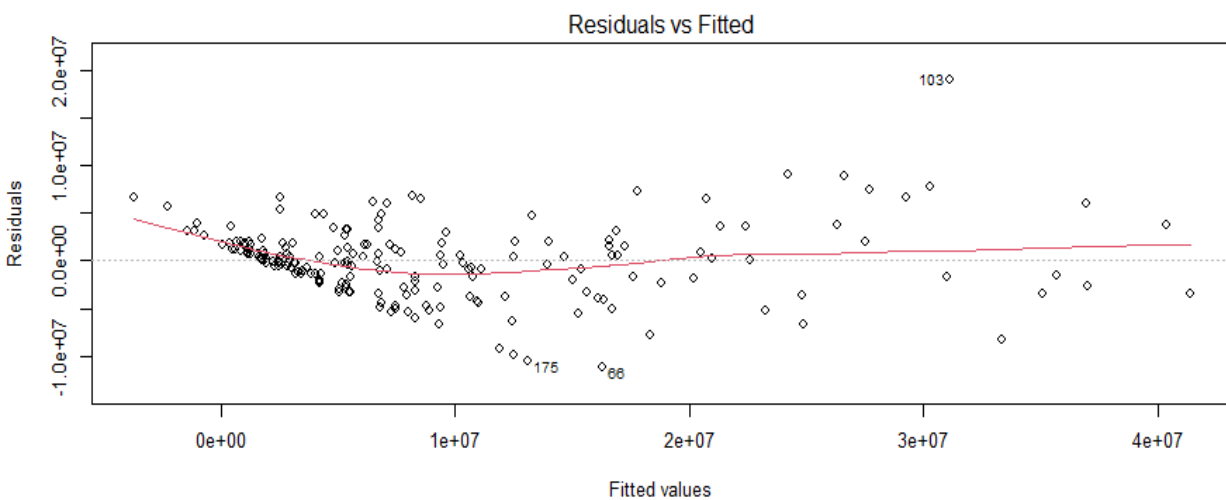
Yellow variables are defensive stats.

The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural network connections. These include vertical and diagonal lines ending in small circles, resembling nodes or components.

**Now let's look at
the residual plots.**

Histogram of the Residuals





Does the model really work?

- I divided the data for the model into two subsets.
- 70% went to a training subset and 30% went to a test subset.
- 70% of the data is used to “train” the model in order to make predictions for the test subset.

88.93% ACCURATE

IN SUMMARY

- Significant player statistics:
 - **GS, MP, AST, PF, & PTS.**
 - **TRB%, USG%, OBPM, DBPM, & BPM.**
 - **4 offensive statistics.**
 - **1 defensive statistic.**

SOMETHING TO THINK ABOUT...

- "Offense sells tickets, defense wins games, rebounding wins championships."



REFERENCES

- <https://www.spotrak.com/nba/contracts/>
- <https://www.basketball-reference.com/>