STOR320
Group 5 Final Project

Alexandra Lostetter, Ava Konopka, Ryan Brown,
Amy Thalangsy, Heywood Williams-Tracy

In college basketball statistics, what combination of variables best predicts win percentage?

Where did the data come from?

2023 Customizable T-Rank and Tempo-Free Stats

RK	TEAM	CONF	G	D-I AVG: REC	104.1 ADJOE	ADJDE	0.4946 BARTHAG	EFF. FG% 50.4 EFGD%	TURNOVER% 18.2 TOR	TORD	REBOUND% 28.5 ORB	DRB	FT RATE 31.5 FTR	FTRD	2-PT % 50.1 2P%	2P%D	3-PT % 34 3P%	3P%D	3P RATE 37.3 3PR	3PRD	67.1 ADJ T.	WAB
15	Duke 5 seed, R32	ACC	36	27-9 14–6	113.3 33	92.9 14	.9072 15	51.1 146	46 15	18.3 192	16.5 278	35.9 8	26.1 68	29.9 224	23.7 13	51.5 110	46.2 29	33.5 20	30.5 20	35.3 250	35.6 128	64.9 298	4.5 16
27	Miami FL 5 seed, Final Four	ACC	37	29-8 15–5	119.1 7	100.9 109	.8715 27	54.3 24	51.2 213	16.1 47	18.1 169	31.8 72	28.8 195	30.7 203	24 17	53.8 40	51.8 254	36.8 34	33.5 142	34.2 275	38.5 222	68.6 104	3.7 20
34	Virginia 4 seed, R64	ACC	33	25-8 15–5	110.0 68	94.0 20	.8591 34	51.2 138	48.3 75	13.6 5	19.5 91	25.6 268	24.9 37	34.7 76	26.3 47	50.4 173	46.4 33	35 117	34.1 186	35.6 239	40 278	61.3 360	4.1 18
40	North Carolina	ACC	33	20-13 11–9	111.5 51	96.4 44	.8431 40	49.2 241	48.7 92	15.2 23	14.9 341	30.4 118	23.3 12	37 41	26.3 47	50.8 150	48.4 88	31.1 324	33 109	37.7 169	32 31	68.8 96	-0.7 62
48	North Carolina St. 11 seed, R64	ACC	34	23-11 12–8	113.6 32	99.2 80	.8272 48	51.5 117	49.8 131	13.5 4	18.8 125	30.4 118	26.2 78	25.3 340	34 253	51.1 132	50.9 211	34.7 137	31.6 51	37.1 190	33.5 52	69.2 72	1.1 41
57	Clemson	ACC	34	23-11 14–6	110.9 59	97.7 61	.8096 57	54.1 29	47.3 43	16.1 47	16.4 285	23.4 325	24.3 22	30.5 211	30.7 171	53.7 48	45.6 21	36.4 49	33.2 121	40.3 95	40.8 293	68 137	-0.3 57
72	Pittsburgh 11 seed, R32	ACC	36	24-12 14–6	113.1 35	101.6 123	.7743 72	52.5 70	48.5 83	16.3 55	16.6 268	29.5 150	28.6 182	32.9 122	30.4 162	51.6 117	48.2 83	36 72	32.6 93	43.4 42	39.5 262	67.2 172	-0.3 54
75	Virginia Tech	ACC	33	19-15 8–12	114.0 29	102.6 143	.7710 75	54 31	50.7 182	14.7 11	16.3 291	25.9 255	27.7 140	25.9 334	25.2 27	54.3 30	50.5 189	35.8 78	34 176	39.6 113	35.6 210	66.6 99	-3.2 97
85	Wake Forest	ACC	33	19-14 10–10	111.4 53	101.5 120	.7432 85	54.2 26	51.6 239	17.1 92	17.3 223	24.8 295	26.6 91	32.1 151	26.7 55	54 34	51 216	36.4 49	35 245	45.4 20	38.6 225	69.4 65	-3 94
131	Syracuse	ACC	32	17-15 10–10	109.0 89	105.7 205	.5863 131	51.6 112	51.2 213	17 88	18.7 126	30 133	33.6 346	30.7 203	50.9 4	36.1 168	34.4 211	29.2 339	48.3 359	67.2 176	-5.7 128		
134	Notre Dame	ACC	32	11-21 3–17	108.5 98	105.8 207	.5743 134	52 114	52.1 219	13.4 2	13.6 360	19.6 355	26.5 83	23.3 356	26.5 2	50.6 78	52.4 204	35.8 52	34.3 144	43 304	36.1	64.7 243	-11.6
136	Georgia Tech	ACC	31	15-18 6–14	104.7 159	102.2 135	.5687 136	48.5 280	50.1 146	15.6 32	15.9 307	27.5 214	29.9 242	23.5 355	30.8 172	47.4 288	52 265	33.4 199	31 28	38.9 133	35.2 113	66.5 214	-8.1 174
156	Boston College	ACC	33	16-17 9–11	102.5 208	101.7 128	.5223 156	48.7 269	51.5 235	18.3 192	18.3 156	27.6 210	28.3 172	27.3 309	27.6 76	49.2 226	49.1 122	31.9 293	36.8 323	31.5 317	39.5 262	66.5 213	-6.6 147
205	Florida St.	ACC	32	9-23 7–13	104.0 169	106.6 231	.4299 205	49.3 236	53.1 302	17.6 134	17.1 236	25.9 255	31.5 306	30.4 213	30.9 174	49.2 226	50.8 207	33 218	37.6 343	34.5 267	41.1 301	68.7 100	-11.9 253
263	Louisville	ACC	32	4-28 2–18	101.2 235	109.4 286	.2915 263	48 296	54.1 327	22.3 360	15.2 327	28.2 302	30.1 256	32.9 122	30.6 168	47.3 294	54.9 342	32.9 225	35.2 215	35.8 233	39.7 289	66.5 215	-16.8 337
RK	TEAM	CONF	G	REC	ADJOE	ADJDE	BARTHAG	EFG%	EFGD%	TOR	TORD	ORB	DRB	FTR	FTRD	2P%	2P%D	3P%	3P%D	3PR	3PRD	ADJ T.	WAB

Compiled by Bart Torvik, sourced from NCAA
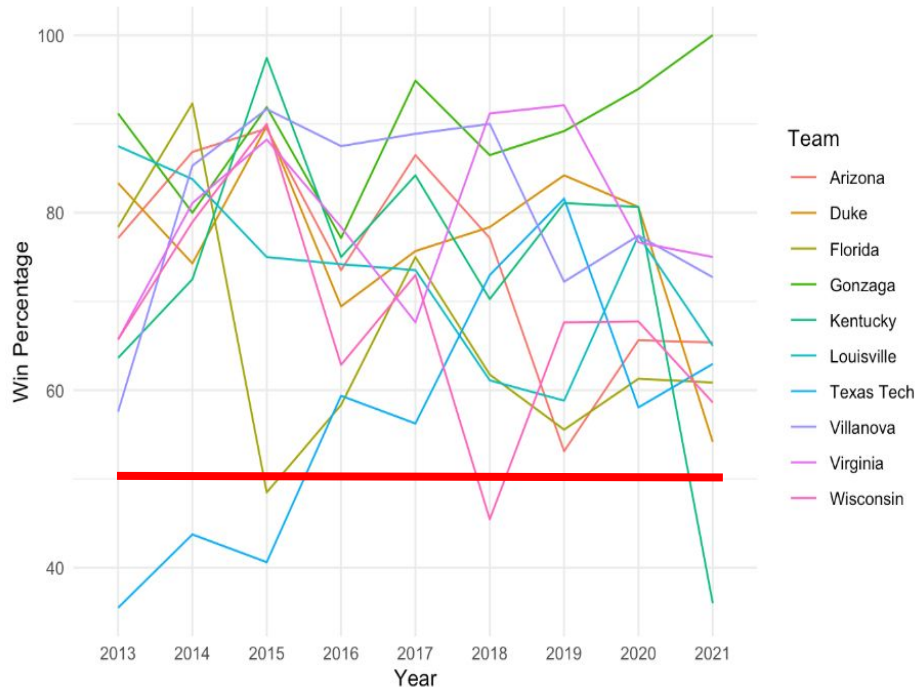
Variables of Interest

- Data mainly includes technical gameplay stats, e.g.
    - Free Throw Rate
    - Rebound Rate
    - Field Goal Shot Percentage
- Does NOT include "non-technical", e.g.
    - Injury Statistics
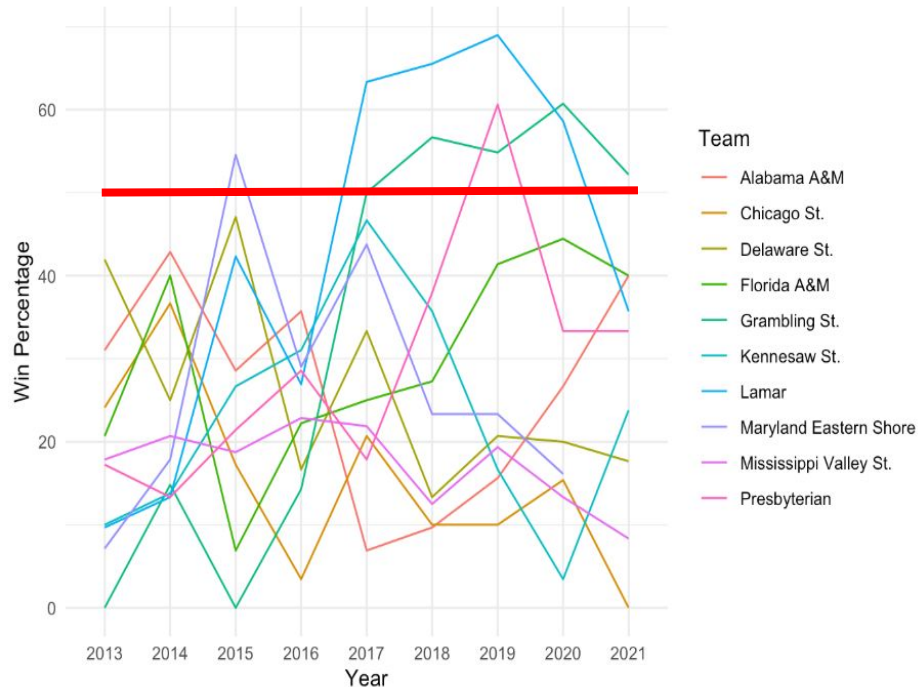   - NCAA Violations
- We constructed Win Percentage variable

Comparing Win Percentages

Win Percentage of Top 10 Teams from 2013-2021

Team
- Arizona
- Duke
- Florida
- Gonzaga
- Kentucky
- Louisville
- Texas Tech
- Villanova
- Virginia
- Wisconsin

Win Percentage of Bottom 10 Teams from 2013-2021

Team
- Alabama A&M
- Chicago St.
- Delaware St.
- Florida A&M
- Grambling St.
- Kennesaw St.
- Lamar
- Maryland Eastern Shore
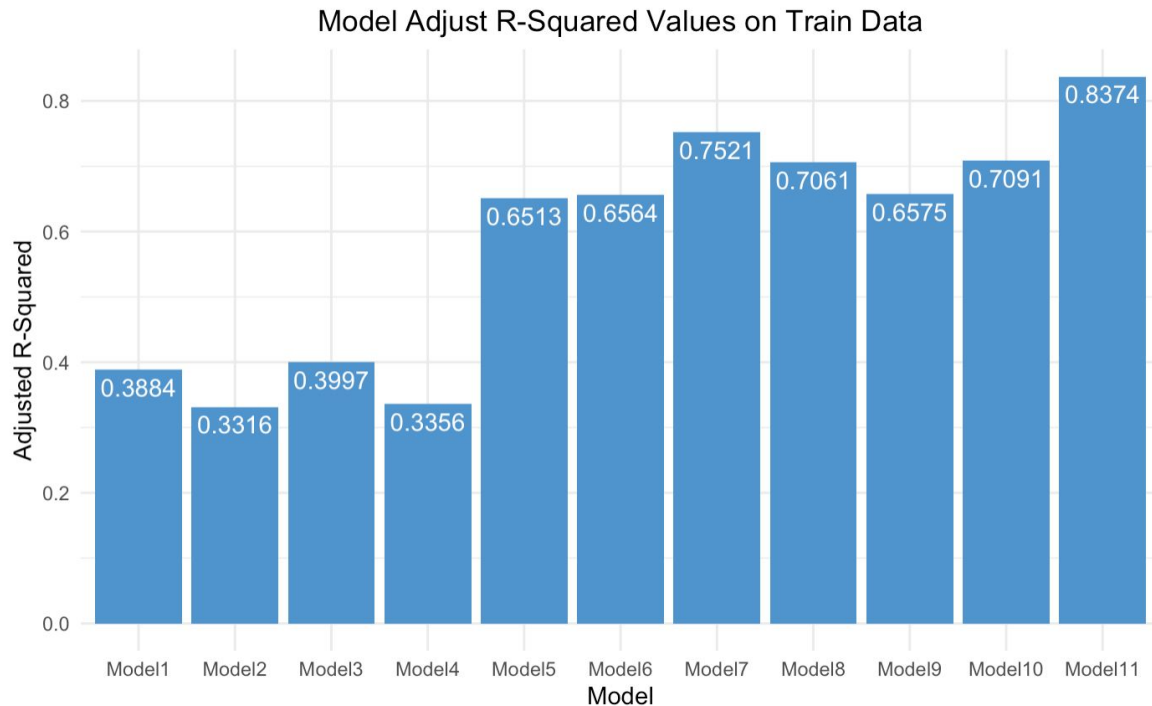- Mississippi Valley St.
- Presbyterian

Methodology

1. Cross validation
2. Candidate models (adjusted R-squared)
3. Regular subsets (Mallows' Cp)
4. Further Analysis (MAE)

Candidate Models

Models build up in complexity, starting at only field goals (O/D) for Models 1 and 2 and including all predictors in 11

Model Adjust R-Squared Values on Train Data

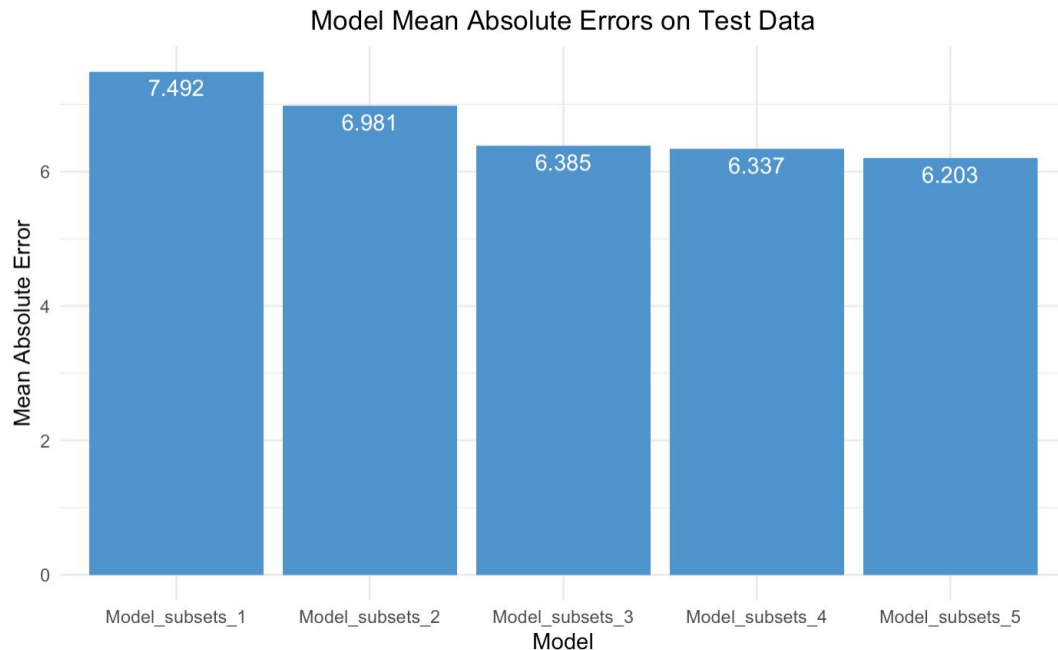| Model	Adjusted R-Squared
Model1	0.3884
Model2	0.3316
Model3	0.3997
Model4	0.3356
Model5	0.6513
Model6	0.6564
Model7	0.7521
Model8	0.7061
Model9	0.6575
Model10	0.7091
Model11	0.8374

Regular Subsets

Adding more predictors decreases Mallows' Cp (1024 vs 38) and MAE

Model 5 includes field goals, free throws, turnovers, and rebounds

Model Mean Absolute Errors on Test Data

Best Model Actual vs Predicted

Win Percentage: Actual vs Predicted on Test Data
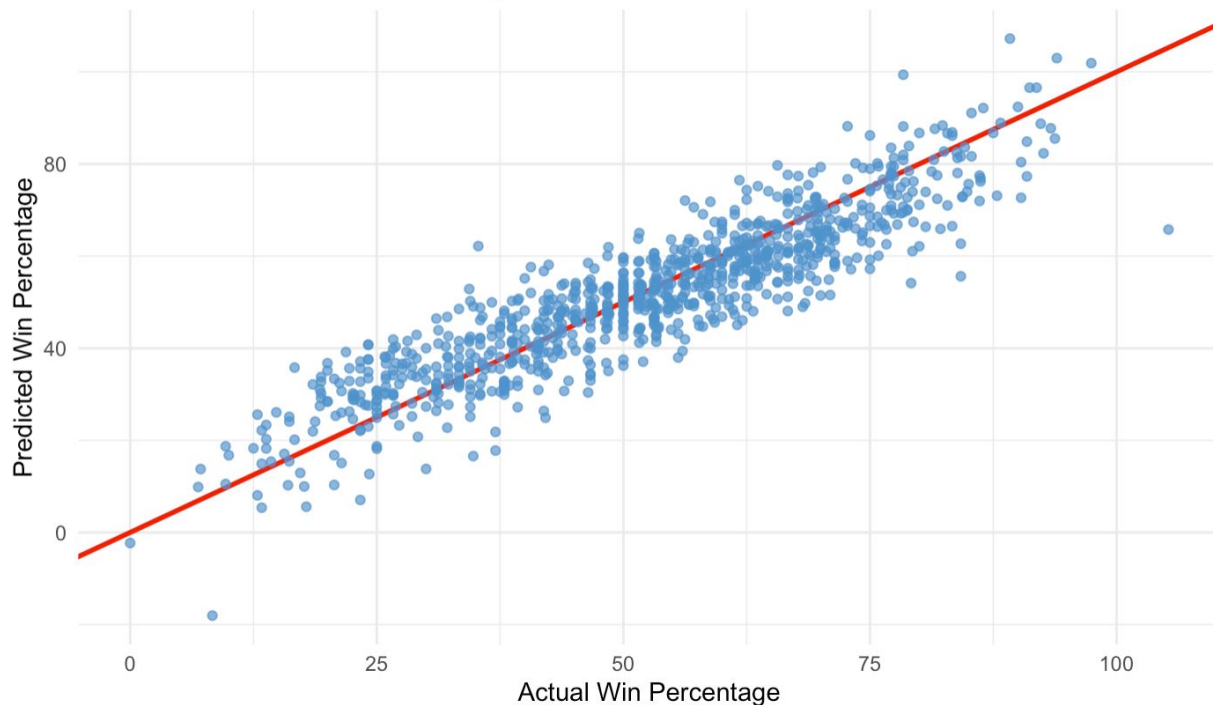
Results

- Found that all skills measured by our data were relevant
- Further analysis likely requires non-technical stats, e.g.
  - Recruiting
  - Injury
  - Program budget