

NO SEASON BETTER

The Phenomenal 1920 Season of George Sisler

Michael Round
Center for
auto SOCRATIC EXCELLENCE



www.rationalsys.com
(913) 515-3911

The benchmark for a good season, batting-average wise, is 300. A great hitter hits 400. Most people know the statistic: Ted Williams was the last 400 hitter - 406 in 1941. Each year, excitement is rampant as a player or two toys with the magic 400 level, before usually collapsing in late-season.

But how would one gauge "the best season"? Sure, one could rank annual batting champions, and search for the highest average. That'd be easy - maybe even right.

This analysis, based on data from Retrosheet, suggests the 1920 season of George Sisler stands above all others.



BASEBALL BATTING CHAMPIONS

Clearly, by the table, Nap Lajoie's 426 average makes this the best hitting season, followed by the great Rogers Hornsby at 424, and George Sisler and Ty Cobb, each with 420 seasons. *But is this right?*

Has a 400 hitter with 400 at-bats, for example, had a better season than a player hitting 380 with 600 at-bats? As note above, several players have flirted with the magical "400", only to fall off as the season progresses. Is our 400 hitter above any different? These are only two variables - average and # of at-bats. Surely, there are others: on-base %, slugging %, home-runs, etc.

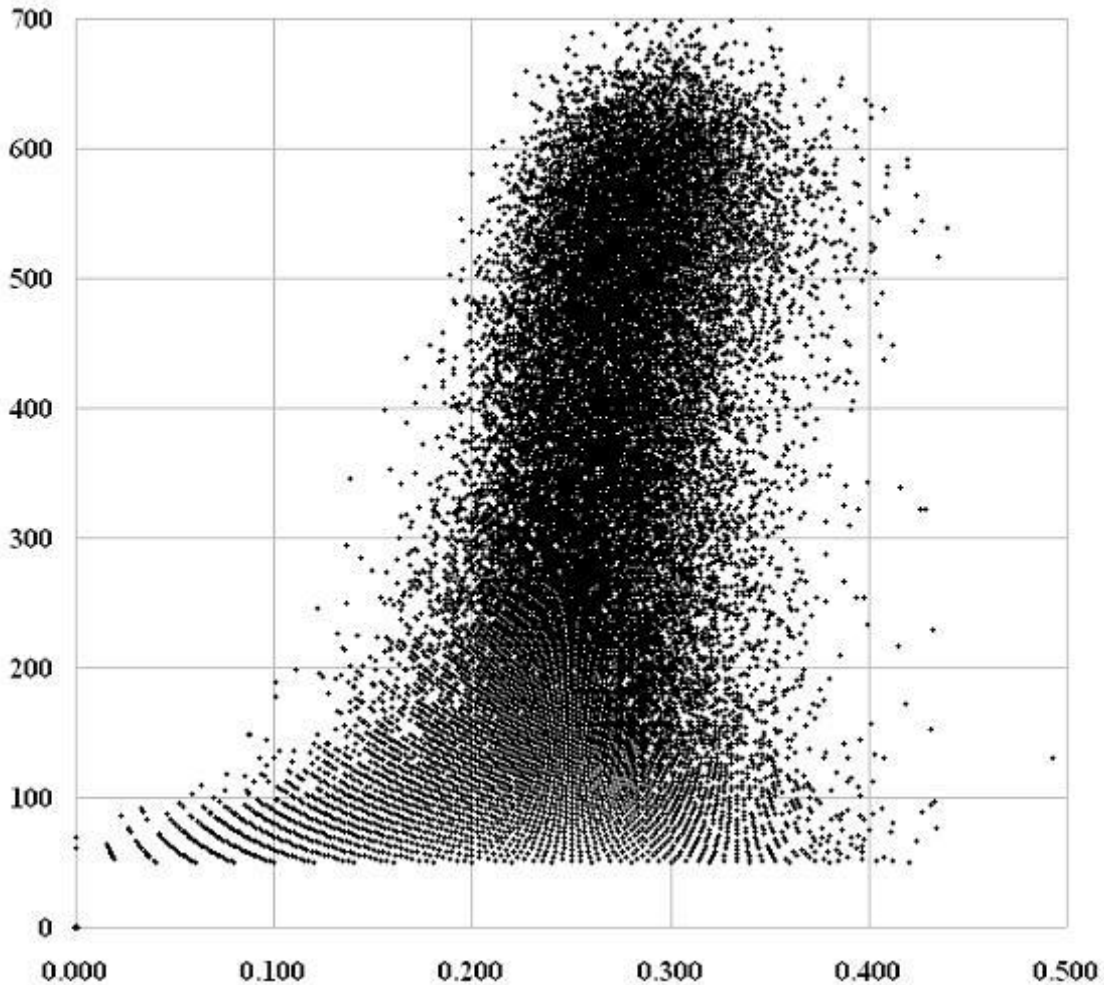
Let's assume, for this sake of this argument, I'm interested only in batting average. Wouldn't it be neat, before answering the question above of who's best, to look at the performance of all batters over the years? That is, plot all players' statistics.

But to make sure the data is not skewed by a player getting 10 hits in 20 at-bats. let's limit the number of at-bats: choose a random number: 50.

AMERICAN			NATIONAL		AMERICAN			NATIONAL	
Year	Player	Avg	Player	Avg	Year	Player	Avg	Player	Avg
1901	Lajoie	0.426	Burkett	0.376	1951	Fain	0.344	Musial	0.355
1902	Delahanty	0.376	Beaumont	0.357	1952	Fain	0.327	Musial	0.336
1903	Lajoie	0.344	Wagner	0.355	1953	Vernon	0.337	Furillo	0.344
1904	Lajoie	0.376	Wagner	0.349	1954	Avila	0.341	Mays	0.345
1905	Flick	0.308	Seymour	0.377	1955	Kaline	0.340	Ashburn	0.338
1906	Stone	0.358	Wagner	0.339	1956	Mantle	0.353	Aaron	0.328
1907	Cobb	0.350	Wagner	0.350	1957	Williams	0.388	Musial	0.351
1908	Cobb	0.324	Wagner	0.354	1958	Williams	0.328	Ashburn	0.350
1909	Cobb	0.377	Wagner	0.339	1959	Kuenn	0.353	Aaron	0.355
1910	Cobb	0.385	Magee	0.331	1960	Runnels	0.320	Groat	0.325
1911	Cobb	0.420	Wagner	0.334	1961	Cash	0.361	Clemente	0.351
1912	Cobb	0.409	Zimmerman	0.372	1962	Runnels	0.326	Davis	0.346
1913	Cobb	0.390	Daubert	0.350	1963	Yastrzemski	0.321	Davis	0.326
1914	Cobb	0.368	Daubert	0.329	1964	Oliva	0.323	Clemente	0.339
1915	Cobb	0.369	Doyle	0.320	1965	Oliva	0.323	Clemente	0.329
1916	Speaker	0.386	Chase	0.339	1966	Robinson	0.316	Alou	0.342
1917	Cobb	0.383	Roush	0.341	1967	Yastrzemski	0.326	Clemente	0.357
1918	Cobb	0.382	Wheat	0.335	1968	Yastrzemski	0.301	Rose	0.335
1919	Cobb	0.384	Roush	0.321	1969	Carew	0.332	Rose	0.348
1920	Sisler	0.407	Hornsby	0.370	1970	Johnson	0.329	Carty	0.366
1921	Heilmann	0.394	Hornsby	0.397	1971	Oliva	0.337	Torre	0.363
1922	Sisler	0.420	Hornsby	0.401	1972	Carew	0.318	Williams	0.333
1923	Heilmann	0.403	Hornsby	0.384	1973	Carew	0.350	Rose	0.338
1924	Ruth	0.378	Hornsby	0.424	1974	Carew	0.364	Garr	0.353
1925	Heilmann	0.393	Hornsby	0.403	1975	Carew	0.359	Madlock	0.354
1926	Manush	0.378	Hargrave	0.353	1976	Brett	0.333	Madlock	0.339
1927	Heilmann	0.398	Waner	0.380	1977	Carew	0.388	Parker	0.338
1928	Goslin	0.379	Hornsby	0.387	1978	Carew	0.333	Parker	0.334
1929	Fonseca	0.369	O'Doul	0.398	1979	Lynn	0.333	Hernandez	0.344
1930	Simmons	0.381	Terry	0.401	1980	Brett	0.390	Buckner	0.324
1931	Simmons	0.390	Hafey	0.349	1981	Lansford	0.336	Madlock	0.340
1932	Alexander	0.367	O'Doul	0.368	1982	Wilson	0.332	Oliver	0.331
1933	Foxx	0.356	Klein	0.368	1983	Boggs	0.361	Madlock	0.323
1934	Gehrig	0.363	Waner	0.362	1984	Mattingly	0.343	Gwynn	0.351
1935	Myer	0.349	Vaughan	0.385	1985	Boggs	0.368	McGee	0.353
1936	Appling	0.388	Waner	0.373	1986	Boggs	0.357	Raines	0.334
1937	Gehring	0.371	Medwick	0.374	1987	Boggs	0.363	Gwynn	0.370
1938	Foxx	0.349	Lombardi	0.342	1988	Boggs	0.366	Gwynn	0.313
1939	DiMaggio	0.381	Mize	0.349	1989	Puckett	0.339	Gwynn	0.336
1940	DiMaggio	0.352	Garns	0.352	1990	Brett	0.329	McGee	0.335
1941	Williams	0.406	Reiser	0.343	1991	Franco	0.341	Pendleton	0.319
1942	Williams	0.356	Lombardi	0.330	1992	Martínez	0.343	Sheffield	0.330
1943	Appling	0.328	Musial	0.357	1993	Olerud	0.363	Galarraga	0.370
1944	Boudreau	0.327	Walker	0.357	1994	O'Neill	0.359	Gwynn	0.394
1945	Stirnweiss	0.309	Cavarretta	0.355	1995	Martínez	0.356	Gwynn	0.368
1946	Vernon	0.353	Musial	0.365	1996	Rodriguez	0.358	Gwynn	0.353
1947	Williams	0.343	Walker	0.363	1997	Thomas	0.347	Gwynn	0.372
1948	Williams	0.369	Musial	0.376	1998	Williams	0.339	Walker	0.363
1949	Kell	0.343	Robinson	0.342	1999	Garcia	0.357	Walker	0.379
1950	Goodman	0.354	Musial	0.346	2000	Garcia	0.372	Helton	0.372
					2001	Suzuki	0.350	Walker	0.350
					2002	Ramírez	0.349	Bonds	0.370
					2003	Mueller	0.326	Pujols	0.359
					2004	Suzuki	0.372	Bonds	0.362
					2005	Young	0.331	Lee	0.335
					2006	Mauer	0.347	Sanchez	0.344
					2007	Ordóñez	0.363	Holliday	0.340

MAJOR LEAGUE BASEBALL: 1871-2007

50+ At-Bats in A Season



With average on the x-axis and at-bats along the y-axis, some interesting things show up: few players last long if they're not hitting well. If you're only hitting 100, you're not likely to get many at-bats. One can easily see how hard it is to hit 400!

To my question, the points in the upper-right become individually distinctive - and they show # of hits as the product of at-bats and average.

The Great George Sisler

Two of these individual points are the great George Sisler, in 1920 and 1922. In 1920, he had 257 hits in only 154 games! Both points, being individually distinguishable and upper-most towards the right, tell me these are among the two greatest batting-average seasons ever.

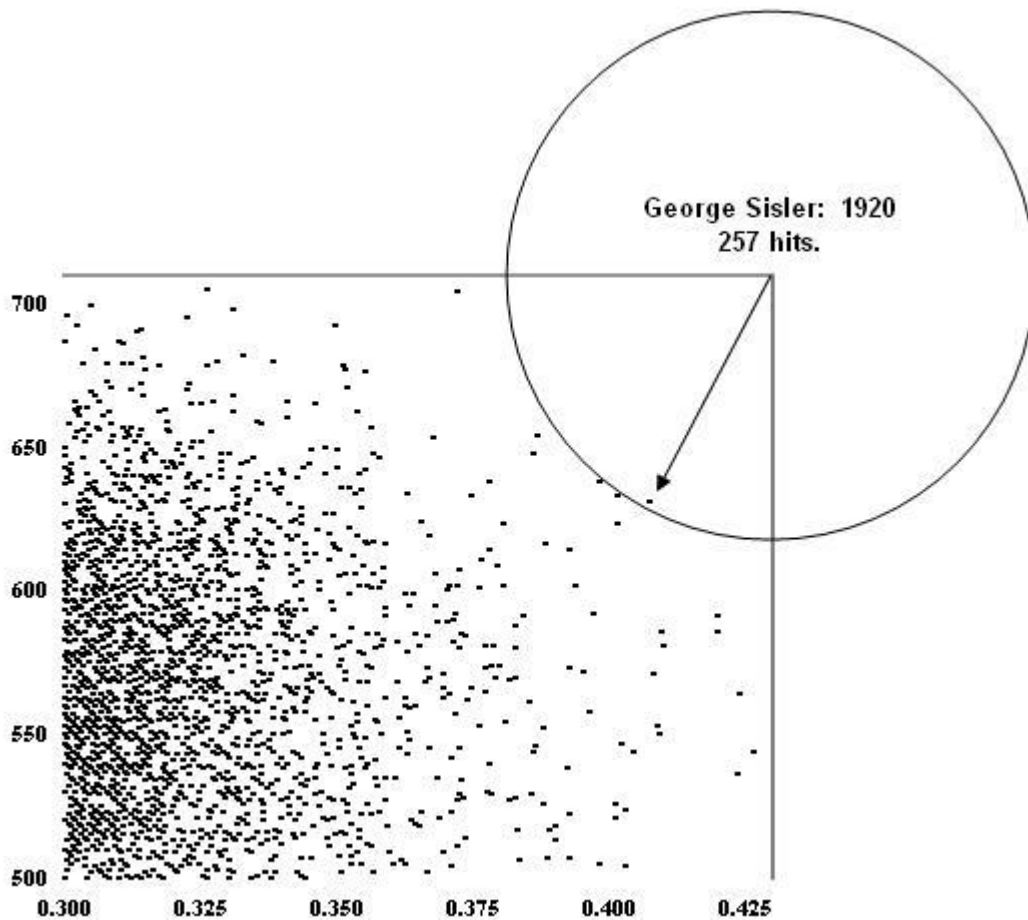
But are they the *best*?

In Search of a Quality Algorithm

One way to find out is "which point is in the 'upper-right-most' sector of the grid. One way to determine this is to establish a new grid, and see which is 'closest' to the intersection of these two axis.

Let's establish an "optimal" year as having a 430 batting average with 710 at-bats. No one has ever done both simultaneously, but both levels themselves have been approached separately. Perhaps a reasonably good "optimal" starting point.

Now, how can I establish which points are "closest"? I could calculate the distance by way of the Pythagorean Theorem - that would be pretty easy. Time-consuming as well. What if I simply draw a circle from the intersection of the two axis, and see which point I hit first - that would show me which point is closer to the optimal point than any other.



Indeed, the 1920 season of Sisler, coupling at-bats with average (producing 257 hits), *was* the greatest batting season of all time! What are some of the others?

THE TEN GREATEST HITTING SEASONS

MAJOR LEAGUE BASEBALL: 1871 - 2007

rank	year	Name	G	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	AVG	OBP	SLG
1	1920	Sisler	154	631	137	257	49	18	19	122	42	17	46	19	.407	.449	.632
2	1929	O'Doul	154	638	152	254	35	6	32	122	2		76	19	.398	.465	.622
3	1930	Terry	154	633	139	254	39	15	23	129	8		57	33	.401	.452	.619
4	1925	Simmons	153	654	122	253	43	12	24	129	7	14	35	41	.387	.419	.599
5	1930	Klein	156	648	158	250	59	8	40	170	4		54	50	.386	.436	.687
6	1922	Hornsby	154	623	141	250	46	14	42	152	17	12	65	50	.401	.459	.722
7	2004	Suzuki	161	704	101	262	24	5	8	60	36	11	49	63	.372	.414	.455
8	1928	Manush	154	638	104	241	47	20	13	108	17	5	39	14	.378	.414	.575
9	1930	Herman	153	614	143	241	48	11	35	130	18		66	56	.393	.455	.678
10	1977	Carew	155	616	128	239	38	16	14	100	23	13	69	55	.388	.449	.570

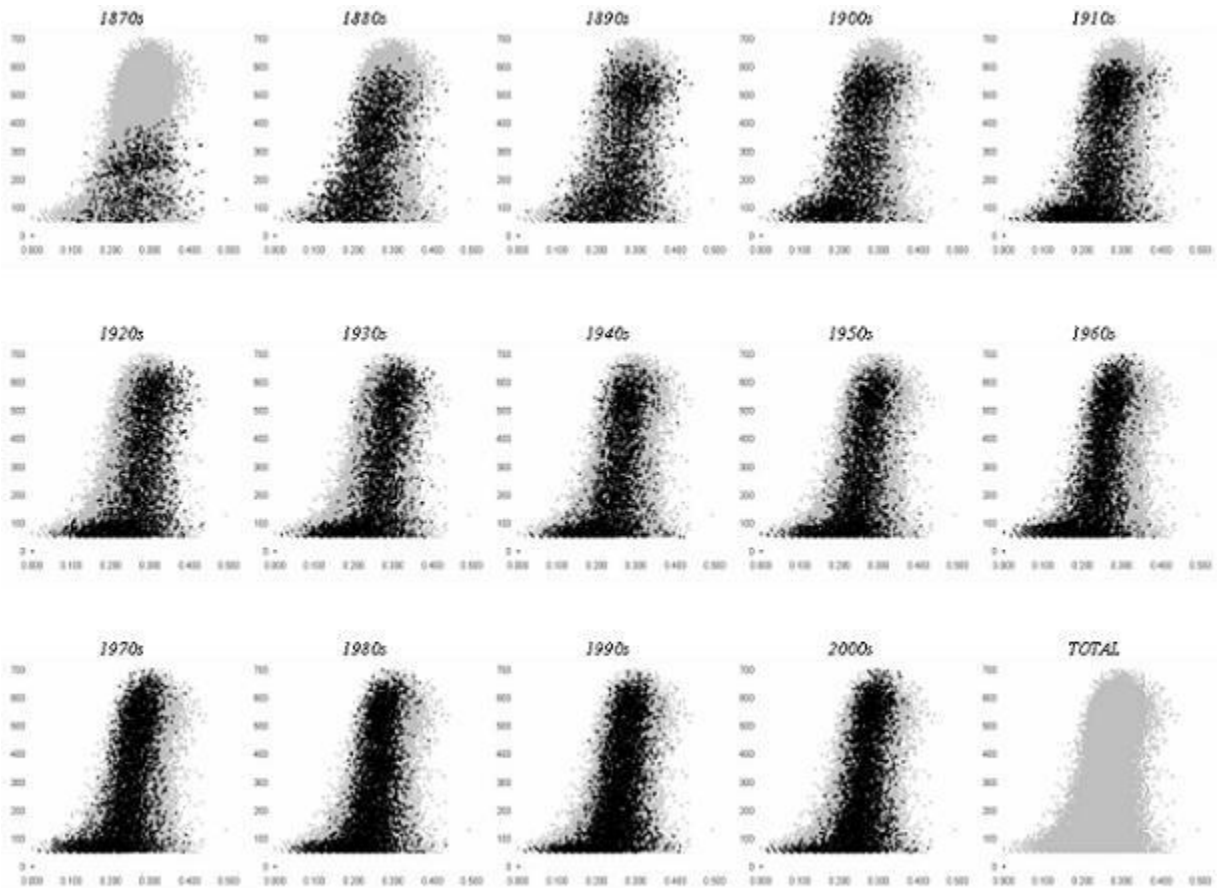
Sadly, Sisler missed the entire 1923 season with poisonous sinusitis. More sadly, despite a career average of .340 and 2,812 hits, he initially received only 34% of the votes for entry to the Hall-of-Fame (75% is required to be inducted), and it wasn't until the fourth try before he finally gained entrance to the Hall.

Surprisingly, those listed above as potential candidates for "greatest season" are nowhere on our table. Why? Those great seasons were the result of fewer at-bats than those seasons of the players in our table.

But does this mean "greatest season" is just a measure of "hits"? Partly - and why not? "Number of hits" is a good barometer of both hitting ability and season-longevity. The table above, however, is a blend of the two, as evidenced by the seventh-place ranking of Ichiro Suzuki and his phenomenal year of 2004, where he had 262 hits while batting .372!

By Decade

I've got all this data well-organized - what else can I do with it? Above, I looked at the history of baseball, lumping everything together. How do the individual *decades* compare? To make sure there is an available comparison, let's superimpose each decade with the totals above.

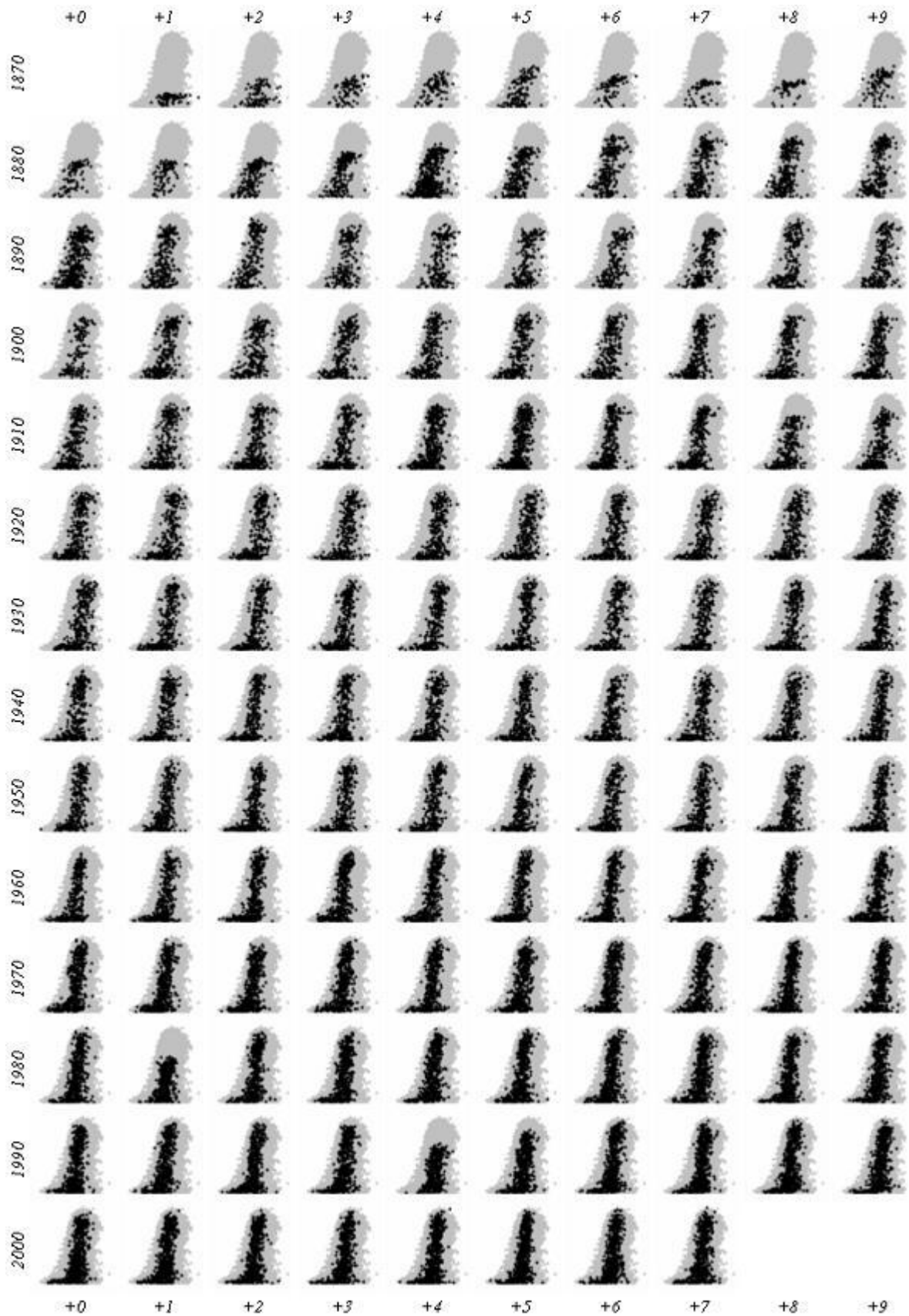


Notice the 60s: this was the "pitcher's decade", where Yastrzemski won the triple crown in 1968 with a meager 301 average. The 1920s, on the other hand, was a hitter's decade: compare the 20s with the 60s and you can easily see the number of "great hitters". You can also start to see the introduction of the 162-game schedule in the 60s.

Finally, the similarity of the decades, despite flickers of change, are remarkably similar.

Let's continue!

By Year



The strike-shortened seasons of 1981 and 1994 are obvious, as is the continuation of the self-similar nature of performance over more than a century!

Looking further at the "top-10" list, three of the ten are from 1930! What a spectacular hitting year that was - as is demonstrated in the graphic above. Compare 1930 with 1968, the year of the pitcher, to see remarkable differences.

Of course, there's many ways to describe "best hitting" season - on base %, slugging percentage, batting average, etc. In future editions of "Sports Forensics", I'll present an interesting graph of all three, which have Baby Ruth and Barry Bonds head-and-shoulders above the crowd.