# 데이터베이스의 활용

레먼 데이터베이스

### 레먼데이터베이스 소개

- 숀 레먼이 만든 MLB 데이터베이스
- 1871년부터 2022년까지의 MLB 데이터 보유(타격, 투구, 수비 등)
- 매시즌 종료 후 갱신
- 웹사이트: <a href="http://seanlahman.com/download-baseball-database/">http://seanlahman.com/download-baseball-database/</a>
- 우리는 이 수업에서 2019 시즌 데이터까지 쌓아둔 MySQL 버전의 데이터를 사용할 것

```
Through 2019 season

2019 - MS Access version
2019 - comma-delimited version
2019 - R Package

2019 - MySQL version
2019 - SQL Lite
```

# MySQL에 레먼데이터베이스 생성하기

#### MySQL에 레먼데이터베이스 생성하는 방법

- 1. <a href="https://drive.google.com/file/d/19vB07gSRPIVuxpjj6ZF\_kiujkknLQbS7/view?usp=drive\_link">https://drive.google.com/file/d/19vB07gSRPIVuxpjj6ZF\_kiujkknLQbS7/view?usp=drive\_link</a> 에서 lanman-mysql-dump.sql 다운로드하여 구름IDE에 업로드
- 2. 레먼데이터베이스 생성
  - mysql -u아이디 -p비밀번호 < lahman-mysql-dump.sql
- 3. 잘 생성되었는지 확인
  - mysql 접속: mysql -u아이디 -p비밀번호
  - 데이터베이스 목록 확인: show databases;
  - 레먼 데이터베이스 사용: use lahmansbaseballdb;
  - 테이블 목록 확인: show tables;
- 4. 이후 작업을 위해 DBeaver로 MySQL 서버에 접속

```
mvsql> show tables:
 Tables in lahmansbaseballdb
 allstarfull
 appearances
 awardsmanagers
 awardsplayers
 awardssharemanagers
 awardsshareplayers
 batting
 battingpost
 collegeplaying
 divisions
 fielding
 fieldingof
 fieldingofsplit
 fieldingpost
 halloffame
 homegames
 leagues
 managers
 managershalf
 parks
 people
 pitching
 pitchingpost
 salaries
 schools
 seriespost
 teams
```

### 테이블 목록

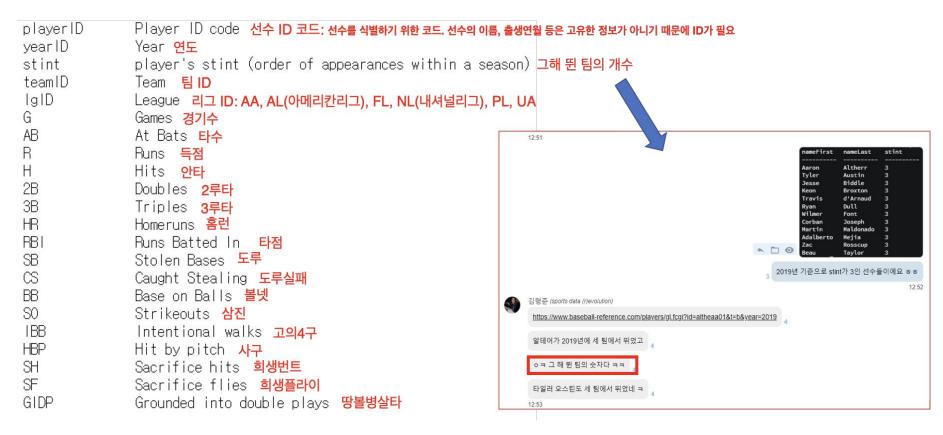
#### 주요 테이블

- People: 선수 메타 데이터
- Batting: 타자 데이터
- Pitching: 투수 데이터
- Fielding: 수비 데이터

전체 테이블--

AllStarFull - All-Star appearances HallofFame - Hall of Fame voting data Managers - managerial statistics Teams - yearly stats and standings BattingPost - post-season batting statistics PitchingPost - post-season pitching statistics TeamFranchises - franchise information FieldingOF - outfield position data FieldingPost-post-season fielding data FieldingOFsplit - LF/CF/RF splits ManagersHalf - split season data for managers TeamsHalf - split season data for teams Salaries - player salary data SeriesPost - post-season series information AwardsManagers - awards won by managers AwardsPlayers - awards won by players AwardsShareManagers - award voting for manager awards AwardsSharePlayers - award voting for player awards Appearances - details on the positions a player appeared at Schools - list of colleges that players attended CollegePlaying - list of players and the colleges they attended Parks - list of major league ballparls HomeGames - Number of homegames played by each team in each ballpark

# 테이블 명세 - Batting



# 테이블 명세 - Pitching

```
Player ID code
plaverID
vearID
                Year
stint
                player's stint (order of appearances within a season)
teamID
                 Team
IgID
                League
                Wins 승
                Losses III
                Games 경기수
GS
                Games Started선발 횟수
                Complete Games 완투
Shutouts 완봉
CG
SHO
                Saves 세이브
SV
                Outs Pitched (innings pitched x 3) 아웃카운트(이닝*3)
IPOuts
                Hits 피안타
                Earned Runs 자책점
                Homeruns <mark>홈런</mark>
Walks볼넷
HR
BB
                Strikeouts 삼진
S0
                Opponent's Batting Average <mark>상대타자 타율</mark>
Earned Run Average<mark>평균자책점</mark>
BA0pp
FRA
                Intentional Walks 고의4구
Wild Pitches 폭투
IBB
                Batters Hit By Pitch 사구
HBP
BK
                Balks 보크
                Batters faced by Pitcher 타석, 상대한 타자 수
GF
                Games Finished 마무리 경기수
                Runs Allowed 실점(자책 + 비자책)
Sacrifices by opposing batters 희생번트
SH
SF
                Sacrifice flies by opposing batters 희생플라이
GIDP
                Grounded into double plays by opposing batter병살타
```

# 테이블 명세 - Fielding

```
playerID
              Player ID code
vearID
              Year
stint
              player's stint (order of appearances within a season)
teamID
              Team
IgID
              League
              Position 포지션
Pos
              Games 경기수
GS
              Games Started 선발횟수
              Time played in the field expressed as outs 플레이시간(아웃카운트)
InnOuts
              Putouts 자살
Ρ0
              Assists 보살
              Errors 실책
              Double Plays <mark>더블플레이</mark>
PB
              Passed Balls (by catchers) 패스트볼(포일): 포수가 충분히 잡을 수 있는 공을 놓친 경우
WP
              Wild Pitches (by catchers)
SB
              Opponent Stolen Bases (by catchers)포수도루 허용
CS
              Opponents Caught Stealing (by catchers)포수 도루 저지
              Zone Rating ZR = 자기 구역에서 아웃 처리한 공 / 자기 구역에 떨어진 공
```

## 추신수 선수 메타 데이터 조회하기

#### WHERE 활용

#### 전략

- 1. 선수 메타 데이터이므로 People 테이블을 활용해야 함
- 2. 우리가 알고 있는 것은 추신수 선수의 이름이므로 이름을 활용하여 추신수 선수의 정보를 조회

#### SQL문 작성

#### **SELECT** \* **FROM** people **WHERE** nameLast = 'Choo';

	<b>№</b> playerID	-	123 birthYear	•	123 birthMonth		123 birthDay	•	RBC birthCountry	-	ABC birthState	•	ABC birthCity	•
1	choosh01		1,98	32		7		13	South Korea		Busan		Busan	

# Quiz8. MLB 한국 출생 선수 조회

다음과 같이 조회되게 할 것

조건

birthCountry 컬럼에

한국이 어떤 명칭으로

들어가 있는지

모른다고 가정하고 접근해들

		RBC nameFirst ▼	RBC nameLast 🔻	RBC birthCountry	RBC birthCity ▼	123 weight	123 height 🔻
<b>:</b>	1	Cha-Seung	Baek	South Korea	Busan	190	76
	2	Jung	Bong	South Korea	Seoul	175	75
1	3	Hee-Seop	Choi	South Korea	Hwasun	235	77
	4	Ji-Man	Choi	South Korea	Incheon	250	73
	5	Jin Ho	Cho	South Korea	Jeonju	175	72
	6	Shin-Soo	Choo	South Korea	Busan	210	71
	7	Jae-gyun	Hwang	South Korea	Seoul	215	72
	8	Jung Ho	Kang	South Korea	Gwangju	210	72
	9	Byung-Hyun	Kim	South Korea	Gwangju	176	71
	10	Hyun Soo	Kim	South Korea	Seoul	210	74
	11	Sun-Woo	Kim	South Korea	Incheon	180	74
	12	Dae-Sung	Koo	South Korea	Daejeon	185	73
╘	13	Dae-ho	Lee	South Korea	Busan	250	76
를	14	Sang-Hoon	Lee	South Korea	Seoul	190	73
	15	Chang-Yong	Lim	South Korea	Gwangju	175	71
	16	Seunghwan	Oh	South Korea	Jeongeup	205	70
	17	Byung Ho	Park	South Korea	Seoul	220	73
	18	Chan Ho	Park	South Korea	Gongju	210	74
	19	Tommy	Phelps	South Korea	Seoul	190	75
	20	Rob	Refsnyder	South Korea	Seoul	200	72
	21	Hyun-Jin	Ryu	South Korea	Incheon	255	75
	22	Jae Kuk	Ryu	South Korea	Choon Chung Do	210	75
	23	Jae Weong	Seo	South Korea	Gwangju	215	73

# 추신수 선수 시즌 총합 및 평균 볼넷, 삼진 개수 조회

#### 집계함수 활용

#### 전략

- 1. 타자 데이터이므로 Batting 테이블을 활용해야 함
- 타자 데이터에서 추신수 선수에 관한 정보는 playerID이므로, People 테이블에서 확인한 playerID를 가지고 조회해야 함
- 3. 집계함수 중 SUM, AVG를 활용해야 함

#### SQL문 작성

SELECT SUM(BB), AVG(BB), SUM(SO), AVG(SO) FROM batting WHERE playerID = 'choosh01';

	<u></u>	123 SUM(BB)	•	123 AVG(BB)	_	123 SUM(SO)	-	123 AVG(SO)	•
	1	8	355	53.43	375	1,	546	96.	625
1									

### 추신수 선수 전 시즌 [연도, 팀ID, 타율, 안타, 홈런] 데이터 조회하기

#### AS 활용

#### 전략

- 1. 타자 데이터이므로 Batting 테이블을 활용해야 함
- 타자 데이터에서 추신수 선수에 관한 정보는 playerID이므로, People 테이블에서 확인한 playerID를 가지고 조회해야 함
- 3. 타율 = 안타 / 타수
- 4. 타율은 AVG로 조회되도록 할 것

#### SQL문 작성

**SELECT** yearID, teamID, H/AB **AS AVG**, H, HR **FROM** batting **WHERE** playerID = 'choosh01';

	123 yearID 🔻	RBC teamID	123 AVG	•	123 H	•	123 HR	-
1	2,005	SEA	 0.05	556		1		0
2	2,006	SEA	0.09	909		1		0
3	2,006	CLE	0.29	945		43		3
4	2,007	CLE	0.29	941		5		0
5	2,008	CLE	0.30	091		98		14
6	2,009	CLE	0.30	002		175		20
7	2,010	CLE		0.3		165		22
8	2,011	CLE	0.25	88		81		8
9	2,012	CLE	0.28	326		169		16
10	2,013	CIN	0.28	347		162		21
11	2,014	TEX	0.24	418		110		13
12	2,015	TEX	0.27	757		153		22
13	2,016	TEX	0.24	416		43		7
14	2,017	TEX	0.2	261		142		22
15	2,018	TEX	0.26	343		148		21
16	2,019	TEX	0.26	647		149		24

### 추신수 선수 출루율 기준으로 시즌 기록 정렬하기

#### ORDER BY 활용

#### 전략

- 1. 타자 데이터이므로 Batting 테이블을 활용해야 함
- 타자 데이터에서 추신수 선수에 관한 정보는 playerID이므로, People 테이블에서 확인한 playerID를 가지고 조회해야 함
- 3. 출루율 = (안타 + 볼넷 + 사구) / (타수 + 볼넷 + 사구 + 희생플라이)
- 4. 출루율은 OBP로 조회되도록 할 것. On Base Percentage

#### SQL문 작성

**SELECT** yearID, teamID, (H + BB + HBP) / (AB + BB + HBP + SF) **AS** OBP **FROM** batting **WHERE** playerID = 'choosh01' **ORDER BY** OBP **DESC**;

	123 yearID 🔻	RBC teamID ▼	123 OBP ▼
1	2,013	CIN	0.4231
2	2,010	CLE	0.4009
3	2,008	CLE	0.3973
4	2,009	CLE	0.3942
5	2,018	TEX	0.3765
6	2,015	TEX	0.3748
7	2,006	CLE	0.3735
8	2,012	CLE	0.3732
9	2,019	TEX	0.3712
10	2,016	TEX	0.3571
11	2,017	TEX	0.357
12	2,007	CLE	0.35
13	2,011	CLE	0.3436
14	2,014	TEX	0.3403
15	2,005	SEA	0.1905
16	2,006	SEA	0.1667

## 추신수 선수 출루율 기준으로 시즌 기록 정렬하기 (확인)

#### https://www.baseball-reference.com/

Stan	dar	d E	Bat	ting	She	ow Mi	nors	Gam	ies b	y Po	ositio	<u>n</u>	Shar	e & I	Expo	rt ▼	Glos	sary	Hide	Parti	al Rows	3							
Year	Age	Tm	Lg	G	PA	АВ	R	н	2B	3B	HR	RBI	SB	cs	вв	so	ВА	ОВР	SLG	OPS	OPS+	тв	GDP	НВР	SH	SF	IBB	Pos	Awards
2013	30	CIN	NL	154	712	569	107	162	34	2	21	54	20	11	112	133	.285	.423	.462	.885	145	263	3	26	3	2	5	*8/7H	MVP-12
2010	27	CLE	AL	144	646	550	81	165	31	2	22	90	22	7	83	118	.300	.401	.484	.885	147	266	11	11	0	2	11	*9/D	MVP-14
2008	25	CLE	AL	94	370	317	68	98	28	3	14	66	4	3	44	78	.309	.397	.549	.946	151	174	5	5	0	4	4	97HD	
2009	26	CLE	AL	156	685	583	87	175	38	6	20	86	21	2	78	151	.300	.394	.489	.883	136	285	9	17	0	7	5	*97D/H8	
2018 *	35	TEX	AL	146	665	560	83	148	30	1	21	62	6	1	92	156	.264	.377	.434	.810	114	243	11	10	1	2	2	D97/H	AS
2015	32	TEX	<u>AL</u>	149	653	555	94	153	32	3	22	82	4	2	76	147	.276	.375	.463	.838	125	257	7	15	2	5	1	*9/HD	
2006	23	CLE	AL	45	167	146	23	43	11	3	3	22	5	3	18	46	.295	.373	.473	.846	119	69	2	1	1	1	2	9/7HD	
2012	29	CLE	AL	155	686	598	88	169	43	2	16	67	21	7	73	150	.283	.373	.441	.815	129	264	10	14	0	1	0	*9/H	
2019	36	TEX	AL	151	660	563	93	149	31	2	24	61	15	1	78	165	.265	.371	.455	.826	109	256	6	18	0	1	3	D97H	
2006	23	тот	AL	49	179	157	23	44	12	3	3	22	5	3	18	50	.280	.360	.452	.812	110	71	3	2	1	1	2	9/7H8D	
2016	33	TEX	AL	48	210	178	27	43	7	0	7	17	6	3	25	46	.242	.357	.399	.756	99	71	1	7	0	0	1	9/DH	
2017	34	TEX	AL	149	636	544	96	142	20	1	22	78	12	3	77	134	.261	.357	.423	.780	100	230	18	7	3	5	1	9D/H	
2007	24	CLE	AL	6	20	17	5	5	0	0	0	5	0	1	2	5	.294	.350	.294	.644	72	5	0	0	0	1	1	/7H9	
2011	28	CLE	AL	85	358	313	37	81	11	3	8	36	12	5	36	78	.259	.344	.390	.733	107	122	7	6	0	3	3	9	
2014	31	TEX	AL	123	529	455	58	110	19	1	13	40	3	4	58	131	.242	.340	.374	.714	100	170	9	12	0	4	3	7D9/H	
2020	37	TEX	AL	33	127	110	13	26	3	0	5	15	6	2	13	33	.236	.323	.400	.723	99	44	1	2	0	2	1	7D/H9	
2005	22	SEA	AL	10	21	18	1	1	0	0	0	1	0	0	3	4	.056	.190	.056	.246	-28	1	0	0	0	0	0	/8H	
2006	23	SEA	AL	4	12	11	0	1	1	0	0	0	0	0	0	4	.091	.167	.182	.348	-8	2	1	1	0	0	0	/8	

### MLB 2019 타율 TOP 5 조회하기

#### ORDER BY, LIMIT 활용

#### 전략

- 1. 타자 데이터이므로 Batting 테이블을 활용해야 함
- 타자 데이터에서 추신수 선수에 관한 정보는 playerID이므로, People 테이블에서 확인한 playerID를 가지고 조회해야 함
- 3. 타율 = 안타 / 타수
- 4. 타율과 같은 비율 스탯은 보통 규정타석를 채운 시즌의 기록만 놓고 비교.
  - a. MLB 규정타석 = 전체 경기수(162) \* 3.1, 소수 반올림 = 502타석
  - b. 타석 = 타수 + 볼넷 + 사구 + 희생번트 + 희생 플라이 + (타격방해) + (주루방해)

#### SQL문 작성

SELECT playerID, H/AB AS AVG FROM batting WHERE yearID = 2019 AND (AB + BB + HBP + SH + SF) >= 502

	REC playerID ▼	123 AVG ▼
1	☑ anderti01	0.3353
2	☑ yelicch01	0.3292
3	☑ marteke01	0.3286
4	☑ lemahdj01	0.3272
5	☑ rendoan01	0.3193

Rk	Name	Age	Tm	Lg	G	PA	AB	R	н	2B	3B	HR	RBI	SB	CS	ВВ	so	BA
1	Tim Anderson	26	CHW	AL	123	518	498	81	167	32	0	18	56	17	5	15	109	.33
2	Ketel Marte#	25	ARI	NL	144	628	569	97	187	36	9	32	92	10	2	53	86	.329
3	Christian Yelich*	27	MIL	NL	130	580	489	100	161	29	3	44	97	30	2	80	118	.329
4	DJ LeMahieu	30	NYY	AL	145	655	602	109	197	33	2	26	102	5	2	46	90	.32
5	Anthony Rendon	29	WSN	NL	146	646	545	117	174	44	3	34	126	5	1	80	86	.319

## Quiz9. 류현진과 커쇼 전 시즌 ERA 조회

#### 다음과 같이 조회되게 할 것

#### 조건

류현진 선수와 커쇼 선수의 playerID를 각각 people 테이블에서 찾은 후 pitching 테이블에서 ERA를 조회할 것

	메 playerID	•	123 yearID	•	123 ERA	•	123 W		123 L	•
1	☑ kershcl01		2,0	800		4.26		5		5
2	☑ kershcl01			009		2.79		8		8
3	☑ kershcl01		2,	010		2.91		13		10
4	☑ kershcl01		2	,011	2	2.28		21		5
5	☑ kershcl01		2,	012		2.53		14		9
6			2,	013		1.83		16		9
7			2,	014		1.77		21		3
8			2,	015		2.13		16		7
9			2,	016		1.69		12		4
10			2,	017		2.31		18		4
11			2,	018		2.73		9		5
12			2,	019	;	3.03		16		5
13	☑ ryuhy01		2,	013		3		14		8
14	☑ ryuhy01		2,	014		3.38		14		7
15	☑ ryuhy01		2,	016	1	1.57		0		1
16	☑ ryuhy01		2,	017		3.77		5		9
17	☑ ryuhy01		2,	018		1.97		7		3
18	☑ ryuhy01		2,	019		2.32		14		5

# 추신수 선수 전 시즌 홈런 기록 이름과 함께 조회(홈런 많은

'\_\_\_\_/ LEFT OUTER JOIN, ORDER BY 활용

#### 전략

- 1. Batting 테이블과 People 테이블을 LEFT OUTER JOIN 해야 함
- 2. 홈런 개수로 ORDER BY, DESC 해줘야 함

#### SQL문 작성

#### **SELECT**

batting.yearID, batting.teamID, batting.HR, people.nameFirst, people.nameLast

**FROM** batting

**LEFT OUTER JOIN** people

**ON** batting.playerID = people.playerID

WHERE batting.playerID = 'choosh01'

**ORDER BY** batting.HR **DESC**;

	123 yearID 🔻	RBC teamID 🔻	123 HR 🔻	RBC nameFirst ▼	RBC nameLast 🔻
1	2,019	TEX	24	Shin-Soo	Choo
2	2,010	CLE	22	Shin-Soo	Choo
3	2,015	TEX	22	Shin-Soo	Choo
4	2,017	TEX	22	Shin-Soo	Choo
5	2,013	CIN	21	Shin-Soo	Choo
6	2,018	TEX	21	Shin-Soo	Choo
7	2,009	CLE	20	Shin-Soo	Choo
8	2,012	CLE	16	Shin-Soo	Choo
9	2,008	CLE	14	Shin-Soo	Choo
10	2,014	TEX	13	Shin-Soo	Choo
11	2,011	CLE	8	Shin-Soo	Choo
12	2,016	TEX	7	Shin-Soo	Choo
13	2,006	CLE	3	Shin-Soo	Choo
14	2,005	SEA	0	Shin-Soo	Choo
15	2,006	SEA	0	Shin-Soo	Choo
16	2,007	CLE	0	Shin-Soo	Choo

### MLB에서 가장 많은 홈런을 친 선수 TOP 10 기록 조회하기

#### GROUP BY 활용

#### 전략

- 1. 타자 데이터이므로 Batting 테이블을 활용해야 함
- 2. playerID로 GROUP을 지은 후 해당 선수의 홈런의 개수를 모두 합해야 함
- 3. 상위 10명의 기록만 조회해야하므로 ORDER BY와 LIMIT를 활용

#### SQL문 작성

**SELECT** playerID, **SUM**(HR) **FROM** batting **GROUP BY** playerID **ORDER BY SUM**(HR) **DESC LIMIT** 10;

<ul><li>☑ bondsba01</li><li>☑ aaronha01</li><li>☑ ruthba01</li><li>☑ rodrial01</li></ul>	762 755 714
☑ ruthba01	714
rodrial01	
- Todridio i	696
☑ mayswi01	660
🗹 pujolal01	656
☑ griffke02	630
thomeji01	612
☑ sosasa01	609
☑ robinfr02	586
	☑ mayswi01 ☑ pujolal01 ☑ griffke02 ☑ thomeji01 ☑ sosasa01

### MLB에서 700 홈런 이상을 친 선수 홈런 기록 조회하기

#### GROUP BY, HAVING 활용

#### 전략

- 1. 타자 데이터이므로 Batting 테이블을 활용해야 함
- playerID로 GROUP을 지은 후 해당 선수의 홈런의 개수를 모두 합해야 함
- 3. 합한 홈런 개수가 **700**이상이라는 조건을 **HAVING**으로 줘야 함

#### SQL문 작성

SELECT playerID, SUM(HR) FROM batting GROUP BY playerID HAVING SUM(HR) >= 700 ORDER BY SUM(HR) DESC;

	REE playerID ▼	123 SUM(HR)	•
1	☑ bondsba01		762
2	🗹 aaronha01		755
3	🗹 ruthba01		714

### MLB에서 가장 많은 홈런을 친 선수 TOP 10 기록 선수 이름과 함께 조회하기

#### LEFT OUTER JOIN, 서브쿼리 활용

전략

people 테이블의 playerID와

batting 테이블의 playerID를 매칭하여 JOIN

#### SQL문 작성

#### **SELECT**

p.nameFirst, p.nameLast, t.TOTAL\_HR

**FROM** 

(SELECT playerID, SUM(HR) AS TOTAL\_HR FROM batting GROUP BY playerID ORDER BY TOTAL\_HR DESC LIMIT 10) AS t

**LEFT OUTER JOIN** people **AS** p

**ON** t.playerID = p.playerID;

	RBC nameFirst ▼	RBC nameLast ▼	123 TOTAL_HR ▼
1	Barry	Bonds	762
2	Hank	Aaron	755
3	Babe	Ruth	714
4	Alex	Rodriguez	696
5	Willie	Mays	660
6	Albert	Pujols	656
7	Ken	Griffey	630
8	Jim	Thome	612
9	Sammy	Sosa	609
10	Frank	Robinson	586

### Quiz10. MLB 2019년 투수 탈삼진 TOP 10 기록 이름과 함께 조회

#### 다음과 같이 조회되게 할 것

#### 조건

서브쿼리 사용하지 않고, LEFT OUTER JOIN만 활용할 것

	RBC nameFirst	•	RBC nameLast	•	123 TOTAL_SO	•
1	Gerrit	I	Cole			326
2	Justin		Verlander			300
3	Shane		Bieber			259
4	Jacob		deGrom			255
5	Trevor		Bauer			253
6	Stephen		Strasburg			251
7	Lance		Lynn			246
8	Max		Scherzer			243
9	Charlie		Morton			240
10	Patrick		Corbin			238