REPORT

Baseball MVP Analysis (1986 MLB Season)

Kaylie Le

April 21, 2025

Table of Content

Introduction

The report examines player performance in the 1986 MLB season using three core metrics: On-Base Percentage (OBP), Home Runs (HR), and Runs Batted In (RBI). The analysis aims to determine the most statistically deserving candidate for the Most Valuable Player (MVP) award.

This report reflects interpretations grounded in data and analytical reasoning, combining objective results with an evaluation of broader social and performance trends.

Key Findings

Baseball MVP Analysis (1986 MLB Season)

In the analysis of the 1986 Major League Baseball season, player performance was evaluated using three primary offensive metrics: On-Base Percentage (OBP), Home Runs (HR), and Runs Batted In (RBI). Each of these statistics offers a different perspective on a player's contribution to team offense, and together they provided a well-rounded view of overall performance.

The data showed that Wade Boggs excelled in OBP, demonstrating exceptional consistency at the plate, with 207 hits and 107 runs scored. Tim Raines also stood out for his on-base ability and base-stealing skills, leading the league with 70 stolen bases. However, while both players performed well in terms of reaching base and speed, they did not appear among the leaders in HR or RBI—two metrics that more directly reflect scoring impact.

In contrast, players like Jesse Barfield and Joe Carter displayed notable power. Barfield led the league with 40 home runs and contributed 108 RBIs. Carter posted a league-best 121 RBIs, alongside 29 home runs and 29 stolen bases, highlighting his all-around offensive capabilities.

That said, based on a combined view of all three metrics, Mike Schmidt emerged as the most consistently high-performing player. He ranked second in both home runs (37) and RBIs (119), while also maintaining a strong OBP. What made his performance even more impressive was his ability to deliver at a high level at age 36, showing both durability and leadership. For these reasons, the analysis strongly supports Mike Schmidt as the most deserving candidate for the 1986 MVP.

Baseball MVP Analysis 12

Download the baseball.csv data set that represents batting statistics from the 1986 Major League Baseball season. Read this data set in a variable called baseball.

```
baseball <- read_csv("baseball.csv")
```

The result is:

```
> baseball <- read_csv("baseball.csv")
Rows: 771 Columns: 16
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr  (2): Last, First
dbl (14): Age, G, PA, AB, R, H, 2B, 3B, HR, RBI, SB, CS, BB, SO

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
>
```

Spend time with the data using various exploration functions to get a general feel for what you are working with.

View the first 6 rows

```
head(baseball)
```

The result is:

```
> head(baseball)
# A tibble: 6 × 16
  Last    First   Age     G    PA    AB     R     H  `2B`  `3B`    HR   RBI    SB
  <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Acker   Jim      27    21    28    28     1     3     1     0     0     0     0
2 Adduci  Jim      26     3    13    11     2     1     1     0     0     0     0
3 Aguayo  Luis     27    62   146   133    17    28     6     1     4    13     1
4 Aguile… Rick     24    32    57    51     4     8     0     0     2     6     0
5 Akerfe… Darr…    24     1     0     0     0     0     0     0     0     0     0
6 Aldrete Mike     25    84   256   216    27    54    18     3     2    25     1
# ℹ 3 more variables: CS <dbl>, BB <dbl>, SO <dbl>
>
```

See structure and types

```
glimpse(baseball)
```

The result is

```
> glimpse(baseball)
Rows: 771
Columns: 16
$ Last  <chr> "Acker", "Adduci", "Aguayo", "Aguilera", "Akerfelds", "Aldrete", …
$ First <chr> "Jim", "Jim", "Luis", "Rick", "Darrel", "Mike", "Doyle", "Andy", …
$ Age   <dbl> 27, 26, 27, 24, 24, 25, 35, 24, 33, 27, 33, 25, 29, 22, 32, 23, 3…
$ G     <dbl> 21, 3, 62, 32, 1, 84, 18, 101, 102, 8, 48, 92, 15, 1, 121, 15, 12…
$ PA    <dbl> 28, 13, 146, 57, 0, 256, 45, 324, 230, 11, 7, 241, 12, 0, 453, 60…
$ AB    <dbl> 28, 11, 133, 51, 0, 216, 38, 293, 196, 11, 6, 216, 11, 0, 425, 55…
$ R     <dbl> 1, 2, 17, 4, 0, 27, 2, 30, 29, 0, 0, 31, 1, 0, 40, 9, 24, 0, 0, 6…
$ H     <dbl> 3, 1, 28, 8, 0, 54, 8, 66, 43, 1, 0, 53, 1, 0, 112, 20, 81, 0, 0,…
$ `2B`  <dbl> 1, 1, 6, 0, 0, 18, 1, 7, 7, 0, 0, 9, 0, 0, 21, 5, 15, 0, 0, 18, 0…
$ `3B`  <dbl> 0, 0, 1, 0, 0, 3, 0, 3, 2, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 2, 0, 0,…
$ HR    <dbl> 0, 0, 4, 2, 0, 2, 0, 1, 7, 0, 0, 1, 0, 0, 11, 0, 7, 0, 0, 1, 0, 4…
$ RBI   <dbl> 0, 0, 13, 6, 0, 25, 5, 29, 27, 0, 0, 15, 0, 0, 58, 7, 38, 0, 0, 2…
$ SB    <dbl> 0, 0, 1, 0, 0, 1, 0, 10, 11, 0, 0, 5, 0, 0, 0, 1, 1, 0, 0, 13, 0,…
$ CS    <dbl> 0, 0, 1, 0, 0, 3, 0, 1, 4, 0, 0, 1, 0, 0, 3, 2, 0, 0, 0, 7, 0, 1,…
$ BB    <dbl> 0, 1, 8, 3, 0, 33, 0, 14, 30, 0, 0, 22, 0, 0, 24, 3, 39, 2, 0, 36…
$ SO    <dbl> 21, 2, 26, 12, 0, 34, 8, 36, 38, 4, 3, 39, 4, 0, 77, 13, 56, 3, 0…
```

See column names

```
names(baseball)
```

The result is:

```
> names(baseball)
 [1] "Last"  "First" "Age"   "G"     "PA"    "AB"    "R"     "H"     "2B"
[10] "3B"    "HR"    "RBI"   "SB"    "CS"    "BB"    "SO"
>
```

Remove (filter) from baseball any player with 0 at bats (AB). Store the result in baseball

```
baseball <- baseball %>%
  filter(AB > 0)
```

The result is:

```
> baseball <- baseball %>%
+   filter(AB > 0)
>
```

Data

baseball | 726 obs. of 16 variables

Le_Project2.R* | baseball | best_freedom_df | top_ten_df | no_freedom_d

Filter

	Last	First	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB
1	Acker	Jim	27	21	28	28	1	3	1	0	0	0	
2	Adduci	Jim	26	3	13	11	2	1	1	0	0	0	
3	Aguayo	Luis	27	62	146	133	17	28	6	1	4	13	
4	Aguilera	Rick	24	32	57	51	4	8	0	0	2	6	
5	Aldrete	Mike	25	84	256	216	27	54	18	3	2	25	
6	Alexander	Doyle	35	18	45	38	2	8	1	0	0	5	
7	Allanson	Andy	24	101	324	293	30	66	7	3	1	29	
8	Almon	Bill	33	102	230	196	29	43	7	2	7	27	
9	Amelung	Ed	27	8	11	11	0	1	0	0	0	0	
10	Andersen	Larry	33	48	7	6	0	0	0	0	0	0	
11	Anderson	Dave	25	48	241	216	31	53	9	0	1	15	
12	Anderson	Rick	29	15	12	11	1	1	0	0	0	0	
13	Armas	Tony	32	121	453	425	40	112	21	4	11	58	
14	Asadoor	Randy	23	15	60	55	9	20	5	0	0	7	

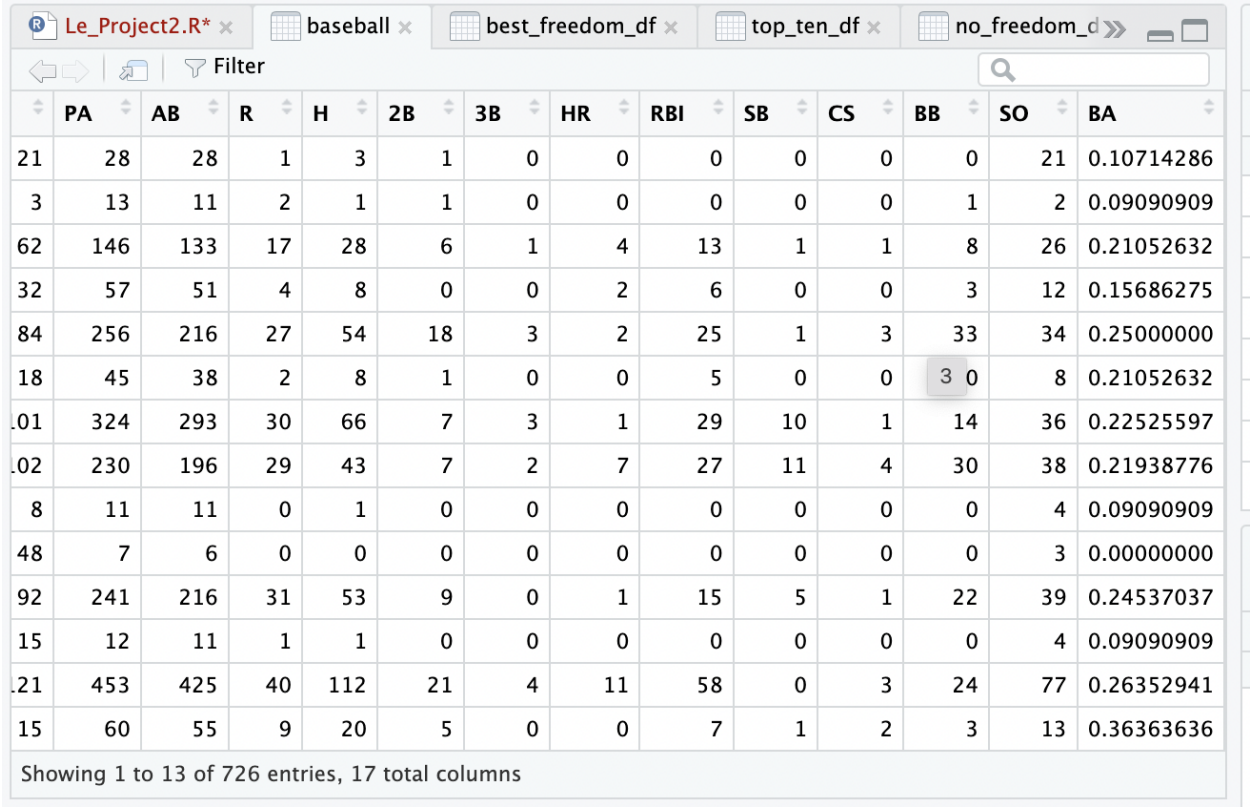
Showing 1 to 14 of 726 entries, 16 total columns

Add a new column batting average called BA. Batting average is computed by the number of hits (H) divided by the number of at bats (AB). Store the result in baseball.

```
baseball <- baseball %>%
  mutate(BA = H / AB)
```

The result is:

```
> baseball <- baseball %>%
+   mutate(BA = H / AB)
>
```



Le_Project2.R* | baseball | best_freedom_df | top_ten_df | no_freedom_d

Filter

	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA
21	28	28	1	3	1	0	0	0	0	0	0	21	0.10714286
3	13	11	2	1	1	0	0	0	0	0	1	2	0.09090909
62	146	133	17	28	6	1	4	13	1	1	8	26	0.21052632
32	57	51	4	8	0	0	2	6	0	0	3	12	0.15686275
84	256	216	27	54	18	3	2	25	1	3	33	34	0.25000000
18	45	38	2	8	1	0	0	5	0	0	3	0	8	0.21052632
.01	324	293	30	66	7	3	1	29	10	1	14	36	0.22525597
.02	230	196	29	43	7	2	7	27	11	4	30	38	0.21938776
8	11	11	0	1	0	0	0	0	0	0	0	4	0.09090909
48	7	6	0	0	0	0	0	0	0	0	0	3	0.00000000
92	241	216	31	53	9	0	1	15	5	1	22	39	0.24537037
15	12	11	1	1	0	0	0	0	0	0	0	4	0.09090909
.21	453	425	40	112	21	4	11	58	0	3	24	77	0.26352941
15	60	55	9	20	5	0	0	7	1	2	3	13	0.36363636

Showing 1 to 13 of 726 entries, 17 total columns

On-base percentage (OBP) is arguably a better statistic than batting average. Create a column called OBP that computes this stat as (H + BB) / (AB + BB). Store the result in baseball.

```
baseball <- baseball %>%
  mutate(OBP = (H + BB) / (AB + BB))
```

The result is:

```
> baseball <- baseball %>%
+   mutate(BA = H / AB)
> baseball <- baseball %>%
+   mutate(OBP = (H + BB) / (AB + BB))
>
```

Le_Project2.R* | baseball | best_freedom_df | top_ten_df | no_freedom_d

Filter

AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP
28	1	3	1	0	0	0	0	0	0	21	0.10714286	0.10714286
11	2	1	1	0	0	0	0	0	1	2	0.09090909	0.16666667
133	17	28	6	1	4	13	1	1	8	26	0.21052632	0.25531915
51	4	8	0	0	2	6	0	0	3	12	0.15686275	0.20370370
216	27	54	18	3	2	25	1	0	33	34	0.25000000	0.34939759
38	2	8	1	0	0	5	0	0	0	8	0.21052632	0.21052632
293	30	66	7	3	1	29	10	1	14	36	0.22525597	0.26058632
196	29	43	7	2	7	27	11	4	30	38	0.21938776	0.32300885
11	0	1	0	0	0	0	0	0	0	4	0.09090909	0.09090909
6	0	0	0	0	0	0	0	0	0	3	0.00000000	0.00000000
216	31	53	9	0	1	15	5	1	22	39	0.24537037	0.31512605
11	1	1	0	0	0	0	0	0	0	4	0.09090909	0.09090909
425	40	112	21	4	11	58	0	3	24	77	0.26352941	0.30289532
55	9	20	5	0	0	7	1	2	3	13	0.36363636	0.39655172

Showing 1 to 13 of 726 entries, 18 total columns

Determine the 10 players who struck out the most this season. Store these results as strikeout_artist.

```
strikeout_artist <- baseball %>%
  arrange(desc(SO)) %>%
  slice(1:10)
```

The result is:

```
> strikeout_artist <- baseball %>%
+   arrange(desc(SO)) %>%
+   slice(1:10)
>
```

strikeout_artist | 10 obs. of 18 variables

Le_Project2.R* | strikeout_artist | baseball | best_freedom_df | top_ten_d

Filter

	Last	First	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB
1	Incaviglia	Pete	22	153	606	540	82	135	21	2	30	88	
2	Deer	Rob	25	134	546	466	75	108	17	3	33	86	
3	Canseco	Jose	21	157	682	600	85	144	29	1	33	117	1
4	Presley	Jim	24	155	660	616	83	163	33	4	27	107	
5	Tartabull	Danny	23	137	578	511	76	138	25	6	25	96	
6	Balboni	Steve	29	138	562	512	54	117	25	1	29	88	
7	Barfield	Jesse	26	158	671	589	107	170	35	2	40	108	
8	Samuel	Juan	25	145	633	591	90	157	36	12	16	78	4
9	Murphy	Dale	30	160	692	614	89	163	29	7	29	83	
10	Strawberry	Darryl	24	136	562	475	76	123	27	5	27	93	2

To be eligible for end-of-season awards, a player must have either at least 300 at bats or appear in at least 100 games. Keep only the players who are eligible to be considered and store them in a variable called eligible_df.

```
eligible_df <- baseball %>%
  filter(AB >= 300 | G >= 100)
```

The result is:

```
> eligible_df <- baseball %>%
+   filter(AB >= 300 | G >= 100)
```

eligible_df | 251 obs. of 18 variables

Le_Project2.R* | eligible_df | strikeout_artist | baseball | best_freedom_d

Filter

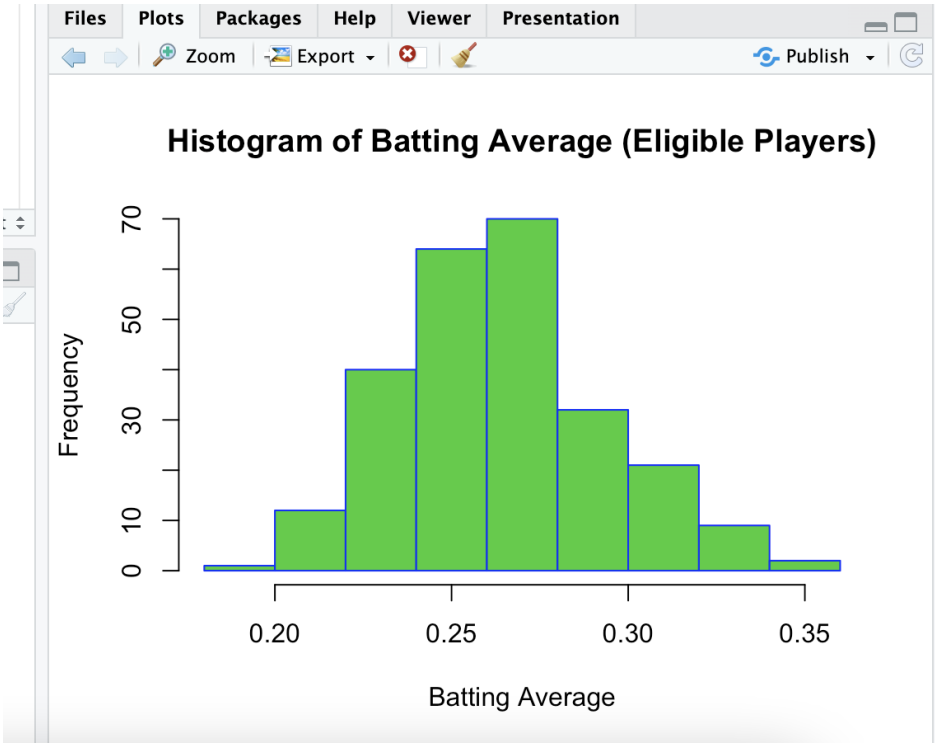
	Last	First	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB
1	Allanson	Andy	24	101	324	293	30	66	7	3	1	29	
2	Almon	Bill	33	102	230	196	29	43	7	2	7	27	
3	Armas	Tony	32	121	453	425	40	112	21	4	11	58	
4	Ashby	Alan	34	120	361	315	24	81	15	0	7	38	
5	Backman	Wally	26	124	440	387	67	124	18	2	1	27	
6	Baines	Harold	27	145	618	570	72	169	29	2	21	88	
7	Balboni	Steve	29	138	562	512	54	117	25	1	29	88	
8	Barfield	Jesse	26	158	671	589	107	170	35	2	40	108	
9	Barrett	Marty	28	158	713	625	94	179	39	4	4	60	
10	Bass	Kevin	27	157	640	591	83	184	33	5	20	79	
11	Baylor	Don	37	160	687	585	93	139	23	1	31	94	
12	Bell	Buddy	34	155	655	568	89	158	29	3	20	75	
13	Bell	George	26	159	690	641	101	198	38	6	31	108	
14	Belliard	Rafael	24	117	350	309	33	72	5	2	0	31	

Showing 1 to 13 of 251 entries, 18 total columns

For eligible players, create a histogram of batting average.

```
hist(eligible_df$BA,
     main = "Histogram of Batting Average (Eligible Players)",
     xlab = "Batting Average",
     col = "limegreen", border = "blue")
```

The result is:



Choose MVP

Step 1: Sort players by OBP (highest first)

```
eligible_df %>%
  arrange(desc(OBP)) %>%
  slice(1:10)
```

```
# A tibble: 10 × 18
   Last    First   Age     G    PA    AB     R     H  `2B`  `3B`    HR   RBI    SB
   <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Boggs   Wade     28   149   693   580   107   207    47     2     8    71     0
 2 Raines  Tim      26   151   664   580    91   194    35    10     9    62    70
 3 Herna…  Keith    32   149   652   551    94   171    34     1    13    83     2
 4 Kruk    John     25   122   327   278    33    86    16     2     4    38     2
 5 Hassey  Ron      33   113   393   341    45   110    25     1     9    49     1
 6 Phelps  Ken      31   125   441   344    69    85    16     4    24    64     2
 7 Murray  Eddie    30   137   578   495    61   151    25     1    17    84     3
 8 Brett   Geor…    33   124   529   441    70   128    28     4    16    73     1
 9 Matti…  Don      25   162   742   677   117   238    53     2    31   113     0
10 Bradl…  Phil     27   143   615   526    88   163    27     4    12    50    21
# ℹ 5 more variables: CS <dbl>, BB <dbl>, SO <dbl>, BA <dbl>, OBP <dbl>
```

This metric shows how often a player reaches base, which is essential for scoring and creating opportunities.

- Wade Boggs leads this category with an extremely high OBP. He had 207 hits, 107 runs, and 71 RBIs, making him an excellent table-setter for his team.
- Tim Raines also stands out with a high OBP and an outstanding 70 stolen bases, showing his impact on the bases.
- Keith Hernandez contributes with both high OBP and 83 RBIs, showing balance between getting on base and driving in runs.

Sort players by HR (highest first)

```
eligible_df %>%
  arrange(desc(HR)) %>%
  slice(1:10)
```

	Last	First	Age	G	PA	AB	R	H	`2B`	`3B`	HR	RBI	SB
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Barfi…	Jesse	26	158	671	589	107	170	35	2	40	108	8
2	Schmi…	Mike	36	160	657	552	97	160	29	1	37	119	1
3	Kingm…	Dave	37	144	604	561	70	118	19	0	35	94	3
4	Gaetti	Gary	27	157	661	596	91	171	34	1	34	108	14
5	Canse…	Jose	21	157	682	600	85	144	29	1	33	117	15
6	Deer	Rob	25	134	546	466	75	108	17	3	33	86	5
7	Baylor	Don	37	160	687	585	93	139	23	1	31	94	3
8	Bell	Geor…	26	159	690	641	101	198	38	6	31	108	7
9	Davis	Glenn	25	158	654	574	91	152	32	3	31	101	3
10	Matti…	Don	25	162	742	677	117	238	53	2	31	113	0

```
# ℹ 5 more variables: CS <dbl>, BB <dbl>, SO <dbl>, BA <dbl>, OBP <dbl>
```

This category highlights players who generate offense through power.

- Jesse Barfield led with 40 HR and had 108 RBIs, showing both power and run production.
- Mike Schmidt followed with 37 HR and an impressive 119 RBIs, appearing again as a high-impact player.
- Dave Kingman had 35 HR but fewer RBIs (94) and lower other metrics compared to others.

Step 3: *Sort players by RBI (highest first)*

```
eligible_df %>%
  arrange(desc(RBI)) %>%
  slice(1:10)
```

```
# A tibble: 10 × 18
   Last    First   Age     G    PA    AB     R     H  `2B`  `3B`    HR   RBI    SB
   <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Carter  Joe      26   162   709   663   108   200    36     9    29   121    29
 2 Schmi…  Mike     36   160   657   552    97   160    29     1    37   119     1
 3 Canse…  Jose     21   157   682   600    85   144    29     1    33   117    15
 4 Parker  Dave     35   162   700   637    89   174    31     3    31   116     1
 5 Matti…  Don      25   162   742   677   117   238    53     2    31   113     0
 6 Rice    Jim      33   157   693   618    98   200    39     2    20   110     0
 7 Barfi…  Jesse    26   158   671   589   107   170    35     2    40   108     8
 8 Bell    Geor…    26   159   690   641   101   198    38     6    31   108     7
 9 Gaetti  Gary     27   157   661   596    91   171    34     1    34   108    14
10 Presl…  Jim      24   155   660   616    83   163    33     4    27   107     0
# ℹ 5 more variables: CS <dbl>, BB <dbl>, SO <dbl>, BA <dbl>, OBP <dbl>
```

This shows how effective a player is in scoring teammates, a direct measure of offensive impact.

RBIs reflect a player's ability to produce runs and support team scoring.

- Joe Carter leads with 121 RBIs, along with 29 HR and 29 SB, showing an all-around offensive threat.
- Mike Schmidt ranks second with 119 RBIs and 37 HR, proving his consistent offensive output.
- Jose Canseco contributes 117 RBIs and 33 HR, reflecting his power and run-driving ability.

Step 4: Conclusion

To choose the MVP, I would consider below perspectives:

- Wade Boggs and Tim Raines are elite in OBP and speed but lack home run power and RBI dominance.
- Joe Carter is balanced in all areas but doesn't lead in multiple categories.
- Mike Schmidt appears in both the HR and RBI top 2, proving both power and run production.
- At age 36, Schmidt maintains elite stats, showing leadership and longevity.

Therefore, Mike Schmidt consistently appears in the top rankings across key categories (HR and RBI), proving his power, productivity, and value to the team. His stats show that he not only hits for power but also delivers when it matters most — by driving in runs.

Conclusion and Recommendations

This project showed how exploring data can uncover meaningful patterns, even in subjects as different as global happiness and professional baseball. By digging into the numbers, clear insights emerged—such as which parts of the world report the highest levels of happiness, and which player stood out the most in the 1986 MLB season.

The analysis suggests that when it comes to understanding well-being, policymakers shouldn't focus solely on economic data. Social factors like freedom, support from family, and generosity also play a major role and deserve more attention. In the same way, looking at a range of baseball statistics led to a well-rounded choice for MVP. Based on the data, Mike Schmidt stood out across all key performance categories and should be recognized as the Most Valuable Player of 1986.

References

Baseball Reference. (1986). *Major League Baseball standard batting statistics*. https://www.baseball-reference.com/leagues/majors/1986-standard-batting.shtml