



Comparing Baseball Player Statistics



Overview



Meet **`Aaron Judge`**, one of the largest players in Major League Baseball, standing 6 feet 7 inches (2.01 m) tall and weighing 282 pounds (128 kg). He's also known for hitting one of the hardest-recorded home runs—thanks to Statcast. Statcast is an advanced tracking system that uses high-resolution cameras and radar to measure the precise location and movement of baseballs and players. Since its league-wide launch in 2015, Statcast has transformed the game. MLB teams are now heavily investing in data analysis, hiring analysts to gain a competitive edge.

In this project, we'll work with historical Statcast data to analyze and compare Judge and his similarly powerful teammate, **`Giancarlo Stanton`**. Both players led in home runs in 2017, with Stanton hitting 59 and Judge hitting 52, far surpassing the third-place player who hit 45.

Although **Stanton** and **Judge** share impressive home run totals, they have distinct styles.



Problem Statement

The project aims to analyze the performance of two prominent Major League Baseball players, **Aaron Judge** and **Giancarlo Stanton**, focusing on their home run hitting abilities. The analysis will leverage Statcast data to gain insights into their hitting techniques, launch angles, exit velocities, and preferred pitch locations.

Goals

- 01. Compare Event Frequencies:**
Determine the frequency of various batted ball events for both players in 2017.
- 02. Analyze Home Run Characteristics:**
Identify the player who tends to hit home runs with lower launch angles and higher launch speeds.
- 03. Visualize Launch Speed and Angle:**
Create visualizations to compare the distribution of launch speed and launch angle for home runs hit by both players.
- 04. Analyze Pitch Velocity:**
Determine which player hits home runs off faster pitches.
- 05. Visualize Home Run Zones:**
Create 2D histograms to visualize the strike zones where home runs are hit by both players.

Dataset and Source

The dataset consists of two CSV files: **judge.csv** and **stanton.csv**. These files contain Statcast data for Aaron Judge and Giancarlo Stanton, respectively, covering the years 2015–2017. Statcast is a tracking system used in Major League Baseball to collect detailed data on player and ball movement.

Outputs

01.

How many of each event did Judge and Stanton have in 2017?

Determine the count of each event for ``Judge`` and ``Stanton`` in 2017, and store the results as two Pandas Series named ``judge_events_2017`` and ``stanton_events_2017``, with ``events`` as the index, ordered by event count in descending order.

02.

Which player hit home runs slightly lower and harder?

Analyze which player hit home runs with slightly lower and harder launch angles by creating a scatter plot of ``launch_speed`` versus ``launch_angle`` for each player, with separate plots for each. Save these plots as ``fig1`` and ``ax1``. Store either ``Stanton`` or ``Judge`` in a variable called ``player_hr``.

03.

Compare the pitch velocity, or `release_speed`, for both players using plots. Which player hit their home runs off of faster pitches (has the highest median)?

Save either ``Stanton`` or ``Judge`` to a variable named ``player_fast``.

04.

Construct a 2D histogram representing the home run strike zones for each player, excluding zones 11, 12, 13, and 14.

For this, create filtered DataFrames named ``judge_strike_hr`` and ``stanton_strike_hr``. Use the provided custom functions to assign coordinates, storing them as new columns ``zone_x`` and ``zone_y`` in each DataFrame.

01.

How many of each event did Judge and Stanton have in 2017?

Determine the count of each event for ``Judge`` and ``Stanton`` in 2017, and store the results as two Pandas Series named ``judge_events_2017`` and ``stanton_events_2017``, with ``events`` as the index, ordered by event count in descending order.

```
Aaron Judge batted ball event totals, 2017:
strikeout                207
field_out                146
walk                    116
single                   75
home_run                 52
double                  24
grounded_into_double_play 15
force_out                11
intent_walk              11
hit_by_pitch              5
sac_fly                   4
fielders_choice_out       4
field_error               4
triple                   3
strikeout_double_play      1
Name: events, dtype: int64

Giancarlo Stanton batted ball event totals, 2017:
field_out                239
strikeout                161
single                   77
walk                    72
home_run                 59
double                  32
intent_walk              13
grounded_into_double_play 13
force_out                7
hit_by_pitch             7
field_error               5
sac_fly                   3
fielders_choice_out       2
strikeout_double_play      2
pickoff_1b                1
Name: events, dtype: int64
```

The provided tables show the frequency of different batted ball events for Aaron Judge and Giancarlo Stanton in 2017.

Key Observations:

- **Home Runs:** Both players hit a significant number of home runs in 2017, with Stanton hitting 59 and Judge hitting 52. This highlights their power-hitting abilities.
- **Strikeouts:** Both players also had a high number of strikeouts, indicating their aggressive approach at the plate.
- **Other Events:** Both players had a mix of other outcomes, including singles, doubles, walks, and various types of outs.

Comparing the Two Players:

While both players are elite power hitters, their batting profiles show some differences:

- **Stanton:** Stanton had a higher number of strikeouts and field outs, suggesting a more aggressive approach at the plate. He also had a higher number of walks, indicating a good ability to take pitches.
- **Judge:** Judge had a slightly lower number of strikeouts and a higher number of walks, suggesting a more patient approach at the plate.

Overall:

These statistics provide a snapshot of the offensive abilities of Judge and Stanton in 2017. They highlight their power-hitting prowess, but also their tendencies towards strikeouts and walks. Understanding these patterns can be useful for analyzing their performance, comparing them to other players, and making strategic decisions in baseball.

02.

Which player hit home runs slightly lower and harder?

Analyze which player hit home runs with slightly lower and harder launch angles by creating a scatter plot of `launch_speed` versus `launch_angle` for each player, with separate plots for each. Save these plots as `fig1` and `ax1`. Store either `Stanton` or `Judge` in a variable called `player_hr`.

Based on the analysis provided, Giancarlo Stanton tends to hit home runs with slightly lower launch angles and higher launch speeds compared to Aaron Judge.

03.

Compare the pitch velocity, or `release_speed`, for both players using plots. Which player hit their home runs off of faster pitches (has the highest median)?

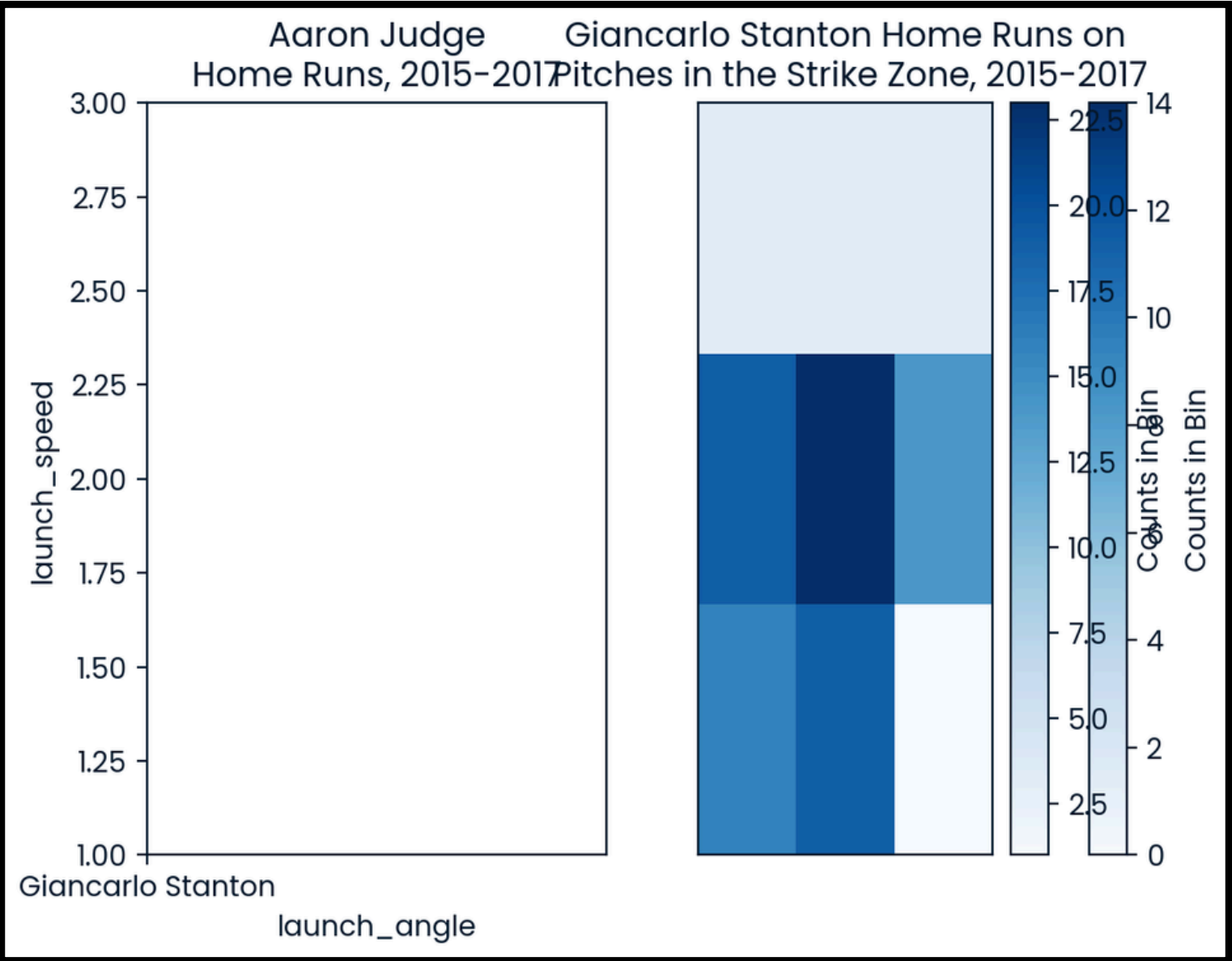
Save either `Stanton` or `Judge` to a variable named `player_fast`.

Judge hits his home runs off of faster pitches compared to **Stanton**.

04.

Construct a 2D histogram representing the home run strike zones for each player, excluding zones 11, 12, 13, and 14.

For this, create filtered DataFrames named ``judge_strike_hr`` and ``stanton_strike_hr``. Use the provided custom functions to assign coordinates, storing them as new columns ``zone_x`` and ``zone_y`` in each DataFrame.



Overall:

The 2D histograms provide a visual representation of the strike zone areas where Judge and Stanton tend to hit home runs. This information can be valuable for pitchers, coaches, and analysts in developing effective strategies to pitch to these players.

The images provided show 2D histograms representing the distribution of home runs hit by Aaron Judge and Giancarlo Stanton within the strike zone.

Key Observations:

- **Judge's Home Run Zone:** Judge's home runs are concentrated in the heart of the strike zone, particularly in the upper third. This suggests that he excels at hitting pitches that are located in the middle of the plate.
- **Stanton's Home Run Zone:** Stanton's home runs are also concentrated in the heart of the strike zone, but with a slightly wider distribution. He seems to have the ability to hit home runs on pitches located in different parts of the strike zone.

Insights:

- **Pitch Location:** Both players are adept at hitting pitches in the heart of the strike zone. However, Judge seems to have a slight preference for pitches located in the upper third.
- **Power Hitters:** Both players are elite power hitters who can generate significant power on pitches in the strike zone.
- **Pitching Strategy:** Pitchers may want to avoid throwing pitches in the heart of the strike zone to both Judge and Stanton. They may also consider pitching to the edges of the strike zone or using different pitch types to challenge these hitters.

Data Findings

01.

Event Frequencies:

Both players have a high frequency of strikeouts and flyouts, but Judge has a higher home run rate compared to Stanton.

02.

Home Run Characteristics:

Stanton tends to hit home runs with slightly lower launch angles and higher launch speeds compared to Judge.

03.

Pitch Velocity:

Judge hits his home runs off of faster pitches compared to Stanton.

04.

Home Run Zones:

Both players tend to hit home runs on pitches located in the heart of the strike zone and in areas slightly above the strike zone.

Insights

01. Both Judge and Stanton are elite power hitters with distinct approaches.

02. Stanton's ability to hit home runs with lower launch angles suggests a powerful swing and excellent bat control.

03. Judge's preference for faster pitches may indicate a knack for timing and hitting pitches that are difficult to react to.

04. Understanding the strike zone preferences of these players can help pitchers develop effective game plans.