Baseball Pitches and Their Outcomes

Susan Orgera

May 12, 2025

Introduction

Baseball is a beloved sport across the world. Like most human endeavors in this data-driven era, statistical analysis has become a significant component of the baseball experience. It was in fact one of the first sports to use analytics [2].

Within the plethora of baseball statistics, this report looked at data associated with pitching and its outcomes in the 2024 MLB season. This includes basic summary statistics and an exploration of relationships between pitch type (e.g. four-seam fastball) and other pitching-related variables (e.g. strikeouts). Also, a selection of physics-related pitching variables were analyzed to confirm the aerodynamic theory of the Magnus effect, using the 2024 season's actual data. Let's begin by looking at some basic physics concepts related to pitching.

Pitching Physics

Pitching physics involves kinematics, dynamics, and aerodynamics. Figure 1 shows the various forces and angular components that affect a ball during flight. The following section provides an overview of some of the equations describing the forces and torques during a pitch [4]. Of particular interest is the Magnus effect.

The Magnus effect is a sideways force acting on a spinning object moving through a fluid, such as air or water [3]. In baseball, it explains how the spin of a pitch alters airflow around the ball, generating lift or sideways movement that causes the ball to rise, drop, or curve. This occurs because the spin creates a pressure differential: air moves faster on one side of the ball and slower on the other, according to Bernoulli's principle. The resulting pressure imbalance produces a force perpendicular to both the spin axis and the direction of motion.

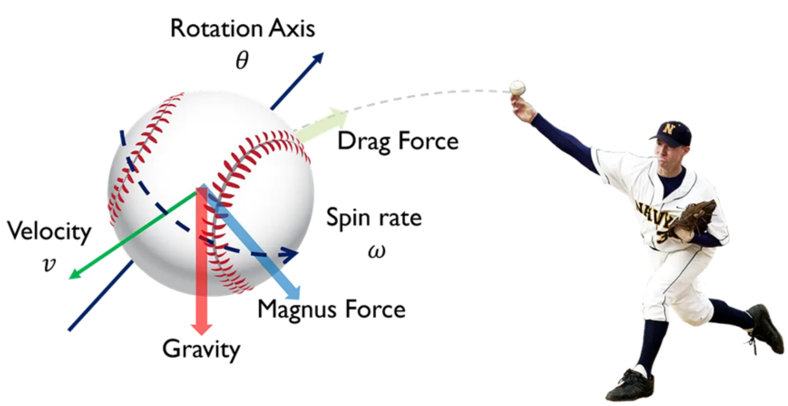

Figure 1: The forces acting on a baseball.

There are several formulas that describe the physical components of a pitch:

1. Newton's Second Law

The motion of the baseball during the pitch is governed by Newton's second law:

$$\vec{F}_{\text{net}} = m\vec{a} \tag{1}$$

Where:

- m is the mass of the baseball ($0.145\,\text{kg}$),
- $\vec{a}$ is the acceleration vector,
- $\vec{F}_{\text{net}}$ is the net force acting on the ball.

2. Torque and Arm Motion

The pitcher's arm applies torque to accelerate the ball:

$$\tau = I\alpha \tag{2}$$

Where:

- τ is the torque,
- I is the moment of inertia of the arm + ball system,
- α is the angular acceleration.

3. Air Resistance (Drag Force)

While the ball travels toward the plate, it experiences air resistance:

$$F_D = \frac{1}{2} C_D \rho A v^2 \tag{3}$$

Where:

- C_D is the drag coefficient (typically ~ 0.3–0.5 for a baseball),
- ρ is the air density ($1.225\,\text{kg/m}^3$),
- A is the cross-sectional area of the baseball,
- v is the velocity of the baseball.

4. Kinematic Equation (Release to Plate)

Assuming constant deceleration due to drag:

$$x(t) = v_0 t - \frac{1}{2} a t^2 \tag{4}$$

5. Magnus Force (Spin-Induced Lift)

Aerodynamic Form

$$F_{\text{Magnus}} = \frac{1}{2}\rho A C_L v^2 \tag{5}$$

Where:

- ρ: air density (kg/m^3)
- A: cross-sectional area of the baseball (m^2)
- C_L: lift coefficient (function of spin)
- v: velocity of the ball (m/s)

Methods

Data for a subset of pitches of the 2024 MLB season were downloaded from https://baseballsavant.mlb.com/ into a csv file. (The intention was to download every pitch of the season but it is unclear what filters prevented this from happening. This was discovered too late to redo!) Using PositCloud (a cloud-based platform that provides a browser-based environment for data science tasks), the file was uploaded into an R environment for statistical analysis. Table 1 provides a summary of the pitching types and their associated abbreviations. The pitches with low counts in the dataset, below 100 for the season, were excluded from some analyses (and marked with an asterisk).

Abbreviation	Pitch Type
CH	Changeup
CU	Curveball
FC	Cutter
EP	Eephus*
FO	Forkball*
FF	Four-Seam Fastball
KN	Knuckleball*
KC	Knuckle-curve
SC	Screwball*
SI	Sinker
SL	Slider
SV	Slurve*
FS	Splitter
ST	Sweeper

Table 1: Abbreviations and names of baseball pitch types.

Exploratory analyses were used to look at correlational relationships, using chi-squared analysis, between pitching type and outcomes. Finally, the Magnus effect was calculated and explored.

Results

Summary Statistics

The following is a summary of pitch types in the dataset compared to typical counts and averages during a baseball season (using total numbers from the 2022 MLB season). Comparing the counts between the 2024 data and the ”typical” 2022 data, it is clear that not every pitch’s data was download for the 2024 dataset (this is where I discovered my mistake). However, the percentages between the datasets appear fairly similar.

Table 2: Comparison of Pitch Types: 2024 Dataset vs. MLB Averages

Pitch Type	2024 Data Count	\%	MLB Typical Count	MLB Typical \%
FF	1862	31.52	232974	32.96
SI	1093	18.50	111351	15.75
SL	826	13.98	136930	19.37
CH	712	12.05	80252	11.35
FC	501	8.48	50234	7.11
CU	308	5.21	56681	8.02
KC	207	3.50	14194	2.01
ST	188	3.18	0	0.00
FS	180	3.05	10376	1.47
SV	18	0.30	0	0.00
FO	10	0.17	0	0.00
SC	2	0.03	0	0.00

Statistical Analysis

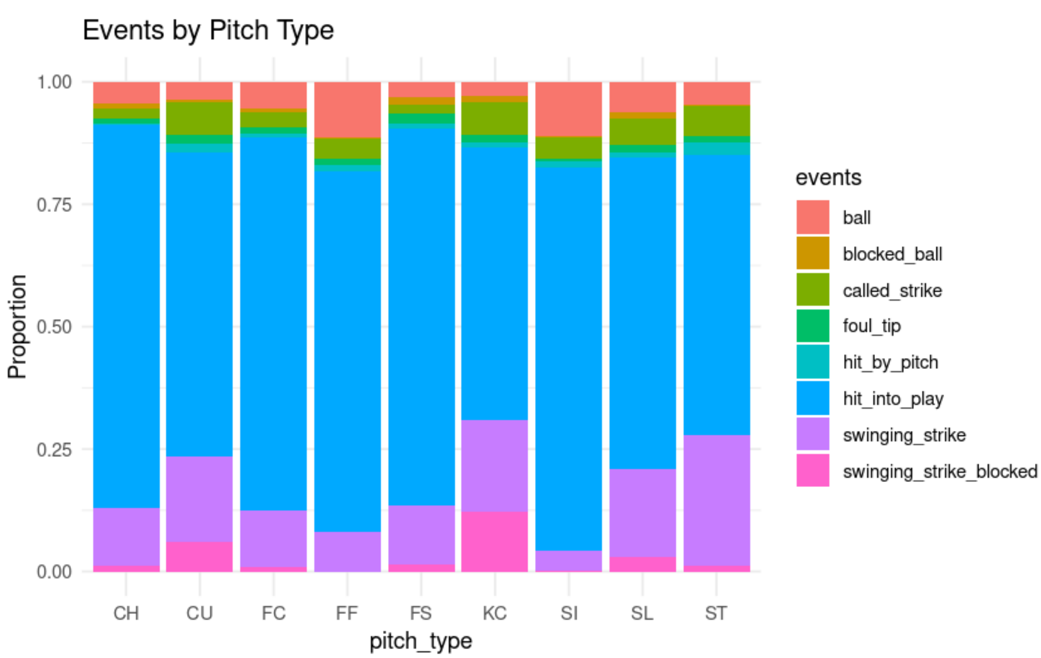

Figure 2: Proportion of events for each pitch type.

Figure 2 demonstrates the proportion of "events" for each pitch type. All pitches were primarily hit into play (though again I am suspicious that my downloaded 2024 dataset is skewed). The sweeper (ST), slider (SL), knuckle-curve (KC), and curveball (CU) result in more swinging strikes. Whereas the four-seam fastball (FF) and sinker (SI) result in more balls. Further analysis is required to determine if these results are statistically significant.

Two key outcomes of an at-bat are the home run and strike out. The following figure shows what percentage of the pitches resulted in either of these outcomes. The knuckle-curve (KC) and sweeper (ST) had the highest percentage of strikeouts, and the sinker (SI) had the lowest percentage. The cutter (FC) may result in the greatest percentage of home runs but the event rate and error bar overlap with other pitch types fairly closely. Additional analysis would be beneficial to determine if cutters do statistically result in more home runs.

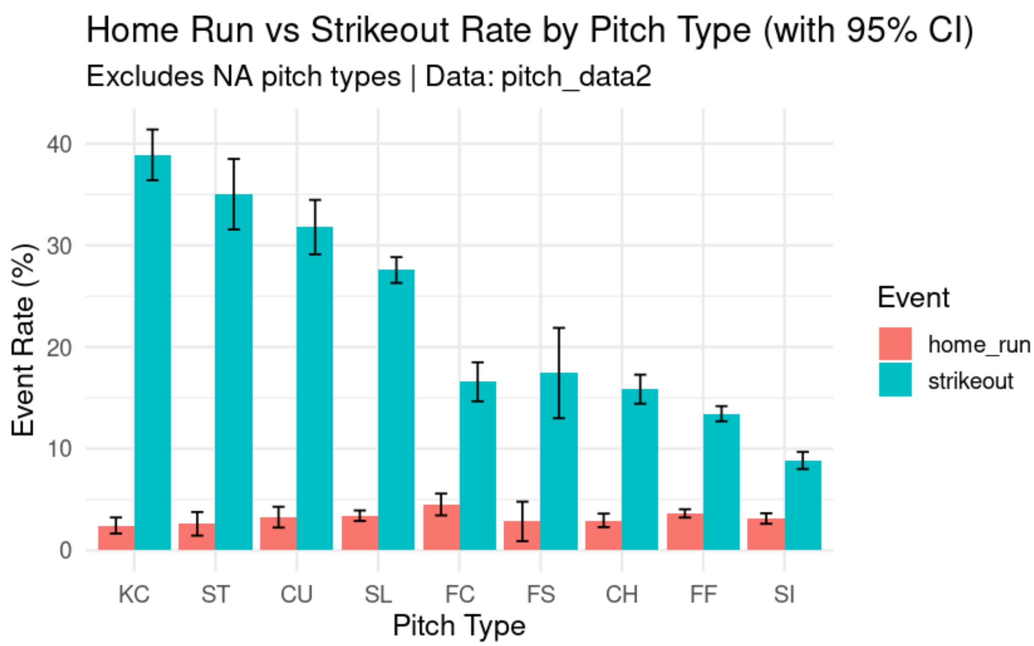

Figure 3: For each pitch type, the graph shows which percentage of its outcomes are either home runs or strikeouts.

A chi-square analysis ($p < 2.2 \times 10^{-16}$) showed that there is a statistically significant relationship between pitch type and all event outcomes in the database.

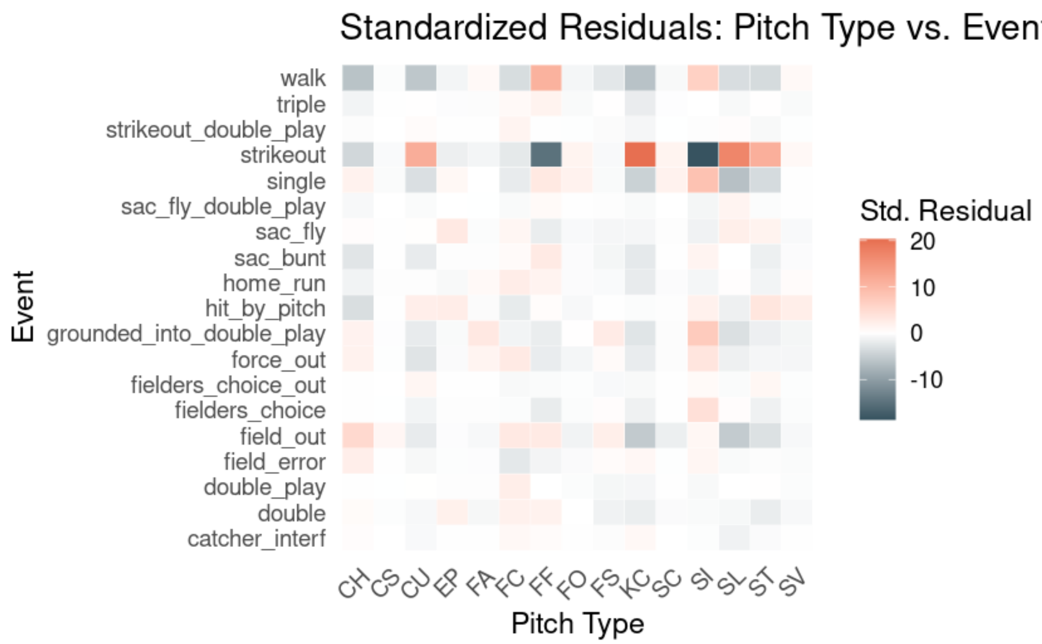

Figure 4: For each pitch type/event pairing, we see how how much they are expected or not. Stronger colors show that the pair is not expected.

The standardized residuals heatmap figure 4 shows which pitches and outcomes deviated the most from the expected value. A value greater than |2| indicates significant deviation from the expected value; a positive value is more likely than expected, and a negative value is less likely than expected. We can see from the heatmap above that strikeouts are more likely than expected when the pitch is a knuckle curve (KC) or a slider (SL) and less likely than expected when the pitch is a four-seam fastball (FF) or a sinker (SI). In fact, many combinations were considered more likely than expected and are shown in the table below. This table shows those pairs with a standardized residual greater than |3| (to keep the table to one page).

Pitch Type & Event Combinations with |Standardized Residual| > 3

Pitch_Type	Event	Std_Residual
KC	strikeout	20.32
SI	strikeout	-18.68
SL	strikeout	17.21
FF	strikeout	-15.07
CU	strikeout	11.59
ST	strikeout	11.35
FF	walk	10.93
SI	single	8.60
SI	grounded_into_double_play	7.38
SI	walk	6.52
SL	single	-6.41
KC	walk	-6.26
CH	walk	-6.16
CU	walk	-5.72
KC	field_out	-5.38
CH	field_out	5.34
SL	field_out	-5.29
KC	single	-4.62
SI	fielders_choice	4.36
CH	strikeout	-4.02
ST	walk	-3.69
ST	single	-3.67
SI	force_out	3.64
ST	hit_by_pitch	3.59
FC	walk	-3.48
SL	walk	-3.45
CH	hit_by_pitch	-3.37
FA	grounded_into_double_play	3.23
CU	single	-3.22
FF	single	3.16
SL	grounded_into_double_play	-3.15
FC	field_out	3.06
EP	sac_fly	3.05
ST	field_out	-3.04

Magnus Effect

The last goal of the report was to compare the expected value of the Magnus effect with an approximation calculated from the 2024 dataset. This was done using Equation 5 (provided again here):

$$F_{\text{Magnus}} = \frac{1}{2} \cdot \rho \cdot A \cdot C_L \cdot v^2 \tag{6}$$

Starting with the total vertical acceleration (a variable in the dataset), acceleration due to gravity was subtracted, and then an estimate for drag force (calculated from variables in the dataset) was subtracted. From this value, the Magnus force was estimated. This value is highly influenced by the ball's velocity [1]. So to compare this value across pitches, we can remove the influence of velocity by comparing the estimated lift coefficients instead. Figure 5 shows the results of this comparison. The calculated values from the database tended to be lower than expected values, across pitch types. Further exploration of the model and data might indicate ways to improve these estimates.

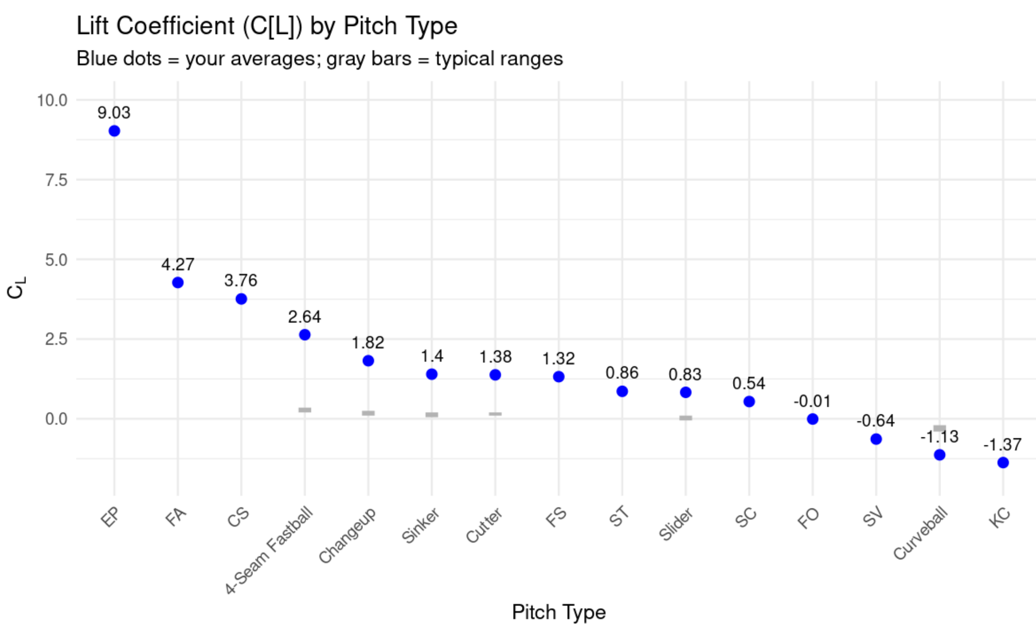

Figure 5: Comparing calculated and typical lift coefficients across all pitch types.

For further exploration, a scatterplot of lift coefficient vs spin rate (in rpm) across pitch types was generated. A fit line for each pitch type was included. This plot shows whether higher spin rates actually result in more effective lift. Flat trend lines show that the spin rate doesn't directly increase lift. Some of the trend lines drop, showing that as spin rate increases, the amount of spin contributing to lift and the Magnus effect decreases. Further analysis can determine if this is an artifact of sparse data at higher spin rate values.

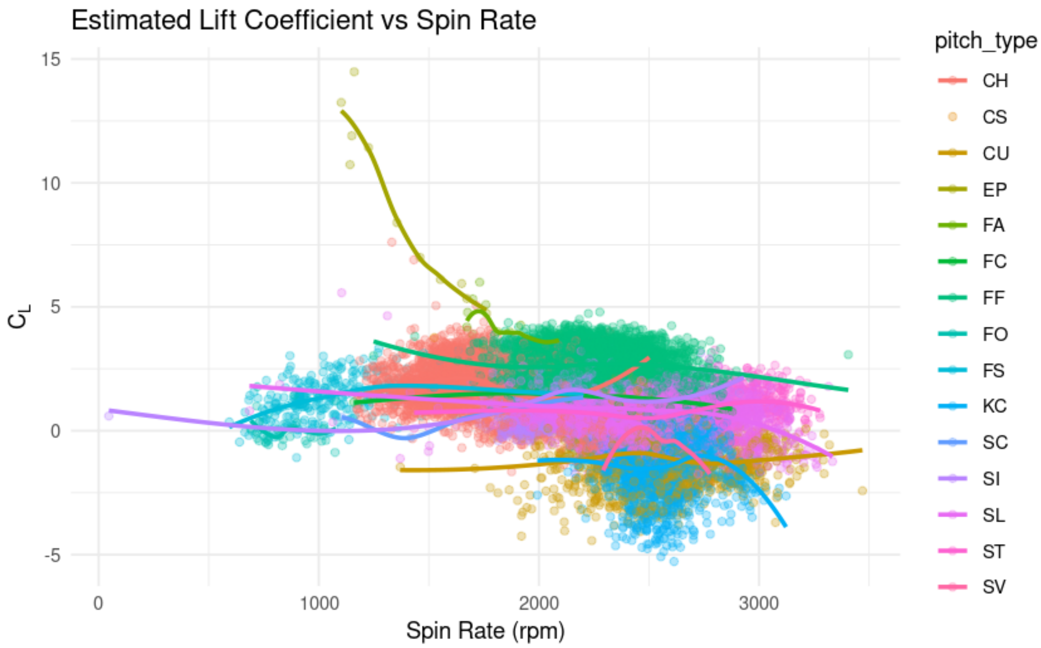

Figure 6: Lift coefficient vs spin rate for each pitch type

Conclusion

This report examined various types of pitching and their relationship to at-bat outcomes. An attempt was made to estimate the Magnus effect and explore its correlation with pitch types. While this effort only begins to touch on the vast field of pitching statistics and the physics of baseball, it shows how statistical analysis can further one's understanding of the game. As more sophisticated methods and richer datasets become accessible, both professional and amateur analysts can continue to deepen their understanding and appreciation of the game. Moreover, a stronger grasp of the physics behind pitching can support more effective performance evaluations and contribute to biomechanical optimization for pitchers and their coaches.

References

[1] Robert K Adair. The physics of baseball. *Physics Today*, 48(5):26–31, 1995.

[2] Jim Albert, Jay Bennett, and Jay Bennett. *Curve ball: Baseball, statistics, and the role of chance in the game.* Springer, 2003.

[3] Britannica. Magnus effect, 1998. Accessed on May 5, 2025.

[4] ChatGPT. Conversation with chatgpt, 2025. OpenAI. Personal communication, May, 2025.