Does Speed Kill…Base Hits?

Exploring speed as a proxy for defensive ability

By Ethan Rendon

2022-07-28

I.   Introduction

"Defense usually doesn't make many headlines, but it goes a long way towards winning baseball games." - **Derek Jeter**

Jeter is the perfect candidate for a Last Dance-esque docu-series for baseball. Not just because of his unrivaled fame, or stature in the record books, but also how Jeter is viewed encapsulates so much of the balance between tradition and modernism of the nation's pastime.

Tradition: For many growing up with Jeter, he was *The Captain* — a hitter's hitter, not a slugger, known for his clutchness and iconic defensive plays.

Modernism: For others who believe in DRS (Defensive Runs Saved), Jeter is the worst defensive shortstop of all time — whose flip play in the 2001 ALDS that stunned the Oakland A's was just a flash in the pan.

A key to being a baseball analyst is testing long-standing traditions, versus numbers, to better understand the game.

Scouting grades are one of these traditions often read on a 20-80 scale where 50 is average. For hitters, grades numerically assess the current and future value of five tools: **Hit, Power, Speed, Arm, and Field.**

Thanks to innovation and access to tracking technologies, tools are no longer solely graded by the "eye test."

- *Hit* can be graded based on swing decisions and bat path.
- *Power* can be graded based on bat speed and exit velocity.
- Speed can be measured using a stopwatch.
- *Arm* can be measured using a radar gun.
- ***Field, however,* is ever-evasive and complex.**

A scout cannot simply capture fielding capabilities by pointing a device at a fielder. Though athleticism tests can serve as a proxy for ability, deferring to the eye test is often the most convenient option.[1]

Jeter's defensive ability serves as an anecdote for the pitfalls of the eye test. The eye test doesn't aggregate, it selects a small sample and labels skill in a binary way, at best "good" or "bad" defender, or at worst arbitrarily — what does the difference between a 50- or 60-grade fielder mean?

Athleticism tests are not as convenient as other measurables like exit velocity. In an effort to unlock an accessible, aggregate, and accurate way to measure fielding and given that quick-burst acceleration strongly correlates with athleticism — this paper aims to test speed as a proxy for defense.

---

[1] Example of an athleticism test done by Loden Sports https://lodensports.com/use-case/mlb-draft-spotlight-spencer-jones-an-outlier-athlete/

I. Constraints

The scope of the question is limited to plays made by centerfielders. Outfielders run further distances on plays thus acceleration is most needed. In particular, center fielders make the lion share (~ 40%) of outfield plays.[2]

The data used is anonymized minor league tracking data provided by SMT. It is filtered to remove foot speeds (ft/s) that are not possible at the cut-off of 32 ft/s, as the highest average in the MLB since tracking began is 30.9 ft/s by Byron Buxton in 2015. This filter helps control for tracking errors on both the ball and players.

Two other filters that assure data quality are removing route efficiencies > 100% and plays where the ball is not tracked more than 50 times — this is to have a reliable sample of where the ball is, when and also to ensure that any mislabeling or tracking errors can be removed. After the data is filtered there are **588 plays made by 76 fielders**!

Some caveats, the dataset does not have wall height and stadium dimensions provided, but overlaying a play onto a MLB stadium is possible. "Defense" in this case is limited to the catching of the ball, but there are other important defensive traits such as arm strength and framing which are crucial to other specific positions.

Based on a small poll (n=76), it is very much up to debate whether or not speed kills….base hits!


---

[2] https://www.mlb.com/news/statcast-looks-at-value-of-center-field-skill-c265821200
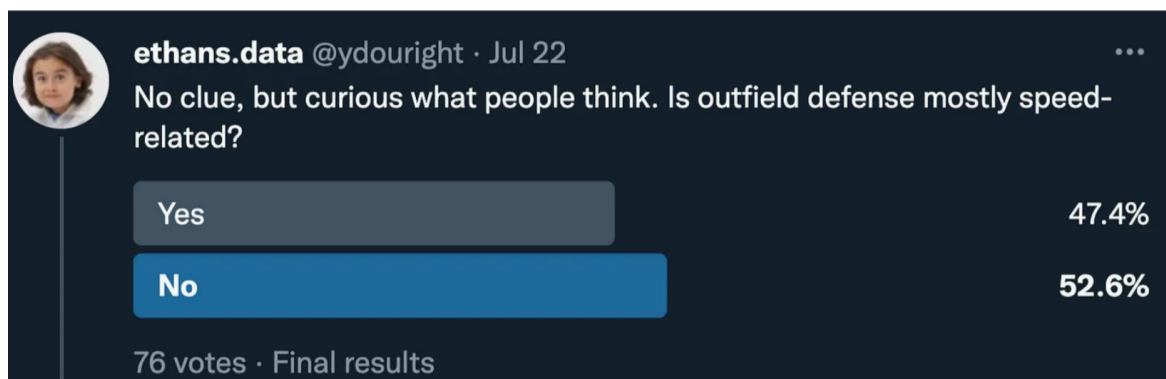
II. Factors

In this paper **four factors are largely used to measure defensive value and speed**, below goes into depth as to their context to being a center fielder, how they are used in the paper and the defects that come with the respective stats.

Defensive value factors

The Bias of Long Distances Between Line Distance vs Route Efficiency
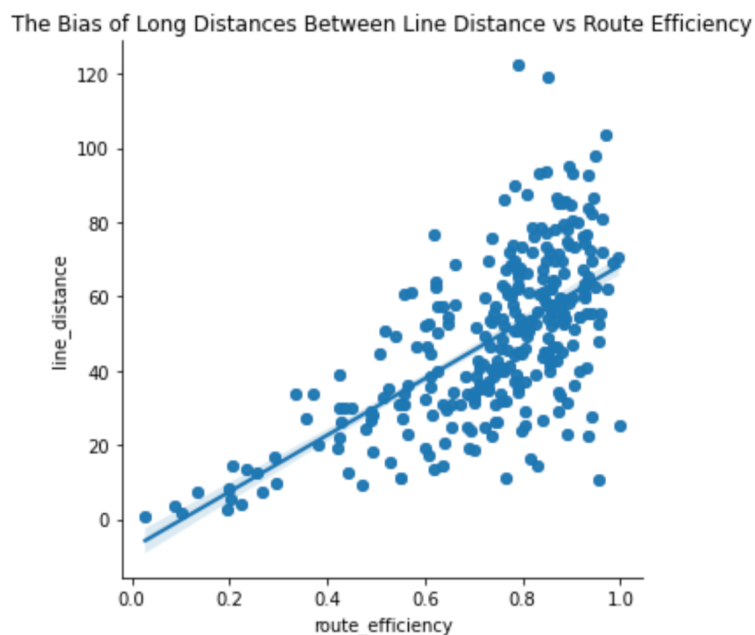
1. **Route efficiency** is the distance covered by a fielder divided by the straight-line distance between a player at the point of contact and where the ball is fielded. So, a route efficiency of 1 or 100% means a player ran in a straight line to the ball.[3]

    Route efficiency is not perfect, it does not account for hang time and is biased towards larger distances. The figure above particularly exposes this bias. But, also mathematically this makes sense as shown by the table below.

    That said, route efficiency combined with other metrics can provide nuance on a play and this paper only uses it to demonstrate relationships rather than player evaluation.

Line distance (ft)	Fielder distance (ft)	Route efficiency (%)
50	100	50
100	150	66.67

---

[3] https://www.mlb.com/glossary/statcast/distance-covered

2. **Range** is the area a player can cover based on catches made in their career.

Having substantial range is crucial to being a solid defensive center fielder as the "captain" of the outfield a center fielder is tasked to cover sizable ground relative to the left or right fielder. Additionally, being able to cover a larger area allows a fielder to make up for deficiencies in route efficiency.

Max range is not very predictive of a fielder's capability because it is sample-based — if a fielder only has one play tracked at 50ft who is to say they can't get a ball 55ft. Under the data set, arises for fielders with smaller samples of plays tracked.

Looking at range under the context of hang time and catch probability is where its value lies.

3. *Good first step rate%* is the percentage of times a player gets closer to the ball's final location in their first second of movement.

A common theory for fielding is having a good first step and direction you run in early on significantly increases the catchability of a play. Tracking data permits an opportunity to test this versus other defensive metrics!
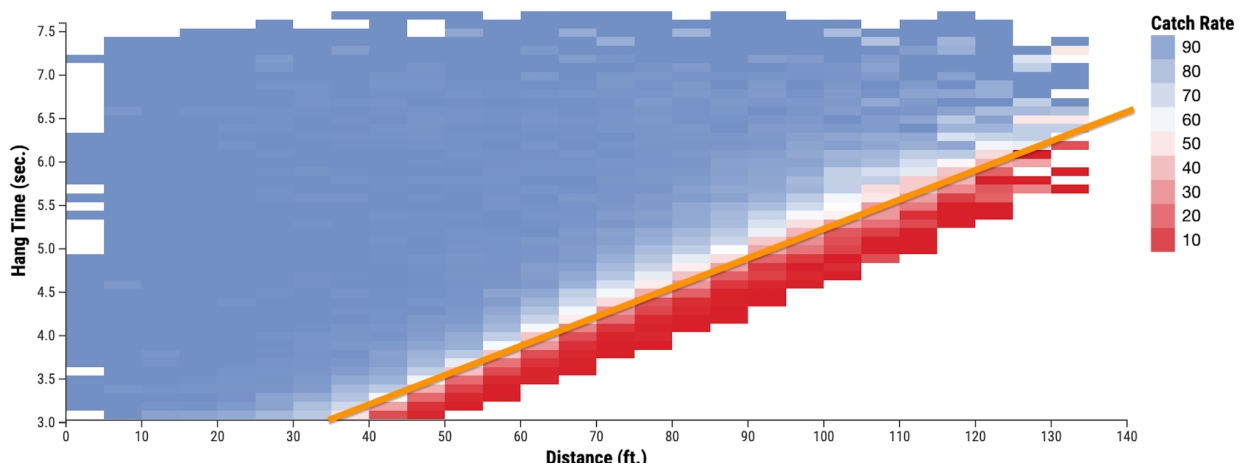

4. **Easy catch%** is the percentage of catches made by a player that were above a 50% catch probability threshold.

This threshold is defined by MLB data, the figure above demarcates an orange line to represent this threshold.[4] If a play is above the line? It is deemed an easy catch. Below the line? Good luck, it is not an easy catch!

To make this distinction the code uses the equation $y > 0.033x + 1.833$ where x is the straight-line distance (ft) a fielder must travel and y is the hang time (sec).

---

[4] Original chart is from https://baseballsavant.mlb.com/statcast_catch_probability, the orange line is added for clarity.

Speed factors
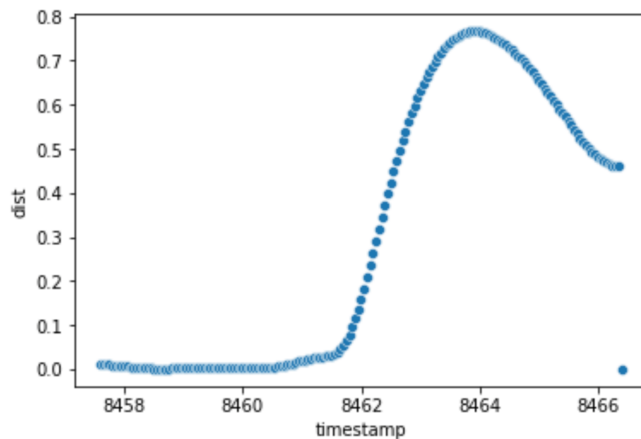
The key distinction for the following speed factors is the use of average versus max. Using both allows for context about a player's career (average) and also for a player's potential (max).

The figure above displays the importance of accounting for max speed, because speed over a given play is not linear. When a play starts little movement occurs, then the ball is hit and the fielders speed up!

Illustrating the relationship between potential and defense metrics is key to understanding whether or not speed is a suitable proxy for ability.
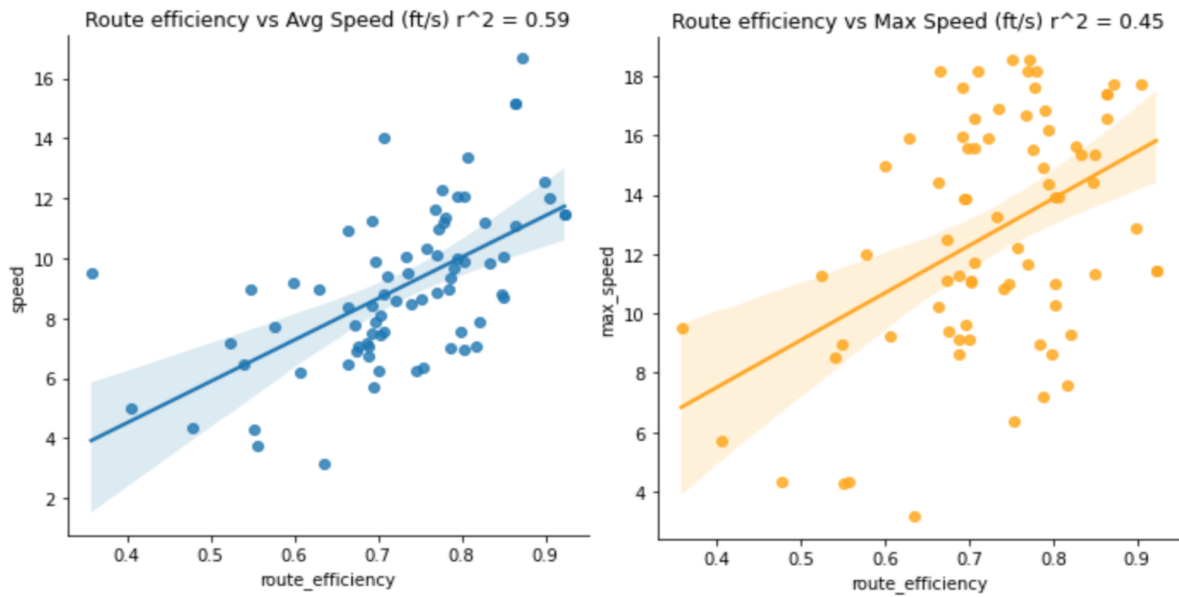
1. **Average speed** is an average of speeds on plays made.
2. **Max speed** is the quickest a player moved on plays made.
3. **Average first step speed** is an average of speeds by a player in their 1st second of movement on plays made.
4. **Max first step speed** is the quickest a player moved in their 1st second of movement on plays made.
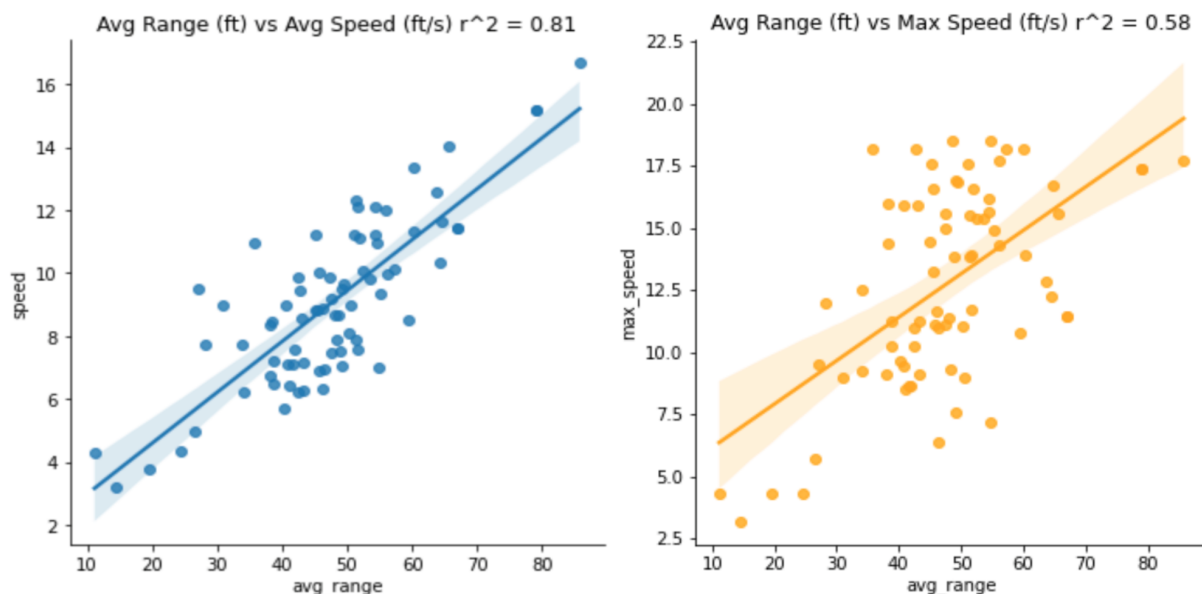
III.   Relationships explored

The data used in the following plots is the player's career data and the plots display the player's defensive ability on the horizontal axis versus average (blue) and max speed (orange).

R^2 in the title of the plots represents how much the defensive metric on the horizontal axis is explained by the speed metric on the vertical axis.
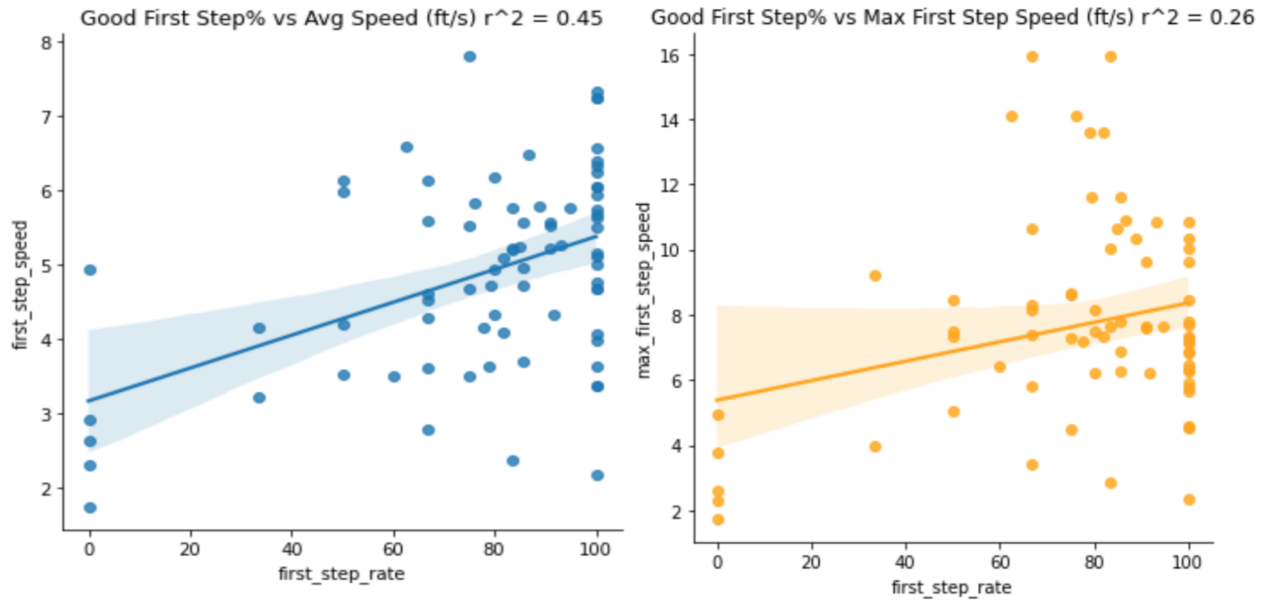
Route efficiency

Range

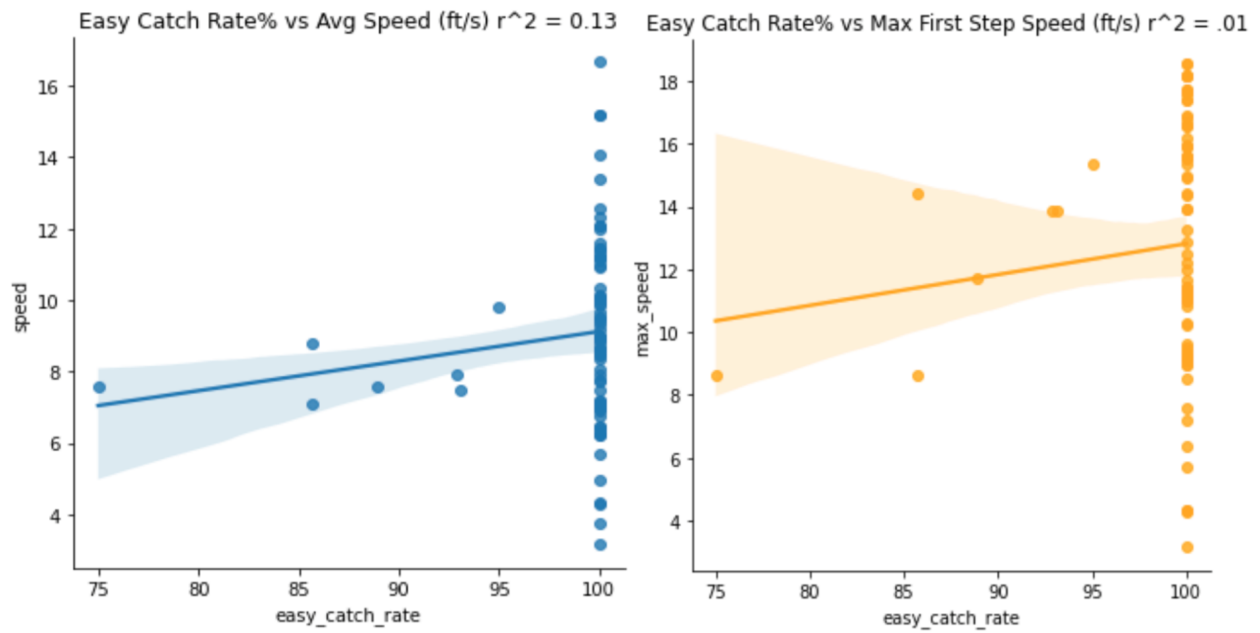# First step

Good First Step% vs Avg Speed (ft/s) r^2 = 0.45

Good First Step% vs Max First Step Speed (ft/s) r^2 = 0.26

Easy catch%

Easy Catch Rate% vs Avg Speed (ft/s) r^2 = 0.13

Easy Catch Rate% vs Max First Step Speed (ft/s) r^2 = .01

IV. Findings

We have defined and plotted metrics for comparing defensive ability to speed. This paper enumerates the following conclusions based on the plots above.

1. The relationship between **route efficiency to speed and max speed is moderate** as speed and max speed respectively explain **59% and 45%** of route efficiency.

Qualitatively, this makes sense as a fielder must get to where the ball location is under the time it is in the air to make a play — thus being quick can translate to having a better route.

2. The relationship between **range to speed is strong** as speed explains **81%** of range abilities. Whereas, **max speed correlation is moderate** and only explains **58%** of range.

Given that range and speed are both average quantities, and max speed is not that explains the considerable difference in correlations — but max speed in combination with max range can hint at the potential of a fielder.

3. The relationship between **a good first step to first step speed is moderate** as speed explains **45%** of having a good first step.

Logically, as speed is only one element to getting closer to the ball and does not account for directions this makes sense and also suggests a link between reaction time on a play to speed.

4. Lastly, **speed metrics cannot at all explain the easy catch rate**. This is reasonable as easy catch rate is situational based on play and independent of player ability as it only takes into account ball flight: hang time and distance from the fielder.

V. Takeaways

Given the **tracking data suggests that speed can moderately and strongly** explain defensive metrics, it can be used as a proxy for center field defensive ability.

Additionally, having a **good first step explains 37% of catchability at increased distance**. This moderately confirms the value of having a good first step!

This paper is only the tip of the iceberg (or diamond). Some ways this research could be used in the future include:

1. **Proxy for infielder defense** with speed and arm strength.
2. **Modeling the potential range** of draft prospects based on biomechanical data as the paper suggests the importance of athleticism.
3. At a developmental level, organizations could place **emphasis on training for speed and reaction time** — as the paper reveals the significance of a good first step.
4. **Research into the fatigue of fielders** over the course of the game since defense is so speed driven. Do fielders fatigue over the course of the game and lose value like pitchers' stuff worsening as workload increases?

VI.   Acknowledgments

VII.   References

1. "Defensive Runs Saved (DRS): Glossary." MLB.com. https://www.mlb.com/glossary/advanced-stats/defensive-runs-saved.
2. Driveline Baseball. "What Is the Countermovement Jump (CMJ) Test?" Driveline Baseball, July 14, 2022. https://www.drivelinebaseball.com/2020/11/what-is-the-counter-movement-jump-cmj-test.
3. Grove, Cameron. Twitter Post. December 19, 2021, 10:39 AM. https://twitter.com/Pitching_Bot/status/1472637658119786504?s=20&t=iGPq6Dw4X3ZK7seGZ5VA8g.
4. McBride, J. M., Blow, D., Kirby, T. J., Haines, T. L., Dayne, A. M., & Triplett, N. T. (2009). Relationship between maximal squat strength and five, ten, and forty yard sprint times. *Journal of strength and conditioning research*, *23*(6), 1633–1636.
5. McDaniel, Kiley. "Scouting Explained: The 20-80 Scouting Scale." FanGraphs Baseball, August 26, 2018. https://blogs.fangraphs.com/scouting-explained-the-20-80-scouting-scale/.
6. Nathan, Alan M.. "Characterizing the performance of baseball bats." *American Journal of Physics* 71 (2003): 134-143.
7. Tango, Tom. "Introducing Infield Outs above Average." Medium. MLB Technology Blog, January 13, 2020. https://technology.mlblogs.com/introducing-infield-outs-above-average-6467e61a98dc.

VIII.   Code Linked <u>Here</u> Hold-Click to Open New Tab

Correlation Heatmap

	center_field	avg_range	range	route_efficiency	speed	max_speed	first_step_speed	max_first_step_speed	first_step_rate	easy_catch_rate
center_field	1.000000	-0.170311	-0.124639	-0.171390	-0.064928	-0.084713	0.074677	-0.023177	-0.027695	0.100032
avg_range	-0.170311	1.000000	0.632859	0.751956	0.811885	0.584908	-0.101155	0.041402	0.274150	0.035388
range	-0.124639	0.632859	1.000000	0.433197	0.469557	0.742876	-0.025947	0.364166	0.372274	-0.013852
route_efficiency	-0.171390	0.751956	0.433197	1.000000	0.589517	0.450989	0.070530	0.112056	0.191133	-0.084312
speed	-0.064928	0.811885	0.469557	0.589517	1.000000	0.753798	-0.056817	-0.076780	0.321752	0.125450
max_speed	-0.084713	0.584908	0.742876	0.450989	0.753798	1.000000	0.104964	0.227830	0.471109	0.099014
first_step_speed	0.074677	-0.101155	-0.025947	0.070530	-0.056817	0.104964	1.000000	0.530835	0.454958	0.120567
max_first_step_speed	-0.023177	0.041402	0.364166	0.112056	-0.076780	0.227830	0.530835	1.000000	0.263783	0.041587
first_step_rate	-0.027695	0.274150	0.372274	0.191133	0.321752	0.471109	0.454958	0.263783	1.000000	-0.051460
easy_catch_rate	0.100032	0.035388	-0.013852	-0.084312	0.125450	0.099014	0.120567	0.041587	-0.051460	1.000000