

## Sumo *Honbashi* Data Paper

Japanese Ministry of Education, Culture, Sports, Science and Technology

Applicant: Laurel Schwaebe on behalf of the Japan Sumo Association

In partnership with SumoDB and Sumo Hositori

Peter Organisciak

Data Curation LIS-4220-1 | Winter 2023

University of Denver

## Abstract

The Japan Sumo Association does not have a complete database of sumo records. By partnering with SumoDB and Sumo Hositori, the Japan Sumo Association is building a database that dates back to the Edo period. Records are limited to *makuuchi* division *rikishi*. The earliest tournament recorded in the database was the October 1757 *basho*, and the latest is the January 2023 *basho*. The data points are on the wrestlers' *shikona*, *heya*, *shusshin*, birthdate, *hatsu*, *intai*, height, weight, highest rank, that tournament's rank and record. Data included was only collected from *honbasho*. The data is often incomplete and needs to be cross referenced against other records, including historical borders and *heya* activity dates, and this is likely going to be an ongoing process as the database grows. The JSA will maintain this database in partnership with SumoDB and Sumo Hositori moving forward in order to provide sumo fans and researchers with as complete records as possible.

*Keywords:* sumo, sports analytics.

## (1) Overview

Sumo, a fairly simple sport in which two wrestlers wearing only *mawashi* attempt to push one another out of the *dohyō* or to the ground, may date as far back as 23 BC and has been a treasured Japanese sport for over 2000 years (Manzenreiter, 2014). While today sumo is seen as a sport abroad, its origins are as a Shinto ritual, with legend that the first sumo bout was actually a fight between the Shinto gods Takemikazuchi and Takeminakata. Naturally, we do not have records of matches dating back to sumo's start, but dedicated followers, be them shrine officials, observers, or fans, have kept note of these bouts longer than the *Nihon Sumō Kyōkai*, called Japan Sumo Association (JSA) in English, has existed. Once the JSA was formed, this organization was responsible for all *honbashi* held within Japan.

As a sport, religious ceremony, and Japanese cultural treasure, it is essential that as much data on sumo be collected and preserved in a centralized, searchable, and accessible place as possible. Unfortunately, while the JSA was formed in 1925, its database only goes back to 2013 and is somewhat sparse when it comes to details on the bouts and the wrestlers themselves. Additionally, there is no way to compare wrestlers, various bouts or tournaments, or otherwise view larger collections of information.

This database is a gathered collection of sumo wrestling statistics from *honbashi* dating back to the Edo period. While the statistics for *honbashi* for the last 97 years are all reported from JSA Grand Sumo Tournaments, much of this information, as well as data from years prior, was not part of the JSA's database. It is for this reason that it is not only the labors of the Japan Sumo Association that have developed such a complete dataset but rather its joint efforts with the researchers at SumoDB and Sumo Hositori, which have maintained their sumo databases since 2007 and 2003 respectively. This database is intended to allow sumo fans, researchers, Shinto officials, and others to access the most complete compilation of data on the sport and to increase knowledge of this Japanese cultural treasure.

The JSA, as part of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) has proposed to acquire the data that SumoDB has gathered and incorporate it and crosscheck it with its own records. The purpose of this integration is to compile the majority of existing data on sumo, which SumoDB has already compiled much of from across the Japanese and English web. This will provide one nearly complete resource for sumo fans and researchers. Additionally, adding the data to the JSA database provides it with a more secure home, as currently SumoDB is an individual research site, and will likely only last until the end of that individual interest or life. This is especially concerning, as many of SumoDB's own sources have vanished from the web. The JSA, as part of the MEXT, is funded indefinitely and will be able to support this database for the foreseeable future, as well as provide greater access as it has a team of translators, researchers, and coders to support it. This will be a multi-step project that will ultimately span all available *rikishi* records and will be available in Japanese and English. The first phase of this project covers the top division and will only be in English.

## (2) Methods

### Data Gathering

This database was created using a combination of research from the Japan Sumo Association, SumoDB, and Sumo Hositori. As mentioned previously, all *honbashi* data after 1926 was created from JSA Grand Sumo Tournaments. Data gathered by SumoDB and Sumo Hositori was cross-referenced against JSA analog records (as the JSA has only digitized its data as of the January 2013 *bashi*).

The data contains key information on *rikishi* gathered at each tournament. The earliest tournament recorded in the database was the October 1757 *bashi*, and each *rikishi*'s record is filled with data on their *shikona*, *heya*, *shusshin*, birthdate, *hatsu*, *intai*, height, weight, highest rank, that tournament's rank and record. Data included was only collected from *honbashi*, or "true" tournaments, which are currently held six times a year. Since the formation of the JSA, all *honbashi* have been organized by the JSA under the English title "The Grand Sumo Tournament." Any data gathered from additional tournaments was discarded. Additionally, only data for *makuuchi*, or top division *rikishi* was included in the database. Other JSA-registered *rikishi* from lower divisions will be added as a later part of the project. This leaves 42 *rikishi* in any given *bashi*, which both provides the most in-demand data to users, but also significantly cuts down on the labor needed for the first part of this project, as the following divisions would add several hundred *rikishi* per *bashi*.

The bulk of the oldest recorders came together as a part of Sumo Hositori's research. Much of the data for the earliest *bashi* can be attributed to Tadamasaka Sakai's *Nihon sumoshi*, or "Japanese sumo history," volumes 1 and 2. While Sakai was best known as the 15th Japanese Minister of Agriculture and Forestry, as well as a Member of the House of Peers, he was also the first chairman of the Grand Sumo Wrestling Yokozuna Deliberation Committee and the first chairman of the Japan Professional Wrestling Association (not to be confused with the Japan Sumo Association). Additionally, his vast collection of sumo-related artifacts went on to form the Sumo Museum, for which he also served as the first director. This is to say, his collection of research, which was then formatted into Sumo Hositori's database, is extremely well-respected. However, records from Edo *bashi* are often incomplete, and as such, Sakai, Sumo Hositori, and now the JSA, have opted to include the largest number of records, rather than only those that are complete.

Later data was generally produced by the JSA itself and was much more standardized. While this data was not digitized for its online database, it was possible to compare the data points gathered from Sumo Hositori and SumoDB's third-party sources, such as newspapers and magazines, against the JSA's paper records. The most recent data, from 2013 onward, was all digital and could be incorporated easily into the database.

As this is only the first phase of this project, the data is currently only available in English and with limited measurement options. Plugins will soon be available to provide better accessibility to the data for a larger international audience.

## Quality Control

As previously mentioned, the data gathered, particularly that dating prior to the creation of the JSA, is not complete. Not all names, weights, heights, or other data points were recorded for the older *basho*. Most frequently missing include *hatsu* and *intai*, as well as height and weight. Most birthdates in the Edo period are listed as month and year, though it is worth noting that Japanese traditions around birth dates have changed since that time.

Prior to the Meiji Restoration, Japan utilized the *shakkanhō*, or traditional Japanese system of measurement (U.S. War Department, 1995). This may lead to some discrepancies in heights and weights, as they would originally have been recorded in traditional measurements before being converted into the metric system.

Related to this, Japan's borders have changed in the time since the first record, so some wrestlers may be listed as Japanese citizens during their *basho* but are currently associated with other countries. For example, Wakanami, who competed in the May 1940 *basho*, was from Karafuto. At the time, this was a Japanese territory, but in 1945, the island became part of Russia. Today, this island is known as Sakhalin and is part of Russia. His *shusshin* is labeled Sakhalin, but his nationality is still listed as Japanese, as at the time, he was from Japanese territory. This becomes more complex with *rikishi* from wartime-occupied lands in Korea and Mongolia, who were ultimately listed as from their respective countries, despite Japan's claim over them at the time. Additionally, even within Japan's borders, prefectures, regions, and domains have changed a great deal over the past 300 years. When possible, *rikishi* are matched to their historical *shusshin*, even if these places do not currently exist under those names.

Furthermore, the *honbasho* format has changed several times after the formation of the JSA, from twice a year to four times a year, to the current six times a year. Because the earlier tournaments were not JSA organized, there may be discrepancies in registration for the tournaments, and regulations around sumo were not standardized until the JSA was organized. *Heya* close regularly, often following the retirement of the head coach, and retired wrestlers frequently start their own stables, sometimes under the same name as previously existing stables. Most *heya* from prior to the last 100 years no longer exist, and some names have seen multiple iterations, leading to a great deal of confusion. For example, there are at least three different Tagonoura stables, one of which was active in the Edo period, one that was active between 2000 and 2012, and one that opened in 2013. There is no direct connection between these *heya*.

Because of these difficulties, cross-referencing every detail of the data and reconciling the names of stables, locations, and even *rikishi* with external data is key to keeping all information as accurate as possible and is likely going to be an ongoing process.

### (3) Dataset description

#### Object name

Sumo Honbasha Database.

#### Format names and versions

One CSV table, adjusted in OpenRefine. Wikidata reconciliation.

#### Data standardization

- Rikishi: the wrestler's *shikona*
- Heya: the wrestler's stable
- Shusshin: the wrestler's home country, or if in Japan, region
- Birth Date: in the format DD/MM/YYYY
- Hatsu: debut date in the format MM/YYYY
- Intai: retirement date in the format MM/YYYY
- Height: in cm
- Weight: in kg
- Highest Rank: formal rank title, in an abbreviated format
- Tournament Rank: formal rank title, in an abbreviated format
- Tournament Record: wins-losses-withdraws

#### Data creation dates

1757-10 to 2023-01-22.

#### Dataset creators

Japan Sumo Association. Project administration, data curation, and investigation.  
Affiliation: Japanese Ministry of Education, Culture, Sports, Science and Technology

Schwaebe, Laurel. Data curation, visualization. Affiliation: University of Denver.

SumoDB. Data curation, research. Affiliation: none.

Sumo Hositori. Data curation, research. Affiliation: none.

#### Language

English.

License

CC-BY.

Repository name

National Diet Library Data Repository.

国立国会図書館のデータリポジトリ。

Publication date

2023-03-12.



#### (4) Reuse potential

##### Sports Analytics

As with any sports data, sumo statistics are key for analysis that may provide insight into individual wrestlers and into greater trends in the sport. Savvy fans can review previous matchups to predict the outcomes of future matches. Anything from the *rikishi's* size to preferred moves and wrestling style may give him the leading edge, according to the data. "Pure-hearted fans" may be the desired audience in this demographic, but it is also worth noting that betting on sumo is also very popular (sometimes to its detriment, as Manzenreiter notes of sumo's match-fixing scandal). Because the JSA wishes to cater to all sumo fans, it must be aware that this data may also be used to help gamblers place bets. As this is legal in Japan and otherwise unheard of outside of Japan, this should not prove to be a legal problem. However, the JSA should continue to monitor its wrestlers and sponsors for match-fixing.

##### Historical Data

Few sports can claim to have such consistent records dating as far back as sumo. Though much of the older data is incomplete, these records provide valuable information regarding the time and the sport for researchers. Matched with the context of the period, this data may provide unique insight.

##### Religious Research

Sumo is uniquely situated between hundreds of years of history, popularity as a sport, and cultural significance as a religious ceremony. While modern-day sumo is more popular as a sport than a ritual, its connection persists and remains relevant throughout its history. These records are likely to be of use to Shinto researchers, particularly those related to older *basho* when sumo was more closely entwined with Shintoism.

## (5) Notes

## Dictionary

*Basho*: tournament.

*Dohyō*: the sumo ring, typically a circle made of partially buried rice-straw bales.

*Edo period*: the period in Japan between 1603 and 1867, also known as the Tokugawa period.

*Hatsu*: debut or beginning.

*Heya*: stable, or the center where the *rikishi* lives and trains. *Rikishi* will stay with one *heya* for his entire career. This word may also be pronounced “*beya*” in some contexts.

*Honbasho*: Grand Sumo Tournament, also known as the “real” or “main” sumo tournament. This tournament is held six times a year and has been organized by the Japan Sumo Association since 1926.

*Intai*: retirement.

*Makuuchi*: the top division of sumo wrestling, made up of the top four “titleholder” ranks, in descending order of rank, *yokozuna*, *ōzeki*, *sekiwake*, and *komusubi*, as well as *maegashira*, which is the lowest rank in this division. *Maegashira* ranks also feature descending numbers, which indicate the *rikishi*’s rank within *maegashira*.

*Mawashi*: the loincloth worn by wrestlers.

*Rikishi*: a professional sumo wrestler, may also be plural.

*Shikona*: a name assumed by sumo wrestlers for sumo. This may or may not be related to their actual names.

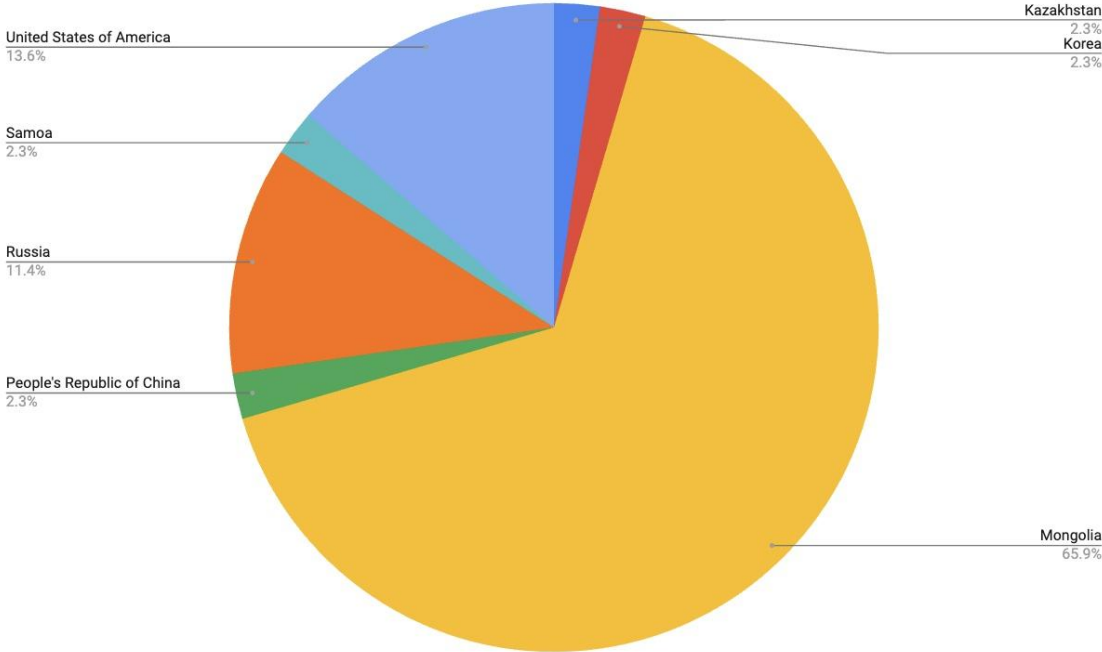
*Shusshin*: the *rikishi*’s home country or region, used to refer to one’s hometown in Japanese.

## Rank Abbreviation Guide

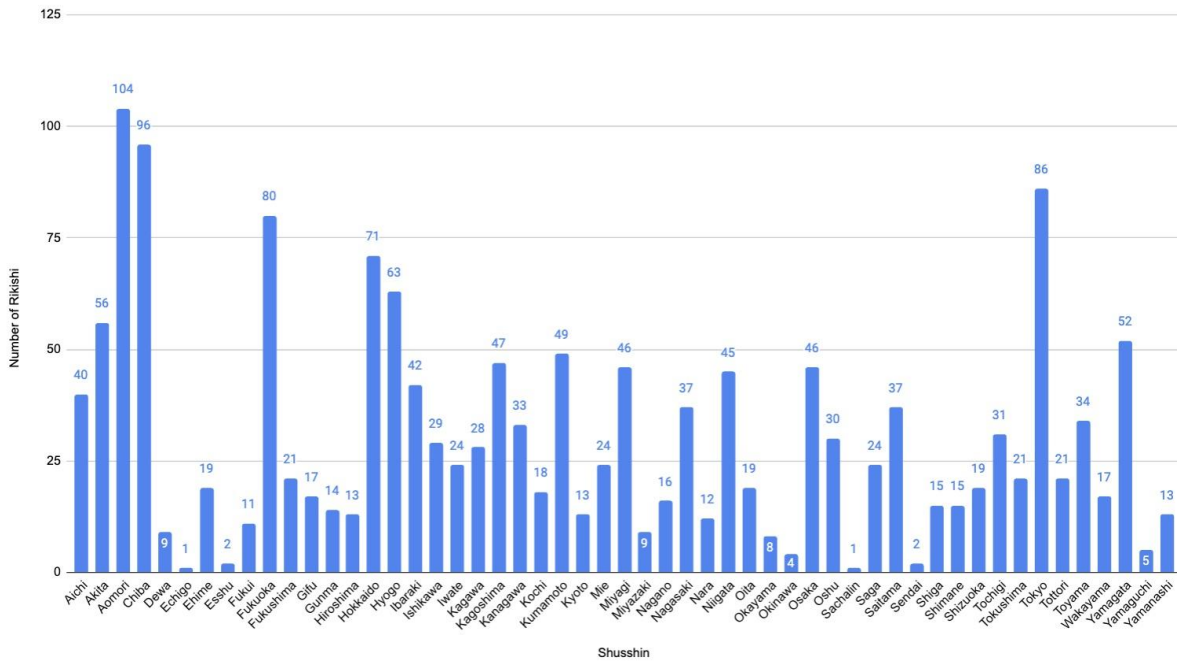
Abbreviation	Rank	Notes
Y	<i>Yokozuna</i>	Top rank; cannot be demoted
O	<i>Ozeki</i>	Second highest rank; like the lower ranks, can be demoted
S	<i>Sekiwake</i>	Third highest rank
K	<i>Komusubi</i>	Fourth highest rank; lowest of the "titleholder" ranks
M	<i>Maegashira</i>	Lowest of the ranks in the <i>makuuchi</i> division; each is numbered in descending order of rank (ex, <i>Maegashira1</i> or M1)

Samples of Data

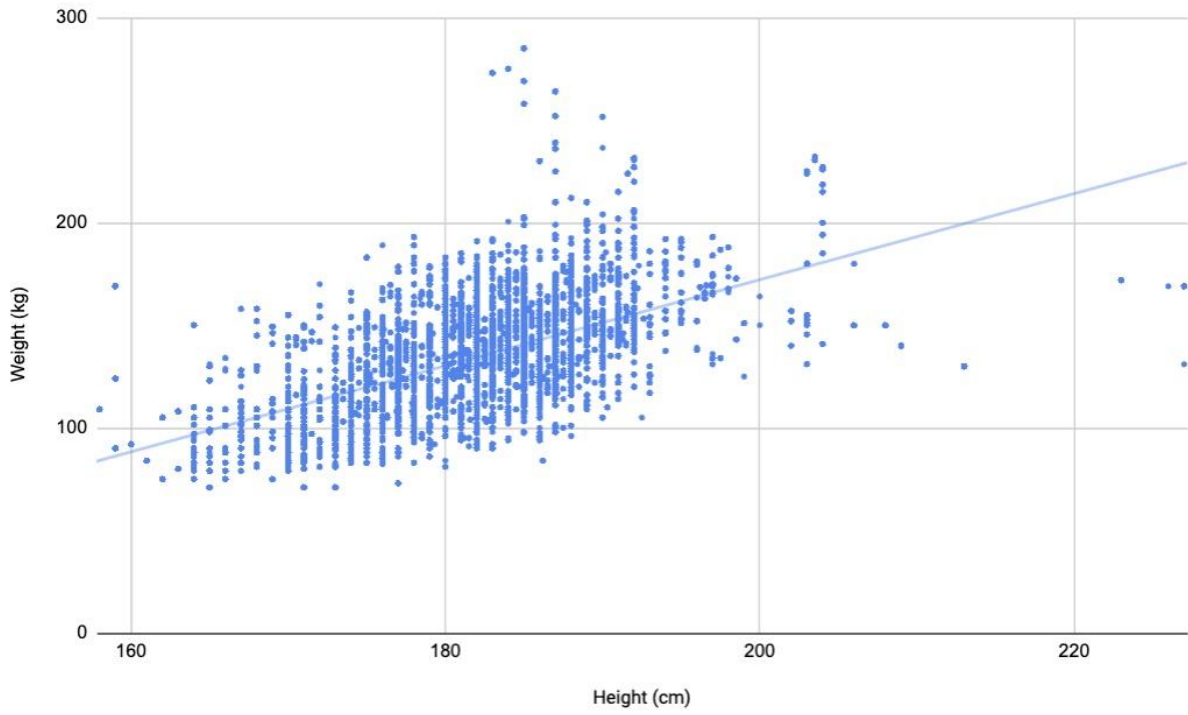
International *Rikishi* by Percent



*Shusshin by Number of Rikishi*



*Rikishi* Height by Weight Scatterplot



## References

- Department, U. S. W. (1995). Handbook on Japanese Military Forces. United States: LSU Press.
- Fry, Michael J., and Jeffrey W. Ohlmann. "Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications." *Interfaces*, vol. 42, no. 2, 2012, pp. 105–08. *JSTOR*, <http://www.jstor.org/stable/41472739>. Accessed 13 Mar. 2023.
- gans. (2023, January 23). *Hositori*. Sumo Hositori. Retrieved February 20, 2023, from <https://sumo-hositori.com/index.html>.
- Guttmann, A., & Thompson, L. (2001). The Modernization of Indigenous Sports. In *Japanese Sports: A History* (pp. 96–116). University of Hawai'i Press. <http://www.jstor.org/stable/j.ctt6wqsmj.8>.
- Japan Sumo Association. (2023). *に誤りがあります*. sumo.or.jp. Retrieved February 18, 2023, from <https://www.sumo.or.jp/>.
- Kasumigaseki, Chiyoda, Tokyo, & Masayasu, I., <https://www.jpo.go.jp> (2008). Japan Patent Office. Retrieved December 16, 2020, from [https://www.jpo.go.jp/e/news/kokusai/developing/training/textbook/document/index/Copyright\\_Law.pdf](https://www.jpo.go.jp/e/news/kokusai/developing/training/textbook/document/index/Copyright_Law.pdf).
- Manzenreiter. (2014). Cracks in the Moral Economy of Sumo: Beasts of Burden, Sport Heroes, National Icons and Living Gods in Disgrace. *International Journal of the History of Sport*, 31(4), 459–473. <https://doi.org/10.1080/09523367.2013.869213>
- Sakai, T. (1956). Nihon sumoshi (Vol. 1). Dai Nihon Sumo Kyokai.*

*Sakai, T. (1964). Nihon sumoshi (Vol. 2). Dai Nihon Sumo Kyokai.*

*Sumo Reference.* (2023, January). Retrieved January 25, 2023, from <http://sumodb.sumogames.de/Default.aspx>.