**Visualization Project**

**Jerry Mischel, DS 745**

**Visualization Summary**

In 2017, 6,105 home runs (HRs) were hit in Major League Baseball (MLB) – the most ever in history and a 46% increase since 2014. This inspired a thoughtful look at how the number of HRs per season has changed since the first major league baseball game in 1871 and exploration into the potential impactful factors, including the average player’s weight each season and steroid usage.

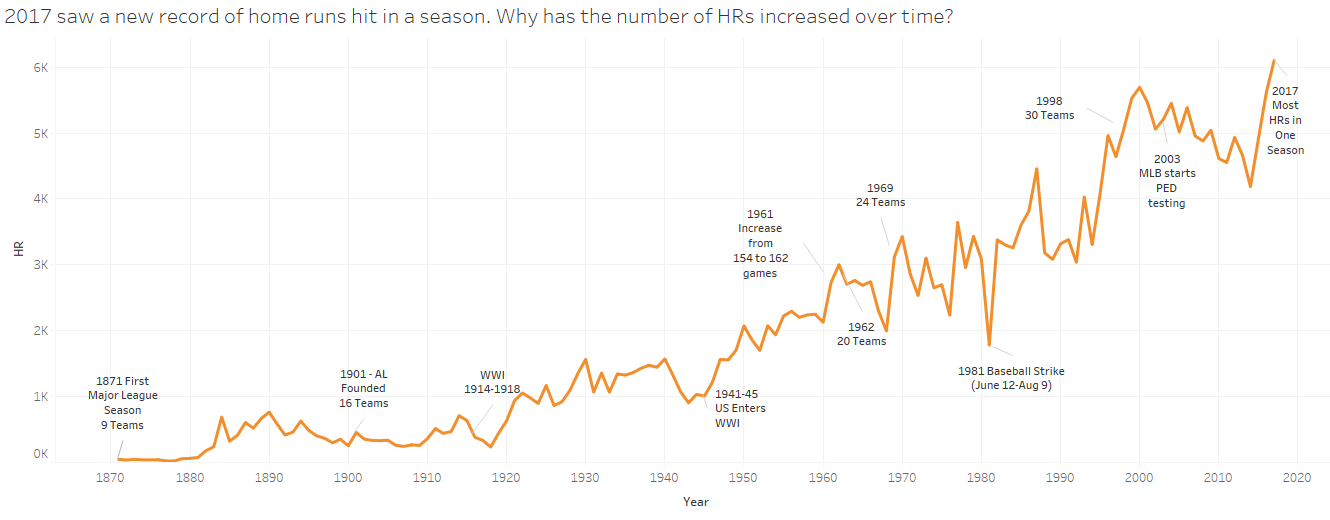
This visualization shows two time series graphs covering the time span from 1871 to 2017 – one (in orange) highlighting the number of HRs hit each season and a second (in blue) showing the increase in average player weights. Various MLB milestones or key dates—like adding teams, changing the number of games and a period called the “steroid era”—are referenced, using vertical dashed lines, to show a comprehensive story of the HR increases over time and explore why this may be happening. In this visualization, you should clearly see as players get stronger (which I’ve equated with weight) they hit more home runs. There are other interesting insights like an increase in home runs and weight during the “steroid era” and a decline in home runs after the MLB started to test for Performance Enhancing Drugs (or PEDs).

This specific dataset was provided by Sean Lahman, an award-winning journalist, on [www.seanlahman.com](http://www.seanlahman.com). Sean creates historical databases covering a gamut of sports, including baseball, pro and college football, pro and college basketball, auto racing, tennis, boxing and the Olympic games. For this visualization, I merged two spreadsheets together called Batting.csv (which included HRs) and People.csv (which included weights). There were a handful of missing values for player weights but not enough to skew the results.

**Iterative Process**

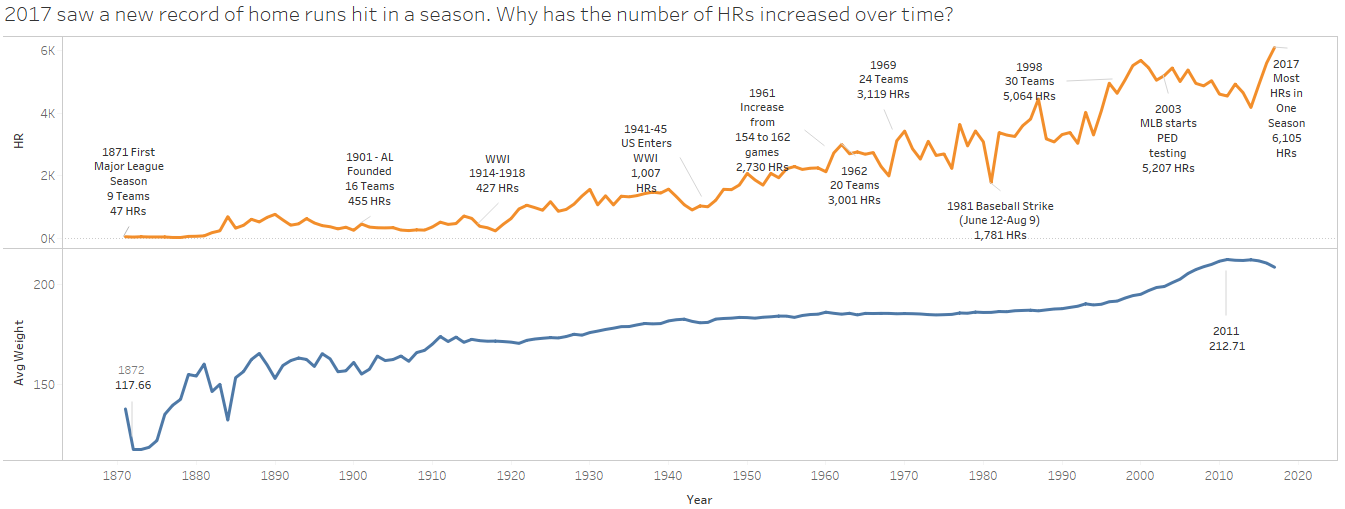
***Version 1 (See Version1.PNG)***

* Pros:
  + Easy to read graph with labels for milestones
  + Clear and accurate title
  + Color difference between graph and labels allows graph to “pop”
* Cons
  + I don’t love the labels as they seem to be floating and unconnected to graph, even though there are lines pointing to each mark.
  + While there are milestones which could impact number of HRs, there’s a missing element. There’s nothing describing the physical attributes of these players, like weight. This could provide key additional context for the reader.
  + Graph lines are a little distracting though they do highlight the number of HRs for each of the milestones. Remove gridlines and include numbers in the labels to help reader advance through the story.



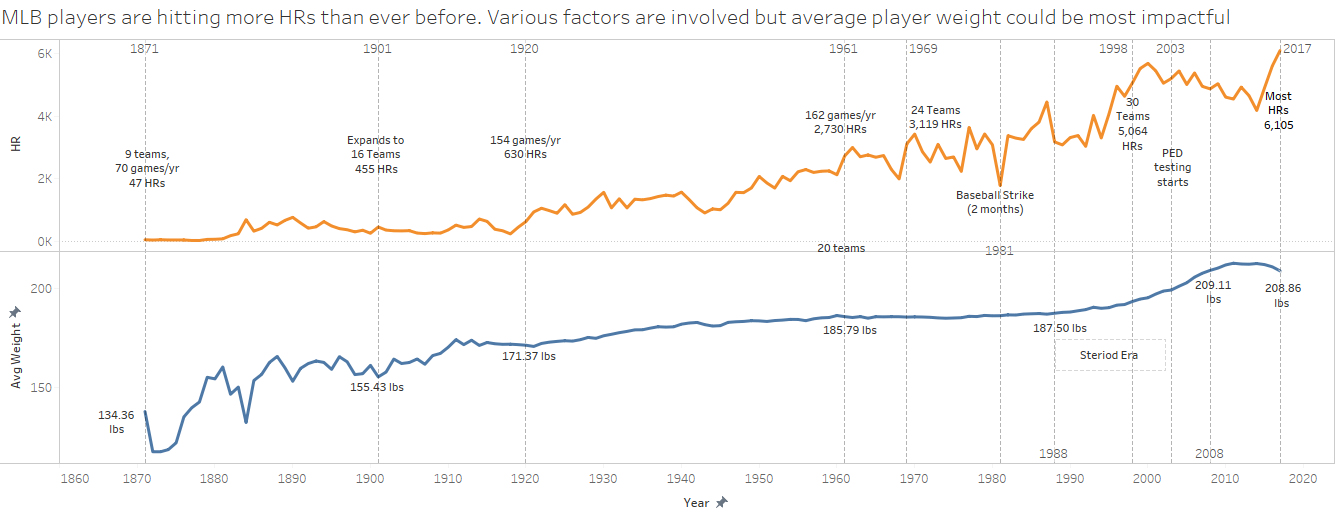
***Version 2 (See Version2.PNG)***

* Pros:
  + Addition of weight variable demonstrates a potential correlation with the increase in HRs. The stronger players become (i.e., the more average weight gain), the more HRs are hit.
  + Weight labels for minimum and maximum show progression over time.
  + Addition of HR numbers for each label help reader to understand the increase in HRs at each milestone.
  + Removing the gridlines provides a cleaner look and allow reader to focus on the data.
* Cons:
  + Two graphs seem to be floating and not connected. Is there a way to connect them together?
  + Too much clutter in HR graph with labels and too many words.
  + Too much empty space with weights. Need additional info to be more impactful.
  + Both graphs are on the same x-axis (years), but wondering if different y-axes (lbs vs HRs) misconstrue trend line.
  + There’s nothing linking the increase in HRs to potential steroid use, which has been called a primary reason for weight gain and increased strength to hit HRs.



***Version 3 (See Version3.PNG)***

* Pros:
  + Dashed reference lines, instead of faint lines marking points, helps to unclutter graph, link x-axis to activity on the y-axis, connect the two graphs and helps to mark milestones in baseball.
  + Additional weight references in bottom graph help to display increase in weight over time for reader. Also connects average player weight to milestone and HR count for that year.
  + “Steroid Era” reference point emphasizes this timeframe for reader, who will see an 11% increase in weight and a significant increase in HRs. This is a key era for HR hitting, so needs to “pop” more for reader. Thicker, blacker border helps to produce this effect.
* Cons:
  + The upper right-hand side of the HR graph is still a little congested, even though reference lines have helped reduce some of the congestion.
  + I don’t love that the years for the reference lines are located in different areas of the graph (top, middle and bottom) but I wanted to focus readers’ attention on one area for each reference line, if at all possible.
  + With the addition of weight and the steroid era box, a title change helped to describe the visualization more. While the title is sufficient to describe visualization, it could use a bit more punch.
  + “Steroid Era” timeframe box would’ve been better served to be drawn around the specific areas on both graphs, but Tableau doesn’t allow such “painting” functionality.



**References**

Arthur, Rob and Tim Dix. (2018). We X-Rayed Some MLB Baseballs. Here’s What We Found. *FiveThirtyEight.* Retrieved from <https://fivethirtyeight.com/features/juiced-baseballs/>.

Field Level Media. (2018). Major League Baseball report on home runs concludes no changes to ball structure. *Reuters.* Retrieved from <https://www.reuters.com/article/us-baseball-mlb-homeruns/major-league-baseball-report-on-home-runs-concludes-no-changes-to-ball-structure-idUSKCN1IP3F6>.

Lahman, Sean. Baseball database. Retrieved from <http://www.seanlahman.com/>.

Wikipedia. Doping in baseball. Retrieved from <https://en.wikipedia.org/wiki/Doping_in_baseball>.

Wikipedia. Timeline of Major League Baseball. Retrieved from <https://en.wikipedia.org/wiki/Timeline_of_Major_League_Baseball>.