**Project Title: Assessing MLB Salaries and Players’ Performance - Write-up Below**

Team Members:

* David Phillips
* Dan McCabe
* Priscilla Maddela

Project Description/Outline:

* Correlation between player salary and WAR per year – determine who was over/under paid for batters (potentially pitchers as well)
* 1986—2016 seasons only

Research Questions to Answer:

* Have teams been doing a better job of accurately paying players based on production over the last 25 years?
  + Are certain teams consistently better than others? If so, does that result in wins?
  + Regular season performance vs. postseason performance
* WAR stat vs Salary

Data Sets to be Used:

* War\_daily\_bat.csv from baseball-reference.com

Rough Breakdown of Tasks:

* Cleaning .csv file
* Analyze data
* Build visualizations
* Presentation

Considerations:

* If a player changes teams mid-season, their WAR score was split between the two teams. So we summed the WAR to get the total WAR for the season.
* If a player changes teams mid-season, their Salary is recorded with the team they played the most games with.
* Found the CPI library to adjust salaries for inflation (originally going to do 1975--2015, but we found that the salary data prior to 1985 was incomplete and found it to be skewing the data (only higher paid players’ salaries were recorded). We then changed the dates range to focus on 1986--2016.
* Some franchises rebranded or moved cities, so we merged those teams. EX: California Angels > Anaheim Angels > LA Angels For those teams, we renamed those to the team which exists today (2019). Full list of affected teams:
  + "CAL": "LAA", "ANA": "LAA", "MON": "WSN", "TBD": "TBR", "FLA": "MIA"
* Some players signed multiple contracts in the same year. There were very few that we found, however they were paid the minimum salary, performed at the minimum WAR (basically our hypothetical replacement player)
* DEFINE WAR SCORE: Wins Above Replacement
  + No clearly established formula exists for WAR. Sources that provide the statistic calculate it differently.
  + Baseball Reference uses six components to calculate WAR for position players:The components are batting runs, baserunning runs, runs added or lost due to grounding into [double plays](https://en.m.wikipedia.org/wiki/Double_play) in double play situations, fielding runs, positional adjustment runs, and replacement level runs (based on playing time). The first five factors
  + are compared to league average, so a value of 0 represents an average player.
    - *Source:* [Wikipedia](https://en.m.wikipedia.org/wiki/Wins_Above_Replacement)
  + WAR is based on previous performance - EX: Trout

**Conclusions:**

1. WAR and Salary correlation is not very strong, however it is still positive. So even though there isn’t a strong correlation, better players are paid bigger contracts.
2. Average cost per WAR (overall chart red line) shows the that players salaries are going up over time (payroll is getting more expensive)
3. Overall, with the ‘moneyball’ movement in the MLB teams are also looking for inefficiencies in the market for players, and are finding ways to pay good players as little as possible, which will create weaker/negative correlation between salary and WAR.
4. Where teams fail is paying a lot of money to unproductive players, this typically happens at the end of players contracts where they are making the same amount of money but their productivity has drastically dropped.
5. The trendline and correlations are flattening out in recent years, probably because teams are so focused on obtaining young, highly productive players that they don’t have to pay much money. Teams focus on this because it then allows them to spend money on veteran players that seem like more of a sure bet to be productive, however that does not always pan out. In order to build a highly successful.championship caliber team they seem to need to get highly productive young players then spend the remaining money wisely on the free agent market.

**Presentation:**

What are you interested in presenting?

**PRESENTATION OUTLINE:**

1. **Introductions (ALL)**
2. Why did we choose this topic? **(DAN)**
3. What is WAR? **(DAN)**

WAR is an all-encompassing stat that stands for wins above replacement. It is meant to show how valuable a player is above a ‘replacement level’ player. This hypothetical replacement player is readily available to be signed or called up from a team’s minor league system.

1. Questions we set out to answer **(DAN)**
   1. Is there a correlation between a player’s WAR score and their Salary, i.e. does a player with a higher WAR score receive a higher salary?
      1. Which players are “overpaid” based on their WAR score?
      2. Which players are “underpaid” based on their WAR score?
   2. The assumption is that teams with higher average overall WAR score but also lower payroll (salary) are better at managing their ROI. Are there teams who are able to
2. Where did we find the data? **(DAN)**
   1. Data exploration - found our source on BaseballReference.com (<https://www.baseball-reference.com/about/war_explained.shtml>)
3. Cleaning the data - some things we found:**(PRISCILLA)**
   1. Baseball reference is an AWESOME source for data - wealth of knowledge!
   2. Lucky to find one spreadsheet that had all the data we needed - just need to clean up prior to our investigation
4. Once we had the final data set, we had to do some more cleaning **(PRISCILLA)**

* If a player changes teams mid-season, their WAR score was split between the two teams. So we summed the WAR to get the total WAR for the season.
* If a player changes teams mid-season, their Salary is recorded with the team they played the most games with.
* Found the CPI library to adjust salaries for inflation (originally going to do 1975--2015, but we found that the salary data prior to 1985 was incomplete and found it to be skewing the data (only higher paid players’ salaries were recorded). We then changed the dates range to focus on 1986--2016.
* Some franchises rebranded or moved cities, so we merged those teams. EX: California Angels > Anaheim Angels > LA Angels For those teams, we renamed those to the team which exists today (2019). Full list of affected teams:
  + "CAL": "LAA", "ANA": "LAA", "MON": "WSN", "TBD": "TBR", "FLA": "MIA"
* Some players signed multiple contracts in the same year. There were very few that we found, however they were paid the minimum salary, performed at the minimum WAR (basically our hypothetical replacement player)
* **Overall, as we worked through the data, we found more opportunities to clean up the data in order to continue with our analysis. So, working through our data, even being a single .csv, we still had several cleanup activities.**

1. Analysis - what did we set out to find? **(DAVID)**
   1. **MAIN QUESTION:** Have Teams gotten better at more accurately paying players based on their performance (WAR)
   2. OTHER QUESTIONS WE CONSIDERED:
      1. What is the correlation between WAR and salary?
      2. How does that correlation trend over time? (1986--2016)
         1. Look at the scatterplots for 1986, 1996, 2006, 2016
         2. 3 axis chart - WAR, Salary, over time
         3. Avg cost per WAR over time (we made up - formula is, salary/WAR. We did this for players and for teams overall)
         4. Yearly cost per WAR per team
         5. How do teams trend individually over time, based on the average cost per WAR stat
      3. Are teams that have higher WAR scores on average have higher payrolls?
      4. Do players have to wait long before they get the big offer?
2. Conclusions **(DAVID)**
   1. WAR and Salary correlation is not very strong, however it is still positive. So even though there isn’t a strong correlation, better players are paid bigger contracts.
   2. Average cost per WAR (overall chart red line) shows the that players salaries are going up over time (payroll is getting more expensive)
   3. Overall, with the ‘moneyball’ movement in the MLB teams are also looking for inefficiencies in the market for players, and are finding ways to pay good players as little as possible, which will create weaker/negative correlation between salary and WAR.
   4. Where teams fail is paying a lot of money to unproductive players, this typically happens at the end of players contracts where they are making the same amount of money but their productivity has drastically dropped.
   5. The trendline and correlations are flattening out in recent years, probably because teams are so focused on obtaining young, highly productive players that they don’t have to pay much money. Teams focus on this because it then allows them to spend money on veteran players that seem like more of a sure bet to be productive, however that does not always pan out. In order to build a highly successful.championship caliber team they seem to need to get highly productive young players then spend the remaining money wisely on the free agent market.
3. Future work/analysis: **(DAVID)**
   1. If we had a little more time, we would have been able to focus in on specific teams. That could be an interesting next step for this analysis - which teams are really good at this, which are really bad?
   2. Another thing we wanted to focus in looking at player career vs calendar years and see how this stat changes over time. When do players peak? When to they start to drop off? What year is a ‘sweet-spot’ for a player to get their big contract?
   3. It would be interesting to focus on only more impact players, possible only ones who played 140+ games in a season and played for 10+ seasons to eliminate many fringe players or ‘replacement’ players and look at how they were compensated over their careers and if that was more accurate than the total player population