

# AL & NL Batting Totals (HR) During the “Steroid Era.”

Final Capstone | Thinkful | Data Analytics  
By: Lisa Alcock

# Background



- Sammy Sosa, Mark McGwire, Barry Bonds, Jason Giambi, A-Rod, Jose Canseco, Johnny Damon, Manny Ramirez, David Ortiz, Aaron Boone, Gary Sheffield, etc. All of these men have at least one thing in common: failed MLB drug tests and/or suspensions.
- Because of the use of performance enhancing drugs (PEDs) by several players, the decade of 1998-2008 became defined as “The Steroids Era.”

# Datasets

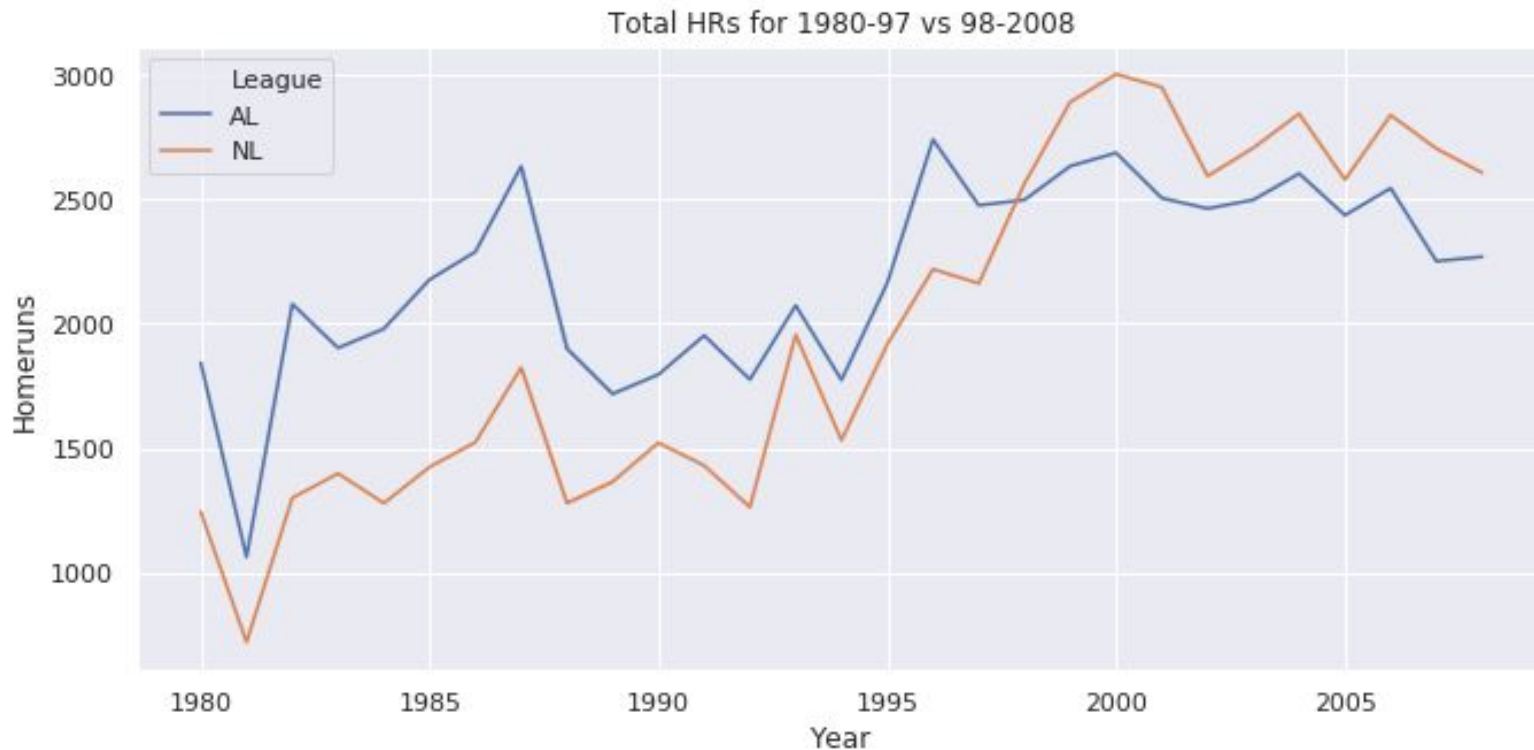
- I decided to look at AL & NL batting totals datasets that came from [Baseball Reference](#) . (I am focusing on HR's because other stats are already averaged in the data.)
- I also grabbed data from a widely-publicized list of players that failed drug testing & were suspended. (Players also named in The Mitchell Report & the Federal grand jury investigation of BALCO).

# Research Questions

- My research question(s):
  - Is there a statistical significance when different years between 1998-2008 are compared?
  - Is there statistical significance between players who tested positive/were suspended for PEDs vs. those that didn't take any PEDs?
- My null hypotheses are that drug use had no effect on the HR numbers over this time period AND, that players who used PEDs had HR numbers that weren't any different from players who didn't take PEDs.

# Total HRs: 1980-1997 vs. 1998-2008

If you look at the HR numbers from the decade prior and 98-2008, in tables online, you can see there's an increase in HRs in both leagues and then a decline for this time period.



# Findings--Statistical Significance

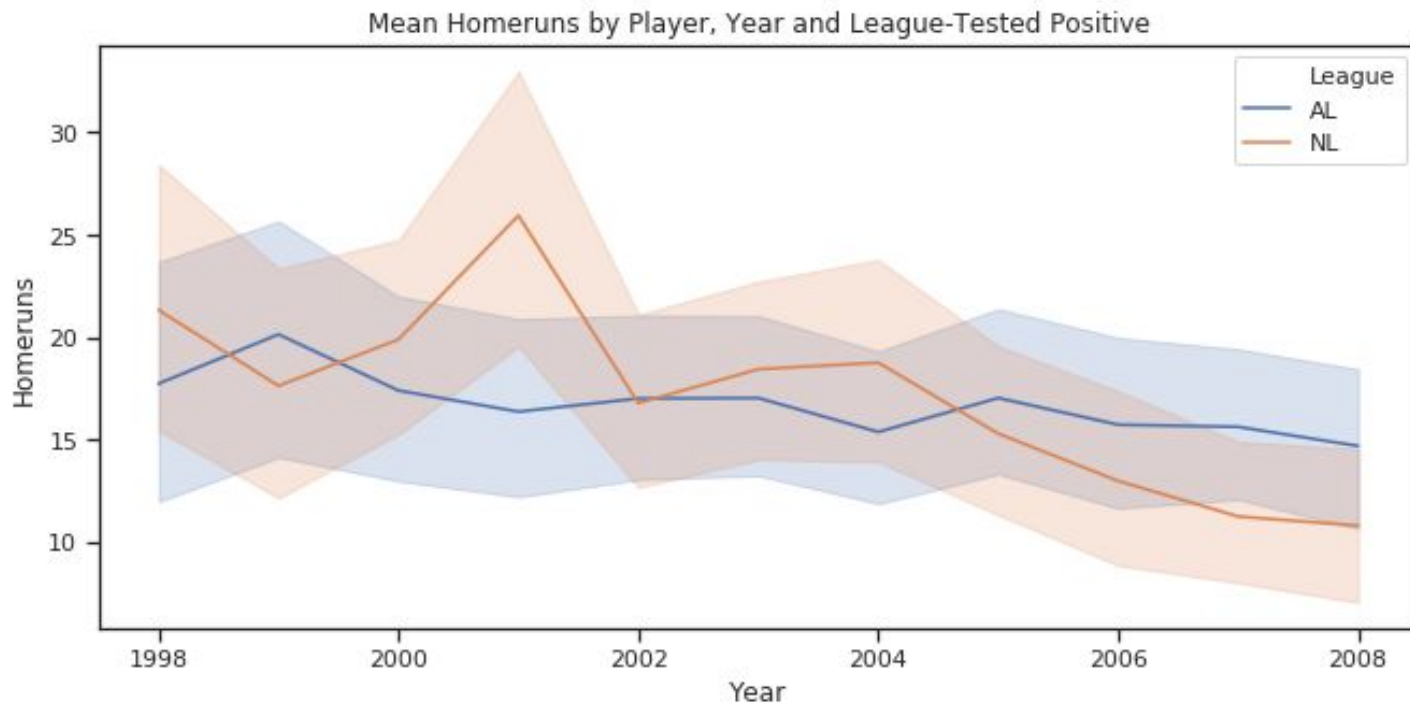
After hypothesis testing, I found statistically significant differences between certain years.

The samples with significance:

- Positive vs. No PEDs, 1998-2008
- AL vs. NL batting (HR's), 1998-2008
- 1999 vs. 2007
- 2000 vs. 2008
- 2004 vs. 2008

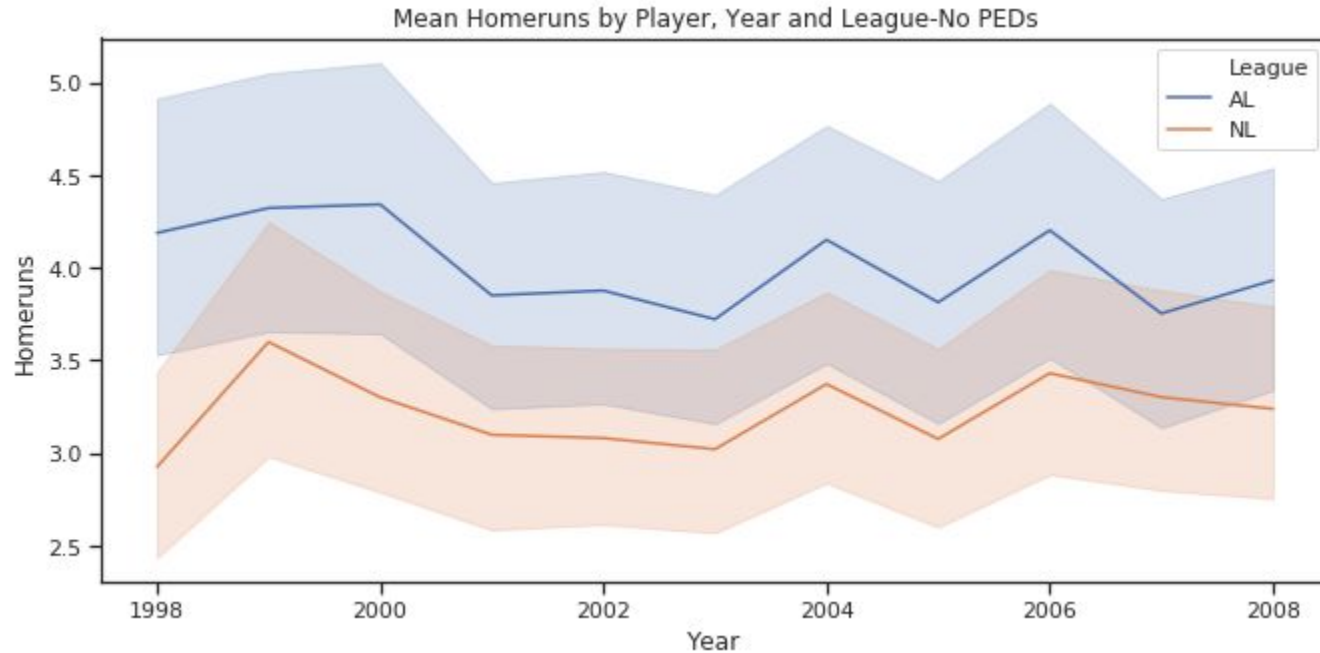
# HR's Players Who Tested Positive and/or Were Suspended

**Mean HRs: 16.95**



# Players, No PEDs

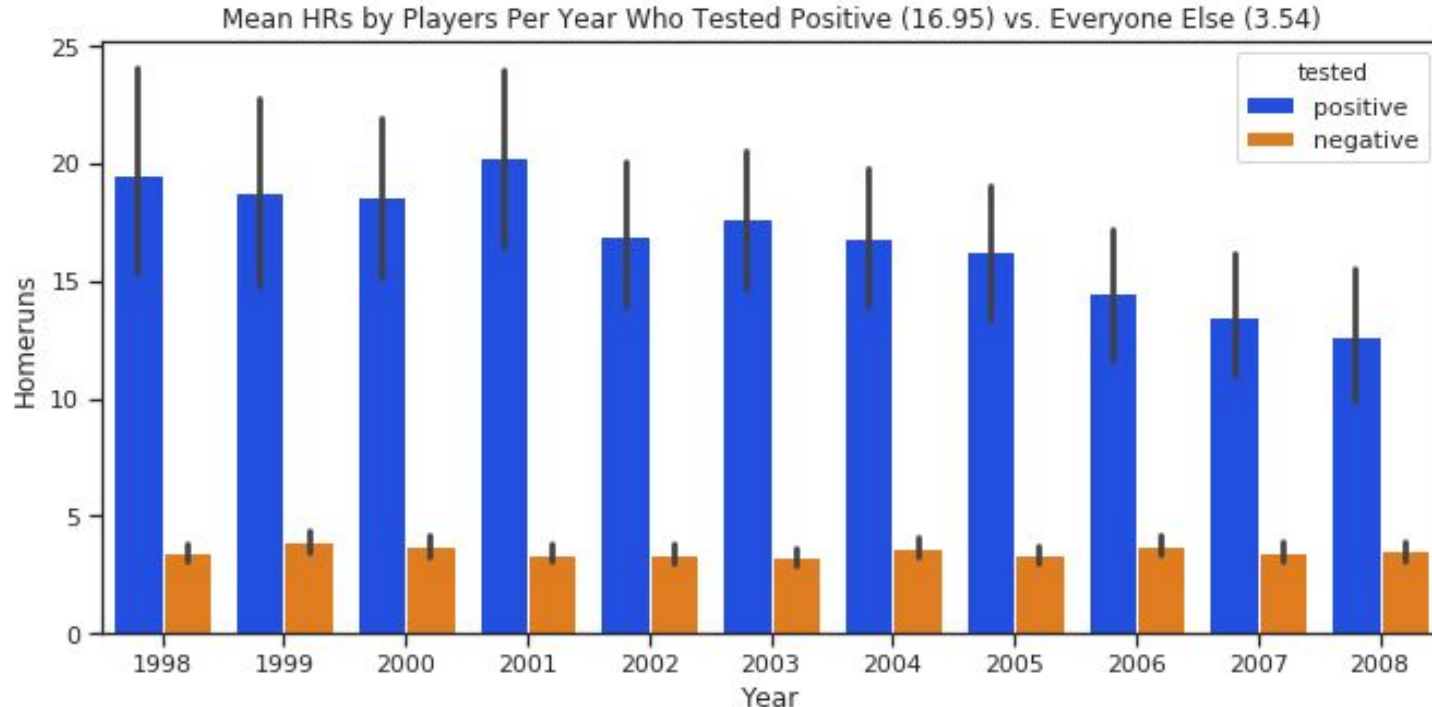
Mean HRs: 3.54





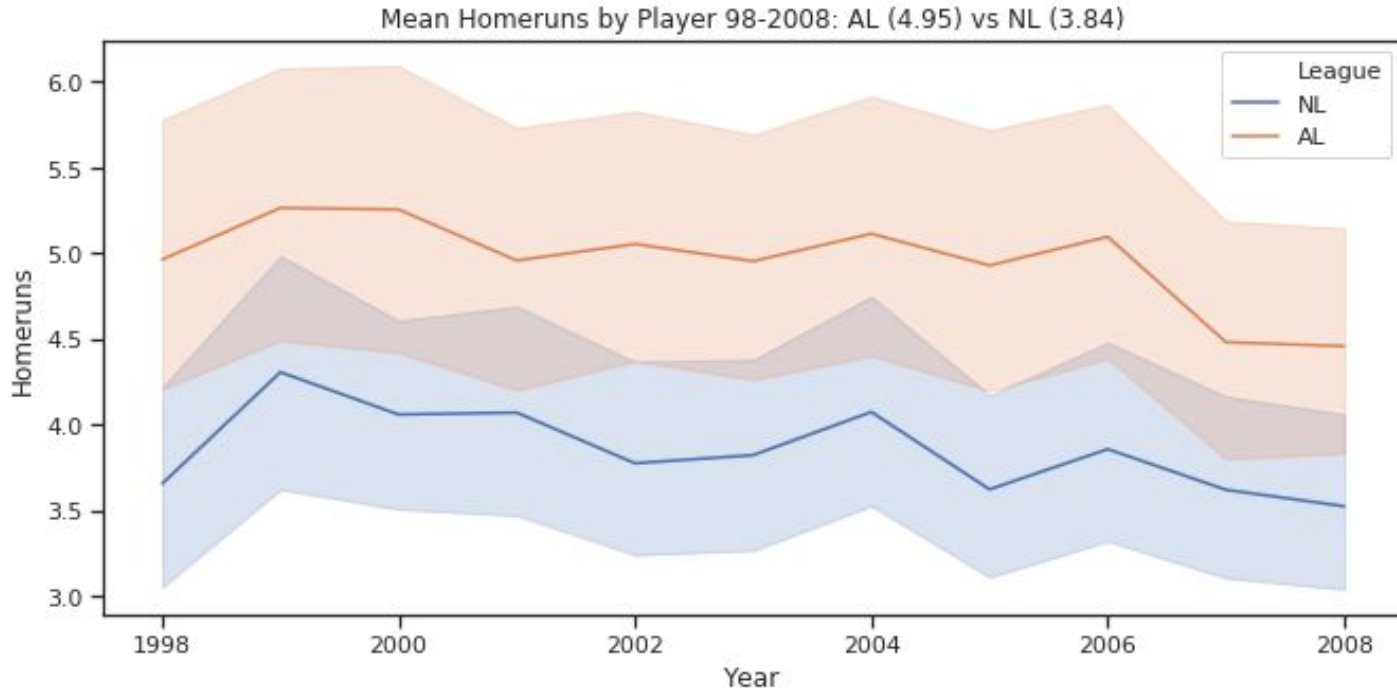
# Players Who Tested Positive vs. No PEDs: 98-2008

**p=0.000, @95%: -14.423 & -12.38, CD: 1.19**



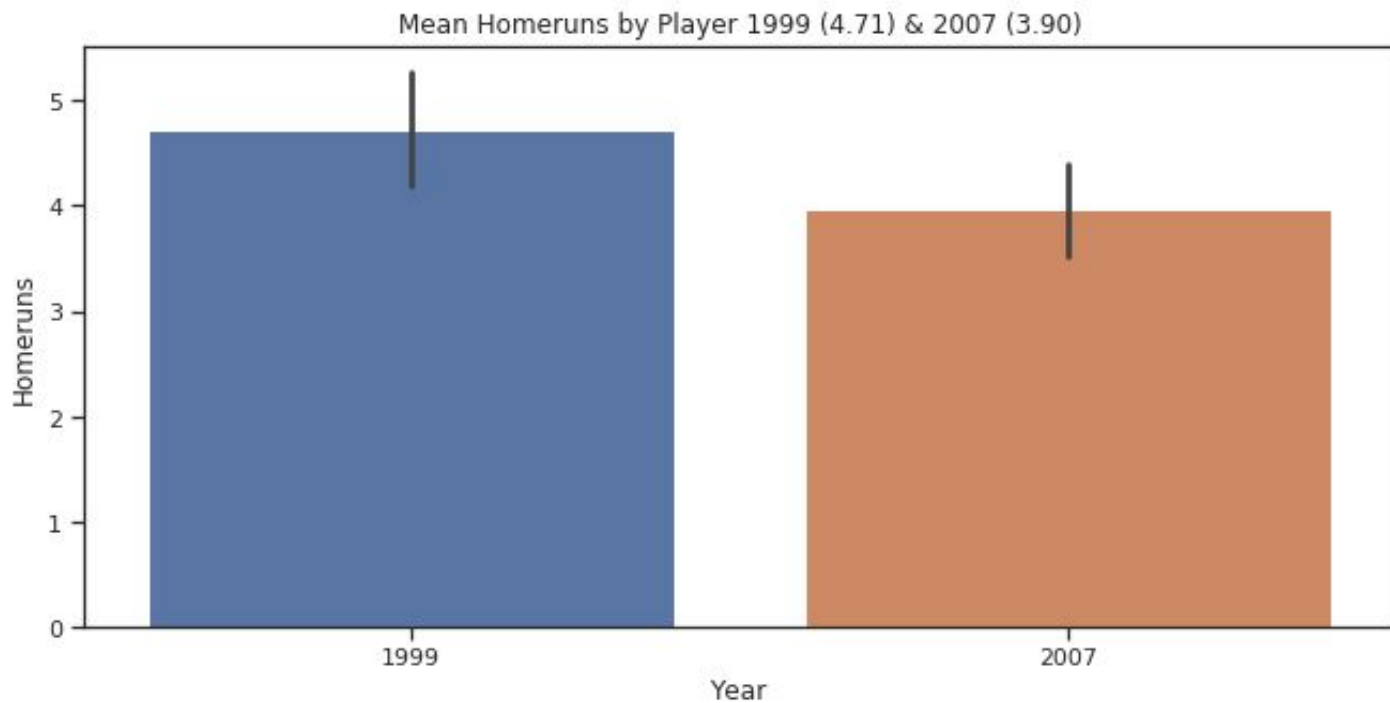
# AL vs. NL HRs 1998-2008

$p=2.242e-14$ , @ 95%: 0.819 & 1.394, CD: 0.13



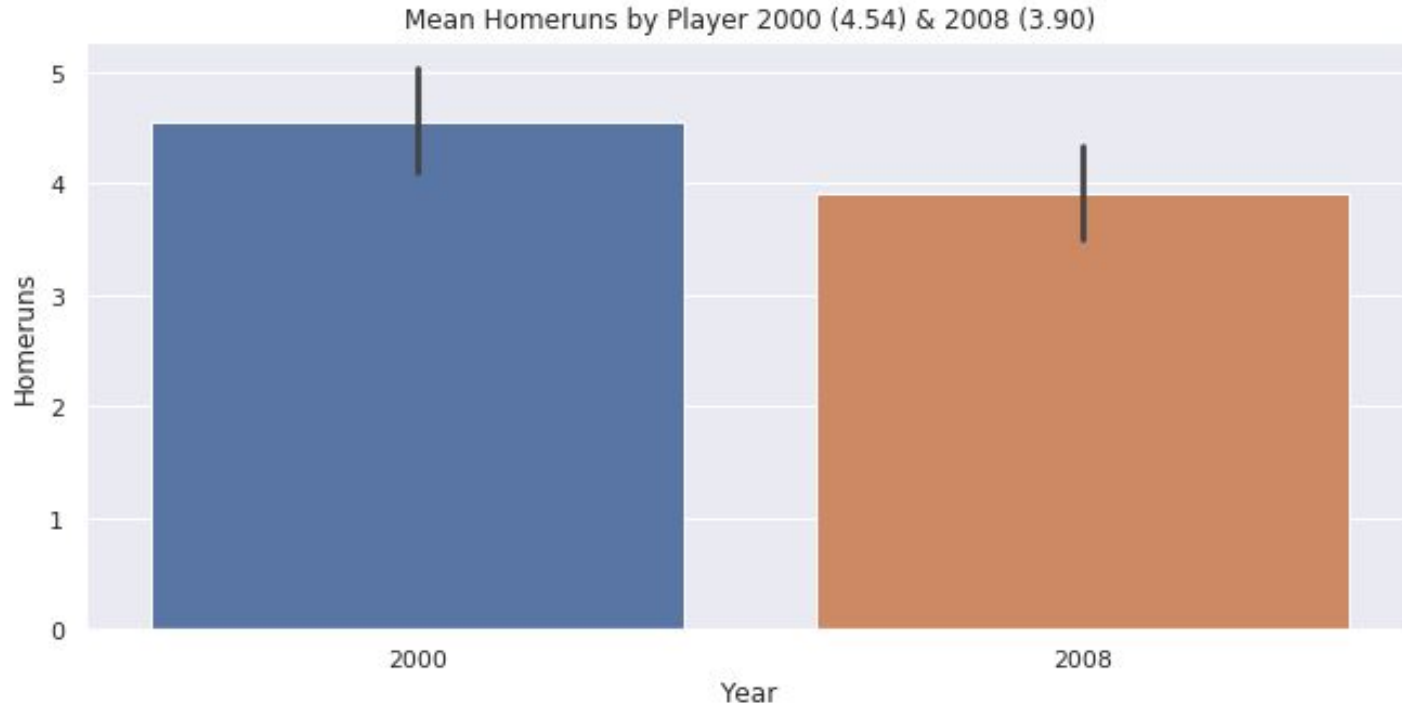
# 1999 vs. 2007

**p=0.027, @ 95% confidence: -1.419 & -0.079, CD= 0.08**



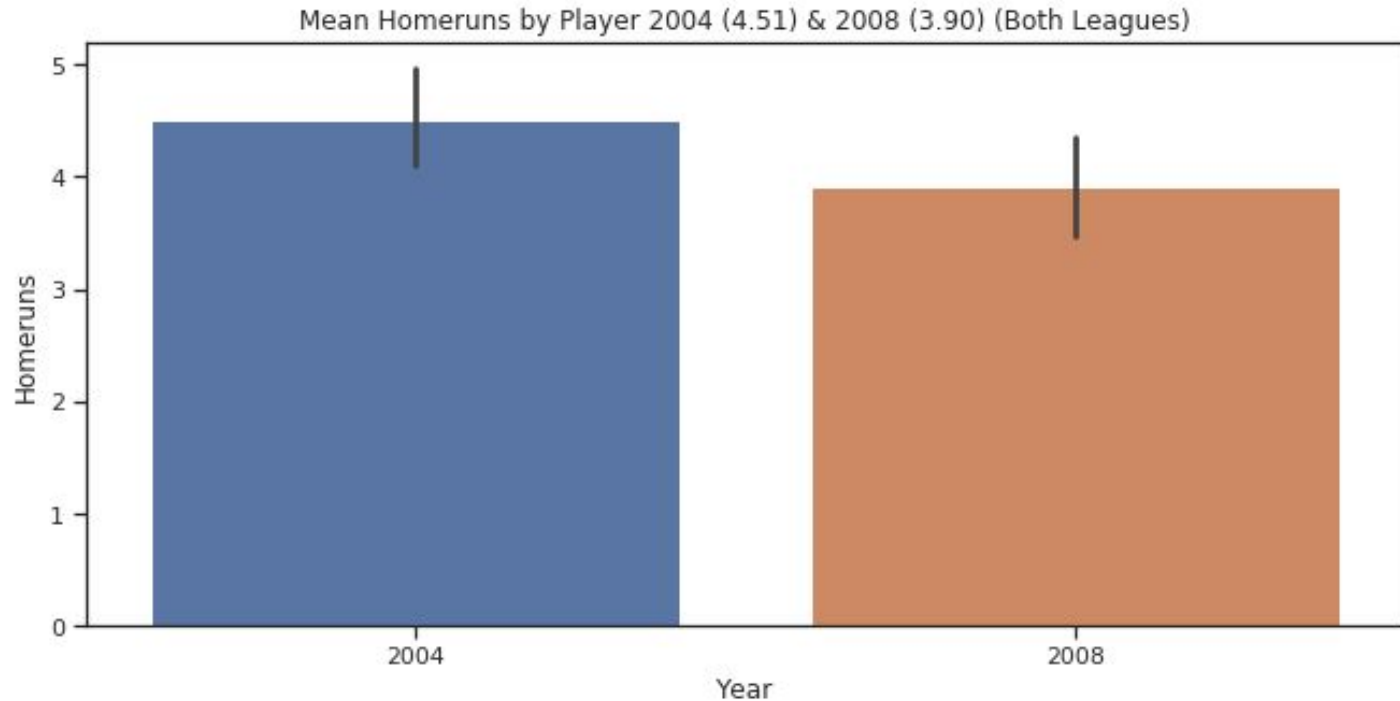
# 2000 vs. 2008

**p=0.047, @ 95% confidence: -1.27 & -0.008 HRs, CD: 0.07**



# 2004 vs. 2008

**p=0.05, @ 95%: difference in HR means is between -1.237 & 0.020. CD: 0.07**



# Final Thoughts, Things to Think About

- While I didn't find significance for all years, I did find 5 different instances of statistical significance in the data. However, the Cohen's  $d$  numbers for several were quite small in their effect sizes between the variables.
- I think my findings add strength to the argument that PEDs did affect HR numbers.
- I am still encouraged by my findings here and I will continue to study this subject. I'd like to create some more datasets, post-2008-present and run statistical tests on them, as well as some for the 1980's-90's.

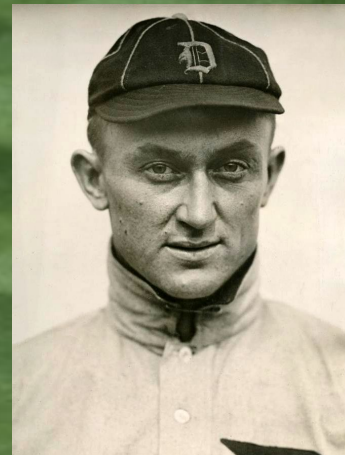
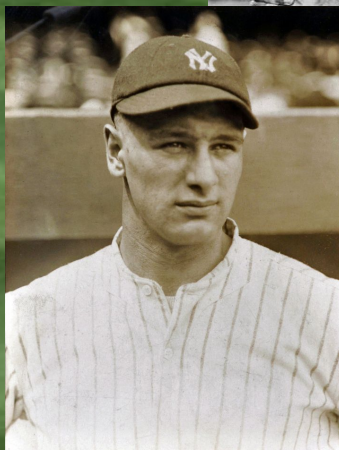
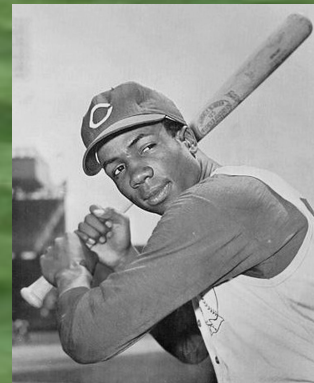
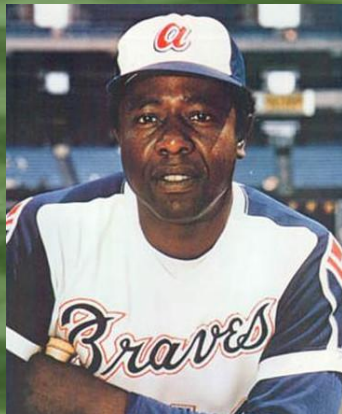
# The Effect on MLB

In 2007, I think The Mitchell Report summed it up perfectly:

*“The illegal use of performance-enhancing substances poses a serious threat to the integrity of the game. Widespread use by players of such substances unfairly disadvantages the honest athletes who refuse to use them and raises questions about the validity of baseball records.”*



# Thank you for attending!





# Appendix - MLB Drug Policy - (for reference)

## PED Policy.

### A Brief History of MLB drug testing that's relevant to this study:

- **1991-** Steroids made the list of banned substances. However, there wasn't official testing in MLB.
- **2003-2004-** The Players Assoc. (MLBPA) & owners agreed on a new (anonymous) testing policy. Urine tests for banned substances began. A first positive test results in treatment and a second in a 15-day suspension without pay or fine of up to 10K. Penalties increase to a 25-day suspension or fine of up to 25K for a third positive test, a 50-day suspension or a fine of up to 50K for a fourth and a one-year suspension or fine of up to 100K for a fifth.' (USA Today, 2014).

# Appendix

\*Also, of note: It was later learned that the results of these tests from 2003 were then turned over to an Federal Grand Jury investigation into BALCO (Bay Area Laboratory Co-Operative). A list of players was leaked and became published online in many venues.

- **2005-** The players & owners agreed again on stronger penalties: 1st time: 50-day suspension, 2nd time: 100-day ban, 3rd: lifetime
- **2006-** Then Commissioner of the MLB, Bud Selig, ordered an investigation into illegal substances in MLB. He appointed Sen. George Mitchell, who led a 20 month investigation into steroid and HGH (human growth hormone, also banned) use in MLB. Former batboys, clubhouse attendants and trainers were questioned. Out of 5 players that were asked to co-operate b/c of their public comments about testing and drug use in MLB, only one, Frank Thomas gave testimony.

# Appendix

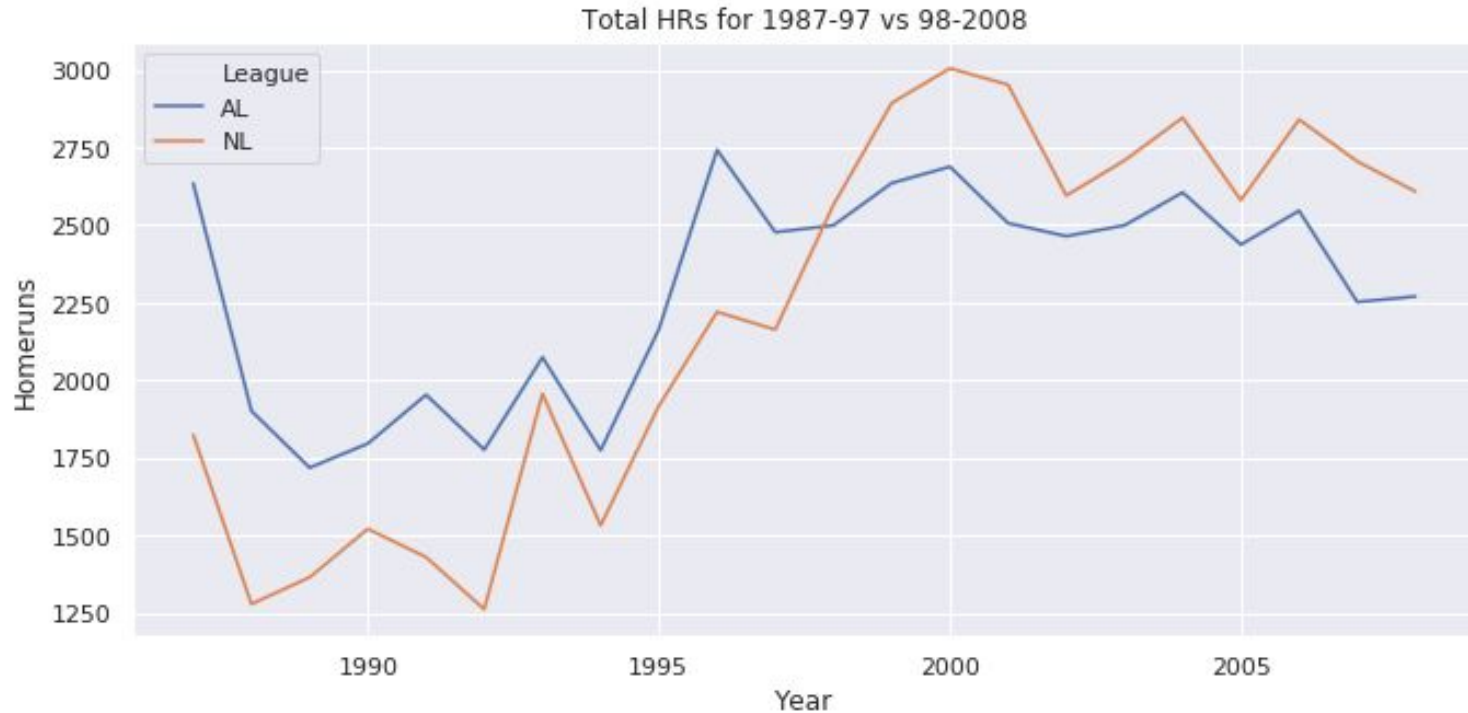
- The results of the investigation were published in a report in December 2007. Mitchell's investigation focused on "high profile" players. (online, Wikipedia, ESPN, etc.)
- The drug policy has changed several times since the Report came out and penalties have become more strict, including heavy fines for not attending treatment programs.
- In Jose Canseco's 2005 book, "Juiced," he named several players--McGwire, Rafael Palmeiro, Pudge Rodriguez, Juan Gonzalez, Jason Giambi. All players claimed his allegations were false. Canseco also said that amphetamines were more widely used in MLB.
- Bonds & McGwire also testified during the federal grand jury investigation of [BALCO](#).

# Appendix - Data References

- Baseball Reference [site](#).
- Players named in the Mitchell Report, [here](#).
- [The Mitchell Report](#).
- Players suspended, [here](#).
- The list leaked in 2009 of the players who tested positive in 2003, [here](#).
- MLB Drug Policy Timeline Coverage, [here](#).
- Interesting article about players who took steroids & HOF chances, [here](#).

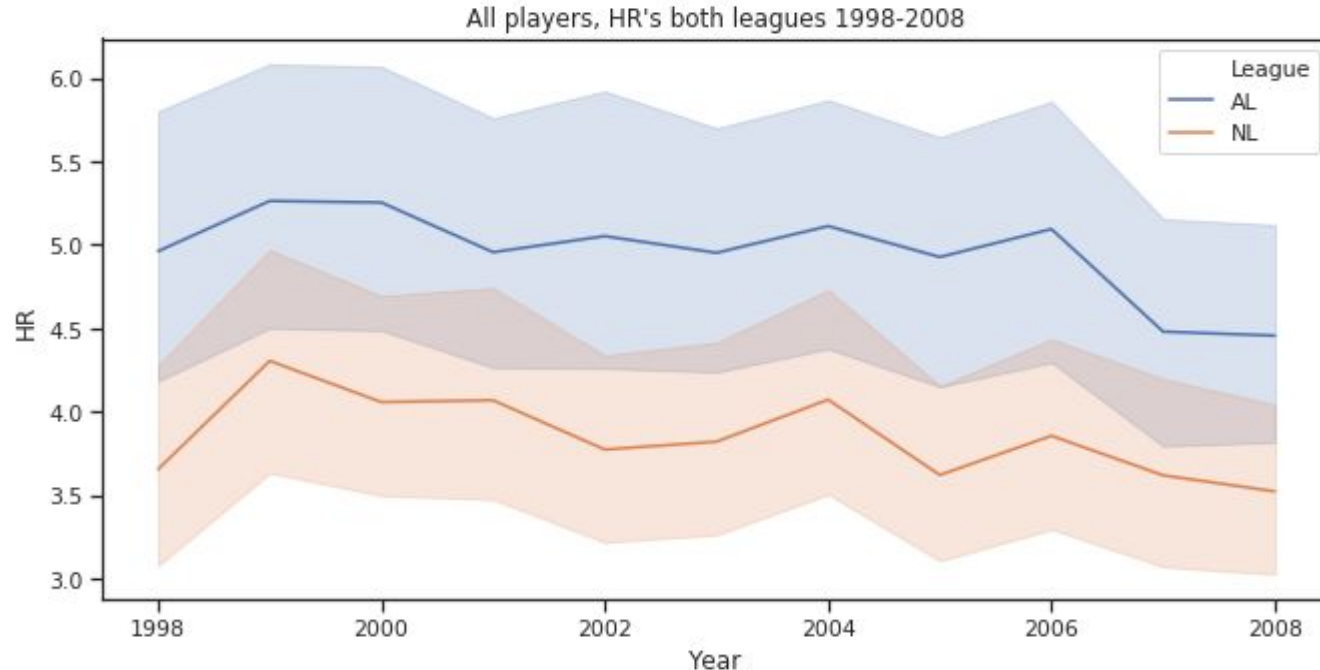
# Appendix: Total HRs 1987-97 vs. 1998-2008

P < 0.05, Difference in the means of HRs @ 95% is 547 & 927, CD = -2.29



# Appendix: Looking at Overall HR Numbers (mean)

If you look at the HR numbers from 1998-2008 in tables online, you can see there's an increase in HRs in both leagues and then a decline for this time period.



# Appendix

## Footnotes (slide 2):

1 Maris's record was controversial at the time b/c he beat out Babe Ruth, who had previously held the record with 60 HR in 154 games while Maris hit 61 in 162 games.

2 I also realize that other possible factors that could affect HR's--like weather (ie: in Colorado with high altitude) or, stadiums that are called "hitters ballparks", ie: the fact that many ballparks have reduced the sizes of left and right fields by bringing in the fences, and the existence of pitcher's parks. However, I'm not going to look at those possible variables here.