Response to Pachter's Review

Joshua Paik and Igor Rivin

1/22/2020

Contents

> *"Academic politics are so vicious precisely because the stakes are so small."* - Henry Kissinger
>
> *"The value of preprints is in their ability to accelerate research via the rapid dissemination of methods and discoveries."* - Lior Pachter

Introduction

Lior Pachter published a review of our paper claiming that age was the greatest contributor to the observed difference in citations between signers of Letters A, B, and C. Any characterization of our paper which says we do not account for age is *false*. In this response, we will clarify a few points in our paper, repeat the relevant analyses, and show that citations per year is an age agnostic metric to compare mathematicians. We will also go to some length to address potential objections to this new analysis, and also show they are categorically false. **To clarify, when comparing mean and median citations per year amongst R1 Math Professors, A < B < C.** This result still stands, as does the rest of our analysis, post some revision to our data. Finally, we will clearly demonstrate Professor Pachter's mistake, namely parameter tuning (hypertuning) of his arbitrary age cutoff, to acheive a result which supports his incorrect interpretation.

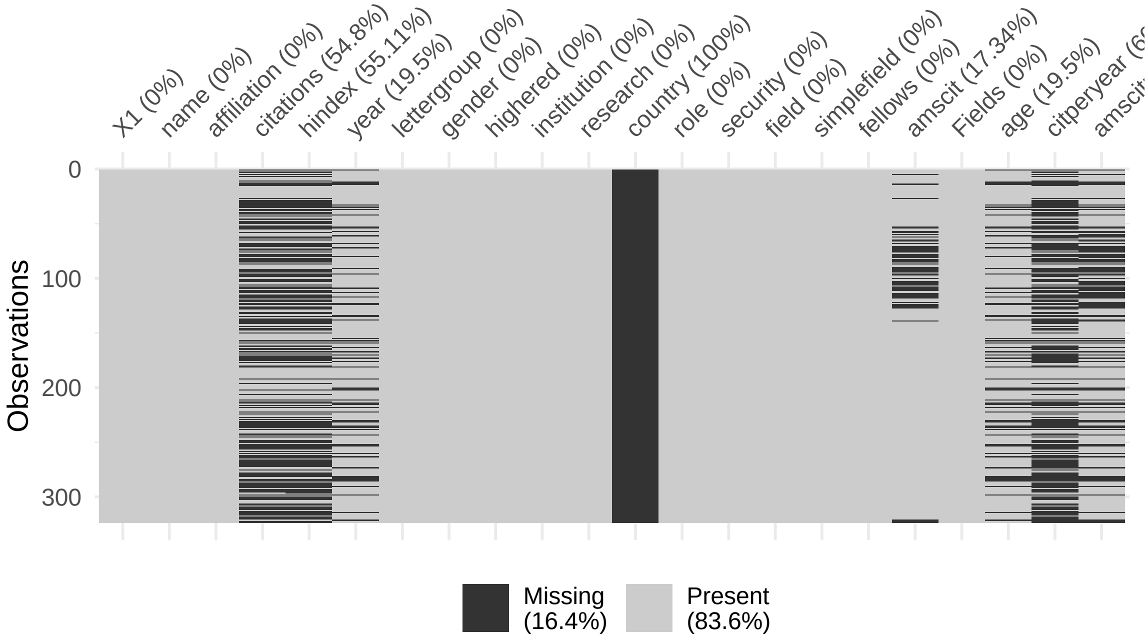
We appreciate Professor Pachter's review - it gives us a chance to make our analysis stronger. We would note that a revision was already in the works and that in a normal review process we would have had three months to respond. However, as our character and ability as scientists were attacked, we thought it was appropriate to reply as quickly as possible.

Corrections and Clarifications

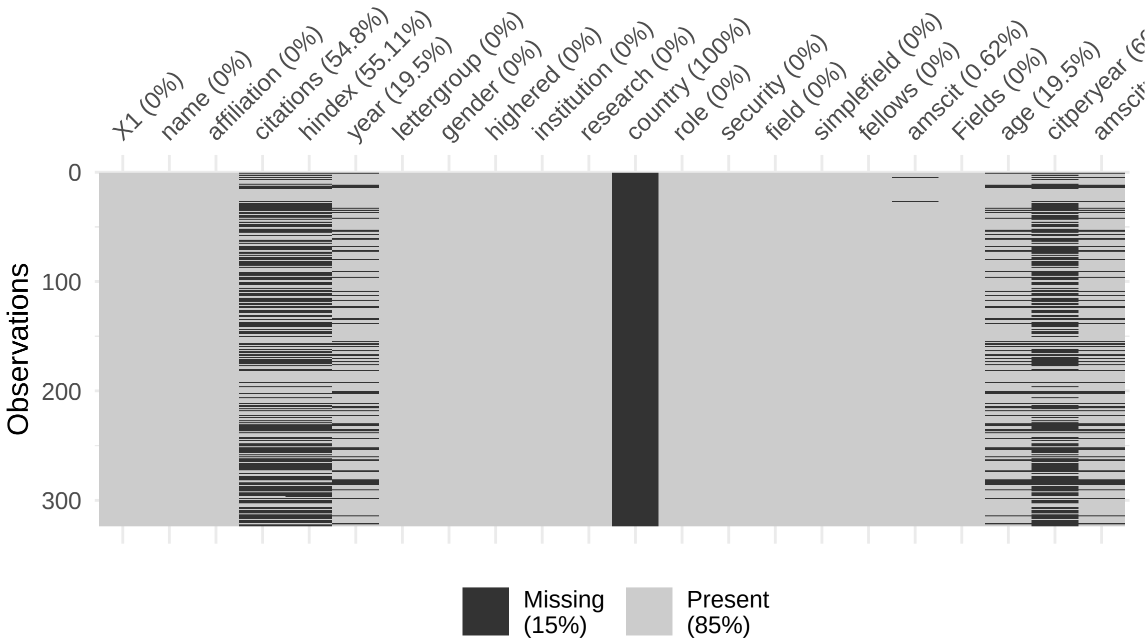
We would like to thank Lior for finding the bug in our appendix which pushes the mean Google Scholar citations of B further away from A. We agree that the sentence - "while this is not optimal, a quick sample size calculation shows that one needs 303 samples or 21% of the data to produce statistics at a 95% confidence level and a 5% confidence interval." - is ridiculous.

We should explain exactly how the data collection took place. We used the scholarly api to initially collect our Google Scholar citations data. But the issues were that the scraper did not accurately differentiate between those who had a generic name and the observed fact that older mathematicians (like Cheeger or Gromov) do not have Google Scholar citations. To assure data quality, we manually checked the google scholar citations of every single letter signer, comparing publications when necessary.

However, the empirical difference in citations was staggering and we could predict an objection. More professors from R2 (teaching focused universities) signed A, so it could have pushed the average down. We had already spent so much effort collecting Google Scholar citations, so we made a choice to only collect MathSciNet data on R1 Math Professors, which is why Professor Pachter did not have MathSciNet citations in our dataset. This choice was not made explicitly enough in our first version. Let us look at the NaN values of those who are full math professors at R1 univesities.



One sees that 17.34% of the Math Sci Net citations data is missing. It appears there was some sort of systematic but unintentional error in the data collection from MathSciNet. We report 3 NaNs on A and B, 3 on A only, 0 on B and C, 50 on B Only, and 0 on C only. We manually checked the missing data and find that all but Marta Civil and Jeffrey X Watt (who are math educators) have Math Sci Net entries. The remaining omissions are fixed and we visually check for NaNs again.



65/323 is empty for AMS citations per year. While visually the nans appear uniform, we will impose a stricter signifance level, say 2%, to assess the difference in AMS citations per year.

Now that we are comparing apples to apples, we reperform the main results.

The Main Result of Paik-Rivin: R1 Math Professors Citations and Citations per Year

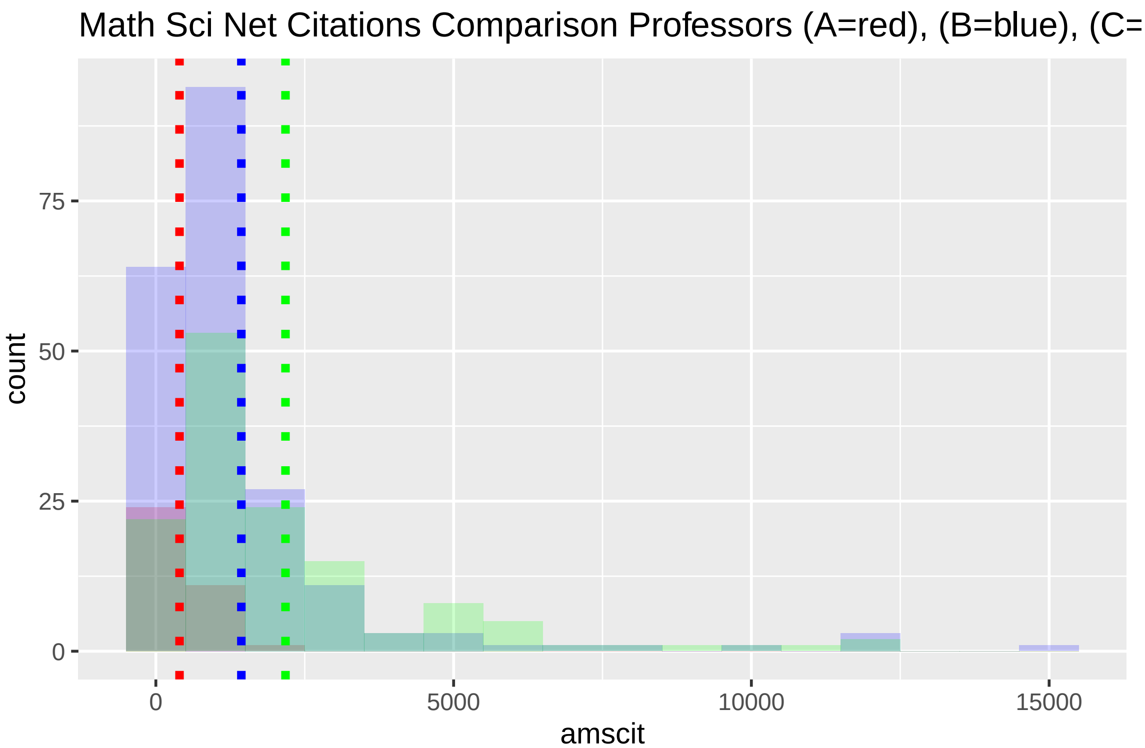
We will compare the mean number of citations and citations per year between signers of Letters A, B, and C. We will validate the significance of the difference between signers using a permutation test.

The following is copy and pasted from our paper (with edit in parentheses):

A permutation test is a nonparametric way of assessing the difference in mean between two populations. We are interested in whether an observed difference in mean is due to chance, and we can assess this in the following way.

0. Record the true difference in mean (or median) ($d\mu$).
1. $H_0 : d\mu = 0, H_1 : d\mu < 0$
2. Sample without replacement 1/2 of the combined data set (X) and what is left (Y)
3. Take the mean (or median) of X and Y and record the difference
4. Repeat 10,000 times and plot the histogram
5. Record the number of points (m) in the induced distribution that is more extreme than or equal to the observed $d\mu$. The probability m/10,000 is the probability that what was observed was due to chance.

MathSciNet Citations for R1 Math Professors

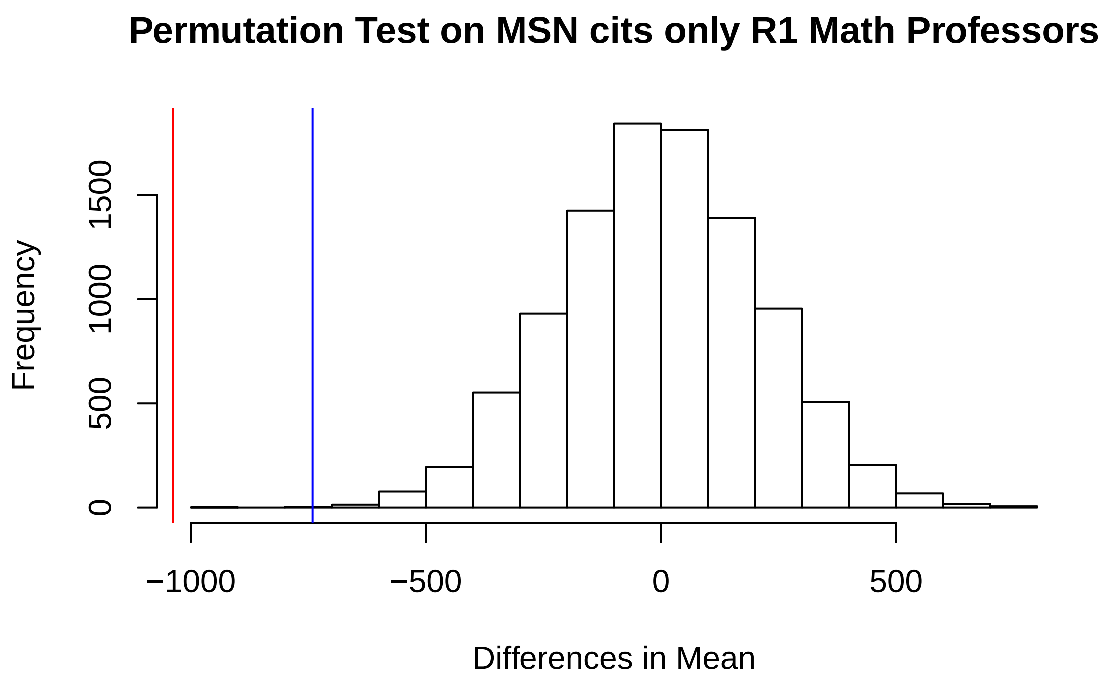


The mean number of citations for signers of letter A is 397 and the median is 261. The mean number of citations for signers of letter B is 1435 and the median is 915. The mean number of citations for signers of letter C is 2177 and the median is 1353.

The three hypotheses we would like to assess are:

1. $H_0 : \mu(A) = \mu(B), H_1 : \mu(A) < \mu(B)$
2. $H_0 : \mu(B) = \mu(C), H_1 : \mu(B) < \mu(C)$
3. $H_0 : \mu(A) = \mu(C), H_1 : \mu(A) < \mu(C)$

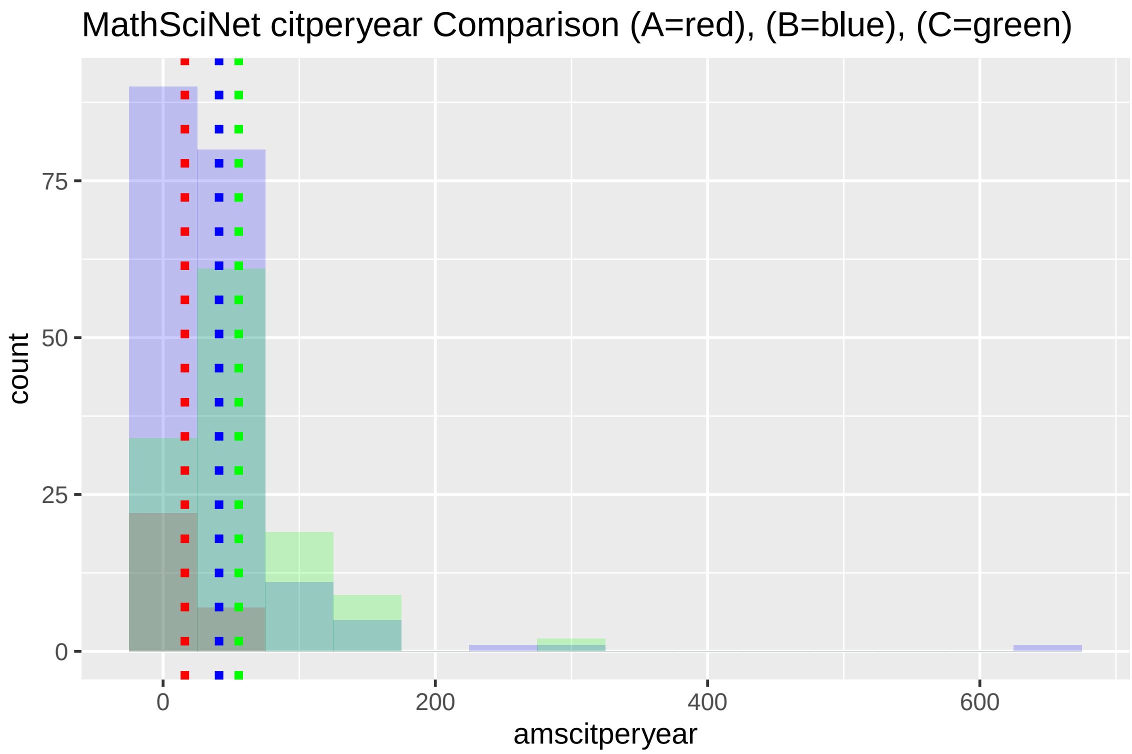
In the following figure, the vertical red bar is the observed difference for hypothesis 1 in the induced distribution, blue is the observed difference for hypothesis 2, and green is the observed difference for hypothesis 3.



The induced p-value for hypothesis 1 is 0. The induced p-value for hypothesis 2 is 0.0002. The induced p-value for hypothesis 3 is 0. Hence we reject all three null hypotheses in favor of the alternative, and $\mu(A) < \mu(B) < \mu(C)$.

MathSciNet Citations per Year for R1 Math Professors

Of course, we considered the fact that citations grow with age, so we calculated citations per year. There may be objections to this - one could hypothesize that citations per year grow with age - but we will soon thoroughly reject this claim.

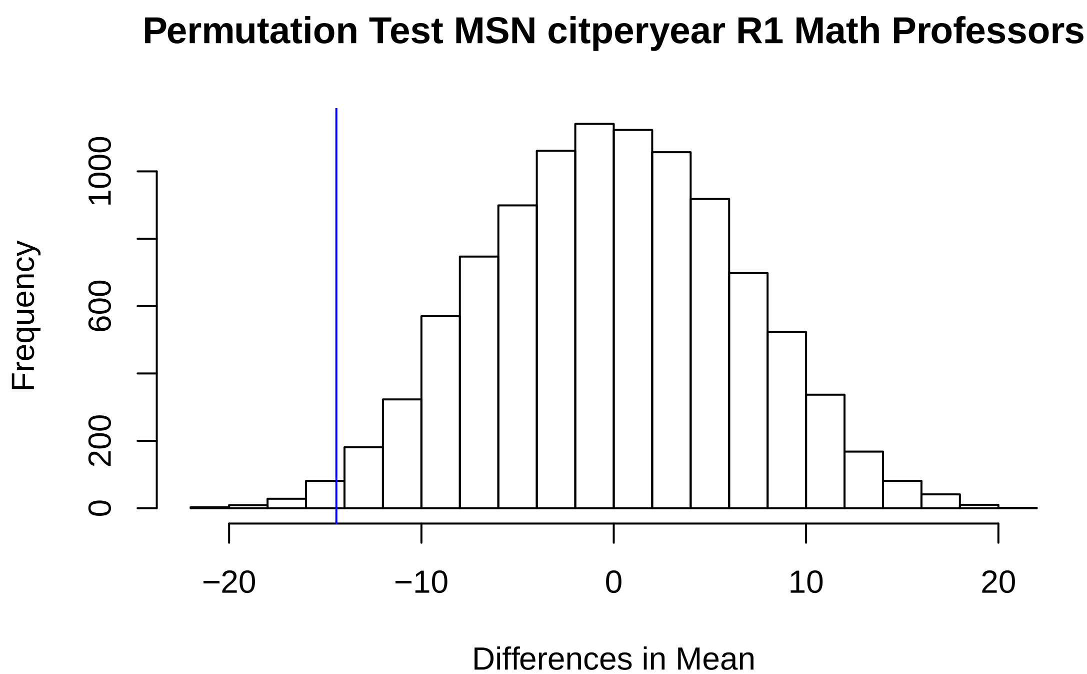


The mean number of citations per year for signers of letter A is 16 and the median is 11. The mean number of citations per year for signers of letter B is 41 and the median is 26. The mean number of citations per year for signers of letter C is 56 and the median is 42.

The three hypotheses we would like to assess are:

1. $H_0 : \mu(A_{citperyear}) = \mu(B_{citperyear}), H_1 : \mu(A_{citperyear}) < \mu(B_{citperyear})$
2. $H_0 : \mu(B_{citperyear}) = \mu(C_{citperyear}), H_1 : \mu(B_{citperyear}) < \mu(C_{citperyear})$
3. $H_0 : \mu(A_{citperyear}) = \mu(C_{citperyear}), H_1 : \mu(A_{citperyear}) < \mu(C_{citperyear})$

As above, the vertical red bar is the observed difference for hypothesis 1 in the induced distribution, blue is the observed difference for hypothesis 2, and green is the observed difference for hypothesis 3.

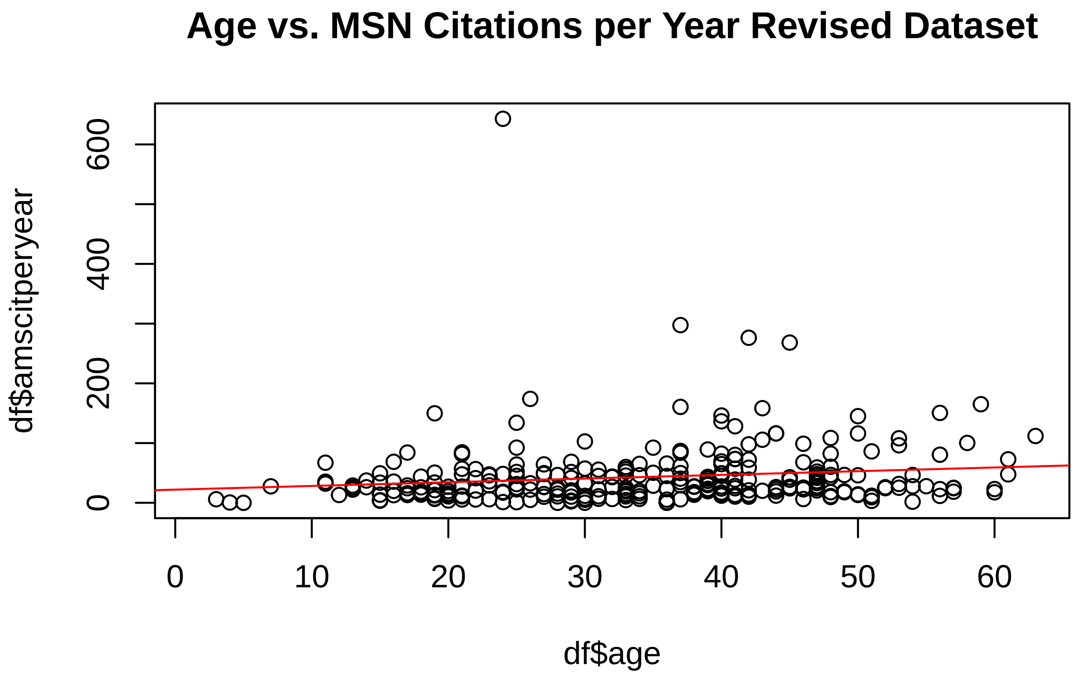


The induced p-value for hypothesis 1 is 0. The induced p-value for hypothesis 2 is 0.0099. The induced p-value for hypothesis 3 is 0. Hence we reject all three null hypotheses, even assessing the induced p-value at a 2% significance level, in favor of the alternative, and we conclude that $\mu(A_{citperyear}) < \mu(B_{citperyear}) < \mu(C_{citperyear})$.

One may object of our usage of the mean here, opposed to the median. So we reperform the permutation test with the median. All three induced p-values are 0 so we can reject all three null hypotheses when comparing median citations per year.

There is no evidence that Citations per Year grows with age

Let us check whether there is a relationship between age and citations per year in our limited dataset.




```r
confint(linearmodel1)
```

```
##                 2.5 %    97.5 %
## (Intercept) 3.7459618 40.351608
## df$age      0.1149979  1.120528
```

The slope of the regression line is slightly positive (0.6178, 95% Confidence Interval = (0.115, 1.12)), but the R^2 values (Adjusted = 0.01662), are tragically low. So there is really no correlation. However one could object that we do not have enough data (n = 258), to assess that there is no correlation between citations per year and age. We know this, but thought it would more appropriate to analyze this in a separate paper. However, as noted above, our honor and ability as scientists were attacked so…

Presenting citations data on every R1 Math Professor with MathSciNet citations

(plus the Institute of Advanced Studies and UC Merced)

We manually collected the citations and year of first publication of every R1 full math professor by consulting wikipedia, going to the relevant faculty pages and then collecting MathSciNet citations. We then anonymized it. The 2787 professors we collected data on is in line with data collected by the AMS, after taking into account that about half of universities in the US are classified R2. Great lengths were taken to assess the accuracy of this data, including correlating publications, PhD years, etc. Of course errors in data collection, especially manual typing errors, happen, but by no means are these errors systematic.

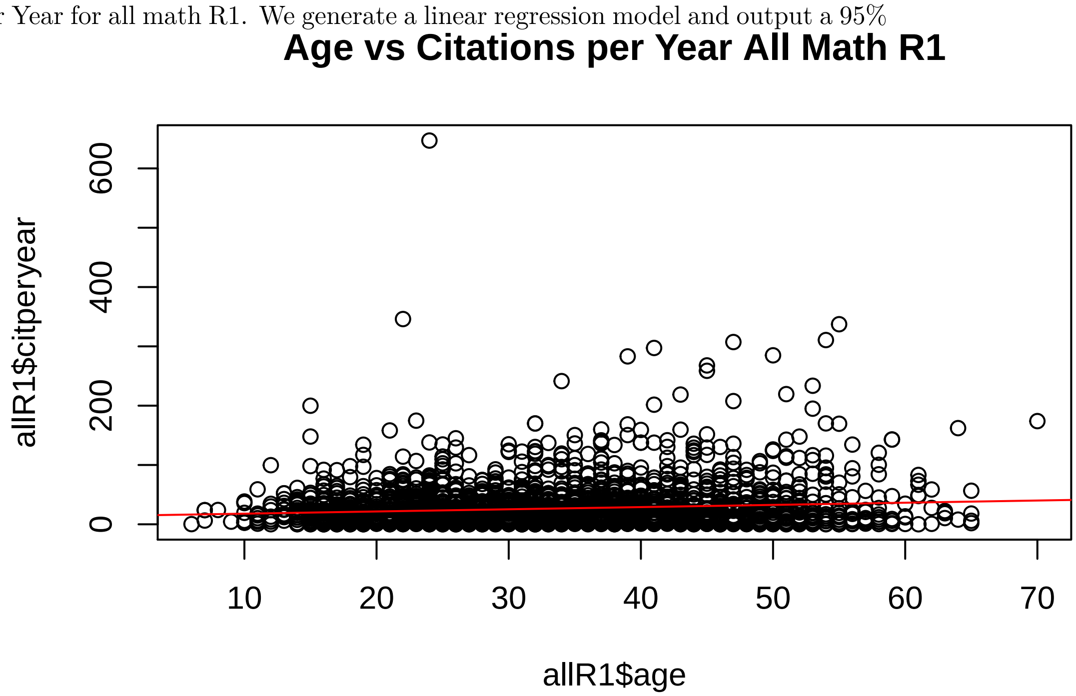
Exercise 1: Pick your favorite R1 institution, go to MathSciNet, and check how similar our data is to what you determined.

Exercise 2: Determine every university without a female professor. We will note that the University of Colorado - Denver has a very strong female professor, but she does not have MathSciNet citations. There should be at least one surprise (and a few non surprises).

Many aspects of this dataset can, should, and will be analyzed. For now, the following will suffice.

Citations per Year vs Age

We plot the Age vs Citations per Year for all math R1. We generate a linear regression model and output a 95% confidence interval for the slope.



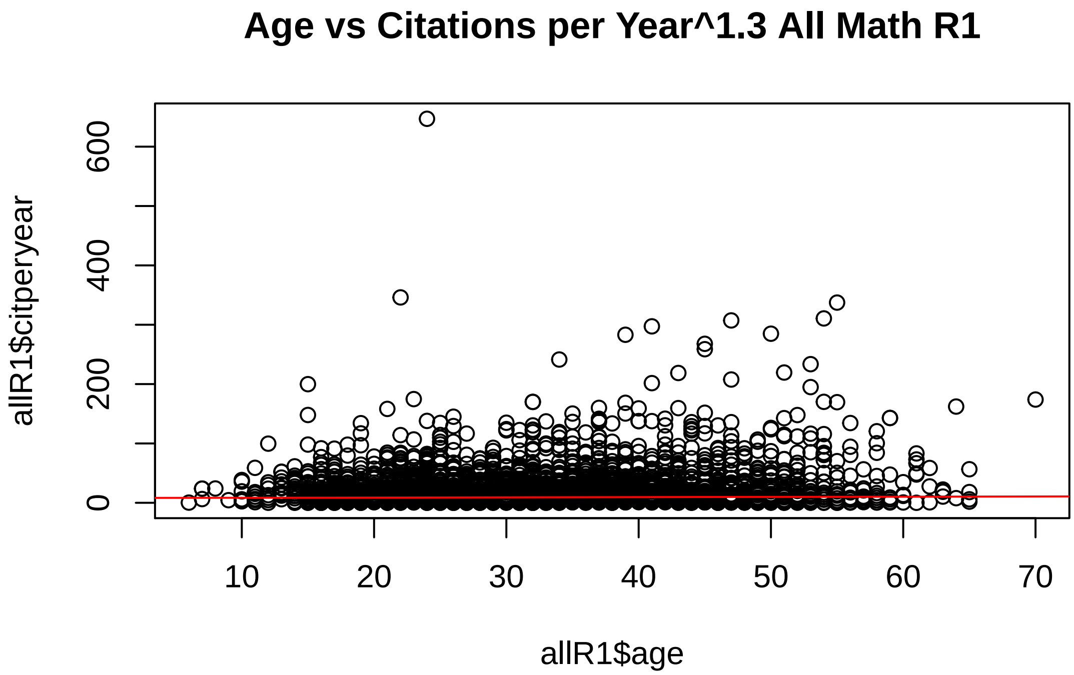

```r
confint(linearmodel2)
```

```
##                 2.5 %   97.5 %
## (Intercept) 10.1588495 18.33147
## allR1$age    0.2515584  0.48675
```

So while visually it appears that there is no correlation between citations and citations per year, one may object and say, the slope is positive! Which leads to the following question.

Question: To what power must we raise age to get zero within the confidence interval of slope.

We object to this question, because the implication of the question is, by how much should we discount the accomplishments of those who are older. Nevertheless, we proceed.



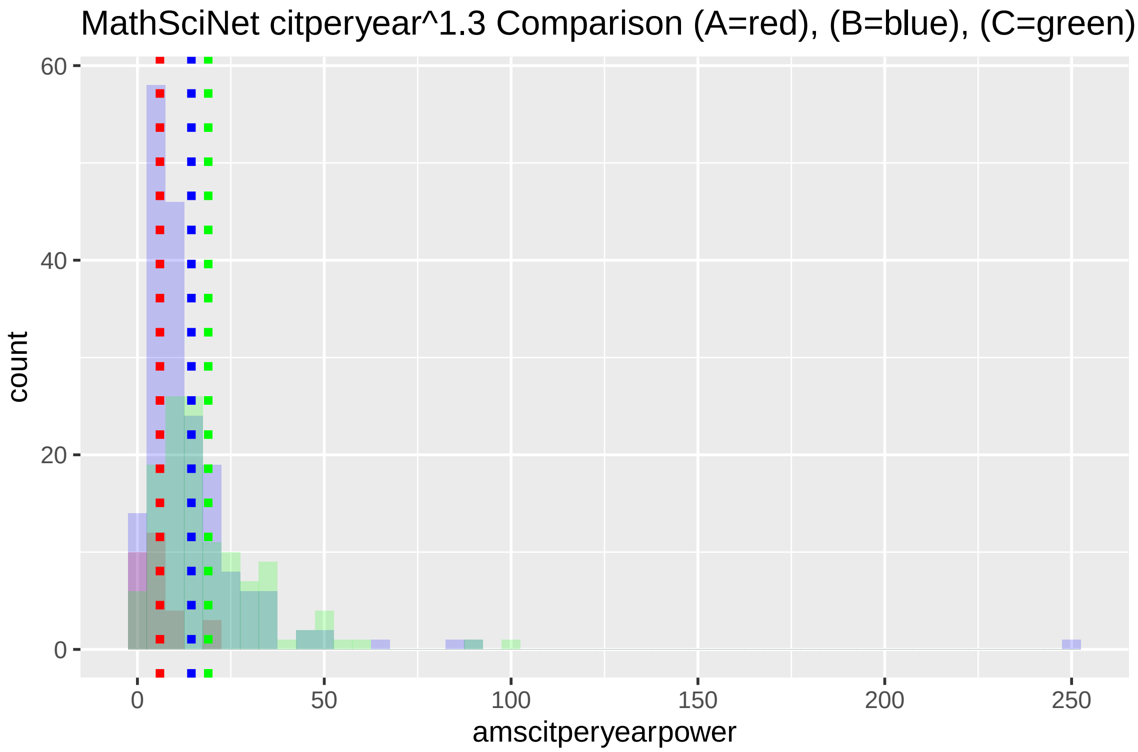

```
confint(linearmodel3)
```

```
##                     2.5 %     97.5 %
## (Intercept)  6.715962422 9.56734674
## allR1$age   -0.004656401 0.07740065
```

It seems raising age to the 1.3 will do the trick.

We will reperform the permutation test comparing citations per yearˆ1.3.

Citations per Year adjusting for fitted handicap on age

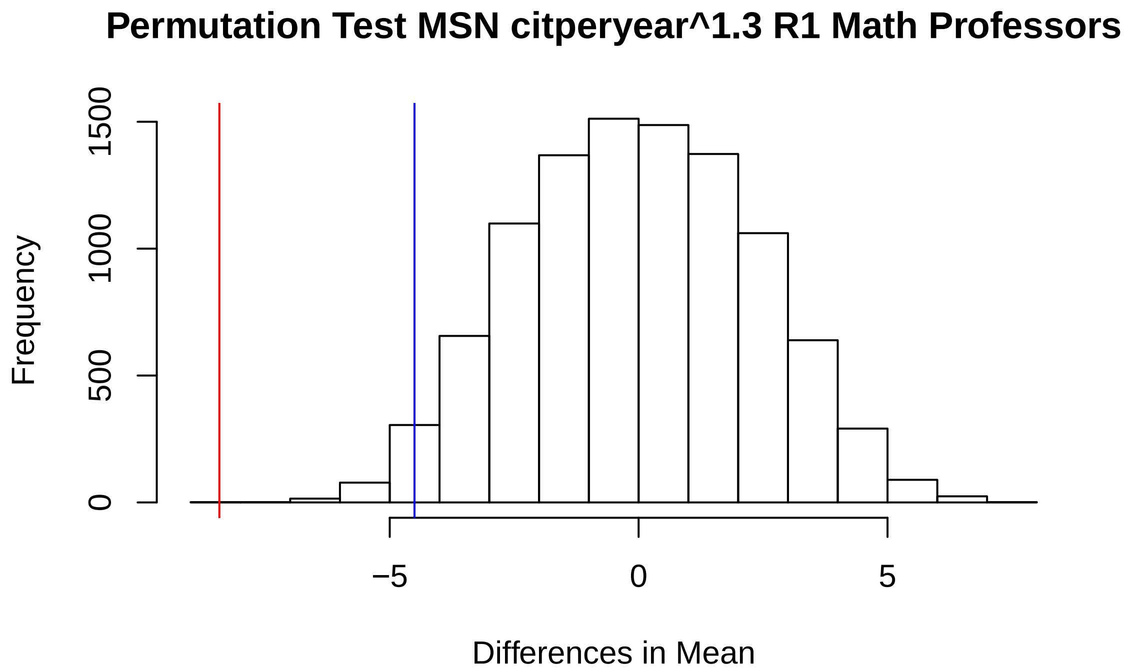


The mean number of citations per year^1.3 for signers of letter A is 6 and the median is 4. The mean number of citations per year^1.3 for signers of letter B is 14 and the median is 9. The mean number of citations per year^1.3 for signers of letter C is 19 and the median is 14.

The three hypotheses we would like to assess are:

1. $H_0 : \mu(A_{citperyear^{1.3}}) = \mu(B_{citperyear^{1.3}}), H_1 : \mu(A_{citperyear^{1.3}}) < \mu(B_{citperyear^{1.3}})$
2. $H_0 : \mu(B_{citperyear^{1.3}}) = \mu(C_{citperyear^{1.3}}), H_1 : \mu(B_{citperyear^{1.3}}) < \mu(C_{citperyear^{1.3}})$
3. $H_0 : \mu(A_{citperyear^{1.3}}) = \mu(C_{citperyear^{1.3}}), H_1 : \mu(A_{citperyear^{1.3}}) < \mu(C_{citperyear^{1.3}})$

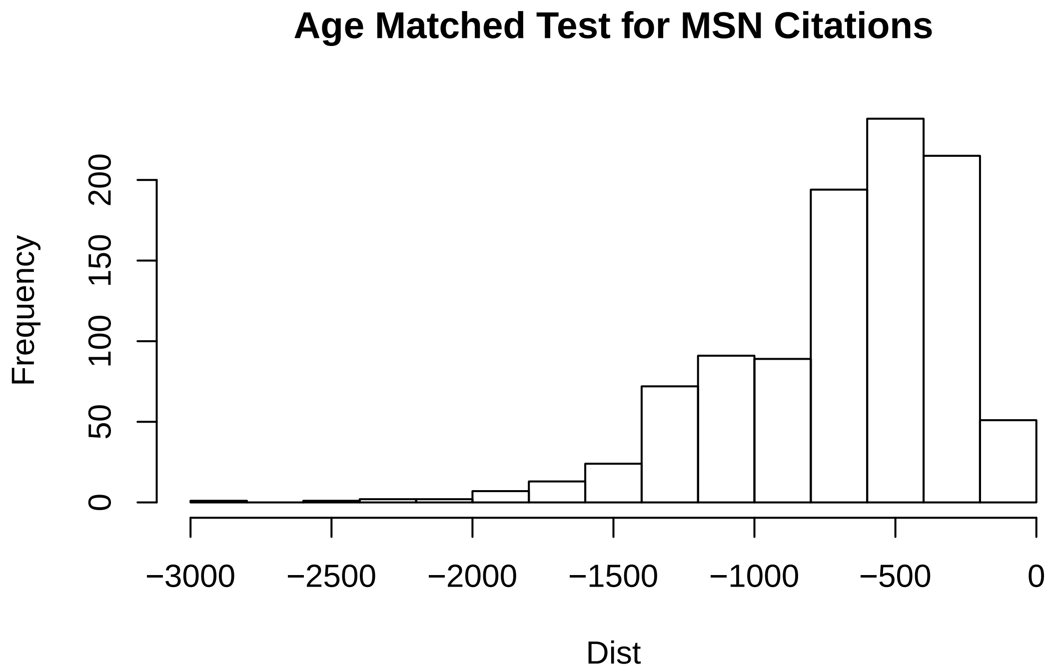
As above, the vertical red bar is the observed difference for hypothesis 1 in the induced distribution, blue is the observed difference for hypothesis 2, and green is the observed difference for hypothesis 3.



The induced p-value for hypothesis 1 is 0. The induced p-value for hypothesis 2 is 0.0207. The induced p-value for hypothesis 3 is 0. Hence we fail to reject hypothesis 2 at a 2% significance level and reject hypotheses 1 and 3 in favor of the alternative. We conclude that after adjusting for age $\mu(A) < \mu(B) \leq \mu(C)$.

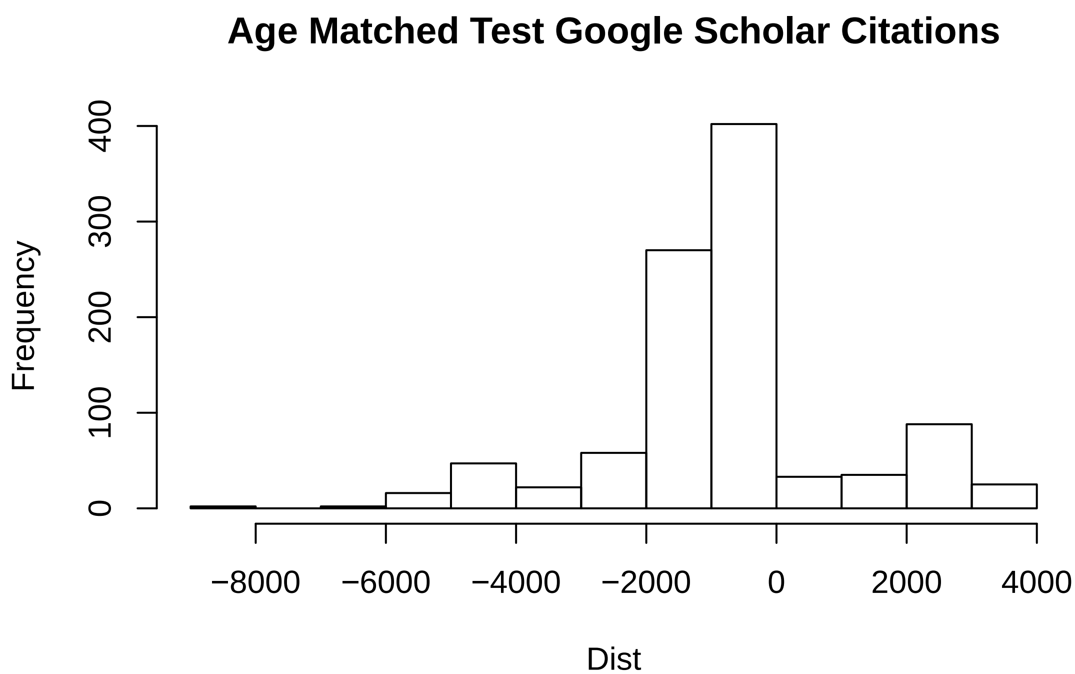
One more check that age is irrelevant when comparing citations

This method was suggested by a friend as a final check to eliminate any question that age was the greatest confounder. We want to show that $\mu(A) < \mu(B \cup C)$. We will randomly sample a population of 20 from A, called X. For each member $x \in X$, we will find every person from B and C that is within a two year age interval from x. We will randomly sample one, and induce a new population Y. Then we will compare the means by storing X-Y. We repeat this 1,000 times and plot a histogram of the induced values. If 0 is within this new distribution, then maybe there is a chance, a totally slim one after above, that in fact age is a confounder. If the distribution is primarily negative, then X < Y. Otherwise X > Y. We perform this analysis with both AMS citations and Google Scholar citations.

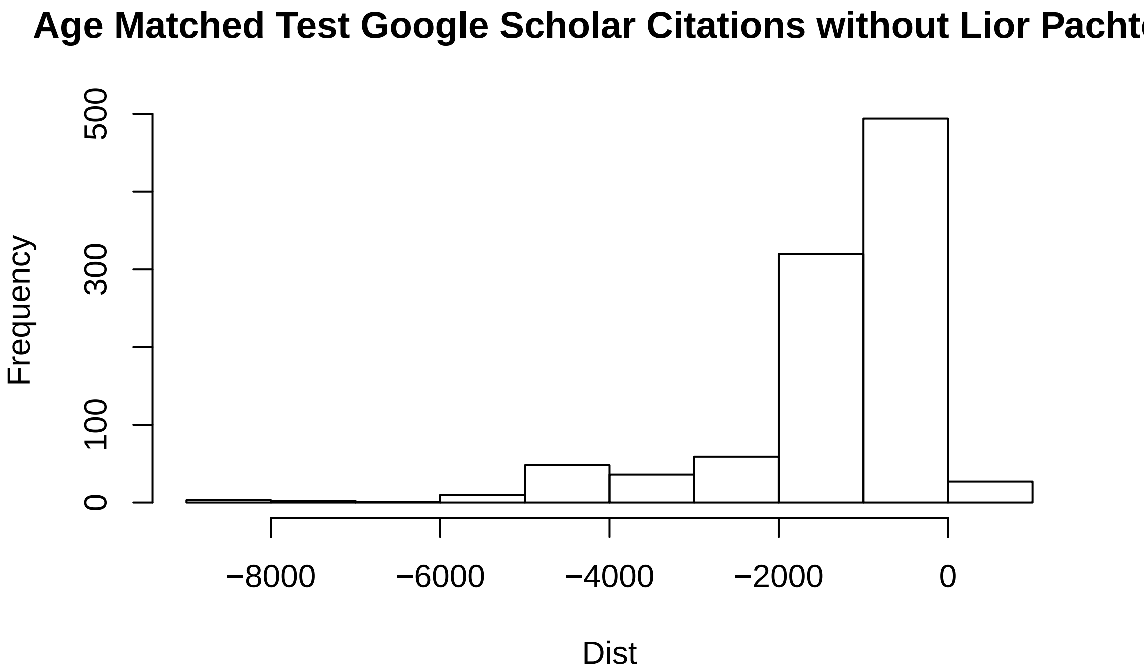


When comparing mathscinet citations with this age matched randomization test, we see that none of the induced distribution is greater than or equal to zero. So when comparing similarly aged apples to apples, $A < B \cup C$

We perform the same analysis with Google Scholar citations.



When comparing Google Scholar citations with this age matched randomization test, we see that 18.1% of the induced distribution is greater than or equal to zero. So when comparing similarly aged apples to apples, it is inconclusive if $A < B \cup C$. Of course, we wonder if this is actually Lior Pachter.

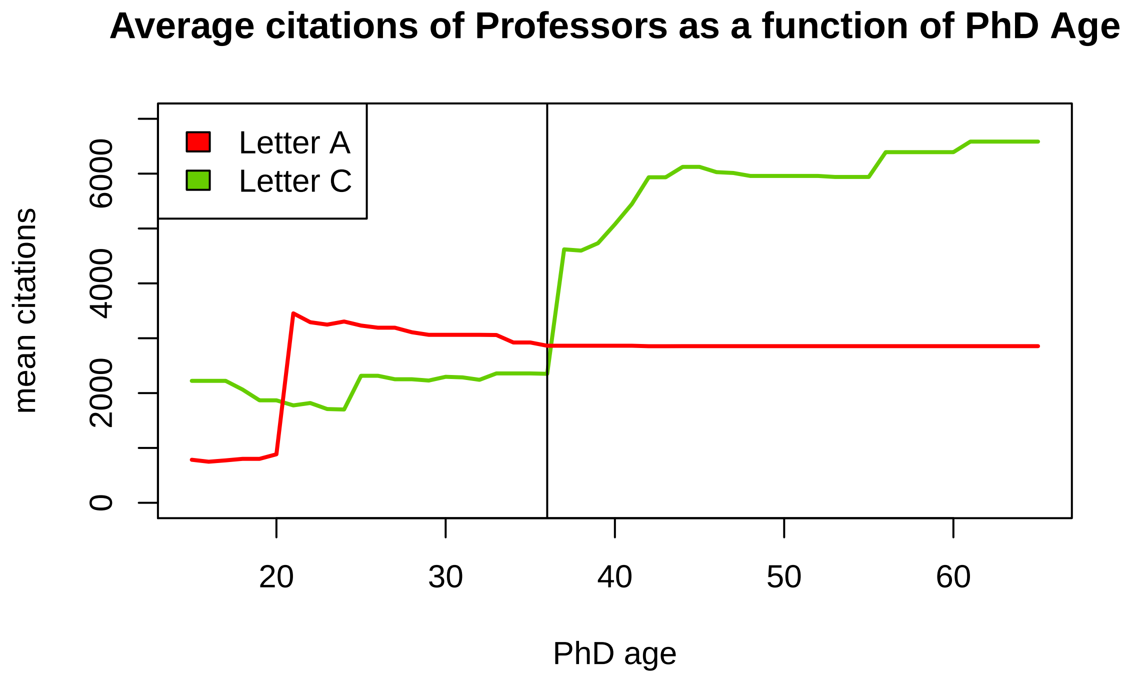


When comparing Google Scholar citations, removing Pachter, with this age matched randomization test, we see that 2.7% of the induced distribution is greater than or equal to zero. So when comparing similarly aged apples to apples, it indeed seems that $A < B \cup C$.

Pachter's Magic Trick: Hypertuning

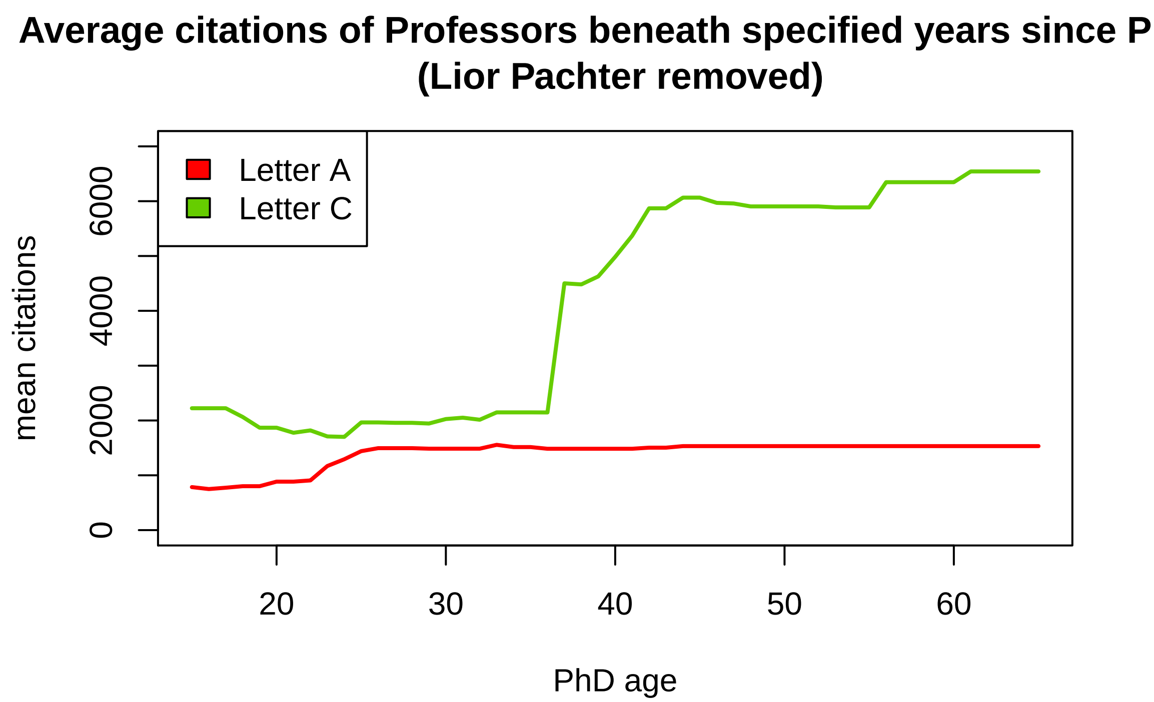
A note about Pachter's final, "damning," (it is not), figure. He chose a cutoff of age 36, and compared the average Google Scholar citations of letter signers. He finds that if one does this cutoff, the mean citations of A is greater than B. We found this choice of 36 to be curious and somewhat arbitrary. It smelled like parameter tuning, but we wanted to investigate.

We plot the average citations per year and note with a vertical line, the 36 (PhD) age cutoff.



The maximum age since PhD of a letter signer of A is 49. If he were to cutoff his comparison at that point, clearly $C > A$. If he were to cutoff his comparison at 38, $C > A$. Any further left of 36, he would be accused of being biased.

Notice the spike at age 21. This is caused by Lior Pachter. What would happen if we removed Pachter?



So it is clear that Pachter's analysis was some sort of magic trick, potentially a thought experiment, and a fraudulent one. It is highly unlikely that a tenured and respected expert in computation and statistics did not know the above result, expecially when a student he suggests take an introductory statistics course immediately spotted it. We claim and assert that he purposefully chose his 36 cutoff to try to undermine our results.

Tier Rankings

In our excel sheet, (which we understand is the bane of reproducibility), and through the magic of pivot tables, we rank R1 departments by calculating average department citations per year (since first publication).

The top 11 departments using this ranking are:

1. Princeton
2. Institute of Advanced Studies
3. Harvard
4. Stanford
5. University of Chicago
6. University of California - Los Angeles
7. Massachussetts Institute of Technology
8. Columbia University
9. New York University
10. University of Miami
11. University of California - Berkeley

We calculate the average citations/average year since PhD of letters A, B, and C, and compare them to our ranked list.

The average Math Sci Net Citations per year (PhD Age) is:

1. For letter A - 15.98726
2. For letter B - 41.86467
3. For letter C - 55.3615

Temple has an average citations per year of 12.33, so we retract our claim that letter A is comparable to Temple. It is closer to the University of Massachusetts - Amherst which has an average citations per year of 16.17. By US News, University of Massachusetts - Amherst's Math Department has a rank of 55. Rutgers has an average citations per year of 35.01, so we retract our claim that letter B is comparable to Rutgers. It is closer to the University of Minesotta which has an average citations per year of 42.07 and a US News Ranking of 19. For Letter C, we claimed that it was another tier higher - indeed it is closer to the University of Chicago, which has an average citations per year of 56.27, ranked 6 by US News.

An astute observer would notice we are not exactly comparing apples to apples. Presumably ones first publication could be before one finishes their PhD. So even with the boost, the order amongst letter signers stands.

Discussion and Conclusion

We have debunked the claim that age is the greatest contributor for the difference in citations and citations per year between signers of Letter A, B, and C. Indeed, the least meritorious of mathematicians as a whole signed letter A, whereas the more meritorious signed letters B and C, with merit judged by citations. If one was not willing to believe citations impose even a small order on merit, one could replace citations with Fields medals, AMS Fellowships, or many other metrics. One could make extrapolations by coupling this analysis with work by Topaz, but we refrain from doing so.

In this analysis, we have addressed most of the criticisms in Pachter's review, acknowledging our errors when pointed, while rejecting his false claim that age was the greates confounder. The only one we have not addressed is his point that, “several p-values are computed and reported without any multiple testing correction.” After consultation with a respected statistician, we do not see what the issue is. We reported every p-value and he is welcome to change the set.seed in our code, which he applauds us as easily reproducible.

Regarding the Russians, one cannot help but make parallels with what is happening now in the USA with what happened in the USSR and China. Mandatory diversity statements are being used as a proxy for affirmative action. Affirmative Action is wrong and is totally pernicious. Advocates of affirmative action seem to believe there is a tradeoff between ability and background, and that someone’s race, class, gender, sexual orientation, etc. is an excuse for bad research. This is totally false. There are many strong female mathematicians, there are many strong black mathematicians, there are many strong gay mathematicians. There is nothing about someone’s background which can justify bad science.

Pachter ends his review by saying “under-represented minorities and women routinely face discrimination and worse. This is completely unacceptable.” We agree. But we do not believe that the solution is to allow for more diverse faculty regardless of their ability as researchers. What we do believe is that increasing opportunities for under represented minorities at an early stage and fairer hiring practices is the solution, not reverse discrimination advocated by some.

We conclude by reiterating our thanks to Pachter. We truly appreciated your review.

Data and Code

All code and data used for this report is available at. https://github.com/joshp112358/Response-to-Pachter

References

Lior Pachter’s Blog Post - Diversity Matters - January 17, 2020 https://liorpachter.wordpress.com/2020/01/17/diversity-matters/

Chad Topaz’s Paper - Version 10 - https://osf.io/preprints/socarxiv/fa4zb/

Our original Paper - Version 1 - https://arxiv.org/pdf/2001.00670.pdf

In Preparation - A Citations Analysis of R1 Math Departments by Joshua Paik and Igor Rivin

Miscellany

There seems to be some squabbles in the comments of Pachter’s blog whether the paper is Paik-Rivin or Rivin-Paik. In mathematics, we follow the Hardy-Littlewood rule, namely all authors are first authors and we list authors alphabetically.