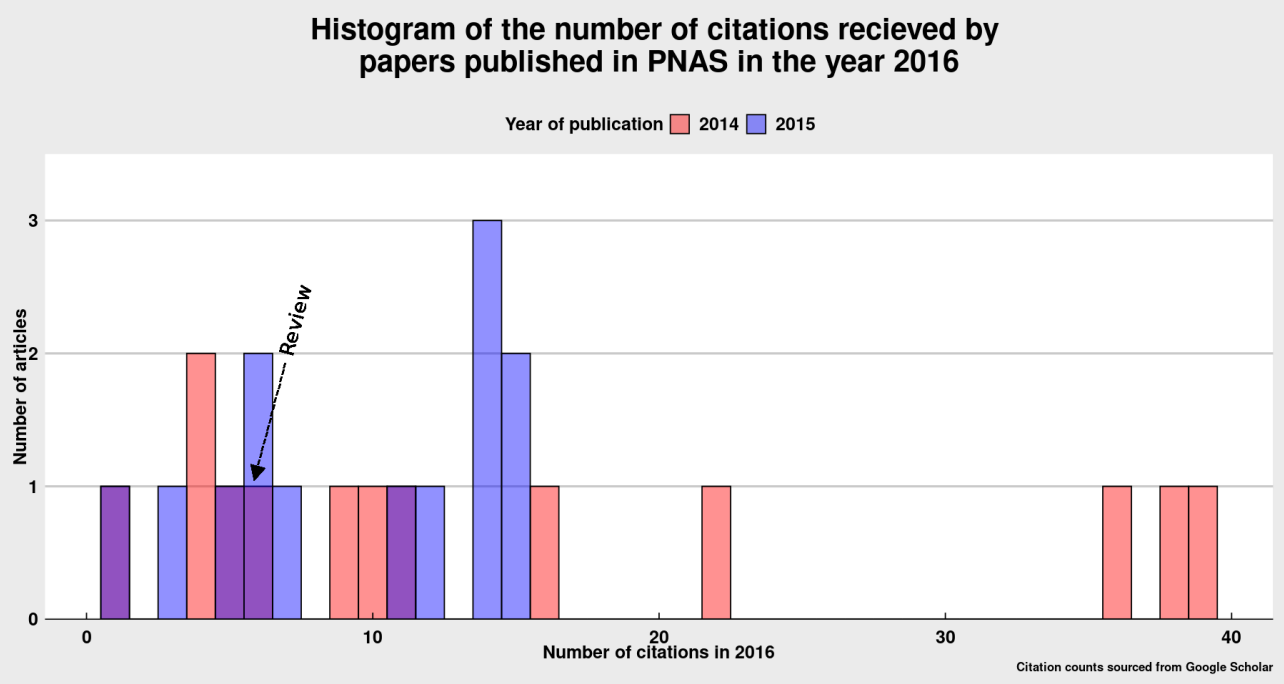
BIO310 Assignment

Ananda Shikhara Bhat
ID: 20181024
20th January 2021

1.
a) I obtained a list of all articles published in PNAS in the years 2014 and 2015 from SCOPUS. I used the dplyr function 'slice_sample' to select 13 random articles from 2014 and 13 random articles from 2015.

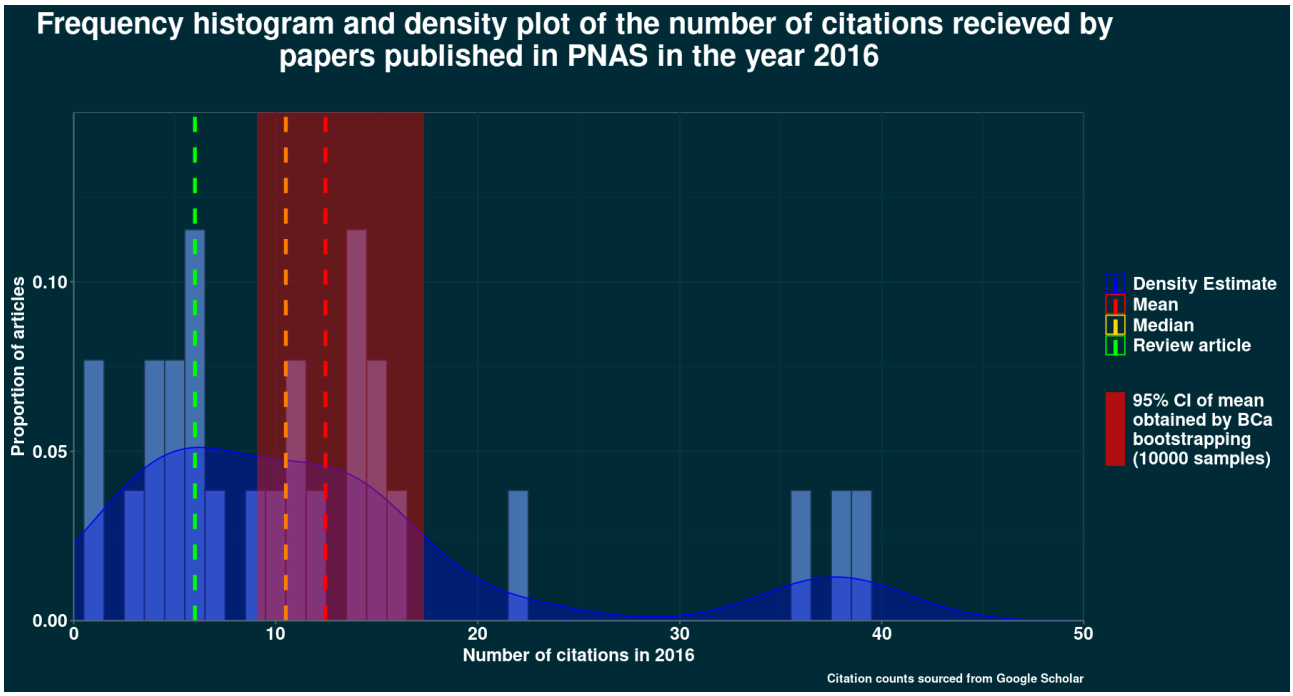
b) Table is provided along with code below

2.



3. The mean number of citations is 12.46, and the median is 10.5. The median is less than the mean. This is because the data has a positive skew. The median would be a better central tendency here, because it is less affected by outliers. (See plots present along with code.)
We could also calculate a confidence interval around the mean using bootstrap techniques.

4. My sample only contained a single review, and this review recieved citations below the lower bound of the 95% confidence interval of the mean, suggesting that reviews may get lower citations than research articles, atleast in the short term. However, this must be taken with a grain of salt, since the sample only contained a single review. A more appropriate way to test this hypothesis would be using a permutation test or a Mann-Whitney test, if the sample was a more balanced mix of reviews and research articles.

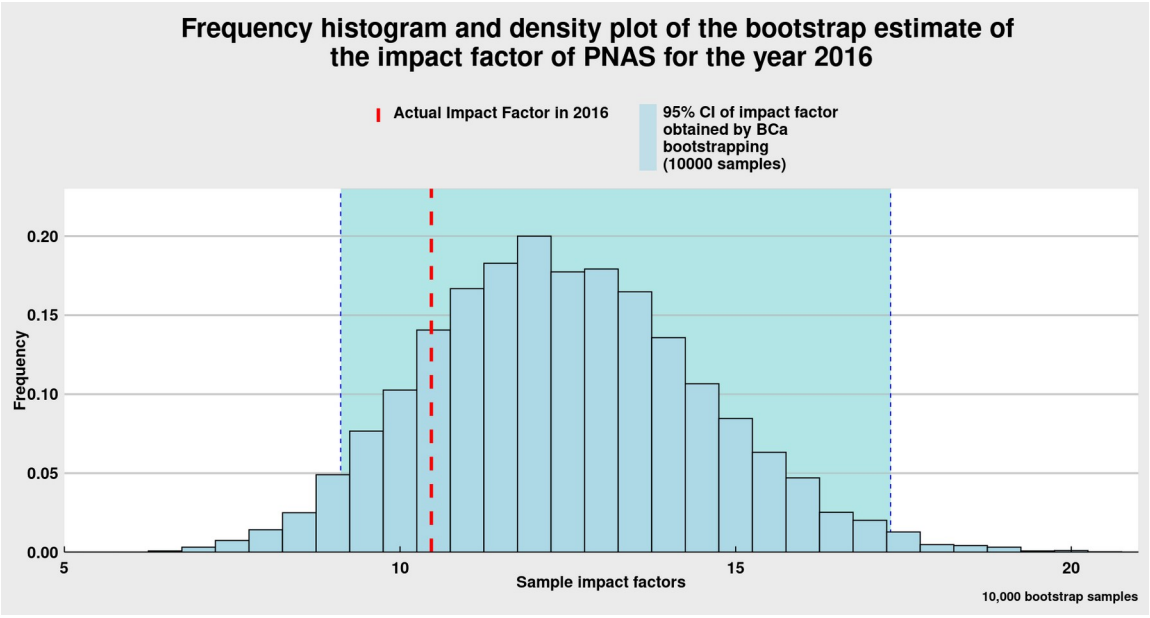


5.

i) Estimate 1: Confidence interval as mean +- 1.96*SE, where SE is the standard error of the mean.
ii) Estimate 2: Bootstrap confidence interval.

I estimated both, results are presented along with the code below.

6. 95% Confidence interval for the impact factor of PNAS in 2016, for our sample, is [9.115385 , 17.30769]



Note: I'll also upload all the code given below as a jupyter notebook into a public github repository once the deadline has passed (I'm guessing you wouldn't want me to upload it before then, since other people may just download it and copy the methods)