

Author – Jean G. Collazo

Date – December 19, 2021

Reproducing Findings of Hurricane Maria Mortality Study

Kishore et al. set out to investigate using surveying and statistical methodology to estimate the mortality caused by Hurricane Maria in 2017. The official death count reported was 64. However, Kishore et al. reported an estimate of 4645 excess deaths during the period of September 20 through December 31, 2017, which is more than 70 times the official estimate.

In this project I set out to reproduce the findings of Kishore et al. study. Reproducing the finding for a computational experiment implies that a different independent group can obtain the same result using the author's won artifacts (as defined by the ACM Artifact Review Badging webpage). In this study the researchers used R for their statistical finding and the majority of their figures in order to produce their results and draw out conclusions. For the purpose of reproducing their findings I will use the artifacts provided by dr. Irizarry's GitHub repository and run them using RStudio from my own personal device.

Population Estimates in the study were conducted with methods outside of R which I'm unable to reproduce. But the formulas used are provided in the study.

The **excess deaths estimate** was accurate and reflected the same number found by the researchers. Below are the numbers I gathered. In the left the estimate reflects the unadjusted survey estimate for excess deaths with the lower limit and upper limit. In the right we have the adjusted survey estimate that accounted for the fact that single-person household only reflect a mortality rate of 0. The reading was

hh_size	hh_size
4645.298	5739.649
hh_size	hh_size
792.9268	1505.522
hh_size	hh_size
8497.669	9889.095

accurate to the authors numbers as well as the lower and upper limits.

The calculations for the adjustment rates appear to correspond to those in the data tables provided by Dr. Irizarry’s repository.

Figures 1 through 4 showed an accurate reading compared to the authors findings. My images are on the left while the ones provided by the authors are on the right

Figure 1

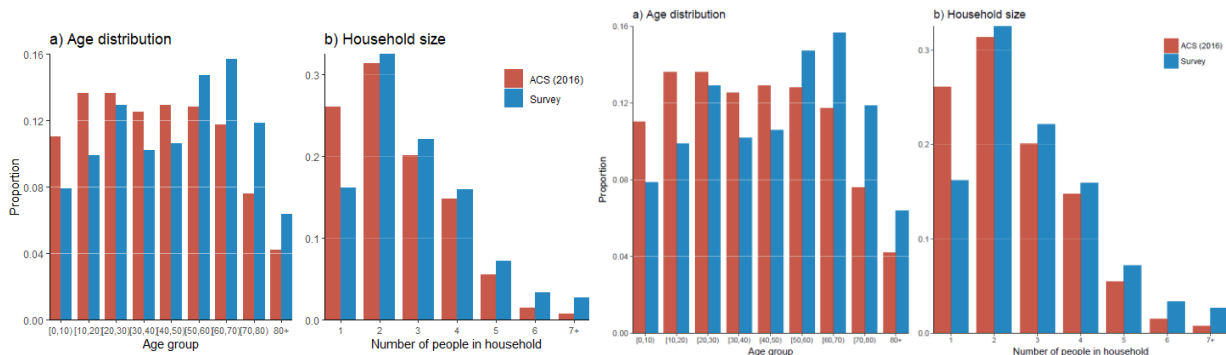


Figure 2

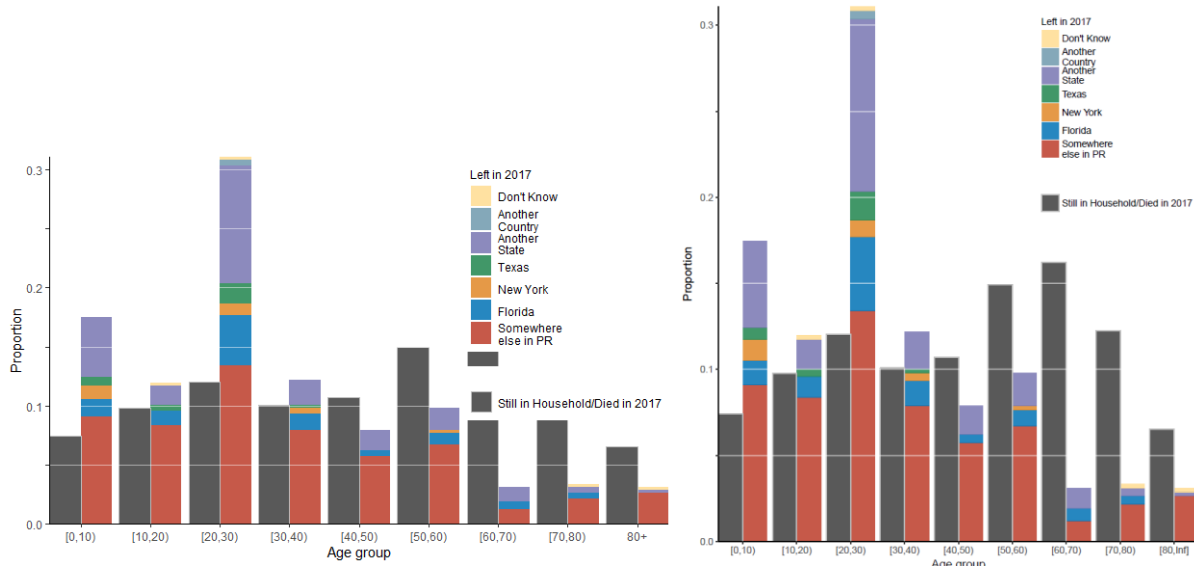


Figure 3

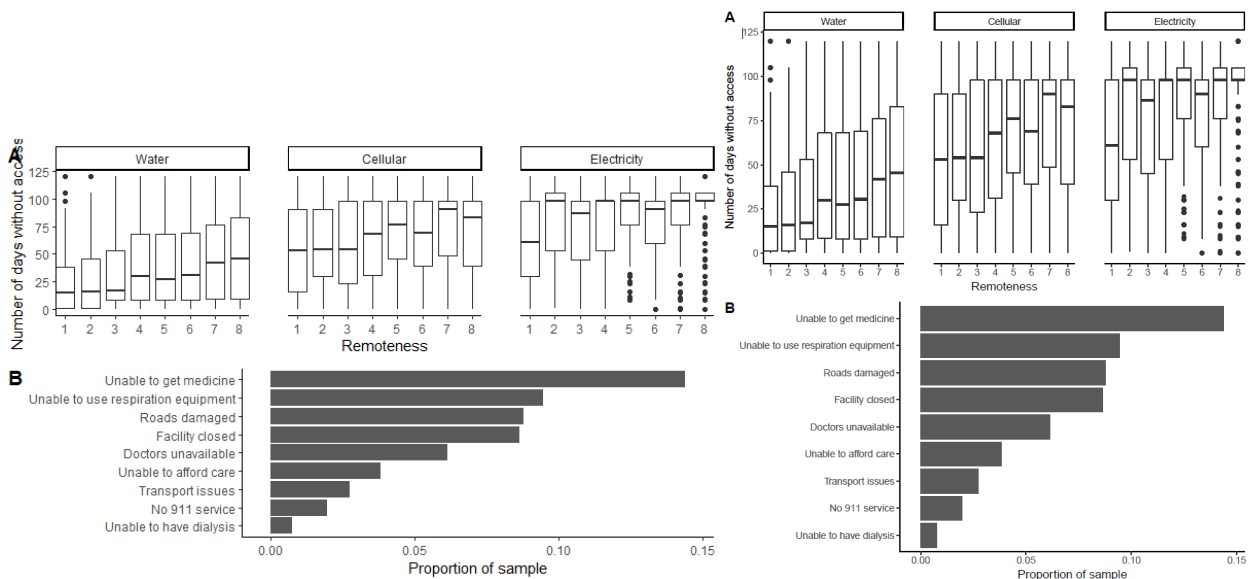
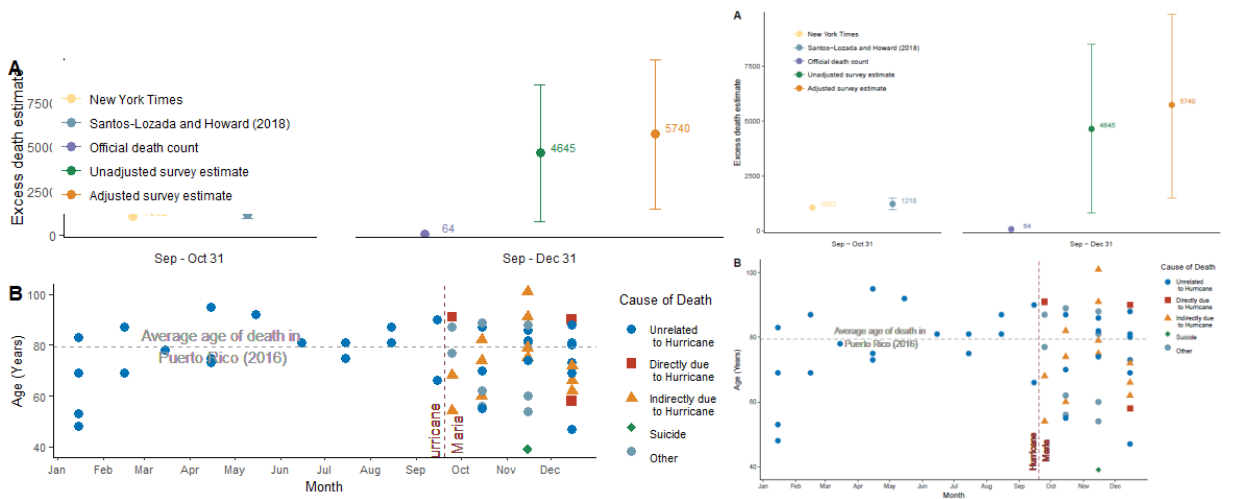


Figure 4.



Supplementary Figures 2 – 6

Figure S2

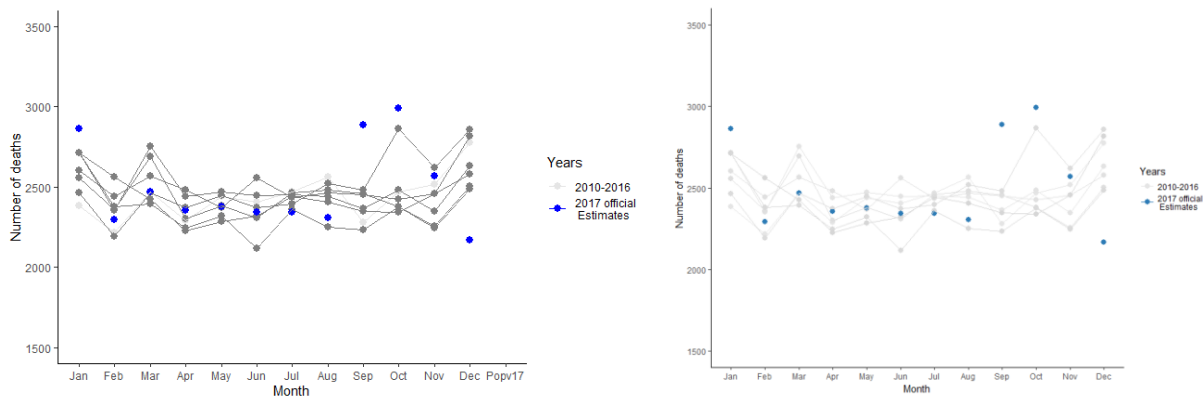


Figure S3A

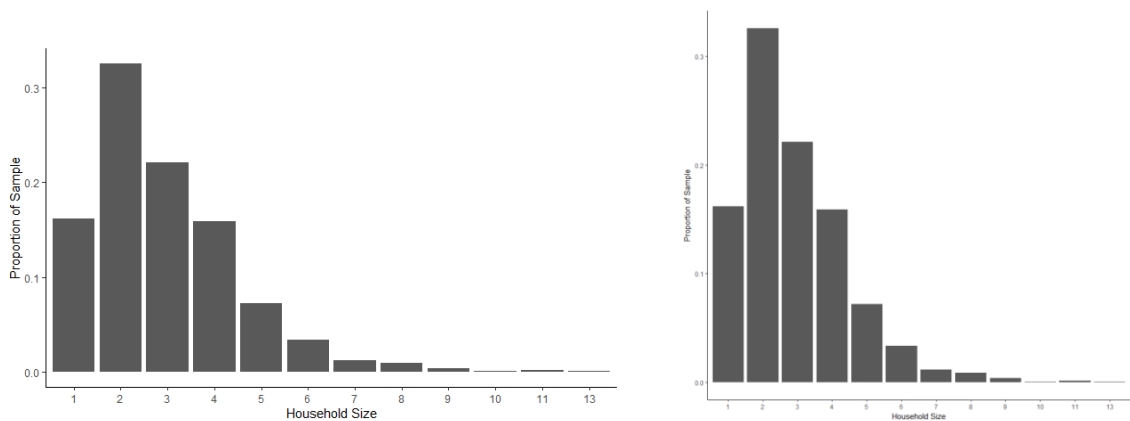


Figure S3B

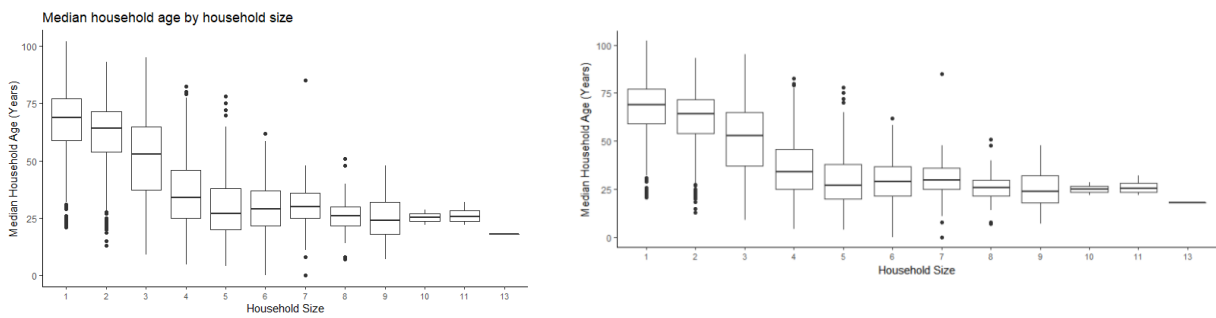


Figure S4

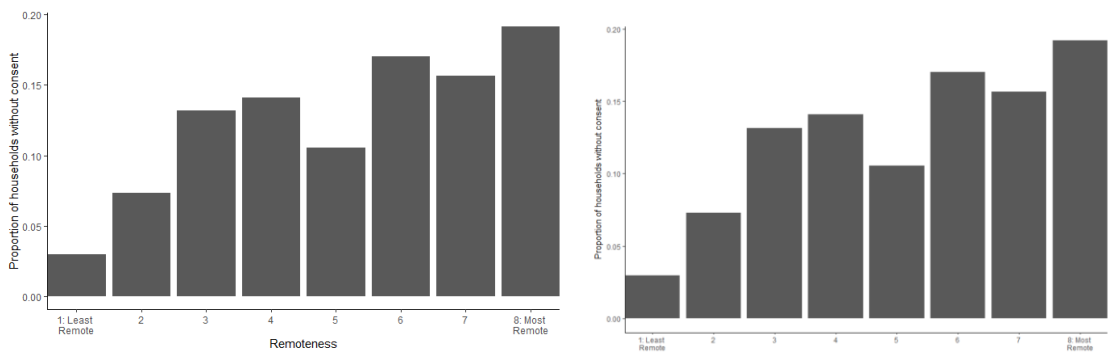


Figure S5

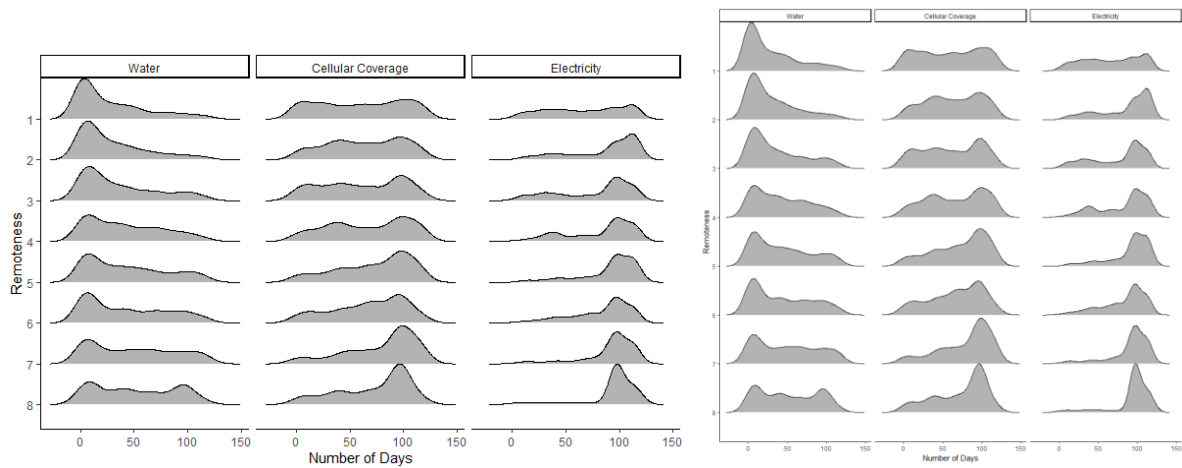
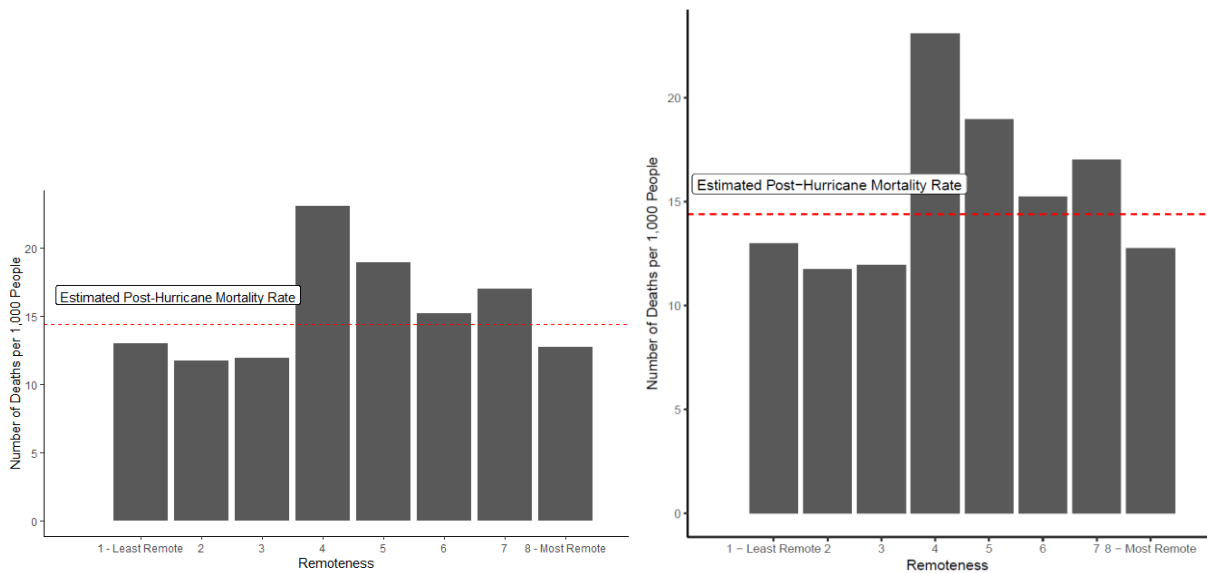


Figure S6



Supplementary Tables 1-3

Table S1 following the author’s artifacts

	Variable	WeightedEst	SE	lower	upper
count	Households	1052957.88800141	284035.552742904	4.96e+05	1.61e+06
hh_size	Population	3030307.15490978	797768.914656318	1.47e+06	4.59e+06
age1	Median Age	49	NA	NA	NA
gender1	Proportion Female	48.8578676504695	0.329208201059147	4.82e+01	4.95e+01
hh_size.1	Mean Household Size	2.87789966668233	0.0596367848640547	2.76e+00	2.99e+00

Finalized author's Table S1

Table S1: Survey Demographics compared to the American Community Survey 2016				
	<i>Survey</i>	<i>Population Estimate</i>	<i>95% Confidence intervals</i>	<i>ACS 2016</i>
Households	3,299	1,052,957	[496,248- 1,609,668]	1,251,554
Population	9,522	3,030,307	[1,466,680 - 4,593,934]	3,529,385
Median age	48.00	49.00	N/A	39.40
Proportion female	48.60%	48.86%	[48.2% - 49.5%]	52.20%
Mean household size	2.88	2.88	[2.76 - 2.99]	2.82

The numbers appear to be accurate on both sides, but the author's table includes information from their Survey and the ACS 2016.

Table S2

Table S2 generated from code only shows the mean and standard deviation for missing utilities divided by the remoteness factors. Meanwhile the finalized version of the table presented by authors has the utility in percentages divided by months (September – December) as well as divided by the remoteness factor.

Table S3

The numbers generated from the authors code are accurate to those in the finalized author's version but in only includes the proportion of medical care access and utility access. Meanwhile the finalized author's table has the contains the lower bound of estimated mortality rate of neighbors.

Conclusion

While several of the figures and numbers are reproducible as demonstrated here some numbers were calculated apart from the code provided in R like the population estimate. In order to reproduce that estimate one would need to know the total number households of barrios where the surveys were conducted, the total population in each remoteness category, and the number of barrios in each remoteness category. This information is not readily available in author's artifacts. Another issue was found for Table S2 where the proportions for utility availability included information separated in months while the code only provides information for means and standard deviations as derived from the survey information. And finally, Table S3 includes information for the lower bound if estimated rates if neighbors which there is no R code provided for that calculation.

References

Kishore, Nishant, et al. "Mortality in Puerto Rico after Hurricane Maria." *New England Journal of Medicine*, vol. 379, no. 2, 2018, pp. 162–170., <https://doi.org/10.1056/nejmsa1803972>.

(Kishore et al.)