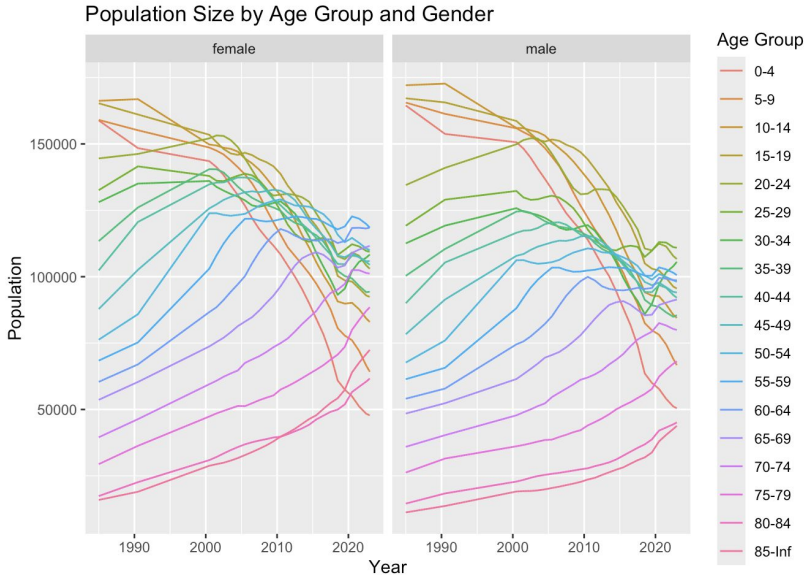
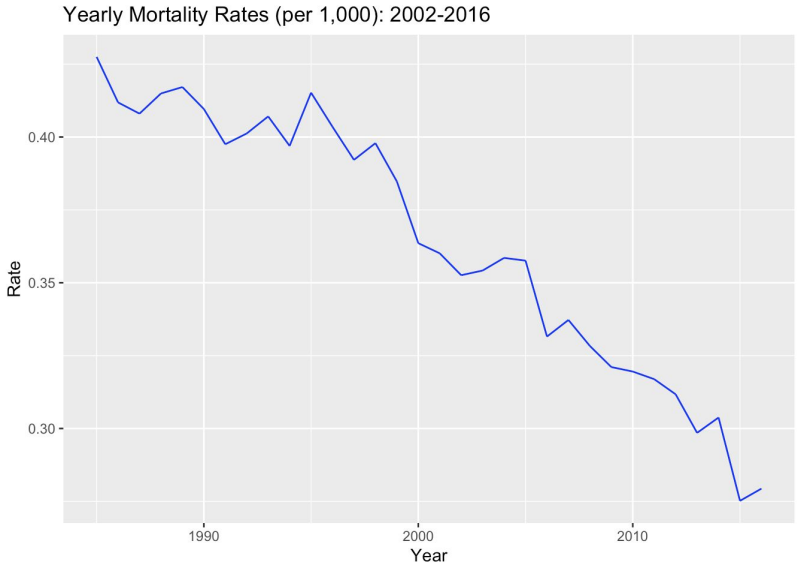
Analyzing Excess Mortality in Puerto Rico after Hurricane Maria

Megan Kern & Zoe Lu

Population of Puerto Rico: Patterns in Data





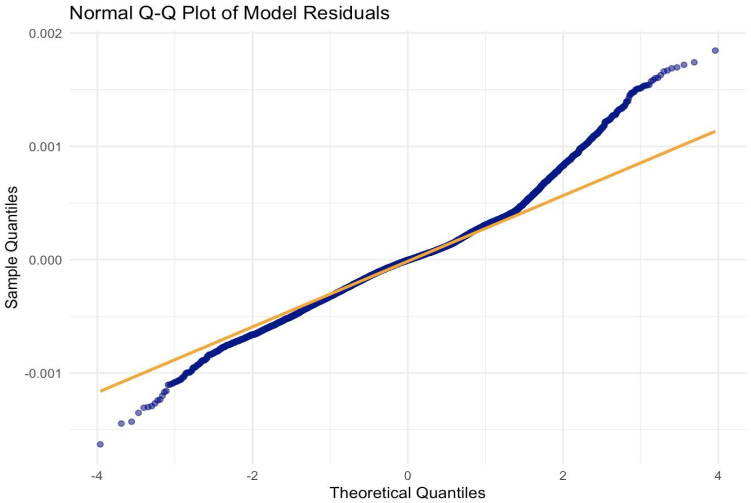
- Population Composition: From 1985 to 2022, younger age groups are decreasing and older age groups are increasing in size
- Mortality rates: decrease in overall mortality rates over time
  - Could help explain why we see increased population sizes for older age groups
  - Healthier lifestyles, improvements in medical care access and quality

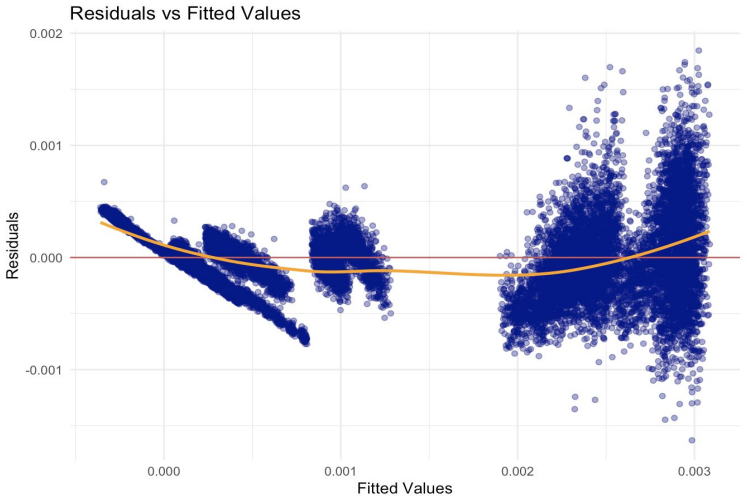
Estimating Expected Mortality Before 2017

Goal: Find the best model to predict mortality rate to get an expected value to compare to observed data

- Better predictions → more likely to detect true periods of excess mortality when they occur

Linear Model: $$Y_i \sim \mathrm{N}(\mu,\sigma^2),\ Y_i = \beta_0 + \beta_1 * Population_i + \beta_2 * Week_i + \beta_3 * Age_i + \beta_4 * Sex_i + \epsilon_i$$





- Normality and constant variance clearly violated, model may not be good for predicting
- 8% of fitted values negative for non-negative outcome (rate)

Estimating Expected Mortality Before 2017

Log-linear Model: $Y_i \sim \text{Poisson}(\mu_i),\ log(\mu_i) = \beta_0 + \beta_1 * Week_i + \beta_2 * Age_i + \beta_3 * Sex_i + log(Population_i)$

- Using an offset for population, we can predict mortality rate rather than death count to compare across groups with different population sizes
- AIC: 125,630
- Residual Deviance: 49,066 on 13,354 degrees of freedom
  - Ratio of residual deviance to degrees of freedom: 3.67, indicating overdispersion
  - Violates model assumption that the mean be directly equal to the variance

Negative Binomial Model:

$$Y_i \sim \text{NegBin}(\mu_i, \theta),\ log(\mu_i) = \beta_0 + \beta_1 * Week_i + \beta_2 * Age_i + \beta_3 * Sex_i + \beta_4 * Sex * Age_i + log(Population_i)$$

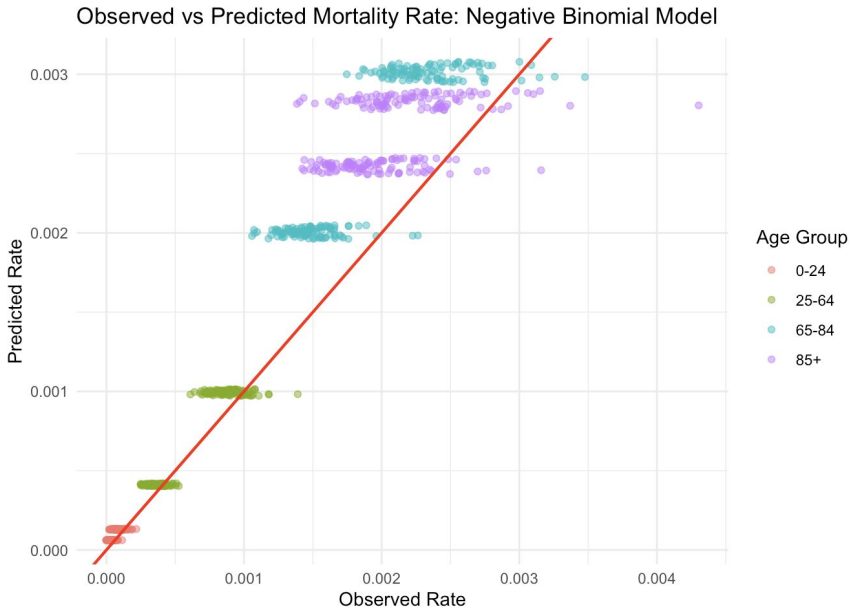
- Negative binomial model corrects for this via a dispersion parameter θ, increasing standard errors when the variance is estimated to be greater than the mean:

$$\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

- AIC: 101,244
- Residual Deviance: 13,764 on 13,351 degrees of freedom (ratio much closer to 1 👍)

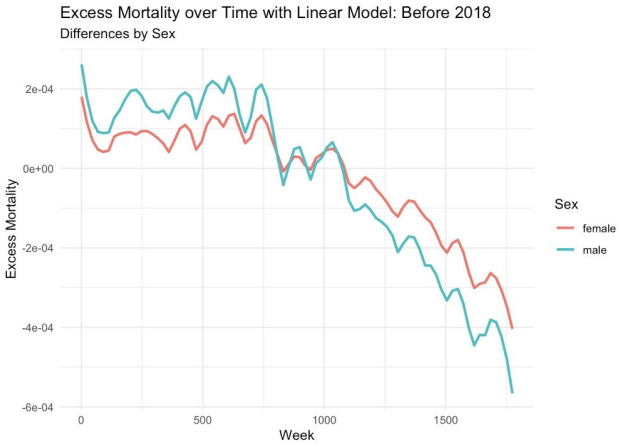
Estimating Expected Mortality

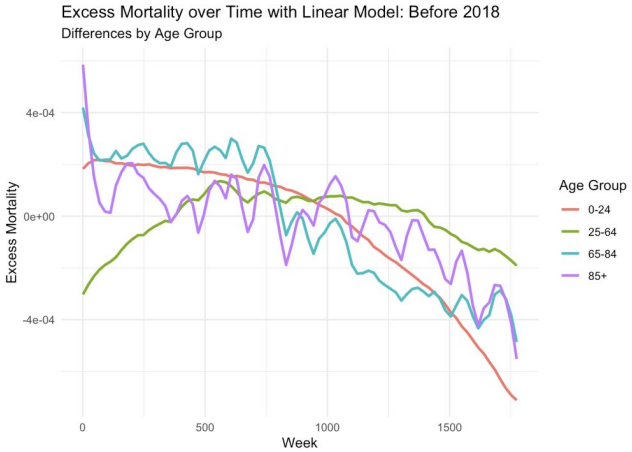
- Negative Binomial model most successful at predicting mortality rate

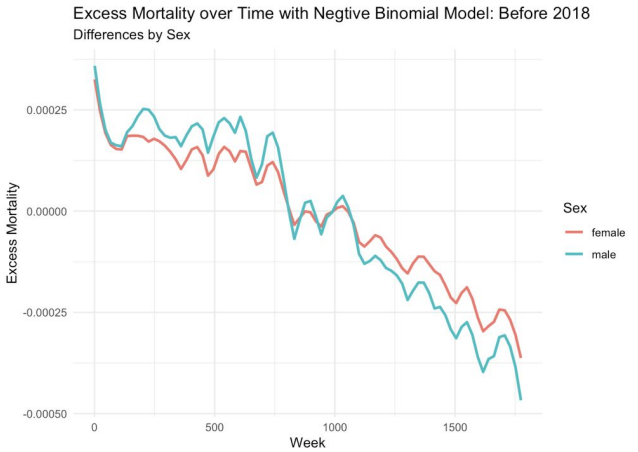


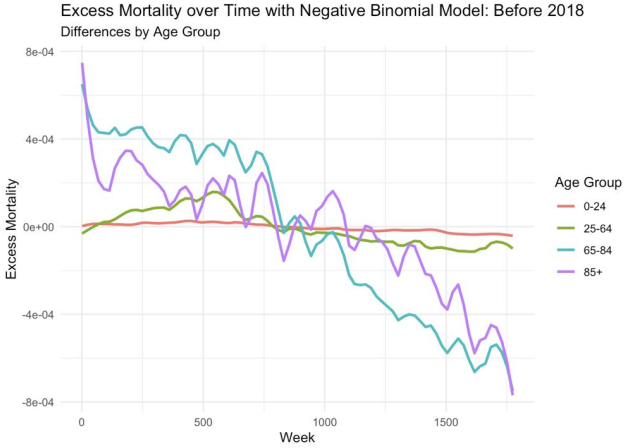
- Still over-predicting for older age groups, significant variation left unexplained - could be fixed by accounting for trends over time

Periods with Excess Mortality Before 2018



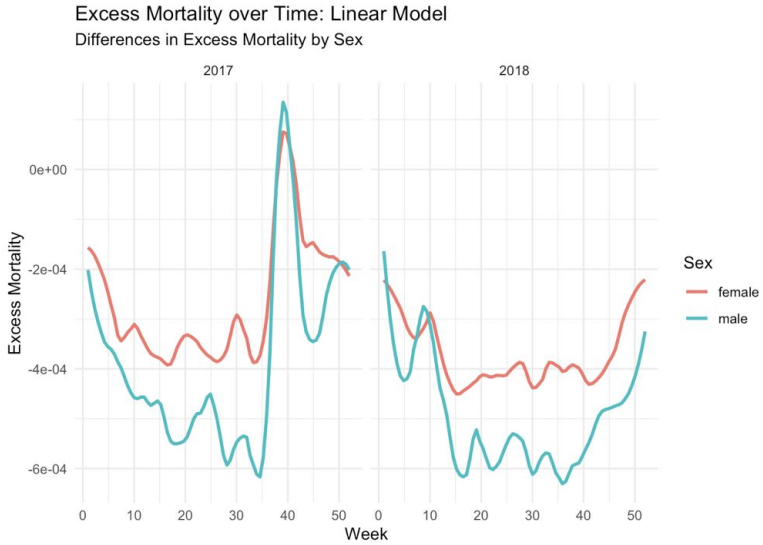


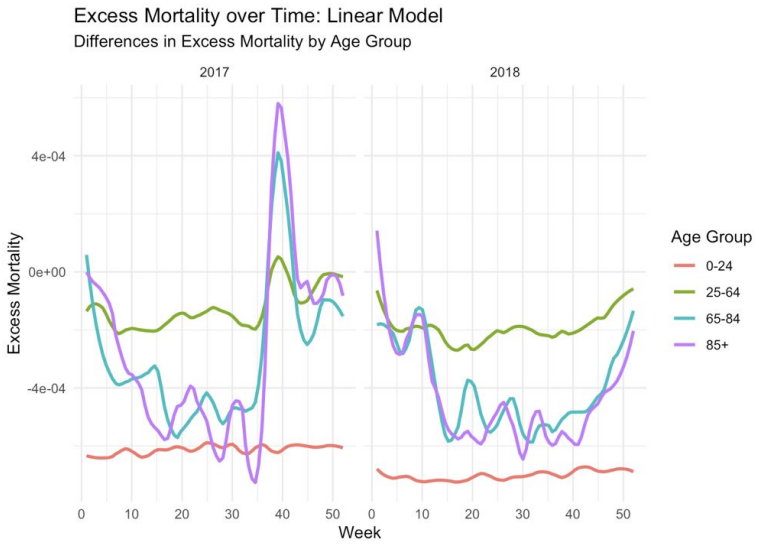




- Excess Mortality: **Observed Rate - Predicted Rate**
- Higher levels of excess mortality around 1985, potentially due to the Puerto Rico Floods in 1985
- Removing times around the 1985 Floods did not change results much

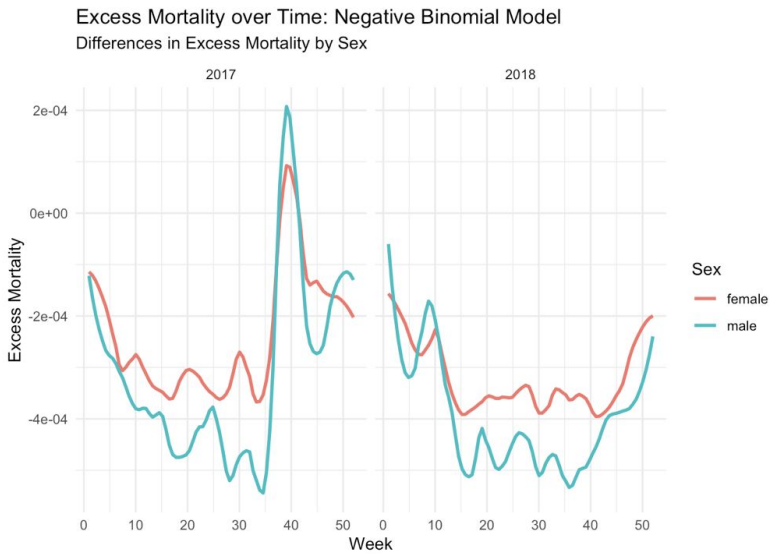
Excess Mortality for 2017-2018: Linear Model

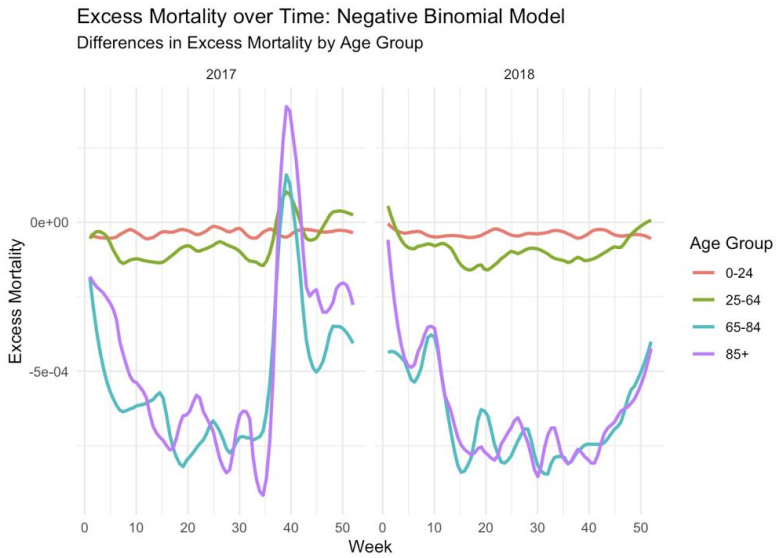




- Model consistently overestimated mortality rates for both genders and age groups, especially for 0-24 age group
- Spike in change in excess mortality around the time Hurricane Maria made landfall for both gender and age groups besides 0-24 age group

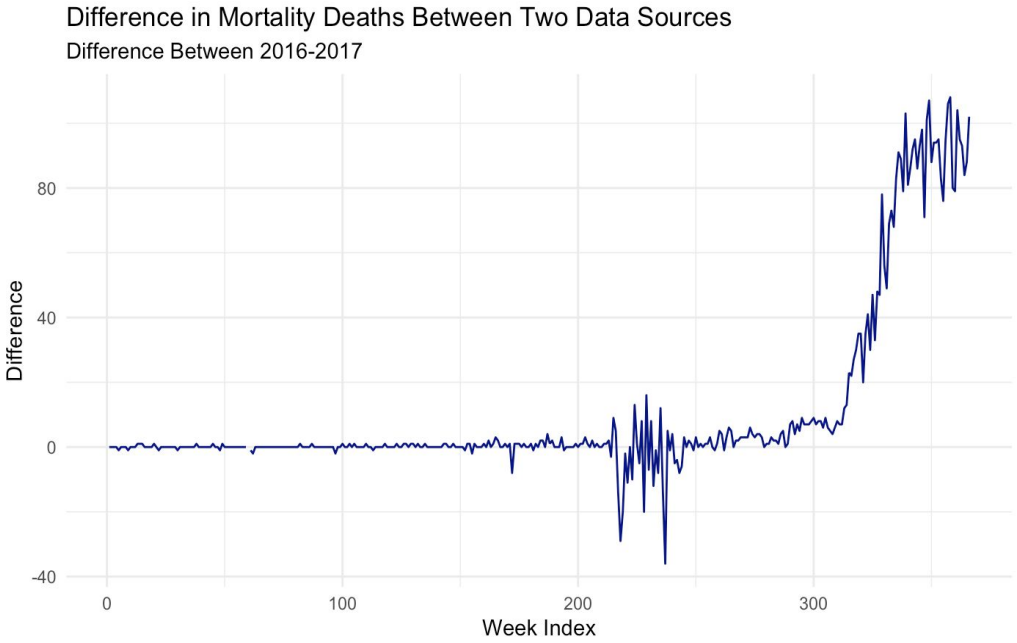
Estimating Mortality for 2017-2018: Negative Binomial





- Model still overestimated mortality rates for both gender but the difference between the two genders is smaller
- Model did pretty well at predicting mortality rates for the younger age groups, not as good for older age groups with similar predictions for 65-84 and 85+ age groups
- Improvement from linear model predictions

Alternative Data Sources - NYT Data



- Both the excessmort and NYT data reported similar data for most months
- Some dates in August had data discrepancies
- NYT data did not have any data on December 2017 resulting in the spike in differences
- Highlights importance of data quality and source

Future Directions

- Using hierarchical models to account for strong age group effect

- Time series models to account for week/seasonal effects
  - Allowing death count to be a smooth function of time could help us predict mortality more effectively using splines:

$$Y_t \sim \text{NegBin}(\mu_t, \theta), log(\mu_t) = \text{linear terms} + f(t) + log(Population_t)$$

- Integrate additional data sources to account for other potential confounders
  - Socioeconomic status
  - Geographic locations
  - Healthcare access