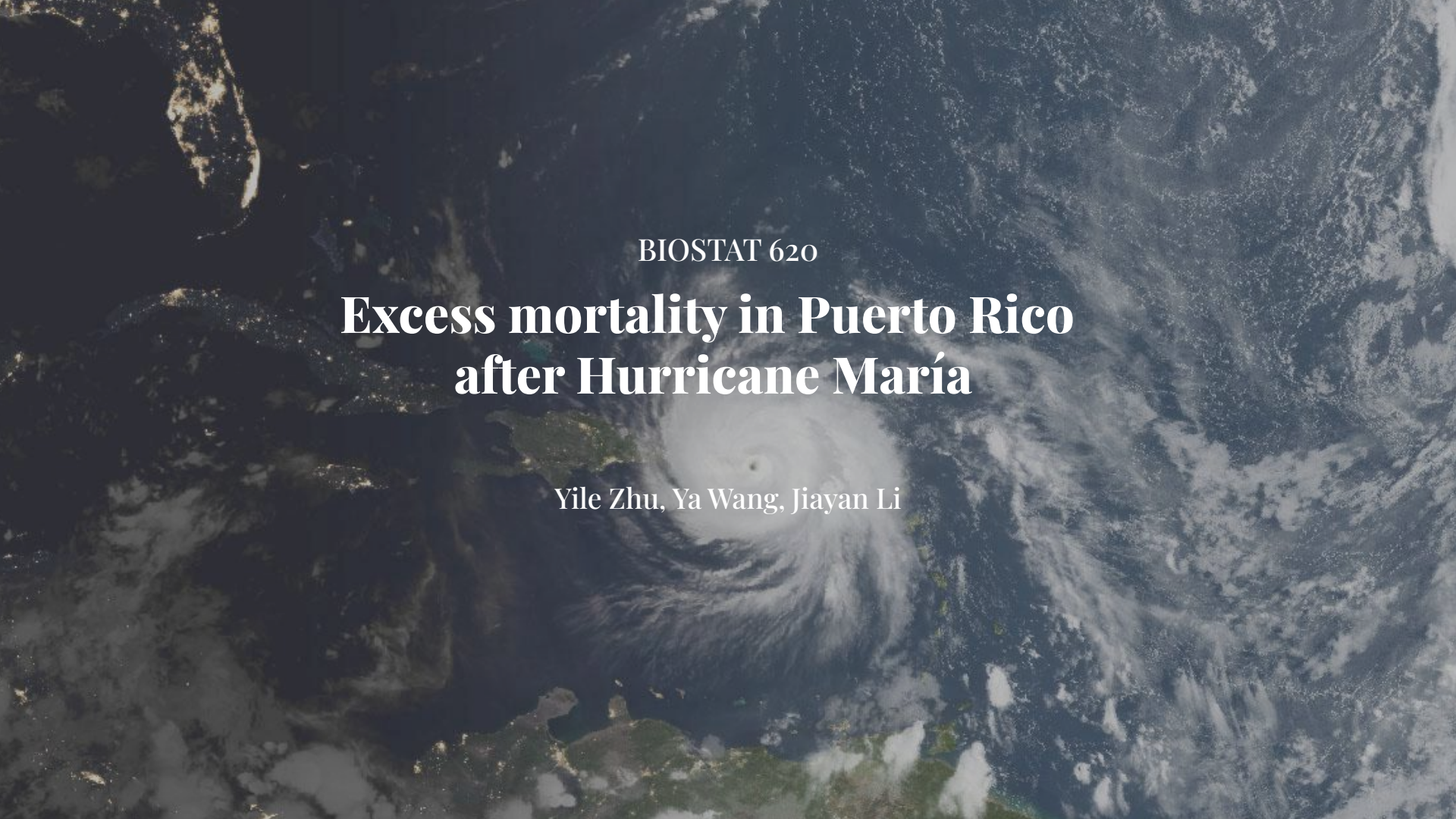


BIOSTAT 620

Excess mortality in Puerto Rico after Hurricane María

Yile Zhu, Ya Wang, Jiayan Li

Introductio

Objective:

To quantify excess mortality following Hurricane María in Puerto Rico and assess its demographic impact using statistical modeling

Key Tasks:

1. Population Patterns - Analyzed age/sex distribution
2. Baseline Mortality Modeling - Used pre-2017 data to calculate expected weekly deaths
3. Historical Excess Mortality Check - Check unusual death periods before 2017
4. Post-María Impact - Calculated weekly excess deaths from Sept 20, 2017
5. Data Validation - Compared NY Times PDF with our model

Data Sources

- Puerto Rico weekly mortality & population data from excessmort R package (1985–2018)
- Daily mortality tables extracted from PDF released via The New York Times (2017 only)

Data Preparation

- Cleaned and aggregated weekly deaths by age group and sex
- Combined granular age groups into broader categories (e.g., 0–24, 25–49...) based on similar mortality patterns
- Removed anomalous years (1985, 2016–2017) for baseline model training

Modeling Approach

- Statistical model

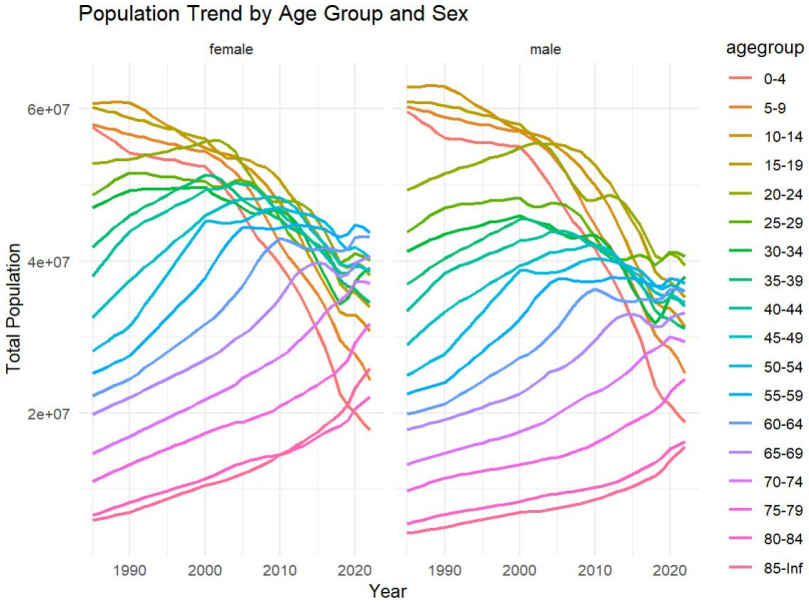
$$Y_i \sim \text{NegBin}(\mu_i, \theta), \quad \text{with } \log(\mu_i) = \mathbf{x}_i^\top \beta$$

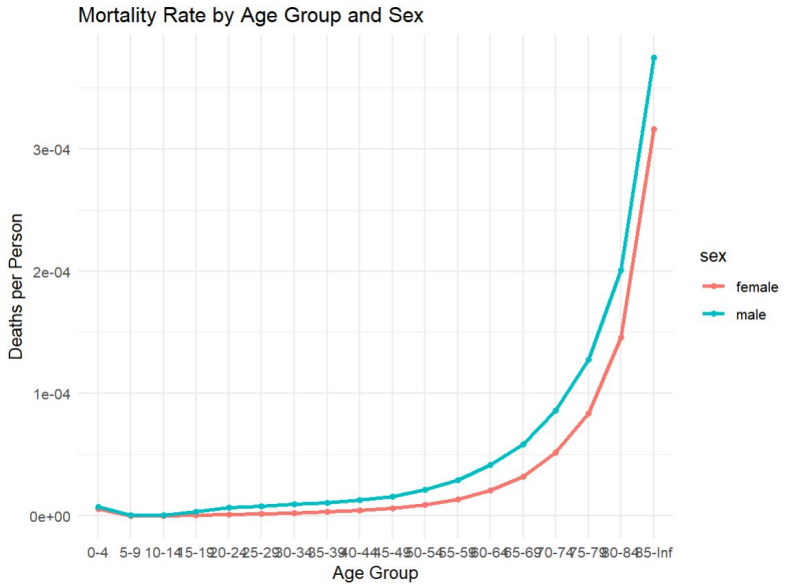
- Excess mortality estimate

$$\hat{\Delta}_{[t_0,t_1]} = \sum_{t=t_0}^{t_1} \hat{\mu}_t \hat{f}(t).$$

Result: Task 1

Examine the population sizes by age group and sex. Describe any interesting patterns.





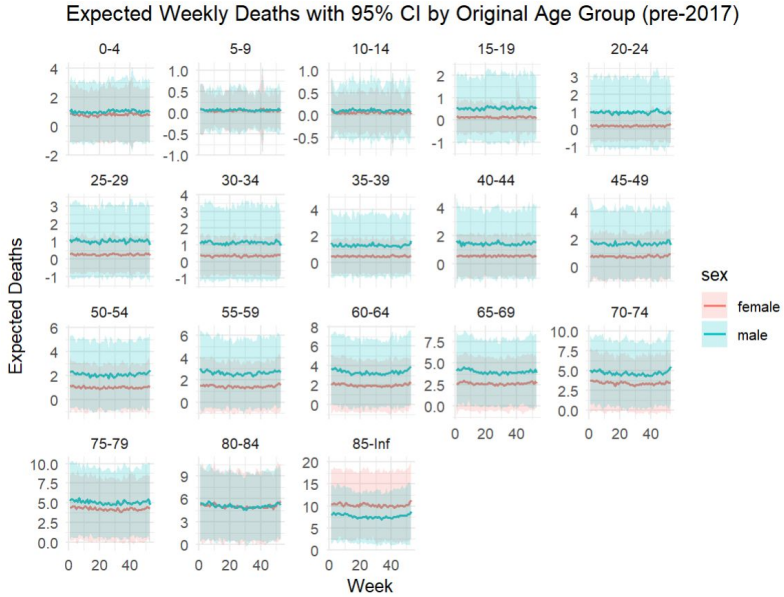
Largest groups: The majority of the population is concentrated in the 25–64 age range.

Sex distribution: Female population tends to be higher in older age groups (65+), likely due to longer life expectancy.

Smallest groups: Children under 5 and elderly over 85 constitute the smallest proportions.

Task 2

Use data from before 2017 to estimate expected mortality and a standard deviation for each week by age group and sex.

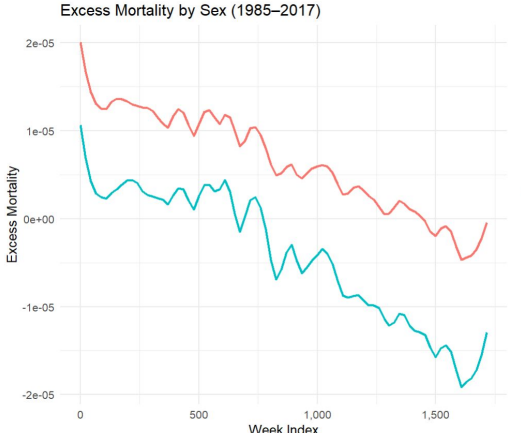


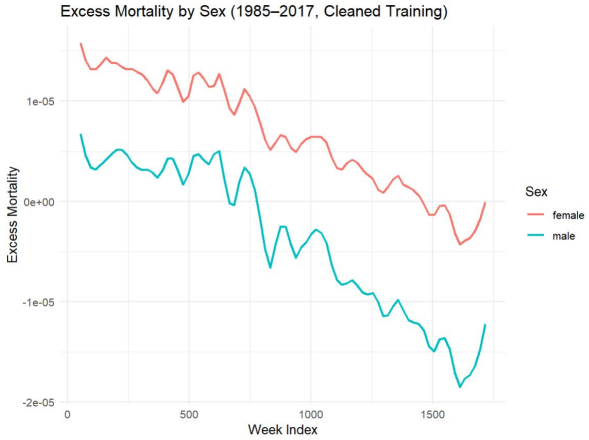
Estimated expected deaths and standard deviations up to 2017, with observations showing **similar mortality and standard deviations** in some age groups, combining the original ages into the following five groups: 0–24, 25–49, 50–64, 65–84, 85+. Then recalculate the mean and standard deviation of the merged age groups and conduct modeling.

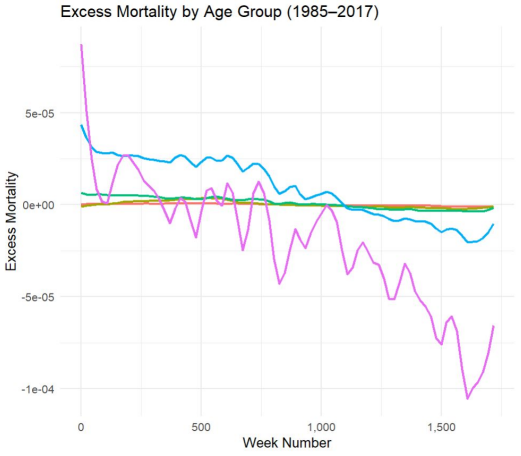
Result patterns: Seasonal mortality patterns were observed; older groups showed higher baseline mortality.

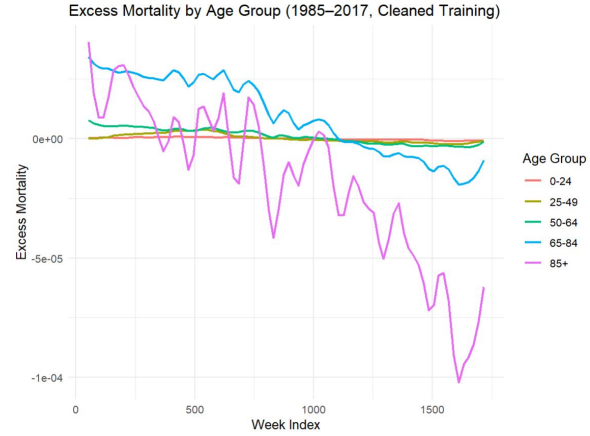
Task 3

- Excess deaths in 1985 **among the elderly (aged 65 and above) especially 85+**
- The elderly population is especially vulnerable to external shocks such as public health crises during historical cycles.
- Gender chart: From 1985, the excess mortality rates for both women and men were positive, with especially for women.
- Cleaned model: A new model **excluding these outliers** was used for 2017–2018 prediction.

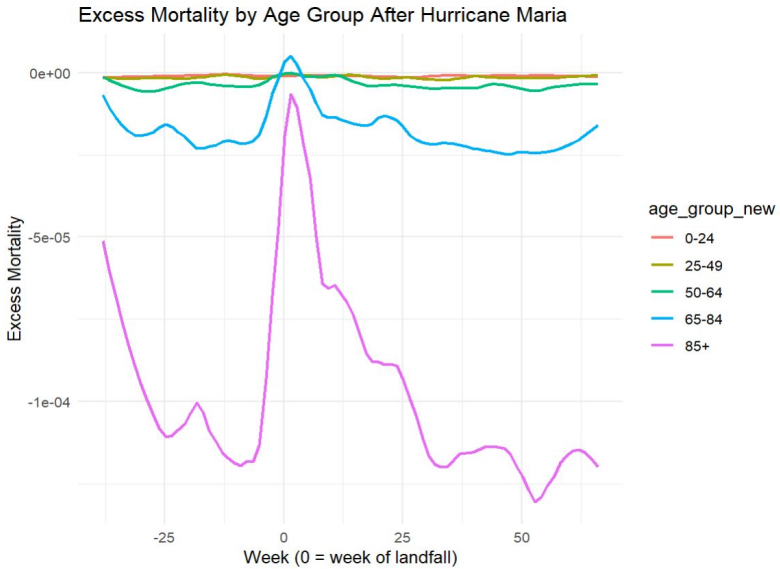






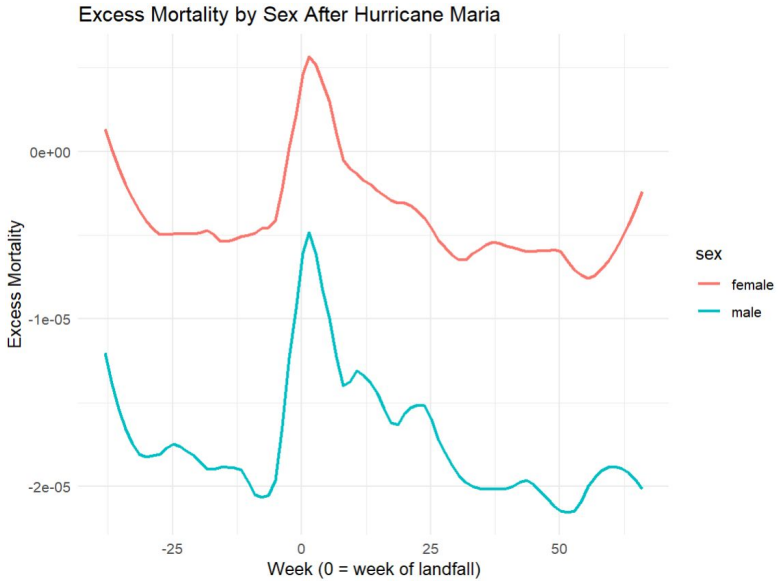


Task 4



We observed **a sharp increase** in excess mortality following the landfall of Hurricane Maria on 20 September 2017, **particularly in older age groups**.

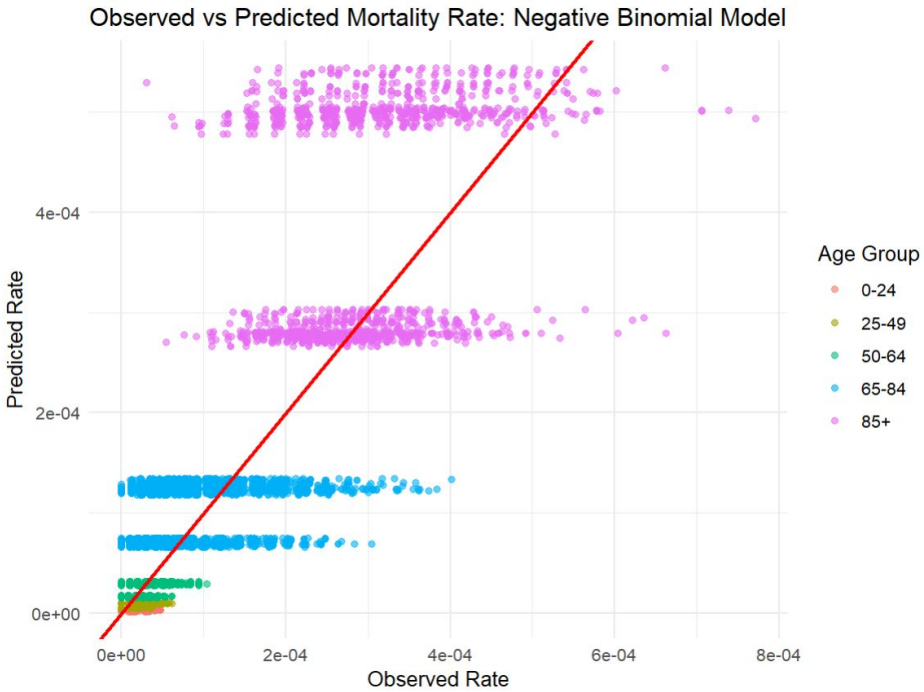
The elderly, especially those **aged 85 and above**, experienced the highest excess mortality.



Males were more affected than females both in the magnitude and duration of excess deaths.

Task 4

Model fitting comparison chart（observed vs predicted）



Accurate fit across age groups: Each age group formed distinct horizontal clusters, with most points closely aligned with the identity line, suggesting the model effectively captured group-specific mortality risk.

Higher mortality groups (65–84 and 85+) showed particularly strong alignment, indicating the model's robustness in estimating rates for more vulnerable populations.

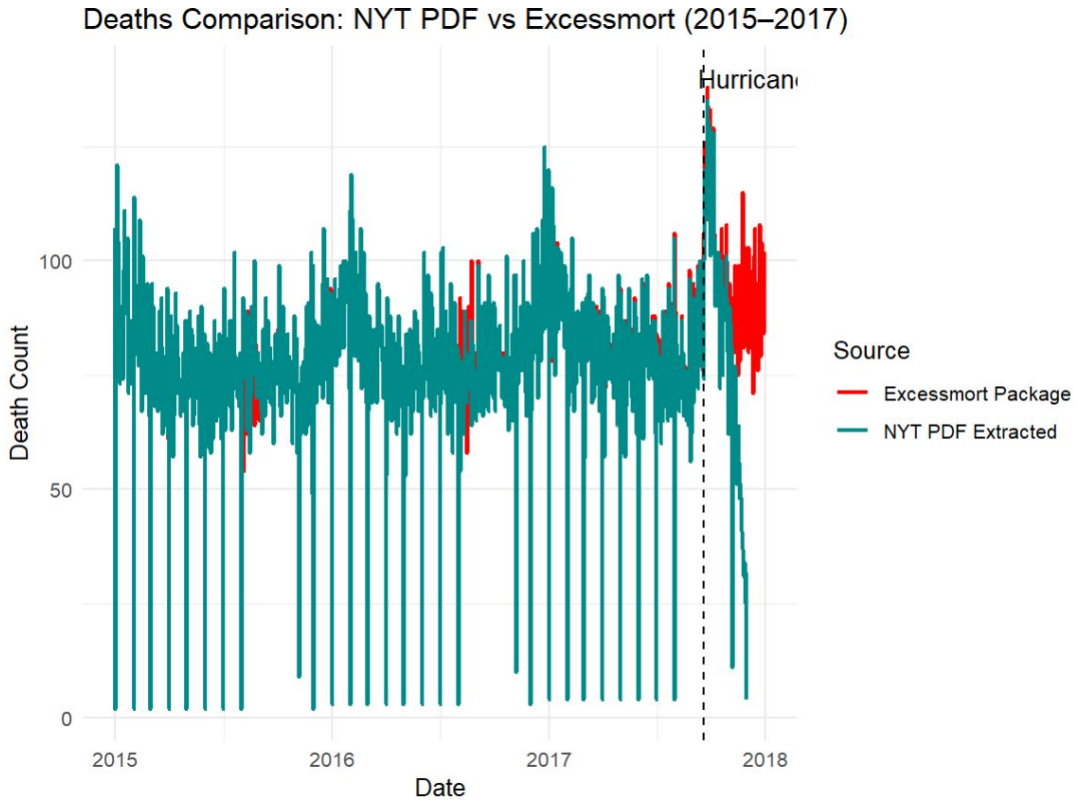
Younger groups (0–24, 25–64) had lower mortality rates, and though their data concentrated near the origin, their predictions also fell tightly along the identity line.

Overall, the predictive accuracy observed in this plot supports the model's validity for downstream estimation of expected mortality, especially in excess mortality calculations post-Hurricane Maria.

Task 5

Both data sources reflect a **sharp spike in daily deaths immediately after the hurricane**, peaking in late September to early October 2017.

However, the **NYT PDF (official registry)** data appears more variable and occasionally lower in some days than the smoother estimates from the excessmort model, which may have applied smoothing or modeled expectations based on historical baselines.



🧠 Extended discussion of age and gender differences

1. Older age groups (65–84 and 85+) have much higher excess mortality after Hurricane Maria than other age groups.

- **Vulnerability of older age groups**

The elderly population already has many underlying diseases (such as cardiovascular, diabetes, etc.), and their physiological adaptability to disasters is poor.

In the aftermath of the disaster, Puerto Rico's infrastructure collapsed (water outages, power outages, communication outages) and prevented the elderly from receiving timely care and medicines for chronic diseases.

2. Excess mortality is generally higher in men than in women at multiple ages, especially in older age groups.

- **Possible reasons for the higher mortality rate in men**

Some studies have shown that men are less able to manage psychological stress after disasters, which may lead to an increase in acute events (eg, stroke, heart attack).

🧠Comparative analysis with NYT providing PDF data

The daily death data taken from NYT's report is generally consistent with the trends we predicted based on the excessmort packet, but there is a slight deviation in absolute terms at some time points, especially 3–4 weeks after the hurricane

Discussion of potential sources of discrepancy and historical background

- **Data latency and reporting lag**

Maria's post-Puerto Rico statistical system was severely paralyzed, and many of the death data were entered with delays of weeks or even months, while NYT may have been completed through post-tracking or special channels.

excessmort uses the CDC's official death statistics, and the delay may be greater due to a more standardized process.

- **Differences in statistical caliber and methodology**

NYT data is based on "registered deaths" daily population, including hospital and community deaths; Excessmort, on the other hand, is based on merged monthly files, which is more standardized but slightly less accurate.

Limitations and room for improvement

The predictive model does not control for seasonal trends or concurrent events, such as other natural disasters or influenza.

Data were aggregated and lacked individual-level variables, limiting the evaluation of interventions.

Future Work can try time series modeling or Bayesian models to improve responsiveness to extreme events.


Thank you!