# Predicting Hurricane Deaths

Group 2 - Adi Arifovic, Oliver Baldwin Edwards, Leah Johnson
May 13th, 2020

# Purpose

- To predict the number of hurricane deaths
  - Using variables such as maximum sustained wind speed, atmospheric pressure, and property damage
- Hypothesis:
  - The number of deaths would *increase* with an increase in maximum sustained wind speed and property damage.
  - Intuitively it would make sense for more destructive hurricanes to also be more deadly.

# Data

# Source

- Dataset “hurricNamed” from the “DAAG” R package
  - 94 named US hurricanes from 1950 to 2012
  - Includes: number of deaths, damage, weather statistics
  - Cited by a research paper

RESEARCH ARTICLE

**Female hurricanes are deadlier than male hurricanes**

Kiju Jung, Sharon Shavitt, Madhu Viswanathan, and Joseph M. Hilbe

PNAS June 17, 2014 111 (24) 8782-8787; first published June 2, 2014 https://doi.org/10.1073/pnas.1402786111

Edited* by Susan T. Fiske, Princeton University, Princeton, NJ, and approved May 14, 2014 (received for review February 13, 2014)

# Variables:

## Response Variable

- `deaths` describes the number of human deaths that occurred due to each hurricane (ranging from 0 to 1846).

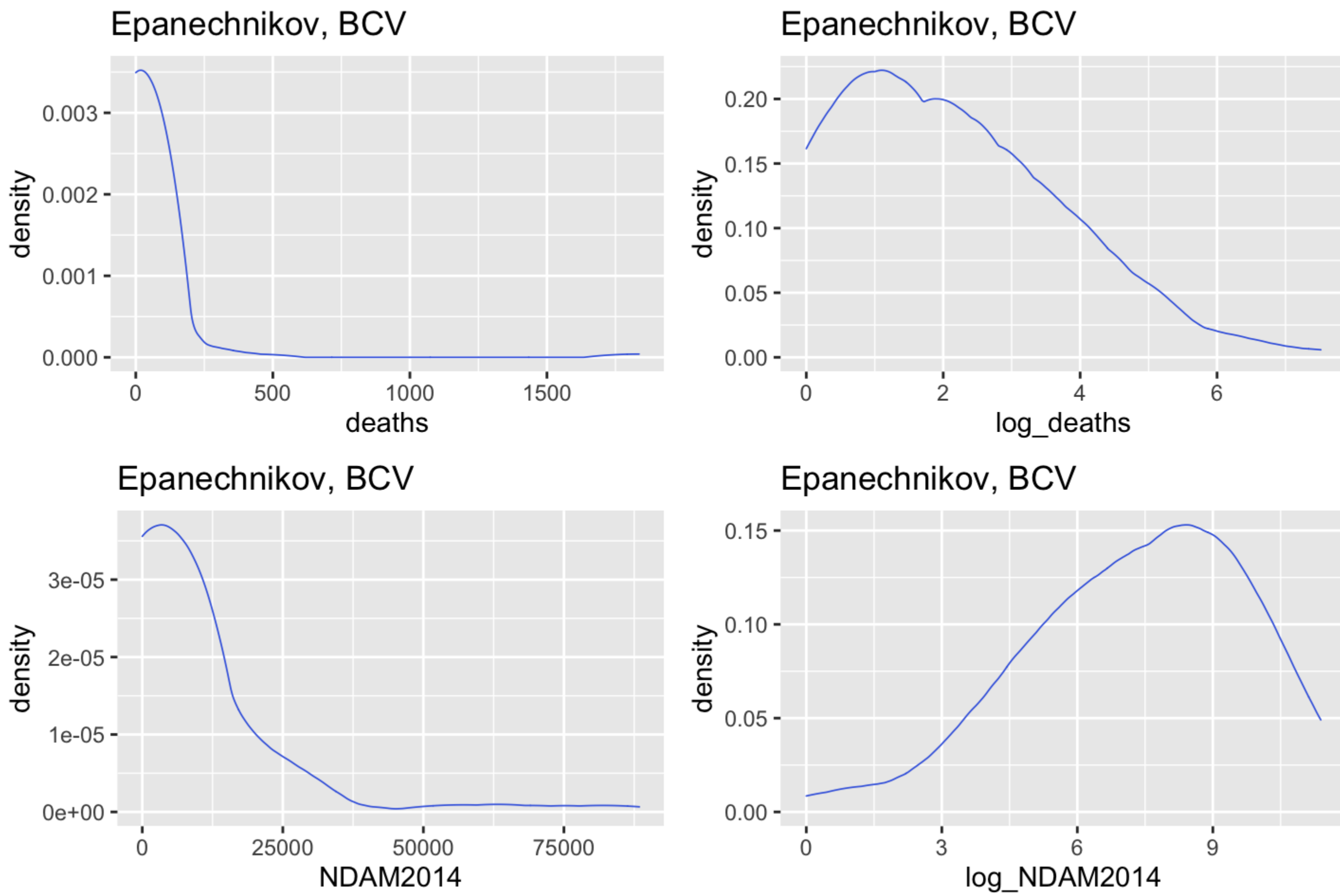
## Explanatory Variables

- `LF.WindsMPH`: max wind speed (mph)
- `LF.PressureMB`: atmospheric pressure (millibars)
- `LF.times`: number of times the hurricane made landfall
- `BaseDam2014`: property damage (millions of 2014 US dollars)
- `NDAM2014`: normalized damage (millions of 2014 US dollars)

# Exploratory Analysis

# Kernel Density Estimation

- `deaths`, `NDAM2014`, and `BaseDam2014` had very skewed kernel density estimates.
- Adding a `log()` transformation to `deaths`, `NDAM2014`, and `BaseDam2014` resulted in more normal kernel density estimates and minimized the effects of the outliers.


Epanechnikov, BCV
density
0.003
0.002
0.001
0.000
0
500
1000
1500
deaths
Epanechnikov, BCV
density
0.20
0.15
0.10
0.05
0.00
0
2
4
6
log_deaths
Epanechnikov, BCV
density
3e-05
2e-05
1e-05
0e+00
0
25000
50000
75000
NDAM2014
Epanechnikov, BCV
density
0.15
0.10
0.05
0.00
0
3
6
9
log_NDAM2014

# Correlation Tests

- We use Kendall's $\tau$ at a significance level $\alpha = 0.05$
- Hypotheses tested:
  - The null hypothesis of no association (independence) $H_0 : \tau \leq 0$
  - The alternative hypothesis of positive association $H_A : \tau > 0$.
  - (We also tested for negative associations with the set of hypotheses $H_0 : \tau \geq 0$ and $H_A : \tau < 0$.)

# Correlation Table

| Variable | Hypothesis Tested | $p-$value | Correlation Estimate |
|---|---|---|---|
| NDAM2014 | $H_A : \tau > 0$ | $8.073 \cdot 10^{-16}$ | 0.578 |
| LF.times | $H_A : \tau > 0$ | 0.008542 | 0.209 |
| LF.PressureMB | $H_A : \tau > 0$ | 1 | $-0.431$ |
| LF.PressureMB | $H_A : \tau < 0$ | $1.915 \cdot 10^{-9}$ | $-0.431$ |
| LF.WindsMPH | $H_A : \tau > 0$ | $7.606 \cdot 10^{-6}$ | 0.331 |
| BaseDam2014 | $H_A : \tau > 0$ | $7.691 \cdot 10^{-15}$ | 0.558 |

Table 1: Correlation Test Results

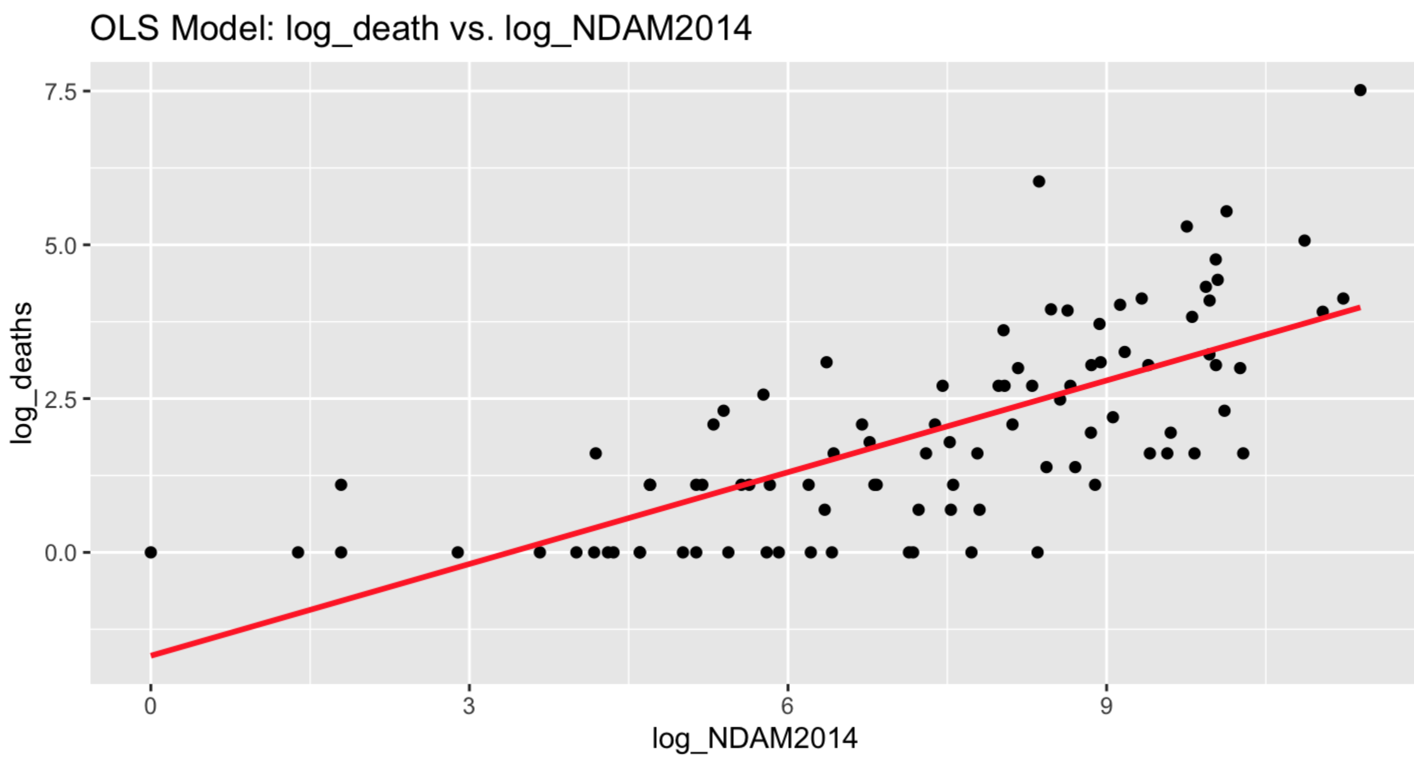
# Model Fitting

# Ordinary Least Squares Regression

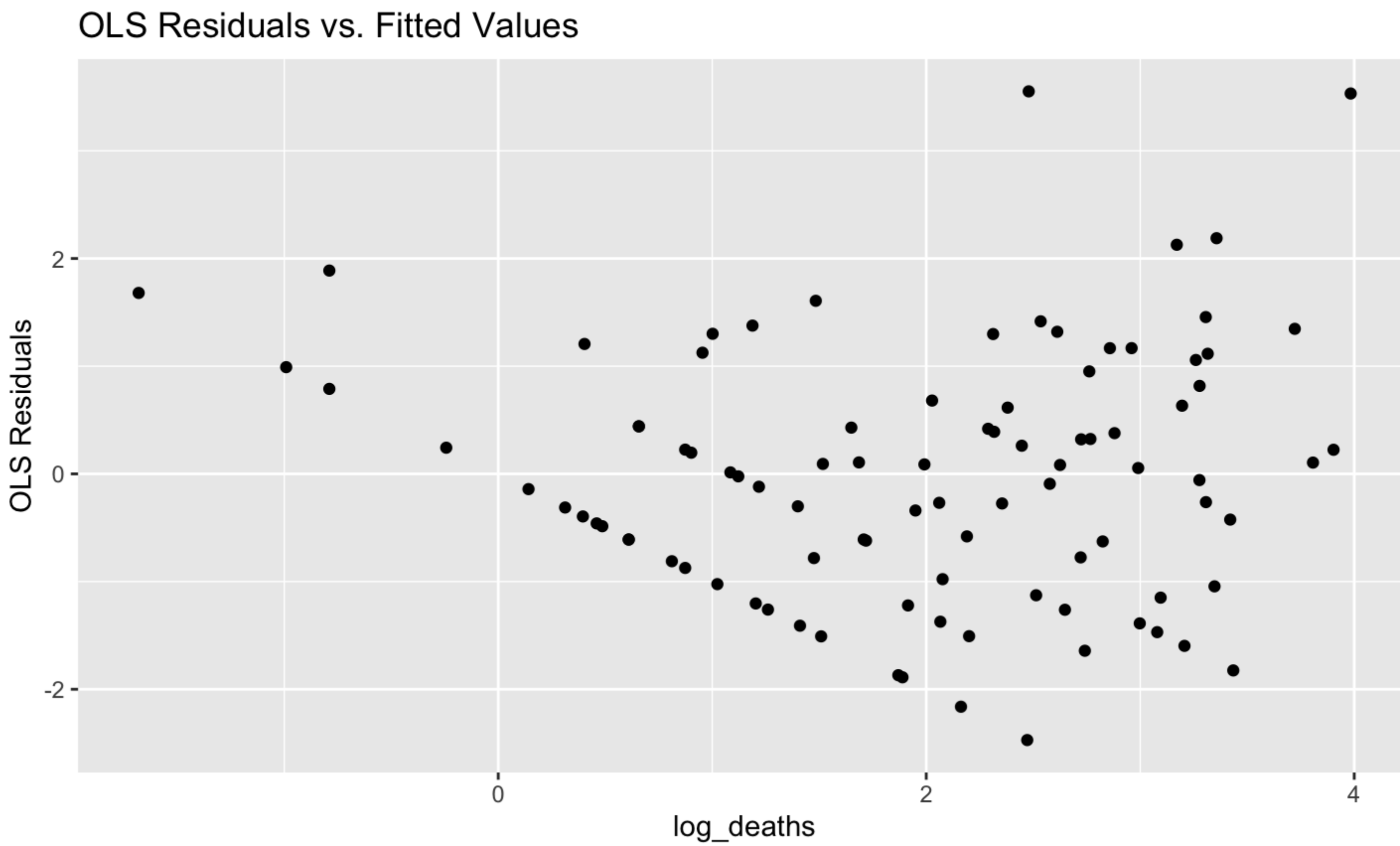
- Carried out best subsets selection method and series of nested F tests
- Assumptions for OLS:
    - Independence
    - Linearity & Equal Variances
    - Normality of Residuals

# Best OLS model

- The best OLS model was the model predicting `log_deaths` with simply `log_NDAM2014` (normalized damage)
- AIC = 301.11, and $R^2_{adj}$ = 50.08%.



# OLS Residuals vs. Fitted



# Kolmogorov-Smirnov Test

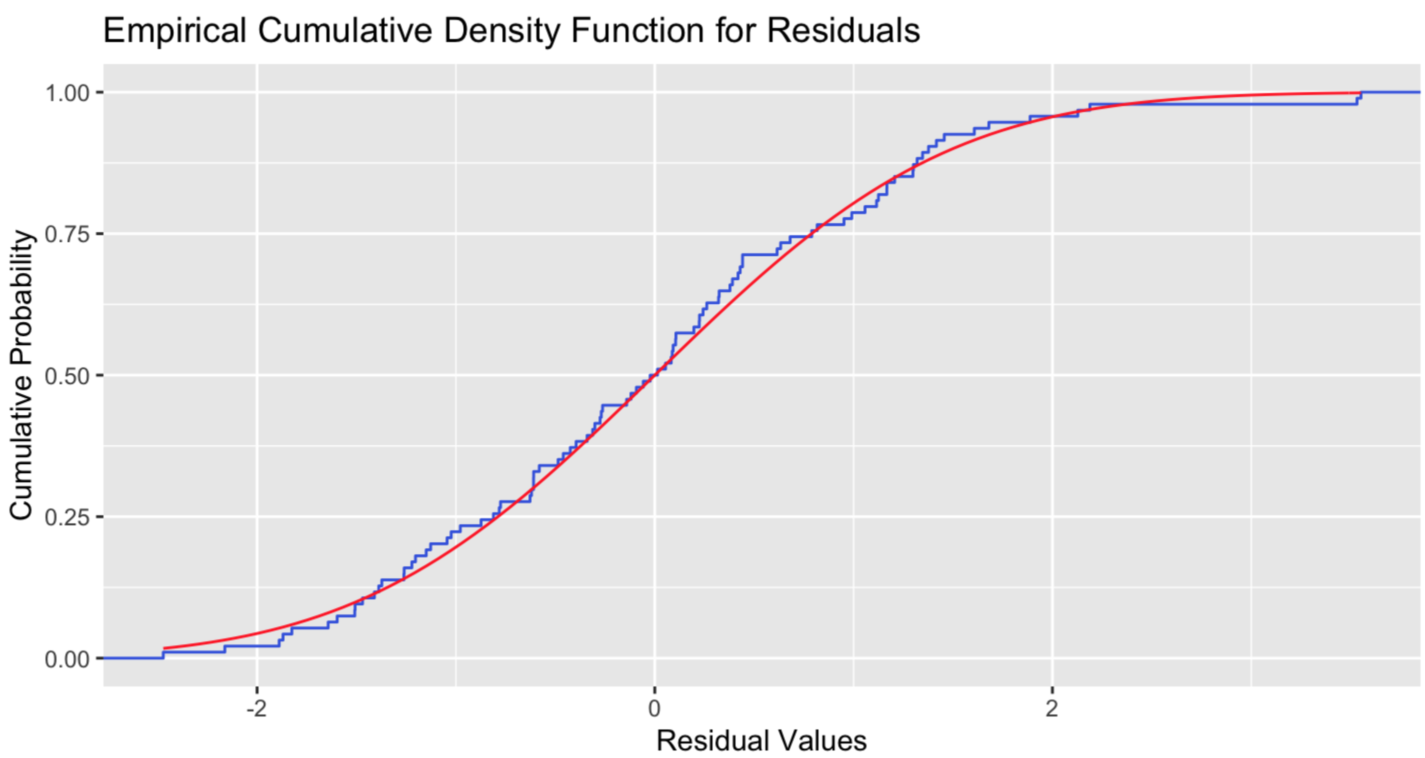
- Ran this test to check the assumption of normally distributed residuals for our above selected linear regression model.
- Hypothesis:

$$
\begin{aligned}
H_0 &: F(t) = F^*(t) \\
H_A &: F(t) \neq F^*(t) \text{ for at least one } t
\end{aligned}
$$

Where $F(t)$ refers to the estimated CDF of the distribution of residuals of our linear model, and $F^*(t)$ is the CDF of the normal distribution.

# Kolmogorov-Smirnov Test Results

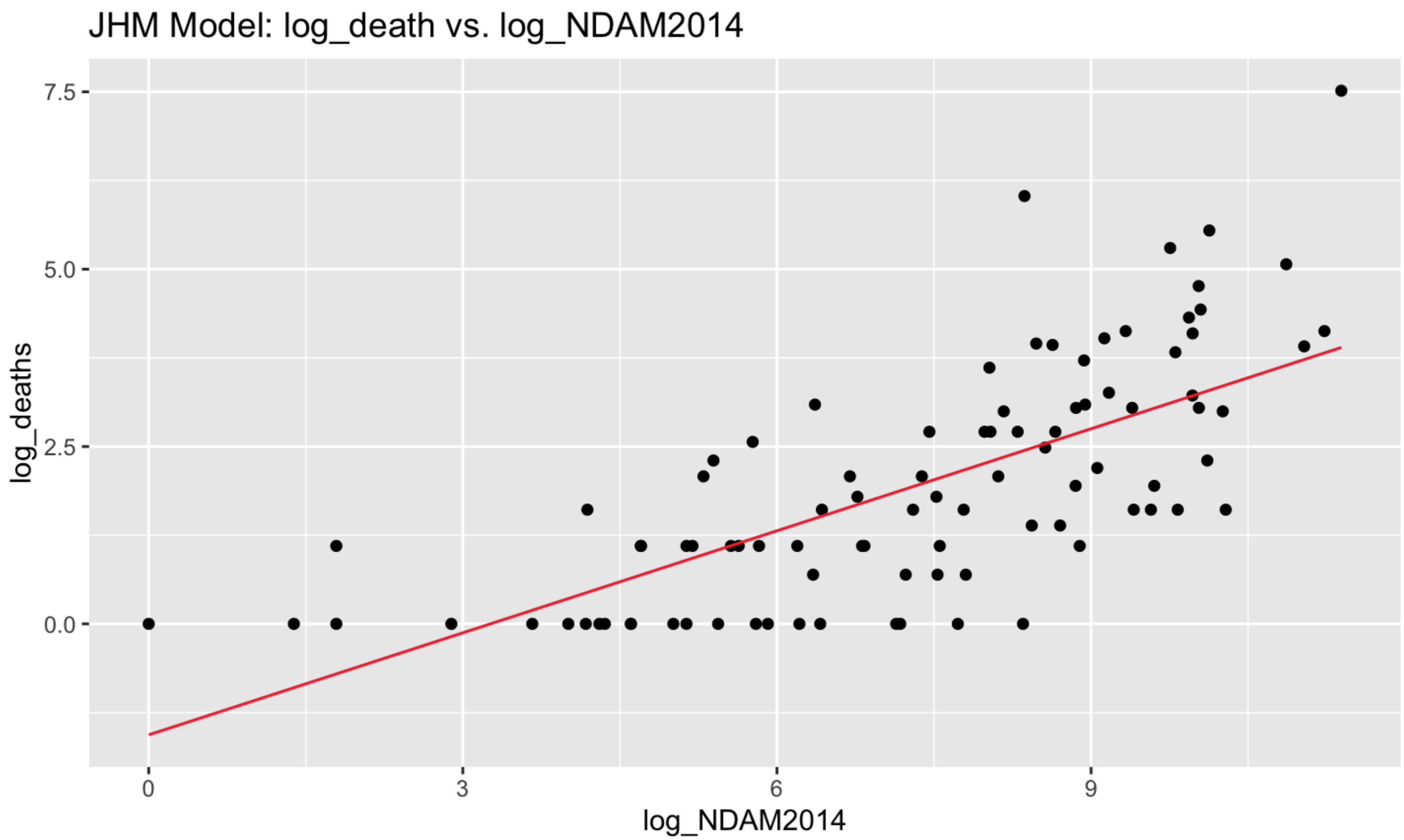
- $p-$value of $0.7095 > 0.05$



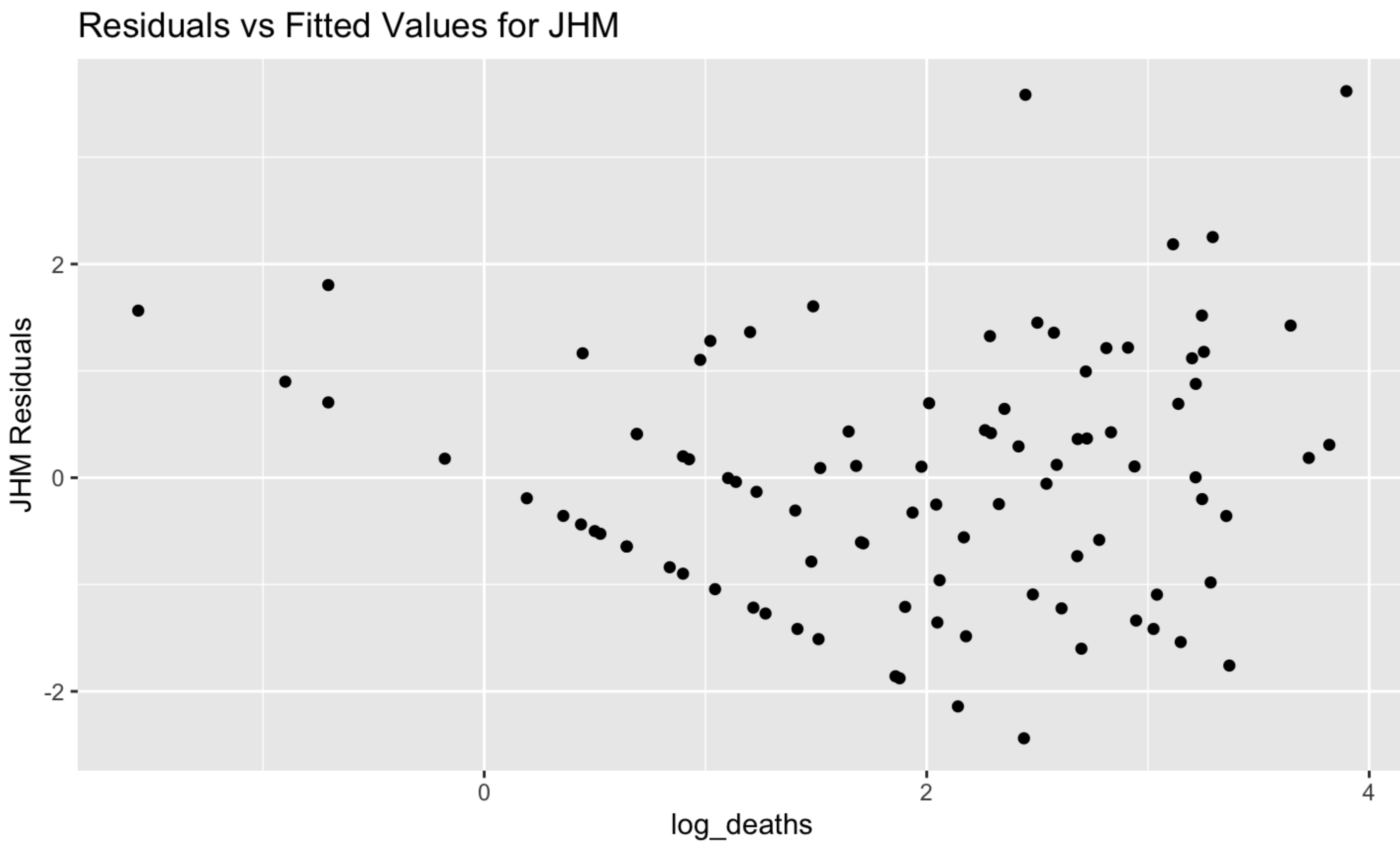
# Rank-based JHM Model

- Fit due to the issue with the equal variances condition in the OLS regression model
- Performed drop in dispersion tests to ensure the significance of each predictor in the model.
- Found that the best model predicting `log_deaths` was again one that used only `log_NDAM2014`.
  - $R^2$ = 48.07%

# Best JHM Model



# JHM Residuals vs. Fitted

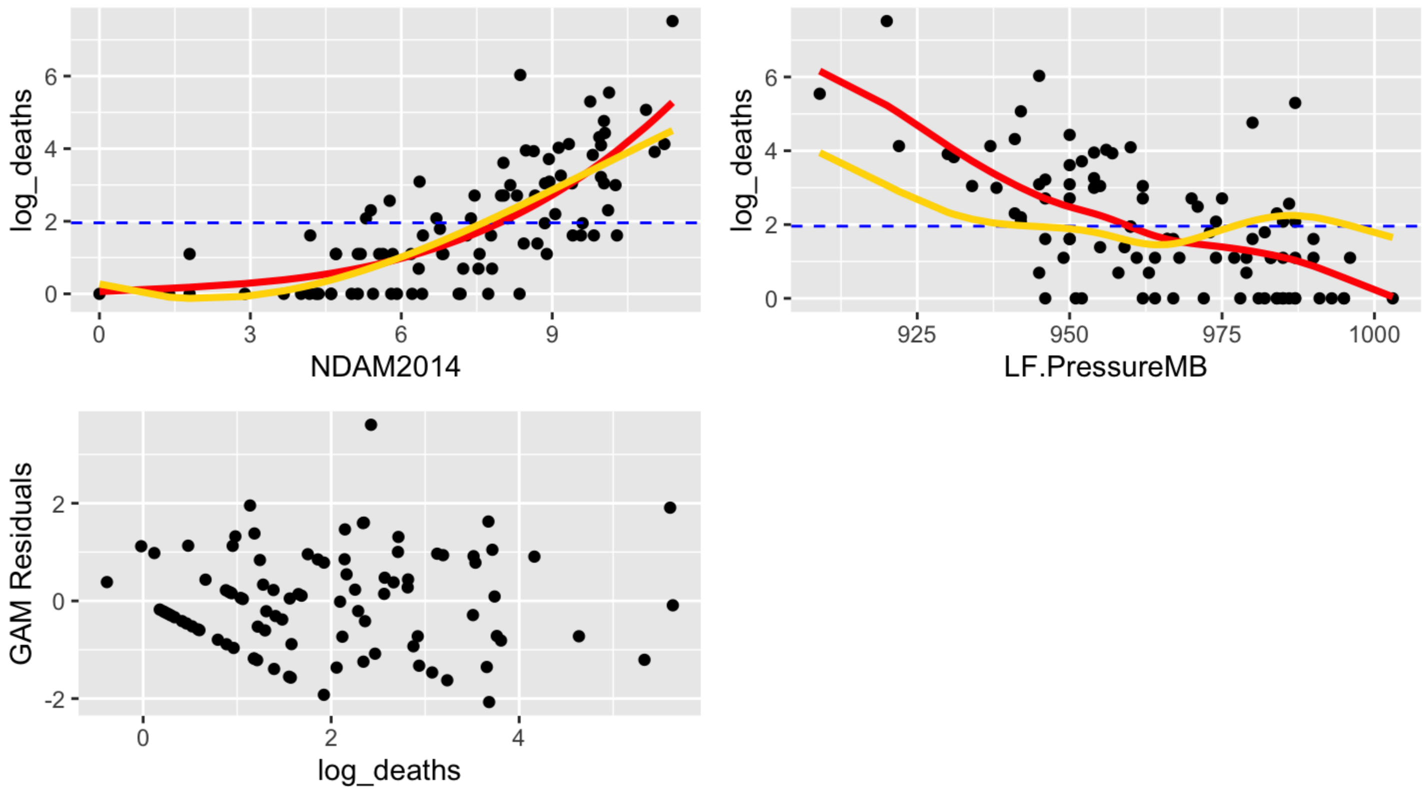


# Generalized Additive Model

# Fitting log_NDAM2014 & LF.PressureMB to the GAM

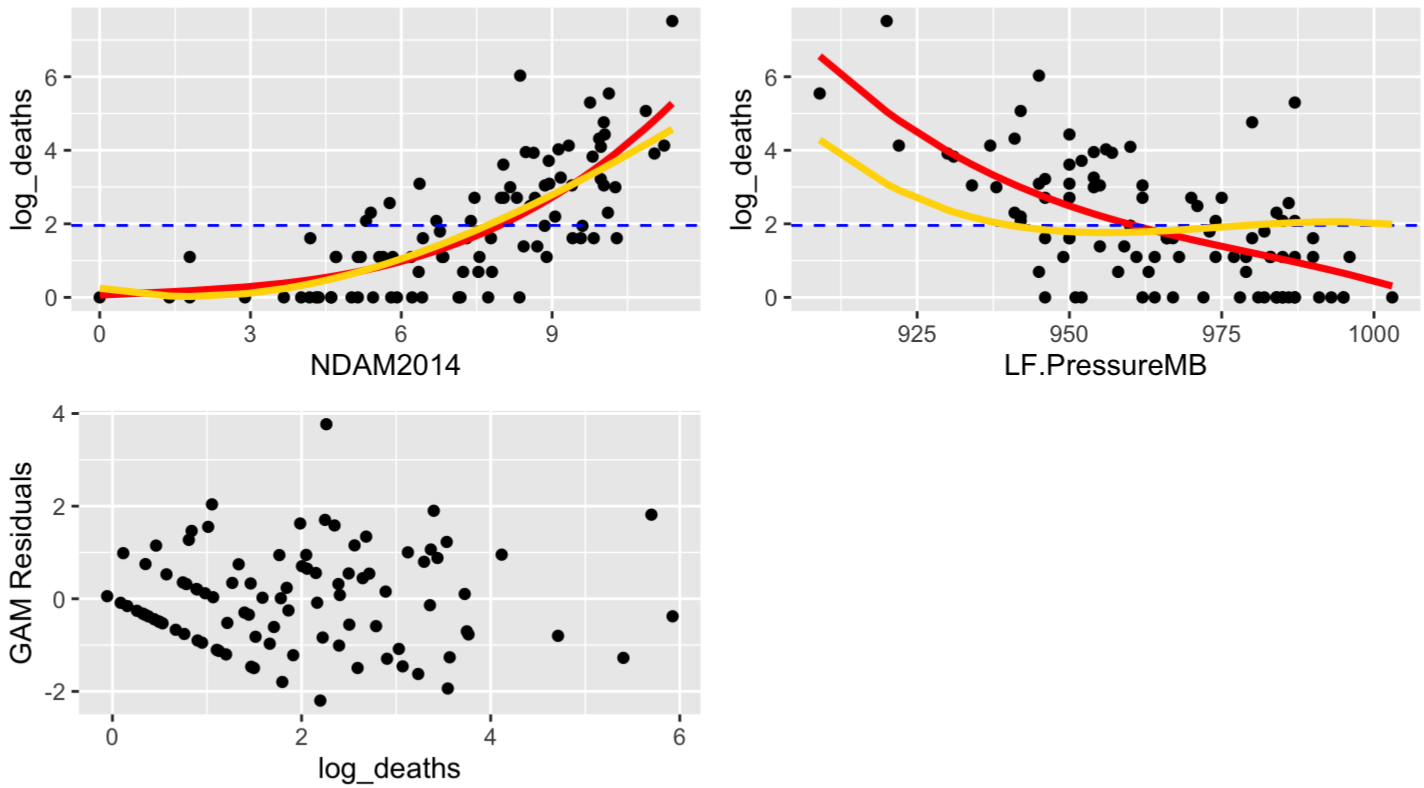
- GAM used to look for possible trends and relationships that were missed by other models primarily due to restrictions by linearity.
- Used manual forward selection to fit the best model.
- The best way to fit `log_NDAM2014` was with a b-spline.
- The best way to fit `LF.PressureMB` was with an s-spline (while keeping the previous b-spline for `log_NDAM2014`).
    - $AIC = 288.015$

# Additional Changes to the GAM




- After plotting, the `LF.PressureMB` scatterplot displayed a poor fit. ( $AIC = 288.015$)

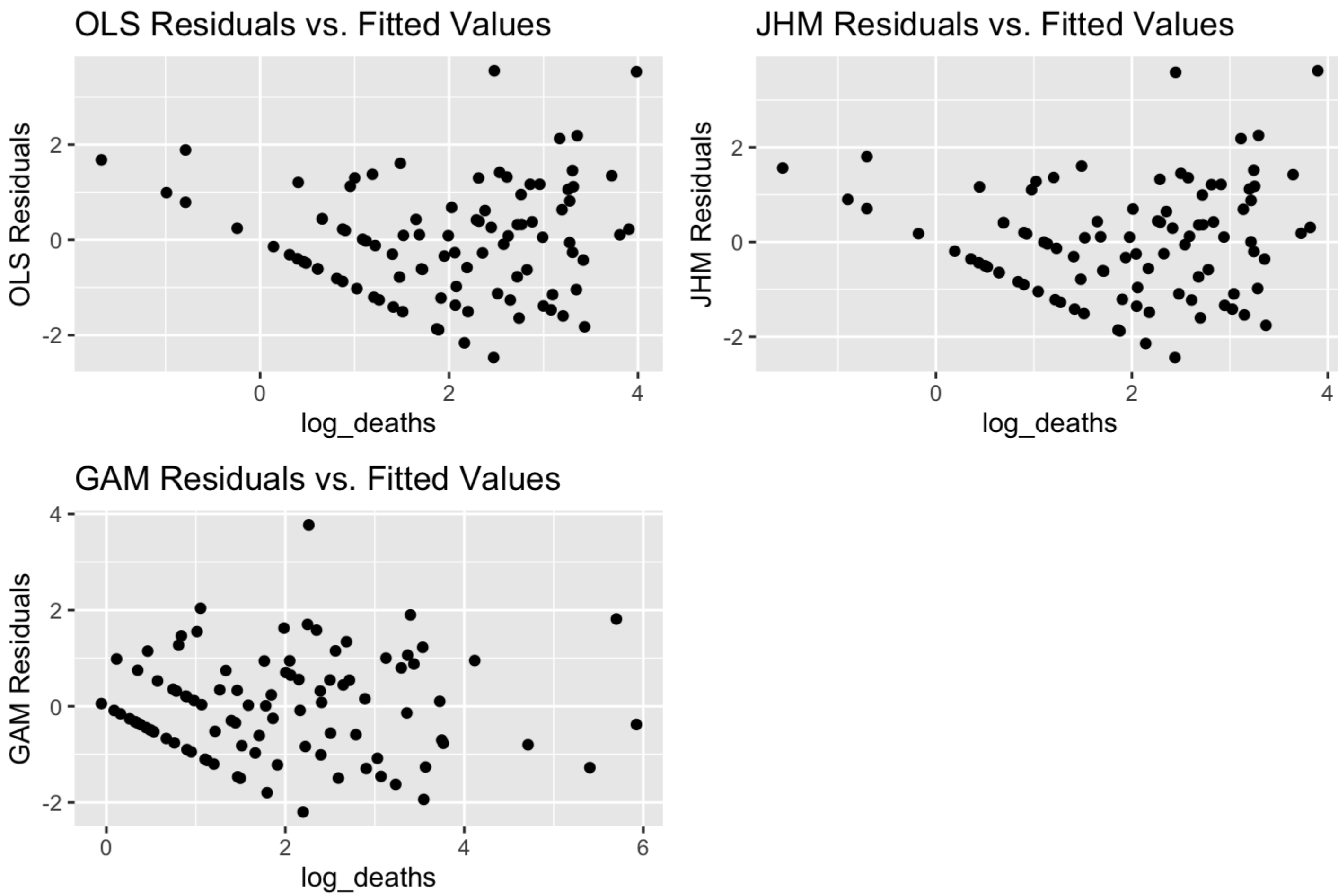
# Final GAM Model




- Applying a cubic fit to `LF.PressureMB` vastly improved the fit. ( $AIC = 290.884$)

# Conclusion

# Comparison of Residuals



# Comparison of Fit

| | $R^2$ | $R^2_{adj}$ | $L1_{prop}$ |
|---|---|---|---|
| Best OLS | 0.5061 | 0.5008 | 0.3195 |
| Best JHM | 0.5054 | 0.5000 | 0.3183 |
| Best GAM | 0.6018 | 0.5743 | 0.3781 |

# Comparison of Fit cont'd

| | cv $R^2$ | cv $R^2_{adj}$ | cv $L1_{prop}$ |
|---|---:|---:|---:|
| Best OLS | 0.4870 | 0.4814 | 0.3083 |
| Best JHM | 0.4867 | 0.4811 | 0.3027 |
| Best GAM | 0.5203 | 0.4872 | 0.3193 |

# Findings

- `gam(log_deaths ~ bs(log_NDAM2014) + poly(LF.PressureMB, 3))`
  - Low AIC = $290.884$
  - Better graphical representation
  - Higher $R^2$, $R^2_{adj}$, $L1_{prop}$
  - Ability to fit local trends
- Matches original hypothesis
  - More severe hurricane = more deaths

# Limitations

- Limited numerical predictors
- Reliance on normalized hurricane damage (`NDAM2014`)
- Reliance on log transformations
  - Potential for better log transformations

# References


- https://www.rdocumentation.org/packages/DAAG/versions/1.24
- https://www.pnas.org/content/111/24/8782#ref-28
- https://ascelibrary.org/doi/abs/10.1061/(ASCE)1527-6988(2008)9:1(29)
- https://www.rhinobldg.com/understanding-barometric-pressure-in-hurricanes/

# Questions?