Access to Drinking Water and Mortality rate

Ryan Díaz-Pérez

Deaths due to unsafe water sources

- Clean drinking water is an essential human right, and an essential resource.
- Regions with the largest rate of population growth rate depict the largest share of deaths by unsafe water sources.
- A factor in this is the lack of access to improved water sources. Improved water sources are defined as pipes or wells that provide water directly to people's homes.
- Sub-Saharan Africa is the region that has the largest share of deaths by unsafe water sources, and the largest population growth rate.

Who would help this analysis

- World governments would are very interested in this analysis.
- If we know what factor has the biggest impact on reducing deaths due to unsafe water sources, governments will have better insight to tackle the problem.
- Humanitarian organizations are also interested in this analysis, because they can aim their efforts towards actions that will have the greatest impact.

Data Sources

- Percentage of the population with access to improved drinking water over time, downloaded from ourworldindata.org
- share of deaths from unsafe water sources over time, downloaded from ourworldindata.org
- The total population by region,downloaded from data.worldbank.org
- The population growth rate by region, downloaded from data.worldbank.org

Models used to fit the data

- Linear Regression Model
- Multi-linear Regression Model
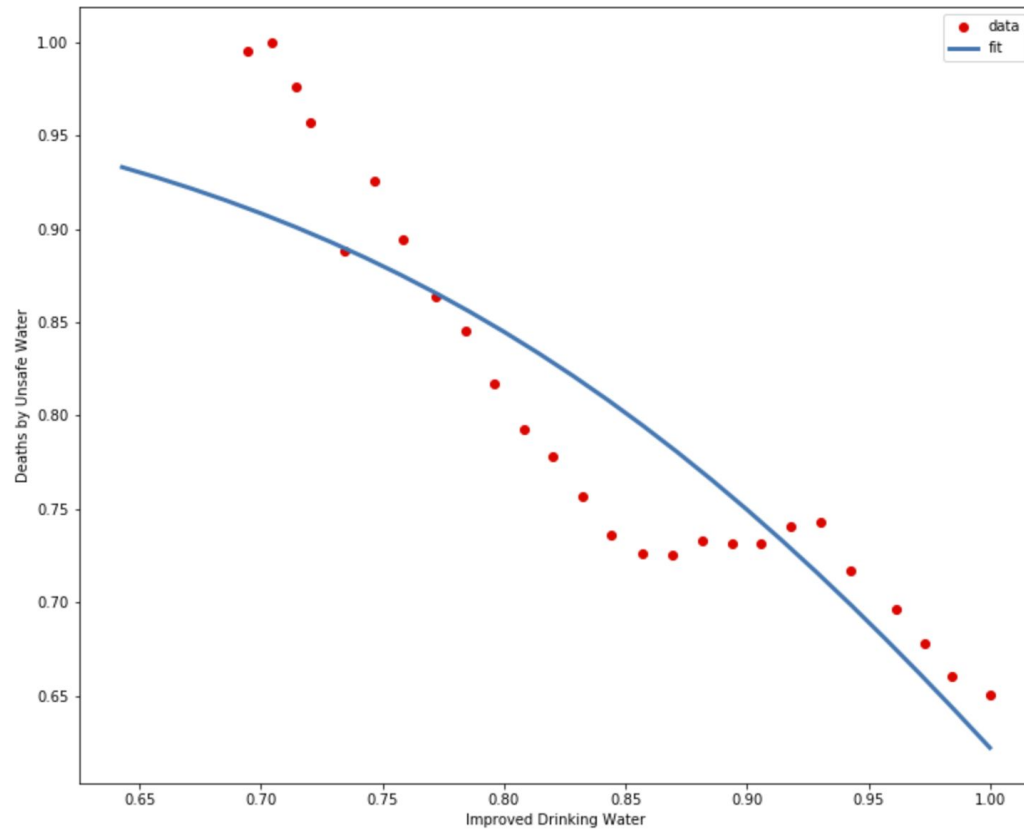- Polynomial Regression Model
- Non-linear Regression Model

Linear Regression Model, distribution plot

Actual vs Fitted Values for Deaths by Unsafe Water
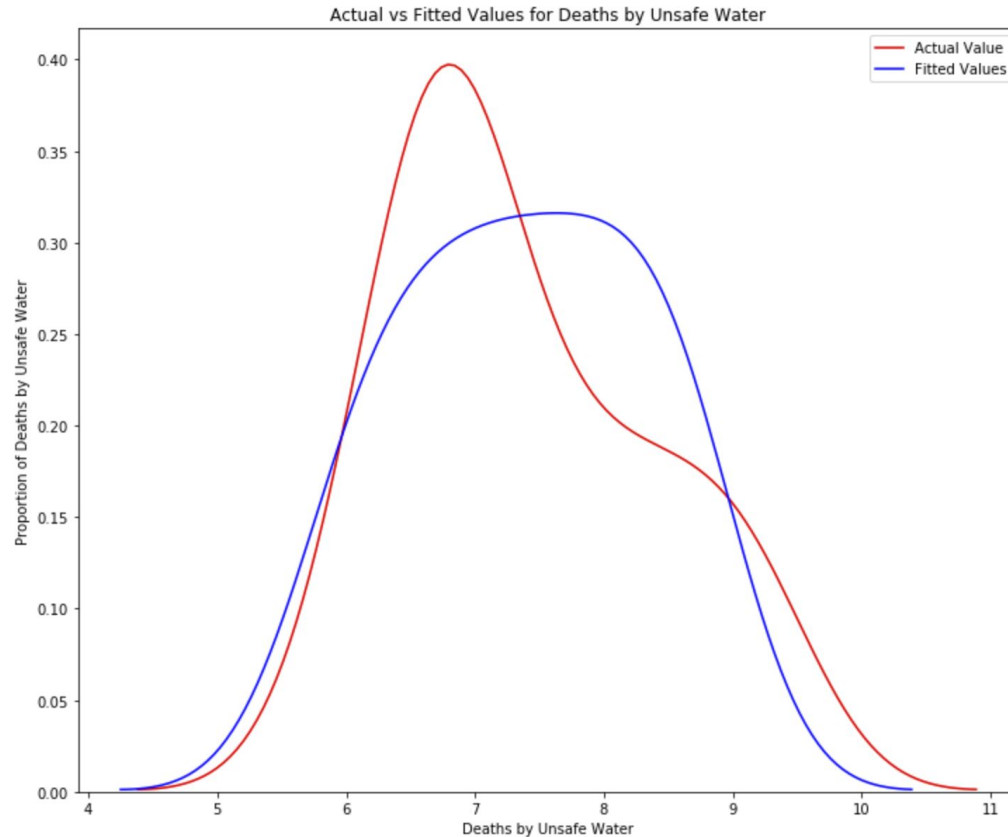
Multi-linear Regression Model, distribution plot

Non-linear Regression Model

Polynomial Regression Model

Polynomial Fit with Matplotlib for Deaths by Unsafe Water
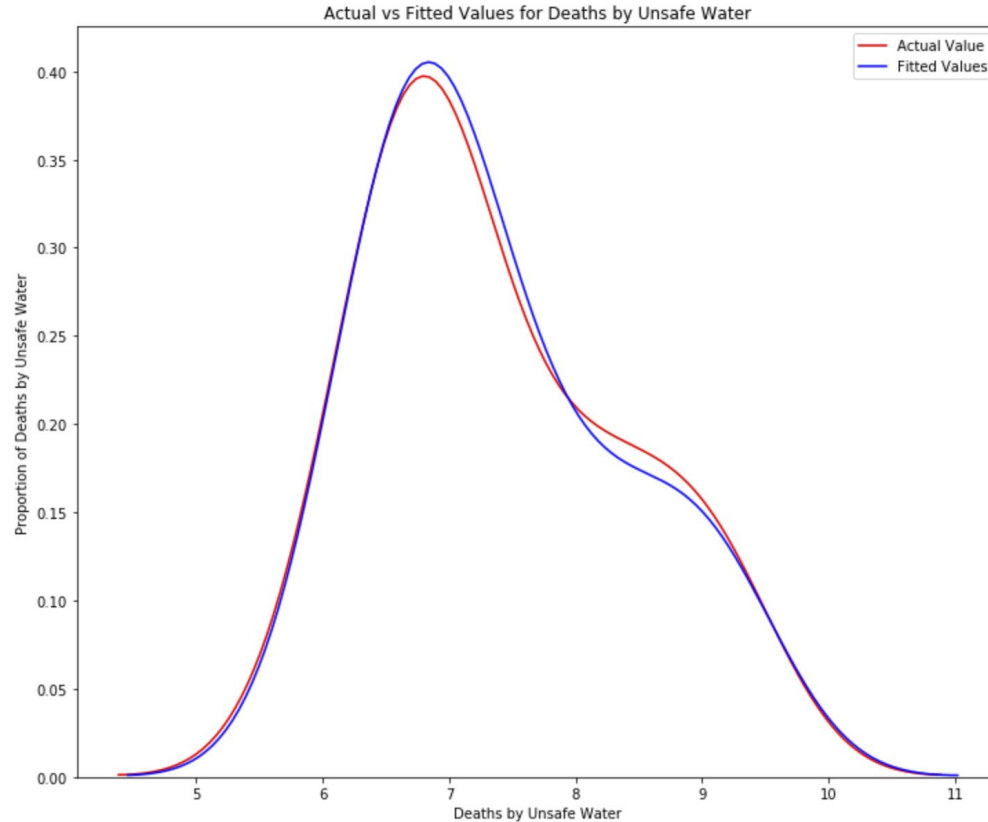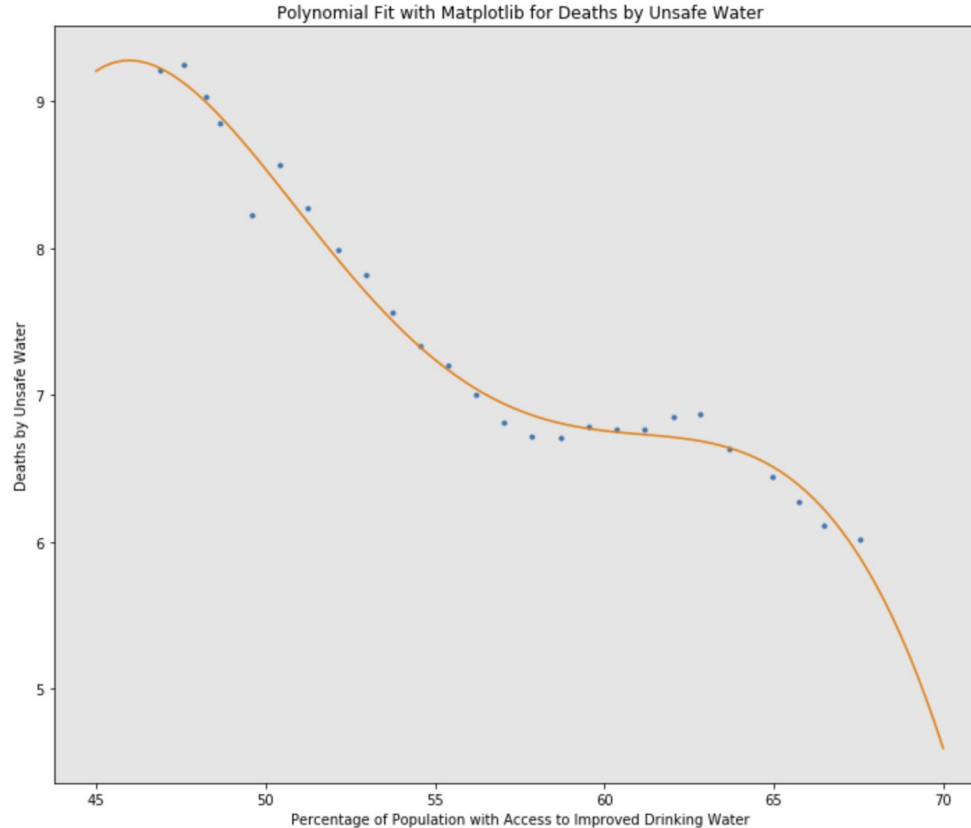
Evaluation of models

	Linear Model	Multi-Linear Model	Non-Linear Model	Polynomial Model
R-squared	0.89	0.97	0.23	0.98
MSE	0.09	0.02	0.01	0.01

The best models are depicted in red. The best R-squared is close to 1, and the best MSE is close to zero.

Conclusion & Discussion

- The best model that fit the share number of deaths due to unsafe water was the polynomial model by using access to improved water sources as the independent variable.
- The amount of data was limited but this can be solve in future work by estimating the share deaths by unsafe water for a longer time period than the 26 years used in this analysis.
- The addition of GDP data would be a great benefit to this analysis because a region can't improve their clean water resources without the financial resources.
- Surprisingly population growth was the least correlated parameter with the share deaths due to unsafe water sources.