second approach

I. Introduction

World leaders are influential people that have the power to impact millions of people's lives. Some leaders are elected for terms while others serve for life, so their expected survival is of great interest. In this paper, we seek to compare Popes, US Presidents, Dalai Lamas, Chinese Emperors, and Japanese Emperors to see how their lifespans compare. Additionally, we want to analyze the impact on survival of a leader's birth year, and whether or not this impact might vary depending on the type of leadership. After we build a model to analyze the effects, we also wish to look at some specific cases and determine the probability that the current 14th Dalai Lama will outlive Pope Francis and the probability that President Obama will outlive Emperor Naruhito. We will accomplish this by using survival analysis. We use Bayesian Inference to estimate the model parameters with the help of the JAGS program.

The rest of the paper is structured as follows. First, we will describe the data used to carry out our analysis. Next, we will describe the methods used in our analysis plan. Then, we will show how we carried out our analysis plan and recount our results. After that, we discuss our conclusions and areas for further research. Finally, our code and some additional information can be foudn in the Appendix.

II. Data

The data used in this analysis contains entries for 177 different world leaders. The types of world leaders present in the data include Popes, US Presidents, Dalai Lamas, Chinese Emperors, and Japanese Emperors. Each leader's birth date type of leader is recorded. For some of these groups, we have data dating all the way back to the 14th century. Additionally, leaders who have passed away also have their death date and age of death recorded. For leaders that are still living, these columns instead contain the date the dataset was created (July 31, 2020) and their current age on that date. In order to further clarify who is dead or alive, there is a column titled "Censored," which takes a value of 0 if the person is still dead and a value of 1 if that person is still alive. Related to this, there is another column called "Fail," which takes on a value of 1 if the person is dead and a value of 0 if the person is alive. In total, there are 11 living leaders in our dataset, 4 of whose age of death we are trying to predict.

III. Methods

III a. Motivating the Model

We use survival analysis as a way to model T_i, the lifespan of a given leader i depending on their year of birth and type of leadership. Survival analysis is useful in that it allows the consideration of "censored" data. This means that we do not always observe the outcome for each data point, for example, we do not know the death date of leaders that are still alive. Thus, we do not know their lifespan, we just know that their survival time T_i will be greater than their current age.

We model the lifespan T_i of an individual i after the Weibull distribution specified below. We choose to use the Weibull distribution because it is often utilized in survival analysis and allows the user to specify a flexible shape parameter for the distribution. The first parameter r, also known as the shape parameter, is a positive scalar, and the second parameter μ, the scale parameter, is a linear function of the covariates (birth year and types of leadership as well as their interactions).

$$T_i \sim Weibull(\alpha, \mu)$$

$$
\begin{aligned}
\log(\mu_i) = \beta_0 + \beta_1(Year\ of\ Birth_i) + \beta_2(Leadership_i = UsPres) + \beta_3(Leadership_i = ChinaEmp) \\
+\beta_4(Leadership_i = DalaiLama) + \beta_5(Leadership_i = JapanEmp) \\
+\beta_6(Year\ of\ Birth_i * (Leadership_i = UsPres)) \\
+\beta_7(Year\ of\ Birth_i * (Leadership_i = ChinaEmp)) \\
+\beta_8(Year\ of\ Birth_i * (Leadership_i = DalaiLama)) \\
+\beta_9(Year\ of\ Birth_i * (Leadership_i = JapanEmp))
\end{aligned}
$$

where Leadership_i = Popes is the baseline for comparison.

III b. Addressing Censored Data

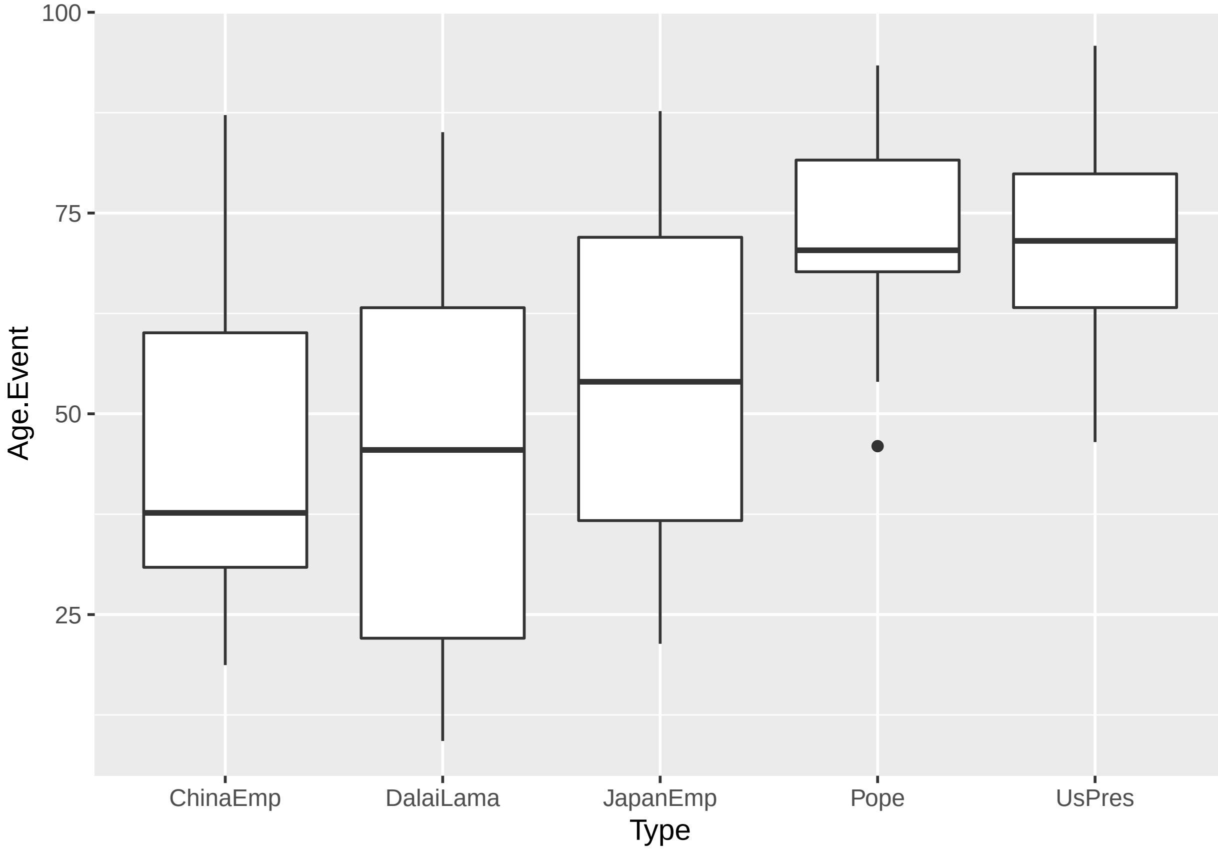
To handle censored observations, we specify their contribution to the likelihood function using the "zeroes trick." Since the sampling distribution of the censored observations is not known to be a standard distribution, we use the Poisson "zeroes trick" where the likelihood of a Poisson(phi) observation equal to zero is exp(-phi), and when observation[i] has a likelihood of L[i], then phi[i] is assigned to -log(L[i]) if our observed data is a set of 0's. This is necessary because we don't know the lifespan of currently living people, so we need a way to model their expected survival times. This method is also generally more computationally efficient (Stander et al, 2018).

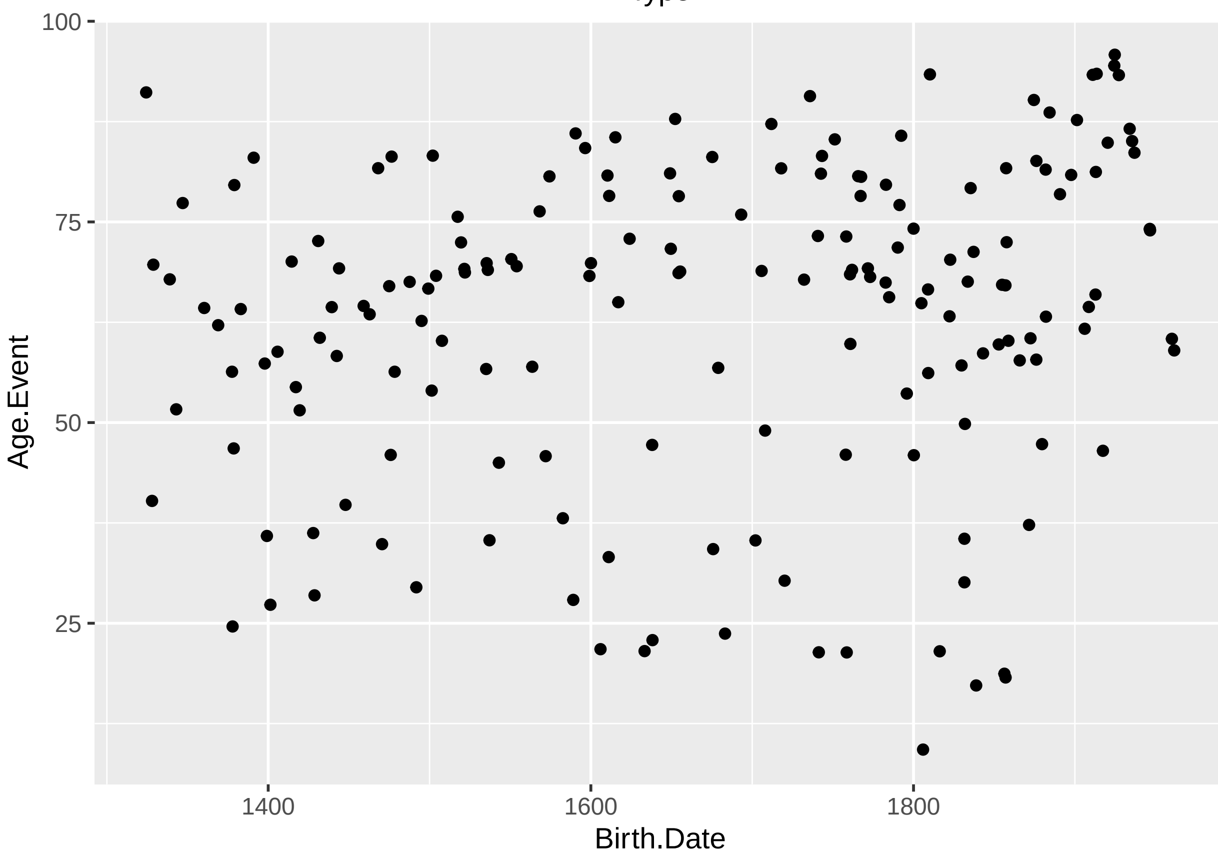
III c. Prior Choice

We assume that our priors are independent because intuitively a given observation could not be two types of leaders. Additionally, we do not have sufficiently strong prior knowledge about the relationship between birth year and type of leadership to specify an informative prior. Therefore, we use uninformative priors for the betas in our model.

For our prior for r, we chose a prior of $exp(1)$, as we believe the hazard of death increases with time.

Boxplot of age (Age.Event) by Type





Data Cleaning

III d. Model Diagnostics

The combination of traceplots, lag-1 scatterplots, and acf plots suggest the chain for each parameters converges, and that 50000 iterations is sufficient. The Rhat's are all close to 1, which is another indicator of converge. All of the effective sample sizes are greater than 200, except for that of beta_0 (the intercept) and r.

IV. Analysis

As Stander et al (2018) pointed out, following this Weibull distribution, $log(T_i)$ is equal in distribution to $\frac{1}{r}(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+...+\beta_9 x_{i1} x_{i5})+\frac{1}{r}\log(\epsilon)$ where $\epsilon \sim exp(1)$. $\alpha_j = -\beta_j/r$, $j = 1, 2, ..., 9$. The interpretation of coefficients depends on the interaction terms (i.e. both year of birth and the type of leadership).

For example, if a Pope is born one year later, he is expected to live longer by a multiplicative factor of $exp(\alpha_1)$, or his lifespan is expected to increase by a percentage of $100 * exp(\alpha_1 - 1)$. If a U.S. President is born one year later, he is expected to live longer by a multiplicative factor of $exp(\alpha_1 + \alpha_6)$; for a U.S. President and a Chinese Emperor were born in the same year y, the Chinese emperor is expected to live longer by a multiplicative factor of $exp(\alpha_3 - \alpha_2 + (\alpha_7 - \alpha_6) * y)$ and so on.

V. Results

insert model output summary here

Our model predicts that the 85-year-old 14th Dalai Lama is expected to live 92.4 years (95% credible interval: 85.4 - 99.6 years), the 84-year-old Pope Francis is expected to live 92.4 years (95% credible interval: 84.2 - 99.4 years), the 60-year-old Janpanese Emperior Naruhito is expected to live 85.3 years (95% credible interval: 66.4 - 99.5 years) and the 60-year-old former President Barack Obama is expected to live 84.8 years (95% credible interval: 61.5 - 99.4 years).

Posterior Predictive Plots (2 pairs)

Francis and Dalai Lama

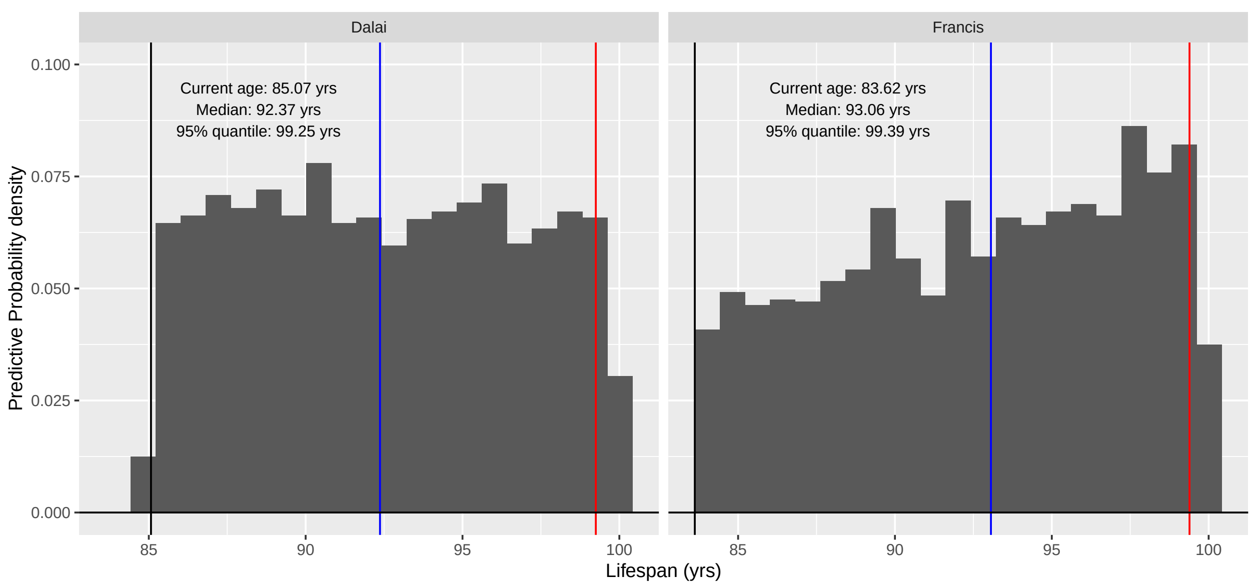


Fig. 3a The posterior predictive probability density function of the lifespan for hypothetical leaders with the same attributes as Pope Francis (right) and the 14th Dalai Lama (left). The black vertical line marks current age, the blue marks posterior median, and the red marks the posterior 95% quantile.

Obama and Naruhito

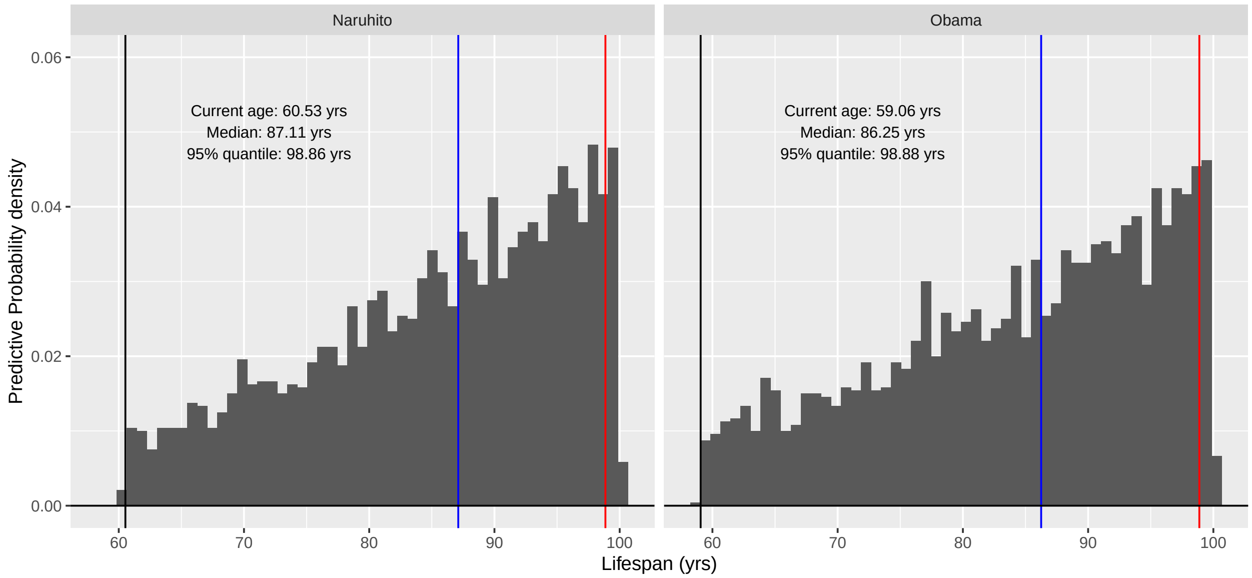
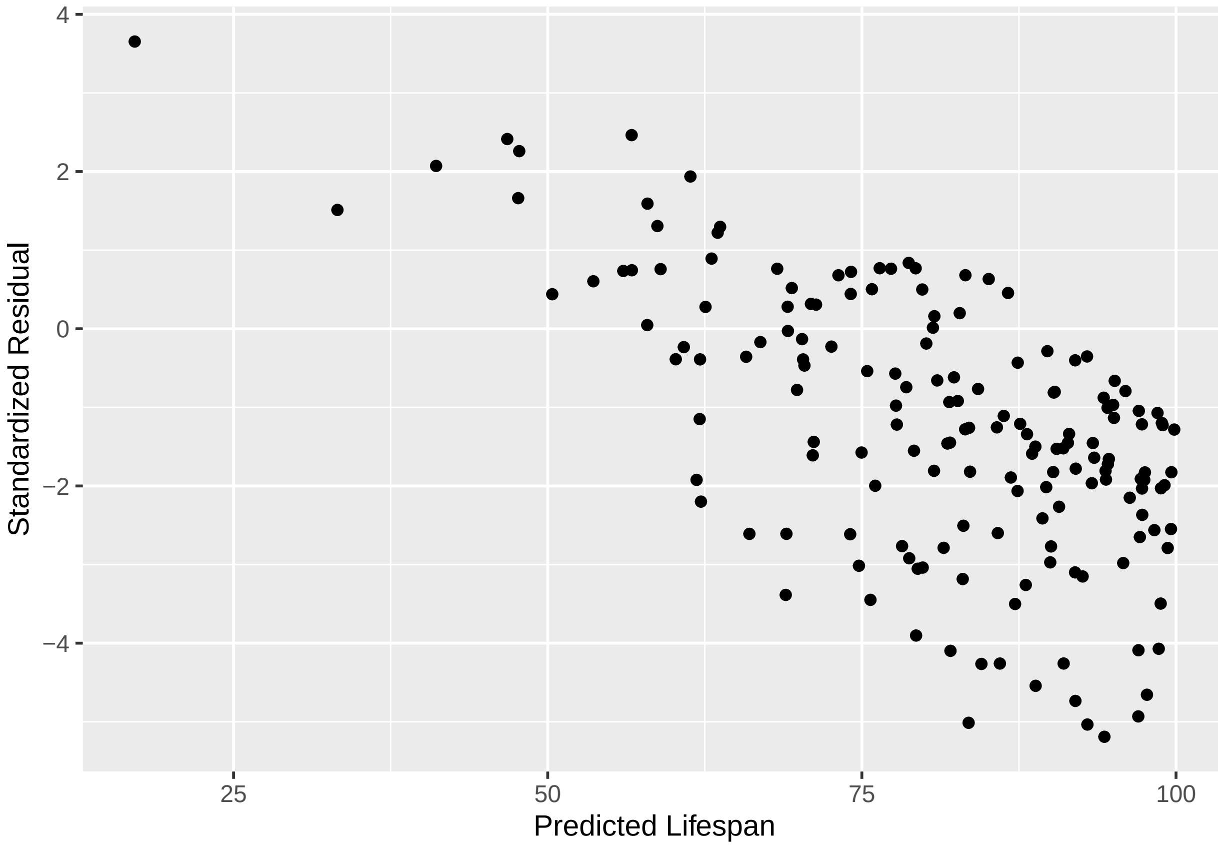


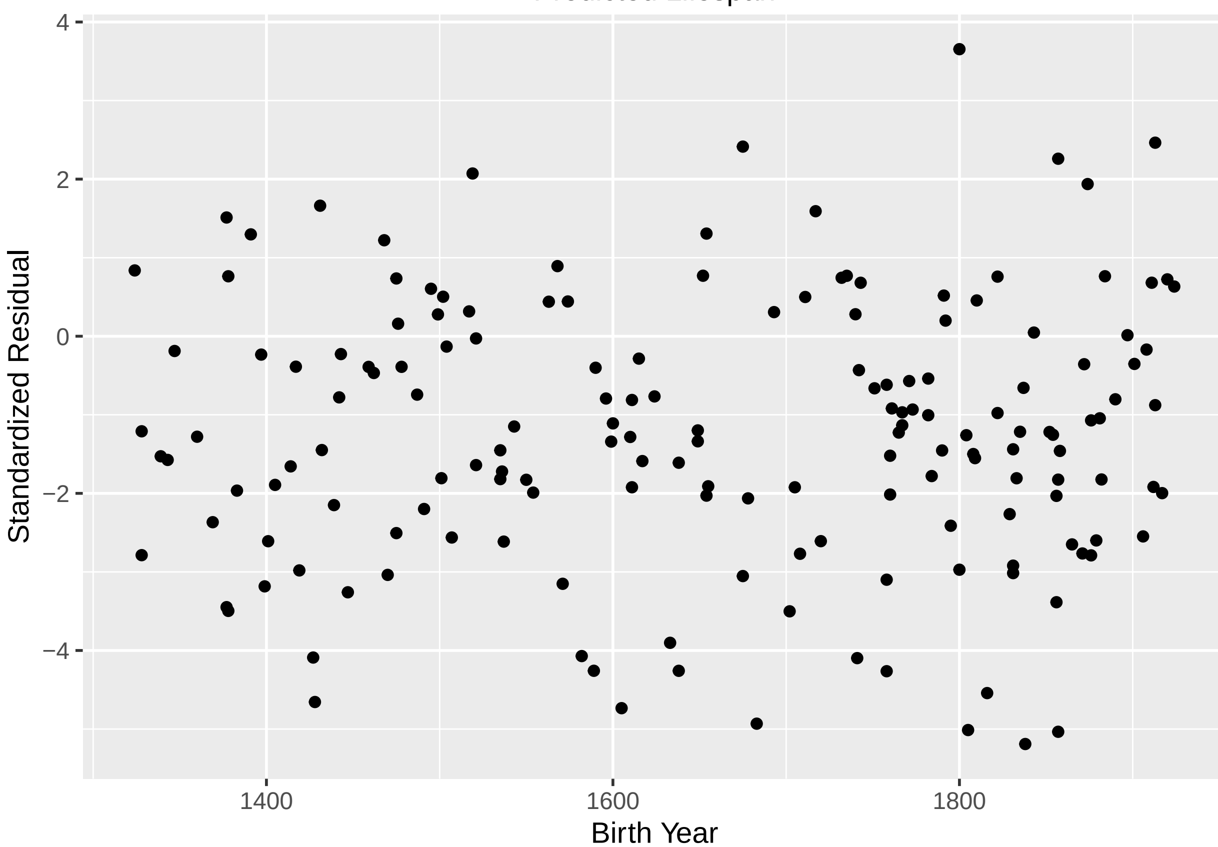
Fig. 3b The posterior predictive probability density function of the lifespan for hypothetical leaders with the same attributes as President Obama (right) and Emperor Naruhito (left). The black vertical line marks current age, the blue marks posterior median, and the red marks the posterior 95% quantile.

The histogram of the posterior predictive distributions of the lifespans of a leaders with the same attributes as Pope Francis and the 14th Dalai Lama are more uniform. The posterior predictive distributions of the lifespans of President Obama and Emperor Naruhito are both left skewed, with the modes in the late 90s.

The probability that the 14th Dalai Lama will have a longer lifespan than Pope Francis is 0.48. The probability that President Obama will have a longer lifespan than Emperor Naruhito is 0.48.

Variable	Mean	Standard Deviation	2.5% Quantile	Median	97.5% Quantile
beta_0	-22.5915879	1.4680033	-25.5221058	-22.5730566	-19.7465000
beta_1	-0.0016443	0.0007534	-0.0031519	-0.0016419	-0.0002007
beta_2	0.6059671	0.4230495	-0.2422546	0.6059599	1.4371459
beta_3	1.3355461	0.2587216	0.8319711	1.3420033	1.8255342
beta_4	1.6900917	0.3171348	1.0439544	1.7043228	2.2750221
beta_5	0.7206409	0.2366567	0.2509796	0.7212037	1.1769018
beta_6	-0.0020307	0.0024308	-0.0067542	-0.0019584	0.0026365
beta_7	0.0004541	0.0013653	-0.0022375	0.0004524	0.0031630
beta_8	0.0067854	0.0017555	0.0033960	0.0067710	0.0104489
beta_9	0.0002448	0.0012659	-0.0023251	0.0002513	0.0026518
r	4.1947291	0.2561452	3.6925186	4.1820203	4.7225935

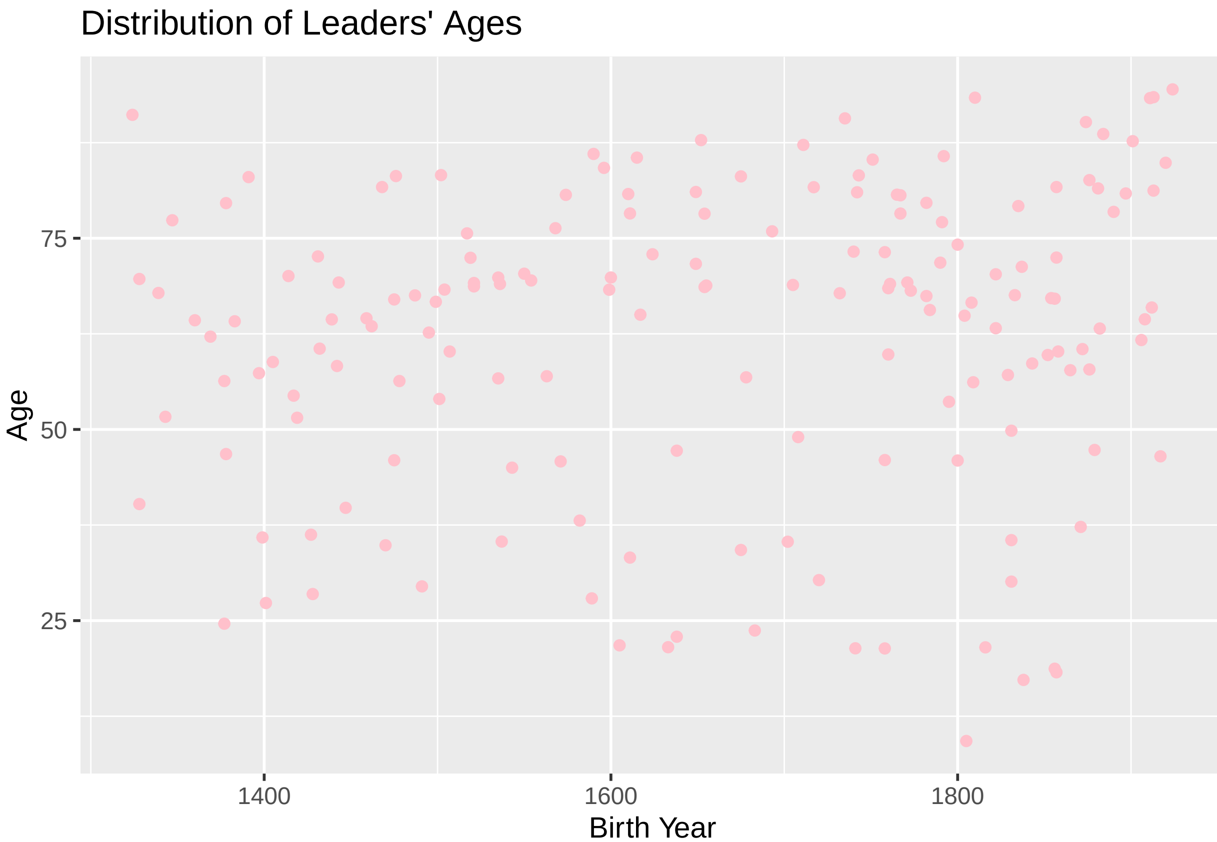

4
2
0
−2
−4
Standardized Residual
25
50
75
100
Predicted Lifespan


4
2
0
−2
−4
Standardized Residual
1400
1600
1800
Birth Year

Posterior Predictive Checks

Lexis Diagram

```
#lg <- lexis_grid(1960, 2000, 30, 80, delta=5)
#leaders_data$Initial.Age <- year(leaders_data$Event.Date) - leaders_data$Birth.Year
#subset <- filter(leaders_data, Censored == 0)
#subset <- subset[0:5,]
#ggplot() + geom_polygon(data = subset, mapping = aes(x = Birth.Year, y = Age.Event, group = Name)) + g
```



VI. Conclusion and Further Discussion

We sought to compare Popes, US Presidents, Dalai Lamas, Chinese Emperors, and Japanese Emperors to see how their lifespans compare. Our model has found that the impact of birth year does not change based on leadership type. Additionally, our model has found that lifespan does not depend on year of birth.

This model could be improved by including more predictors; for example, health widely varies among people, and could lead to a better model for prediction. For further implementations of this research question, it would be valuable to interpret economic data as it concerns the different countries that the leaders grew up in, as the conditions that a person live in lead to changes in a person's lifespan.

VII. References

cite Stander et al

VIII. Appendix