# PRELIMINARY ANALYSIS: Princess Diamond Corona Virus Risk Analysis

Code

# Introduction

This report is a preliminary analysis of the risk of infection (symptomatic and asymptomatic) and death (fatality rate or FR) from the corona virus. The analysis is based on the data derived from the Diamond Princess, a cruise ship which experienced an outbreak of the virus in February 2020. The passengers and crew were forced into a quarantine on the boat for a period of time. The quarantine had the unintended consequence of exposing the people to the virus and dramatically increasing the rates of infection. In total, there were 3,711 people on the boat, 619 became infected and to-date 7 have died.

The current date of this analysis is March 10, 2020.

The Diamond Princess data is unique in that all (or almost all) the passengers and crew were tested for the virus. As far as I know, this is the only meaningful data set where the denominator of the corona risk rates is known. However, there are still a significant number of passengers who have not recovered, or worse, are still in critical condition. Therefore, the numerator of the risk rates of death is only partially known. On the other hand, the risk rates of infection are known.

**Warning**: This analysis is a “naive” analysis, in that I do not adjust for future expected deaths. The fatality rates therefore represent a floor. I will comment on this further in the report.

All code and data can be found on Github (https://github.com/howardnewyork/corona).

# Summary of Results

- The CFR for those below age 60 is very low. The fatality rate increases from age 60 and is very high for those age 70 and above: about 1% for 70s, 8% for the 80s and 16% in the 90s. That is, mortality risk is highly skewed towards the elderly.
- The risk of infection is high across all ages. The belief that young children or young adults are not susceptible to infection is false.
- The risk of becoming symptomatic is also high across all ages.
- The risk of symptomatic and asymptomatic infection increases significantly from age 50 onwards
- Roughly half the infected population is asymptomatic.

# Methodology

Software: I used Greta (https://greta-stats.org/index.html) and Greta GP (https://github.com/greta-dev/greta.gp) to conduct a Bayesian analysis.

Models: The *basic model* is binomial model with a uniform prior assumption on the risk probability. For the youngest and very oldest ages there is very little data and this prior influences the results.

I developed a model combining all age groups and a model that analyzed each age group separately. For the latter, I used two types of models: 1. The first model assumes each age category is independent from the others. 2. The second model assumes a correlation structure between each age category. This was achieved by applying

a latent Gaussian Process over the age-dependent risk parameters, using a RBF (radial basis function) kernel. I tested two GP models.

2.1. The first fixed the $\rho$ parameter of the RBF at 10 years. That is, the correlation between age x and x+d is $e^{-(\frac{d}{\rho})^2}$. So if $d = 10$ and $\rho = 10$, the correlation between two adjacent 10-year age buckets is $e^{-(\frac{10}{10})^2} = 0.36$. If $\rho = 20$, the correlation is 0.778. The tested model fixed $\rho = 10$. 2.2. The second model, treated $\rho$ as a parameter. For this second model, I set the prior for $\rho$ to have a mean of 20 and a standard deviation of 5. The model estimated the posterior mean value of $\rho$ of 22 with a 95% credible interval of 14 to 35.

The GP Model under 2.2 represents an *advanced model* and I think gives a more accurate analysis of the fatality rates. The GP model has the advantage of smoothing out the risk rates across the age buckets and recognizing that the risk rates between age buckets is almost certainly correlated.

# Recommended Actions

Disclaimer: I am not an epidemiologist so my analysis is very much one of an armchair scientist. However, I am an actuary, so I know a thing or two about risk rates.

The analysis suggests the following:

- There will be a lot of people walking around asymptomatic but capable of spreading the disease. Social distancing will be key to avoiding the spread of the disease. Also, if face masks are helpful in stopping an infected person from spreading the diseases, then for sure, people in crowds should be wearing masks even if they feel healthy. The Surgeon General's suggestion not to wear a mask assuming they are available (which they are currently not) is ridiculous.
- The elderly need to exhibit extreme caution.
- Even if you are healthy and young, the risk of an outbreak on a cruise line is very high and the chance of being forced into an involuntary quarantine seems, to me, to be higher than I would be willing to bear.
- Likewise, avoiding large gatherings would be preferred. We are starting to see the closing of college classes and schools. I expect this to accelerate.

# Future Deaths

Based on my quick analysis, it would not surprise me if there will be additional deaths in the exposed population. Most of the infections have now occured at least 20 days ago. I would expect the fatality rates to possibly climb proportiontely by 20% to 50% (e.g. a 10% "naive" fatality rates could translate to an ultimate 12% to 15% rate), but am doubtful these ultimate fatality rates would double over the naive levels. This area still need further analysis (see Russel below for more information).

However, the analysis below is very useful in its own right, even ignoring future deaths.

The next sections go through the detail of the code and the analysis. To skip to the final results, go to the *"Summary of Results"* section at the end of the report.

# Data Sources

Russel et al (https://cmmid.github.io/topics/covid19/severity/diamond_cruise_cfr_estimates.html) National Institute of Infectious Diseases (https://www.niid.go.jp/niid/en/2019-ncov-e.html) Wikipedia (https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_on_cruise_ships#Diamond_Princess)

Note that Russel conducts a similar analysis but adjusts for outstanding deaths. My data of the age distribution differ from Russel and are based on information from Wikipedia. There are 7 deaths, 4 in the 80s, 2 in the 70s and 1 unknown. I assigned the unknown proportionately to the 80s and 70s buckets.

# Initialize Data

r r # Time Series Analysis of Survival from Breast Cancer library(greta) library(greta.gp) library(ggplot2) library(dplyr) library(purrr) library(tidyr) library(readr) library(stringr) library (bayesplot) # MCMC Pars n_samples_base = 10 warmup_base = 10 chains_base=1 # Set Directories working_dir = paste0(getwd(), /) data_dir = paste0(working_dir, /) # Load Data diamond = read_csv(file = paste0(data_dir, .csv)) head(diamond, 10)

| group <chr> | age <dbl> | symptomatic <dbl> | asymptomatic <dbl> | confirmed <dbl> | exposure <dbl> | deaths <dbl> |
|---|---|---|---|---|---|---|
| 00-09 | 5 | 0 | 1 | 1 | 16 | 0.000000 |
| 10-19 | 15 | 2 | 3 | 5 | 23 | 0.000000 |
| 20-29 | 25 | 25 | 3 | 28 | 347 | 0.000000 |
| 30-39 | 35 | 37 | 7 | 34 | 428 | 0.000000 |
| 40-49 | 45 | 19 | 8 | 27 | 334 | 0.000000 |
| 50-59 | 55 | 28 | 31 | 59 | 398 | 0.000000 |
| 60-69 | 65 | 76 | 101 | 177 | 923 | 0.000000 |
| 70-79 | 75 | 95 | 139 | 234 | 1015 | 2.333333 |
| 80-89 | 85 | 27 | 25 | 52 | 216 | 4.666667 |
| 90-99 | 95 | 2 | 0 | 2 | 11 | 0.000000 |

1-10 of 10 rows

# Infection Rate Analysis

## Greta Infection Rate Analysis

### Beta Binom Model: Single Parameter

r r # Identify the data input for Greta using the as_data function categories=1 infected = as_data(diamond $confirmed)exposure = as_data(diamond$exposure) deaths = as_data(diamond $deaths)symptomatic = as_data(diamond$symptomatic) asymptomatic = as_data(diamond$asymptomatic) # Define the prior for the theta parameter theta_infected = beta(shape1 = 1, shape2 = 1, dim = categories) theta_symptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) theta_asymptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) theta_deaths = beta(shape1 = 1, shape2 = 1, dim = categories)

# Define the likehood for the model

distribution(infected) = binomial(size = exposure, prob = theta_infected)
distribution(symptomatic) = binomial(size = exposure, prob = theta_symptomatic)
distribution(asymptomatic) = binomial(size = exposure, prob = theta_asymptomatic) distribution(deaths) = binomial(size = infected, prob = theta_deaths)

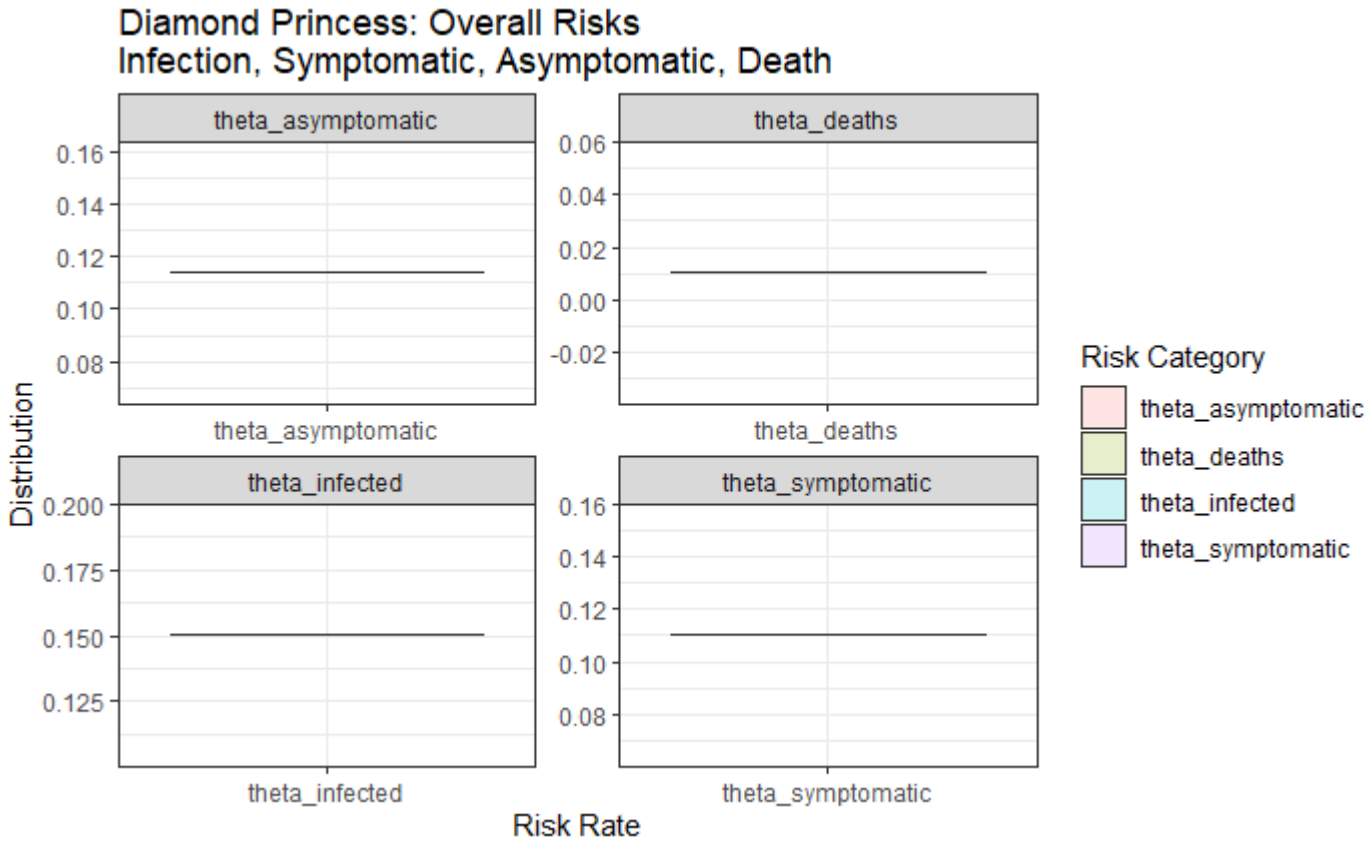
We can now establish and compile the model:

r r # Establish the model m= model(theta_infected, theta_symptomatic, theta_asymptomatic, theta_deaths) r n_samples = n_samples_base; chains=chains_base; warmup=warmup_base S = n_samples * chains # Total number of simulations draws=mcmc(m, n_samples = n_samples, warmup = warmup,chains = chains)

r r # Organize Output for Plotting theta_post = draws %>% reduce(rbind) colnames(theta_post) = paste0(rep(c(*infected*,*symptomatic*,*asymptomatic*, *deaths*), each=categories) , 1:categories) theta_post = cbind(S=1:S,as_tibble(theta_post)) %>% gather(posterior, value, -S) %>% mutate(risk_cat = as.factor(gsub(pattern = _([^_])*$,\,(posterior))), age_cat = as.factor((as.numeric(extract_numeric(posterior)))*10-5) )

```
Calling `as_tibble()` on a vector is discouraged, because the behavior is likely to change in th
e future. Use `tibble::enframe(name = NULL)` instead.
[90mThis warning is displayed once per session.[39mextract_numeric() is deprecated: please us
e readr::parse_number() instead
```

r r combined_risk_rates = ggplot(theta_post) + geom_violin(aes(risk_cat, value, fill=risk_cat), alpha=0.2) + labs(title=Princess: Overall Risks ```



## Beta Binom Model: Vector Parameter

This model assumes that each age category is independent from the others.

r r # Identify the data input for Greta using the as_data function categories = nrow(diamond) exposure = as_data(diamond$exposure) infected = as_data(diamond$confirmed) symptomatic= as_data(diamond$symptomatic) asymptomatic = as_data(diamond$asymptomatic) deaths = as_data(diamond$deaths) # Define the prior for the theta parameter theta_infected = beta(shape1 = 1, shape2 = 1, dim = categories) theta_symptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) theta_asymptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) theta_deaths = beta(shape1 = 1, shape2 = 1, dim = categories)

# Define the likehood for the model

distribution(infected) = binomial(size = exposure, prob = theta_infected, dim = categories)
distribution(symptomatic) = binomial(size = exposure, prob = theta_symptomatic, dim = categories)
distribution(asymptomatic) = binomial(size = exposure, prob = theta_asymptomatic, dim = categories)
distribution(deaths) = binomial(size = infected, prob = theta_deaths, dim = categories)

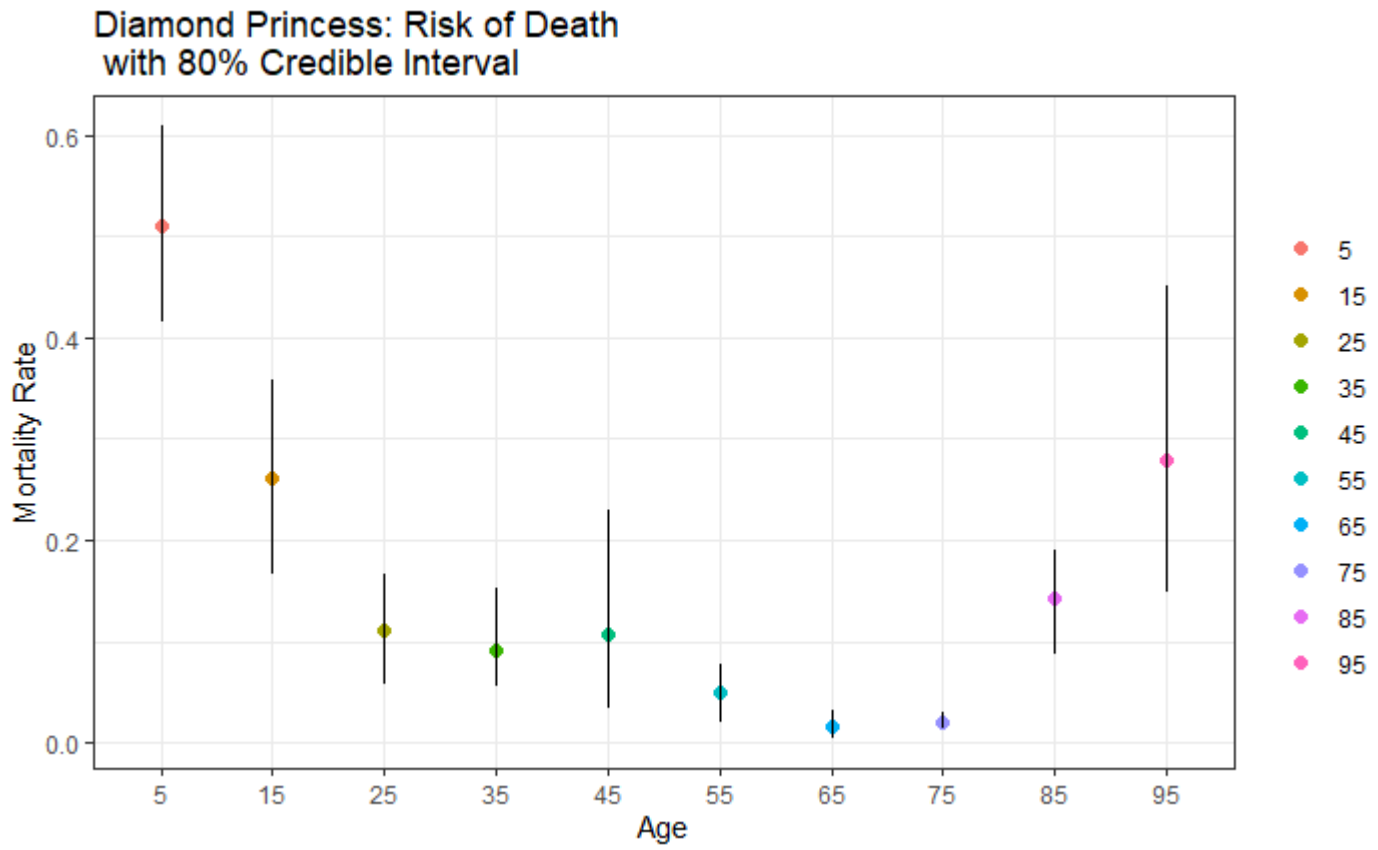
We can now establish and compile the model:

r r # Establish the model m_vec= model(theta_infected, theta_symptomatic, theta_asymptomatic, theta_deaths) r n_samples = n_samples_base; chains=chains_base; warmup=warmup_base S = n_samples * chains # Total number of simulations draws_vec=mcmc(m_vec, n_samples = n_samples, warmup = warmup,chains = chains)

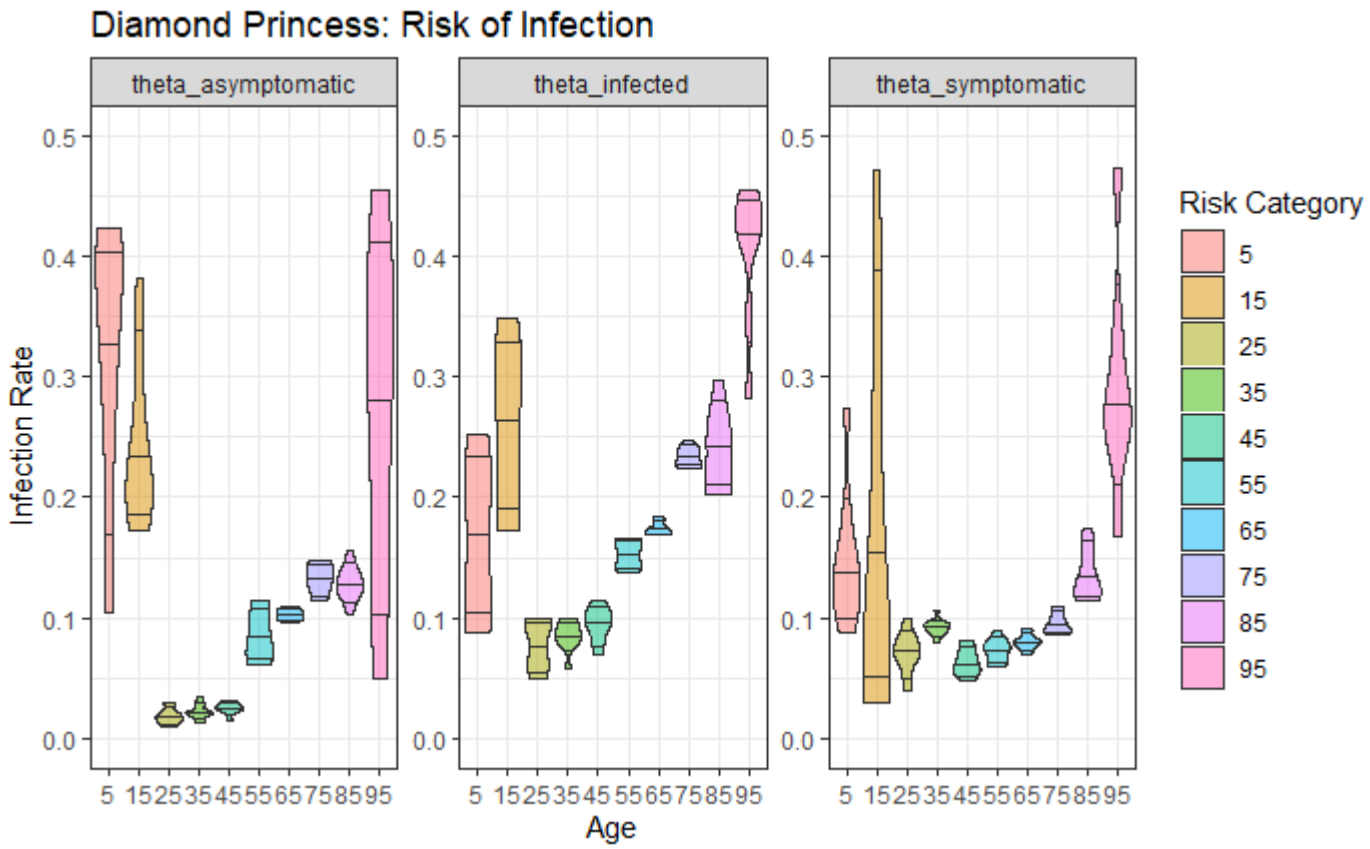
r r # Organize Output for Plotting theta_post_vec = draws_vec %>% reduce(rbind) colnames(theta_post_vec) = paste0(rep(c(*infected*,*symptomatic*,*asymptomatic*, *deaths*), each=categories) , 1:categories) theta_post_vec = cbind(S=1:S,as_tibble(theta_post_vec)) %>% gather(posterior, value, -S) %>% mutate(risk_cat = as.factor(gsub(pattern = _([^_])*$,\,(posterior))), age_cat = as.factor((as.numeric(extract_numeric(posterior)))*10-5) )

```
extract_numeric() is deprecated: please use readr::parse_number() instead
```

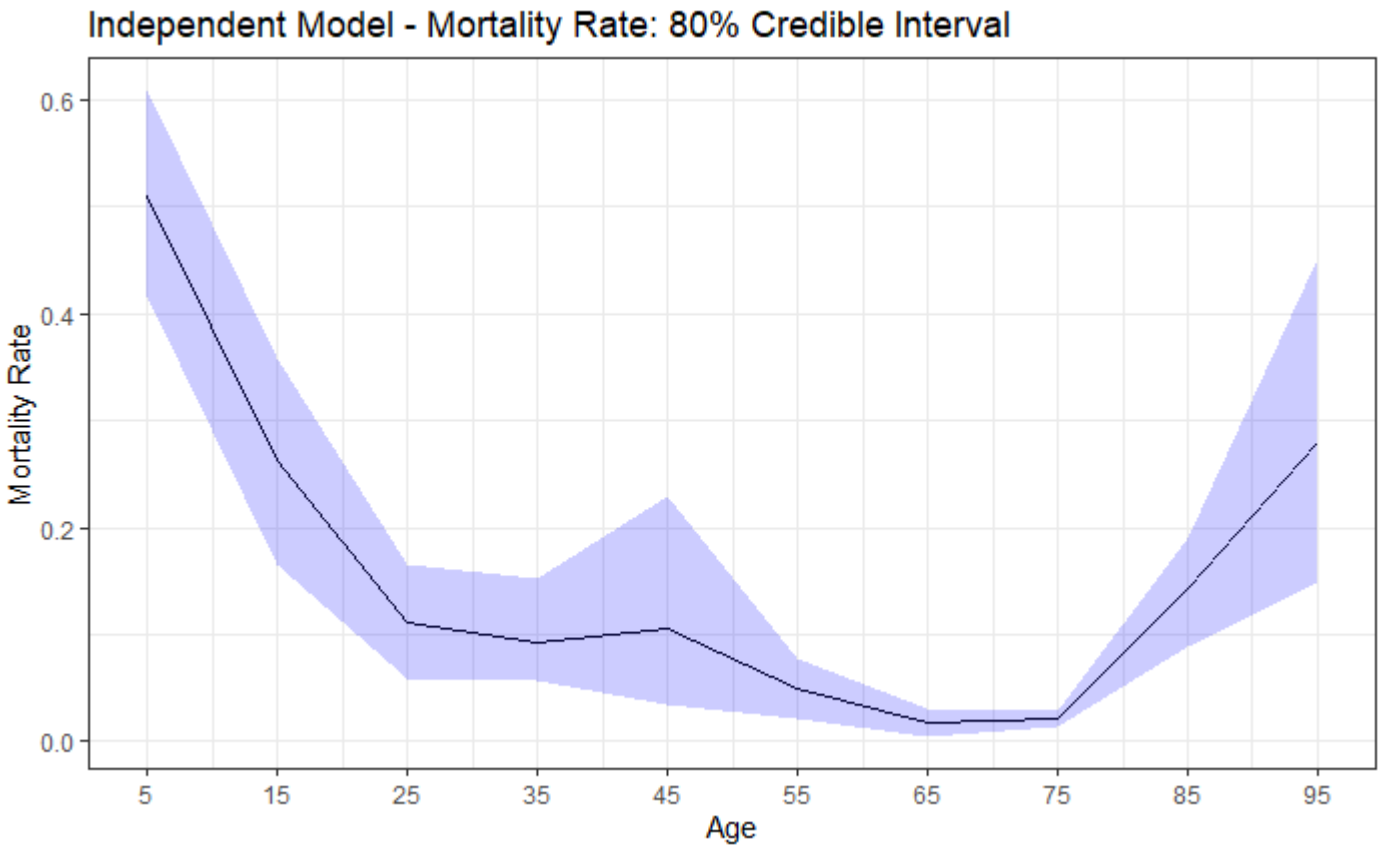
r r theta_post_vec_sum = theta_post_vec %>% filter(risk_cat == _deaths) %>% group_by(age_cat) %>% summarize(mean = mean(value), low = quantile(value, 0.1), high = quantile(value, 0.9) ) %>% gather(range, value, -mean, -age_cat) ggplot(theta_post_vec_sum) + geom_point(aes(age_cat, mean, color = age_cat),size=2) + geom_line(aes(age_cat, value, group = age_cat)) + labs(title=Princess: Risk of Deathwith 80% Credible Interval, color=\, y=Rate, x=) + theme_bw()




r r # ggplot(theta_post_vec %>% filter(risk_cat != _deaths)) + # geom_density(aes(value, fill=(age_cat)), alpha=0.2) + # labs(title=Princess: Risk of Infection, color=\) + # facet_wrap(~risk_cat, scales = )+ # theme_bw() vector_risk_rates = ggplot(theta_post_vec %>% filter(risk_cat != _deaths)) + geom_violin(aes(age_cat, value, fill = age_cat), alpha=0.5, draw_quantiles = c(0.1, 0.5, 0.9), scale = ) + labs(title=Princess: Risk of Infection, color=\, y=Rate, x=, fill=Category) + facet_wrap(~risk_cat, scales = )+ ylim(0,.5)+ theme_bw() vector_risk_rates




r r # ggplot(theta_post_vec %>% filter(risk_cat == _deaths)) + # geom_violin(aes(age_cat, value, fill = age_cat), alpha=0.5, draw_quantiles = c(0.1, 0.5, 0.9), scale = ) + # labs(title=Princess: Risk of Death from Infection, color=\, y=Rate, x=, fill=Category) + facet_wrap(~risk_cat, scales = )+ # ylim(0,.5)+ # theme_bw() risk_band = theta_post_vec %>% filter(risk_cat == _deaths) %>% group_by(age_cat) %>% summarize(mortality_mean = mean(value), mortality_low = quantile(value, probs = .1), mortality_high = quantile(value, probs = .9)) %>% mutate(age= as.numeric(as.character(age_cat))) vector_mortality_rates = ggplot(risk_band ) + geom_line(aes(age, mortality_mean)) + geom_ribbon(aes(age, ymin=mortality_low, ymax=mortality_high), fill= , alpha=.2)+ theme_bw() + labs(title=Model - Mortality Rate: 80% Credible Interval, y= Rate, x=)+ scale_x_continuous(breaks = diamond$age) vector_mortality_rates




r r risk_band

| age_cat <fctr> | mortality_mean <dbl> | mortality_low <dbl> | mortality_high <dbl> | age <dbl> |
|---|---|---|---|---|
| 5 | 0.50960685 | 0.416071076 | 0.60837158 | 5 |
| 15 | 0.26206559 | 0.165569506 | 0.35814373 | 15 |
| 25 | 0.11101261 | 0.058637669 | 0.16616018 | 25 |
| 35 | 0.09204626 | 0.056684223 | 0.15223831 | 35 |
| 45 | 0.10648767 | 0.034756807 | 0.22879223 | 45 |
| 55 | 0.05016280 | 0.020578130 | 0.07682151 | 55 |
| 65 | 0.01711897 | 0.004921642 | 0.03155555 | 65 |
| 75 | 0.02107854 | 0.014470691 | 0.02961811 | 75 |
| 85 | 0.14329560 | 0.087952770 | 0.18938715 | 85 |
| 95 | 0.27878478 | 0.147870044 | 0.45142047 | 95 |

1-10 of 10 rows

Notes: There are very limited data for the age categories 5, 15 and 95. Here the uniform plays an outsize role, the resultant posterior values for these age buckets suggests a uniform prior might not be the most optimal prior, and a prior skewed towards lower risk rates might be more appropriate.

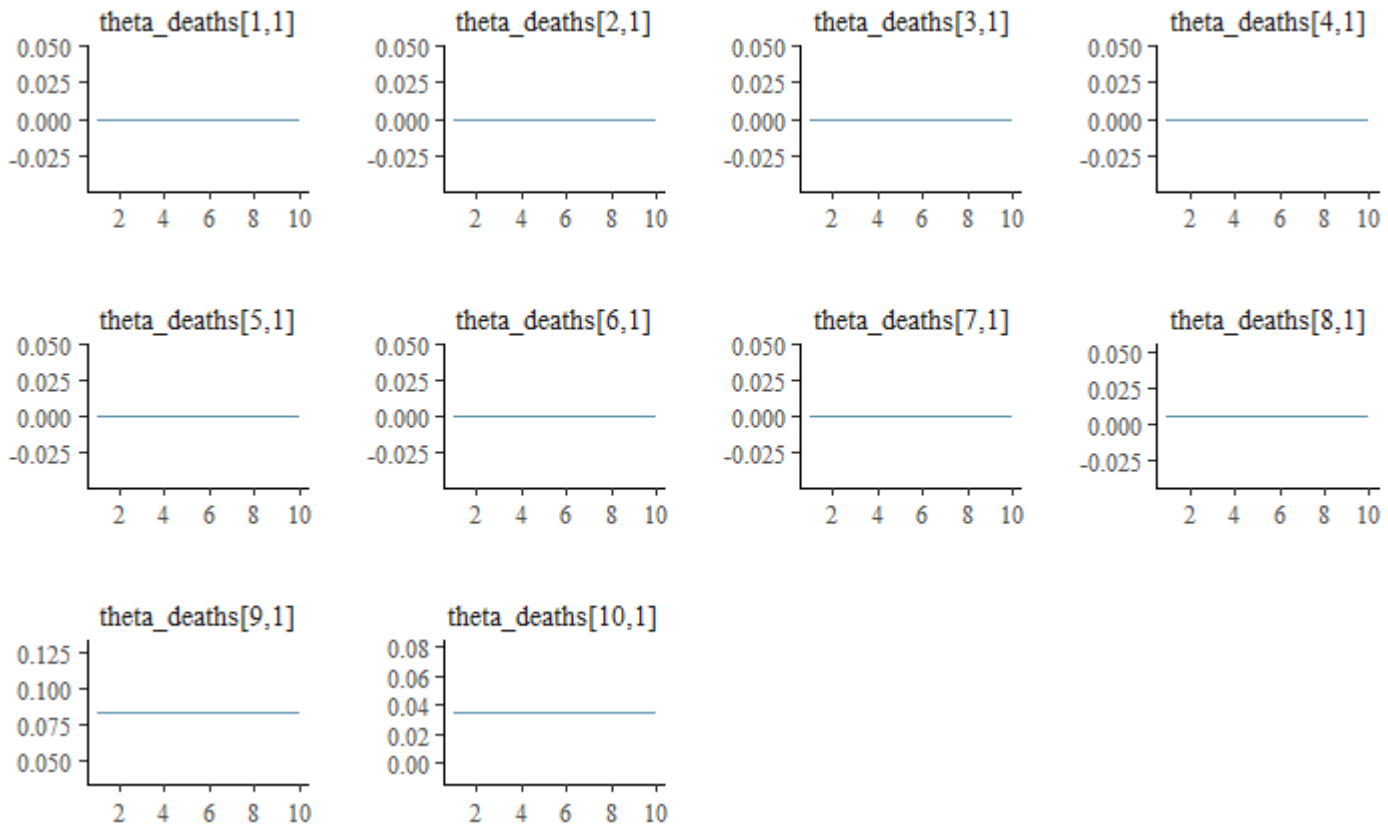
## Gaussian Process Model - Fixed Rho

r r # Identify the data input for Greta using the as_data function categories = nrow(diamond) exposure = as_data(diamond$exposure$)$ages = as_data(diamond$age) infected = as_data(diamond $confirmed$)$symptomatic = as_data(diamond$symptomatic) asymptomatic= as_data(diamond $asymptomatic$)$deaths = as_data(diamond$deaths) # Define the prior for the theta parameter #theta_infected = beta(shape1 = 1, shape2 = 1, dim = categories) #theta_symptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) #theta_asymptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) eta_deaths = lognormal(0, 20) rho_deaths = 10 # kernel & GP kernel = rbf(rho_deaths, eta_deaths) f = gp(ages, kernel) theta_deaths = ilogit(f) # convert to unit interval # Define the likehood for the model #distribution(infected) = binomial(size = exposure, prob = theta_infected, dim = categories) #distribution(symptomatic) = binomial(size = exposure, prob = theta_symptomatic, dim = categories) #distribution(asymptomatic) = binomial(size = exposure, prob = theta_asymptomatic, dim = categories) distribution(deaths) = binomial(size = infected, prob = theta_deaths, dim = categories)

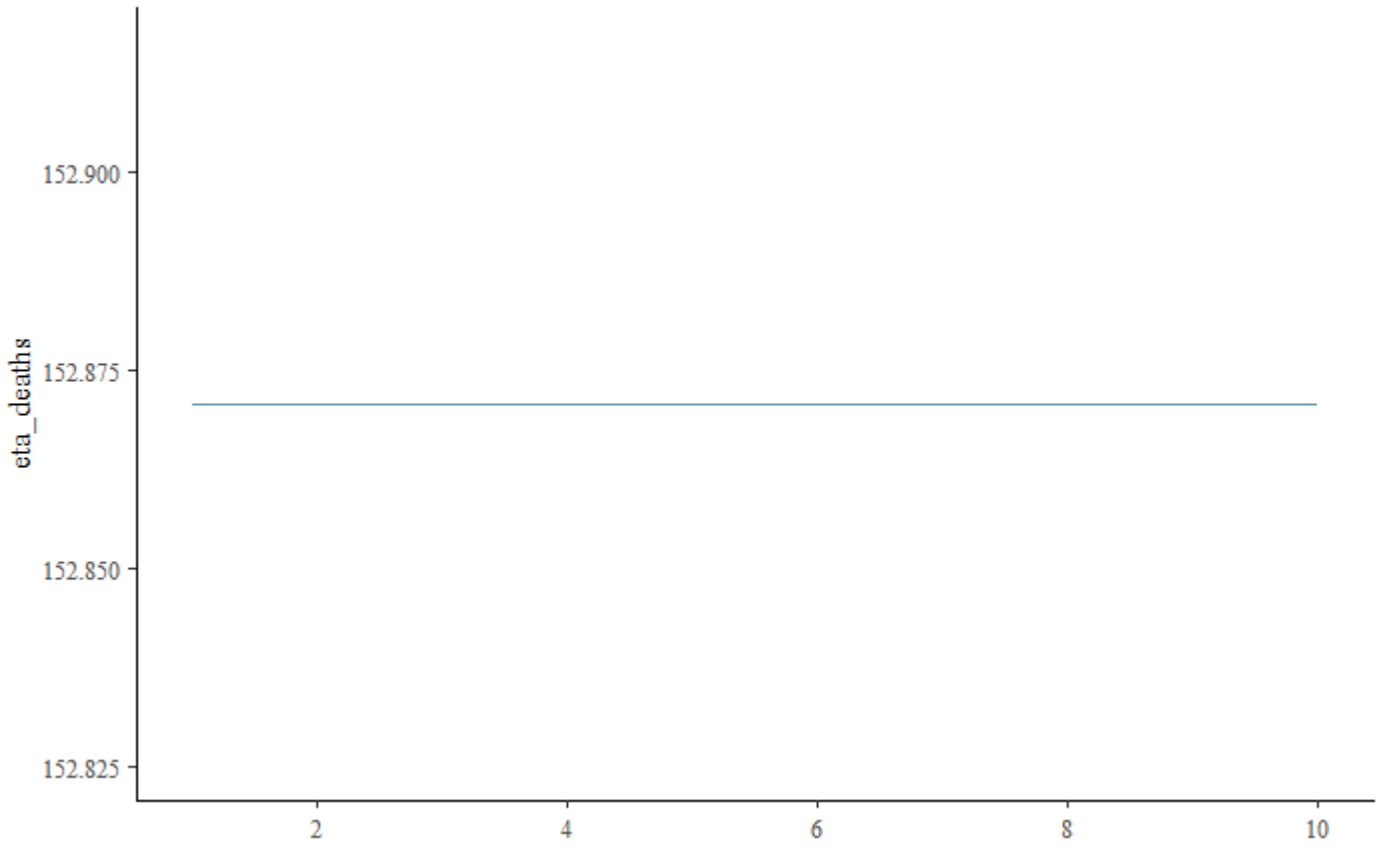
We can now establish and compile the model:

r r # Establish the model #m_gp= model(theta_deaths, eta_deaths, rho_deaths) m_gp= model(theta_deaths, eta_deaths) r n_samples = n_samples_base; chains=chains_base; warmup=warmup_base*4 S = n_samples* chains # Total number of simulations draws_gp=mcmc(m_gp, n_samples = n_samples, warmup = warmup,chains = chains)

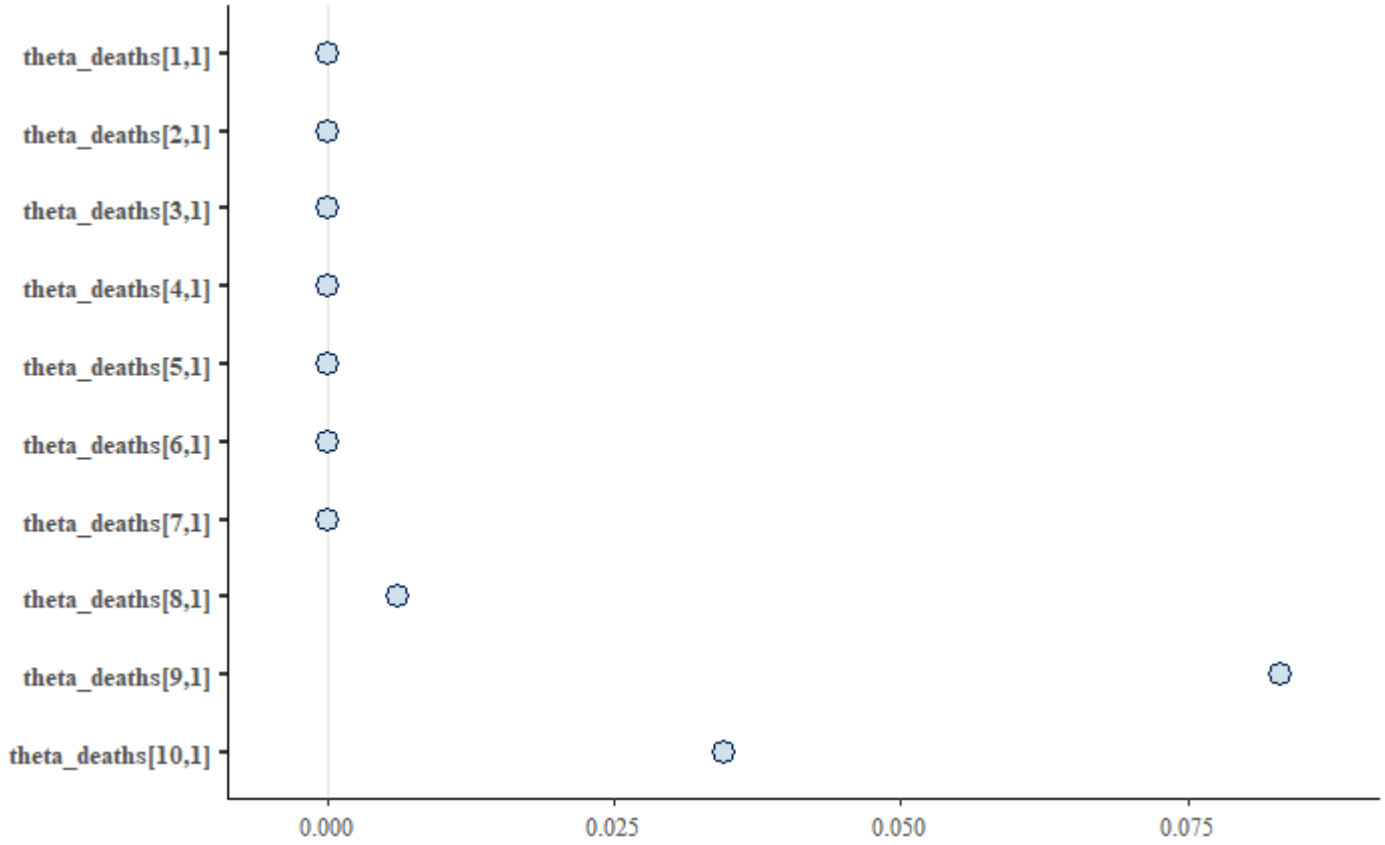
r r # diagnostics theta_names = colnames(draws_gp[[1]]) mcmc_trace(draws_gp, pars = theta_names[1:10])



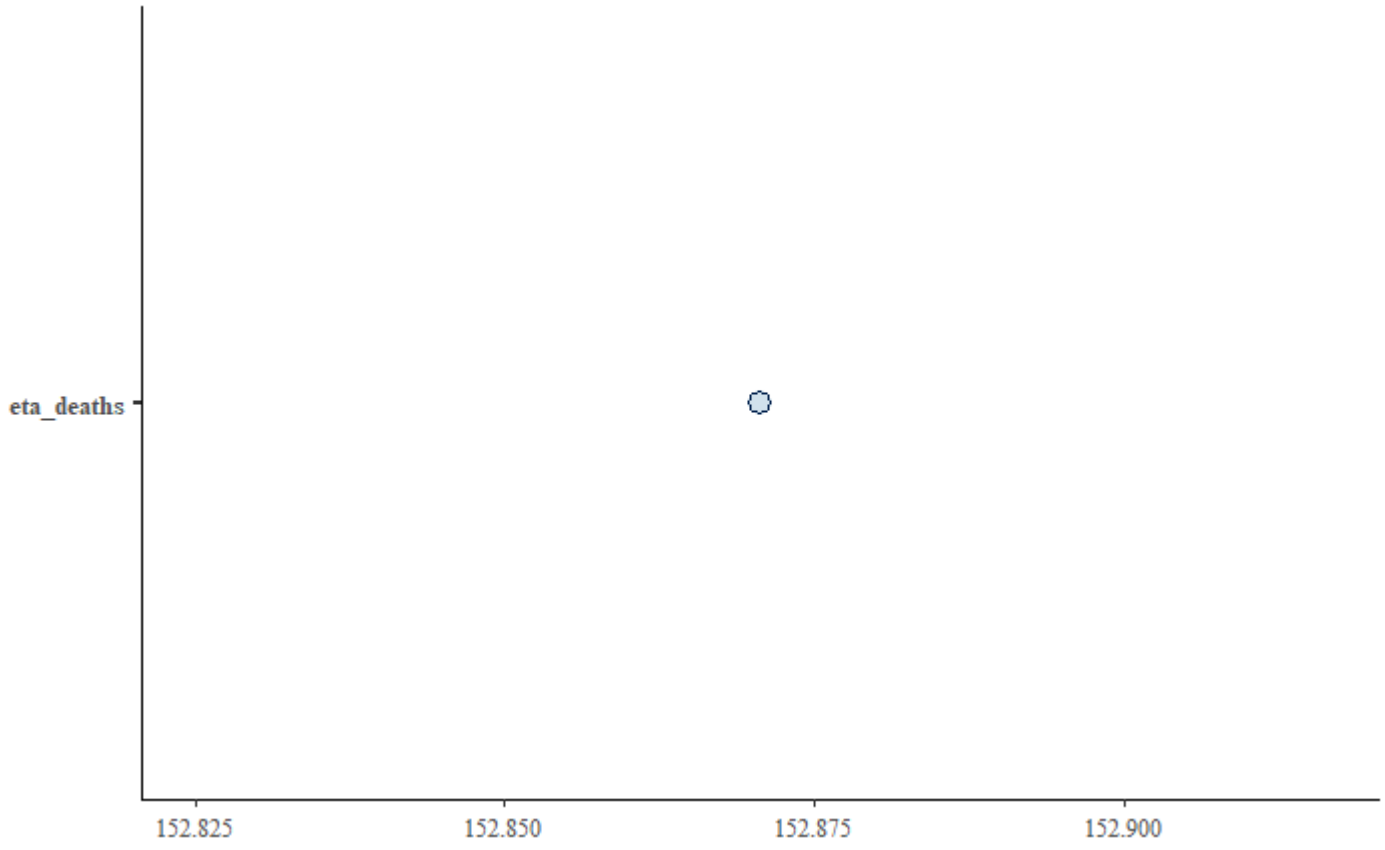
```
r r mcmc_trace(draws_gp, pars = theta_names[11:11])
```




```
r r mcmc_intervals(draws_gp, pars = theta_names[1:10])
```




r r mcmc_intervals(draws_gp, theta_names[11:11])




r r summary(draws_gp)

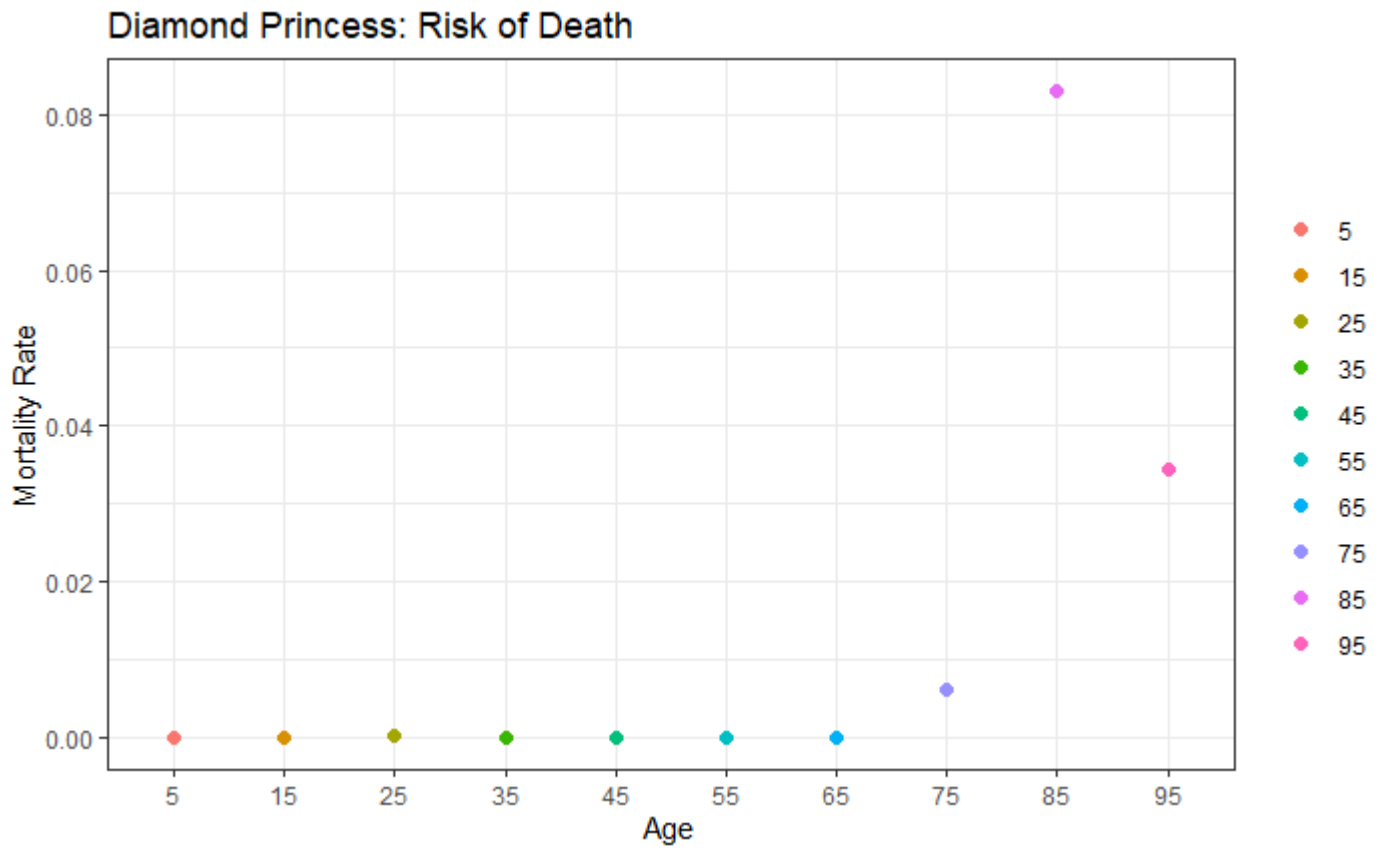
```
Iterations = 1:10
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                      Mean SD Naive SE Time-series SE
theta_deaths[1,1]  1.794e-06  0        0              0
theta_deaths[2,1]  3.678e-05  0        0              0
theta_deaths[3,1]  1.536e-04  0        0              0
theta_deaths[4,1]  1.784e-05  0        0              0
theta_deaths[5,1]  7.120e-06  0        0              0
theta_deaths[6,1]  7.904e-12  0        0              0
theta_deaths[7,1]  1.208e-09  0        0              0
theta_deaths[8,1]  6.195e-03  0        0              0
theta_deaths[9,1]  8.298e-02  0        0              0
theta_deaths[10,1] 3.455e-02  0        0              0
eta_deaths         1.529e+02  0        0              0

2. Quantiles for each variable:

                       2.5%       25%       50%       75%     97.5%
theta_deaths[1,1]  1.794e-06 1.794e-06 1.794e-06 1.794e-06 1.794e-06
theta_deaths[2,1]  3.678e-05 3.678e-05 3.678e-05 3.678e-05 3.678e-05
theta_deaths[3,1]  1.536e-04 1.536e-04 1.536e-04 1.536e-04 1.536e-04
theta_deaths[4,1]  1.784e-05 1.784e-05 1.784e-05 1.784e-05 1.784e-05
theta_deaths[5,1]  7.120e-06 7.120e-06 7.120e-06 7.120e-06 7.120e-06
theta_deaths[6,1]  7.904e-12 7.904e-12 7.904e-12 7.904e-12 7.904e-12
theta_deaths[7,1]  1.208e-09 1.208e-09 1.208e-09 1.208e-09 1.208e-09
theta_deaths[8,1]  6.195e-03 6.195e-03 6.195e-03 6.195e-03 6.195e-03
theta_deaths[9,1]  8.298e-02 8.298e-02 8.298e-02 8.298e-02 8.298e-02
theta_deaths[10,1] 3.455e-02 3.455e-02 3.455e-02 3.455e-02 3.455e-02
eta_deaths         1.529e+02 1.529e+02 1.529e+02 1.529e+02 1.529e+02
```

r r theta_post_gp = draws_gp %>% reduce(rbind) %>% as_tibble() %>% select(-eta_deaths)

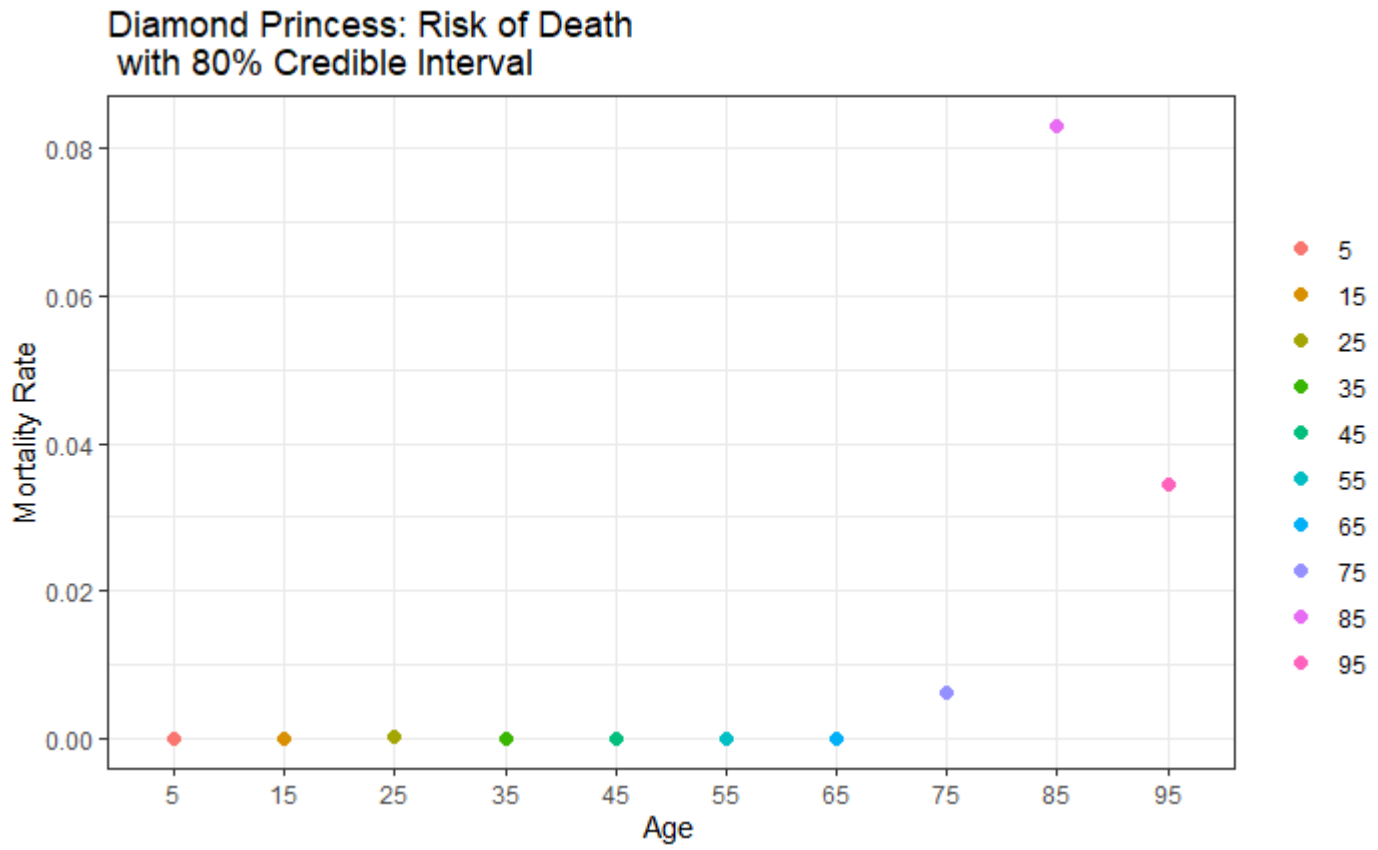
colnames(theta_post_gp) = paste0(rep(c( *deaths*), each=categories) , 1:categories) theta_post_gp = cbind(S=1:S,as_tibble(theta_post_gp)) %>% gather(posterior, value, -S) %>% mutate(risk_cat = as.factor(gsub(pattern = _([^_]*)$,\,(posterior))), age_cat = as.factor((as.numeric(extract_numeric(posterior)))*10-5))

```
extract_numeric() is deprecated: please use readr::parse_number() instead
```

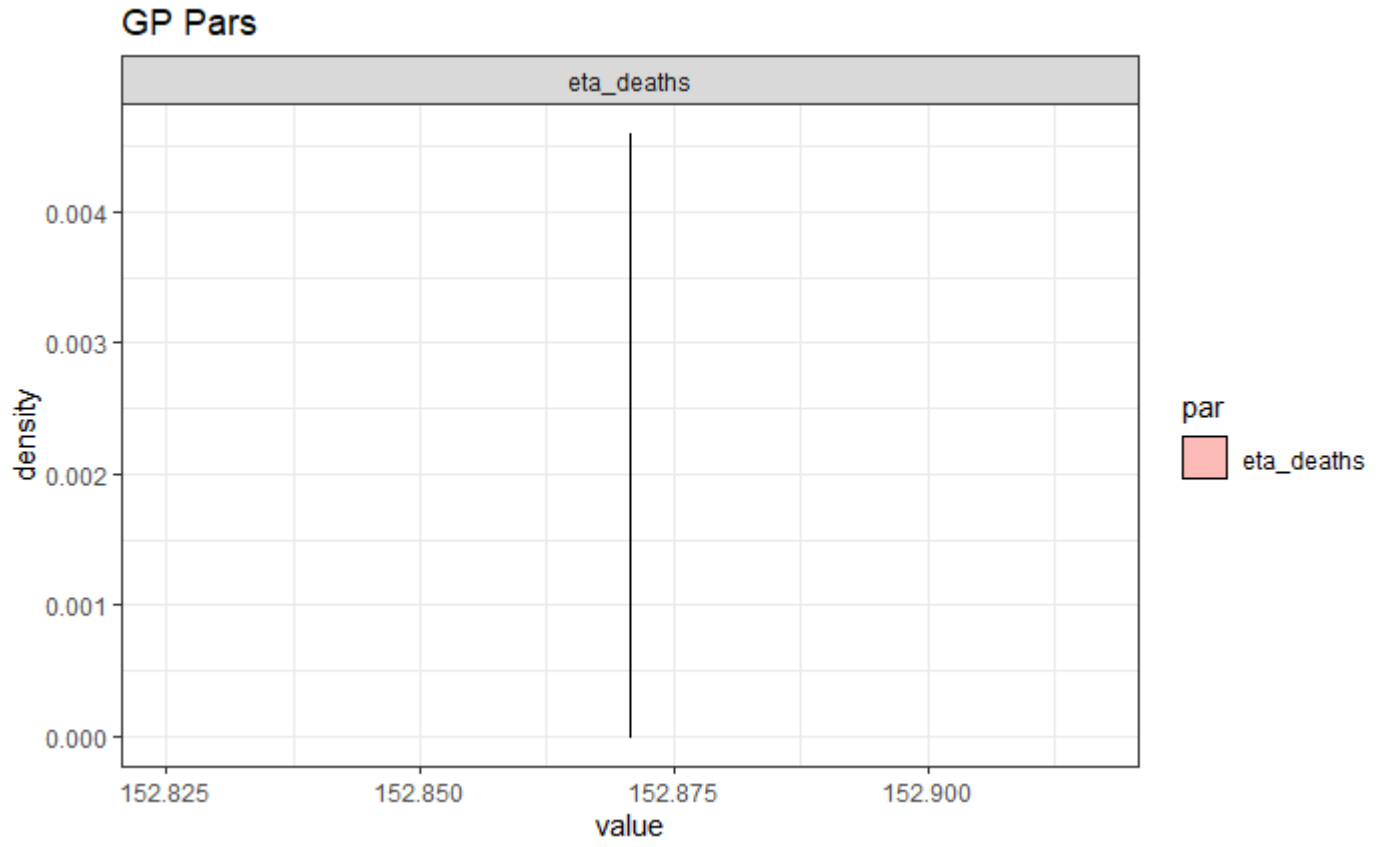
r r theta_post_gp_sum = theta_post_gp %>% group_by(age_cat) %>% summarize(mean = mean(value), low = quantile(value, 0.1), high = quantile(value, 0.9) ) %>% gather(range, value, -mean, -age_cat) ggplot(theta_post_gp_sum) + geom_point(aes(age_cat, mean, color = age_cat),size=2) + geom_line(aes(age_cat, value, group = age_cat)) + labs(title=Princess: Risk of Death, color=\, y=Rate, x=) + theme_bw()




r r # ggplot(theta_post_gp ) + # geom_violin(aes(age_cat, value, fill = age_cat), alpha=0.5, draw_quantiles = c(0.1, 0.5, 0.9), scale = ) + # labs(title=Princess: Risk of Death, color=\, y=Rate, x=) + # ylim(0,.1)+ # theme_bw() ggplot(theta_post_gp_sum) + geom_point(aes(age_cat, mean, color = age_cat),size=2) + geom_line(aes(age_cat, value, group = age_cat)) + labs(title=Princess: Risk of Deathwith 80% Credible Interval, color=\, y=Rate, x=) + theme_bw()

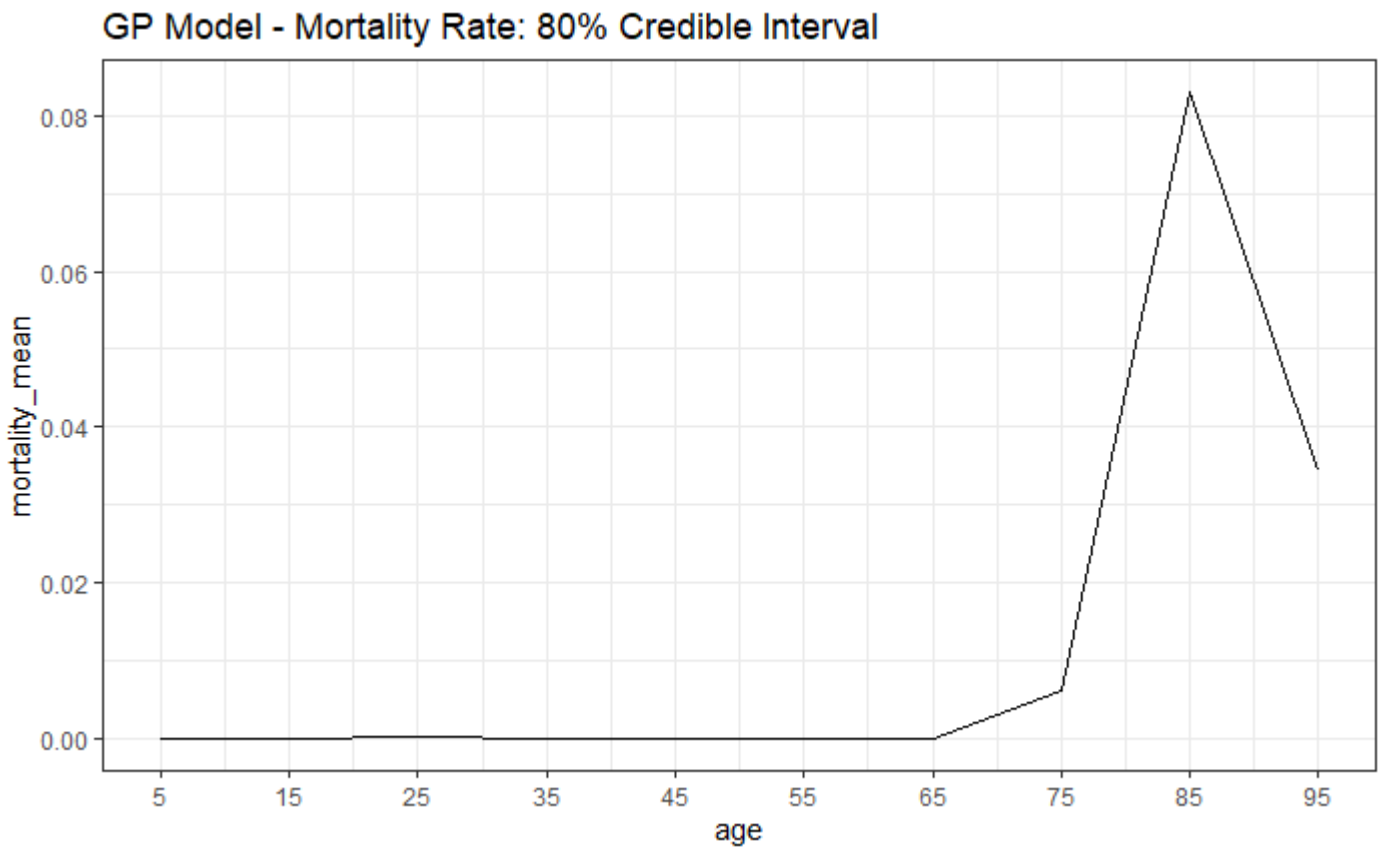



r r gp_par = draws_gp %>% reduce(rbind) %>% as_tibble() %>% select(eta_deaths) %>% gather(par, value) ggplot(gp_par ) + geom_density(aes(value, fill=par), alpha=0.5) + labs(title=Pars) + facet_wrap(~par, scales = )+ theme_bw()



Summary Statistics

r r risk_band_dependent = theta_post_gp %>% group_by(age_cat) %>% summarize(mortality_mean = mean(value), mortality_low = quantile(value, probs = .1), mortality_high = quantile(value, probs = .9)) gp_mortality_fixed_rho = ggplot(risk_band_dependent %>% mutate(age= as.numeric(as.character(age_cat)))) + geom_line(aes(age, mortality_mean)) + geom_ribbon(aes(age, ymin=mortality_low, ymax=mortality_high), fill= , alpha=.2)+ theme_bw() + labs(title=Model - Mortality Rate: 80% Credible Interval)+ scale_x_continuous(breaks = diamond$age) gp_mortality_fixed_rho




r r risk_band_dependent

| age_cat <fctr> | mortality_mean <dbl> | mortality_low <dbl> | mortality_high <dbl> |
|---|---|---|---|
| 5 | 1.794022e-06 | 1.794022e-06 | 1.794022e-06 |
| 15 | 3.678201e-05 | 3.678201e-05 | 3.678201e-05 |
| 25 | 1.536128e-04 | 1.536128e-04 | 1.536128e-04 |
| 35 | 1.783936e-05 | 1.783936e-05 | 1.783936e-05 |
| 45 | 7.120140e-06 | 7.120140e-06 | 7.120140e-06 |
| 55 | 7.903771e-12 | 7.903771e-12 | 7.903771e-12 |
| 65 | 1.208179e-09 | 1.208179e-09 | 1.208179e-09 |

| age_cat <fctr> | mortality_mean <dbl> | mortality_low <dbl> | mortality_high <dbl> |
|---|---|---|---|
| 75 | 6.195068e-03 | 6.195068e-03 | 6.195068e-03 |
| 85 | 8.297854e-02 | 8.297854e-02 | 8.297854e-02 |
| 95 | 3.455170e-02 | 3.455170e-02 | 3.455170e-02 |

1-10 of 10 rows

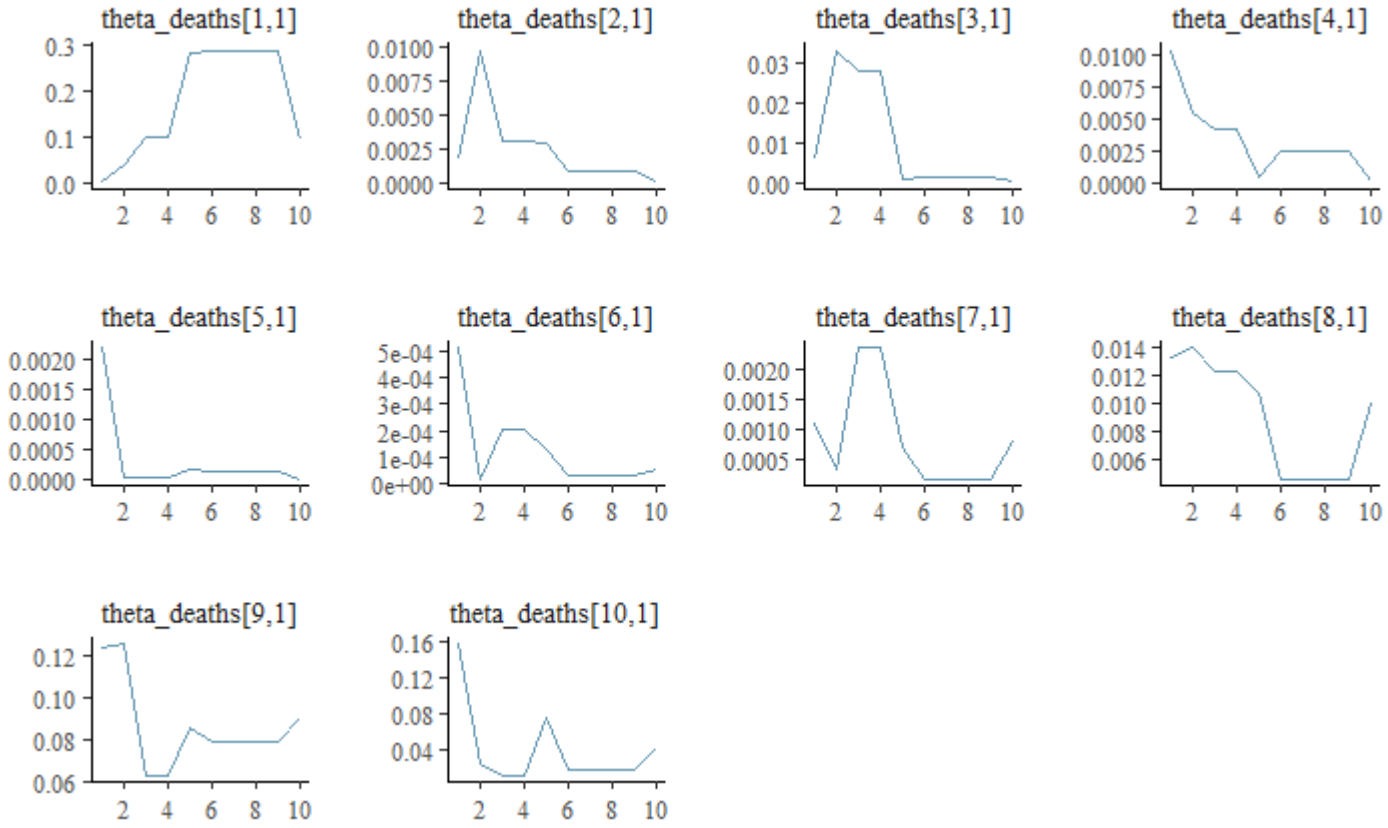
## Gaussian Process Model - Variable Rho

r r # Identify the data input for Greta using the as_data function categories = nrow(diamond) exposure = as_data(diamond$exposure$)$ages $= as_data(diamond$age) infected = as_data(diamond $confirmed)symptomatic = as_data(diamond$symptomatic) asymptomatic= as_data(diamond $asymptomatic)deaths = as_data(diamond$deaths) # Define the prior for the theta parameter #theta_infected = beta(shape1 = 1, shape2 = 1, dim = categories) #theta_symptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) #theta_asymptomatic = beta(shape1 = 1, shape2 = 1, dim = categories) eta_deaths = lognormal(0, 20) rho_mean = 20; rho_sd = 5 rho_deaths = lognormal(log(rho_mean/sqrt(1+rho_sd$^{2/rho_mean}$2)), sdlog = sqrt(log(1+rho_sd$^{2/rho_mean}$2))) # kernel & GP kernel = rbf(rho_deaths, eta_deaths) f = gp(ages, kernel) theta_deaths = ilogit(f) # convert to unit interval # Define the likehood for the model #distribution(infected) = binomial(size = exposure, prob = theta_infected, dim = categories) #distribution(symptomatic) = binomial(size = exposure, prob = theta_symptomatic, dim = categories) #distribution(asymptomatic) = binomial(size = exposure, prob = theta_asymptomatic, dim = categories) distribution(deaths) = binomial(size = infected, prob = theta_deaths, dim = categories)

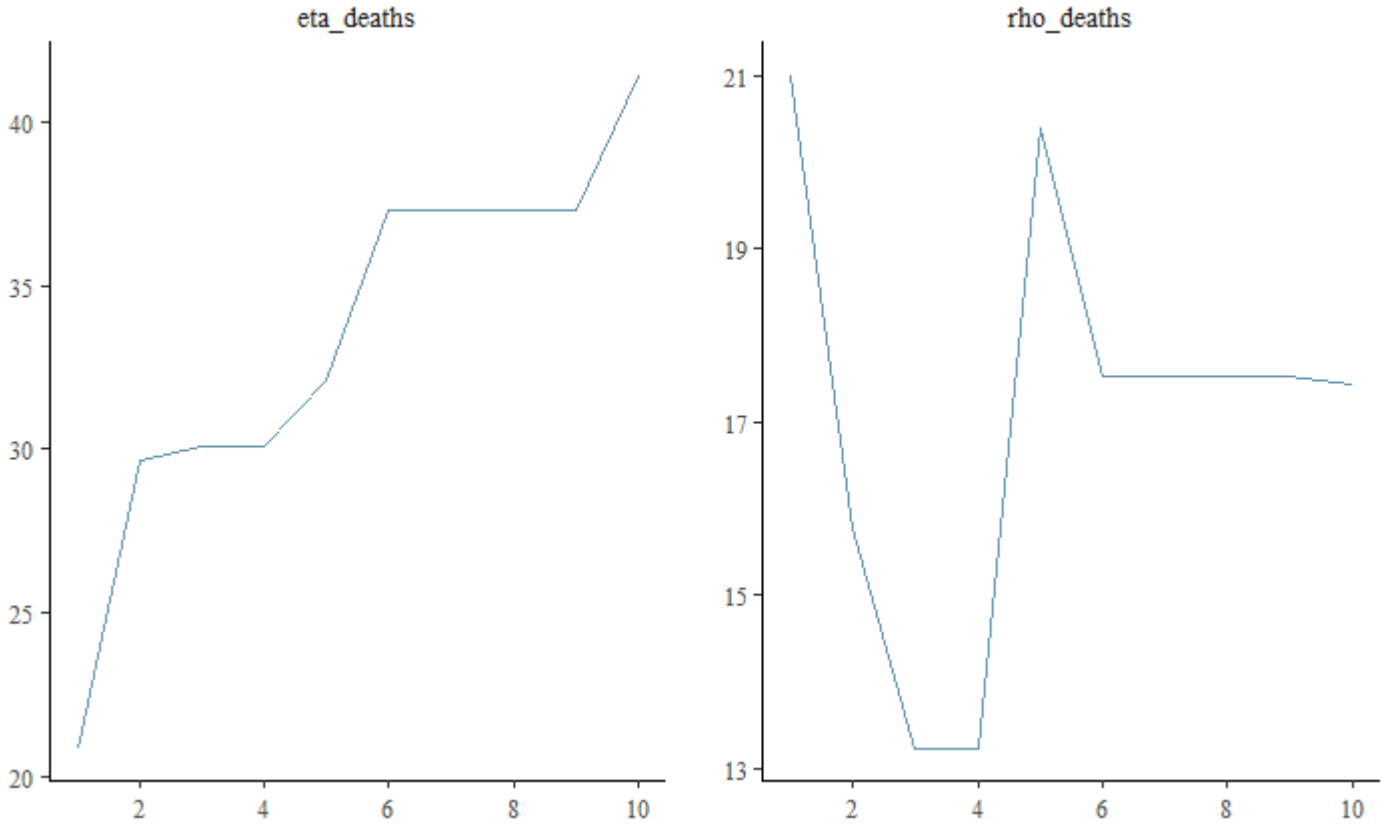
We can now establish and compile the model:

r r # Establish the model #m_gp= model(theta_deaths, eta_deaths, rho_deaths) m_gp_var= model(theta_deaths, eta_deaths, rho_deaths) r n_samples = n_samples_base; chains=chains_base; warmup=warmup_base*4 S* = *n_samples* chains # Total number of simulations draws_gp_var=mcmc(m_gp_var, n_samples = n_samples, warmup = warmup,chains = chains)

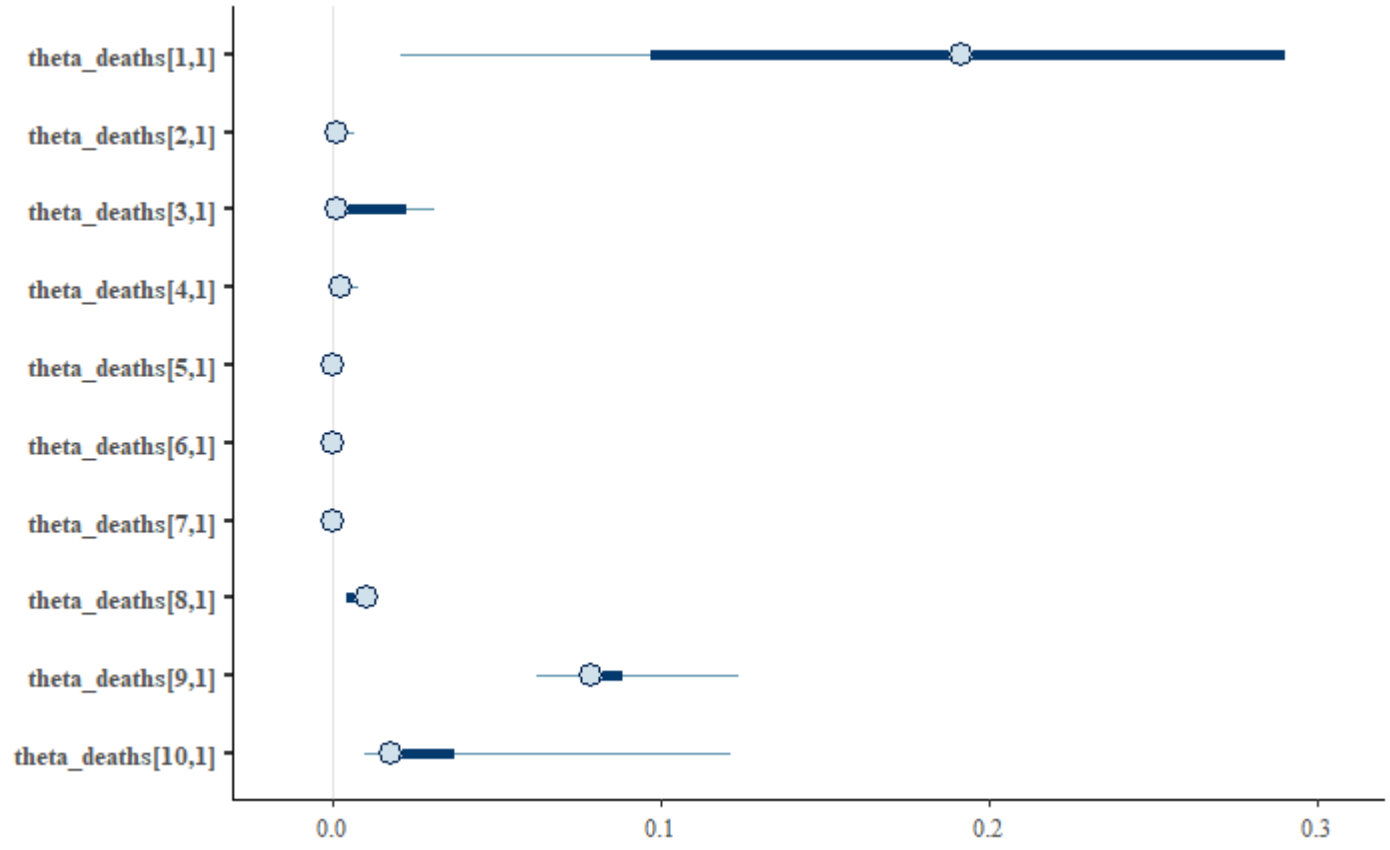
r r # diagnostics theta_names = colnames(draws_gp_var[[1]]) mcmc_trace(draws_gp_var, pars = theta_names[1:10])



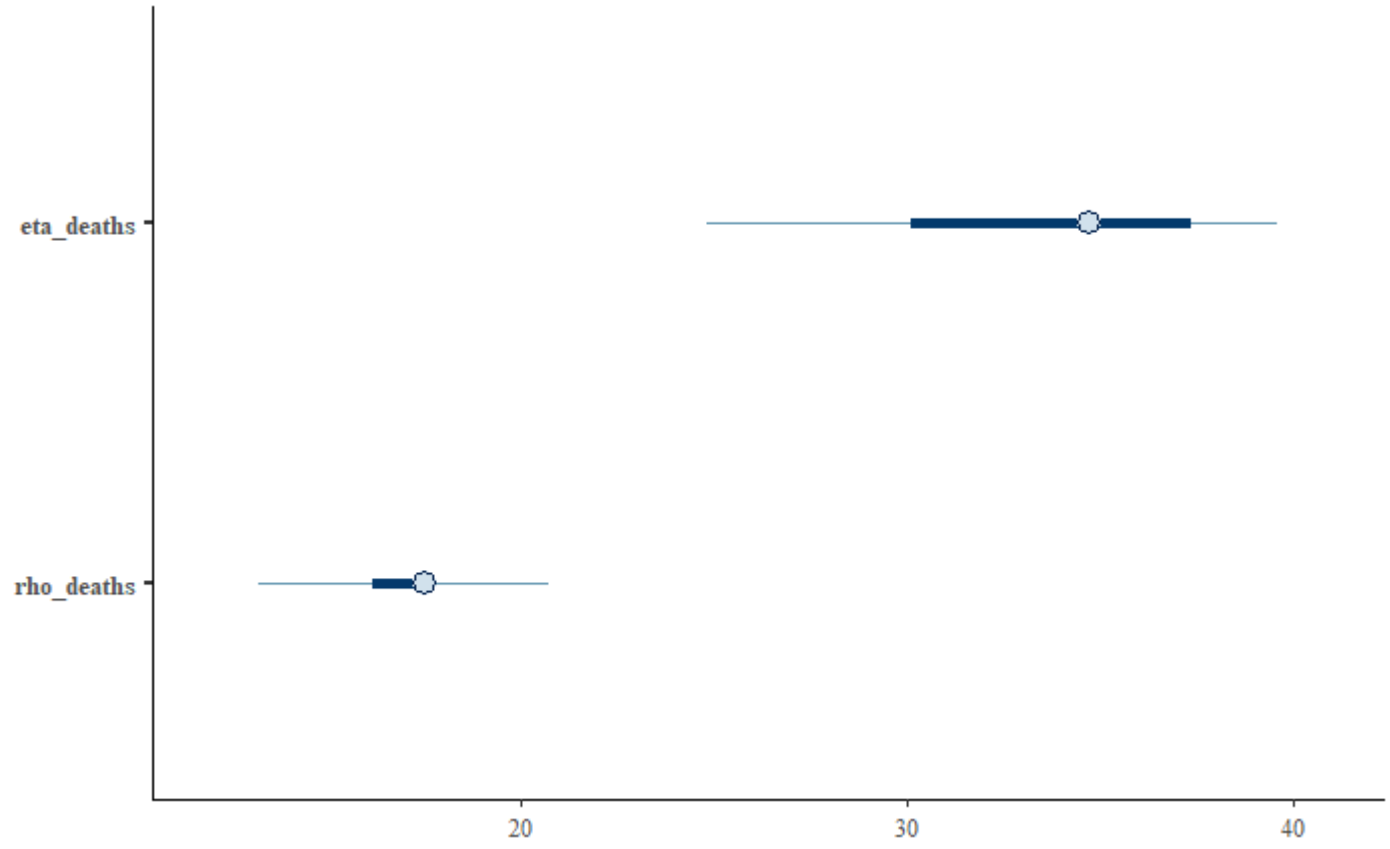

r r mcmc_trace(draws_gp_var, pars = theta_names[11:12])




r r mcmc_intervals(draws_gp_var, pars = theta_names[1:10])




r r mcmc_intervals(draws_gp_var, theta_names[11:12])



r r summary(draws_gp_var)

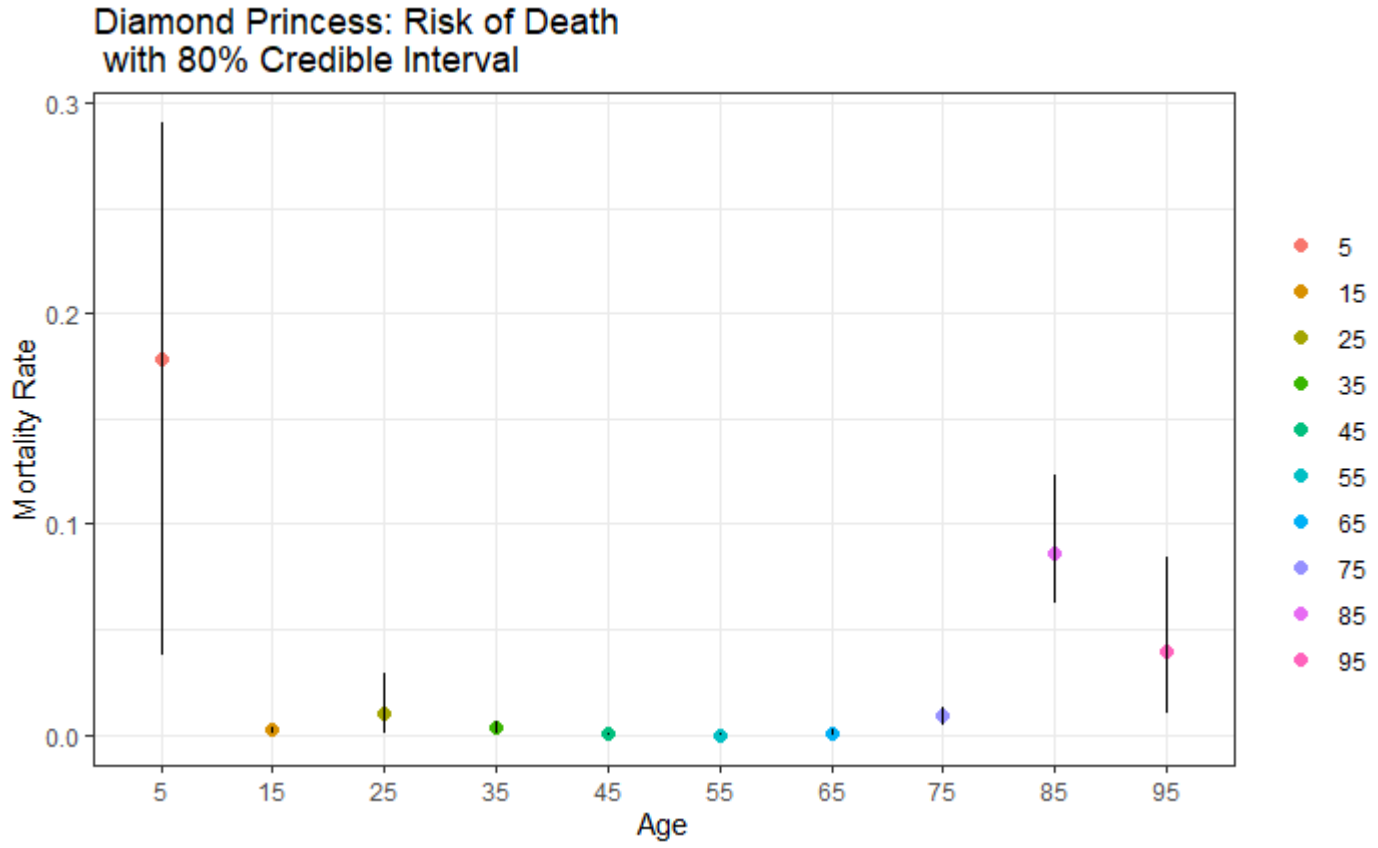
```
Iterations = 1:10
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                      Mean        SD  Naive SE Time-series SE
theta_deaths[1,1]  1.785e-01 0.1196552 3.784e-02      7.574e-02
theta_deaths[2,1]  2.394e-03 0.0028298 8.949e-04      8.949e-04
theta_deaths[3,1]  1.028e-02 0.0138146 4.369e-03      8.169e-03
theta_deaths[4,1]  3.466e-03 0.0029208 9.236e-04      9.236e-04
theta_deaths[5,1]  3.117e-04 0.0006725 2.127e-04      2.127e-04
theta_deaths[6,1]  1.245e-04 0.0001555 4.917e-05      4.917e-05
theta_deaths[7,1]  8.424e-04 0.0008576 2.712e-04      2.712e-04
theta_deaths[8,1]  9.095e-03 0.0039771 1.258e-03      3.111e-03
theta_deaths[9,1]  8.644e-02 0.0216421 6.844e-03      6.844e-03
theta_deaths[10,1] 3.929e-02 0.0464268 1.468e-02      1.468e-02
eta_deaths         3.336e+01 5.9667252 1.887e+00      3.293e+00
rho_deaths         1.712e+01 2.5518966 8.070e-01      8.070e-01

2. Quantiles for each variable:

                      2.5%       25%       50%       75%     97.5%
theta_deaths[1,1]  1.265e-02 9.688e-02 1.914e-01 2.901e-01 2.901e-01
theta_deaths[2,1]  2.074e-04 8.441e-04 1.300e-03 2.992e-03 8.302e-03
theta_deaths[3,1]  1.710e-04 1.465e-03 1.465e-03 2.275e-02 3.236e-02
theta_deaths[4,1]  1.580e-04 2.493e-03 2.493e-03 4.149e-03 9.267e-03
theta_deaths[5,1]  2.867e-05 4.978e-05 1.402e-04 1.402e-04 1.758e-03
theta_deaths[6,1]  2.022e-05 3.155e-05 4.211e-05 1.838e-04 4.463e-04
theta_deaths[7,1]  1.866e-04 1.866e-04 5.336e-04 1.026e-03 2.354e-03
theta_deaths[8,1]  4.662e-03 4.662e-03 1.033e-02 1.221e-02 1.381e-02
theta_deaths[9,1]  6.268e-02 7.888e-02 7.888e-02 8.886e-02 1.247e-01
theta_deaths[10,1] 9.969e-03 1.798e-02 1.798e-02 3.762e-02 1.401e-01
eta_deaths         2.284e+01 3.011e+01 3.473e+01 3.733e+01 4.049e+01
rho_deaths         1.324e+01 1.619e+01 1.752e+01 1.752e+01 2.087e+01
```

r r theta_post_gp_var = draws_gp_var %>% reduce(rbind) %>% as_tibble() %>% select(-eta_deaths, -rho_deaths)

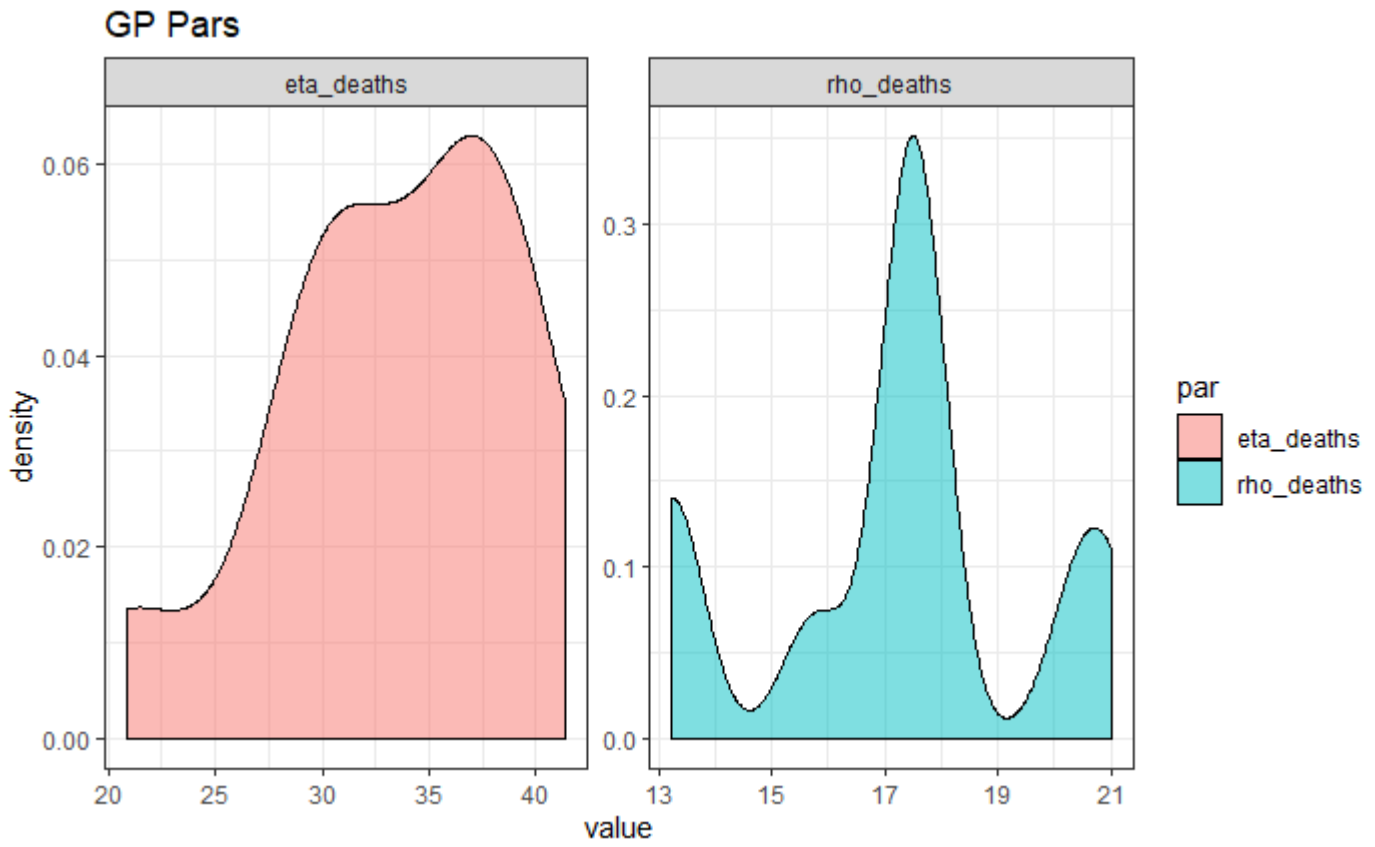
colnames(theta_post_gp_var) = paste0(rep(c( *deaths*), each=categories) , 1:categories) theta_post_gp_var = as_tibble(theta_post_gp_var) %>% mutate(S=1:S) %>% gather(posterior, value, -S) %>% mutate(risk_cat = as.factor(gsub(pattern = _([^_])*$,\,(posterior))), age_cat = as.factor((as.numeric(extract_numeric(posterior)))*10-5))

```
extract_numeric() is deprecated: please use readr::parse_number() instead
```

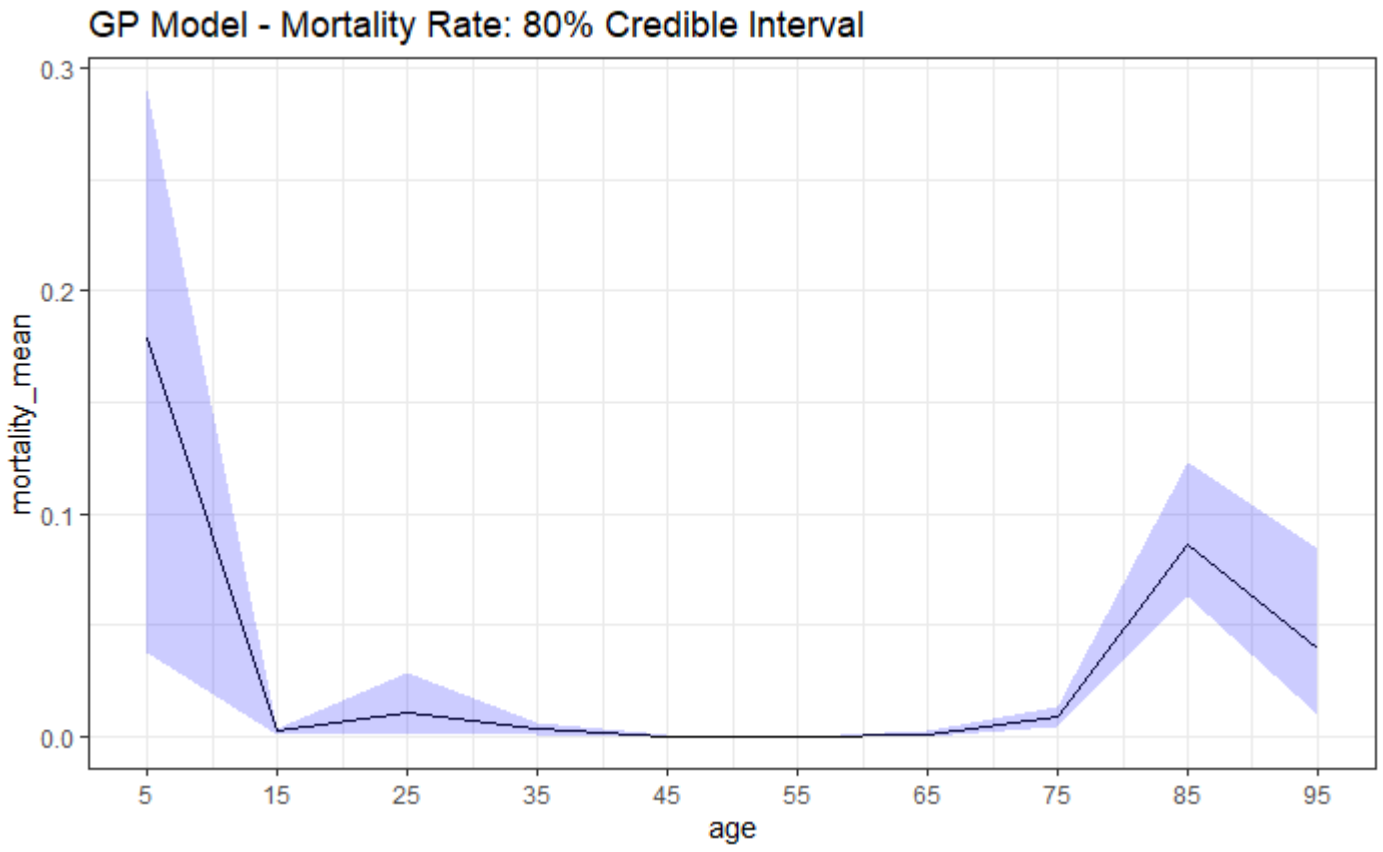
r r theta_post_gp_var_sum = theta_post_gp_var %>% group_by(age_cat) %>% summarize(mean = mean(value), low = quantile(value, 0.1), high = quantile(value, 0.9) ) %>% gather(range, value, -mean, -age_cat) ggplot(theta_post_gp_var_sum) + geom_point(aes(age_cat, mean, color = age_cat),size=2) + geom_line(aes(age_cat, value, group = age_cat)) + labs(title=Princess: Risk of Deathwith 80% Credible Interval, color=\, y=Rate, x=) + theme_bw()




r r # ggplot(theta_post_gp_var) + # geom_violin(aes(age_cat, value, fill = age_cat), alpha=0.5, draw_quantiles = c(0.1, 0.5, 0.9), scale = ) + # labs(title=Princess: Risk of Death, color=\, y=Rate, x=) + # ylim(0,.2)+ # theme_bw() gp_par = draws_gp_var %>% reduce(rbind) %>% as_tibble() %>% select(eta_deaths, rho_deaths) %>% gather(par, value) gp_par_plot = ggplot(gp_par) + geom_density(aes(value, fill=par), alpha=0.5) + labs(title=Pars) + facet_wrap(~par, scales = )+ theme_bw() gp_par_plot




r r risk_band = theta_post_gp_var %>% group_by(age_cat) %>% summarize(mortality_mean = mean(value), mortality_low = quantile(value, probs = .1), mortality_high = quantile(value, probs = .9)) gp_mortality = ggplot(risk_band %>% mutate(age= as.numeric(as.character(age_cat)))) + geom_line(aes(age, mortality_mean)) + geom_ribbon(aes(age, ymin=mortality_low, ymax=mortality_high), fill= , alpha=.2)+ theme_bw() + labs(title=Model - Mortality Rate: 80% Credible Interval)+ scale_x_continuous(breaks = diamond$age) gp_mortality




r r risk_band

| age_cat <fctr> | mortality_mean <dbl> | mortality_low <dbl> | mortality_high <dbl> |
|---|---|---|---|
| 5 | 0.1784818576 | 3.752783e-02 | 0.2901118588 |
| 15 | 0.0023938470 | 7.619264e-04 | 0.0037021905 |
| 25 | 0.0102788319 | 6.289948e-04 | 0.0288424918 |
| 35 | 0.0034662464 | 4.582985e-04 | 0.0059531631 |
| 45 | 0.0003117283 | 4.240483e-05 | 0.0003734893 |
| 55 | 0.0001244725 | 3.009166e-05 | 0.0002345878 |
| 65 | 0.0008423537 | 1.865668e-04 | 0.0023543490 |
| 75 | 0.0090945316 | 4.661952e-03 | 0.0133210052 |
| 85 | 0.0864400057 | 6.268410e-02 | 0.1232989063 |
| 95 | 0.0392922912 | 9.969000e-03 | 0.0846335082 |

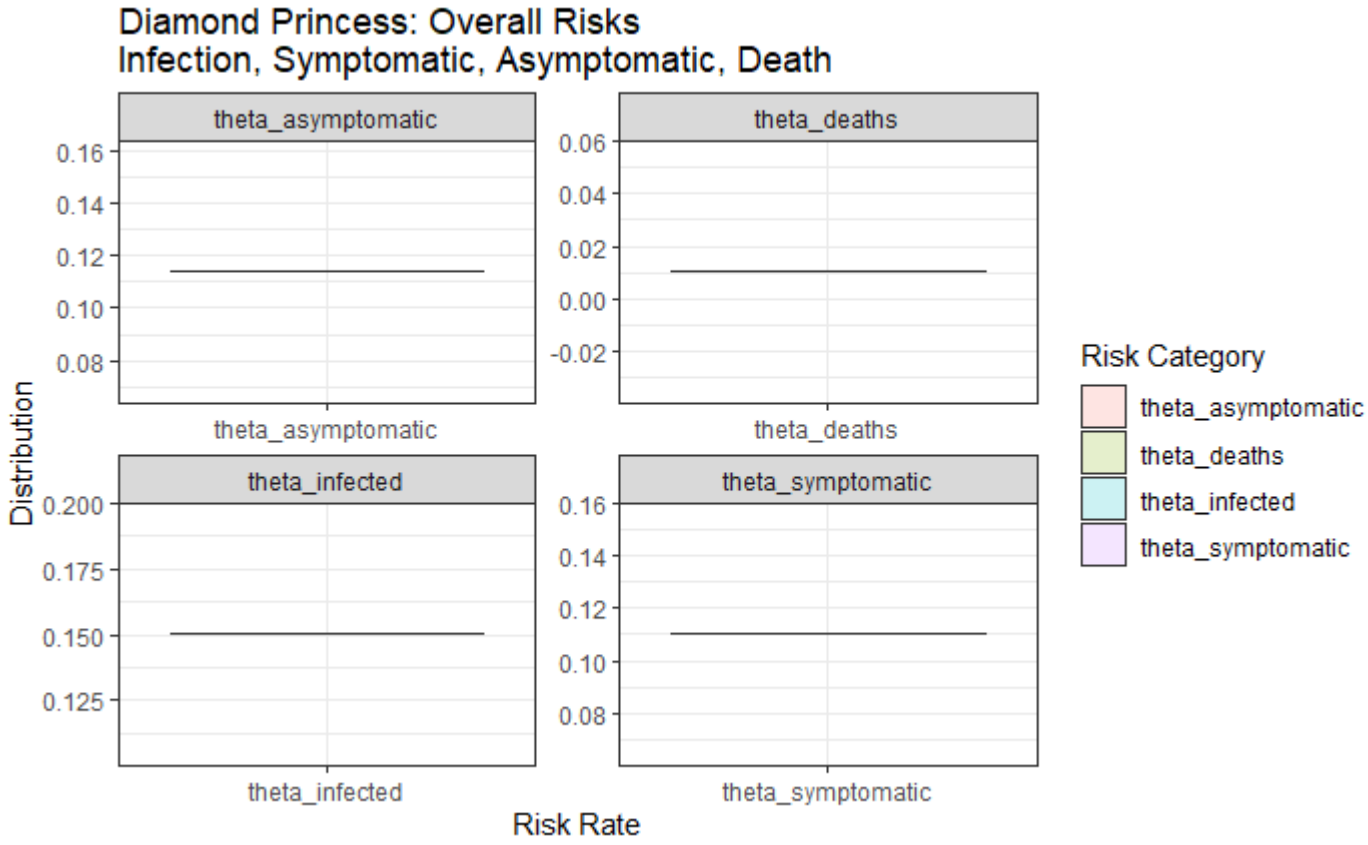
1-10 of 10 rows

# Summary of Results

# A. Combined Age Category Model

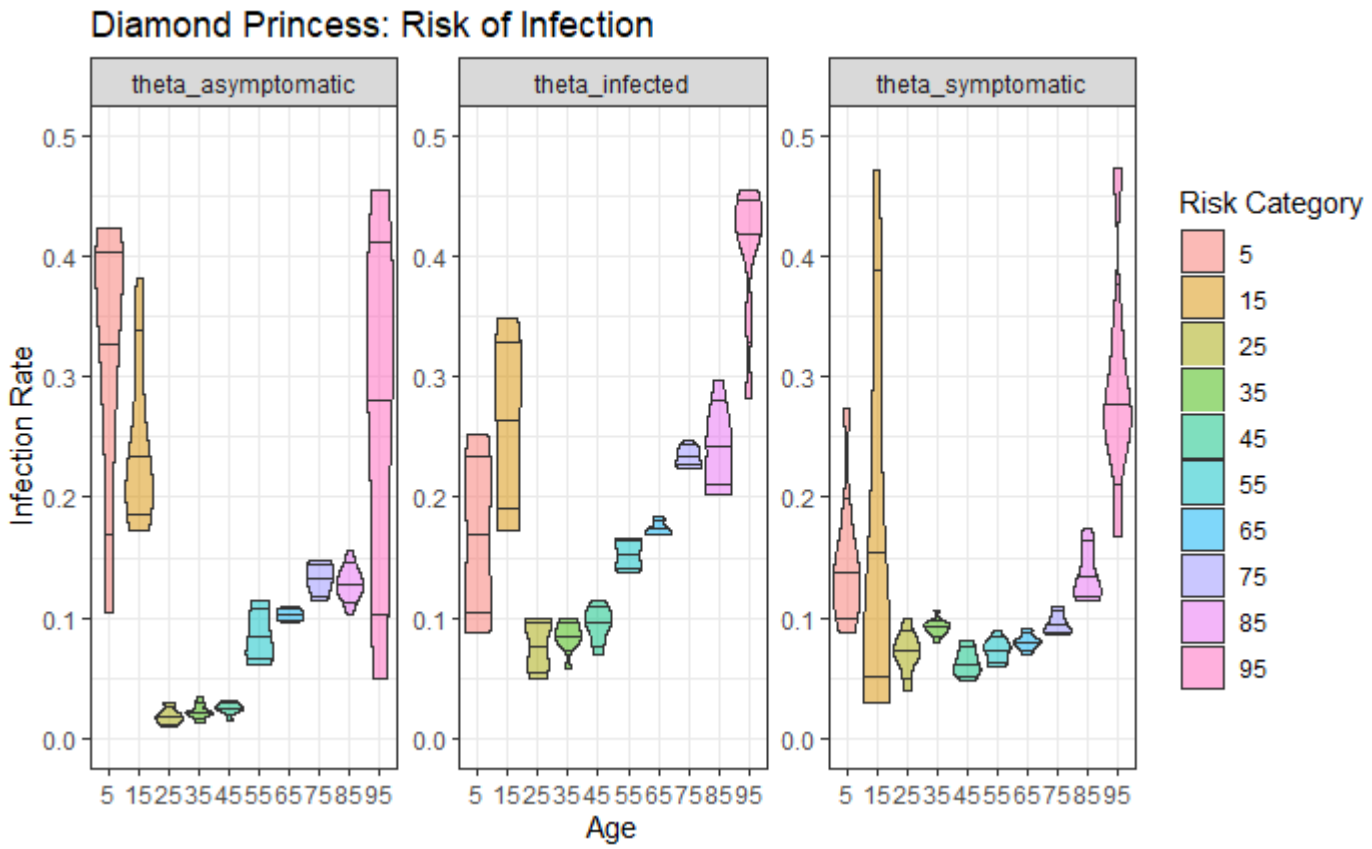
## Risk Rates

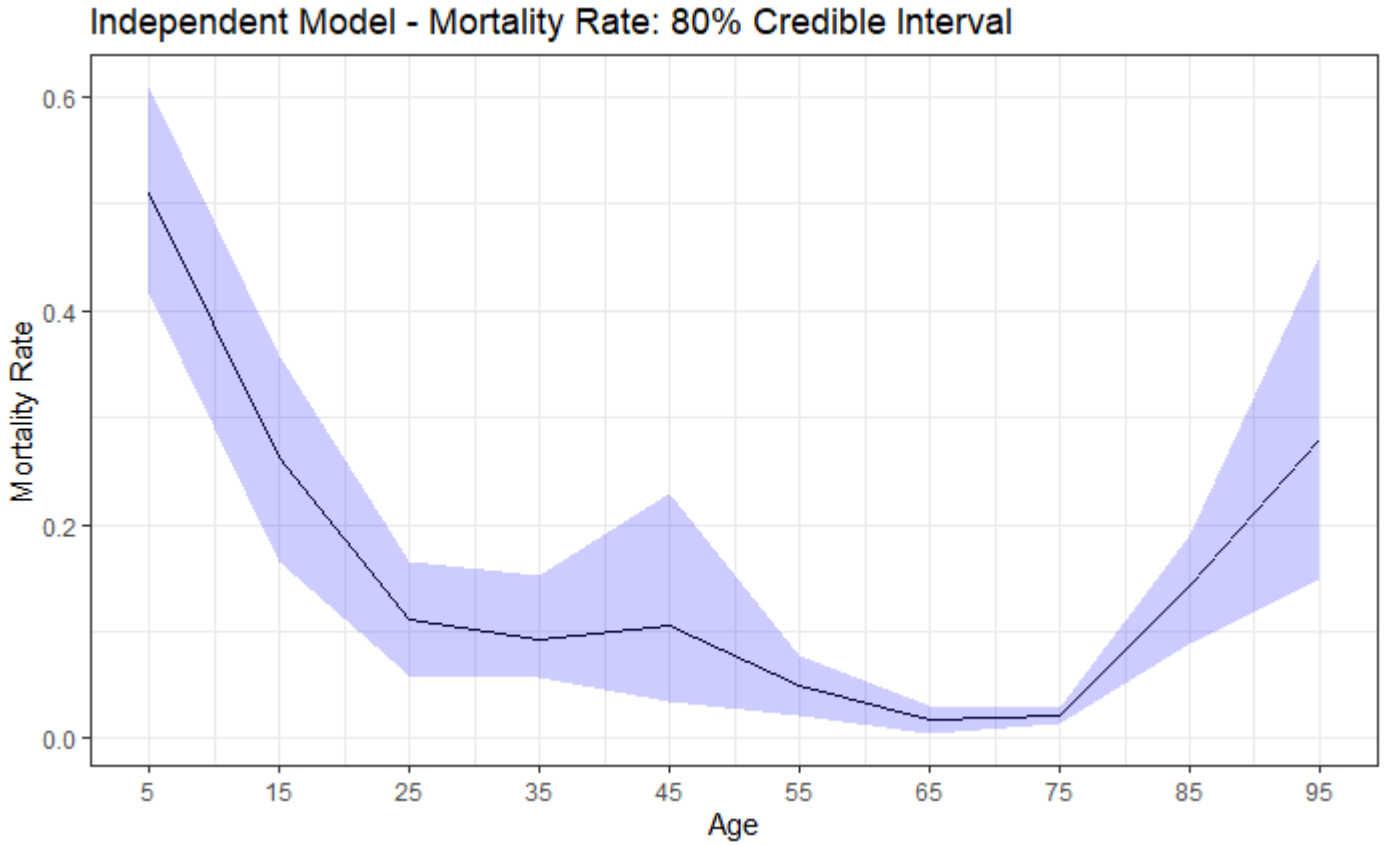
The theta (or $\theta$) amounts represent the parameters of the binomial distribution and are the rates or infection of fatality.




# B. Independent Age Category Model

## Risk Rates



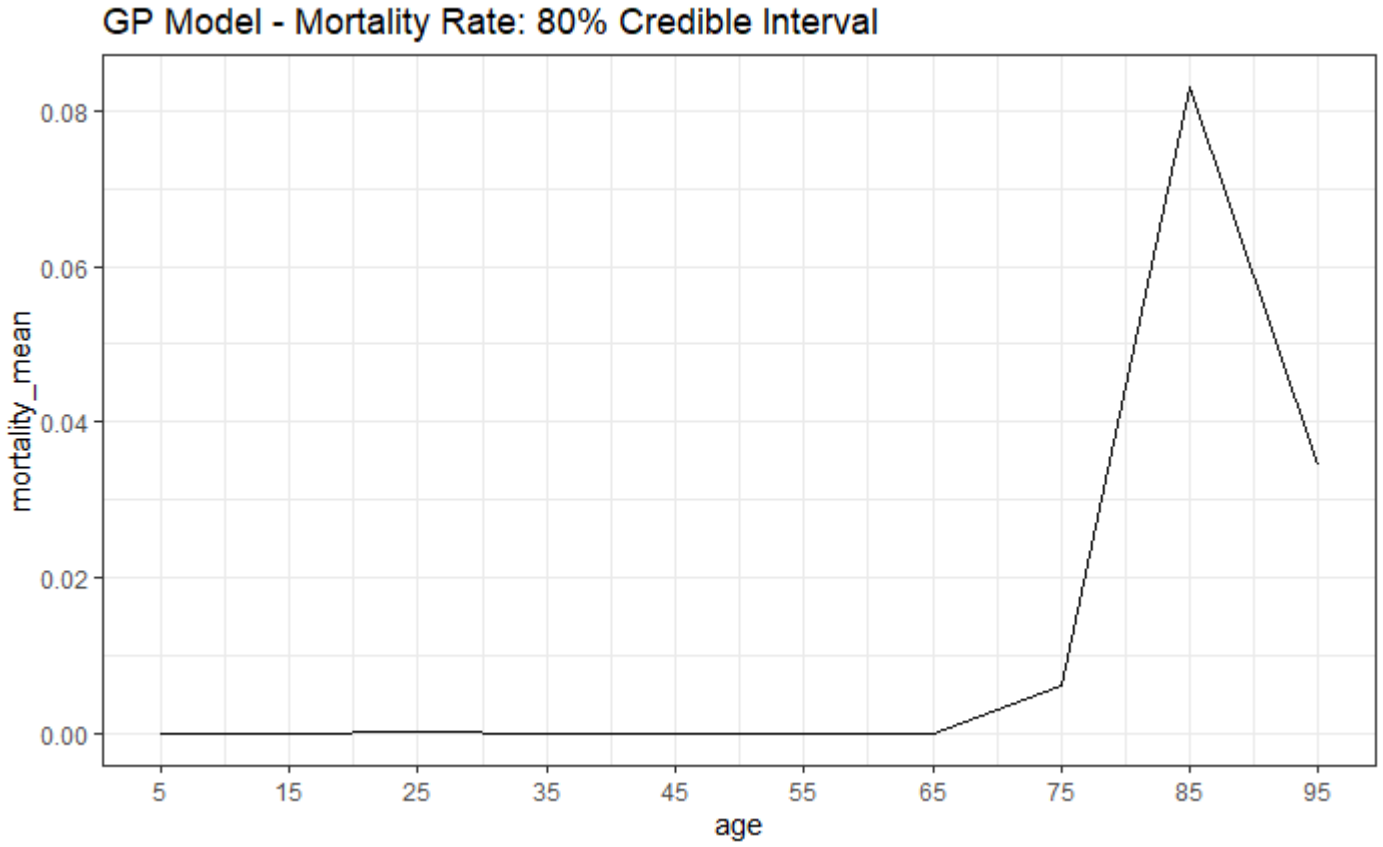




Notes: There are very limited data for the age categories 5, 15 and 95. Here the uniform plays an outsize role, the resultant posterior values for these age buckets suggests a uniform prior might not be the most optimal prior, and a prior skewed towards lower risk rates might be more appropriate.

# C. Gaussian Process Model: Fixed Rho

Fixed $\rho = 10$

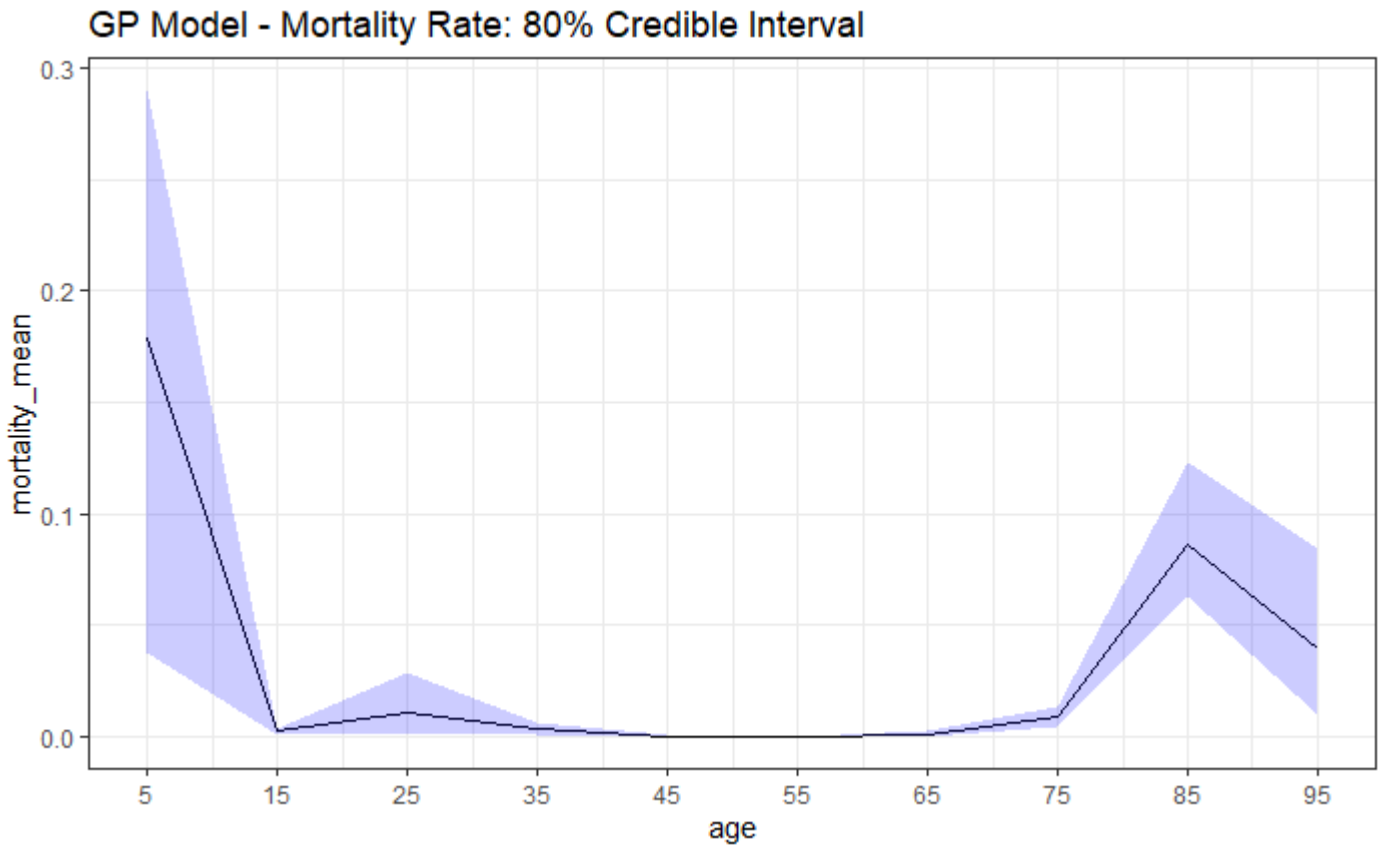
## Fatality Rates




# C. Gaussian Process Model: Variable Rho

Fixed $\rho = 10$

## Fatality Rates




# Model Parameters