

# SEIRD contact matrix

```
library(comomodels)
library(tidyverse)
#> -- Attaching packages ----- tidyverse 1.3.1 --
#> v ggplot2 3.3.4      v purrr 0.3.4
#> v tibble 3.1.2       v dplyr 1.0.6
#> v tidyr 1.1.3        v stringr 1.4.0
#> v readr 1.4.0        v forcats 0.5.1
#> -- Conflicts ----- tidyverse_conflicts() --
#> x dplyr::filter() masks stats::filter()
#> x dplyr::lag()     masks stats::lag()
library(ggplot2)
library(socialmixr)
#>
#> Attaching package: 'socialmixr'
#> The following object is masked from 'package:utils':
#>
#> cite
```

## Introduction

Age-structured compartmental models such as the SEIRD implemented in comomodels use contact matrices to specify the spread of a disease within and between age groups. Given a contact matrix  $C$ , each element  $C_{i,j}$  indicates the expected number of contacts someone from age group  $i$  has per day with people from age group  $j$ .

Comomodels includes estimates of the contact matrix for each country (Prem et al., 2017; with full details available in the data documentation). Separate matrices are available for contacts at home, work, school, and other. Below, we generate a plot of the contact matrices.

```
contact_home <- comomodels::contact_home
contact_work <- comomodels::contact_work
contact_school <- comomodels::contact_school
contact_other <- comomodels::contact_other
population <- comomodels::population

# reformat matrices for plotting
ages <- seq(0, 80, 5)
age_names <- vector(length = 16)
for(i in seq_along(age_names)) {
  age_names[i] <- paste0(ages[i], "-", ages[i + 1])
}

format_matrix <- function(contact_matrix, age_names) {
  colnames(contact_matrix) <- age_names
  contact_matrix$age_infectee <- age_names
  contact_matrix %>%
    pivot_longer(all_of(age_names)) %>%
```

```

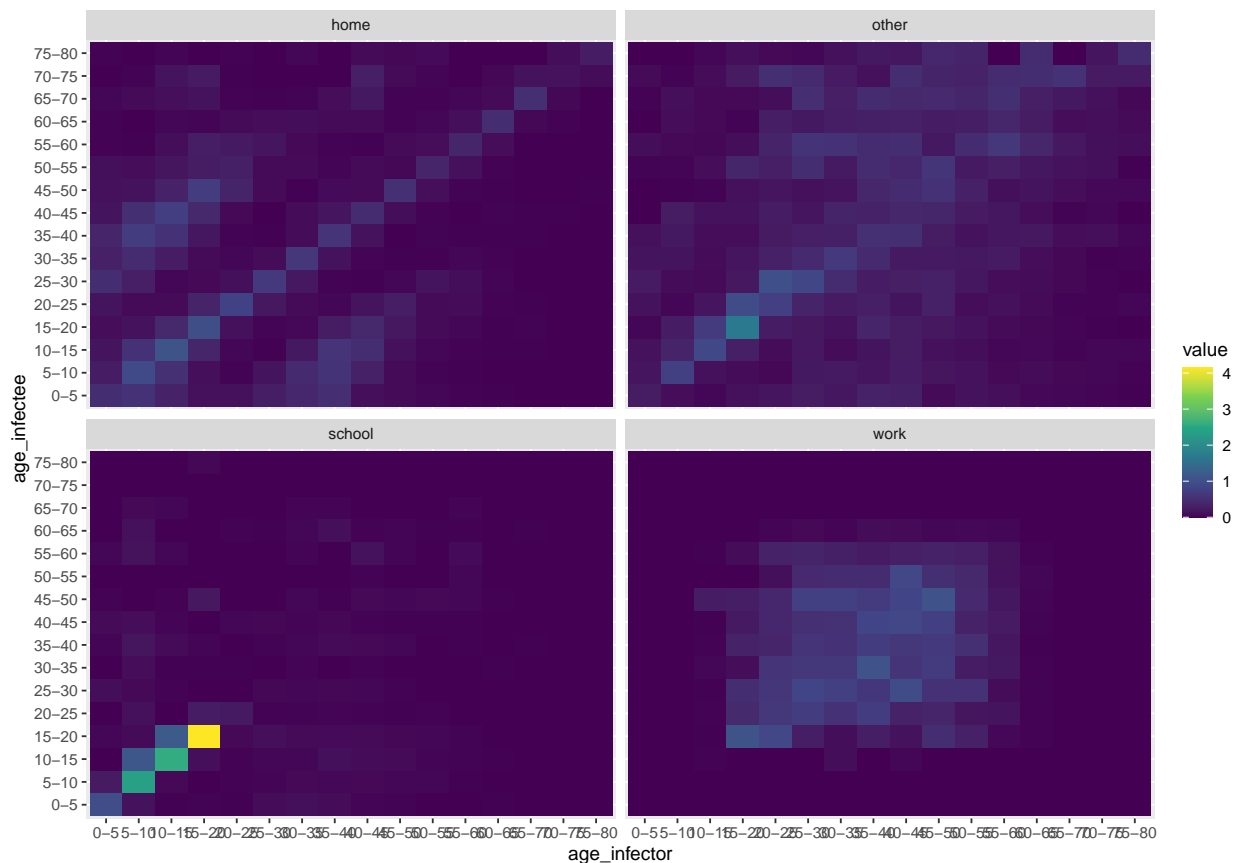
  rename(age_infectior=name) %>%
  mutate(age_infectior=fct_relevel(age_infectior, age_names)) %>%
  mutate(age_infectee=fct_relevel(age_infectee, age_names))
}

c_home <- format_matrix(contact_home$"United Kingdom", age_names) %>% mutate(type="home")
c_work <- format_matrix(contact_work$"United Kingdom", age_names) %>% mutate(type="work")
c_school <- format_matrix(contact_school$"United Kingdom", age_names) %>% mutate(type="school")
c_other <- format_matrix(contact_other$"United Kingdom", age_names) %>% mutate(type="other")

c_all <- c_home %>%
  bind_rows(c_work) %>%
  bind_rows(c_school) %>%
  bind_rows(c_other)

# plot all
c_all %>%
  ggplot(aes(x=age_infectior, y=age_infectee, fill=value)) +
  geom_tile() +
  scale_fill_viridis_c() +
  facet_wrap(~type)

```



While it is possible to construct location-based transmission models, in this study we consider only age structure. Thus, we obtain the total contact matrix for the age-structured SEIRD model by summing the four location specific contact matrices. Further details, including an explanation of the model equations, can be found in the SEIRD\_age\_structured vignette in the comomodels package: <https://github.com/Como->

## Uncertainty in the contact matrix

Typically, when performing simulations of the age-structured SEIR model, or inference for the parameters of the model, the contact matrix is provided as a fixed input. However, using fixed values for the contact matrix neglects the uncertainty which may be present in the contact data.

The purpose of the remainder of this notebook is to investigate the sensitivity of the outputs of the age-structured SEIRD model to the values of the contact matrix. Using bootstrap samples which represent the uncertainty in the contact matrix, we show significant uncertainty in the numbers of infected individuals.

## Accuracy of uncertainty estimates produced by the bootstrapping methods

The bootstrap algorithm is an easy way to obtain some idea of the uncertainty in the contact matrix. However, due to the simplicity of the procedure, its results should be treated as mere approximations of the true uncertainty that may exist in the contact matrix. In particular, the bootstrap algorithm does not account for the possibility that the contact data in the original survey is unrepresentative of the actual population. For example, if contact data is collected primarily from an urban area in a country whose population is mostly rural, the resulting contact matrices may be inaccurate for the country, and the bootstrap cannot account for the lack of information in the original data.

## Bootstrap samples of contact matrices

We obtain samples of the contact matrix using the `socialmixr` library which accesses the POLYMOD data. This library allows us to generate bootstrap samples of the contact matrix for countries covered by the study. The first step is to generate 200 of these samples for the contact matrix in the United Kingdom.

```
# Define age groups and names
ages <- seq(0, 80, 5)
age_names <- vector(length = 16)
for(i in seq_along(age_names)) {
  age_names[i] <- paste0(ages[i], "-", ages[i + 1])
}

# Get population data
pops <- population[population$country == "United Kingdom", ]$pop
pop_fraction <- pops/sum(pops)
pop_fraction[16] <- sum(pop_fraction[16:21])
pop_fraction <- pop_fraction[1:16]
n_ages <- 16

# Load the contact matrix data from POLYMOD and get bootstrap samples
n_bootstrap <- 200
data(polymod)
polymod_data <- contact_matrix(polymod,
                               n=n_bootstrap,
                               countries="United Kingdom",
                               age.limits=ages)

#> Using POLYMOD social contact data. To cite this in a publication, use the 'cite' function
```

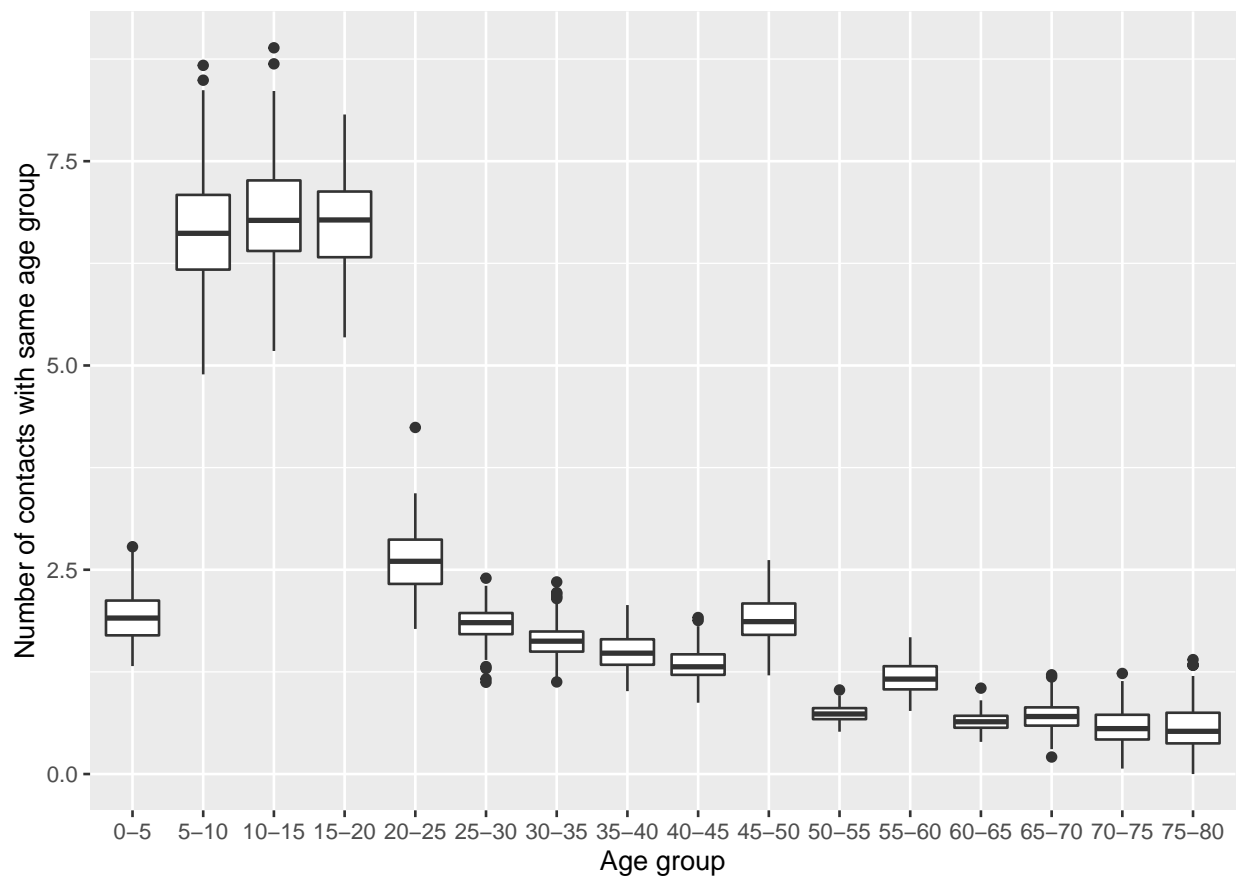
```
# Get the first element of the list, which contains the matrices
matrices <- polymod_data["matrices"][[1]]
```

First, we inspect the range of values in the sampled contact matrices. In the plot below, we look at the distribution of the diagonal elements of the matrix for each age group.

```
contacts_same_age <- c()
ages_list <- c()
for (i in 1:n_bootstrap){
  contacts_same_age <- append(contacts_same_age, diag(matrices[[i]][[1]])[1:16])
  ages_list <- append(ages_list, age_names)
}

data <- data.frame(ages_list, contacts_same_age)
data$ages_list <- factor(data$ages_list, levels=age_names[1:16], ordered=TRUE)

ggplot(data, aes(x=ages_list, y=contacts_same_age), 1) + geom_boxplot() +
  xlab("Age group") + ylab("Number of contacts with same age group")
```



The plot shows the most uncertainty for ages 5–20 (schoolchildren). In most of the other age groups, the uncertainty is small but nowhere does it appear negligible. In the next section, the effect of these uncertainties on the SEIR outputs will be studied.

## SEIRD simulations

In this step, we run the age-structured SEIRD model once for each bootstrap sample of the contact matrix. We use fixed values for the other parameters of the model. In particular, we set  $\beta = 2$ ,  $\kappa = 1$ , and  $\gamma = 1$ , with an initial exposed group compartment equal to 0.1% of the population size, and all other individuals in the susceptible group.

The age-structured model takes another parameter  $\mu$  which specifies the rate at which infected people enter the deceased compartment. For this parameter, we provide a vector giving a separate value for each age group. The values are selected to resemble the death rates of the COVID-19 outbreak in mainland China during January and February 2020 (Verity, R, et al. “Estimates of the severity of coronavirus disease 2019: a model-based analysis.” The Lancet Infectious Diseases 20.6 (2020): 669-677).

```
mu=c(0.000016, 0.000016, 0.00007, 0.00007, 0.00031, 0.00031, 0.0026, 0.0026,
      0.0048, 0.0048, 0.006, 0.006, 0.019, 0.019, 0.043, 0.043)

for (i in 1:n_bootstrap){
  matrix=matrices[[i]][[1]]

  # Remove the column and row names so the model will accept it
  colnames(matrix) <- NULL
  rownames(matrix) <- NULL

  # Keep the data for ages 0-80, in 5 year increments
  matrix <- matrix[1:16,1:16]

  model <- comomodels::SEIRDage(n_age_categories=n_ages,
                                contact_matrix=matrix,
                                age_ranges=as.list(age_names))

  # Set the other parameters of the model
  transmission_parameters(model) <- list(b=2.0, k=1.0, g=1.0, mu=mu)
  initial_conditions(model) <- list(S0=pop_fraction*0.999,
                                    E0=rep(0, n_ages),
                                    I0=pop_fraction*0.001,
                                    R0=rep(0, n_ages),
                                    D0=rep(0, n_ages))

  res <- run(model, time=seq(0, 100, by=1))

  # Get states from results
  res <- res[['states']]

  # Save the data for the I and R compartments
  x = filter(res, compartment %in% c("I", "R", "D"))
  if (i==1)
    all_results <- x
  else
    all_results <- rbind(all_results, x)
}
```

Having obtained the simulation results, we plot the central 90% probability interval of the number in the infected compartment over time for two selected age groups (15–20 and 75–80).

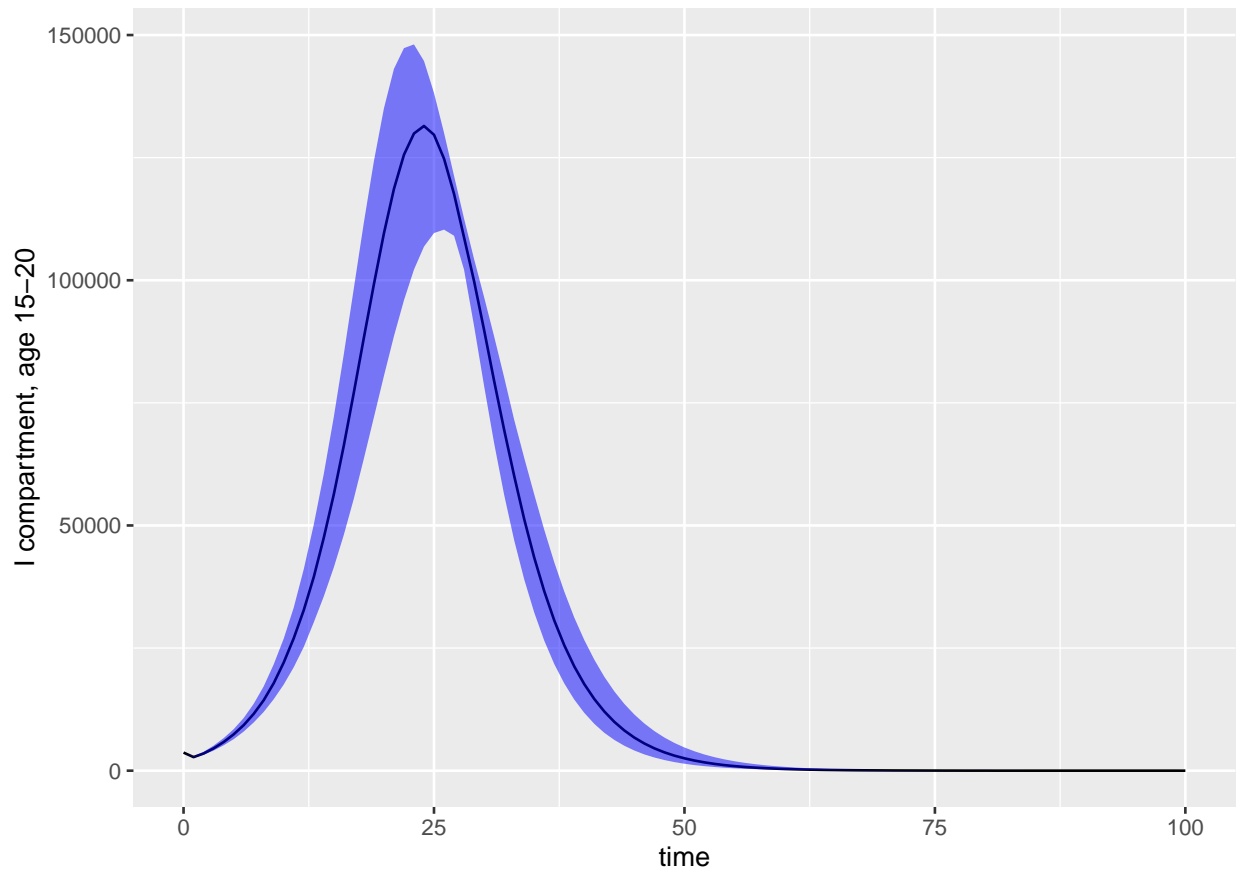
```
I_data <- filter(all_results, age_range=="15-20", compartment=="I")
data <- I_data$value * sum(pops)
```

```

dim(data) <- c(length(I_data$time)/n_bootstrap, n_bootstrap)
quants <- t(apply(data, 1, quantile, probs=c(0.05, 0.5, 0.95), na.rm=TRUE))
quants_df <- data.frame(quants)
quants_df["time"] <- seq(0, 100, by=1)

ggplot(quants_df, aes(x = time)) +
  geom_line(aes(y=X50.)) +
  geom_ribbon(aes(ymin=X5., ymax=X95.), fill="blue", alpha=0.5) +
  ylab("I compartment, age 15-20")

```

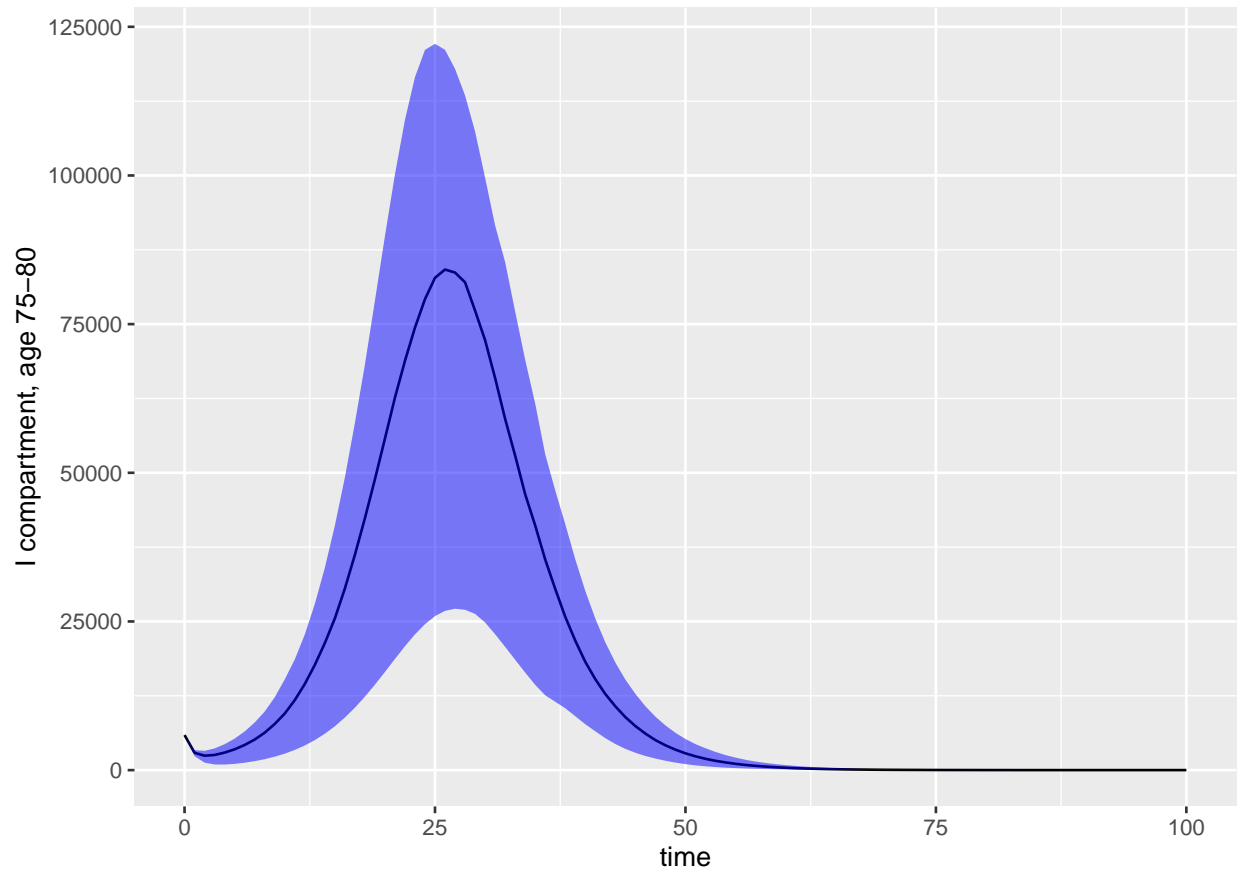


```

I_data <- filter(all_results, age_range=="75-80", compartment=="I")
data <- I_data$value * sum(pops)
dim(data) <- c(length(I_data$time)/n_bootstrap, n_bootstrap)
quants <- t(apply(data, 1, quantile, probs=c(0.05, 0.5, 0.95), na.rm=TRUE))
quants_df <- data.frame(quants)
quants_df["time"] <- seq(0, 100, by=1)

ggplot(quants_df, aes(x = time)) +
  geom_line(aes(y=X50.)) +
  geom_ribbon(aes(ymin=X5., ymax=X95.), fill="blue", alpha=0.5) +
  ylab("I compartment, age 75-80")

```

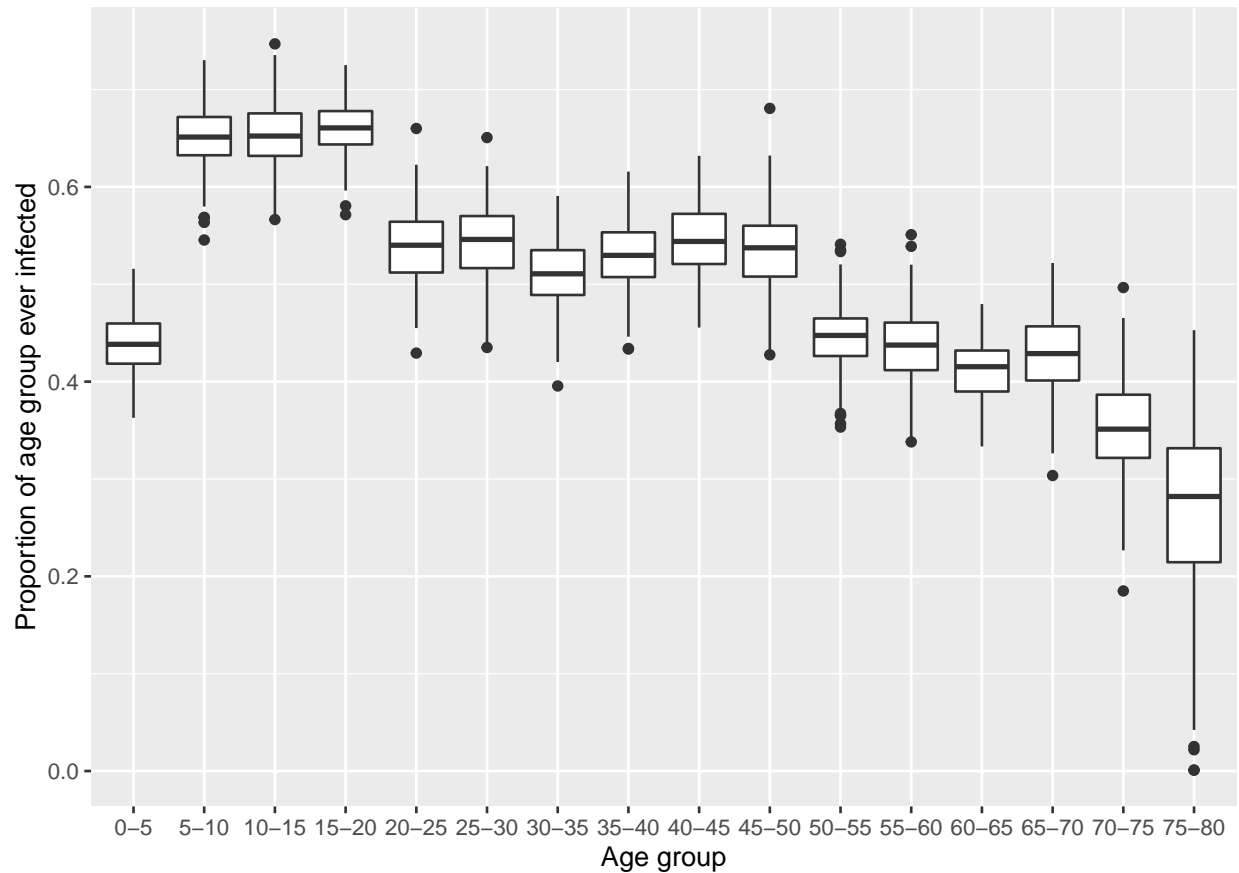


The results show that the while the shape of the epidemic trajectory remains similar for all contact matrices in the bootstrap sample, the number of people in the I compartment exhibits significant uncertainty, particularly near the peak of the epidemic.

Next, we perform the uncertainty analysis for the number in the recovered compartment at the final time point, for all age groups.

```
data <- filter(all_results, compartment=="R", time==100.0)

ggplot(data, aes(x=age_range, y=value/pop_fraction), 1) +
  geom_boxplot() +
  xlab("Age group") +
  ylab("Proportion of age group ever infected")
```



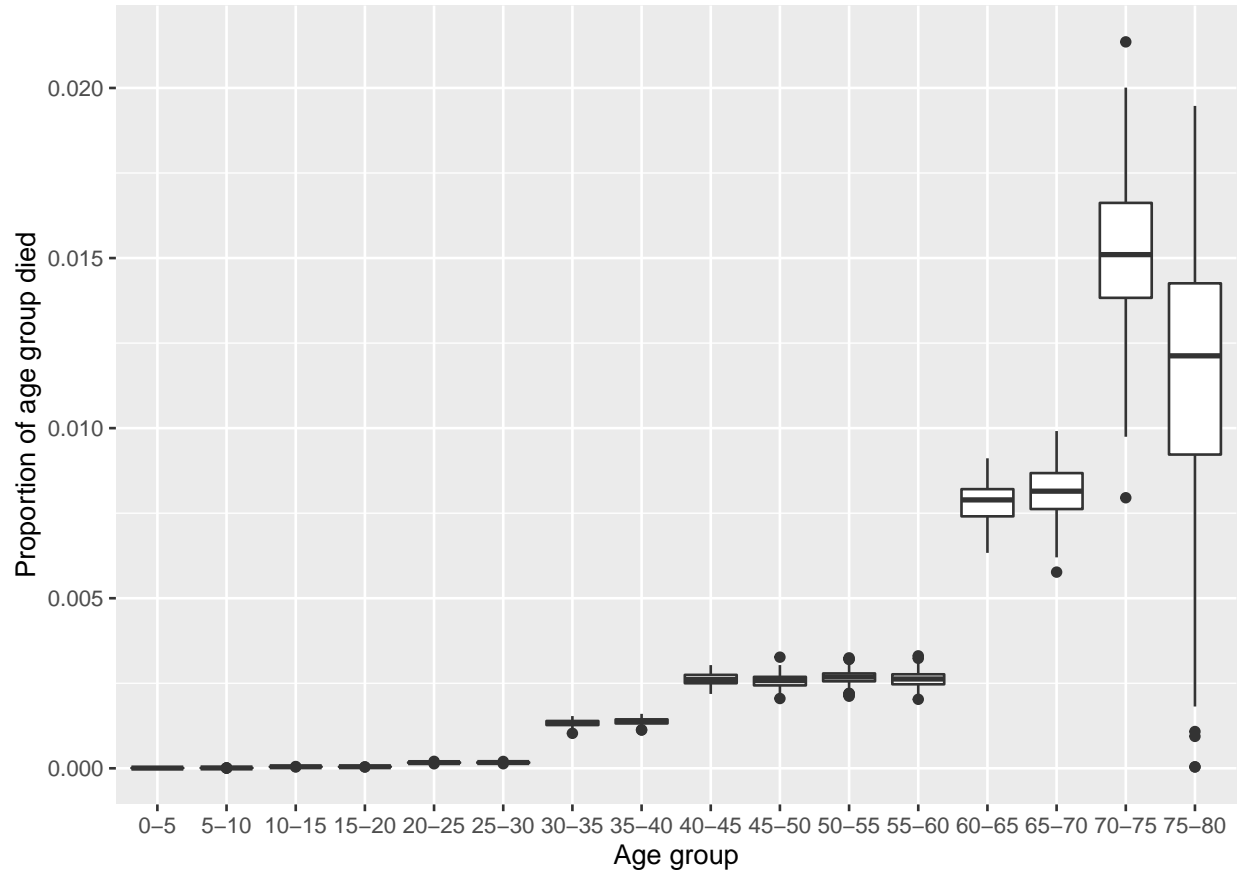
These results show that younger age groups suffer more infections. This trend can be explained by the larger numbers of total contacts for younger age groups, as observed in the contact matrices at the beginning of this notebook.

Finally, we study the effects on death.

```
data <- filter(all_results, compartment=="D", time==100.0)

ggplot(data, aes(x=age_range, y=value/pop_fraction), 1) +
  geom_boxplot() +
  xlab("Age group") +
  ylab("Proportion of age group died")
```





Although younger people are more likely to be infected in this simulation, deaths occur mainly in the elderly, due to the age-structured mortality parameter  $\mu$  described above. The different bootstrap samples of the contact matrix result in a wide range in the number of deaths, particularly in the 75–80 age group.