EbolaSIR_Rproject

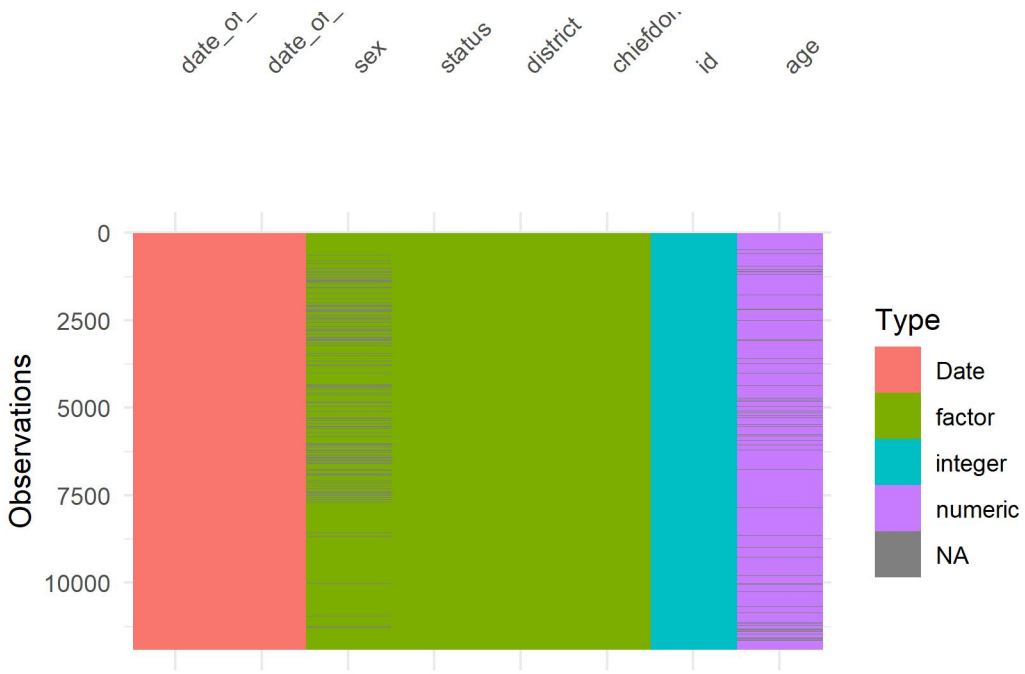
```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union


Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

EDA

Load and visulize the data (ebola_sierraleone_2014)

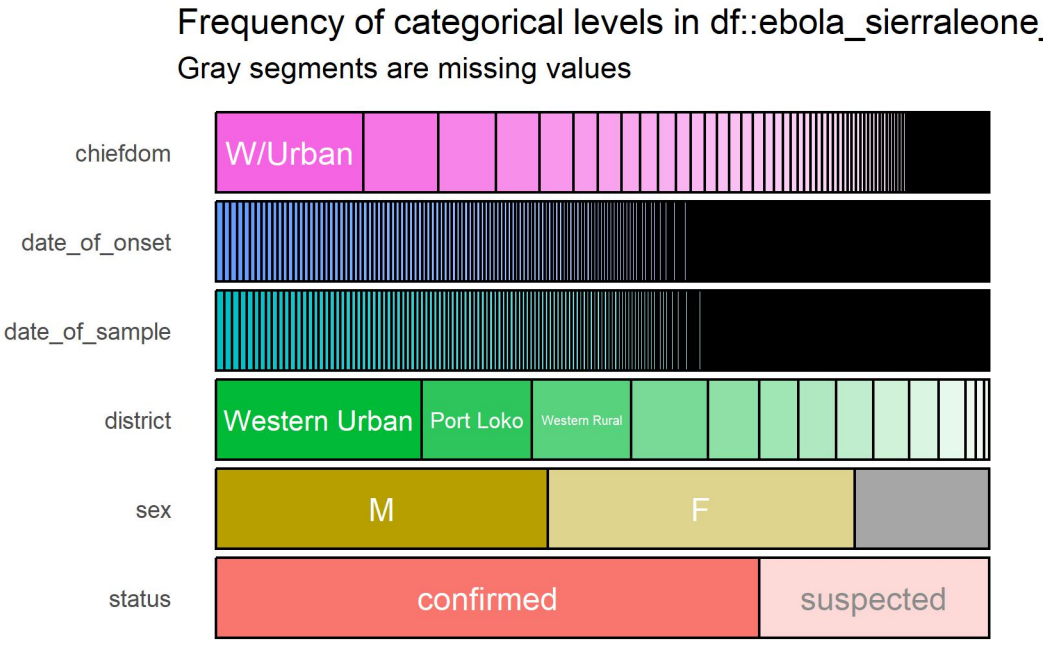
check for missing data

```
  id age sex    status date_of_onset date_of_sample district    chiefdom
1  1  20   F confirmed    2014-05-18     2014-05-23 Kailahun  Kissi Teng
2  2  42   F confirmed    2014-05-20     2014-05-25 Kailahun  Kissi Teng
3  3  45   F confirmed    2014-05-20     2014-05-25 Kailahun Kissi Tonge
4  4  15   F confirmed    2014-05-21     2014-05-26 Kailahun  Kissi Teng
5  5  19   F confirmed    2014-05-21     2014-05-26 Kailahun  Kissi Teng
6  6  55   F confirmed    2014-05-21     2014-05-26 Kailahun  Kissi Teng
```

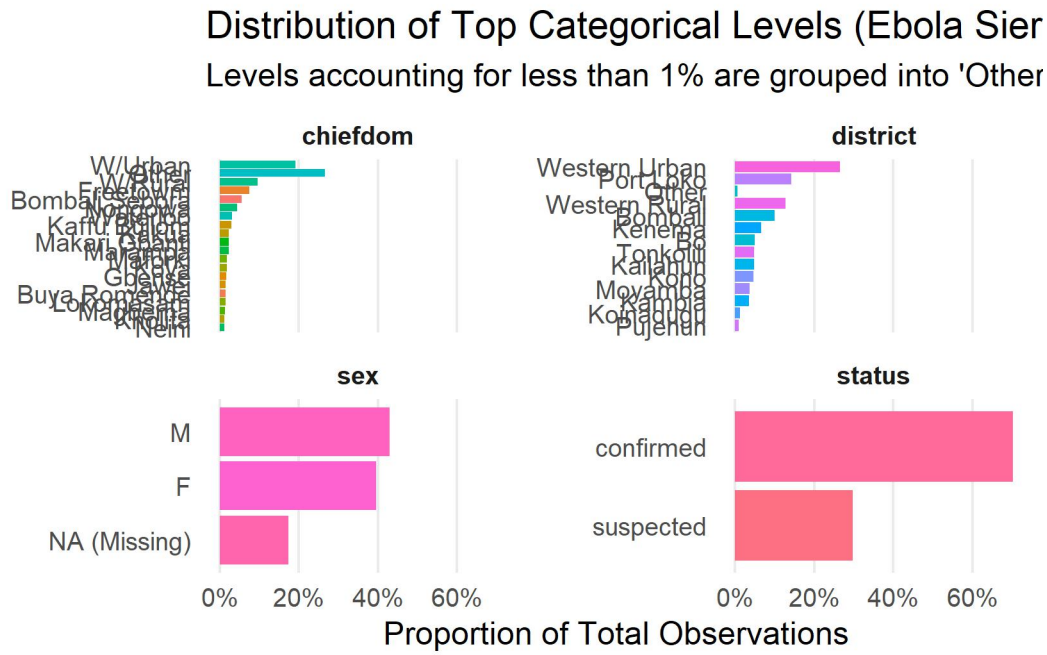


From this figure, we can see sex and age are missing for some cases.

```
Warning in geom_bar(stat = "identity", position = "stack", colour = "black", :
Ignoring unknown parameters: `size`
```



The above figure shows distribution across districts, sex and status.



Data cleaning for POMP

```
        date
1 2014-05-18
2 2014-05-20
3 2014-05-20
4 2014-05-21
5 2014-05-21
6 2014-05-21
```

POMP requires cts data, going to group daily cases by week and sum.

```
# A tibble: 6 x 2
  week_start confirm_cases
  <date>             <int>
1 2014-05-18            13
2 2014-05-25            20
3 2014-06-01            22
4 2014-06-08            84
5 2014-06-15            40
6 2014-06-22            73
```

Let's make the data ready for pomp. Create the time series data.

```
'data.frame':   69 obs. of  2 variables:
 $ time : num  1 2 3 4 5 6 7 8 9 10 ...
 $ cases: int  13 20 22 84 40 73 79 74 53 76 ...
```

Build Stochastic SIR POMP

Stochastic SIR POMP Model

We model the transmission dynamics of the 2014 Ebola outbreak in Sierra Leone using a **stochastic SIR partially observed Markov process (POMP)**. The latent epidemic process is represented by a stochastic SIR model with binomial transition dynamics, while the observation process links the latent epidemic states to reported case counts through a Negative Binomial distribution to account for reporting noise and overdispersion.

The population is divided into epidemiological compartments, and the compartment counts form the latent state vector X_t. The model is formulated in discrete time with weekly resolution and is implemented using the `pomp` package in R.

Transmission and Recovery Rates

The SIR model is governed by two key epidemiological rates:

- β_t: the **time-varying transmission rate**, which controls how frequently infectious individuals transmit infection to susceptible individuals.
- γ: the **recovery rate**, which determines the rate at which infectious individuals recover or are removed from the infectious compartment.

From these quantities, we compute the **effective reproduction number** $\mathcal{R}_t = \beta_t/\gamma$. This quantity is the number of people in a population who can be infected by an individual at a specific time. When $\mathcal{R}_t > 1$, each infectious individual generates more than one secondary infection on average and the epidemic tends to grow. When $\mathcal{R}_t < 1$, transmission is insufficient to sustain epidemic growth and case counts tend to decline. In this framework, $\mathcal{R}_t$ reflects changes in transmission intensity due to behavioral responses and public health interventions rather than depletion of the susceptible population.

State Variables and Time Step

Let N denote the total population size. At time t, let:

- S_t denote the number of susceptible individuals,
- I_t denote the number of infectious individuals (prevalence),
- R_t denote the number of recovered or removed individuals.

In addition, we define C_t as the **incidence**, representing the number of new infections occurring during week t.

The latent state vector is therefore $X_t = \{S_t, I_t, R_t, C_t, \log \beta_t\}$.

Time is discretized into weekly intervals with step size $\Delta t = 1$ week.

Process Model Specification

Infection Flow ($S \to I$)

Let $N_{SI,t}$ denote the number of susceptible individuals who become infected during the interval $(t, t + \Delta t]$. New infections are modeled probabilistically as $N_{SI,t} \sim \text{Binomial}(S_t, p_{SI,t})$, where $p_{SI,t}$ is the probability that a susceptible individual becomes infected during the time step.

The infection probability is derived from the per-capita force of infection, $\lambda_t = \beta_t \frac{I_t}{N}$, yielding $p_{SI,t} = 1 - \exp(-\lambda_t \Delta t) = 1 - \exp\left(-\beta_t \frac{I_t}{N}\right)$, since $\Delta t = 1$.

Weekly incidence is recorded as $C_{t+1} = N_{SI,t}$.

Recovery Flow ($I \to R$)

Let $N_{IR,t}$ denote the number of infectious individuals who recover during the interval $(t, t+\Delta t]$. Recoveries are modeled as $N_{IR,t} \sim \text{Binomial}(I_t, p_{IR})$, where the probability of recovery is derived from the recovery rate, $p_{IR} = 1 - \exp(-\gamma \Delta t)$.

State Update Equations

The compartment sizes evolve according to $S_{t+1} = S_t - N_{SI,t}$,
$I_{t+1} = I_t + N_{SI,t} - N_{IR,t}$,
$R_{t+1} = R_t + N_{IR,t}$.

Stochasticity in the process model arises from these discrete infection and recovery events, reflecting intrinsic demographic variability in disease transmission.

Time-Varying Transmission

We allow the transmission rate β_t to evolve over time. Specifically, β_t follows a random walk on the log scale, $\log \beta_{t+1} = \log \beta_t + \eta_t$,
where $\eta_t \sim \text{Normal}(0, \sigma^2_{\text{proc}})$.

This formulation ensures that β_t remains positive while reflecting more realistic transmission dynamics due to events such as Operation Octopus. The time-varying transmission rate is treated as part of the latent state vector and is estimated jointly with the compartment sizes using particle filtering.

Observation Model

Observed weekly Ebola case counts correspond to incidence, not prevalence. Let Y_{t+1} denote the number of reported cases during week $t+1$. The observation process is modeled using a Negative Binomial distribution, $Y_{t+1} \sim \text{NegBinomial}(\mu_{t+1}, k)$, with mean $\mu_{t+1} = \rho\, C_{t+1}$.

Here, $\rho \in (0, 1)$ denotes the reporting fraction, representing the expected proportion of true infections that are reported, and $k > 0$ is a dispersion parameter allowing for overdispersion relative to a Poisson observation model.

Stochastic Process Implementation

The stochastic SIR process and observation model are implemented in `pomp` using `rprocess` and `dmeasure` Csnippets. These components define the binomial infection and recovery transitions, the random walk evolution of $\log \beta_t$, and the Negative Binomial measurement model. The full latent state vector used in the POMP specification is $X_t = \{S_t, I_t, R_t, C_t, \log \beta_t\}$.

The following csnippet functions define the structure of the POMP object.

X_t = {S,I,R, C, logβ}.

Measurement Density - dmeas

This calculates the likelihood of observing Y_t (reported cases) given the latent incidence C_t using the Negative Binomial distribution.

dnbinom_mu: Negative binomial distribution with number of success size and mean mu

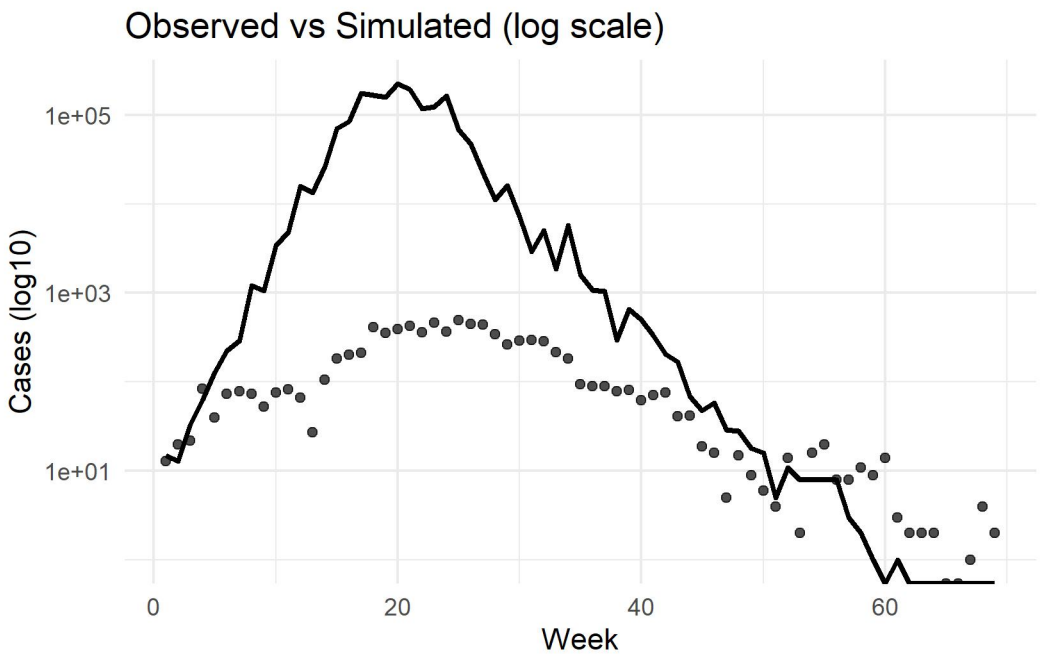
For Poisson Distribution

Initialization

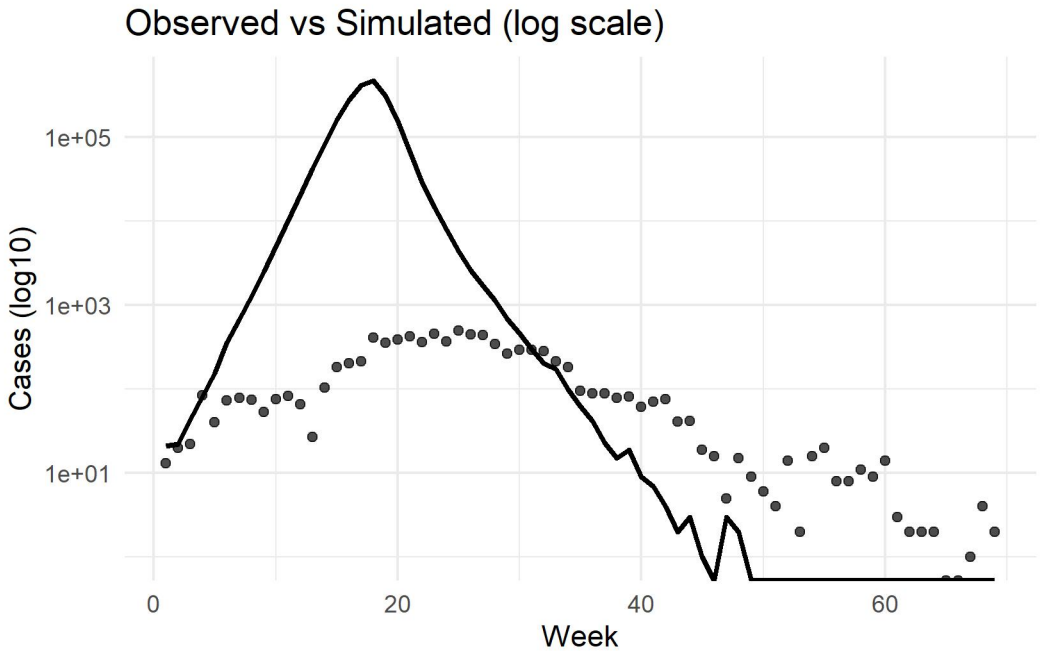
Build the process functions, and the parameter definitions to build the final pomp object.

Build the pomp object negative binomial:

```
Warning in scale_y_log10(): log-10 transformation introduced infinite values.
log-10 transformation introduced infinite values.
```




```
Warning in scale_y_log10(): log-10 transformation introduced infinite values.
log-10 transformation introduced infinite values.
```



R_eff for NB

	time	beta	R_eff	S	I	C	cases
70	1	1.0781058	2.395755	6999895	88	55	15
71	2	1.1148090	2.477282	6999798	157	97	13
72	3	1.0713119	2.380571	6999640	262	158	33
73	4	1.0022271	2.226978	6999392	423	248	62
74	5	0.9684369	2.151769	6998980	701	412	127
75	6	0.9796684	2.176502	6998268	1198	712	222
76	7	0.9956836	2.211708	6997081	2020	1187	289
77	8	1.0309573	2.289380	6995000	3419	2081	1208
78	9	1.1671106	2.590236	6990978	6363	4022	1056
79	10	1.2721971	2.820175	6982841	12417	8137	3453

R_eff for poisson

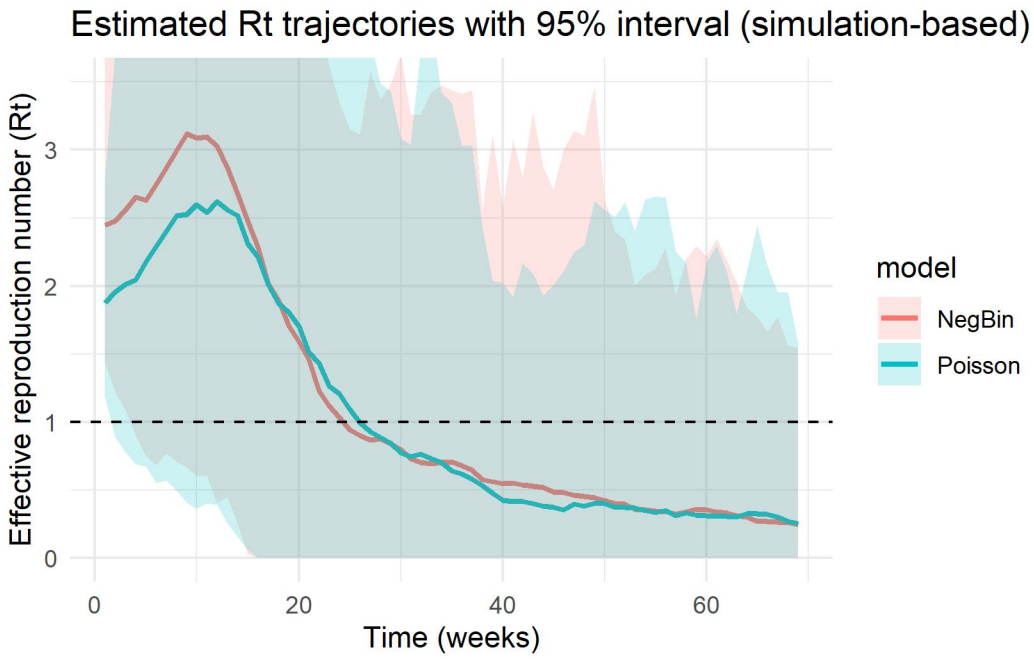
	time	beta	R_eff	S	I	C	cases
70	1	1.125510	2.501100	6999906	71	44	21
71	2	1.217909	2.706397	6999827	128	79	22
72	3	1.119990	2.488758	6999691	216	136	43
73	4	1.149702	2.554687	6999432	400	259	82
74	5	1.238842	2.752553	6998907	800	525	153
75	6	1.347722	2.993976	6997752	1686	1155	357
76	7	1.325654	2.943992	6995472	3416	2280	688
77	8	1.260253	2.797014	6991133	6598	4339	1295
78	9	1.251609	2.774558	6982898	12635	8235	2513
79	10	1.321415	2.922265	6966121	25151	16777	5106

Negative Binomial coefficients:

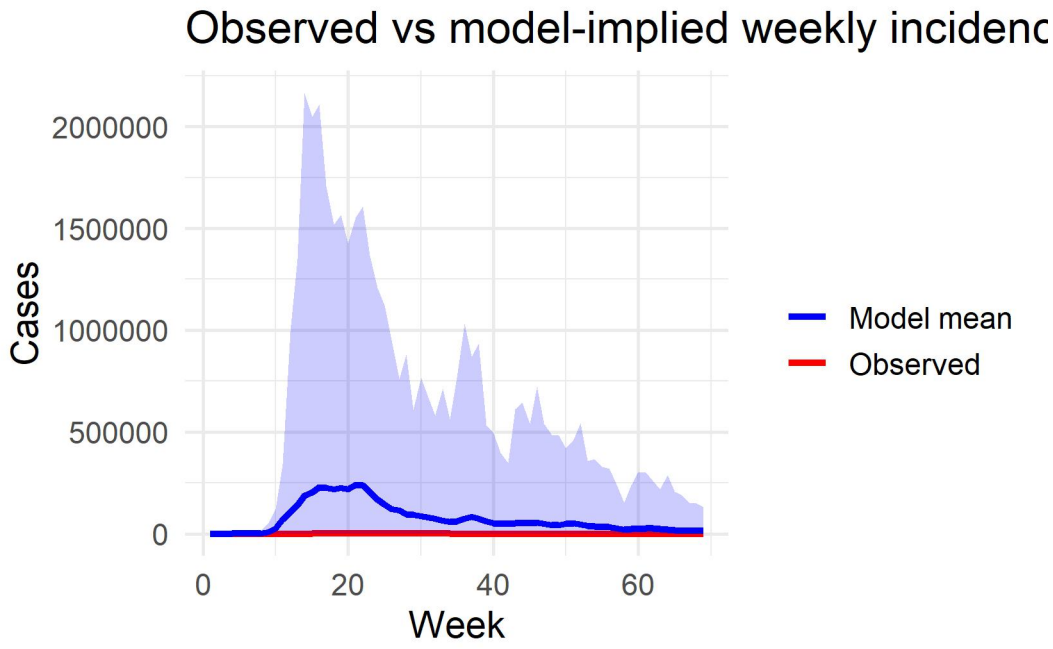
gamma	rho	k	N	logbeta0
2.748356e-01	5.002242e-01	2.326257e+01	7.000000e+06	-4.202177e-01
sigma_proc	**S0**	**I0**		
2.258461e-01	6.999950e+06	5.000000e+01		

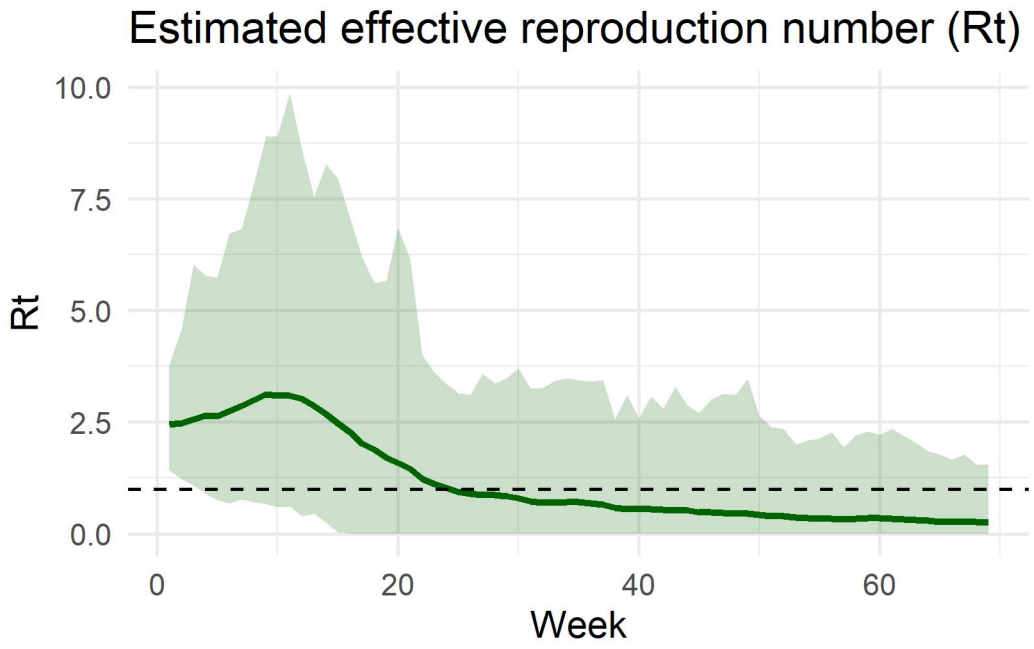
Poisson coefficients:

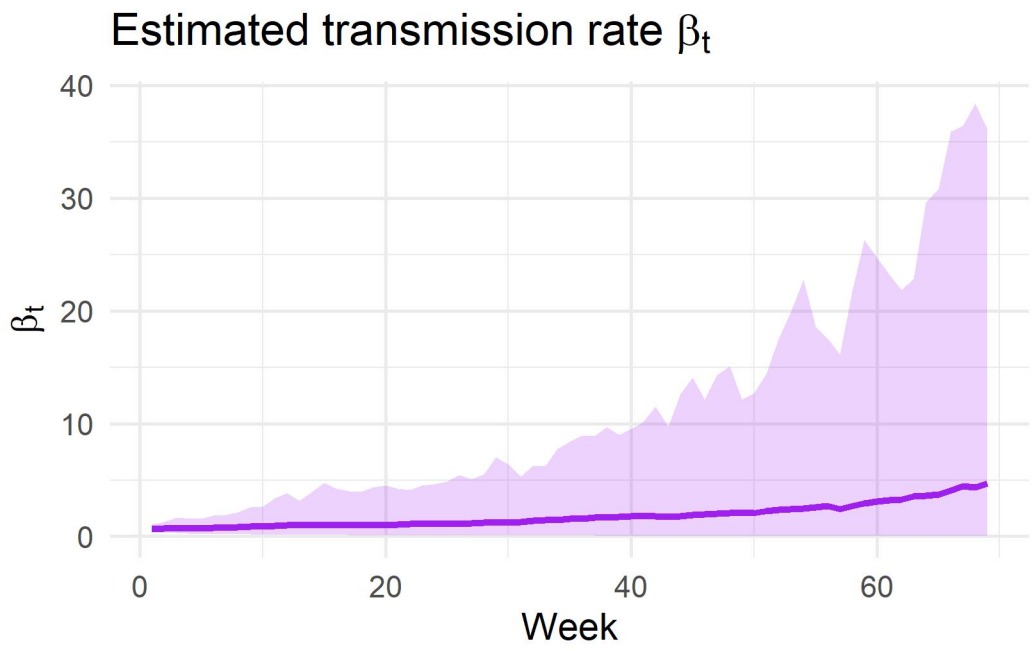
gamma	rho	N	logbeta0	sigma_proc
2.894346e-01	8.155383e-01	7.000000e+06	-5.871847e-01	2.608304e-01
S0	**I0**			
6.999950e+06	5.000000e+01			



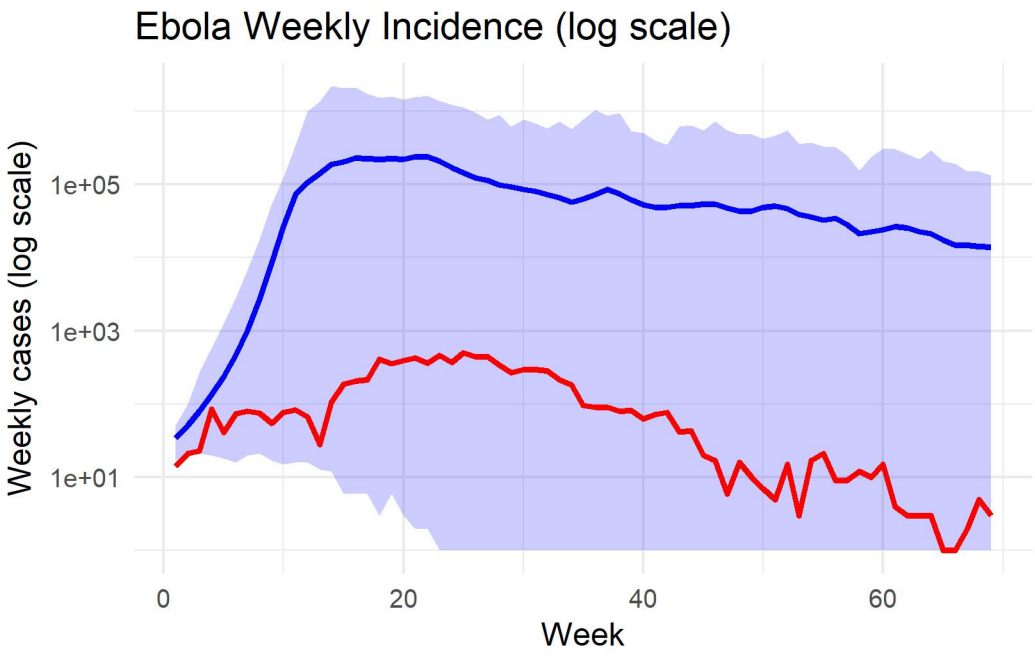

```
# A tibble: 10 x 2
    time observed
   <dbl>    <dbl>
 1     1       13
 2     2       20
 3     3       22
 4     4       84
 5     5       40
 6     6       73
 7     7       79
 8     8       74
 9     9       53
10    10       76
```

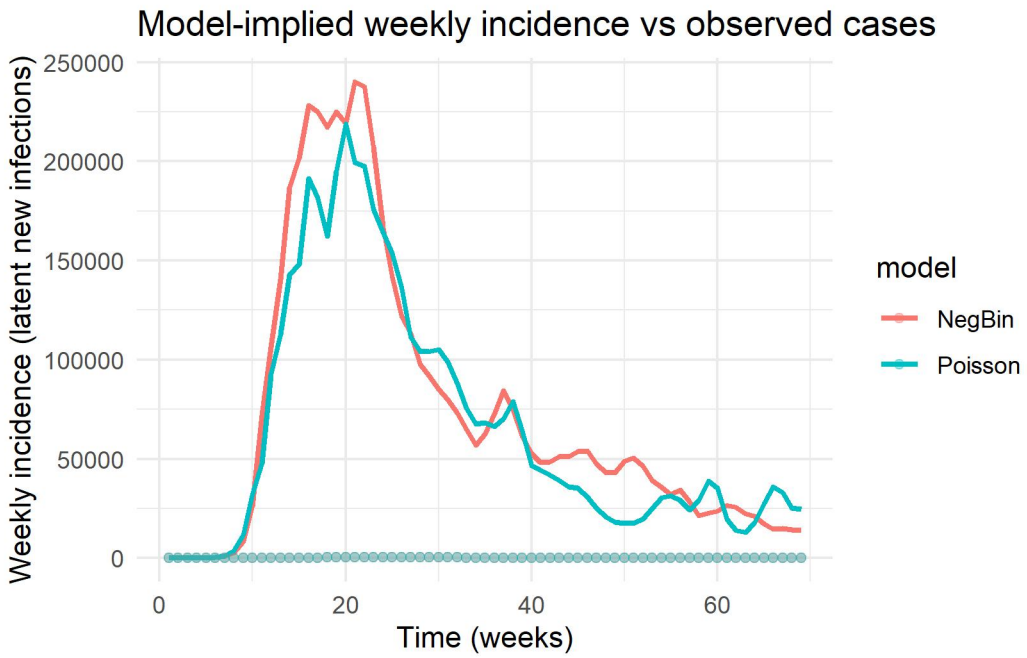

Observed vs model-implied weekly incidenc
2000000
1500000
1000000
500000
0
Cases
0
20
40
60
Week
Model mean
Observed


Estimated effective reproduction number (Rt)
10.0
7.5
5.0
2.5
0.0
Rt
0
20
40
60
Week

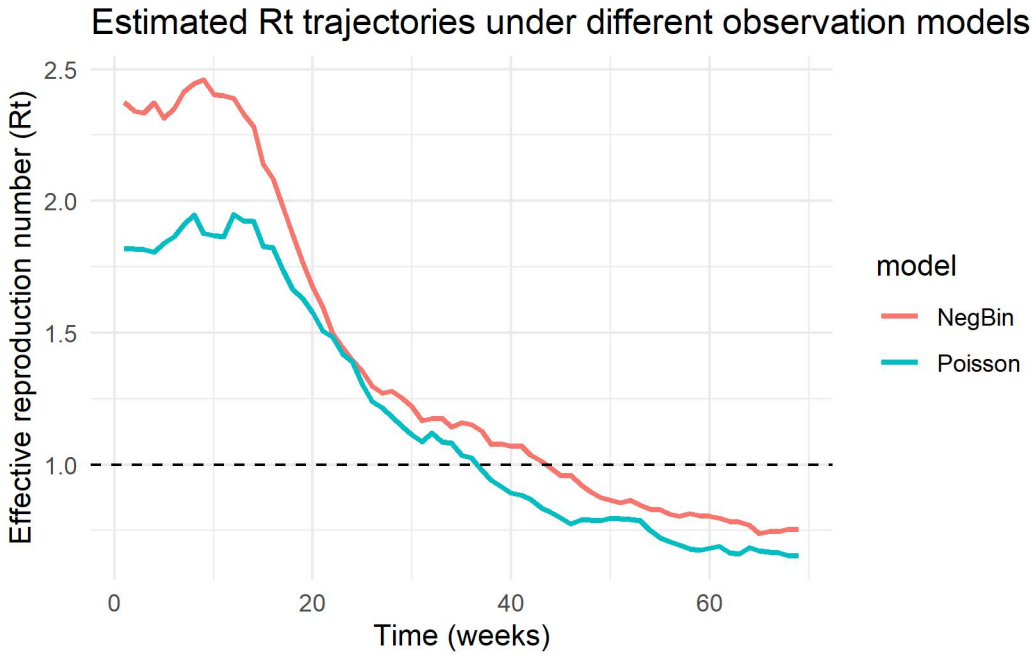


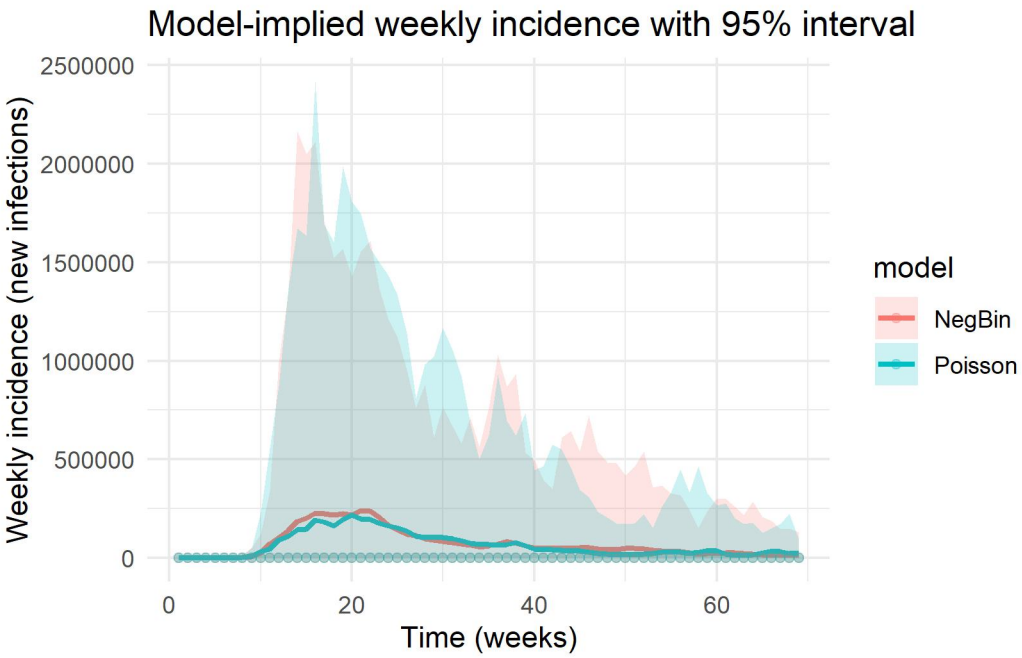
For interpretability, consider log-transformations.

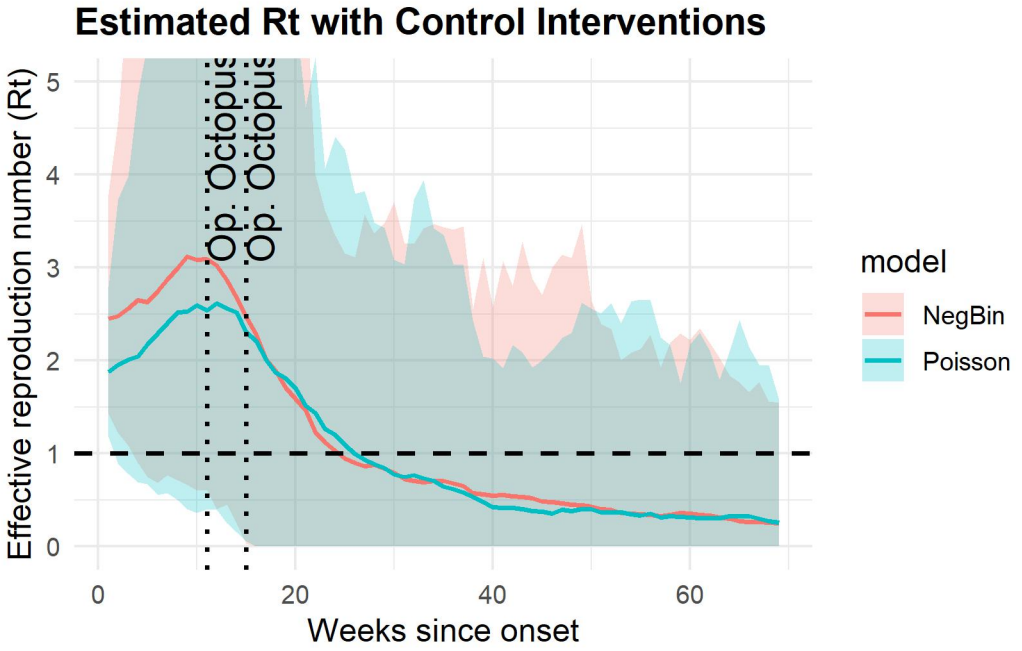





```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



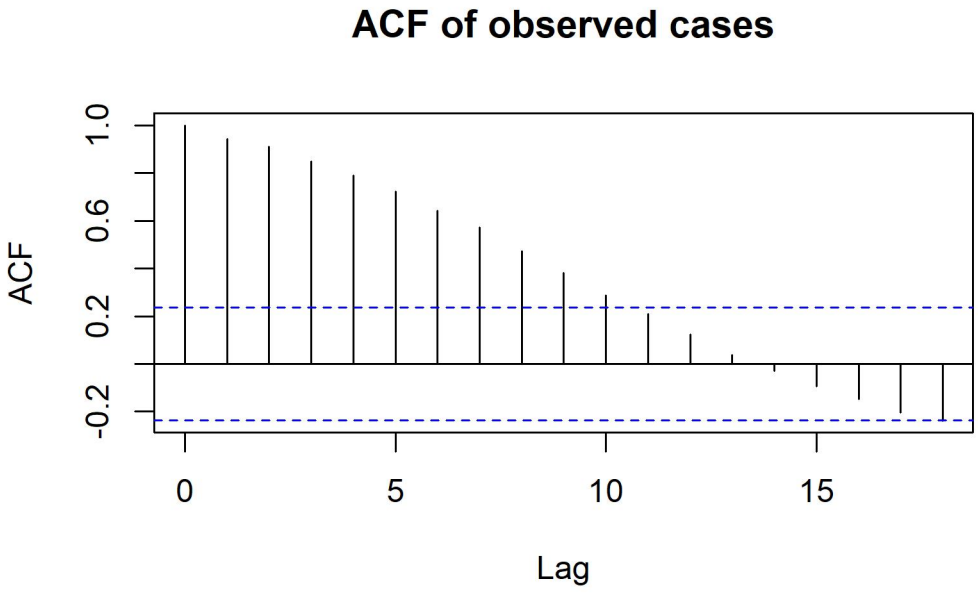


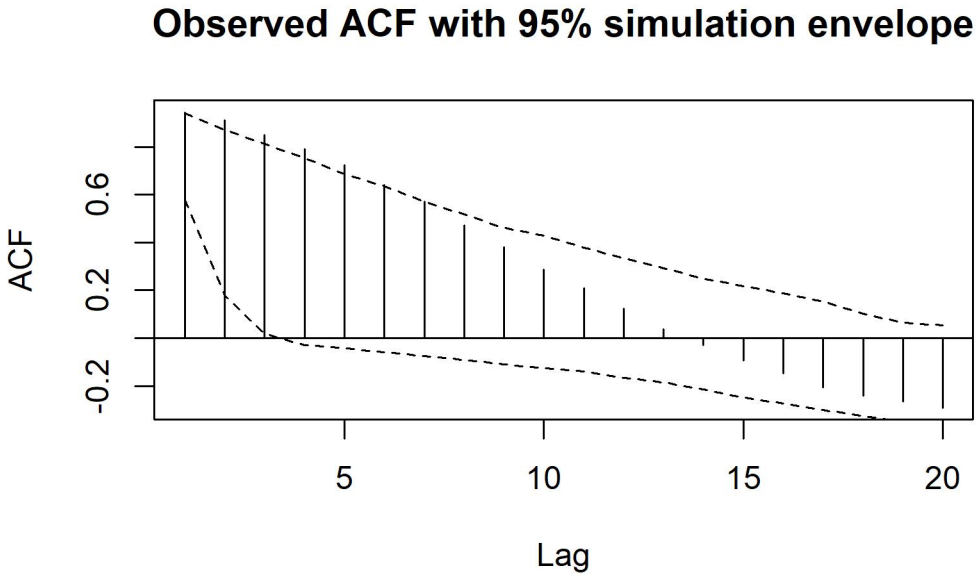


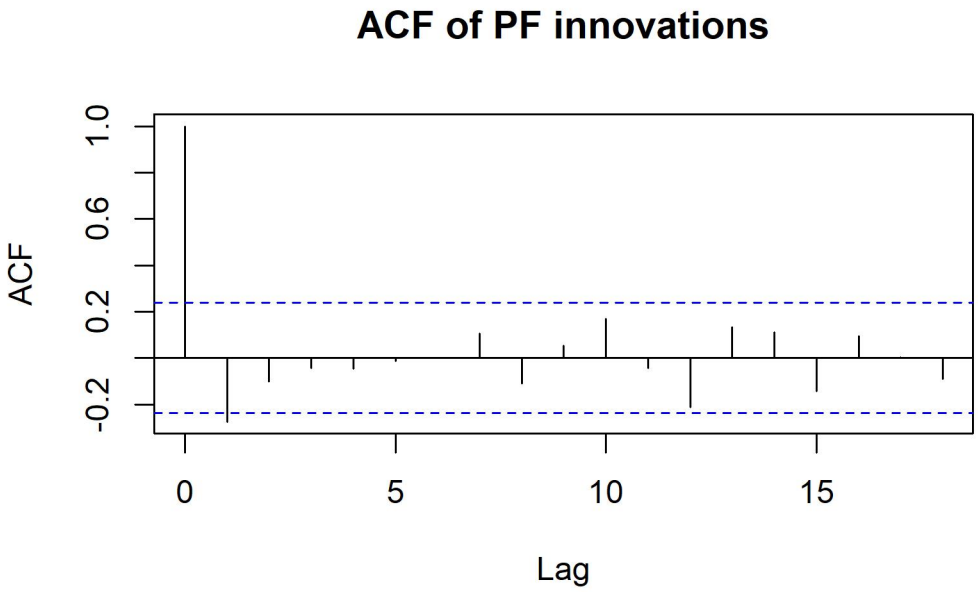
gamma	rho	k	N	logbeta0
1.316314e+00	6.225120e-01	1.267083e+10	7.000000e+06	-4.202177e-01

sigma_proc	S0	I0
1.253383e+00	6.999950e+06	5.000000e+01

```
        gamma           rho             k             N      logbeta0
-1.291582e+00  8.968548e-04  3.146846e+00  7.000000e+06 -4.202177e-01
   sigma_proc            S0            I0
-1.487902e+00  6.999950e+06  5.000000e+01
```






ACF of PF innovations
ACF
1.0
0.6
0.2
-0.2
0
5
10
15
Lag