A Stochastic SIR POMP Analysis of the 2014 Ebola Epidemic in Sierra Leone

Deepan Islam
https://github.com/atpabuser/STATS-506-repo/tree/main

December 15, 2025

1 Introduction

The 2014 Ebola virus disease (EVD) outbreak in West Africa was one of the most devastating epidemics of the 21st century, resulting in widespread morbidity, mortality, and severe economic and societal disruption. Among the affected countries, we focus on Sierra Leone because it experienced prolonged community transmission, implemented well-documented national control measures, and provides a relatively consistent time series of reported cases suitable for state-space modeling.

We analyze the 2014 Ebola epidemic in Sierra Leone using a stochastic susceptible–infectious–recovered (SIR) model formulated as a partially observed Markov process (POMP), applied to publicly available surveillance data aggregated at the national level.

The primary objectives of this analysis are twofold. First, we assess whether a stochastic SIR POMP can reproduce the observed temporal structure of the Ebola epidemic despite substantial underreporting and observation noise. Second, we examine how inferred time-varying reproduction numbers evolve over the course of the outbreak and how these estimates align temporally with the timing of major control efforts. To evaluate robustness, we perform a sensitivity analysis comparing Poisson and Negative Binomial observation models and conduct diagnostic checks to assess overall model adequacy.

2 Data

We use publicly available Ebola surveillance data from the `ebola_sierraleone_2014` dataset in the `outbreaks` R package. The dataset records confirmed Ebola cases in Sierra Leone, including the date of report and administrative district, providing a longitudinal record of the 2014 epidemic.

For modeling purposes, case counts are aggregated to the national level and summarized at a weekly frequency. Weekly aggregation aligns with the discrete-time formulation of the stochastic SIR model. National-level aggregation avoids the need to model spatial transmission explicitly while preserving the dominant temporal dynamics of the epidemic.

Observed data correspond to reported incident cases rather than disease prevalence. As a result, weekly case counts are treated as noisy observations of latent incidence generated by the underlying transmission process. Substantial underreporting and variability in reporting are expected, motivating the use of an overdispersed observation model in the POMP framework.

3 Methods

We model the transmission dynamics of the 2014 Ebola outbreak in Sierra Leone using a stochastic susceptible–infectious–recovered (SIR) model formulated as a POMP. The latent epidemic process is governed by stochastic infection and recovery dynamics, while the observation process links latent incidence to reported case counts through an overdispersed measurement model. The model is formulated in discrete time with weekly resolution and implemented using the `pomp` package in R.

3.1 Latent Process Model

Let N denote the total population size. At week t, the latent state consists of susceptible, infectious, and removed individuals, denoted by S_t, I_t, and R_t, respectively. Weekly incidence is denoted by C_t, representing the number of new infections occurring during week t. The full latent state vector is $X_t = \{S_t, I_t, R_t, C_t, \log \beta_t\}$, where β_t denotes the time-varying transmission rate.

New infections during the interval $(t, t+1]$ are modeled as $N_{SI,t} \sim \text{Binomial}(S_t,\ 1 - \exp(-\beta_t I_t/N))$, derived from a per-capita force of infection $\lambda_t = \beta_t I_t/N$. Weekly incidence is recorded as $C_{t+1} = N_{SI,t}$. Recoveries during the same interval are modeled as $N_{IR,t} \sim \text{Binomial}(I_t,\ 1 - \exp(-\gamma))$, where γ is the recovery rate. Compartment counts evolve according to $S_{t+1} = S_t - N_{SI,t}$, $I_{t+1} = I_t + N_{SI,t} - N_{IR,t}$, and $R_{t+1} = R_t + N_{IR,t}$.

3.2 Time-Varying Transmission and Reproduction Number

To accommodate temporal variation in transmission intensity, the transmission rate β_t evolves as a random walk on the log scale, $\log \beta_{t+1} = \log \beta_t + \eta_t$, with $\eta_t \sim \mathcal{N}(0, \sigma^2_{\text{proc}})$. From the estimated transmission and recovery rates, the effective reproduction number is computed as $R_t = \beta_t/\gamma$.

3.3 Observation Model

Observed weekly Ebola case counts correspond to incident infections rather than disease 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, $Y_{t+1} \sim \text{NegBinomial}(\rho C_{t+1}, k)$, where $\rho \in (0, 1)$ is the reporting fraction and $k > 0$ is a dispersion parameter. This formulation accommodates substantial reporting variability and overdispersion relative to a Poisson observation model. As a sensitivity analysis, results are compared to those obtained under a Poisson observation model.

3.4 Inference and Implementation

The stochastic process and observation models are implemented using `pomp` `rprocess` and `dmeasure` Csnippets. Model parameters and states are estimated via sequential Monte Carlo and iterated filtering (`mif2`), with uncertainty quantified using simulation-based diagnostics.

4 Results

4.1 Model Fit and Incidence Dynamics

Model-implied weekly incidence reproduces the overall temporal structure of the epidemic, including rapid early growth, peak incidence, and subsequent decline (Appendix Figure 2). On the log scale, latent incidence substantially exceeds reported case counts throughout the outbreak.

This discrepancy reflects structural modeling assumptions rather than model misspecification. In particular, latent incidence is inflated by both substantial underreporting and the assumption that the entire national population is initially susceptible, whereas transmission in practice was geographically and socially heterogeneous. The reporting fraction absorbs much of this discrepancy, allowing the model to reproduce the timing and shape of the epidemic while differing in absolute magnitude.

4.2 Time-Varying Transmission and Reproduction Number

Estimated time-varying reproduction numbers exhibit a clear decline over the course of the epidemic (Appendix Figure 1). Early in the outbreak, inferred values of R_t exceed 2, indicating sustained transmission and rapid epidemic growth. Over time, R_t declines steadily and crosses below the epidemic threshold $R_t = 1$, after which transmission is no longer self-sustaining.

The decline in R_t aligns temporally with the implementation of major national control measures, including district quarantines and the initiation of Operation Octopus. While the model does not explicitly encode intervention effects, this temporal consistency suggests a substantial reduction in transmission intensity over the course of the outbreak. Estimates obtained under Poisson and Negative Binomial observation models yield similar mean trajectories, though uncertainty intervals are wider under the Negative Binomial model.

4.3 Late-Epidemic Behavior and Identifiability

Although estimated transmission rates β_t increase late in the epidemic (Appendix Figure 13), inferred reproduction numbers remain well below one. Late-epidemic transmission parameters are weakly identified due to sparse case counts, and variation in β_t during this period primarily reflects uncertainty rather than renewed epidemic growth.

4.4 Model Diagnostics and Sensitivity Analysis

Autocorrelation diagnostics indicate that the fitted model adequately captures the temporal dependence in the data. The observed autocorrelation structure is well reproduced by simulations from the fitted model, and particle filter innovations exhibit no significant residual autocorrelation (Appendix Figures 3–4).

Sensitivity analysis comparing Poisson and Negative Binomial observation models shows that qualitative conclusions regarding transmission dynamics are robust to assumptions about reporting noise. Both models recover similar mean trajectories for R_t and identify the same periods of declining transmission. However, the Poisson model produces systematically narrower uncertainty intervals, particularly during periods of high incidence and late in the epidemic, whereas the Negative Binomial model yields wider intervals that better reflect variability in reported case counts. These differences are reflected primarily in uncertainty quantification rather than in the estimated mean transmission dynamics.

5 Conclusion

This analysis shows that a stochastic SIR model formulated within a POMP framework can capture the temporal dynamics of the 2014 Ebola epidemic in Sierra Leone despite substantial underreporting and observation noise. By separating latent transmission from the reporting process, the model yields plausible time-varying reproduction numbers and reproduces key features of the epidemic trajectory.

Estimated reproduction numbers decline over time and remain below one during the later stages of the outbreak. Sensitivity analyses indicate that qualitative conclusions regarding transmission dynamics are robust to assumptions about reporting noise, with differences between observation models primarily affecting uncertainty quantification.

Overall, these results illustrate the usefulness of stochastic state-space models for inference in partially observed epidemic systems.

References

[1] Na, W., Park, N., Yeom, M., and Song, D. (2015). Ebola outbreak in Western Africa 2014: what is going on with Ebola virus? *Clinical and Experimental Vaccine Research*, 4(1), 17–22. doi:10.7774/cevr.2015.4.1.17.

[2] Centers for Disease Control and Prevention (CDC). (2014). Update: Ebola Virus Disease Outbreak — West Africa, October 2014. *MMWR Morbidity and Mortality Weekly Report*. Available at https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6343a3.htm.

[3] King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical Inference for Partially Observed Markov Processes via the R Package `pomp`. *Journal of Statistical Software*, 69(12), 1–43. (Updated PDF) https://kingaa.github.io/pomp/vignettes/pompjss.pdf.

[4] King, A. A. (n.d.). Iterated filtering: principles and practice. Online tutorial. https://kingaa.github.io/short-course/mif/mif.html.

[5] World Bank Staff. (2021). *An Introduction to Deterministic Infectious Disease Models*. World Bank. https://documents1.worldbank.org/curated/en/888341625223820901/pdf/An-Introduction-to-Deterministic-Infectious-Disease-Models.pdf.

[6] Blackwood, J. C. and Childs, L. M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*, 5(1), 195–221. doi:10.1080/23737867.2018.1509026. Available at https://www.stat.cmu.edu/~kass/covid/SEIRmodelingINTRO.pdf.

[7] ETH Zurich, Theoretical Biology. (n.d.). Stochastic simulation of epidemics. Online learning module. https://tb.ethz.ch/education/learningmaterials/modelingcourse/level-2-modules/stochSIR.html.

A Additional Figures and Diagnostics

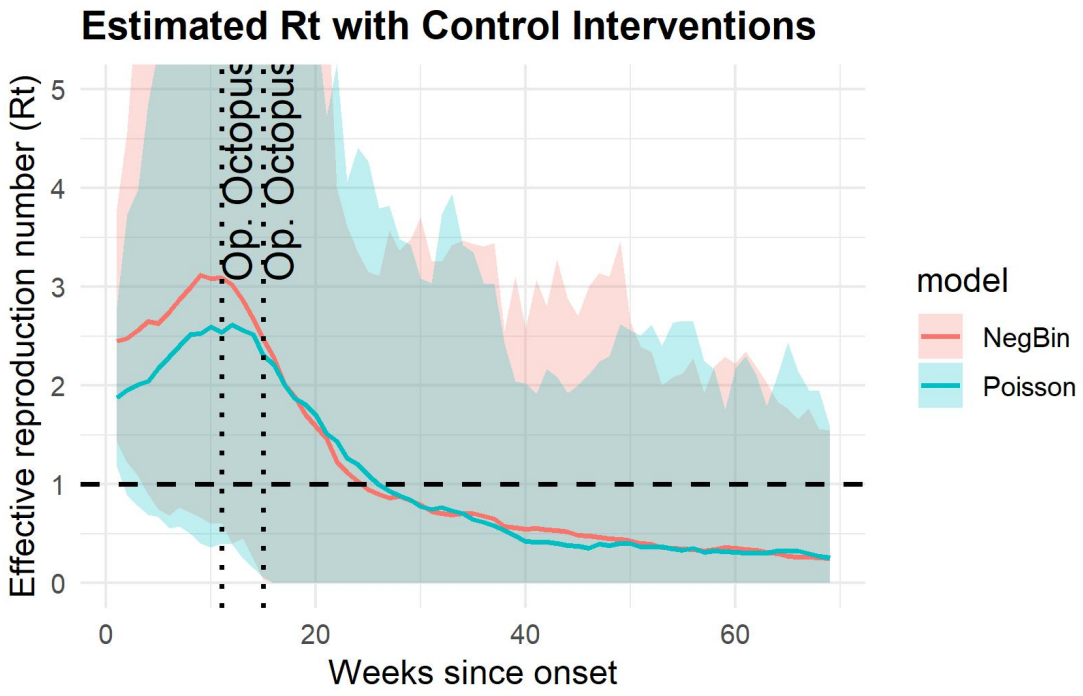

Figure 1: Estimated R_t trajectories under Poisson and Negative Binomial observation models. Vertical lines indicate the timing of major control interventions (e.g., Operation Octopus).

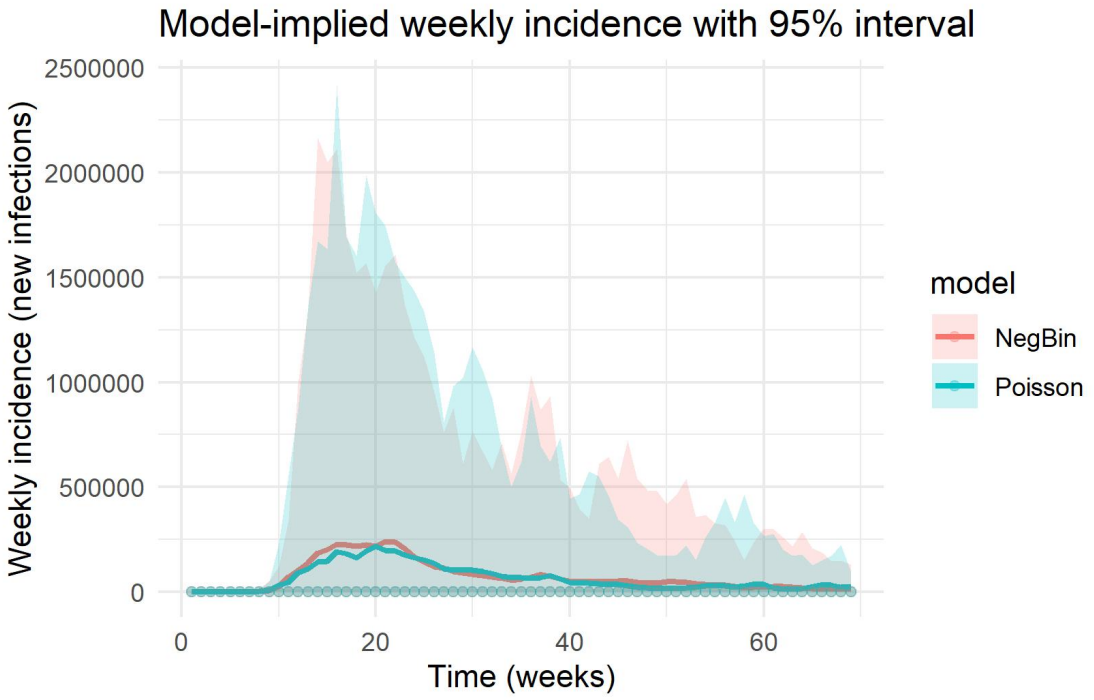

Figure 2: Observed weekly Ebola cases compared to model-implied weekly incidence with 95% uncertainty intervals.

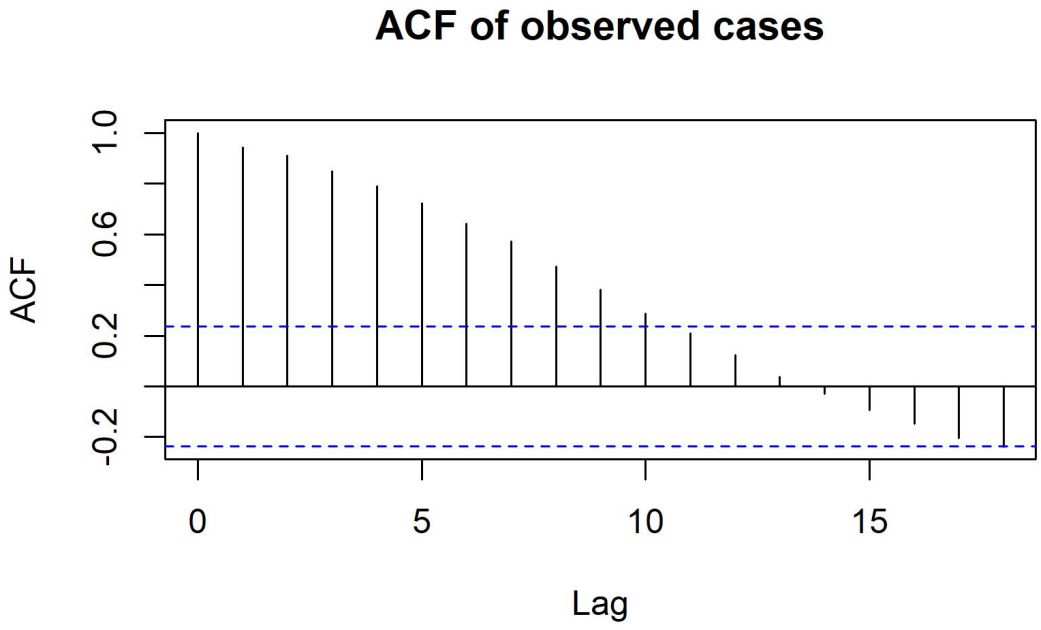

Figure 3: Autocorrelation function of observed weekly Ebola case counts.

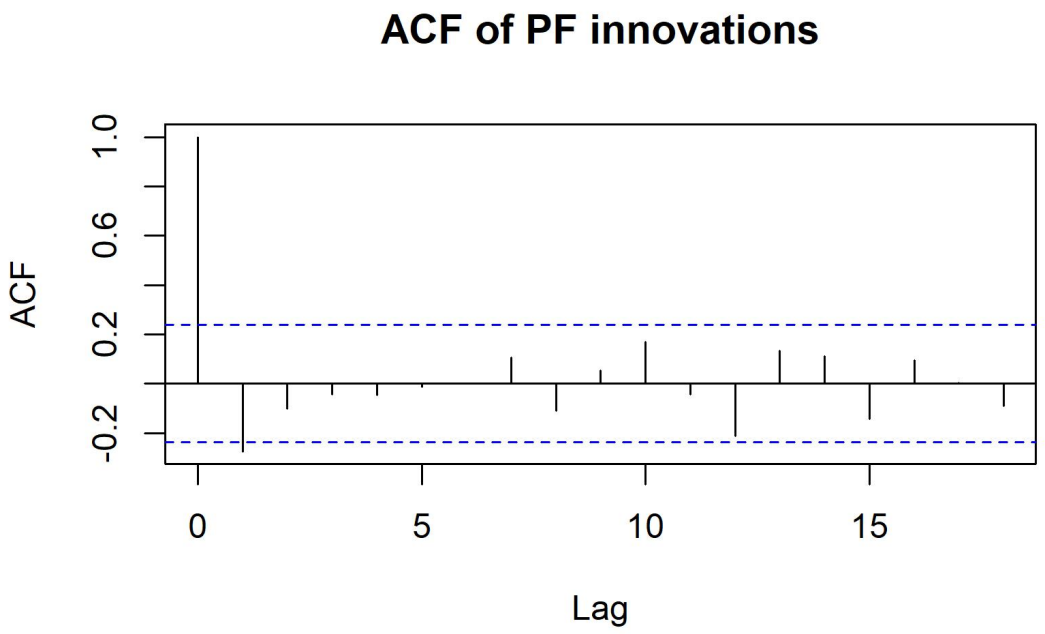

Figure 4: Autocorrelation function of particle filter innovations.

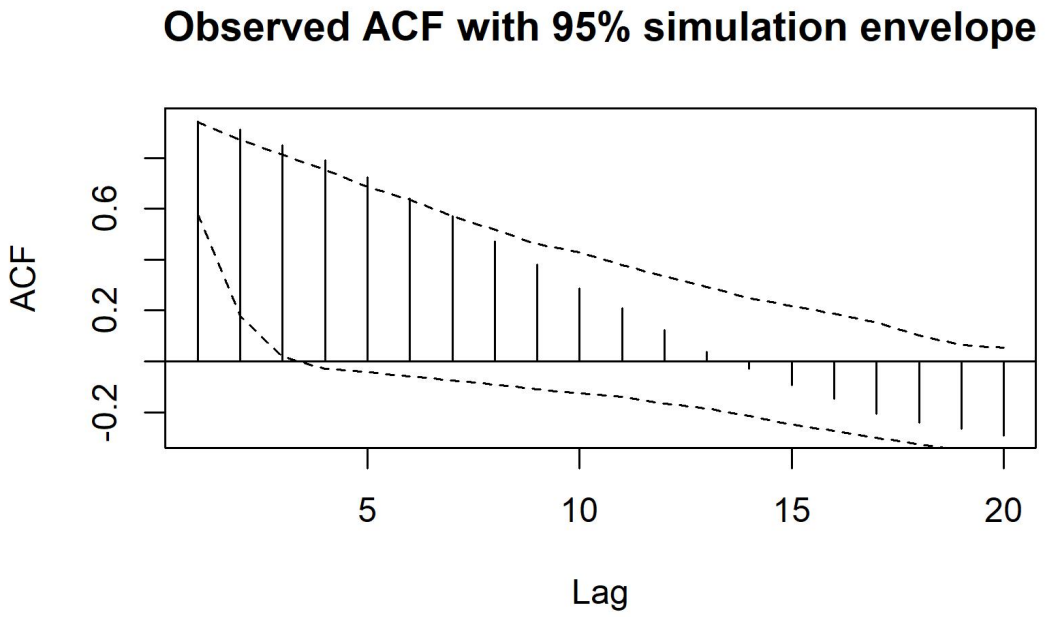

Figure 5: Observed autocorrelation function with 95% simulation-based envelope from the fitted stochastic SIR POMP.

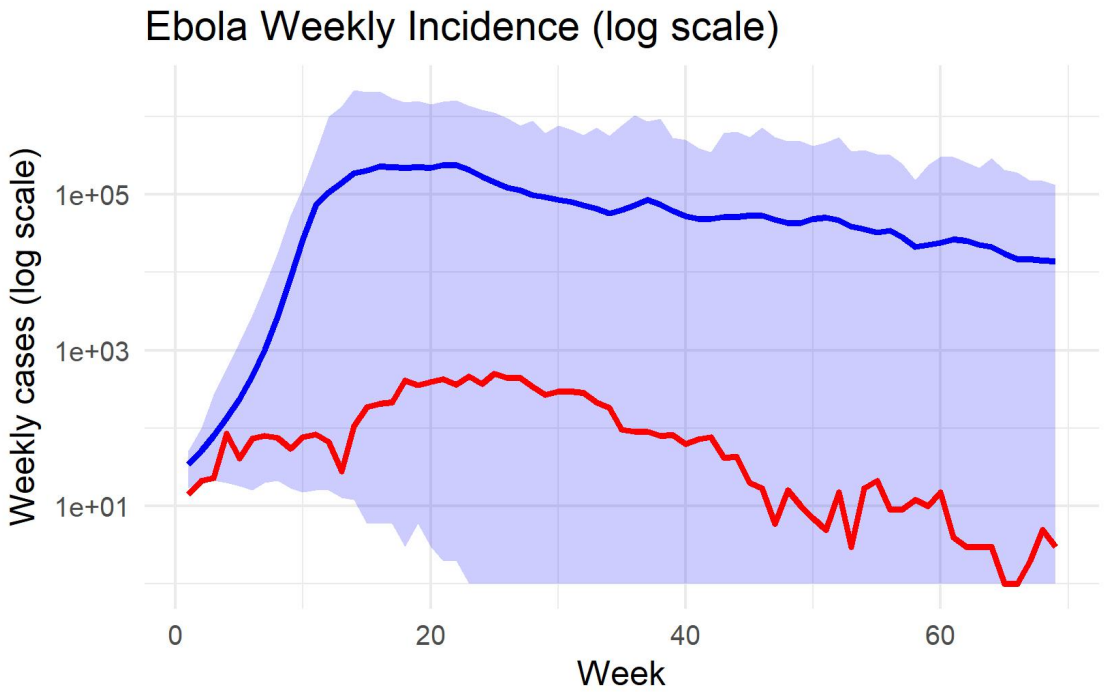

Figure 6: Weekly Ebola incidence on the log scale: observed data and model-implied trajectory.

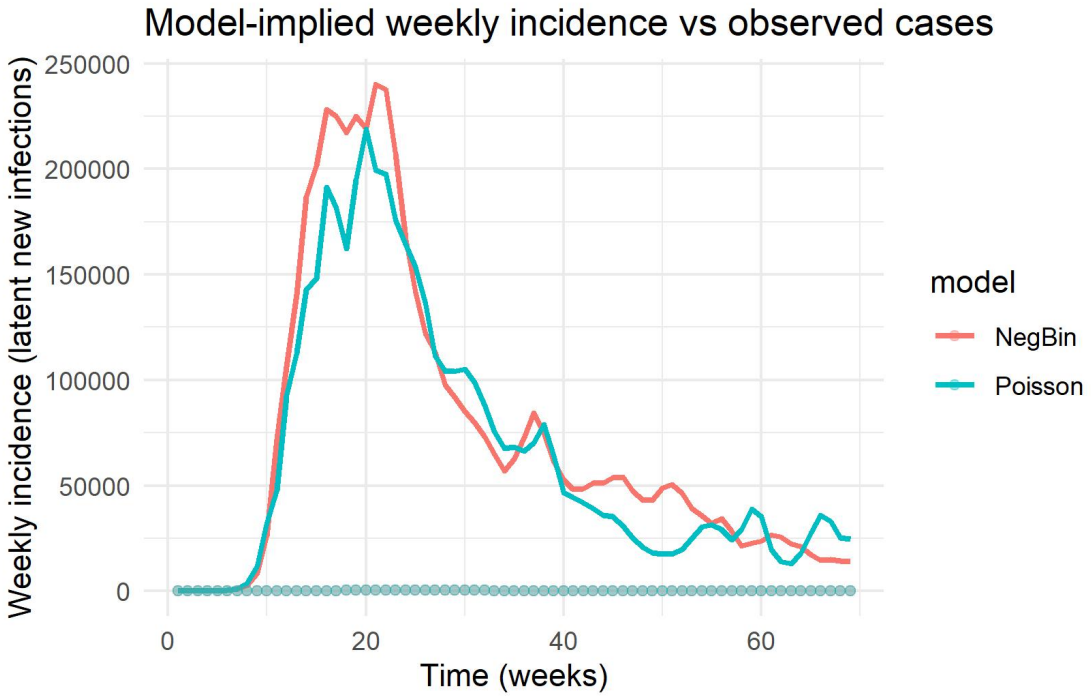

Figure 7: Model-implied weekly incidence compared to observed cases under both observation models.

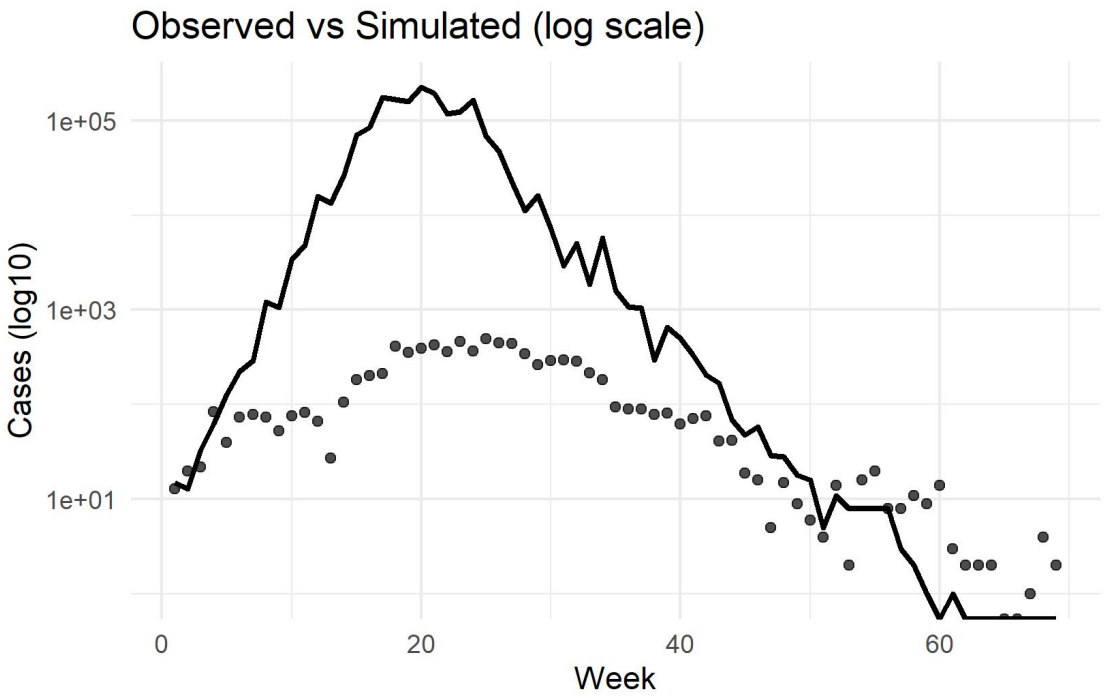

Figure 8: Observed versus simulated weekly incidence under the Negative Binomial observation model (log scale).

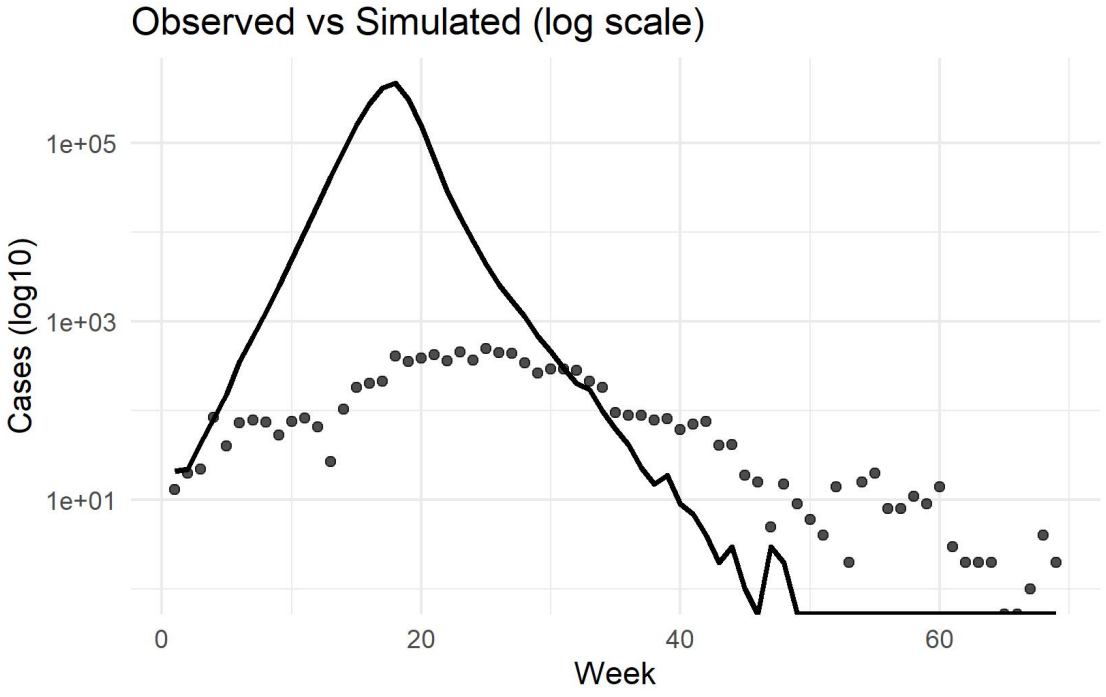

Figure 9: Observed versus simulated weekly incidence under the Poisson observation model (log scale).

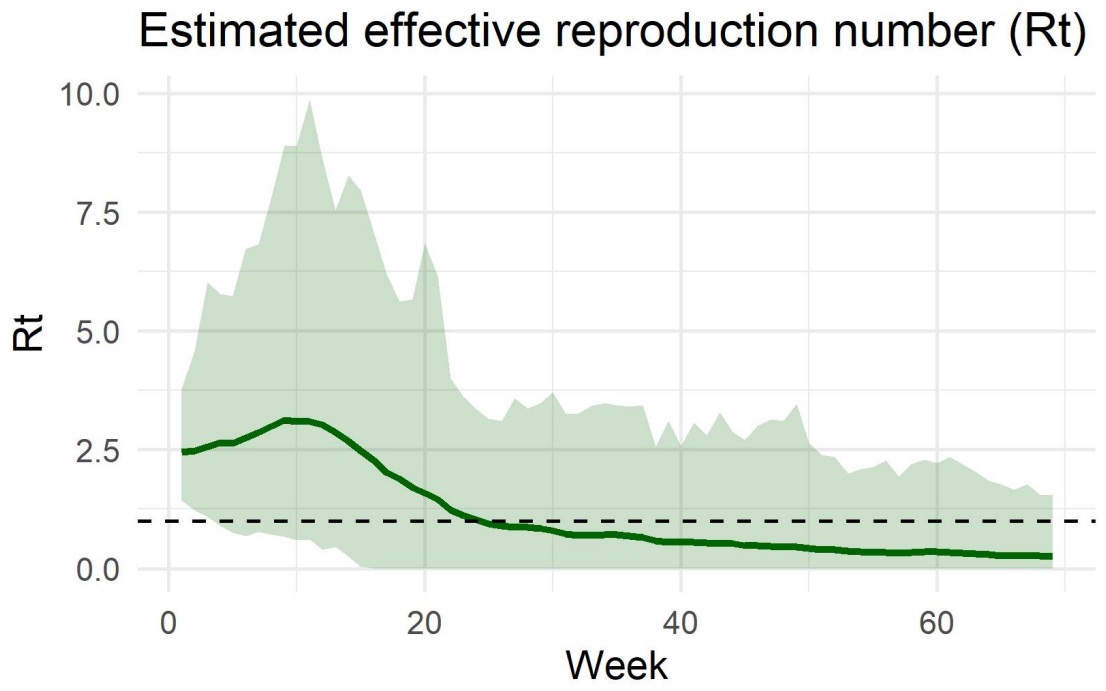

Figure 10: Estimated effective reproduction number R_t per week with uncertainty band.

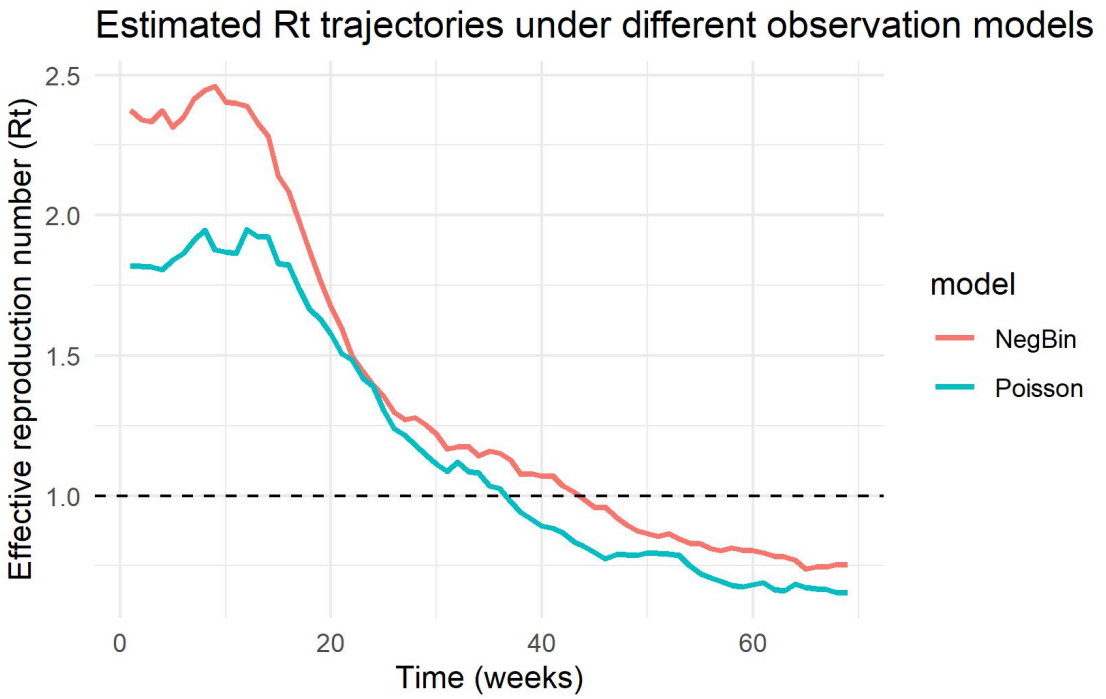

Figure 11: Estimated R_t trajectories under Poisson and Negative Binomial observation models.

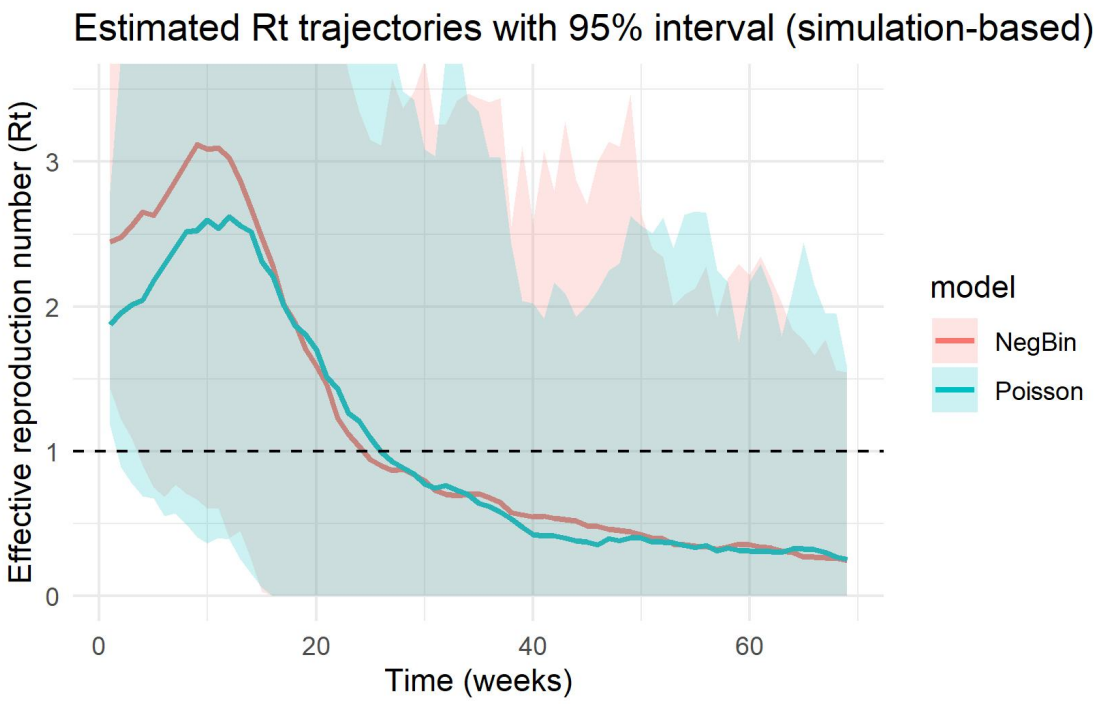

Figure 12: Estimated R_t trajectories with 95% simulation-based intervals.

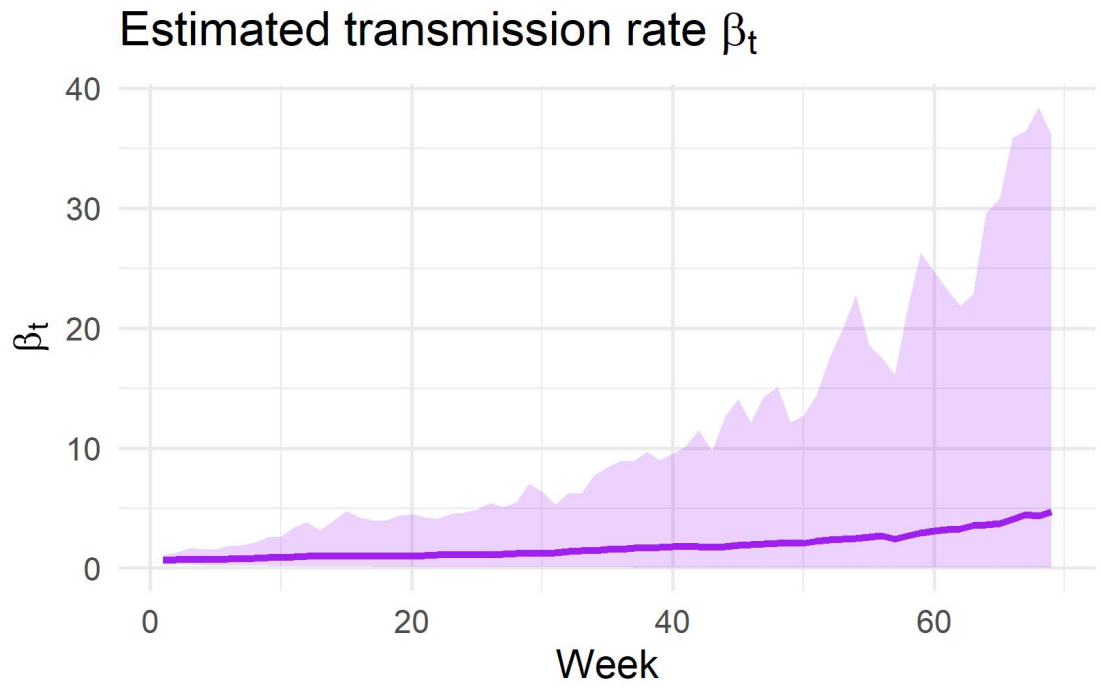

Figure 13: Estimated time-varying transmission rate β_t with uncertainty band.

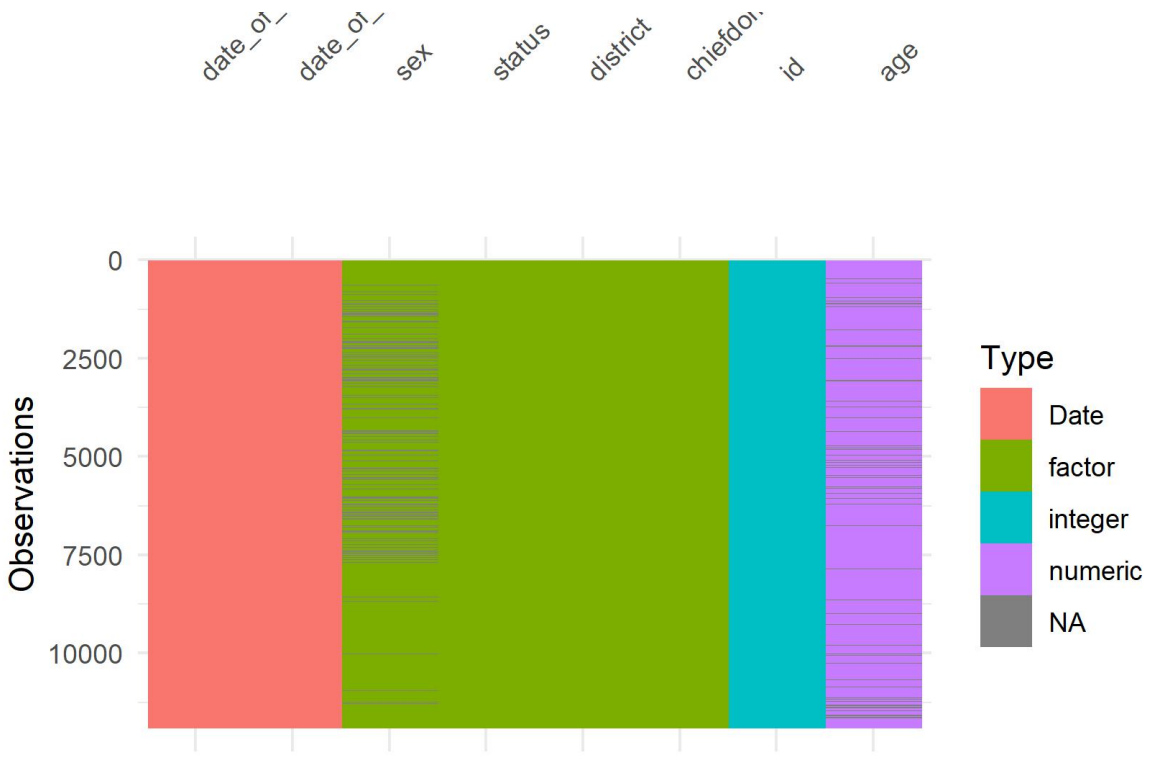

Figure 14: Missingness structure in the raw Ebola dataset.

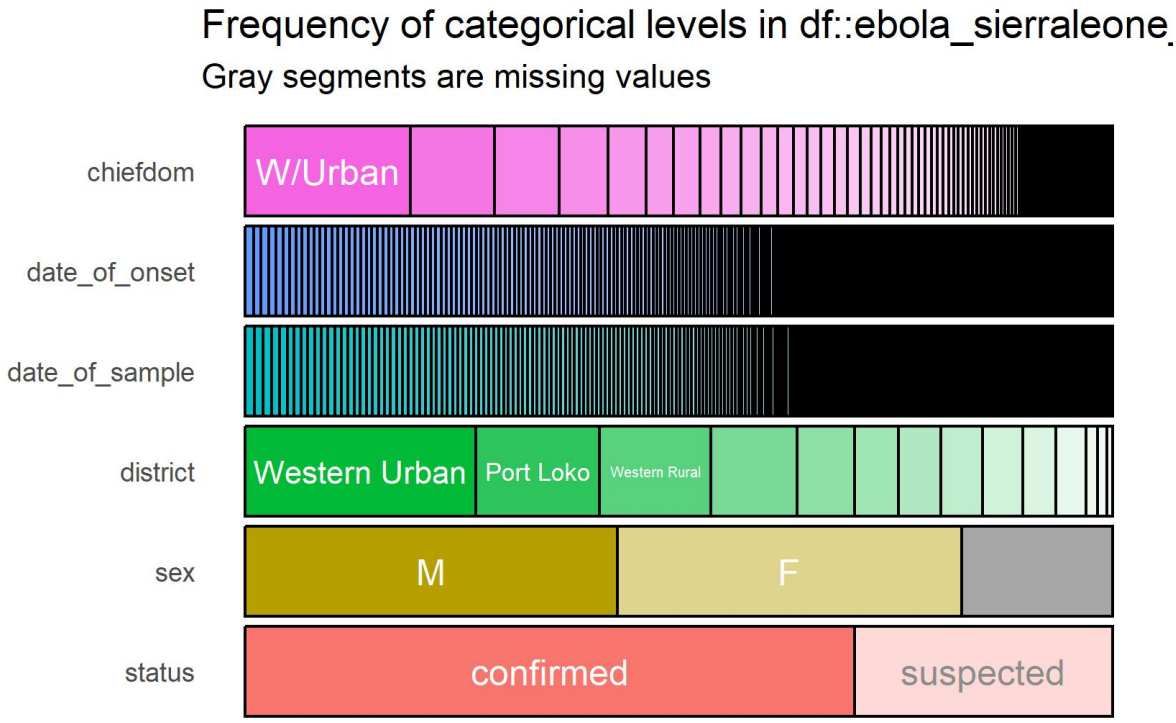

Figure 15: Frequency of categorical variable levels in the Sierra Leone Ebola dataset.

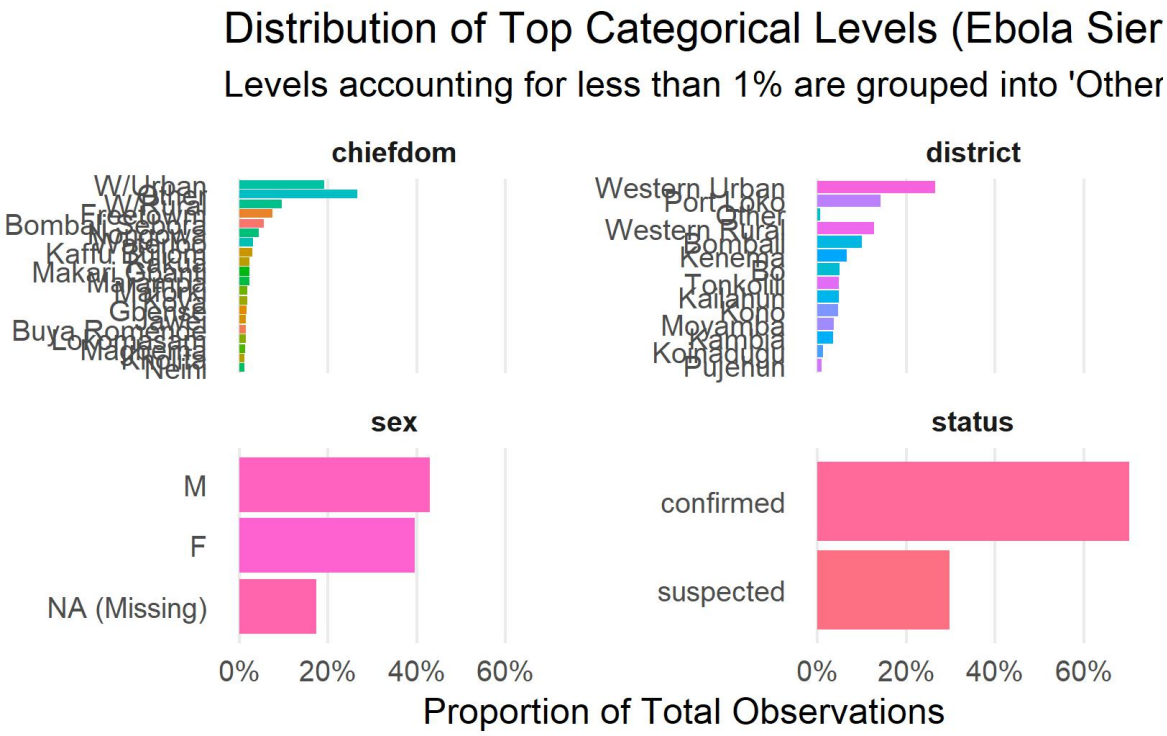

Figure 16: Distribution of the most frequent categorical variables in the dataset.