Pandemic Flu Simulation

Nathan Kurtz-Enko

Table of contents

Abstract

- Problem
- Previous approaches
- Novelty
- Key findings

Background

Understanding the spread of infectious diseases, like the flu, is critical for advising policy making with respect to public health and wellness. For example, studying the behavior of virus transmission and recovery rates allows health professionals and policymakes to define the quantity of antiviral agents that should be readily available, regions and demographics that should be prioritized for vaccination, and other strategies intended to prevent a pandemic. Furthermore, being able to communicate the results of an infectious disease model/simulation is very important, as this will facilitate the adoption of containment measures.

Many researchers studying the spread of infectious diseases explicitly cite the 1918 influenza outbreak as a motivating factor for their work (Berkessel et al. 2021; Germann et al. 2006; Johnson and Mueller 2002; Valle, Mniszewski, and Hyman 2012). In the case of Johnson and Mueller (2002) and Patterson and Pyle (1991), the objective was to look back through previous records and data to verify morbitity of the 1918 pandemic. The authors made the observation that each time the statistics are reviewed, the total deaths from the pandemic seem to increase. In the 1920s, the record was approximately 21.5 million deaths; in 1991, Patterson and Pyle (1991) updated the estimate to 24.7-39.3 million; most recently, Johnson and Mueller (2002) estimated the death toll is on the order of 50 million. However, Johnson and Mueller (2002) does not make this claim definitively. Instead, Johnson allows for speculation for future estimates because there are many places affected by the 1918 influenza outbreak for which we do not have any data or the means of estimating today.

Clearly, the capacity for large-scale damage from infectious disease outbreaks is a very real threat. Hence, understanding how these things spread and the impact that different mitigation strategies have on the rate of spread and severity of illness is critical.

Literature Review

Infectious disease epidemic models have been the subject of study since the 1980s. It was Russian scientists that first began to implement models for estimating spread patterns of influenza and other viruses (Longini, Fine, and Thacker 1986). Much of that work is relatively inaccessible due to the fact that it is written in Russian. Nonetheless, eventually their work made its way into academia across the globe.

Generally, there are two approaches for modeling pandemic flu: deterministic or stochastic models. Early models were deterministic, and they described the movement of viruses using differential equations and partitioned populations into a number of subgroups. These are referred to as SIR (susceptible-infected-recovered), SEIR (susceptible-exposed-infected-recovered), and other derivatives.

Initial reactions to these determinsitic models by academics were skeptical for a few reasons. First, geographic regions being studied were relatively isolated, hence the ability to apply these models for global study was uncertain (Longini, Fine, and Thacker 1986). Furthermore, for the

time, the models were computationally expensive. Longini, Fine, and Thacker (1986) noted that over 90% of time was dedicated to running the algorithms themselves. However, this is less of an issue today because high-performance computing is readily available.

While relatively old, these determinstic models still have applications today. Valle, Mniszewski, and Hyman (2012) used an SIR model to show that an individual's behavior will likely change in response to an outbreak; some individuals will adopt new behavior to protect themselves, thereby influencing the probability of transmission per contact.

More recently, stochastic models have become prevelant because they are able to capture more complex spreading patterns. Colizza et al. (2007), Cooper et al. (2006), Ferguson et al. (2005), Germann et al. (2006), Samsuzzoha, Singh, and Lucy (2013), and Tan et al. (2021) all used some stochastic process to model the spread of influenza or other infectious diseases and analyze different mitigation strategies. Colizza et al. (2007) found that a large-scale application of prophylaxis antiviral agents could successfully contain a pandemic H5N1 influenza given a viral reproduction ratio of 1.9 or lower. Cooper et al. (2006) found that significant air travel restrictions could delay the spread of influenza, but not contain an epidemic altogether. The findings of Germann et al. (2006) determined that rapid and preferential vaccination of children could contain an outbreak when the reproduction ratio of influenza is at or below 1.9.

Others have explored the relationship between geographic variables (e.g., altitude and solar radiation) and incidence rates for COVID-19, as well as severity of illness (Arias-Reyes et al. 2021; Stephens, Chernyavskiy, and Bruns 2021). Arias-Reyes et al. (2021) concluded that above 1000 meters above sea level, incidence rates and severity for COVID-19 decline. Through analysis of public geographic and COVID-19 data, Stephens, Chernyavskiy, and Bruns (2021) determined that for increases of 495 meters in elevation above LA county, and for areas of equivalent population density, infection rates were 12.82%, 12.01%, and 11.72%. However, mortality rates were not statistically significantly different between high and low elevation areas. Stephens, Chernyavskiy, and Bruns (2021) notes that other environmental variables such as increased solar radiation, larger swings of high/low temperatures throughout the day, and long-term exposure to altitude hypoxia could explain these observed differences.

Since influenza and similar viruses spread via aerosols, Sagripanti and Lytle (2007) proposed solar radiation as a contributing factor to the seasonality of influenza. Specifically, the authors calculated the expected inactivation of influenza by UV radiation in several cities and different times of the year. They estimated that a full day of sunlight exposure will reduce influenza by 99% in regions at a similar latitude to Mexico City, and by 90% in regions at a similar latitude to Miami. However, Weber and Stilianakis (2008) note that the impact on survivability of influenza by humidity and temperature are as significant as solar radiation. Furthermore, it is unlikely that outdoor transmission is a significant mechanism for the spread of influenza because it is so sensitive to wide range of environmental variables.

Regardless, it is clear that higher altitude and higher solar radiation could theoretically influence the spread of influenza. The purpose of this paper will be to simulate pandemic flu

spread and identify critical thresholds for these covariates in order to minimize the risk of a severe outbreak.

Modeling Theory

As previously mentioned, there are two very general approaches to simulating pandemic virus spread. We will be discussing the SIR model (Débarre, n.d.; "An Introduction to Deterministic Infectious Disease Models," n.d.) because it is relatively simple, and establishes a foundation for considering other approaches.

The SIR model categorizes a population into three groups: susceptible, infected, and recovered. The rate of change of each of these groups is defined using differential equations, and we can account for mortality and birth rates in their definitions as well.

$$\begin{aligned}\frac{dS}{dt} &= \delta - \frac{\beta \times S \times I}{N} - \mu \times S \\ \frac{dI}{dt} &= \frac{\beta \times S \times I}{N} - (\gamma + \mu) \times I \\ \frac{dR}{dt} &= \gamma \times I - \mu \times R\end{aligned}$$

Here, S, I, R stand for the susceptible, infected, and recovered populations respectively. The total population is denoted by N, and the rates of transmission, recovery, mortality, and birth are denoted by β, γ, μ, and δ.

The outbreak status is described by the reproduction ratio, R_0, which is the ratio between the transmisison rate and recovery rate. When $R_0 = \beta/\gamma > 1$, more people are becoming infected than are recovering, hence the system is in a state of outbreak. And, supposing that there is some percentage of the population that is already vaccinated (i.e., already in the recovered group), this ratio becomes $(\beta \times S(0))/(\gamma \times N)$, where $S(0)$ is the number of people in the susceptible group at time $t = 0$.

These facts make it easy to define different transmission, recovery, and vaccination target rates to avoid a significant outbreak. Hence, the SIR is a simple model, but effective at clearly communicating actionable results to policy makers and authority figures.

Method

In this analysis, we used Epiverse-TRACE (Gupte, Eggo, and Van Leeuwen 2025; Lambert, Kucharski, and Tamayo 2025; Gupte, Van Leeuwen, and Kucharski 2025), Tidyverse (Wickham et al. 2025; Wickham and Henry 2025; Wickham, Vaughan, and Girlich 2025; Wickham, Hester, and Bryan 2024; Wickham 2016), and other tools (Rudis 2022; Walker and Herman

2025; Xie 2025) in the R programming language. The objective was to use contact and population estimates to simulate the spread of influenza in the United States of America.

Data

We used contact data collected by Breen, Mahmud, and Feehan (2022) for different age groups across the United States of America. Additionally, we accessed U.S. Census results from 2022 via Walker and Herman (2025) in order to estimate populations for each of the age groups defined by Breen, Mahmud, and Feehan (2022). Census population estimates for contact data age groups were not provided by Walker and Herman (2025). However, we aggregated the state-wide counts for all ages ranging from one to 100 in bins defined in the contact data. The data are displayed in Table 1 and Table 2.

Table 1: USA National Contact Matrix

	[0,18)	[18,25)	[25,35)	[35,45)	[45,55)	[55,65)	[65,100]
[0,18)	47402.285	24645.765	31405.06	43077.563	25249.87	9696.327	5821.438
[18,25)	10550.575	43999.421	15290.51	9758.546	10319.81	5849.805	3840.709
[25,35)	19076.694	21417.483	55106.90	23122.012	17274.64	13386.274	6885.919
[35,45)	23716.460	12519.840	21174.81	45537.642	20427.31	11326.252	9290.538
[45,55)	14583.902	13840.380	16543.69	21371.308	34937.29	17599.507	9497.310
[55,65)	5625.521	7913.416	12956.29	11994.823	17783.26	20681.963	12011.882
[65,100]	4015.851	6154.390	7929.90	11646.666	11359.92	14203.840	27637.752

Table 2: USA National Population Estimates

Age Group	Population
[0,18)	72450827
[18,25)	31328131
[25,35)	45501300
[35,45)	43695365
[45,55)	40431645
[55,65)	42085437
[65,100]	57794852

Initial Conditions

There are two approaches to defining initial conditions (e.g., number of members in each model sub-group at initialization, influenza spread parameters, etc.) for the model: define

some reasonable values or derive values from current/historic flu data. For the sake of brevity, we will use the former approach and use values for rates discussed in literature.

Nikbakht, Baneshi, and Bahrampour (2018) provide multiple estimates for the basic reproduction ratio of influenza in the United states given 2017-2018 data. Specifically, using Maximum Likelhood Estimation, they estimate $R_0 = 1.8$ with a 95% confidence interval of $[1.78, 1.8]$. While this estimate is approaching seven years old, it should be sufficient for the purpose of this paper. Hence, we used this figure for our simulation.

Virlogeux et al. (2015) provide an estimate for the incubation period of influenza. Specifically, they note that incubation period follows a Weibull distribution with shape = 2.101 and scale = 3.839. We used the expected value of the incubation period $E[\text{Incubation Period}] = \text{scale} \times \Gamma(1 + \frac{1}{\text{shape}}) = 3.400167$ days. Tang et al. (2020) note that the inverse incubation period is equivalent to the infection rate, hence $\alpha = 0.2941032$.

According to the CDC, the infectious period is about five to seven days typically. For the purpose of this paper, we assumed a worst case scenario and uesd an infectious period of seven days. Turinici (2025) states that the inverse of the mean infectious period is equivalent to the recovery rate, hence $\gamma = 0.1428571$.

The final parameters are collected and dispayed in Table 3.

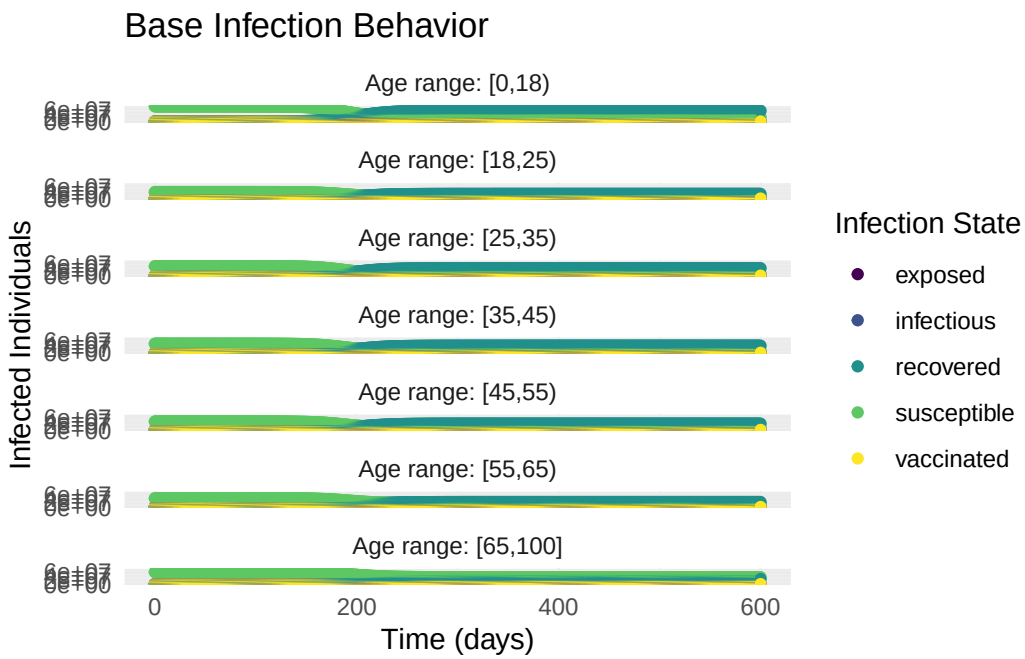
Table 3: Parameters for SEIRV Model

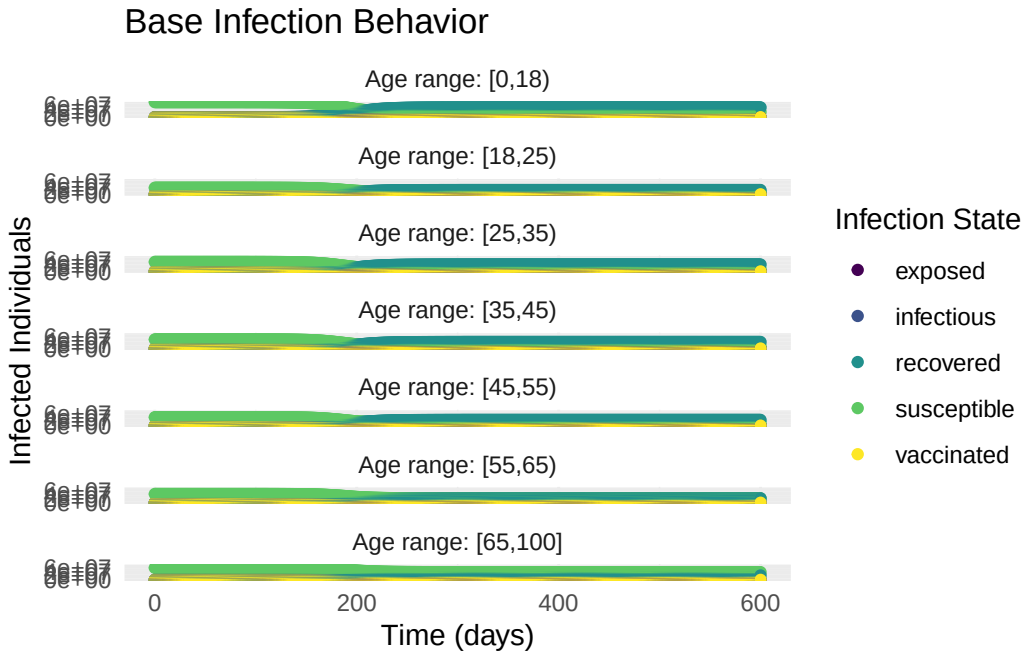
Name	Symbol	Value
Reproduction Ratio	R_0	1.8000000
Infection Rate	α	0.2941176
Recovery Rate	γ	0.1428571
Transmission Rate	$	
beta$	0.2571429	

In order for the flu to progress through a population, there must be some members in the infected group at the start of the simulation. So, we assumed that 0.0001% of the population is already infected. Furthermore, we assumed that this proportion of infected individuals was consistent across all age groups.

Simulation Results

We are using a Susceptible-Exposed-Infected-Recovered-Vaccinated (SEIRV) model. We are comparing a base output in which there are no intervention measure taken with alternative outcomes in which particular interventions are done. The objective is control the spread of influenza and minimize the final size of the pandemic.


Base Infection Behavior
Age range: [0,18)
Age range: [18,25)
Age range: [25,35)
Age range: [35,45)
Age range: [45,55)
Age range: [55,65)
Age range: [65,100]
Infected Individuals
0
200
400
600
Time (days)
Infection State
exposed
infectious
recovered
susceptible
vaccinated


Base Infection Behavior
Age range: [0,18)
Age range: [18,25)
Age range: [25,35)
Age range: [35,45)
Age range: [45,55)
Age range: [55,65)
Age range: [65,100]
Infected Individuals
0
200
400
600
Time (days)
Infection State
exposed
infectious
recovered
susceptible
vaccinated

Discussion

Future Work

- Underexplored covariates or regions.
- Need for interdisciplinary models combining epidemiology, climatology, and public health.
- Potential of machine learning to enhance traditional SIR frameworks.

Conclusions

- Summary of key findings.
- Implications for pandemic preparedness and policy.
- Call for more nuanced, data-rich modeling approaches.

References

"An Introduction to Deterministic Infectious Disease Models." n.d. https://documents1.worldbank.org/curated/en/888341625223820901/pdf/An-Introduction-to-Deterministic-Infectious-Disease-Models.pdf.

Arias-Reyes, C., F. Carvajal-Rodriguez, L. Poma-Machicao, F. Aliaga-Raduán, D. A. Marques, N. Zubieta-DeUrioste, R. A. Accinelli, et al. 2021. "Decreased Incidence, Virus Transmission Capacity, and Severity of COVID-19 at Altitude on the American Continent." *PLOS ONE* 16 (3): e0237294. https://doi.org/10.1371/journal.pone.0237294.

Berkessel, J. B., T. Ebert, J. E. Gebauer, T. Jonsson, and S. Oishi. 2021. "Pandemics Initially Spread Among People of Higher (Not Lower) Social Status: Evidence from COVID-19 and the Spanish Flu." *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506211039990.

Breen, C. F., A. S. Mahmud, and D. M. Feehan. 2022. "Novel Estimates Reveal Subnational Heterogeneities in Disease-Relevant Contact Patterns in the United States." *PLOS Computational Biology* 18 (12): e1010742. https://doi.org/10.1371/journal.pcbi.1010742.

Colizza, V., A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani. 2007. "Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions." *PLoS Medicine* 4 (1): e13. https://doi.org/10.1371/journal.pmed.0040013.

Cooper, B. S., R. J. Pitman, W. J. Edmunds, and N. J. Gay. 2006. "Delaying the International Spread of Pandemic Influenza." *PLoS Medicine* 3 (6): e212. https://doi.org/10.1371/journal.pmed.0030212.

Débarre, F. n.d. "SIR Models of Epidemics Level 1 Module in 'Modelling Course in Population and Evolutionary Biology'." https://ethz.ch/content/dam/ethz/special-interest/usys/ibz/theoreticalbiology/education/learningmaterials/701-1424-00L/sir.pdf.

Ferguson, N. M., D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. 2005. "Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia." *Nature* 437 (7056): 209–14. https://doi.org/10.1038/nature04017.

Germann, T. C., K. Kadau, I. M. Longini, and C. A. Macken. 2006. "Mitigation Strategies for Pandemic Influenza in the United States." *Proceedings of the National Academy of Sciences* 103 (15): 5935–40. https://doi.org/10.1073/pnas.0601266103.

Gupte, Pratik, Rosalind Eggo, and Edwin Van Leeuwen. 2025. *Epidemics: Composable Epidemic Scenario Modelling*. https://github.com/epiverse-trace/epidemics.

Gupte, Pratik, Edwin Van Leeuwen, and Adam Kucharski. 2025. *Finalsize: Calculate the Final Size of an Epidemic*. https://github.com/epiverse-trace/finalsize.

Johnson, N. P. A. S., and J. Mueller. 2002. "Updating the Accounts: Global Mortality of the 1918-1920 'Spanish' Influenza Pandemic." *Bulletin of the History of Medicine* 76 (1): 105–15. https://doi.org/10.1353/bhm.2002.0022.

Lambert, Joshua W., Adam Kucharski, and Carmen Tamayo. 2025. *Epiparameter: Classes and Helper Functions for Working with Epidemiological Parameters*. https://doi.org/10.5281/zenodo.11110881.

Longini, I. M., P. E. M. Fine, and S. B. Thacker. 1986. "Predicting the Global Spread of New Infectious Agents." *American Journal of Epidemiology* 123 (3): 383–91. https://doi.org/10.1093/oxfordjournals.aje.a114253.

Nikbakht, Roya, Mohammad Reza Baneshi, and Abbas Bahrampour. 2018. "Estimation of the Basic Reproduction Number and Vaccination Coverage of Influenza in the United States (2017-18)." *Journal of Research in Health Sciences* 18 (September): e00427. https://pmc.ncbi.nlm.nih.gov/articles/PMC6941634/.

Patterson, K. D., and G. F. Pyle. 1991. "The Geography and Mortality of the 1918 Influenza Pandemic." *Bulletin of the History of Medicine* 65 (1): 4–21. https://www.jstor.org/stable/44447656.

Rudis, Bob. 2022. "Cdcfluview: Retrieve Flu Season Data from the United States CDC FluView Portal." https://github.com/hrbrmstr/cdcfluview; GitHub.

Sagripanti, J.-L., and C. D. Lytle. 2007. "Inactivation of Influenza Virus by Solar Radiation." *Photochemistry and Photobiology* 83: 1278–82. https://doi.org/10.1111/j.1751-1097.2007.00177.x.

Samsuzzoha, Md., M. Singh, and D. Lucy. 2013. "Parameter Estimation of Influenza Epidemic Model." *Applied Mathematics and Computation* 220: 616–29. https://doi.org/10.1016/j.amc.2013.07.040.

Stephens, K. E., P. Chernyavskiy, and D. R. Bruns. 2021. "Impact of Altitude on COVID-19 Infection and Death in the United States: A Modeling and Observational Study." *PLOS ONE* 16 (1): e0245055. https://doi.org/10.1371/journal.pone.0245055.

Tan, Y., D. Cator III, M. Ndeffo-Mbah, and U. Braga-Neto. 2021. "A Stochastic Metapopulation State-Space Approach to Modeling and Estimating COVID-19 Spread." *Mathematical*

Biosciences and Engineering 18 (6): 7685–7710. https://doi.org/10.3934/mbe.2021381.

Tang, Biao, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu. 2020. "Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions." *Journal of Clinical Medicine* 9 (2). https://doi.org/10.3390/jcm9020462.

Turinici, Gabriel. 2025. "The Impact of Recovery Rate Heterogeneity in Achieving Herd Immunity." *Bollettino Dell'Unione Matematica Italiana*, February. https://doi.org/10.1007/s40574-025-00471-w.

Valle, S. Y. D., S. M. Mniszewski, and J. M. Hyman. 2012. "Modeling the Impact of Behavior Changes on the Spread of Pandemic Influenza." In *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*, 59–77. https://doi.org/10.1007/978-1-4614-5474-8_4.

Virlogeux, Victor, Ming Li, Tim K. Tsang, Luzhao Feng, Vicky J. Fang, Hui Jiang, Peng Wu, et al. 2015. "Estimating the Distribution of the Incubation Periods of Human Avian Influenza a(H7N9) Virus Infections." *American Journal of Epidemiology* 182 (8): 723–29. https://doi.org/10.1093/aje/kwv115.

Walker, Kyle, and Matt Herman. 2025. *Tidycensus: Load US Census Boundary and Attribute Data as 'Tidyverse' and 'Sf'-Ready Data Frames*. https://walker-data.com/tidycensus/.

Weber, T. P., and N. I. Stilianakis. 2008. "A Note on the Inactivation of Influenza a Viruses by Solar Radiation, Relative Humidity and Temperature." *Photochemistry and Photobiology* 84: 1601–2. https://doi.org/10.1111/j.1751-1097.2008.00416.x.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2025. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, and Lionel Henry. 2025. *Purrr: Functional Programming Tools*. https://purrr.tidyverse.org/.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2025. *Tidyr: Tidy Messy Data*. https://tidyr.tidyverse.org.

Xie, Yihui. 2025. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. https://yihui.org/knitr/.