

My title*

Impact Analysis of Japanese Occupation on Population Shifts Across Shanghai's Districts During WWII

Sandy Yu

December 2, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Analyses and findings in this paper are structured into several sections: Section 2 – Data, ?@sec-model – Model, ?@sec-results – Results, and ?@sec-discussion – Discussion. The Data section examines all variables and attributes kept for analysis, followed by an explanation of their measurement and purpose. The Model section defines the generalized linear model used for further analysis, explain its components, and presents model justifications. The Result section focuses on visualizing and presenting the model results through data presented in Data section. The Discussion section further evaluate the interpretations behind the model results presented in the previous section, and touches on any expanded topics as well as weaknesses and next steps.

*Code and data are available at: https://github.com/Jingying-yu/Shanghai_population_change

2 Data

2.1 Overview

Data is cleaned and analyzed using the open-source statistical programming language R (R Core Team 2023) and supporting packages including tidyverse (Wickham et al. 2019), arrow (Richardson et al. 2024), readr (**readr?**), knitr (Xie 2023), kableExtra (**kableExtra?**), modelsummary (**modelsummary?**), rstanarm (Goodrich et al. 2022), dplyr (Wickham et al. 2023), and broom.mixed (**broom.mixed?**). These tools facilitate data manipulation, visualization, modeling, and reporting. Detailed descriptions of each dataset are provided in the subsections below.

2.2 Data Sources

- talk about where the data in my primary reference book is gathered
- mentions the credibility of source and instability of recording instrument & methodology

2.3 Historical Background

Place: Shanghai, China

When: 1936-1942

Who: Chinese population in Shanghai

Define: 1. give 1-2 sentence broad overview of China's state of unrest 2. THREE districts in Shanghai: Chinese District, International Settlement, French Concession - who controlled each district and the level of governance each authority have in comparison to Chinese government 3. Outline area (%) of each district (do not get into specifics, put that in Results section)

Important Event Timeline

1. 1937-08-13: Japanese armed forces entered Shanghai
2. 1937-11-12: Japanese armed forces claims occupation of Shanghai → ends Chinese district
- 1940-06-22: France surrendered to Germany → French Concession stopped taking refugees
3. 1942-01: Japanese armed forces claims authority over International Settlement (which was mainly under the governance of U.K and U.S prior to this date)
4. 1945-08-15: Japan surrendered in WWII
5. 1945-10: most Japanese armed forces withdrew from Shanghai

2.4 Measurement & Methodology

- how population is recorded
- Why choose 1936 - 1942 (ex. if an event occurred in Nov of 1937, would I take values of 1937 as a variable for prior to event occurrence or after?)

Survey Difficulties - Population Volatility caused by Warfare: Population is extreme volatile within this time period, population shifts constantly as proceed - Regional Recording Limitations: Survey response collect difficulty varies between districts. - Resistance due to Political Risk

2.5 Variables of Interest

Table 1: Explanatory models of flight time based on wing width and wing length

Year	Pop CHN District	Pop IS	Pop FC
1936	2155717	1180969	477629
1937	2155717	1218630	477629
1938	2074693	1272552	504731
1939	2098331	1257703	495942
1940	1479726	1233394	477629
1941	957339	446692	835523
1942	1049403	1585673	854380

2.5.1 Outcome variable

- outcome variable is *population*

2.5.2 Predictor variables

Time variable: year Observations used in modelling and other analysis in this paper is ranged between 1936 to 1942. No observation is selected before 1936 as the growth pattern varies significantly from the event of our interest (Japanese occupation of Shanghai's different districts), and no value is elected beyond 1942 due to data constraints. No annual population records exists on a district bases beyond 1942. Foreign districts in 1949 as new political party came to power and the People's Republic of China was declared.

2.5.2.1 Binary Indicators

District indicators: - Variable *district_is* and *district_fc* are two binary variables used to indicate the district identity of an observation. There are 3 possible district identities in Shanghai during 1930s-40s: Chinese-administrated District, International Settlement, and French Concession.

The Chinese-administrated District is governed by local Chinese authority. Land type in this area is primarily rural, most of the local population are farmers. Percentage of foreign (out-of-province) immigrants or refugees is low. ((sh?)) In this analysis paper, Chinese-administrated District takes value of *district_is* = 0 and *district_fc* = 0 (it is neither International Settlement nor French Concession).

The International Settlement district is a generalized term for section of lands that allowed the occupation of foreign (non-Chinese) residence. Foreigners in Shanghai were not allowed to reside in areas outside of the bounds of this district. Governance power of foreign population within the district – including the policing, taxing, etc, are transferred to foreign authorities. This district takes up most of the central-city in Shanghai during the 1930s-40s, with foreign investment attracting many business opportunities and subsequently job opportunities. Due to the complicated political authorities within the district, the part of Shanghai remains initially untouched by the Japanese forces as they entered Shanghai's Chinese-administrated District. International Settlement district takes the value of *district_is* = 1 and *district_fc* = 0 in this paper.

French Concession district is a land district given solely to the French. French residency are given the authorization to self-govern within this area. Due to political reasons on the Western battlefield, this district remains untouched by the Japanese forces all the way until they withdrew from Shanghai. The French Concession district takes value of *district_is* = 0 and *district_fc* = 1 for all analysis in the paper.

Historic Event indicators: - *cd_occupied*, *french_surrender*, and *is_occupied* are binary variables that mark the whether or not a major historic event that may have impacted Shanghai's population during the observation period (1936-1942) has occurred yet. 1. The Japanese armed forces declared occupation of the Chinese-administrated district on 1937-11-12. With expectation that Shanghai's population will internally shift out of the Chinese-administrated district, the binary variable *cd_occupied* is used to measure the impact of this event on the population change in Shanghai. 2. As WWII impacted the Eastern globe, on the primary battlefield (Western Battlefield), France surrendered to Germany on 22 June 1940. Since Japan sided with Germany in WWII, France's whiteflagging to Germany led to Shanghai's Japanese forces promising not to invade the French Concession on the condition that the district stop taking refugees and support Japanese governance. The paper uses the binary variable *french_surrender* to evaluate the impact this event has on the overall population in Shanghai. 3. In January of 1942, the Japanese forces occupied the International Settlement. Local authority handed over its governing power to the invading forces. This event's impact on Shanghai's population is measuring using the binary variable *is_occupied*.

Table 2: Explanatory models of flight time based on wing width and wing length

Year	Population	district_is	district_fc	cd_occupied	french_surrender	is_occupied
1936	477629	0	1	0	0	0
1937	2155717	0	0	1	0	0
1938	1272552	1	0	1	0	0
1940	477629	0	1	1	1	0
1941	957339	0	0	1	1	0
1942	1585673	1	0	1	1	1

3 Model

The goal of our modelling strategy is to evaluate the impact of 2 historical events in Shanghai during WWII.

Here we briefly describe the Difference-in-Difference analysis model used to investigate the impact of the Japanese forces taking over the Chinese district and International Settlement in November 1937 and December 1941 on population shift between different Districts in Shanghai.

Background details and diagnostics are included in Appendix Section B.

3.1 Model set-up

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_{i,t}, \sigma)$$

$$\alpha \sim \text{Normal}(0, 2.5)$$

$$\beta_k \sim \text{Normal}(0, 2.5) \quad \text{for } k = 1, \dots, 11$$

$$\sigma \sim \text{Exponential}(1)$$

Where:

- y_i : Observed population in district i at year t.
- $cd_occupied$: Indicator for whether the Chinese-administered district was occupied by Japanese forces in year t (1 if $t \geq 1937$, 0 otherwise).

- *french_surrender*: Indicator for the French Surrendering to Germany in 1940, leading to the French Concession rejecting refugees in Shanghai (1 if $t \geq 1940$, 0 otherwise).
- *is_occupied*: Indicator for the Japanese occupation of the International Settlement in 1942 (1 if $t \geq 1942$, 0 otherwise).
- *district_type*: Dummy variable for International Settlement districts (1 if i is IS, 0 otherwise).
- *district_type*: Dummy variable for French Concession districts (1 if i is FC, 0 otherwise).

Interaction Terms:

The interaction terms in the model are essential because they capture how the impact of historical events (e.g., the 1937 occupation of Chinese-administered districts, the 1940 refugee rejection in the French Concession, and the 1942 occupation of the International Settlement) varies across different district types. By including interactions between event indicators and district types, the model accounts for district-specific responses to each event, controlling for the fact that the effect of an event on population changes may not be uniform across all districts. For instance, the 1937 occupation might lead to population declines in Chinese-administered districts but population increases in the International Settlement due to refugee influx. These terms ensure that the model can differentiate and estimate the unique effects of each event within each district type, thereby improving precision and interpretability.

- *cd_occupied x district_type*: Captures the effect of the 1937 occupation on the International Settlement in comparison to the Chinese-administered district (base district).
- *cd_occupied x district_type*: Captures the effect of the 1937 occupation on the French Concession in comparison to the base district.
- *french_surrender x district_type*: Effect of the 1940 refugee rejection on the International Settlement.
- *french_surrender x district_type*: Effect of the 1940 refugee rejection on the French Concession.
- *is_occupied x district_type*: Effect of the 1942 occupation on the International Settlement.
- *is_occupied x district_type*: Effect of the 1942 occupation on the French Concession.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ .

4 Results

- Data Trend Visualization

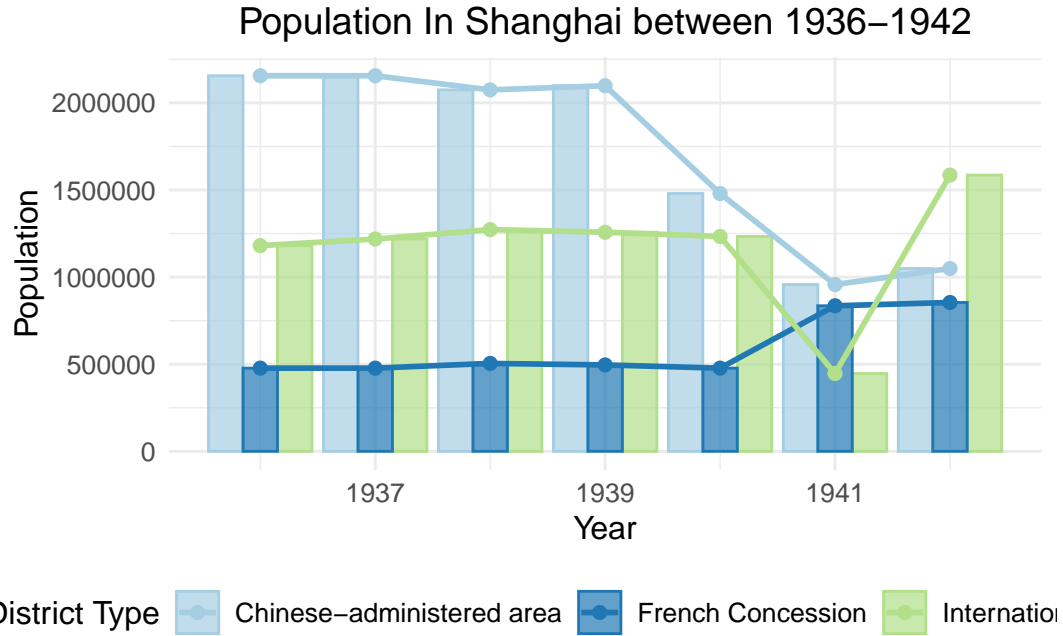


Figure 1: Population in Shanghai between 1936 - 1942

- Model results are summarized in Table 6 in Appendix (Section B).

4.0.0.1 District Impact on Population

Table 3: Explanatory models of flight time based on wing width and wing length

Parameters	Estimated Effect	Standard Error
(Intercept)	117928165.6	189545438.8
district_is	-943281.6	381299.4
district_fc	-1642126.9	350808.4

4.0.0.2 Historical Events

Table 4: Explanatory models of flight time based on wing width and wing length

Parameters	Estimated Effect	Standard Error
cd_occupied	89350.01	350665.6
french_surrender	-735223.28	340852.1
is_occupied	-88982.32	361506.9

4.0.0.3 Historic Events' Impact on Each District

```
# A tibble: 6 x 3
  term                estimate std.error
  <chr>                <dbl>    <dbl>
1 is_occupied         -88982.    361507.
2 district_is:cd_occupied    89431.    414727.
3 district_fc:cd_occupied    38397.    410271.
4 district_is:french_surrender 470618.    352216.
5 district_fc:french_surrender 1044471.    352470.
6 district_is:is_occupied    934059.    476934.
```

Table 5: Explanatory models of flight time based on wing width and wing length

Parameters	Estimated Effect	Standard Error
is_occupied	-88982.32	361506.9
district_is:cd_occupied	89430.69	414726.8
district_fc:cd_occupied	38396.87	410271.2
district_is:french_surrender	470617.65	352215.9
district_fc:french_surrender	1044470.54	352469.9
district_is:is_occupied	934059.10	476934.2

5 Discussion

5.1 Interpretation of the Event Impacts

5.1.1 Impact of Occupation in Chinese District

5.1.2 Impact of French Surrendering to Germany on Shanghai

5.1.3 Impact of Occupation in International Settlement

5.2 Reason why the French Concession was Exempted from Japanese Occupation

- During WWII, on the European battlefield, France surrendered to Germany on June 22nd of 1940 (Check this & reference with a credible source). Since Japan was on the same side as Germany during that time, Japanese armed forces decided allow the French Concession to retain its own governance.
- However, the French Concession is still constantly under the watch of the Japanese forces that has occupied the rest of Shanghai since 1942. ,etc.(can reference to video summary here)

5.3 Refugees from outside of Shanghai Province

- industrialization was not so popular back in the 1930s, many provinces are still mostly rural. But warfare messed with the land and the yields, causing many farmers to starve and ultimately have to seek refugee (find job) in industrialized city (where income is not dependent on land yield)
- many people flooded to Shanghai for this reason (can insert calculation in data table 20 from reference book here)

5.3.1 Job Oppurtunities in Shanghai

Aside from the instability caused by the Japanese armed forces, there are a few other reasons for Chinese population to move into the international settlement & French Concession. 1. Job opportunities - Heart of Shanghai city in the 1930-1940s, high pop density lead to boom of economy = more job opportunities - many factories are located with the settlements (less so in concession)

5.3.2 Barrier to Entry in Job Market

2. Barrier to Entry for certain job types

- —

5.4 Instability after Japanese Occupation ends in WWII: Chinese Civil War

- after the Japanese armed forces withdrew from Shanghai, what follows is not recovery and rest for the local population
- Competition for power within China's 2 political parties caused full-scale political warfare, now known as the *Chinese Civil War*
- unlike Western political parties, Chinese political parties in the 1900s are more like parties of a throne. They each have their own ideology and an armed forces that follow. Most of all of the provinces are effected by this internal warfare, which ended in 1949 as lost and withdrawn to Taiwan.

5.5 Weaknesses and next steps

5.5.1 Data Measurement Weaknesses

- data was recorded during wartime, many numbers were missing
- base unit for population records is not ppl, it is instead "household". # of ppl in a household is estimated based on historical data
- many people chose to not report or partially report numbers due to convenience or economic reasons, data may not be accurate

5.5.2 Next Steps

Appendix

A Additional data details

B Model details

B.1 Model Summary Table

B.2 Posterior predictive check

In Figure 2a we implement a posterior predictive check. This shows...

In Figure 2b we compare the posterior with the prior. This shows...

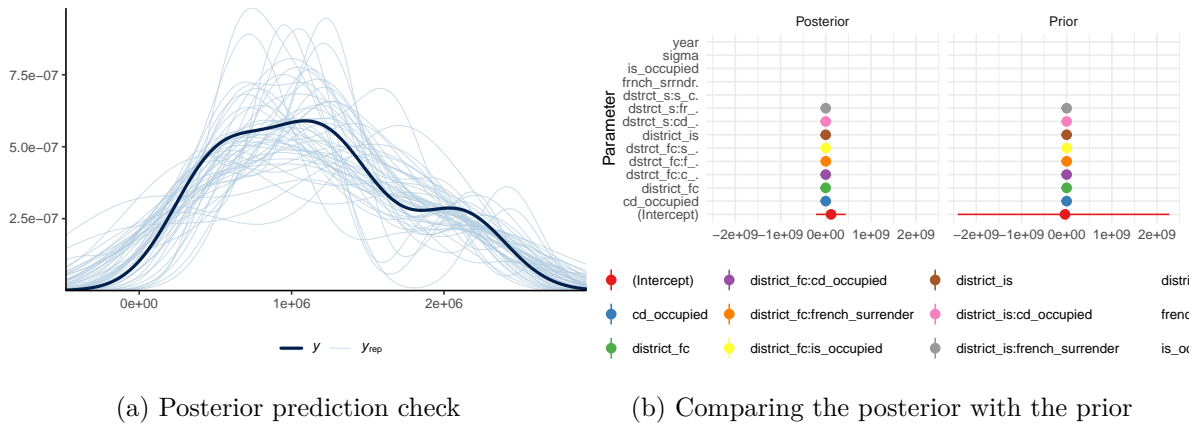


Figure 2: Examining how the model fits, and is affected by, the data

B.3 Diagnostics

Figure 3a is a trace plot. It shows... This suggests...

Figure 3b is a Rhat plot. It shows... This suggests...

C Appendix B: Idealized Survey Methodology

This appendix presents an idealized design for a population sampling survey, assuming that a data-recorder has unlimited time and authority to collect data within the historical timeframe of 1936–1942. The proposed methodology outlines a systematic approach to gathering detailed

Table 6: Explanatory models of flight time based on wing width and wing length

	Population model
(Intercept)	117 928 165.58 (189 545 438.85)
year	−59 841.02 (98 298.59)
district_is	−943 281.63 (381 299.41)
district_fc	−1 642 126.91 (350 808.44)
cd_occupied	89 350.01 (350 665.57)
french_surrender	−735 223.28 (340 852.10)
is_occupied	−88 982.32 (361 506.88)
district_is \times cd_occupied	89 430.69 (414 726.84)
district_fc \times cd_occupied	38 396.87 (410 271.18)
district_is \times french_surrender	470 617.65 (352 215.89)
district_fc \times french_surrender	1 044 470.54 (352 469.86)
district_is \times is_occupied	934 059.10 (476 934.16)
district_fc \times is_occupied	377 436.02 (466 616.20)
Num.Obs.	21
R2	0.837
Log.Lik.	−289.106
RMSE	162 825.22

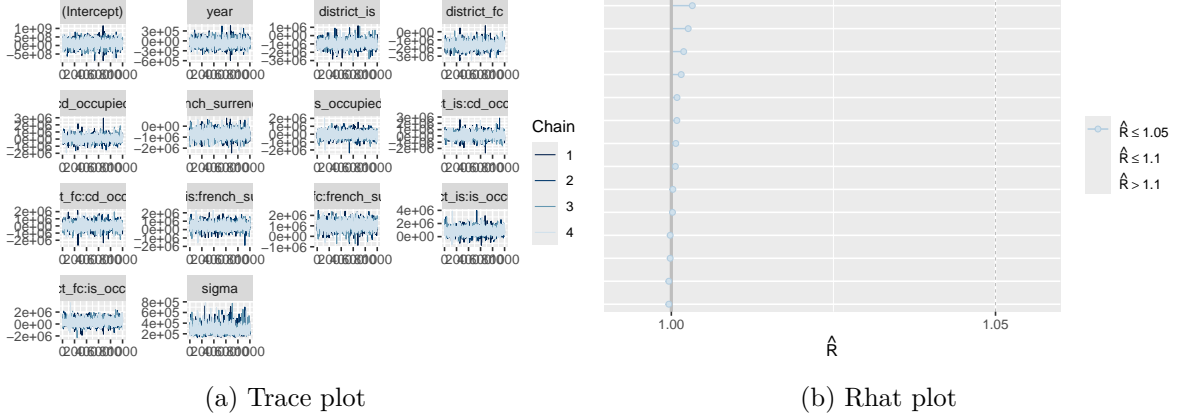


Figure 3: Checking the convergence of the MCMC algorithm

population data from Shanghai’s districts during pivotal wartime events, with considerations for the practical constraints and technology available during that period. While acknowledging the inherent challenges of conducting such a survey, this design aims to provide a comprehensive framework for understanding population dynamics within the Chinese-administered areas, International Settlement, and French Concession.

C.1 Objective of the Survey

The purpose of this survey is to evaluate the relationship between key historical events and population shifts within Shanghai’s districts between 1936 and 1942. Specifically, the study aims to document the population changes in three major districts—Chinese-administered areas, International Settlement, and French Concession—during pivotal events such as the 1937 Japanese occupation of Chinese districts, the 1940 rejection of refugees in the French Concession, and the 1942 occupation of the International Settlement.

C.2 Target Population

The survey targets residents and local representatives from the three major districts of Shanghai. These include long-term residents who were present before and during the historical events under study, as well as recent migrants and refugees arriving in these districts as a result of wartime displacement. The goal is to achieve representative coverage across the three districts to capture the diversity in population trends and movement patterns.

C.3 Sampling Design

The sampling design employs stratified random sampling to ensure representation across the three major districts. Each district is treated as a separate stratum: Chinese-administered areas, International Settlement, and French Concession. A fixed number of households per district are selected based on district size, population density, and anticipated population shifts. Special attention is given to refugee populations within the International Settlement and French Concession to capture the impact of displacement. Enumerators use household lists and local community organizations, such as neighborhood associations and refugee aid committees, to randomly select respondents within strata.

C.4 Recruitment Strategy

The recruitment strategy involves direct engagement with local leaders, community groups, and refugee aid organizations. Enumerators collaborate with influential figures in each district to gain trust and permission to conduct the survey. Recruitment booths are set up in public spaces such as markets and parks to engage with potential respondents. Additionally, refugee aid centers are utilized to access displaced populations. Recruitment materials are printed in both Chinese and English to accommodate the linguistic diversity of Shanghai's residents.

C.5 Survey Instrument

The survey instrument consists of a structured questionnaire covering demographic information, migration history, wartime impact, and population and housing details. The demographic section collects age, gender, occupation, family size, and household composition. Migration history includes the origin, year of migration to the current district, and reasons for relocation. Wartime impact focuses on displacement, property loss, and wartime experiences. Lastly, the population and housing section gathers data on the number of residents, housing type, and living conditions. Open-ended questions are also included to capture respondents' personal narratives about wartime displacement and resettlement.

C.6 Data Collection Procedures

Data collection occurs between 1936 and 1942, with enumerators visiting households, public spaces, and refugee shelters. Enumerators are trained to conduct interviews in Chinese, Shanghaiese, or English based on respondent preferences. Responses are recorded on pre-printed forms, which are later transcribed and stored securely. To ensure accuracy, enumerators cross-check entries at the end of each survey day. Each enumerator is assigned a specific district to maintain consistency in data collection practices.

C.7 Limitations and Challenges

Some areas, particularly those under military occupation, may not be accessible for data collection, leading to incomplete coverage. Additionally, respondents may underreport family size or migration history due to privacy concerns or fear of government reprisal. Refugee populations are often transient and may not remain in one location long enough for follow-up interviews. Despite these limitations, the survey employs robust sampling and recruitment strategies to mitigate potential biases.

C.8 Ethical Considerations

Informed consent is obtained by informing respondents about the purpose of the survey and how their data will be used before they agree to participate. Personal identifying information is anonymized to ensure respondent privacy and security. Participation is entirely voluntary, and respondents can opt out at any time without consequences. The survey complies with ethical standards to respect the rights and dignity of all participants.

References

- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. <https://github.com/apache/arrow/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.