My title*

My subtitle if needed

Tian Yi Zhang

03 April 2022

Abstract

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

1 Introduction

The usage of contraception has become an important part of modern society. Various contraceptive methods have been invented throughout the years to aid couples in their family planning efforts. These contraceptive methods range from hormonal methods and intrauterine methods to barrier methods, providing many options for both male and female users. Notably, according to the Centers for Disease Control and Prevention, the barrier method known as condom was the most popular male contraceptive used in the United States a few years ago, with it being around 8.7% of total contraceptives used (Daniels and Abma 2018). This is also the case in the country of Kenya, with around 1% of the surveyed individuals reporting to have used condoms, as denoted in the 1998 DHS report (The Demographic and Health Surveys Program 1999).

Another interesting statistic in the studies mentioned above are the high usage rates of female hormonal contraceptives. 12.6% of contraceptive methods used in the United States are of the oral hormonal type (Daniels and Abma 2018), and around 9% of surveyees in Kenya report using oral contraceptives as well (The Demographic and Health Surveys Program 1999). This statistic is perplexing in that the usage rate of hormonal contraceptives are rather high despite the noticeable side effects: Various studies have suggested in the past that female contraceptives often lead to high-risk side effects. For example, the U.S. Department of Health & Human Services notes that oral contraceptives that rely on hormonal controls can lead to higher blood pressure and more blood clots (Office on Women's Health 2019). As such, one would wonder why male contraceptives do not see more wide usage despite a relative lack of harmful side effects in comparison to female hormonal contraceptives.

This paper attempts to analyze potential reasons for the relative lack of condom usage as a contraceptive method in Kenya utilizing statistics presented in the 1998 DHS final report on Kenya (The Demographic and Health Surveys Program 1999). Specifically, the paper focuses on finding a, of lack thereof, relationship between the age, geographical location, and the knowledge of condom(TODO)

WRITE SOMETHING ABOUT FINDINGS; WRITE SOMETHING ABOUT THE STURCTURE OF THE PAPER

You can and should cross-reference sections and sub-sections. For instance, Section 2. R Markdown automatically makes the sections lower case and adds a dash to spaces to generate labels, for instance, Section **??**.

*Code and data supporting this paper are available at: AOIWDJAWOIDJ

2 Data

2.1 Raw Data Origin and DHS Methodology

The dataset used in this paper is obtained from the PDF "FR102," titled "Kenya Demographic and Healthy Survey 1998" and published in 1998 by the Demographic and Health Surveys Program.[1]

The PDF where the dataset is parsed from is a report compiled by the DHS program using the third Demographic and Health Survey conducted in Kenya, the previous two of which were conducted in 1989 and 1993. (The Demographic and Health Surveys Program 1999). According to the information provided on the DHS Program website, Standard DHS Surveys are described as surveys with a sample size that consists of 5000 to 30000 households conducted on a five year cycle (The Demographic and Health Surveys Program 2022). Since there is around a five year gap between each consecutive Kenya DHS survey, we can safely assume that the PDF report containing the raw data is a standard DHS survey.

The survey used by the DHS Program was distributed in nearly all districts in Kenya. While the total population of the districts not included in the survey only amounts to less than four percent of the country's population, it is still a substantial percentage of the national population that was not sampled from. The PDF specified that stratified sampling was used, which greatly contributes greatly to the increase in precision of the findings from the survey and generally leads to a sample that better represents the population than a simple random sampling. However, it was unclear to the readers what characteristics or attributes were used to divide the population into strata. Since it was stated later in the PDF that the survey wished to obtain "district-level data for planning purposed" (The Demographic and Health Surveys Program 1999), it should be safe to assume that the strata were constructed based on geographical locations. It is also worth noting that the survey included men from age 15 to 54 while only including women from age 15 to 49. A sufficient justification was not found for this discrepancy and it may have an impact on the accuracy of the findings.

According to the DHS report (The Demographic and Health Surveys Program 1999), the survey questions can be divided into 14 categories. Of those 14, the following subset of topic is related to the focus of this paper:

- Background characteristics such as age and education
- Knowledge and use of family planning methods.

The surveys questions were originally constructed fully in English before being translated into various local languages, and the wording of the questions were revised several times by local pretest teams before being sent out, ensuring that there will not be any inaccuracies or awkward wording present due to the language barrier. (The Demographic and Health Surveys Program 1999). This is especially important due to the sensitive nature of the topics asked in the surveys, and helps in making sure that as many participants answer honestly as possible.

2.2 Dataset Parsing and Cleaning

The raw dataset is parsed using the package `pdftools` (Ooms 2022) and the statistical programming language R (R Core Team 2020) from Table 10.14.1 and Table 10.14.2 on page 170 and 171 of the PDF respectively (The Demographic and Health Surveys Program 1999). Two tables are parsed instead of one because the tables each present data on only one gender. Therefore a dataset that combines data from both genders in order to provide a more accurate analysis. As such, a new variable, `gender`, is constructed to help distinguish the data between female and male participants. Additionally, there is an issue in that the rows of the tables corresponded to different variables; for example, some rows contained data for certain age groups while other

[1]The PDF can be obtained at https://dhsprogram.com/publications/publication-fr102-dhs-final-reports.cfm?csSearch=456440_1.

rows contained data for different marital statuses. To simplify the analysis process, only the rows pertaining to age groups, education, and provinces were parsed from the PDF.

The dataset is then cleaned using the `janitor` package (Firke 2021) and the `pointblank` (Iannone and Vargas 2022) package is used to set up validation tests for the class and the content of the datasets. Subsequent graphs and tables shown in this paper also utilize the `tidyverse` (Wickham et al. 2019) and `knitr` (Xie 2021) packages.

The final datasets—clean_data_age.csv, cleaned_data_province.csv and cleaned_data_education.csv—showcased in Table 1, Table 2 and Table 3 respectively, contains the information about the same variables in every column except the first, which is the column that denotes either the age, province or education. The purpose of the other columns are explained in the DHS report (The Demographic and Health Surveys Program 1999), summarized as follows:

- gender: male or female
- know_about_condoms: percentage of participants that know about condoms
- public_sector: percentage of participants that cited government-run health services as their primary source for obtaining condoms
- private_medical_sector: percentage of participants that cited private medical sources as their primary way for obtaining condoms
- private_pharmacy: percentage of participants that cited private pharmacies as their primary source for obtaining condoms
- shop: percentage of participants that cited commercial shops as their primary source for obtaining condoms
- cbd_agent: percentage of participants that cited community-based distribution agents as their primary source for obtaining condoms
- friends_and_relatives: percentage of participants that cited their friends and relatives as their primary source for obtaining condoms
- other_sources: percentage of participants that cited other sources as their primary way for obtaining condoms
- dont_know_a_source: percentage of participants that do not know a source for obtaining condoms
- number_of_people: number of participants that belong in each category

2.3 Preliminary Analysis

Table 1 shows that an overwhelming majority of surveyees know about condoms, with all percentages well over 90%. For women, the percentage of surveyees that know about condoms are relatively lower in age groups 15-19 and 40-49, though there is not a similar trend for men, as more than 97% of men know about condoms in all age groups surveyed. This difference could be due to the fact that condom is a male-contraceptive and is thus less advertised to women. Similarly, Table 1 shows that a larger percentage of men know about condoms than women in every single province survyed, though this difference is minuscule.

The datasets also show that there are more than twice the number of female participants than male participants. Note that according to Faria (Faria 2022), the gender ratio in Kenya in 2000 is about 98.6 males per 100 females, meaning that there is no significant numerical difference between genders in Kenya. Therefore, the unequal sample size in the dataset could lead to unequal variances between the male and female samples and would negatively affect the results of any tests that require the assumption of equal variances.

Table 1: First 5 columns on percentage of participants, categorized by age groups, that know about and/or cite a specific source for condoms, or do not know about condoms

Age	Gender	Know About Condoms	Public Sector	Private Medical Sector
15-19	female	95.9	18.8	4.3
20-24	female	97.7	33.9	7.6
25-29	female	97.4	40.3	9.0
30-39	female	96.3	36.7	8.4
40-49	female	93.8	31.3	6.2
Total	female	96.3	33.8	7.5
15-19	male	99.9	17.3	4.0
20-24	male	99.4	22.5	7.8
25-29	male	99.2	22.6	10.0
30-39	male	99.3	26.1	7.9
40-49	male	97.9	23.2	7.3
50-54	male	97.9	19.5	3.7
Total	male	99.0	22.6	7.3

Table 2: First 5 columns on percentage of participants, categorized by provinces, that know about and/or cite a specific source for condoms, or do not know about condoms

Province	Gender	Know About Condoms	Public Sector	Private Medical Sector
Nairobi	female	99.4	20.5	8.2
Central	female	97.9	41.3	8.1
Coast	female	92.2	34.7	11.4
Eastern	female	97.8	29.7	9.1
Nyanza	female	98.2	43.3	5.6
Rift Valley	female	92.2	26.2	7.9
Western	female	97.0	39.9	3.8
Total	female	96.3	33.8	7.5
Nairobi	male	100.0	19.0	6.3
Central	male	99.4	24.5	11.7
Coast	male	97.9	16.7	3.4
Eastern	male	100.0	17.0	5.0
Nyanza	male	100.0	34.0	4.8
Rift Valley	male	97.5	24.3	13.1
Western	male	97.9	14.4	2.2
Total	male	99.0	22.6	7.3

Table 3: First 5 columns on percentage of participants, categorized by education level, that know about and/or cite a specific source for condoms, or do not know about condoms

Education	Gender	Know About Condoms	Public Sector	Private Medical Sector
No education	female	87.4	26.5	4.5
Primary incomplete	female	95.5	30.6	5.8
Primary complete	female	98.1	34.1	8.4
Secondary+	female	99.8	40.5	10.0
Total	female	96.3	33.8	7.5
No education	male	85.6	13.9	3.7
Primary incomplete	male	99.5	20.1	4.5
Primary complete	male	99.0	20.6	8.5
Secondary+	male	100.0	26.3	8.5
Total	male	99.0	22.6	7.3

2.4 Knowledge of a source for condom: Age Groups

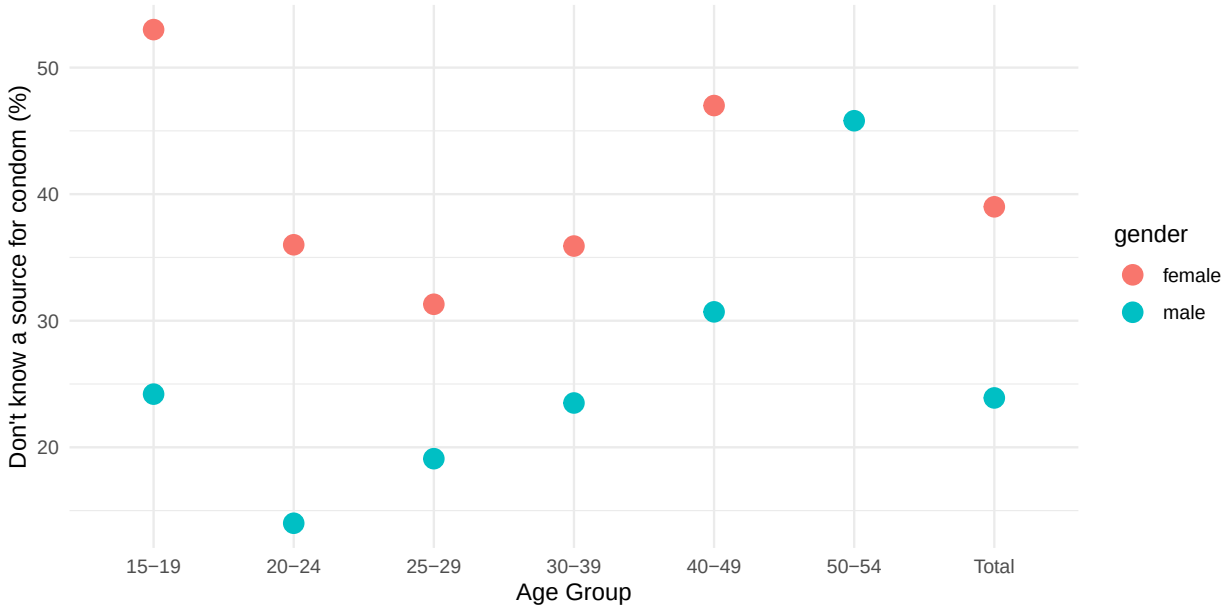

Figure 1: Percentage of Surveyee Without Knowledge of A Source for Condom versus Age Groups

We can observe from Figure 1 that there is a substantially higher percentage of women who possess no knowledge of ways to obtain condoms In all age groups, there are on average twice as many women who do not know a source to obtain condoms compared to men. This difference is much more drastic in the younger population (age 15-19).

Note that the female age group of 50-54 was not included in the sample, thus leaving no way for us to compare the difference in percentage for that specific age group.

2.5 Knowledge of a source for condom: Geographical Location

Figure 2 plots the percentage of male and female surveyees without knowledge on a source for condoms in each province surveyed and in total.

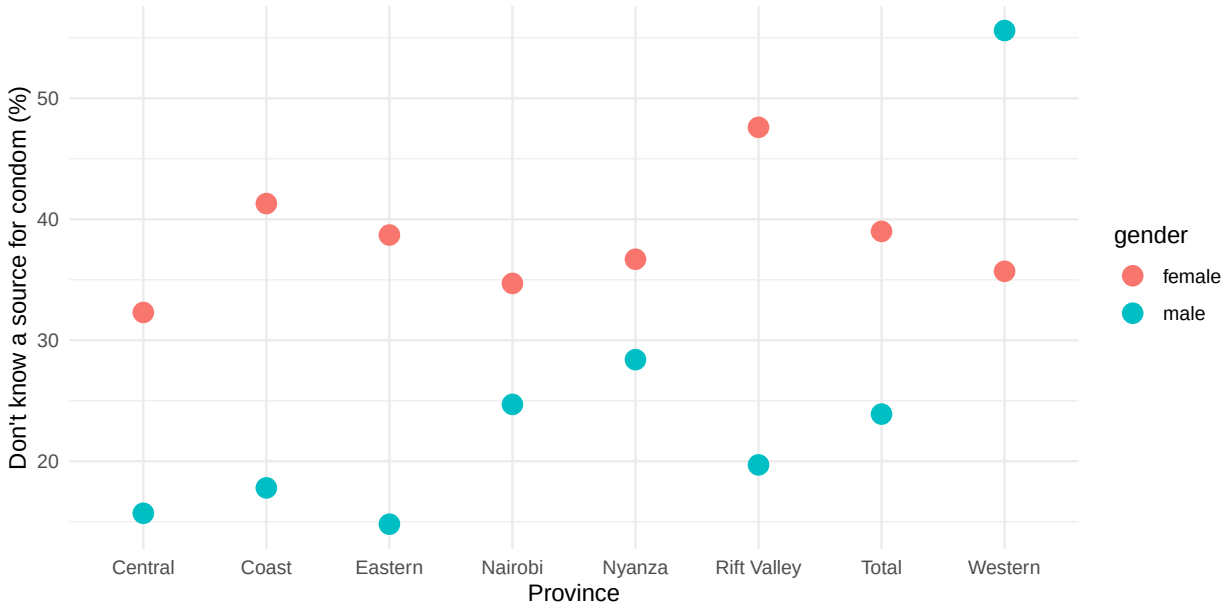

Figure 2: Percentage of Surveyee Without Knowledge of A Source for Condom versus Provinces

Similar to the trend shown in Figure 1, the difference in percentage between female and male participants are at the minimum 10%, with women being far more likely to know no source for condoms.

According to Table 2.4 in the Kenya Population and Housing Census in 2019 (Kenya National Bureau of Statistics 2019), Nairobi, the capital city of Kenya, has a population density of 6247 personnel per square kilometer, dwarfing any other county or province in Kenya. One hypothesis is that this level of population density and urbanization may be a reason as to why the gap between female and male knowledge on condom sources is relatively smaller in Nairobi. However, Nyanza, a province where the population density of its counties range from 393 to 958 personnel per square kilometer, has a difference in female and male knowledge on condom sources very similar to Nairobi's.

2.6 Knowledge of a source for condoms: Education Level

Figure 3 shows a trend that is similar to the previous two; a higher percentage of women do not know a source for condoms compared to men. A notable outlier is when there is a complete lack of education: Both female and male surveyees without any education have nearly identical level of ignorance regarding ways to obtain condoms.

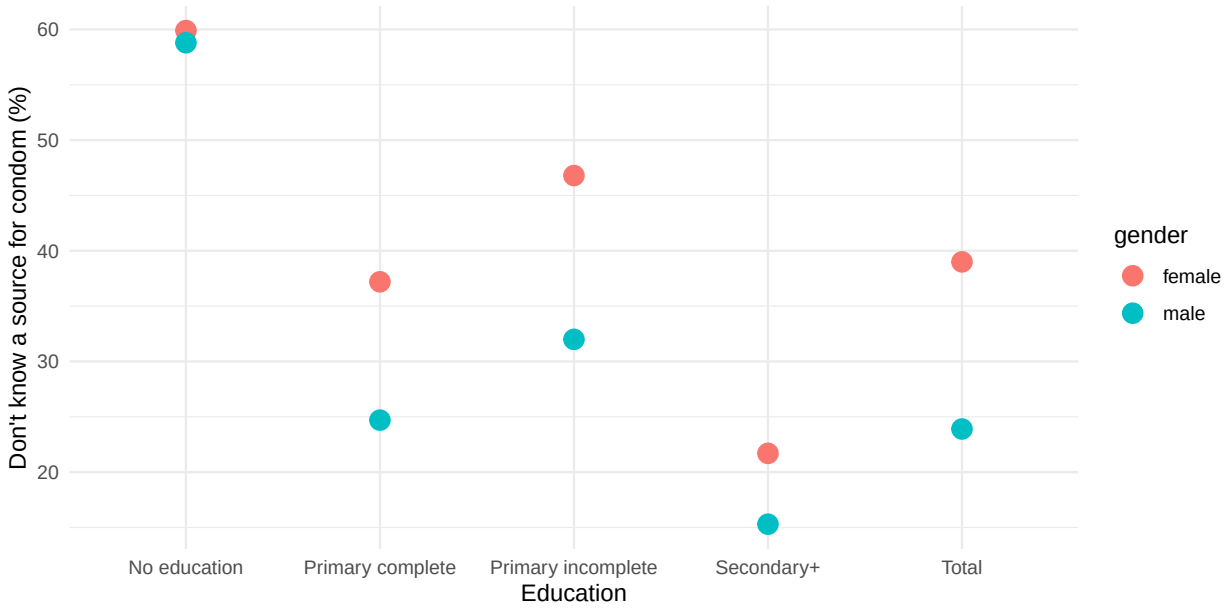

Figure 3: Percentage of Surveyee Without Knowledge of A Source for Condom versus Education Level

3 Results

3.1 Population density and knowledge on condoms

As shown in Figure 2, the difference in ignorance of condom sources between women and men vary by a significant amount between provinces. The initial hypothesis presented in this paper is that the difference could be linked to the population density in each province. However, after reviewing the Kenya Population Census (Kenya National Bureau of Statistics 2019), no apparent pattern is found. Some provinces with relatively massive population density such as Nairobi have a smaller difference in condom source ignorance as hypothesized to be due to its greater urbanization. However, other provinces that contain counties with population density comparable to Nairobi, such as the Coast province, have a much large difference in condom source ignorance between men and women: Mombasa, a county in the Coast province, has the second largest population density at 5495 personnel per square kilometer, yet the overall gap between female and male knowledge on condom sources is still relatively large in the Coast province. On the other hand, Nyanza, a province where the population density of its counties range from 393 to 958 personnel per square kilometer, has a difference in female and male knowledge on condom sources very similar to Nairobi's. It is unclear as to whether population density plays an important role in determining the region's ignorance of sources to obtain condoms.

3.2 Age Group and knowledge on condoms

Figure 1 shows that lack of knowledge on a source for condoms is much higher in the younger age group (age 15 to 19), decreases in older age groups, and once ago rises in the oldest age groups (age 40 to 49 and age 50 to 54).

In general, the disparity between percentage of women and men that do not know a source for condoms remains largely unaffected by age groups, though is worth noting however that younger surveyees (people in the age of 15 to 19) have a relatively higher difference in this regard. In total, the percentage of women that do not know of a source for condoms are 15.1% higher than men.

3.3 Education Level and knowledge on condoms

As demonstrated in Figure 3, a staggeringly high percentage of women and men with no education have no knowledge on where to obtain condoms. The graph also shows that receiving a higher level of education leads to an overall decrease in condom source ignorance in both genders. However, the difference in condom source ignorance between women and men seem to not be correlated to education level, as the difference in condom source ignorance between genders is lower in both the 'No Education' and 'Secondary+' levels compared to the 'Primary Incomplete' and 'Primary Complete' levels.

4 Discussion

4.1 Level of Variance in Population Density

As discussed in the Results section, we seemingly cannot establish a relationship between population density of a province and the difference in condom source ignorance between its female and male residents. The reasoning is that provinces that contain counties with similar levels of population density have very difference disparities in the percentage of women and men that know where to obtain condoms. However, it is important to note that each province is sectioned into different numbers of counties. A possible connection that was not analyzed in this paper is the hypothesis that the aforementioned disparities were caused by the level of variance in population density among the counties for each province. For example, according to the Kenya Population Census (Kenya National Bureau of Statistics 2019), Nyanza, a province where the population density of its counties range from 393 to 958 personnel per square kilometer, has a difference in female and male knowledge on condom sources very similar to that of the Nairobi province. On the other hand, the Rift Valley province contains counties with population density ranging from 14 to 397 personnel per square kilometer, and the Eastern province contains counties with population density ranging from 6 to 236 personnel per square kilometer. In these two provinces which have relatively lower variance in its counties' population densities, the difference in percentage of women and men that know sources for condoms are much larger compared to Nyanza, a province with a higher variance in its counties' population densities. There is a possibility of correlation in here, and this direction could prove to be an interesting add-on or extension of this paper.

4.2 Higher percentage of women with no knowledge of condom sources on average

Across all the tables and graphs presented in this paper, it is apparent that on average, a significantly higher percentage of women have no knowledge sources for condoms. It is unclear as to why this disparity exists. One possible reasoning is that condom is a male-contraceptive and thus less advertised to women.

4.3 Limitations

4.3.1 Unequal sampling size between genders

As discussed in the beginning of the Data section, in the report conducted by the DHS Program that is used to extract the raw data for this paper from, there are more than twice the number of female surveyees than male surveyees, yet, according to Faria (Faria 2022), the gender ratio in Kenya in 2000 is about 98.6 males per 100 females, meaning that there is no significant numerical difference between genders in Kenya. Therefore, there is no feasible justification for this difference in sample size between genders. This disadvantage could negatively affect the accuracy of any tests based on the assumption of equal variances.

4.3.2 Inexplicable lack of data

As showcased in Figure 1, there is a lack of data for women in the age group of 50 to 54 and we could not locate an explanation for it in the DHS Program report. It could be that there were simply not enough female surveyees in the age group of 50 to 54; While it was stated in the DHS Program report (The Demographic and Health Surveys Program 1999) that it used stratified sampling, the report did not specify which characteristic the sample was stratified on, leaving no way for us to figure out the reasoning behind this discrepancy.

4.3.3 Omission of certain counties and estimation for certain counties

The survey used by the DHS Program was distributed in nearly all districts in Kenya. While the total population of the districts not included in the survey only amounts to less than four percent of the country's population, it is still a substantial percentage of the national population that was not sampled from.

Additionally, the DHS Program report (The Demographic and Health Surveys Program 1999) noted that due to a drastic increase in number of counties in the country in the few years prior to the publication of the report, values for certain variables for some rural counties were estimated instead. One of the reasons for doing so, as stated in the DHS Program report, was that these districts were included in the previous two DHS Program reports of the same nature on Kenya, and thus reliable estimates could be achieved. It is unclear as to how much these estimates deviate from the true data or how much effect this has on the overall dataset, but it is nevertheless a potential source of error that sadly cannot be remedied in the scope of this paper alone.

Appendix

A Additional details

References

Daniels, Kimberly, and Joyce C. Abma. 2018. *Current Contraceptive Status Among Women Aged 15–49: United States, 2015–2017*. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/products/databriefs/db327.htm#:~:text=The%20most%20common%20contraceptive%20methods,and%20male%20condom%20(8.7%25).

Faria, Julia. 2022. *Male to Female Ratio of the Total Population in Kenya from 2000 to 2020*. Statista. https://www.statista.com/statistics/1226942/male-to-female-ratio-of-the-total-population-in-kenya/#:~:text=As%20of%202020%2C%20Kenya's%20population,the%20country%20registered%20minimal%20increases.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Iannone, Richard, and Mauricio Vargas. 2022. *Pointblank: Data Validation and Organization of Metadata for Local and Remote Tables*. https://CRAN.R-project.org/package=pointblank.

Kenya National Bureau of Statistics. 2019. *Volume 1: Population by County and Sub-County*. https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county.

Office on Women's Health. 2019. *Birth Control Methods*. U.S. Department of Health & Human Services. https://www.womenshealth.gov/a-z-topics/birth-control-methods.

Ooms, Jeroen. 2022. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*. https://CRAN.R-project.org/package=pdftools.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

The Demographic and Health Surveys Program. 1999. *Kenya Demographic and Healthy Survey 1998*. National Council for Population and Development. https://dhsprogram.com/publications/publication-fr102-dhs-final-reports.cfm?csSearch=456440_1.

———. 2022. *Demographic and Health Survey (DHS)*. USAID. https://dhsprogram.com/Methodology/Survey-Types/DHS.cfm.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.