Analysis of Water Security in Jordan

Jonathan Goodwin

02 April 2022

Abstract

Water source and availability are important factors in determining the health and safety of a society. We obtain data regarding the source of water in Jordan and the time it takes residents to access it. We find that the most common source of water is piped into residents, and that a significant portion of Rural residents are receiving water through a tanker truck. These findings have implication for the further development of infrastrcuture in Jordan.

1 Introduction

Sources of water are important in understanding the living conditions of the population of Jordan. Most of the population that answered the survey do use a piped water source into their residence. Access to water both in a timely manner and low cost is important to allowing people to live a secure life.

We analyzed the sources of water for two subsets of the Jordan population, urband and rural residents. From Figure **??** and Figure 1, we can see that both urban and rural residents primarly source their water from pipes into their residence directly. However we also find that tanker trucks make up a significant source of water for rural residents.

This information is relevant to the further development and enhancement of living conditions for rural and urban residents of Jordan. Gaps in the availability of water for rural and urban residents can be identified and addressed by the Government of Jordan.

In section 2 we talk about the process of gathering and analyzing the dataset. Then in section 3 we discuss the implications of the dataset presented. Finally in section 4 we discuss implications for the Jordan residents, as well as the weakness of the study and the further investigation that may be useful to the topic of water security for the residents of Jordan.

2 Data

The dataset is from the Jordan Population and Family Health Survey(JPFHS). The survey was conducted by the Department of Statistics, (Statistics (DOS) [Jordan] and Inc.(MI) 1998) from June 7, to October 31, of the year 1997,through the Demographic and Health Survey(DHS) program.The survey was funded by the U.S Agency for International Development, while Macro International Inc. provided technical support of the survey.

The survey covers 7600 households, the sample is nationally representative, the intent of the survey was to provide information on levels of fertility, infant and child mortality and family planning. Of this survey, this analysis is focused on the information regarding the source of water and the time availability of water. This information was available in the JPFHS report on page 32, table 2.7.

To compile the data set, the R language was used, (R Core Team 2020), along with the packages, Pointblank (Iannone and Vargas 2022), haven (Wickham and Miller 2021), the paper was compiled using Knitr (Xie 2021) and KableExtra(Zhu 2021) packages. The table was acquired through the PDF document using the package PDFtools (Ooms 2022). Also made use of reshape2 (Wickham 2007) in manipulating the data to creae plots.

Table 1 gives a small look at the dataset.

Table 1: Excerpt of dataset

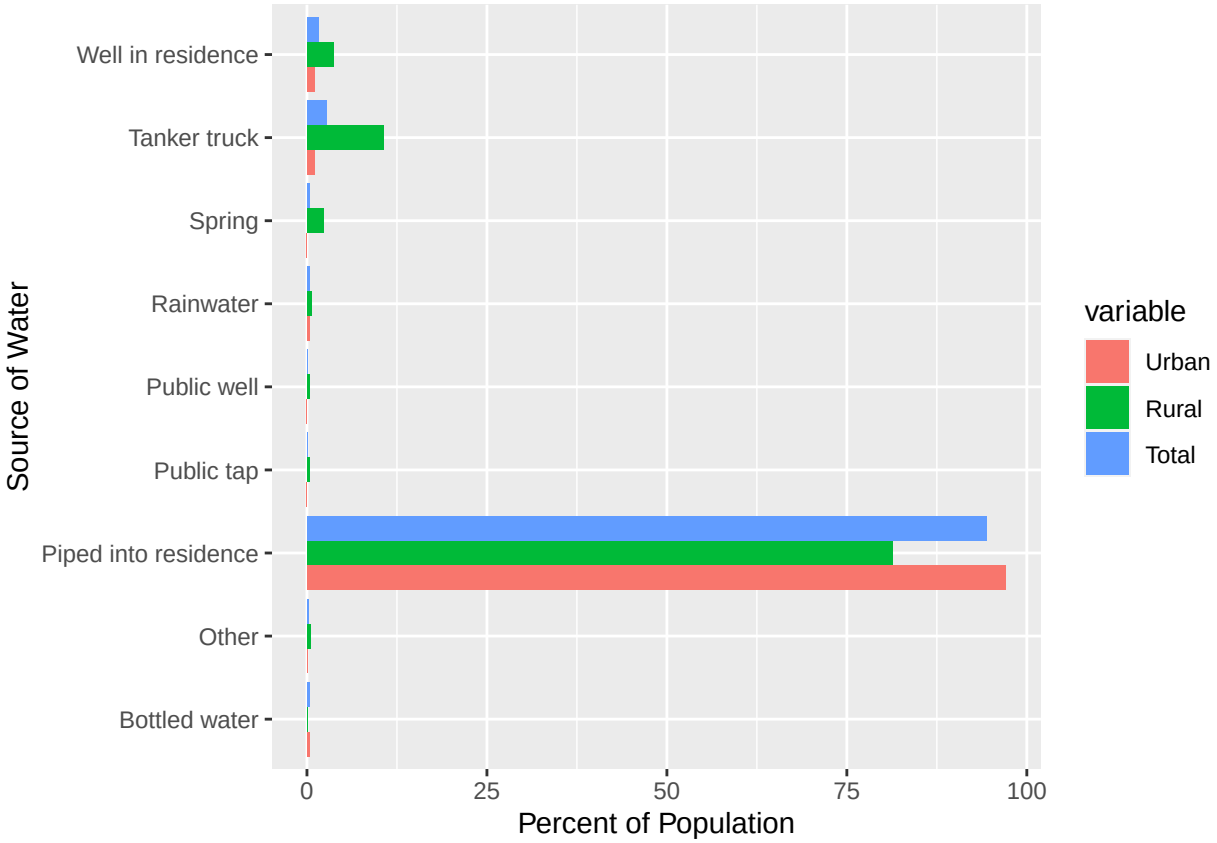
Characteristic	Urban	Rural	Total
Source of drinking water	NA	NA	NA
Piped into residence	97.1	81.4	94.4
Public tap	0.0	0.3	0.1
Well in residence	1.1	3.7	1.6
Public well	0.0	0.4	0.1
Spring	0.0	2.3	0.4

There are a total of 10 different sources of water captured by the survey, Piped into residence, Public tap, well in residence, public well, spring, rainwater, tanker truck, bottled water, and other. Respondents were grouped into two categories, urban and rural.

Additionally the time to the water source is given as an over or under of 15 minutes.

3 Results

From Figure **??** we see a breakdown comparing the urban and rural residents sources of water, along with a total for the population of Jordan. From this table we see that the vast majority of residents from both populations receive their water via a pipe directly into their residence. This is a positive as it means these people have quick and direct access to a water source.



In Figure 1 we take a closer look at a similar plot but with the piped into residence source removed.

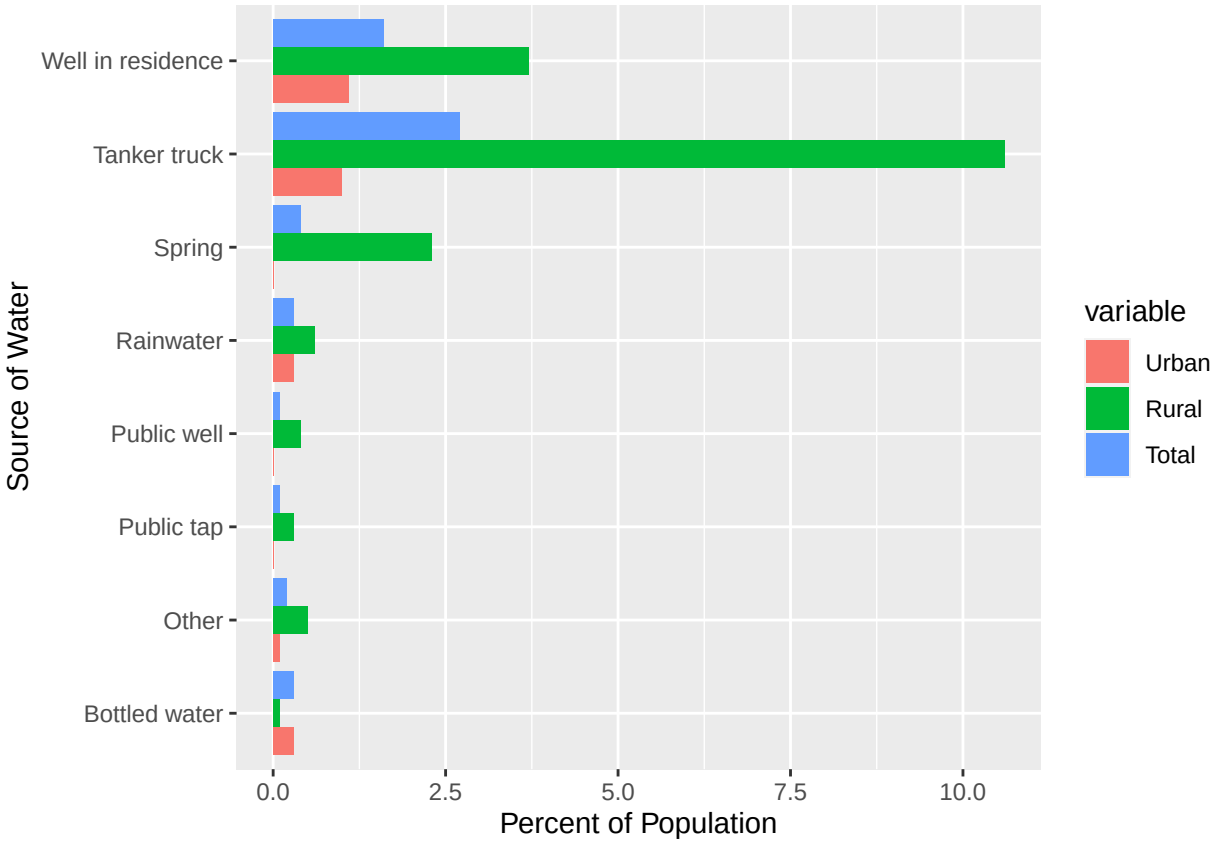


Figure 1: Bar plot without piped into residence response

Unsurprisingly the majority of natural water sources such as wells are being used by rural residents, making up 3.7% of all rural water sources compared to the 1.1% asmong urban residents. Also Springs are only used by rural residents, making up 2.3% of rural sources.

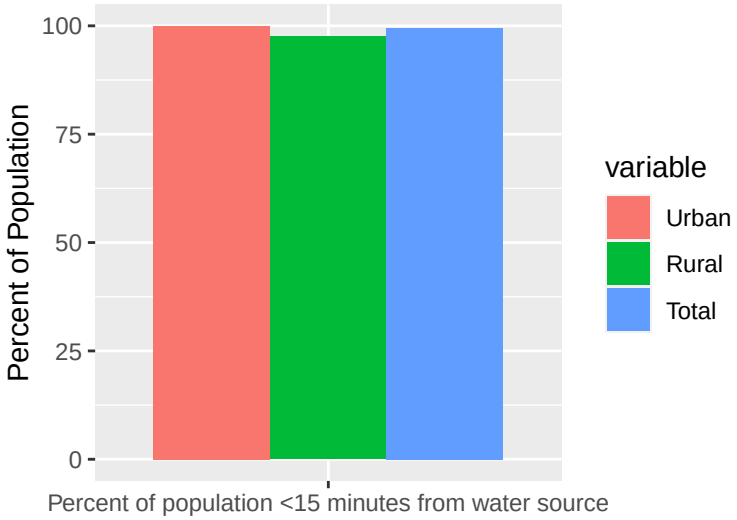


Figure 2: Water available within 15 minutes

As we can see both rural and urban residences can access water quite readily in the country of Jordan.

4 Discussion

4.1 Water in Jordan

According to the United States Agency for International Development(USAID 2021) "Jordan is one of the most water-scarce countries in the world," the country faces problems with it's ground water resources being drained twice as fast as they can be replenished. This explains the lack of use of wells we see in the country in the dataset and the possible over reliance on bottled and tanker truck water among the rural population.

According to (USAID 2021) one of the primary issues affecting Jordan is the loss of water through leaky pipes, theft and underbilling, the metering of water in the population. Since Jordan lacks naturally occurring water resources the effective transportation of their water is going to be the primary way the country can reduce the costs of delivering water to its citizens.

4.2 Weaknesses and next steps

More information regarding the availability deficiencies would be the next steps. With pipes being both the primary source but also according to points made in section 4.1 it's not just important for Jordan that residents have piped water for easy access but also the quality of those piped sources as the country faces issues of water conservation. Further investigation regarding the quality of those pipes would help to establish a better idea of the water

Appendix

DataSheet

Datasheet questions from Gebru et al. (2021)

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
    - The dataset was created to analyze the source and availability of water in different populations of Jordan.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
    - Jonathan Goodwin
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
    - No one.
4. *Any other comments?*
    - No.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
    - Each row of the dataset corresponds to a characteristic, with values of the percent of rural and urban residents of Jordan that fit that characteristic, with a final column giving a total for the entire population.
2. *How many instances are there in total (of each type, if appropriate)?*
    - There are 10 rows of data, 9 correspond to sources of water, one corresponds to the time availability of the water source.
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
    - The dataset is all instances of those that responded to the questionaire, so it does not encompass the entire population of Jordan, the larger set. According to the DHS the sample is representative of the national population.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
    - Each instance consists of a numeric value for the percentage of residents which responded affirmatively to that characteristic in the survey.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
    - No.

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
   - No.
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
   - No.
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
   - No.
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
   - No.
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*
    - The dataset does not rely on any external sources.
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
    - No, the data is all publicly available from the Department of statistics Jordan and required no special permissions to access.
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
    - No.
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
    - The sub populations are urban and rural residents of Jordan.
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
    - No individuals can be identified from the dataset.
15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
    - None of the data is of a sensitive nature.
16. *Any other comments?*
    - No.

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*
    - The data was originally acquired by the Department of Statistics[Jordan], (Statistics (DOS) [Jordan] and Inc.(MI) 1998) via the Jordan Population and Family Health Survey. A table from that report found on page 32, was then used for this analysis.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
    - Subjects for the survey were interviewed in person.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
    - The sample was from the Jordan population and the sampling strategy of multistage stratified design was used in collecting the sample.
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
    - The Department of Statistics in Jordan, as well as Marcus International Inc. under contract funded by the United States Agency for International Development.
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
    - The data was collected from June 7 of 1997 to October 31 of 1997.
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
    - DHS surveys are approved and review by the ICF Institutional Review Board(IRB).(Demographic and Surveys, n.d.).
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
    - Individuals of the survey were contacted directly for the survey.
8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*
    - Participants are notified of all aspects of the
9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
    - According to the DHS, informed consent statements are read to respondents who may accept or decline participation.(Demographic and Surveys, n.d.)

10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
    - Participants of the survey are told they may terminate participation at any time.
11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*
    - The original survey was subject to ethical review this particular subset of that study has not been.
12. *Any other comments?*
    - No

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*
    - In the process of building this dataset
2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*
    - Yes, both the raw data acquired through the table, and the cleaned version of the dataset are available in the repository associated with this analysis.
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
    - The R language and packages associated with the cleaning process are all freely available.
4. *Any other comments?*
    - No.

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
    - Prior to this analysis the dataset was only used as part of the original Survey.
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
    - Yes, it is available on github here
3. *What (other) tasks could the dataset be used for?*
    - The dataset has been minimized for this analysis but the full raw data includes other factors regarding water security that may be of interest.
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*

- No.

5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
   - No.
6. *Any other comments?*
   - No.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
   - No.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
   - The dataset is available via Github.
3. *When will the dataset be distributed?*
   - The dataset is currently available via Github.
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
   - No.
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
   - No.
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
   - No.
7. *Any other comments?*
   - No.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
   - The dataset will be available on Github
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
   - No.
3. *Is there an erratum? If so, please provide a link or other access point.*
   - No.
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*

- No.

5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*

   - No.

6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*

   - No.

7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*

   - No.

8. *Any other comments?*

   - No.

Code

Repository associated with this analysis is available at github]

References

Demographic, and Health Surveys. n.d. *Protecting the Privacy of DHS Survey Respondents*. https://dhsprogram.com/Methodology/Protecting-the-Privacy-of-DHS-Survey-Respondents.cfm.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

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.

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/.

Statistics (DOS) [Jordan], Department of, and Macro International Inc.(MI). 1998. *Jordan Population and Family Health Survey 1997*.

USAID. 2021. *STRENGTHENING WATER SECURITY*. https://www.usaid.gov/jordan/fact-sheets/strengthening-water-security.

Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.

Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.