The Lack of Accecibility to Media Outlets in Tanzinia in 1996 May Explain Absences of Benefits Seen Through Mass Media*

Data largely uses data on females

Oluwabusayomi Adekuajo

11 April 2022

Abstract

The significance of communication on the development of a country cannot be overestimated as the ability to spread information to a large audience supports community and its progression. Lack of progression in developing countries raises the concern of whether beneficial effects of mass media can be observed. In this paper the prevalence of mass media outlets in Tanzania in 1996 was determined, in which most individuals in rural did not have access to any media sources included. Our findings have implications to develop affordable access to media through public spaces.

Contents

*Code and data are available at: https://github.com/rubyzero10/starter_folder-mainPaper4.

1 Introduction

Communication, more than ever, has great significance to civilisation, so much so that is necessary for survival. The constant development of technology has led to the progression of media to what is known as mass media. Its key function is to communicate various messages to a large audience; common outlets include radio, newspapers, television, internet, social media, and more. The ability to bring information to the masses concerns the successiveness of a society, as some communities have the resources, such as easy access to internet, to take advantage of these communication outlets more than others that would face struggles to obtain the same source. As the strength of cooperation, obtained through utilization of mass media, is necessary for the development of a community, after the late 90s, New print, radio, and TV enterprises were established, from commercial media outlets to church and community-based stations, in east African developing countries including Tanzania(Spurk and Dingerkus (2017)). The benefits of this new media was predicted to promote democratic support and enhance, accountability and participation of the public sector (Spurk and Dingerkus (2017)).

Proper implementation of mass media was expected to have led to numerous positive outcomes, however, fulfilment of such expectation had yet to be met despite its investment in east Africa. In this paper, to determine that people are taking advantage of available mass media, we will analyse the proportion of mass media in the different regions of Tanzania. We compare the prevalence of all media outlets, along with discover possible preferences of media outlets. Media use was further compared by gender to visualize possible influence of gender roles on access to media. With this information we tried to determine circumstances in which benefits of media could be seen in developing countries, specifically Tanzania. Findings show that there was a large percentage of both men and women, in multiples regions of Tanzania, that did not use any forms of mass media. Out of those that that did use any or all forms of mass media, the most common use of was radio when compared to reading the newspaper or watching television. Findings are in line with past research that determined that the development of radio in Tanzania has been far more rapid than that of the press, with the potential of reaching the whole country (Mytton (1968)). Radio is the most common source of media in developing countries due to its affordability and use of less electricity. Thus our finings imply that in order to achieve the right environment to support the development of media, economic viability must be considered and optimized.

The remainder of the paper is split into four sections. Section 2 explains the data source and collection methodology, potential bias and ethics issues, as well as our selected data's characteristics. Section 3 presents the findings from our methodology. Section 4 provides discussion of results, the weaknesses as well as next steps to expand on the findings of the report. Section A includes data sheet for the dataset put together.

2 Data

Data was available in, PDF format, from the DHS Final Report from a 1996 survey in Tanzania. In a reproducible way, dataset was transformed, using the R statistical programming language @ citeR, to be manipulated and analysed. Figures and graphs were created with the packages ggplot2 Wickham (2016), knitr Xie (2021) , kableExtra Zhu (2021), tidyr Wickham and Girlich (2022), dplyr Wickham et al. (2021), modelsummary Arel-Bundock (2022), patchwork Pedersen (2020) and tidyverse Wickham et al. (2019). This report uses 20 observations of 13 variables.

2.1 DHS

The 1996 Tanzania Demographic and Health Survey is a nationally representative survey with the main purpose to provide information on fertility, family planning, infant and child mortality, maternal and child health and nutrition, knowledge and attitudes of AIDS, and female circumcision(Bureau of and Inc. (1997)). The TDHS is the third national samples survey of it its kind, practicing surveys since 1991. The general objected of the1996 TDHS, which concerned the data used, is national-level data that will allow the calculation of demographic rates, particularly fertility and childhood mortality rates(Bureau of and Inc. (1997)). The purpose of the TDHS was to provide estimated to the whole country including rural and urban areas.

2.2 Survey's Methodology

The TDHS sample was achieved through a three-stage design consisting of the same 357 enumeration areas (EA) used in 1991-92, 262 from rural and 95 in urban areas(Bureau of and Inc. (1997)). The first of the collection stages involved selecting a ward, in which branch or EA was selected from with those words on the second stage. This type of sampling is refered to as stratified sampling. After a list of households from the EA was selected, households from this list were randomly selected by start number in the third stage.

After the sampling process describes above, 8900 households were selected in which 8141 were occupied as the target sample size. 7969 individuals were interviewed showing a response rate of 98 percent. Out of the eligible 8900 households, 8501 was the target population for females age 15- 49, in which 96 percent were actually interviewed. In contracts, only 2658 was the target population for men age 15-59, from that 85 percent were interviewed. The TDHS thus collected data from 8,120 women aged 15-49 and 2,256 men aged 15-59 in Tanzania in 1996.

2.3 Strength

The field staff conducting the interviews had to undergo intensive training consisting of interviewing techniques, detailed review of questionnaire items, mock interviews and much more, before being selected for the fieldwork. A series of assessments were given and graded from selected interviews and those who shows extra understood and had the ability to detect error in completed questionnaires were later chose to be editors. This training period consisted over the course of three weeks.

Individual men and women were selected from the household interview. The interviewer for this was different than the one who performed the individual interview. This was done to reduce error due to age shift in participants representing the oldest and youngest age group. After interview was completed, team editors, which had previously selected for excellence, checked to ensure that all relevant questions were properly recorded, that the skip pattern instructions were followed, and that responses were internally consistent. Each team was instructed to complete the editing work and resolve all errors found in the questionnaires before household was left. Team supervisors that located the households and assigned them to the interviewers. were required to ensure that all the selected households and eligible women and men in a cluster were interviewed, and that assignment sheets for the interviewers and supervisors were filled out completely and correctly.

2.4 Weakness, Ethics and Bias Issues

The selection of households on the Tanzania DHS surveys was unlike most others. Due to this, if material from other countries represented through the DHS is being compared to Tanzania, there must be cation that error may existed due to mythology. The modified selection process was used to minimise the difficulty encountered in moving from one selected household to another given the scattered nature of households, but this also results in possible error due to clustering of information. This must be cautioned when estimating data onto population.

Table 1: Comaprison of Media Distribution All Media Access and No Access Amoung Genders

Region	No Media Female	No Media Male	All Media Female	All Media Male
Dodoma	55.6	35.0	4.8	7.1
Arusha	58.0	41.5	4.5	16.0
KiIimanjaro	33.6	24.6	4.6	8.7
Tanga	61.6	41.3	2.0	10.7
Morogoro	62.9	45.5	3.4	5.6
Coast	33.9	12.9	2.5	9.7
Dares Salaam	8.0	2.0	32.5	55.9
Lindi	49.7	14.1	1.6	1.4
Mtwara	70.1	14.9	0.7	0.0
Ruvuma	57.3	27.5	1.1	1.0
Iringa	70.7	48.2	0.8	1.5
Mbeya	54.1	31.9	1.0	5.6
Singlda	65.2	35.7	1.3	3.6
Tabora	57.1	25.9	3.5	7.4
Rukwa	73.4	50.0	1.4	10.3
Kigoma	58.9	24.3	0.5	1.4
Shmyanga	69.1	47.0	2.9	8.5
Kagera	51.4	23.2	1.1	10.1
Mwartza	67.7	26.9	1.0	9.0
Mara	59.6	23.6	0.4	7.3

In urban areas one member households were more common, as they were more likely to lock up their houses for safety. This made interviewing more difficult and response rate lower in such areas. Response rate was also lower among men despite the larger range in age and lower target sample size. This was observed due to lower and more frequent absence of men seen in household at time of interview, due to work status. This fact must be considered when comparing results between men and women.

2.5 Variable Selcetion

Interested in understanding the prevalence of mass media outlets in Tanzania households, data showing percentage of women and men who usually read the newspaper once a week, watch television once a week or listen to radio weekly in Tanzania 1996 was selected. The full table can be seen in Figure 1, in which specifically region data was selected for the purposes of this paper. Data was gathered as PDF and read and manipulated through use of pdftools Ooms (2022) and tidyverse Wickham et al. (2019). Table showing dataset attained through PDF can be seen in Table 1. In each in gender, whether they used no media, read newspaper weekly, watched television weekly, listened to radio weekly, or used all three was collected and distributions of each variable is shown in Table 2. The amount of women and men in each region was recorded, represented as numer_f and numberl_m, in which it should be noted that women were interviewed at a much larger capacity. This data represents percentages, making comparison of gender classes possible.

3 Results

It is evident that there are gender influences on access to media as seen through Table 2. In the female category, the largest mean amongst variables is those with no access to media. This case is not observed in males as, listening to radio weekly is the largest mean seen. Access to all three media is observed to be the smallest mean in both genders, though men still occupy twice the mean of their female counterpart. Interested on how access to all three media and no mass media compared, Figures 2 and 3 were made.

Table 2.12 Access to mass media

Percentage of women and men who usually read a newspaper once a week, watch television once a week, or listen to the radio weekly, by selected background characteristics, Tanzania 1996

	Women						Men					
Background characteristic	No mass media	Read newspaper weekly	Watch television weekly	Listen to radio weekly	All three media	Number of women	No mass media	Read newspaper weekly	Watch television weekly	Listen to radio weekly	All three media	Number of men
Age												
15-19	54.1	15.2	10.2	43.0	5.8	1,732	30.9	23.7	19.9	66.2	12.6	488
20-24	47.7	16.7	11.7	49.9	6.2	1,676	23.2	33.1	25.5	74.3	18.0	371
25-29	52.2	15.1	10.1	45.9	6.0	1,440	23.7	30.1	23.4	72.7	16.3	301
30-34	55.6	11.8	7.9	42.1	3.8	1,118	26.2	29.8	17.6	71.8	12.2	272
35-39	56.5	9.3	6.4	42.3	3.9	888	30.3	22.5	15.3	68.3	11.6	251
40-44	60.1	7.9	4.7	39.0	1.3	680	28.8	29.8	16.0	66.8	10.9	206
45-49	68.7	5.3	3.8	30.3	1.5	585	31.7	20.7	11.4	67.2	7.8	149
50-54	NA	NA	NA	NA	NA	NA	41.9	13.1	3.5	57.6	2.4	118
55-59	NA	NA	NA	NA	NA	NA	50.0	14.0	2.6	45.9	1.9	100
Residence												
Mainland	55.5	13.0	8.1	42.5	4.6	7,881	30.1	25.3	16.5	67.3	11.4	2,187
Total urban	22.5	39.2	25.6	73.6	16.6	1,811	7.8	63.6	42.7	89.3	36.6	509
Dar es Salaam city	7.7	60.7	47.4	88.7	35.0	563	1.5	82.4	69.5	94.1	59.9	171
Other urban	29.3	29.5	15.8	66.7	8.3	1,248	11.0	54.2	29.1	86.9	24.8	338
Total rural	65.3	5.2	2.8	33.3	1.0	6,070	36.9	13.7	8.6	60.6	3.8	1,678
Zanzibar	21.1	14.8	34.4	75.1	9.4	239	4.1	49.4	63.7	92.9	40.5	69
Pemba	29.2	6.4	18.6	68.5	1.7	92	7.4	24.1	18.5	90.7	7.4	28
Unguja	16.2	19.9	44.2	79.2	14.2	148	1.9	66.7	94.4	94.4	63.0	41
Region												
Dodoma	55.6	12.7	10.2	40.6	4.8	355	35.0	16.4	18.6	59.3	7.1	96
Arusha	58.0	15.6	7.9	39.9	4.5	589	41.5	28.7	19.1	57.4	16.0	156
Kilimanjaro	33.6	13.2	9.2	64.4	4.6	390	24.6	21.0	15.9	72.3	8.7	119
Tanga	61.6	5.0	8.8	36.2	2.0	464	41.3	14.7	13.3	57.3	10.7	108
Morogoro	62.9	10.9	5.8	36.6	3.4	408	45.5	18.2	14.0	50.3	5.6	95
Coast	33.9	18.1	6.5	65.0	2.5	159	12.9	22.6	29.0	82.3	9.7	45
Dar es Salaam	8.0	58.6	44.1	88.4	32.5	646	2.0	79.6	65.8	93.1	55.9	191
Lindi	49.7	8.2	4.4	48.1	1.6	187	14.1	19.7	2.8	85.9	1.4	54
Mtwara	70.1	2.3	3.4	27.7	0.7	355	14.9	13.9	1.0	84.2	0.0	96
Ruvuma	57.3	4.5	2.8	41.4	1.1	305	27.5	24.5	2.0	72.5	1.0	82
Iringa	70.7	4.9	2.6	28.8	0.8	466	48.2	12.4	2.9	48.9	1.5	100
Mbeya	54.1	12.4	1.6	42.4	1.0	473	31.9	25.0	6.9	66.7	5.6	137
Singida	65.2	9.9	3.8	31.5	1.3	283	35.7	16.7	11.9	60.7	3.6	80
Tabora	57.1	9.1	7.6	40.9	3.5	225	25.9	33.3	11.1	72.2	7.4	82
Rukwa	73.4	5.1	3.4	24.9	1.4	242	50.0	12.8	11.5	48.7	10.3	71
Kigoma	58.9	4.6	2.5	40.3	0.5	351	24.3	14.3	5.7	72.9	1.4	95
Shinyanga	69.1	9.1	4.0	29.6	2.9	686	47.0	21.3	9.1	52.4	8.5	202
Kagera	51.4	7.7	5.6	45.1	1.1	467	23.2	20.3	17.4	73.9	10.1	139
Mwanza	67.7	8.1	1.6	31.9	1.0	573	26.9	17.9	14.1	70.5	9.0	176
Mara	59.6	8.7	0.4	39.7	0.4	257	23.6	25.5	18.2	69.1	7.3	64
Education												
No education	76.3	0.2	2.3	23.2	0.1	2,316	60.8	2.4	5.9	37.5	0.8	304
Primary incomplete	61.1	6.1	4.8	36.9	1.2	1,630	35.5	17.5	11.4	61.5	6.1	664
Primary complete	43.6	18.9	10.4	53.7	5.8	3,732	21.6	29.9	19.6	76.1	14.0	1,066
Secondary+	7.1	57.0	45.1	87.9	33.0	441	5.0	65.6	46.2	91.4	38.6	222
Total women/men	54.5	13.1	8.8	43.5	4.8	8,120	29.3	26.1	18.0	68.1	12.3	2,256

NA = Not applicable

Figure 1: Full raw dataset

Table 2: Summary of numerical ditrubution indicators variables for regions

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
no_mass_media_f	20	0	55.9	15.5	8.0	58.5	73.4	
read_news_weekly_f	19	0	11.4	11.8	2.3	8.9	58.6	
watch_tele_weekly_f	18	0	6.8	9.2	0.4	4.2	44.1	
listen_radio_weekly_f	20	0	42.2	15.1	24.9	40.1	88.4	
all_f	18	0	3.6	7.0	0.4	1.5	32.5	
number_f	19	0	394.1	151.2	159.0	372.5	686.0	
no_mass_media_m	20	0	29.8	13.2	2.0	27.2	50.0	
read_news_weekly_m	20	0	22.9	14.4	12.4	20.0	79.6	
watch_tele_weekly_m	20	0	14.5	14.0	1.0	12.6	65.8	
listen_radio_weekly_m	20	0	67.5	13.0	48.7	69.8	93.1	
all_m	18	0	9.0	11.8	0.0	7.3	55.9	
number_m	17	0	109.4	44.5	45.0	96.0	202.0	

Figures 2 visualizes lack of a mass media source and its distribution among individual regions of Tanznia. It is further shown that females, across all regions have, on average, less access to media compared to males. When comparing to Figure 3 representing amount of those with access to all three media, results show that other than Dares Salaam, an urban city, all regions have less access to all media than those with no access to media. These results indicate lack of resources to economically support all three sources.

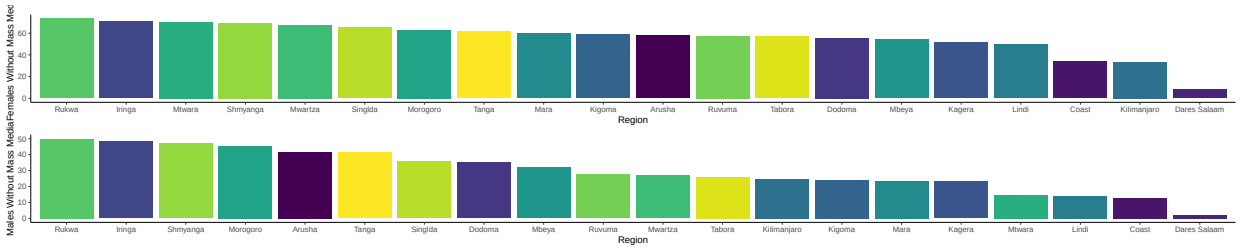

Figure 2: More women have no access to media

When availability of resources increase, we were interested to know individuals pattern in mass media collection. Just looking at male individuals, Figure 4 demonstrates that when males do not have access to all three media, when x axis is 0, they will listen to radio weekly the most. In fact, the other two sources of mass media, are almost not explored by anyone in this category. There is strong positive linear relationship in having all three media, watching tv and reading the newspaper. There is a weak, almost no relationship seen, however between listening to radio weekly and having ample access to mass media. This indicates that as more resources are available to spend on more mass media, these other forms of communication, other than radio, are then explored. It is possible that they are seen not as a necessity to get information but more of a treat. Figure 5 displaying the same information for female observes a different trend. All three variables have a positive correlation with using all three methods of media, though more females will listen to the radio more as their access to media increases.

When made to prioritize a source to invest in personally, which media outlet is most used from participants? When a large amount of the females in the population are listening to the radio, there is a small amount reading the newspaper, and an even smaller amount watching tv as seen in Figure 6. Indicating that watching

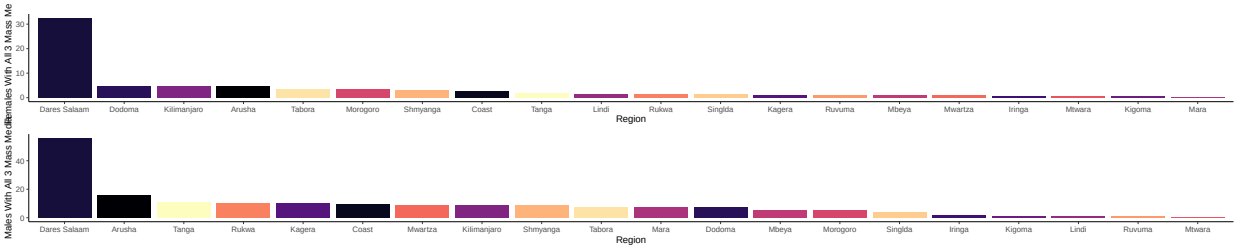

Figure 3: Urban city is shown to have most access to all 3 mass media

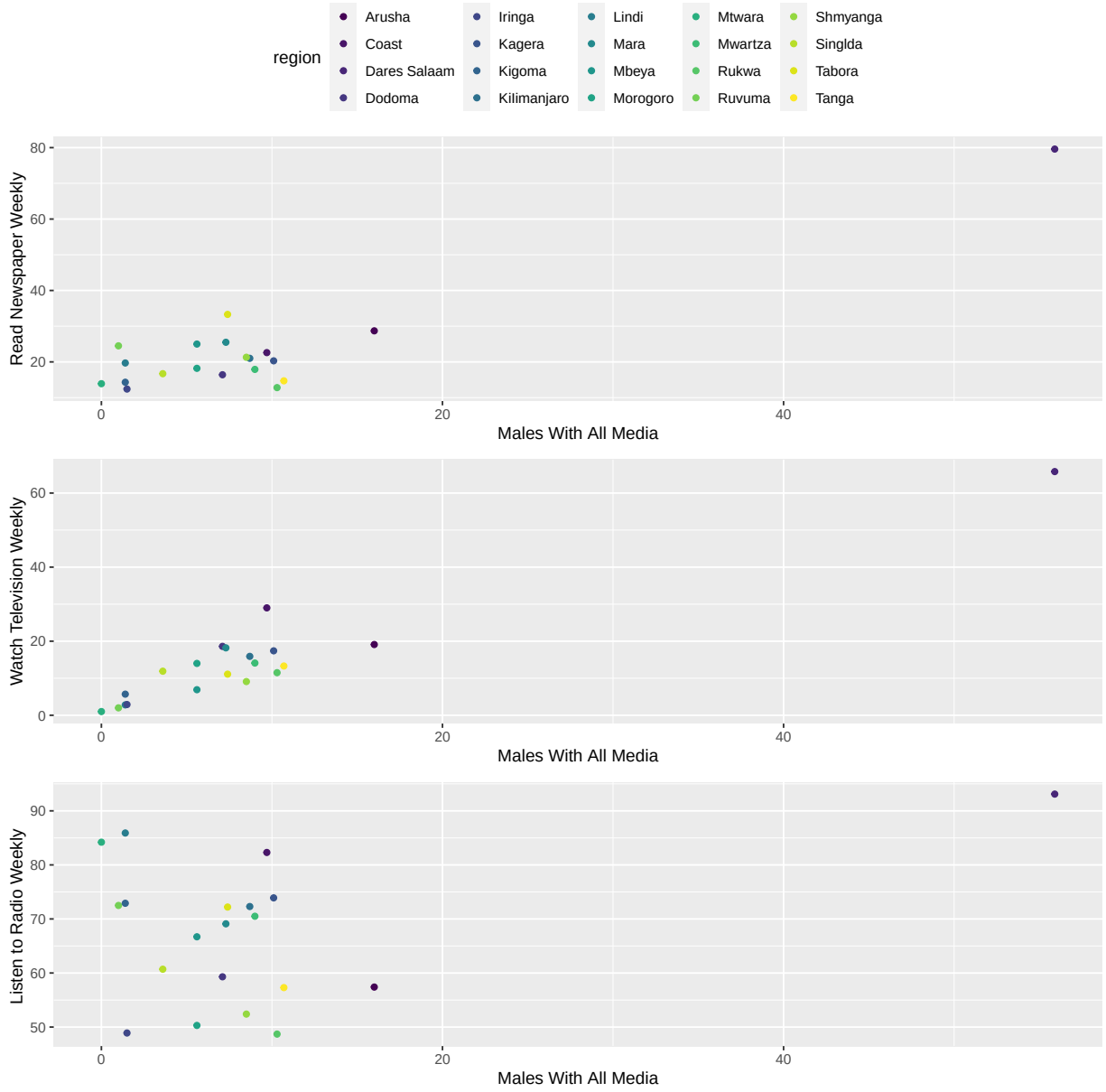

Figure 4: Watching television is desired once eligible

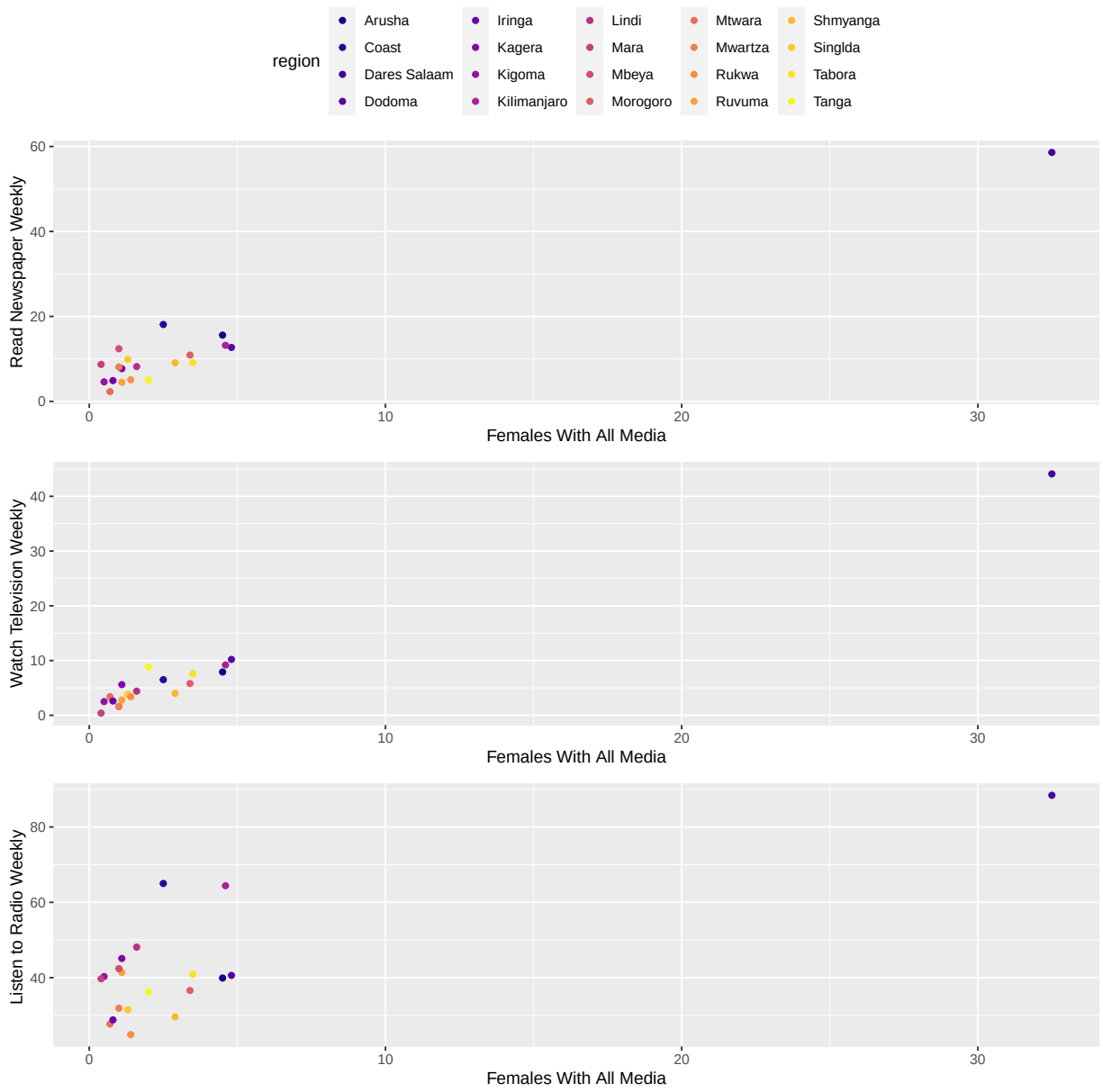

Figure 5: Reading newspaper is next desired mass media

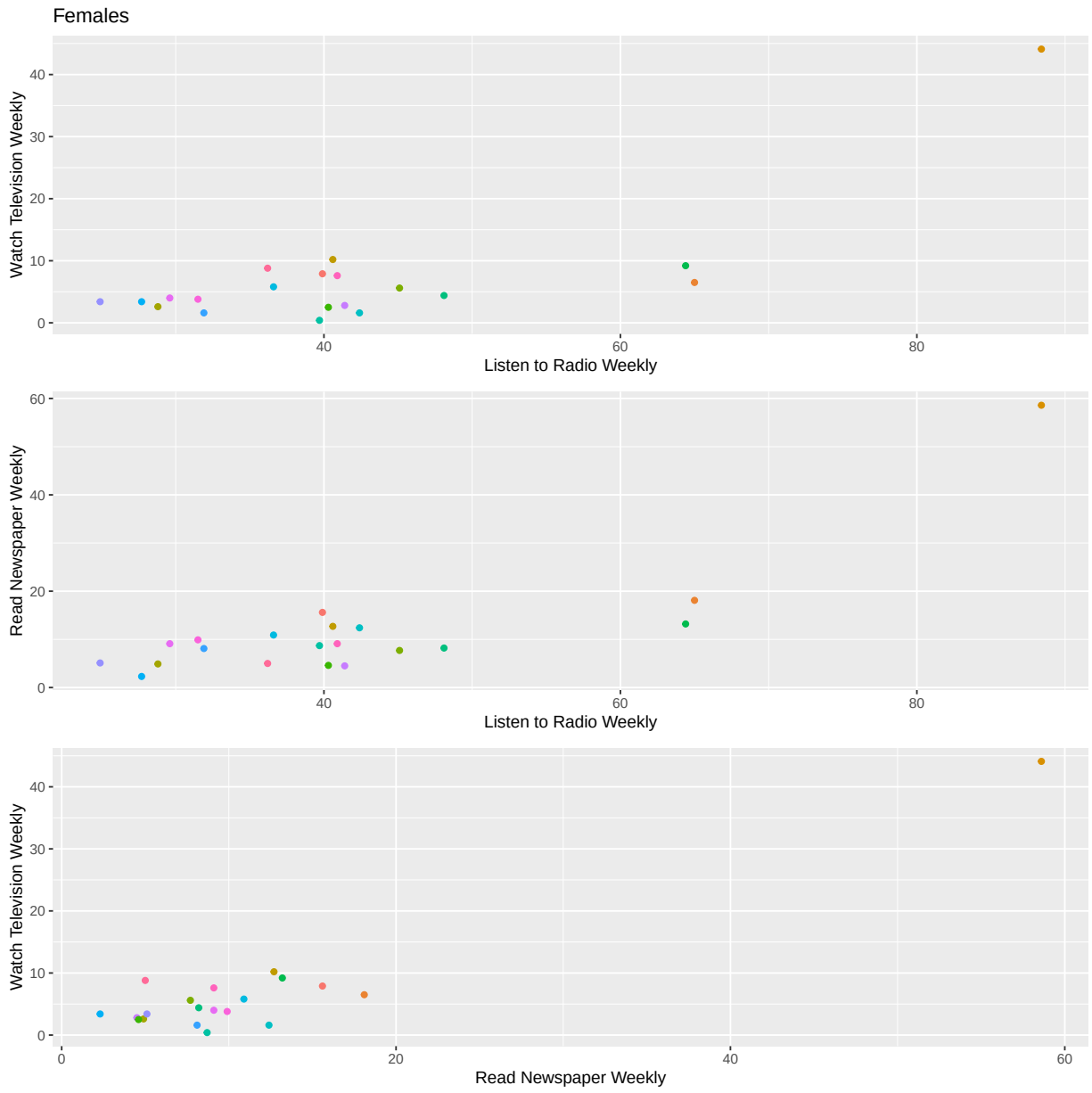

Figure 6: Results confirm reading newspaper is next desired mass media

tv may be seen as the largest investment to this group. Figure 7, however, shows that as a large amount of the male population is listening to the radio, there is more that have starting to watching television than to read the newspaper. Reason for such results were of interest and studied further.

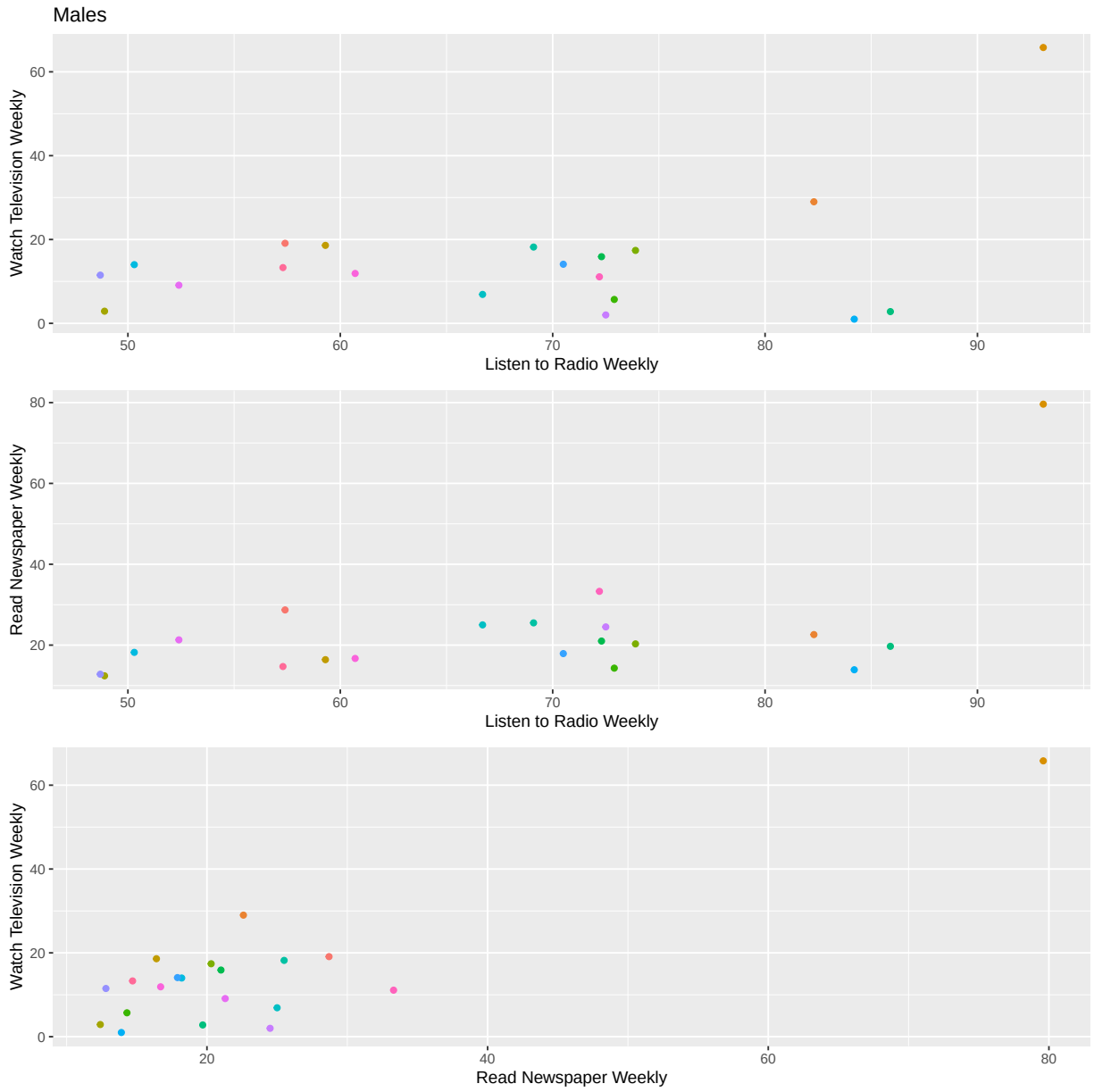

Figure 7: Results confirm watching television is next desired mass media

4 Discussion

4.1 Agriculture

Though there were discrepancies between gender comparison, overall, radio was the most common source of mass media used by those that had used at least one form of media. Radio has been used as a powerful communication tool that has proven to be the most effective media in stimulating agriculture and the development in the rural areas (Bernard and Dulle (2014)). In Tanzania, agriculture plays a huge role in their mixed economy, contributing the largest shared of any sector to the Gross Domestic Product (GDP)(Bureau of and Inc. (1997)). Out of the sources of media studied, radio and television would serve the most useful sources in for farmers. A study found they in fact, out of the sources of media studied, radio and television would serve the most useful sources of agricultural information to farmers in addition to notifying farmers of new agricultural innovations and emergencies(Bernard and Dulle (2014)). Low levels of television access were observed, in both males and females when only one source could be accessed. Though as resources increased in male participants, watching television was the next desired source of media. In terms of affordability, access to a television would require a high source of income as it requires an adequate source of power. It is understandable why radio is most used as it required less electricity and thus is capable of spreading messages to audiences that just have a receiver with a sufficient power. Print, however, though expensive and majority privately owned, would be more affordable than television. Benefits to agricultural information, which in turn leads to a higher economy, may explain why television was this invested in before reading the newspaper, in males, the main worker population.

4.2 Literacy

Newspaper though slightly more popular in the female demographic, was not a heavily sought-after source of media. To be able to effectively get information out of a newspaper, the audience must be able to read and understand its inputs. Tanzania is known, among east African countries, to have a higher literacy rate a literacy rate of 80%(Spurk and Dingerkus (2017)), nevertheless, it rural communities tend to have high poverty levels and thus lower literacy rates. The population of Tanzania is predominantly rural, with the proportion of urban residents increasing from 6 percent in 1967 to 18 percent in 1988(Bureau of and Inc. (1997)). Without proper educational resources to understand the newspaper, it remained utilized as a media choice in a small portion of participants. High illiteracy levels, poverty, lack of electricity, mixed with the high cost of information and communication technologies, resulted in Tanzania's rural areas constrained to a poor and unreliable communication infrastructure(Ogessa and Sife (2017)). This idea is represented in the data as majority of participants did not use any source of media. Though mass media can provide opportunities in developing countries, the lack of actual use of mass media nulls possible effect as a whole.

4.3 Weaknesses and next steps

Data collection mainly represents data on women. This is despite the large emphasis that mass media has on the agriculture demographic, which is mostly male. With the low sample size selected within this groups, external reliability of results is questionable. Implication of this study may be biased towards the female population, however independent analyse on male data provided is possible in the future. It has been established which forms of media are popular, along with the overall use of media in the Tanzanian population.

The results of this study imply that mass media must be more affordable in order to be optimized. As radio play is already popular and efficient, developing more accessible radio communication to the public could increase effectiveness of mass media. This because even though radios are the cheapest option, there are still a multitude of individuals that cannot afford to have one in their home. By providing radio play in a public setting such as a radio station, would alleviate an individual purchasing barrier, resulting in benefits of mass media to be seen. It is possible that further development has occurred since data collection in 1996.

Appendix

A Datasheet

Extract of the 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 establish the effectiveness of mass media in developing, specifically, Tanzanian populations.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
    - The data was created by the Bureau of Statistics, in collaboration with the Ministry of Health. Technical assistance was provide by Macro International Inc.
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.*
    - The TDHS was funded but the U.S. Agency for International Development (USAID) through the worldwide Demographic and Health Surveys programme.
4. *Any other comments?*
    - NA

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.*
    - The instances that the data set represent are background characteristic in categories, age, residence, and region, education, and total.
2. *How many instances are there in total (of each type, if appropriate)?*
    - There are 10 instances of age, 8 of residence, 20 of region, 4 of education,
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 does not contain all possible instances as some are not as related to data collected such as religion.
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 consisted of raw data
5. *Is there a label or target associated with each instance? If so, please provide a description.*
    - The first column of the instance refereed to the background characteristic

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.*
    - There is information missing from the age category for ages above 50, as they were unavaliable
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.*
    - There is no relationship between instances
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.*
    - There are no recommended data splits
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
    - There is not errors, sources of noise, or redundancies in the dataset
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 data set is self-contained
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.*
    - There is not data that might be considered confidential
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.*
    - There is no data that might be offensive, insulting, threatening, or might otherwise cause anxiety
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.*
    - This dataset has sun population divided by instances mentioned earlier.
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.*
    - It is not possible to identify individuals.
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.*
    - Data contain education data that might be considered sensitive in any way.
16. *Any other comments?*
    - NA

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 acquired through interviews done in reported regions.
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?*
    - In person interviews were done.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
    - Dataset used stratified sampling based of wards, and enumeration areas and then further based off randomly selected number.
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)?*
    - Team supervisors located the households and assigned interviewers, one for household interview, one for individual selected.
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 fieldwork for the main survey began in late July 1996 and lasted until November 1996.
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.*
    - No ethical review processes
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
    - Data was obtained through Demographic and Health Surveys website
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.*
    - Notice of collection is unavailable
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.*
    - Individuals consent to collection, exact language to which the individuals consented unavaliable.
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).*
    - A mechanism to revoke their consent is unavaliable
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.*

- An analysis of the potential impact of the dataset and its use on data subject was not conducted

12. *Any other comments?*
    - NA

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.*
    - Data was processed as PDF file and converted using R into a usable data fram
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.*
    - Raw data is saved in outputs/data/raw_data.csv
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
    - Yes R software used to clean the data is available at https://www.r-project.org/
4. *Any other comments?*
    - NA

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
    - The dataset has not been used for any tasks already
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.*
    - https://github.com/rubyzero10/starter_folder-mainPaper4
3. *What (other) tasks could the dataset be used for?*
    - The dataset could be used to see media use amoung age classes.
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?*
    - There is nothing about the composition of the dataset that might impact future uses
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
    - The dataset should not be used to compare recent trends
6. *Any other comments?*
    - NA

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 the dataset will not be distributed to third parties outside of the entity
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 will be distributed throguh Github
3. *When will the dataset be distributed?*
    - The dataset be distributed in Aprl 2022
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.*
    - Dataset will not ne distributed under a copyright or other intellectual property (IP) license
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.*
    - Third parties have not imposed IP-based or other restrictions on the data associated with the instances
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 export controls or other regulatory restrictions apply to the dataset or to individual instances
7. *Any other comments?*
    - NA

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
    - Oluwabusayomi Adekuajo
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
    - Via Github
3. *Is there an erratum? If so, please provide a link or other access point.*
    - There is no erratum
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)?*
    - Datasey wil not be updated
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.*
    - There are no applicable limits on the retention of the data associated with the instances

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.*
    - Older versions of the dataset will no continue to be maintained?
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.*
    - There is no mechanism for other to contribute to the dataset.
8. *Any other comments?*
    - NA

B Additional details

References

Arel-Bundock, Vincent. 2022. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://CRAN.R-project.org/package=modelsummary.

Bernard, RT, and Frankwell W Dulle. 2014. "Access and Use of Mass Media by Small-Scale Farmers in Accessing Agricultural Information for Poverty Alleviation in Tanzania: A Case Study of Kilombero District."

Bureau of, Statistics [Tanzania], and Macro International Inc. 1997. "Tanzania Demographic and Health Survey 1996."

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.

Mytton, Graham L. 1968. "Tanzania: The Problems of Mass Media Development." *Gazette (Leiden, Netherlands)* 14 (2): 89–100.

Ogessa, Catherine M, and Alfred S Sife. 2017. "Newspaper Coverage of Agricultural Information in Tanzania." *University of Dar Es Salaam Library Journal* 12 (1): 12–26.

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

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

Spurk, Christoph, and Filip Dingerkus. 2017. "The Need and the Opportunities for Sustainability–the Case of Local Radio Stations in Tanzania." *Journal of Media Business Studies* 14 (1): 38–59.

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

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.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Maximilian Girlich. 2022. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

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

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.