Spring 5-1-2014

Visualizing the "Hermit Kingdom" Graphing, Mapping, and Analyzing the 2008 North Korean Census

Molly Linhorst

Visualizing the "Hermit Kingdom"

Graphing, Mapping, and Analyzing the 2008 North Korean Census

A Capstone Project Submitted in Partial Fulfillment of the Requirements of the Renée Crown University Honors Program at Syracuse University

Molly Linhorst
Candidate for B.A. Degree
and Renée Crown University Honors
May 2014

Honors Capstone Project in Political Science

Capstone Project Advisor: _____
Professor Stuart Thorson

Capstone Project Reader: _____
Professor Frederick F. Carriere

Honors Director: _____
Stephen Kuusisto, Director

Date: April 23, 2014

Abstract

Discussion rarely takes a dispassionate view of North Korea and its people, and ingrained aversion to the idea of promoting a better understanding of the state distracts from serious, empirical study of the population. This research makes use of the 2008 North Korean census, an extensive set of fifty-three tables, which received little attention and remained in PDF format. This research project first translates and publishes machine-readable formats of the census data and then employs R statistical programming to create visualizations of select tables and conduct simple data analysis, including p-tests and Mantel-Haenzel tests. Males and urban residents were found to be less likely to suffer physical impairment than females and rural residents, allowing us to reject the null hypothesis that there is no proportional variation. Maternal mortality ratios are found to differ across provinces and are lowest in Pyongyang. Governments and humanitarian aid organizations can use these findings to effectively target vulnerable populations. This study sets precedence for further analysis of the 2008 North Korean census.

Table of Contents

Executive Summary

This study emerges in response to the limited analysis and discussion of the 2008 North Korean population census. Discussion rarely takes a dispassionate view of North Korea, and sensationalized reporting on the country's latest nuclear developments often overpowers empirical research on the state and its population. The extensive set of high-quality information presented in the census, which was conducted by the North Korean government in cooperation with the United Nations Population Fund, provides considerable insight into the North Korean population. The census includes information ranging from highest educational attainment, to cooking fuel used by individual households, to age at the respondent's first marriage, and occupation. The data was released publicly by the United Nations as a PDF in 2009 but has never been released in a machine-readable format, like an Excel document or a CSV (comma-separated values) file. Without a machine-readable version of the data, any analysis was labor-intensive.

The first goal of this research project was to translate all the tables (there are fifty-three in total) from a PDF into separate machine-readable files. This required a considerable amount of time to ensure that all numbers, column headings, and row headings transferred correctly. After the data was "cleaned" (spacing, headings, and the structure of the tables were standardized), the data was published online, making analysis of this extensive dataset accessible to anyone with a computer and an Internet connection.

To ensure that all readers, regardless of their level of statistics training and understanding, can engage with the census information, visualizations were made.

The author constructed all visualizations through a free statistical computer program titled *R*. The program, which the author taught herself in anticipation of completing this project, is open-source. This means that the program is publicly available online and anyone is allowed to offer suggestions to improve the program. This program was chosen to produce visualizations and statistical analysis so that the author's work is easily and freely reproducible. Visualizations produced using *R* include maps, bar graphs, and dot plots.

Five data tables from the original census were used to create all the visualizations: table 2, providing population counts for province and the administrative districts within each province; table 18, providing counts of maternal mortality in each province and the place of maternal death; table 42, providing counts of citizens reporting difficulty seeing; table 43, providing counts of citizens reporting difficulty hearing; table 44, providing counts of citizens reporting difficulty walking or climbing stairs. Each of these tables, except table 2, provides insight into health in North Korea. Health therefore became an overarching theme of this research project.

Through the visualizations and simple statistical analysis, a few trends appeared. First of all, maternal mortality ratios – a ratio often used by international organizations to gauge maternal health, a ratio that gives the average number of maternal deaths per 100,000 live births – vary between North Korean provinces. Of the ten provinces, Pyongyang City had the lowest maternal mortality ratio, meaning fewer mothers die from pregnancy complications in Pyongyang than elsewhere in North Korea.

The same trends appeared in the data detailing visual impairment, hearing impairment, and difficulty walking or climbing stairs. In all three cases, a lower proportion of men than women reported impairment, and a lower proportion of urban residents than rural residents reported impairment. Statistical tests showed that these significant differences in proportion were unlikely to appear by chance. Further testing found that the variables gender and urban/rural are not independent from each other.

Research of the available literature and interviews with experts on North Korea revealed certain cultural and social norms that may have influenced the data and could explain some of the trends identified earlier. Among these norms is the expectation of the woman to sacrifice her well-being for the well-being of her husband, father-in-law, and son. If this thought process is applied to healthcare, especially when resources for healthcare are so limited, women may be sacrificing their healthcare to ensure that men receive the appropriate healthcare, thus leaving more women with physical impairments than men. There is also the potential for systematic error if the stigma associated with disability, combined with the government's classification of difficulty seeing, hearing, and walking as disabilities, deterred some respondents from answering honestly that they have difficulty seeing, hearing, or walking for fear that they would be branded as disabled.

Overall, although the analyzed data do not produce an abnormally high maternal mortality ratio nor do they produce an abnormally large number of respondents with physical impairments, there remain significant differences in

women's health and men's health, as well as rural residents' health and urban residents' health. These differences may be explained by redistribution of resources by elites and party officials to benefit themselves and the capital city of Pyongyang. North Korea lacks enough medical resources, and its hospitals sometimes lack electricity, heating, or running water, which has produced a black market for healthcare underneath the socialist North Korean healthcare system.

This research is valuable to any governments or organizations that work in North Korea, providing detailed information on the nation's population. These findings are particularly valuable for humanitarian or development aid organizations that wish to better target vulnerable populations. The data analysis herein, and any data analysis produced from the now freely available machine-readable census files, can help organizations more effectively reach the citizens of North Korea and address their needs.

Professors Stuart Thorson and Fred Carriere provided considerable assistance throughout the research and analysis process.

Preface

When my friends ask me about the work I do relating to North Korea as a research assistant for the US-DPRK Scientific Engagement Consortium, they like to tease me. "How's your buddy Kim Jong-un today?" they might ask. After the newest announcement of some missile test, they ask me how likely it is that North Korea will attack the U.S. in the next few days. They ask me how I can possibly conduct research about the country, or how I would ever engage its people – "How would you possibly study the country? No one can get in or out. Why would you even want to?"

No doubt working with North Korea in any capacity is a challenge, and I have only barely been exposed to the frustration that comes with directing humanitarian aid or academic exchanges with the country. But it's the conversations I have with friends and family, paraphrased above, that keep me engaged. I feel that I have a duty to try to understand what others do not wish to understand. And then, I have a duty to share that understanding.

North Korea is a nation like any other; it has interests, history, relationships with other states, and rational political actors. It is a member of the international community, and within that context it acts in its own interest. As an individual who aspires to continue along a path of diplomacy and cross-cultural exchange, I seek a deeper understanding of North Korea as a state and as a people. The 2008 DPR Korea Census is one way to begin to understand a country that deserves more of our attention.

After studying for four months in Beijing, I was fed up with censorship; I recall trying to read breaking stories about the Newtown shootings and receiving blank Safari pages, annoyed that I could not connect with my home. The daughter of two journalists and the sister of yet another, I cannot stand limited information, and I take great pride in some of America's most basic protections. Whether in China, the U.S., or elsewhere, limited information stifles understanding, collaboration, and innovation. Although a grand opening of the North Korean state is not likely in the near future, little by little, we can engage the nation, explore its complexities, and answer the questions that prevent further engagement.

The following capstone project answers multiple questions:
What can we learn from the 2008 North Korean census?
How can we use open-source programming and data to improve our world?
How can we engage the general public in statistical analysis?
And perhaps most importantly:
What can one student do in the comfort of the Honors Lounge to further our understanding of a "closed" country halfway across the globe?

This project is only the beginning of a larger project of understanding.

Molly Linhorst
April 23, 2014

Acknowledgements

Completing this project speaks not only to the author's hard work, but to her friends', family's and advisors' support. It takes a village to raise a child. Similarly, it takes a village to write a thesis; it cannot possibly be a solitary act.

First and foremost, she thanks Professor Stuart Thorson for his ideas, encouragement and support over the last four years. Professor Thorson recognized some potential in the author during her freshman year, and she is incredibly grateful for the opportunities he has awarded her. She likely would never have pursued an interest in the DPRK without his influence. Furthermore, his assistance with the R scripts herein was invaluable.

The author also heartily thanks Professor Fred Carriere for acting as her Honors Reader and sparking an ongoing curiosity in the DPRK. His passion and understanding of the Koreas is inspiring.

Geunho Ryu, Shin Yon Kim, and Daniela Lopez all contributed to the tedious task of cleaning the census data. The author expresses deep gratitude for their help.

Other professors have played in meaningful role in the capstone process, and have significantly enriched the author's undergraduate career. These professors include Kristi Andersen, Gavan Duffy, Terry Lautz, Robert Rubinstein, Donald Siegel, Zhou Lanjing, Dai Yishin, Colin Elman. Without these professors, the author's path would look dramatically different.

The author expresses gratitude for her roommates' patience. A senior's spring semester is a stressful time, and the author could not have completed this thesis without her roommates' understanding and support. ΦΟΨΡ deserves a special "thanks," as well as Jackie White for her chocolate-covered espresso beans.

Finally, and arguably most importantly, the author thanks her parents and brother for their love and support, as well as for her brother's constructive competitiveness, which has energized her to achieve. Without them, the dream of a college degree would remain merely a dream, and the dream of a rich, rewarding life would never have included four wonderful years at Syracuse University.

Introduction

In stark contrast to the drama of the Kim dynasty, enlivened with stories of violent executions and colorful NBA visitors, the 2008 census of the Democratic People's Republic of Korea (hereinafter North Korea or DPRK)[1] has received little attention. This extensive set of information was publicly released with limited analysis and news coverage.

The census provides considerable insight into the everyday lives of North Koreans yet remains a seldom-used empirical framework for discussion. Ranging from educational attainment, literacy, and rate of maternal deaths, to households' source of water supply and choice in cooking fuel, the census provides a wide window into the "Hermit Kingdom."

This study emerges in response to the lack of analysis and attention paid toward the 2008 census. In an effort to educate laymen, policymakers and academics alike, the author seeks to analyze and create visualizations of the available data. Further, to ensure that the data is publicly and freely available, this project incorporates a comprehensive transfer of all census data into machine-readable formats that are available online. Ultimately, *Visualizing the Hermit Kingdom: Graphing, Mapping, and Analyzing the 2008 North Korean Census* sets a precedence for further study of the available census data.

[1] Democratic People's Republic of Korea is the official state name of North Korea. Publications produced by the North Korean government use either this full name, its abbreviation as DPRK, or the condensed DPR Korea. For the purpose of this paper, the author will primarily use North Korea to refer to the country.

Public Discourse

Countless myths, caricatures, and mysteries surround the nation sometimes termed the "Hermit Kingdom." Americans' perceptions of resonate with images from the popular comedy *Team America*, in which a puppet representation of Kim Jong Il sings in stereotypically broken English about his loneliness in his global domination efforts.[2] Serious discussion of North Korea typically dissolves into generalizations. Ingrained aversion to the idea of promoting a better understanding of the state and potentially working *with* the North Korean leadership inhibits constructive research and discussion. The "Hermit Kingdom" remains a dark spot in public consciousness. The international community forms broad, generalized impressions of North Korea based on limited, widely circulated news reports about the state regime's latest "provocation." Only rarely does discussion take a dispassionate view of the nation and its people.

Politicians' sweeping remarks, like President George W. Bush's 2002 "axis of evil" accusation, and media outlets' alarming announcements of the newest nuclear threat monopolize public understanding of the country. According to a 2013 poll conducted by Gallup, Inc., 83% of Americans perceive the North Korea's nuclear development to be a "critical threat," tied with the nuclear development program of Iran.[3] Gallup's 2014 World Affairs poll found that among Americans, North Korea is the least favorable country. Only 11% of respondents reported their opinions toward the country to be at all favorable. The

[2] "Team America: World Police," Dir. Trey Parker, 2004.
[3] Jones, "In U.S., 83% Say North Korean Nukes Are a Critical Threat."

nation's favorability among the American public has not risen above 15% for more than a decade.[4] With statistics like these paired with (and possibly triggered by) sensationalized reporting, public discourse veers toward broad generalizations of the state and its leadership. Rarely do empirical studies about the lives of North Korean citizens make their way into the global public consciousness.

In the following chapters, the author shifts attention from the usual focus on the state's nuclear program and other much-publicized, highly controversial information, focusing instead on the census data available about the nation's 24 million citizens. Although the author's discussion of her findings in the latter portion of this paper explores relevant concerns about the North Korean leadership and the dissemination of public goods, her analysis remains data-driven.

Implementation of the 2008 Population Census

The United Nations Population Fund (UNFPA) and the Central Bureau of Statistics (CBS) of North Korea cooperatively launched the 2008 North Korean Census in October of 2008, using October 1st at 12:01 AM as the reference point for information gathering. This census followed fifteen years after the nation's first census in 1993. The census follows the de jure method[5], and uses direct interviewing by door-to-door enumerators to collect information.

[4] Wilke, "North Korea Least Favorable Among Nations."
[5] The de jure method of census enumeration collects information about citizens by their place of permanent residence. This method contrasts with the de facto method, which counts residents where they happen to be at the time of enumeration.

The UNFPA was heavily involved in the project from its inception. DPR Korea requested UNFPA support for a national census in 2006. Shortly thereafter, the Population Fund formed a mission to explore the possibility of a census, ultimately agreeing to assist in conducting with census with several stipulations detailed in a memorandum of understanding as follows: the census will observe international standards for data collection; the census will cover the entire country; the United Nations will release the information to the global community; international experts will directly assist in operations, and one chief technical advisor will reside in the country; the UNFPA will have access to all aspects of the census and will mobilize any additional resources deemed necessary.[6] The UNFPA designed the questionnaire, trained enumerators and supervisors, and raised the necessary funds. The fund used for the census included $4 million from South Korea and 500,000 Swiss francs to cover supplies and training. With UN recommendations and training, a system of supervisors acting under the CBS sent 35,000 enumerators across the country. From October 1st through October 15th, 2008, these enumerators conducted house-to-house interviews.

Before sending enumerators out to canvass the nation, the UNFPA and CBS conducted a precautionary pilot census. One year prior to the date of the actual census, 50,000 households across the ten provinces of North Korea were interviewed to test the fairness, sufficiency, and conditions for implementation of the planned census. Additionally, twelve UNFPA-affiliated demographers monitored the census. The Central Bureau of Statistics and the UNFPA released

[6] Engracia, "UNFPA's role in the DPRK's 2008 Population Census."

preliminary census results in early 2009, ultimately releasing a PDF file of the census results – including all 53 tables of data – later in 2009. This 278-page document is freely available for download on the United Nations Statistics Division website.[7]

Accuracy and Reliability of Data

Overall, United Nations organizers found the census data to be reliable and accurate, corresponding whenever possible to international standards for population and housing censuses. UNFPA Representative in China and Country Director in North Korea Dr. Bernard Coquelin reported positively on the census and the professionalism of its enumerators.[8] According to the United Nations Statistics Division (UNSD), the 2008 census produced accurate data:

> All the 5 observation teams observed that the procedures of the census taking followed by the DPRK conform to the general international procedures. No major problems were observed and the errors have been or will be corrected through the supervision mechanism, and at the manual editing and coding stages. The data collected by the enumerators observed is accurate enough to ensure a good census. The teams also observed the commitment from the local authorities and census workers, and the good cooperation and hospitality from the respondents.[9]

Similarly, Luisa Engracia, Chief Technical Advisor for the UNFPA-North Korean collaboration on the census, reports that the collected data is of high quality. Engracia reports widespread cooperation from the public as well as support and

---

[7] Access the PDF here: https://unstats.un.org/unsd/demographic/sources/census/2010_PHC/North_Korea/Final national census report.pdf

[8] Gharzeddine. "UNFPA Helps Plan and Monitor Successful DPRK Census."

[9] "The Democratic People's Republic of Korea 2008 Census." *United Nations Statistics Division.*

commitment from all levels of government. Engracia finds the enumeration process to be well organized and enumerators to be well trained.[10]

To ensure quality data, the organizers also conducted a post-enumeration survey, re-interviewing a sample of North Korean households. From this post-enumeration survey, organizers could estimate the degree of error in the census data. Unfortunately, the United Nations has not publicly released this post-enumeration survey.[11] In her presentation of the enumeration process to the National Committee on North Korea, Engracia cites the net coverage error to be less than 0.1%.[12] Thomas Spoorenberg, a statistician with the UNSD, cited another quality assurance check conducted by the United Nations, which found that "based on demographic assessment comparing the populations that were enumerated in the 1993 and 2008 censuses, the data quality of the 2008 census is also good (if not excellent)."[13]

Like other international humanitarian aid and development organizations, the UNFPA had to cooperate closely with the North Korean government to gain access. Unfortunately, working with the North Korean government can be challenging; interviews with humanitarian aid organizers quickly reveal frustration. Furthermore, because the central government often controls where, when, and how the data is collected, experts advise skepticism when analyzing this data. The central government exercises this control in part because the state is still technically in a state of war with South Korea. The World Food Programme

[10] Engracia, "UNFPA's role in the DPRK's 2008 Population Census."
[11] Spoorenberg, "Student Thesis on 2008 DPRK Census."
[12] Engracia, "UNFPA's role in the DPRK's 2008 Population Census."
[13] Spoorenberg, "Student Thesis on 2008 DPRK Census."

(WFP), for example, did not have access to one third of the country.[14] Monitors with the WFP had to report planned visits for distribution sites to the state as many as six days ahead of time.[15] These restrictions limit the organizations' data collection to sample populations potentially unrepresentative of the whole population, and provide the state enough time to alter the conditions evaluators met on site.

These concerns have not appeared in either UNFPA or UNSD documentation, and personal communications with experts in the field have not revealed significant reason to believe the 2008 census includes inaccurate or unrepresentative data. That said, despite the generally healthy skepticism warranted under these circumstances, statisticians, enumerators, and North Korea experts overwhelmingly find the data to be of high quality.

[14] Bertini, "Collecting Data in North Korea."
[15] Pinkston and Saunders, "Seeing North Korea Clearly."

Methods

Machine-Readable Data

One of the possible explanations for limited analysis of the 2008 census stems from the simple fact that the data was not released in a machine-readable format. Scholars and analysts can conduct minimal analysis on such complex sets of information without importing it directly into statistical computer programs. Although some programs claim to be able to strip data from PDFs and present these data in a machine-readable format, the programs rarely translate the data into numerical values, translating them instead into text that is not usable in statistical software. Furthermore, these programs sometimes provide inconsistent or missing data. After scouring the web, the author could only find the original PDF of the census data. The absence of a machine-readable dataset may have discouraged academics and analysts from studying the census. The census data is publicly available, yet it is not available in a convenient format. To encourage further analysis and discussion about the 2008 DPR Korea Census, ensuring that the data is accessible and downloadable is essential. For this reason, the author has chosen to transcribe the data to be machine-readable. This process is the first component of the project, and the first step to analyzing and visualizing the data with any depth.

Although a seemingly elementary task, translating the 53 tables from a PDF into a machine-readable format, like Excel or CSV, is time-consuming, tedious, and occasionally challenging. What's more, this process could potentially produce inaccurate data. Producing a machine-readable format of this census data

involves both transferring the data to a different form as well as "cleaning" the data to ensure that it is clear and compatible in this new machine-readable format. It is probable that while transferring and cleaning the data, digits are accidentally altered or eliminated. When analyzed, these incorrect data might lead to incorrect discoveries. This process, therefore, is significant for the author's project, and requires great care and patience.

For this project, the author has converted the census data into two popular and adaptable formats: Excel files and CSV (comma-separated values) files. The author created a separate Excel file and a separate CSV file for each individual table, making 106 files overall that are machine-readable, clearly labeled, and consistent with the original census document. After initial transfer, the author tested random portions of each dataset to check its accuracy by comparing it visually to the original census PDF. She also summed random chunks of cell values and compared her summations to the totals provided by the PDF. Nevertheless, the author recommends that users of the machine-readable datasets keep the original PDF of the census at close hand; although she tried to make all headings and titles clearly describe the data therein, some datasets require more context to understand their content. Furthermore, there remains room for unnoticed human error.

Upon finishing the transcriptions, the data files were published publicly online. They are available for download on the Maxwell School's Korean Peninsula Affairs Center website. Interested readers can find the data here.[16]

[16] http://maxwell.syr.edu/moynihan/kpac/Datasets/2008_DPRK_Census_Data/

R Software

This report employs secondary data analysis to provide insight into North Korea's population. Using both visualizations of the data – including bar graphs, dot plots, and political maps – and mathematical analysis, the author explores trends in the North Korean population. Consistent with the author's overarching goal to make the extensive dataset publicly available, the following analysis is conducted using R statistical software.

R is an open-source software freely available online.[17] Using R to create all graphics and conduct all statistical tests allows others to reproduce or improve the author's work. R uses its own programming language, and therefore requires some training. Fortunately, there exists a large virtual community around the open-source program, so nearly any coding question, no matter their complexity, can be answered by a simple Google search. R ensures accessibility as long as a machine and Internet connectivity are available. R therefore embodies the spirit of the author's work, seeking to engage the public discourse in available, empirical, and understandable data analysis.

The author shares her R "scripts" – the coding she has written to conduct statistical analysis or visualize the data – in Appendix 2. This is to ensure reproducibility.

Visualizations

Arguably more telling than numbers, visualizations can present a clear picture of life and population trends in North Korea. The author seeks to provide

[17] R can be downloaded at: http://www.r-project.org/

the public with empirical information in a widely understandable format. Visualizations are her chosen strategy for opening a window on to the Democratic People's Republic of Korea. The audience, regardless of experience with statistical analysis, will thereby have the opportunity to inform its judgment and discourse based on empirical evidence.

R has powerful visualization capabilities. Through the base graphics programmed into R and a series of freely available and installable packages, more complex visualizations are possible, including mapping. The ability to map the census data creates the possibility of analyzing North Korea not only from an external, international perspective, comparing figures across state borders, but from a domestic perspective, visually exploring differences *between* provinces. Mapping could prove valuable domestically to DPRK officials as well as international NGOs that seek to target specific issues among the North Korean population.

The author includes visualizations of all the census tables she analyzes. These visualizations include bar graphs, dot plots, and maps. All visualizations are created using R and are available in Appendix 1. The R scripts used to create the visualizations comprise Appendix 2.

Statistical Analysis

Statistical analysis, testing relationships between variables, is a necessary addition to visualizing the data. Although visualizations may appear to show certain trends, such trends may not be statistically significant. Statistical significance tests will serve as a "check" to the visualizations herein. In laymen's

terms, statistical significance calculations measure the likelihood that a figure could have occurred by chance alone, rather than caused by a relationship between the variables. A statistically significant difference in a sample is one that is unlikely to have occurred by chance. To test significance, the author employs the p-test (the prop.test function in R). For the purpose of this study, p-values of .01 or less will result in calling a relationship statistically significant and rejecting the null hypothesis that there is no relationship between the two variables.

In the discussion section, following the presentation of visualizations and data analysis, the author presents some potential sources of systematic error that cannot be analyzed statistically.

Research Questions

This project seeks to provide a window through which to understand the North Korean population as a whole, as well as the framework around which to understand differences *within* the North Korean population. The goals of this analysis are deliberately open-ended to encourage further research and discussion. However, four specific tables from the 2008 North Korean Census serve as the backbone to this project.

Table 18 provides information about maternal deaths nationwide, including the location of death (at home, at the hospital, or elsewhere). Given the available information, the author creates visualizations of maternal death trends across all ten North Korean provinces. Calculating the maternal mortality ratio – a figure commonly used to make international comparisons – provides insight into maternal health differences globally and between DPR Korea provinces.

Table 42 presents information on visual impairment, measuring North Korean citizens' "difficulty seeing." This table measures different levels of difficulty, ranging from no difficulty to blindness. The data specifies age group, gender, and whether residents reside in urban or rural areas of the country.

Table 43 displays hearing impairment data, measuring "difficulty hearing." Like the previous table, hearing impairment is divided into four levels of difficulty ranging from no difficulty to deaf. The data specifies age group, gender, and whether residents reside in urban or rural areas of the country.

Finally, Table 44 also measures respondents' physical impairment, dividing respondents into four levels of difficulty walking or climbing. The data specifies age group, gender, and whether residents reside in urban or rural areas of the country.

These four datasets alone provide insight into variance across the North Korean population, particularly in their delineations between gender and urban or rural residence. Bar graphs and maps visualizing the tables begin to shape a deeper understanding of the North Korean people and, with some educated guesswork and statistical analysis, the state of healthcare in North Korea. Although comparing maternal mortality rates and select proportions of the physically impaired across national boundaries, the majority of the research herein aims to understand differences within North Korea. Recognizing significant provincial, gender, or urban-rural inequities[18] may have considerable

[18] The published census data does not clearly delineate between "urban" and "rural" areas. The author discusses this further in the *Data and Visualizations* section of the following chapter.

implications for international NGOs planning humanitarian efforts in North Korea. Inequities may also direct researchers to interesting trends that provide insight into North Korean society, economy, and leadership. Understanding and analyzing this data is a starting point to a larger effort.

Given the aforementioned tables and the information therein, several questions drive the following research. First and foremost, does the maternal mortality ratio vary between provinces? Given the residence of the North Korean elite in the capital and more urban centers, as well as the widely held suspicion that the elites redistribute resources to their own communities, it is logical to expect to see a lower maternal mortality ratio in Pyongyang and more urban provinces.

Tables 42, 43, and 44 do not specify respondents' home provinces, but they do delineate between urban and rural areas more generally, prompting the question: Do urban and rural areas, as defined by the 2008 census, observe different proportions of physical impairment? If proportions prove to be significantly different, and assuming the theory that urban areas receive more resources, it is logical to expect to see greater physical impairment in rural areas.

The latter three tables analyzed extensively in this study also delineate by gender. Are there significant inequities in physical impairment between men and women? Given the historical repression of women in Korean society, if inequities do indeed exist, it is reasonable to expect to see higher levels of impairment among women because they do not receive an equal amount of healthcare.

Maternal Mortality

Introduction

Table 18 of the census presents the number of maternal deaths reported in the preceding twelve months as of October 1, 2008. The census data includes women who died during pregnancy or childbirth as well as women who "subsequently died due to maternal cause." Inspection of the original census questionnaire reveals that the data also counts all female deaths occurring within 42 days of a miscarriage, abortion, or giving birth as maternal deaths. The table specifies the location of maternal death, providing three choices: home, hospital, or other. This table serves as a starting point for exploring the issue of maternal health, and healthcare more generally. The 2008 census was the first to collect data on maternal deaths in North Korea.

International institutions such as the United Nations and the World Bank use the measure of maternal mortality ratio (MMR) to gauge maternal health. The MMR is a standardized measure that allows for comparisons across countries and regions. As a result, this study uses MMR as its primary variables in exploring North Korean trends in maternal health. To get this measure, the number of maternal deaths is divided by the total number of live births during the same time period, then multiplied by 100,000. The figure gives the number of maternal deaths per 100,000 live births.

Data and Visualizations

According to the 2008 population census, the national MMR of the time period October 1, 2007 through October 1, 2008 is 77.22365, or roughly 77.

MMRs for individual provinces vary from a low of 68.46531 in Pyongyang to a high of 85.99274 in Ryanggang. The range of provincial MMRs is 17.53. Appendix 1a lists all ten DPR Korea provinces and their corresponding MMR. Appendix 1b presents provincial maternal mortality rates on a political map of the state.[19]

Appendix 1c, while not directly related to maternal mortality, shows the percentage of each province that is considered to be "urban" by the 2008 North Korean census. Because Table 18, which provides maternal mortality data, only defines respondents' residences by the province and not by the degree to which her residence is in an urban or a rural area, this additional map in appendix 1c helps readers correlate province of residence with the degree to which the province of residence is urban or rural. Comparing these two appendices (appendix 1b and appendix 1c) introduces the new variable of "urbanness" to the maternal deaths data. Appendix 1c uses data from Table 2 of the census, which presents population by province and gender. This dataset also divides provinces into more specific administrative districts: city, district and county.

For the purpose of this project, it is important to note how census enumerators differentiate between "urban" and "rural." The released census document gives little explanation, briefly defining "urban" areas as enumeration areas containing an average of 180 households and "rural" areas as enumeration

[19] Note that four administrative districts – Kaesong, Kumgangsan, Rason, and Sinuiju – are delineated in the country map outline used in the appendices but are *not* included in Table 18 (maternal deaths data). Appendix 1b's map therefore does not include these four districts, leaving them colorless. Appendix 1c's map does include these districts because Table 2 (population by province, city, district and county) includes specific data on these districts.

areas containing an average of 150 households. It is unclear, however, when in the enumeration process the enumerators determined these averages. The published hard copy of the census, produced by the North Korea Central Bureau of Statistics, offers more detailed information on how enumerators define "urban" and "rural" areas:

> In the DPRK, the classification of areas into urban and rural areas is based more on administrative standards than on the physical character or features of the area. This classification is applied on the lowest level of administrative units…DPRK classifies all *dong's, up's,* and *ku's* [sic] as urban areas and all *ri's* [sic] as rural area – irrespective of population size, density or features that globally characterize urban areas. Notwithstanding this classification scheme, most *dong's* and *up's* [sic] do have urban characteristics. Similarly, the *ri's* [sic] …are typically rural, in the global sense of the word. The *ku's* [sic]…can actually have urban or rural features…"[20]

The above explanation codes "urban" and "rural" areas by administrative constructions. Because the Central Bureau of Statistics roughly equates these administrative regions with common global characterizations of "urban" and "rural," the author interprets "urban" and "rural" to be associated with their conventional characteristics.[21]

Several bar graphs follow the previous maps of North Korea, offering another way to illustrate provinces' unequal maternal mortality rates. Appendix 1d presents all ten provinces' MMRs in a horizontal bar graph, followed by a dot plot displaying the same information in appendix 1e. Appendix 1f reorders the data in the dot plot to give a different illustration of the data. Appendix 1g incorporates information on the place of maternal death (at home, at a hospital, or

[20] *Population Census Atlas of the DPR Korea*, 2009.
[21] i.e. Urban corresponds to higher populations, higher population densities, more infrastructure.

elsewhere) to form a stacked horizontal bar graph. To make provincial differences in place of death clearer, the author produced two additional horizontal bar graphs. Appendix 1h includes two bar graphs presenting the provincial percentages of maternal deaths at home and at the hospital.

Major Findings

Within North Korea, there exists a wide gap in maternal mortality ratios between provinces. Consistent with the original hypothesis, the capital (Pyongyang) has the lowest MMR. Not all of the other provinces support the original hypothesis that the more urban areas will have lower maternal mortality ratios. Ryanggang, for example, falls between 60% and 70% urban but has the highest maternal mortality ratio, while other provinces that are less urban, like South Hwanghae, between 30% and 40% urban, report a lower MMR. "Urbannness" does not necessarily predict maternal mortality rates. However, it is important to note that the lowest maternal mortality rate is found in the country's capital.

Overall, more women die from childbirth or pregnancy complications at home than at a hospital.[22] Differences in place of maternal death also vary by province. Pyongyang reported the highest percentage of maternal deaths occurring at a hospital. Following a similar trend to maternal mortality ratio, Ryanggang reported the highest percentage of maternal deaths occurring at home. With the exception of Pyongyang, comparison between the "urbanness" of the province

[22] The third possible place of death offered in the census questionnaire was "other." Because of the vagueness of this term, the author has omitted visualization or analysis of maternal deaths occurring at "other."

and the percentage of maternal deaths occurring at a hospital does not present any clear trend. Similarly, no obvious trends appear when comparing maternal deaths at home to a province's percent "urbanness."

Readers should note that there are differences in "urbanness" within provinces as well; the aggregations used to shade the map in appendix 1c do not capture the variance *within* provinces. The print publication produced by the Central Bureau of Statistics includes one map that plots the distribution of urban and rural areas, as defined by their administrative units, providing a more detailed map of "urban" and "rural." Assuming the census indeed does define "urban" and "rural" purely by administrative units, comparing maternal mortality rates to this map could capture the nuances within provinces, and could lead to a more tailored hypothesis about the relationship between "urbanness" and maternal mortality ratios. However, because Table 18 in the published census data does not provide more specific information about geography than the overarching province, detailed analysis within provinces is currently challenging.

Maternal Mortality Data in an International Context

In an international context, North Korea's maternal mortality ratio is not alarmingly high, especially considering the high ratios in developing countries. The CIA World Factbook, citing information gathered in 2010, estimates a North Korean MMR of 81 and ranks North Korea's ratio as the eighty-third highest internationally. According to this data, North Korea is flanked by El Salvador, Cabo Verde, and Argentina with MMRs of 81, 79, and 77, respectively. The CIA

World Factbook gives the United States a 2010 ratio of 21 and South Korea a ratio of 16.[23]

The World Bank's online visualizations of maternal mortality ratios provide insight into regional trends in maternal mortality. Appendix 1i displays a map produced by the World Bank to provide readers with a visual global context of maternal mortality.

[23] "Country Comparison: Maternal Mortality Rate." *Central Intelligence Agency*.

Difficulty Seeing

Introduction

Tables 42, 43, and 44, measuring difficulty seeing, difficulty hearing, and difficulty walking or climbing respectively, do not provide respondents' resident province. Instead, these tables show whether respondents are from urban or rural areas.[24] These tables also include respondents' gender.

Table 42 reports census findings on visual impairment, asking respondents one question: Does _____ have difficulty seeing even if wearing glasses? The census does not provide any further definition for "difficulty," but gives respondents four response levels: no difficulty, some difficulty, a lot of difficulty, and cannot do at all.

Data and Visualizations

The majority of respondents, regardless of gender and whether residing in an urban or rural area, report having no difficulty seeing; 2.40% of respondents reported any difficulty. The author created three visualizations for the data, all bar graphs. Appendix 1j is a horizontal bar graph comparing respondents reporting "a lot of difficulty" seeing. Although percentages are small, there are clear visual differences between urban and rural women as well as between men and women more generally. The same holds true for the following graph, appendix 1h, which graphs respondents reported to be blind, or "cannot see at all." This graph, in which rural women report highest rates of blindness, especially highlights

[24] These tables provide no further insight into how enumerators differentiated between urban and rural areas. Please refer to the author's discussion of the urban-rural delineation in the previous chapter.

gendered sight differences. Finally, aggregating respondents that reported any level of difficulty, appendix 1i continues the trends observed in the previous two bar graphs; a higher percentage of women than men report having difficulty seeing, and a higher percentage of rural residents than urban residents report having difficulty seeing.

Testing statistical significance with the function *prop.test* in R, the author found that at every level of difficulty, men were less likely to report difficulty seeing. Additionally, respondents in urban areas were also less likely to report difficulty seeing. In each case, p-scores were reported to be less than 2.2e-16 and a 95% confidence interval that remains below 0. The outcome suggests that the author can reject the null hypothesis that there is no statistically significant difference between the proportion of visually impaired men and women and that there is no statistically significant difference between the proportions of visually impaired urban and rural residents. The *prop.test* results support the alternative hypothesis that males and urban residents have lower proportions of visually impaired respondents than females and rural residents, respectively.

Major Findings

The statistical analysis returned extremely low p-scores in part because the sample is so large, but the proportional differences in visual impairment depending on gender and "urbanness" of residence are still significant. The visualizations and the statistical analysis support the author's original hypotheses that rural respondents and female respondents both report higher proportions of visual impairment than urban respondents and male respondents. These findings

might suggest systematic error in responses (for example, female and/or rural respondents are more willing to admit visual impairment than male and/or urban respondents). They might also suggest inequities in access to healthcare. The author will further explore these potential explanations in her discussion section.

Difficulty Hearing

Introduction

Table 43 reports census findings on hearing impairment, asking respondents one simple question: Does _____ have difficulty hearing? The census does not provide any further definition for "difficulty," but gives respondents four response levels: no difficulty, some difficulty, a lot of difficulty, and cannot do at all. The census includes no further documentation or explanation about hearing impairment. Like the previous table that detailed citizens' visual impairment, table 43 does not provide respondents' resident province but does include whether respondents are from urban or rural areas.[25] Table 43 also includes respondents' gender.

Data and Visualizations

The majority of respondents, regardless of gender and whether residing in an urban or rural area, report having no difficulty hearing. Only 1.73% of the total population reported any difficulty hearing, making hearing impairment the least common impairment among the three that the author analyzes (seeing, hearing, walking/climbing). The author created three bar graphs to visualize the data. Appendix 1m is a horizontal bar graph comparing respondents reporting "a lot of difficulty" hearing. Again, there are clear visual differences between urban and rural women, urban and rural men, as well as between men and women more generally. Appendix 1n, which graphs respondents reported to be deaf, or "cannot

[25] These tables provide no further insight into how enumerators differentiated between urban and rural areas. Please refer to the author's discussion of the urban-rural delineation in the Maternal Mortality chapter.

hear at all," reports much lower percentages of respondents reporting difficulty. This graph presents rural respondents as those most likely to be deaf. Aggregating respondents that reported any level of difficulty, appendix 1o continues the trends observed in the previous two bar graphs; a higher percentage of women than men report having difficulty hearing, and a higher percentage of rural residents than urban residents report having difficulty hearing.

The author then tested statistical significance with the function *prop.test* in R. At every level of difficulty, men and urban respondents were less likely to report difficulty hearing. The highest p-score was received testing whether the proportion of male respondents reported to be deaf reported to be less than the proportion of female respondents reported to be deaf. The p-score returned from this *prop.test* was 5.444e-11, but all other *prop.test*s returned a p-score less than 2.2e-16. All tests returned a 95% confidence interval that did not cross 0. The outcome suggests that the author can reject the null hypothesis that there is no statistically significant difference between the proportion of hearing impaired men and women and that there is no statistically significant difference between the proportion of hearing impaired urban and rural residents. The *prop.test* results support the alternative hypothesis that males and urban residents have lower proportions of hearing impaired respondents than females and rural residents, respectively.

Major Findings

Like the visual impairment tests, the statistical analysis returned extremely low p-scores in part because the sample is so large, but the proportional

differences in impairment depending on gender and "urbanness" of residence are still significant and suggest interesting differences across gender and "urbanness" of residence. The visualizations and the statistical analysis support the author's original hypotheses that rural respondents and female respondents both report higher proportions of hearing impairment than urban respondents and male respondents. Again, these findings might suggest systematic error in responses. They might also suggest inequities in access to healthcare. The author will further explore these potential explanations in her discussion section.

Difficulty Walking or Climbing

Introduction

Finally, the author explored responses regarding difficulty walking. Table 44 presents census findings on difficulty walking or climbing, asking respondents: Does _____ have difficulty walking or climbing stairs? Like the previously explored impairments, the census does not provide any further definition for "difficulty." The questionnaire provides four potential responses: no difficulty, some difficulty, a lot of difficulty, and cannot do at all. There is no further documentation or explanation about this physical impairment. Table 44 includes whether respondents are from urban or rural areas[26] as well as the respondents' gender.

Data and Visualizations

Overall, while a very low percentage of respondents reported having difficulty walking or climbing stairs (2.48%), a slightly higher proportion of the overall North Korean population reported having difficulty walking or climbing stairs than the proportion of the population reported having difficulty hearing or seeing. Just as in the last two chapters, the author created three bar graphs to visualize the data. Appendix 1p is a horizontal bar graph comparing respondents reporting "a lot of difficulty" walking. Again, there are clear visual differences between urban and rural women, urban and rural men, as well as between men and women more generally. Appendix 1q, which graphs respondents that cannot

[26] These tables provide no further insight into how enumerators differentiated between urban and rural areas. Please refer to the author's discussion of the urban-rural delineation in the Maternal Mortality chapter.

walk at all, reports very low percentages and less obvious differences between gender and "urbanness." Aggregating respondents that reported any level of difficulty, the differences become clearer again. Appendix 1r continues the trends observed in the previous two bar graphs; a higher percentage of women than men report having difficulty hearing, and a higher percentage of rural residents than urban residents report having difficulty hearing.

Testing statistical significance with the function *prop.test*, men and urban respondents at every level except one were less likely to report difficulty walking or climbing stairs. Testing for gendered differences between males and females that reported they are unable to walk ("cannot do at all") returned a p-score of 0.1966 and a 95% confidence interval that comfortably included 0. In this case, the null hypothesis that the proportions of men and women that cannot walk are equal *cannot* be rejected. Every other *prop.test* returned p-scores less than 2.2e-16. This outcome suggests that the null hypotheses – there is no statistically significant difference between the proportion of physically impaired men and women and there is no statistically significant difference between the proportions of impaired urban and rural residents – can be rejected. The *prop.test* results support the alternative hypothesis that males and urban residents have lower proportions of hearing impaired respondents than females and rural residents, respectively. However, this conclusion does not hold true for the gendered proportions that cannot walk at all.

Major Findings

Like the visual and hearing impairment tests, the statistical analysis returned extremely low p-scores in part because the sample is so large, but the proportional differences in impairment depending on gender and "urbanness" of residence are still significant. The one exception in this test adds interest. While the overall trends hold true, and the authors' hypotheses are supported, the *prop.test* shows that the proportion of males and females that cannot walk at all are not necessarily unequal; the gender differences illustrated in appendix 1q could be caused by chance. Again, these findings might suggest systematic error or could suggest inequities in access to healthcare. The author will further explore these potential explanations in her discussion section.

Discussion

The visualizations and data analysis detailed above draw attention to the lives and well-being of the North Korean population. All four of the tables analyzed herein measure health indicators, and therefore the author's findings begin to shape a picture of health in North Korea. In nearly all cases of physical impairment, gender and "urbanness" of a respondent's residence have a direct relationship with the respondent's health. Maternal mortality ratios, although related to only one gender, provide a wealth of information about a nation's healthcare.

Before discussing how best to employ or interpret these statistically significant trends, it is necessary to understand the context in which the data was collected. A brief discussion of both healthcare and disabilities in North Korea follow.

Healthcare in North Korea

Former World Food Programme Executive Director Catherine Bertini does not hesitate to share her opinion: "North Korea is one of the worst places to get sick in."[27] Bertini and other leaders of humanitarian aid programs personally recall visiting hospitals and discovering empty shelves. Interviews with North Korean defectors relay horror stories of receiving operations without anesthesia because the hospital did not have any. In contrast to the well-stocked hospitals of Pyongyang, through which tour guides lead foreign delegations, many North Korean hospitals lack basic resources.

---

[27] Bertini, Catherine. "Collecting Data in North Korea."

As a socialist state, North Korea provides universal healthcare to its citizens. In theory, therefore, all citizens should receive whatever healthcare necessary. All medical procedures and medicines are free, but experts emphasize that this is in theory only.[28] In reality, a modified market system has developed – a "survival strategy"[29]; those that can pay for medicines buy them independently on the black market and bring them to the hospital for administering. The system privileges the elite and party members, who receive preferential services.[30] Many healthcare facilities lack running water, heat, and electricity. At the same time, according to some sources, medical professionals are well-trained and knowledgeable about modern medical technologies.[31]

In theory, authoritarian governments effectively and efficiently deliver public goods, including education and healthcare. In fact, especially early in the life of the northern Korea, the international community admired the success with which the state educated and treated its people. Under Kim il Sung's rule, citizens were required to have regular check-ups, contributing to a strong system of preventative care. A Western doctor that often travels to the DPRK to contribute his professional services praised the state's efficiency in healthcare:

> "For a health care professional, a police state is a paradise. I came with my medical van to a North Korean village, the local official blew a whistle, and in 10 minutes everyone in the village was waiting in front of our van. Every single person! No excuse was tolerated, and nobody dared to evade us. In other developing countries it was so different!"[32]

---

[28] Kim, Jae-young. "The Good Parts to Life in North Korea."
[29] Zellweger, *People with Disabilities in a Changing North Korea*.
[30] Lankov, 2013.
[31] Carriere, Fred. "Urban-Rural and Male-Female Statistical Findings."
[32] Lankov, 2013.

Theoretically, the North Korean healthcare system can effectively distribute resources across the country. Although its healthcare professionals are well-trained, the state lacks the necessary resources to provide adequate healthcare. The census data and the trends determined in earlier chapters may reflect the challenges of this resource-strapped system.

Disabilities in North Korea

According to the 2008 census data, out of a national population of 24 million people, nearly 1.8 million people, or 7.34% of the national population, report having some sort of disability.[33] Although not explicitly stated in the census document, the North Korean government classifies difficulty seeing, walking, hearing, and using one's mental faculty as disabilities. Scholar and former aid worker Katharina Zellweger believes that the North Korean proportion of population with some sort of disability is no different from the proportion in other countries. Zellwegger estimates 8.16% of the population, or 1.96 million North Koreans, are disabled. Comparing this to the global average of 10%-15%, Zellwegger does not believe North Korea's numbers are disproportionate. Zellwegger also argues that while there is a sense of shame surrounding disabilities, this is not unique to North Korea. Although North Korea has received international criticism for housing its disabled outside of Pyongyang, and old laws prohibited the disabled from residing in the nation's capital, Zellwegger argues

[33] In this context, "disability" refers to visual impairment, hearing impairment, difficulty walking, and/or difficulty using mental faculty (tables 42-45). This figure, however, does *not* account for any respondents who have more than one disability. The actual number may therefore be slightly lower.

that in all countries, institutions for the disabled have been pushed to the outskirts of cities.[34]

Although the country certainly cannot claim a perfect track record regarding rights of the disabled, North Koreans with disabilities do not experience an abnormally strong stigma.

Gender Inequities in Health

The visualizations and data analysis suggest that there is a sustained pattern across different physical impairments: women are more likely than men to report difficulty seeing, hearing, or walking. Before exploring the potential cultural or political explanations for this, it is necessary to first recognize that systematic error may be affecting the results. All census data is based on respondents' answers. The data, therefore, is only as accurate as the respondents' answers are honest and informed. The data from tables 42, 43, and 44 were each collected through one question about physical impairment. In each case, this question did not elaborate on the definitions of "difficulty seeing," "difficulty hearing," or "difficulty walking or climbing stairs." There is no available data to ensure that respondents understood the questions asked of them. However, the reported accuracy of the data and the overwhelming approval of the final product by the United Nations, in addition to continual monitoring and supervision throughout the enumeration process, suggest that this issue of understanding the term "difficulty" likely did not compromise the collected data.

[34] Zellweger, *People with Disabilities in a Changing North Korea*.

Supervisors have already spoken to the high quality of the census data, but their preventative steps to ensure accurate data cannot control for respondents' honesty. Systematic error may also skew data if respondents choose to not answer enumerators' questions accurately. This uncertainty is one possible explanation for the gender differences recognized in various physical impairments. It is possible that men, feeling the need to appear strong and fill the proper role of the male gender, felt less comfortable reporting that they had physical impairments, thereby decreasing the gender's proportion of physical impairment. Systematic error like this could stem from social and cultural norms in North Korea.

If the information accurately reflects the true conditions of the North Korean population, social and cultural norms may still play a large role. If females truly have higher proportions of impairment, the difference could stem from the healthcare men and women receive, suggesting that women do not receive the same amount of care as men. Professor Fred Carriere finds this to be a compelling argument, citing ancient thought patterns that persist today. Carriere recalls the story of the "fervent woman," a tradition rooted in historical Confucian preference for men. The "fervent woman" was the honorable image of a self-sacrificing family woman, a Korean woman willing to regularly sacrifice her well-being for the health of her husband, father-in-law, and son. The story originates from one woman's literal self-sacrifice who, during a time of poverty, cooked portions of her buttocks to nourish her ill father-in-law because the family could not afford meat. If the "fervent woman" was deemed honorable enough, the government will approve the construction of a pagoda in her honor. Carriere

describes this as a mostly rural tradition that demonstrated a woman's "true dedication to the system of the ideology."[35] In essence, the woman only existed *through* the males.

Carriere does not see any reason why this same thought pattern cannot manifest itself in healthcare, especially given the scarcity of medical resources. One possible explanation for higher proportions of impaired females is that women are sacrificing their healthcare for the health of their husbands, fathers-in-law, or sons. Limited resources may lead to prioritizing one person's health over another's. If female self-sacrifice remains a strong value, this norm could be producing the disproportionate number of physically impaired women. Indeed, through personal interactions with North Korean officials, Carriere has recognized that this thought pattern continues. In some places and within some families, however, this tradition does not continue; some women have not been encultured with this ancient norm of sacrifice. Carriere hypothesizes that in more urban areas, where ideas are exchanged more often and old ways are discarded sooner, this thought process may not dictate the woman's role.

If Carriere's hypothesis is true, the data should show larger differences between proportions of impaired men and impaired women in rural areas, and closer proportions of impairment across gender in urban areas. To test this hypothesis, the author conducted a Mantel-Haenszel chi-square test. This tests the null hypothesis that two variables – in this case, gender and "urbanness" – are conditionally independent. Conducting three different Mantel-Haenszel

[35] Carriere, Fred. "Urban-Rural and Male-Female Statistical Findings."

chi-square tests across the three types of impairment, all tests returned a p-score of less than 2.2e-16 and none of the 95% confidence intervals included 0. With these results, the null hypothesis can be rejected; gender and "urbanness" are *not* independent from each other. Professor Carriere's hypothesis is supported.

The above presents only one possible explanation for the disproportionate number of physically impaired females, if in fact the data is not skewed by systematic error. Further research on the cultural impact of gender roles on healthcare is necessary to make a stronger argument.

Urban and Rural Inequities in Health

In a market system, superior health professionals flock to profitable cities with large populations and favorable market conditions for doctors. That said, the differences between the proportion of urban North Koreans with disabilities and the proportion of rural North Koreans with disabilities may appear reasonable – excellent healthcare is more readily available in urban centers. However, in a socialist system in which health professionals and medical centers are distributed evenly based on population, the availability of healthcare should not determine levels of disability and other health problems. If the theoretical socialist healthcare system were in effect and functioning properly in North Korea, disparate proportions of disabled people between urban and rural areas would be less likely.

However, the 2008 census data shows higher rates of physical impairment in rural areas. One explanation could stem from the fact that the underground

healthcare system is market-driven. The elites and those favored by the party, most of whom are living in the city, can afford to pay for health services that prevent or treat such impairments while North Koreans in poorer, more rural areas cannot afford black market goods.

Another explanation involves the central party, which distributes resources to the elite, in part as a way to reward behavior. This explanation also accounts for the lower maternal mortality rate in Pyongyang. The capital serves as the home to top party officials that have the power to redistribute medical resources. Pyongyang is also the playground for foreigners who travel to North Korea on leisure, or diplomats who visit Pyongyang on business; Pyongyang is the city that projects North Korea's image to the rest of the world. Like the criticism applied to North Korea's treatment of the disabled and their social stigma, North Korea is not necessarily unique from other developing countries in that its foreign aid begins at the capital and then moves outward.[36]

Maternal Health

Aside from Pyongyang, the hypothesis that more urban areas have lower maternal mortality ratios was not clearly supported by the data. Dividing provinces into smaller administrative units and comparing the maternal mortality ratios of each unit with its "urbanness" could show a clearer trend. More localized data could also provide a more visible trend between the "urbanness" of residential area and the place of death, supporting or rejecting the hypothesis that

[36] Spezza, "Child Nutrition & Medical Training Urgent Issues in N. Korea."

maternal deaths are more likely to occur at home in rural areas merely because hospitals are less accessible.

Two important variables likely to affect maternal mortality ratios were not included in the census data: midwives and nutrition. Between 2000 and 2006, 97% of North Korean deliveries were assisted by "skilled birth attendants".[37] North Korea has an expansive midwifery system that has been credited for the country's relatively low maternal mortality ratio. Nutrition is a considerable variable affecting maternal death rates. A 2006 World Health Organization report determined the leading causes of maternal death in Asia to be hemorrhage – a common cause of maternal deaths among developing countries – followed by anemia, and sepsis.[38] According to a 2009 UNICEF study, 31% of North Korean women were anemic.[39] As a chronically food insecure nation, North Korean citizens' health suffers, and physical impairment and maternal mortality ratios consequently increase.

---

[37] Islam, "Progress Toward Achieving Millennium Development Goal 5 in South-East Asia."
[38] "WHO Analysis of Causes of Maternal Death: A Systematic Review." *The Lancet* (2006).
[39] "DPR Korea 2013: Humanitarian Needs and Priorities." *United Nations* (2013).

Conclusion and Further Study

So what? What does this analysis contribute to academia or to the broader community? Other than the simple argument that increased information about the "Hermit Kingdom" is valuable to our nation's understanding, why explore these variables and trends?

Contrary to popular belief, *hundreds* of non-profit organizations, non-governmental organizations, businesses, and bilateral or multilateral projects have engaged North Korea. Despite the many roadblocks to working in and with North Korea, foreign engagement is possible and, indeed, happening regularly. This includes humanitarian aid, training and education, and development assistance, all of which can be directly improved by the data analysis herein.

The 2008 census is a massive, high-quality dataset that provides insight beyond the numbers and into the intricacies of the North Korean population. Further analysis and visualizations similar to what the author did has the power to inform and improve all projects in North Korea by determining information such as:

- Who are the most vulnerable people (to disease, hunger, etc.)?
- Where are the most vulnerable people? How can they best be reached?
- What resources are most needed?
- What projects would reach the most people?

This capstone project is a starting point to using data more effectively to understand what populations exist in North Korea and what these populations most need. Aid and development organizations can take the information herein

and apply it to their work, ensuring, for example, that they concentrate maternal health projects in provinces with higher maternal mortality ratios. From this analysis, disability advocacy groups may decide to focus their efforts in rural areas of North Korea rather than urban areas, or target females specifically. With the machine-readable data now freely accessible, and example coding used in an open-source statistics program, anyone with an Internet connection and a functioning computer can conduct useful and insightful analysis.

Finally, and most importantly, researchers working with this data or any other North Korean information should seek to collaborate with other interested organizations or individuals. Jiehae Blackman, the founder of EngageDPRK, is a great example of someone with whom a researcher in this field could collaborate. Her website maps all foreign aid projects in North Korea that are publicly detailed online. EngageDPRK employs the power of visualization. Overlaying maternal mortality ratios over her maps of maternal health projects, for example, could show humanitarian organizations where (and whether) their projects correspond to higher maternal mortality ratios.

References

Blackman, Jiehae. *EngageDPRK*. Reah International. U.S. Institute of Peace. Web. 13 Apr. 2014.

Bertini, Catherine. "Catherine Bertini." *Catherine Ann Bertini: Home Page*. Maxwell School of Citizenship and Public Affairs, 2012. Web. 13 Apr. 2014.

Bertini, Catherine. "Collecting Data in North Korea." Personal interview. 11 Apr. 2014.

Carriere, Fred. "Urban-Rural and Male-Female Statistical Findings." Personal interview. 16 Apr. 2014.

Cha, Victor. *The Impossible State*. New York: Harper-Collins, 2012. Print.

Chang, Winston. *R Graphics Cookbook*. Sebastopol: O'Reilly Media, 2013. PDF.

Cohen, Roberta. "World Food Day: The Challenge of North Korea." *The Brookings Institution*. The Brookings Institution, Oct. 2013. Web. 24 Mar. 2014.

"Country Comparison: Maternal Mortality Rate." *Central Intelligence Agency*. Central Intelligence Agency, n.d. Web. 15 Feb. 2014.

The Crumbling State of Health Care in North Korea. Publication. Amnesty International, 2010. Web.

"Data: Maternal Mortality Ratio." *The World Bank*. The World Bank Group, 2010. Web. 10 Apr. 2014.

"The Democratic People's Republic of Korea 2008 Census." *United Nations Statistics Division*. United Nations, n.d. Web. 10 Apr. 2014.

DPR Korea. Central Bureau of Statistics. *DPR Korea 2008 Population Census: National Report*. Pyongyang: 2009. Web. 17 March 2013.

Engracia, Luisa. “UNFPA’s role in the DPRK’s 2008 Population Census.” National Committee on North Korea, United Nations Foundation, Washington, D.C. 16 July 2010. Presentation.

Gharzeddine, Omar. "UNFPA Helps Plan and Monitor Successful DPRK Census." *UNFPA*. UNFPA, 15 Feb. 2010. Web. 17 Jan. 2014.

"Improve Maternal Health by 2015." *The World Bank*. The World Bank Group, 2014. Web. 10 Mar. 2014.

Islam, M. “Progress Toward Achieving Millennium Development Goal 5 in South-East Asia.” *BJOG* (2011): 6-11.

Jones, Jeffrey M. "In U.S., 83% Say North Korean Nukes Are a Critical Threat." *Gallup Politics*. Gallup, Inc., 18 Feb. 2013. Web. 20 Apr. 2014.

Khan, Khalid S. et al. "WHO Analysis of Causes of Maternal Death: A Systematic Review." *The Lancet* (2006): 1066-074. Web. 10 Mar. 2014.

Kim, Jae-young. “The Good Parts to Life in North Korea.” *NKNews.org*. NK News, 2 Oct. 2012. Web. 10 Mar. 2014.

Lankov, Andrei. *The Real North Korea: Life and Politics in the Failed Stalinist Utopia.* New York: Oxford University Press, 2013. Print.

Mason, Margie. "Frigid Hospitals Cloud North Korea's Picture of Health." *The Irrawaddy*. Irrawaddy Publishing Group, 25 June 2013. Web. 15 Apr. 2014.

Morrow, Martha. "DPRK Maternal Health Research." Message to the author. 13 Apr. 2014. E-mail.
Olsen, Harold. “Student Thesis on 2008 DPRK Census, Healthcare.” Message to author. 16 Apr. 2014. E-mail.
Pinkston, Daniel A. and Phillip C. Saunders. "Seeing North Korea Clearly." *Survival* 45.3 (2003): 79-102. Web.
DPR Korea. Central Bureau of Statistics. *Population Census Atlas of the DPR Korea*. Pyongyang: 2009. Print.
R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Spezza, Gianluca. "Child Nutrition & Medical Training Urgent Issues in N. Korea." *NKNews.org*. NK News, 28 Jan. 2014. Web. 10 Apr. 2014.
Spezza, Gianluca. "Does North Korea Have Welfare?" *NKNews.org*. NK News, 27 Jan. 2014. Web. 15 Apr. 2014.
Spoorenberg, Thomas. “Student Thesis on 2008 DPRK Census.” Message to the author. 11 Apr. 2014. E-mail.
Stephen, Elizabeth H. *Demography of a Reunified Korea*. Rep. Washington, D.C.: Center for Strategic and International Studies, 2013. Web.
Team America: World Police. Dir. Trey Parker. Perf. Trey Parker, Matt Stone, Elle Russ. Paramount Pictures, 2004.
United Nations. *DPR Korea 2013: Humanitarian Needs and Priorities*. 15 Aug. 2013.
United States. Library of Congress. Federal Research Division. *Country Profile: North Korea*. Washington, D.C.: Library of Congress, 2007. Print.
Weingartner, Erich. "My Bumpy Road to Pyongyang." *38 North*. US-Korea Institute at SAIS, 22 Oct. 2010. Web. 13 Apr. 2014.
Weingartner, Erich. "My First Monitoring Trip." *38 North*. US-Korea Institute at SAIS, 26 June 2011. Web. 13 Apr. 2014.
Weingartner, Erich. "My Introduction to Nampo Port." *38 North*. US-Korea Institute at SAIS, 15 Dec. 2010. Web. 13 Apr. 2014.
Wilke, Joy. "North Korea Least Favorable Among Nations." *Gallup Politics*. Gallup, Inc., 19 Feb. 2014. Web. 20 Apr. 2014.
Zellweger, Katharina. “DPRK Census Atlas and Student Thesis.” Message to the author. 15 Apr. 2014. E-mail.
Zellweger, Katharina. *People with Disabilities in a Changing North Korea*. Working paper. Stanford: Shorenstein Asia-Pacific Research Center, 2014. Web.
Zellweger, Katharina. “Aid and Development Co-operation and North Korea: A Window of Opportunities or Pushing at a Closed Door.” Korea Institute, Harvard University, Cambridge, MA. 29 March 2012. Forum.

Appendix 1: Visualizations

Maternal Mortality

a. Maternal Mortality Ratio by Province Table
b. Maternal Mortality Ratio Political Map
c. DPRK Percent Urban by Province Political Map
d. Maternal Mortality Ratio Horizontal Bar Graph
e. Maternal Mortality Ratio Dot Plot
f. Reordered Maternal Mortality Ratio Dot Plot
g. Stacked Maternal Mortality Ratio Bar Graph by Death Location
h. Maternal Mortality Ratio Bar Graphs by Death Location
i. World Bank Maternal Mortality Ratio Visualization

Visual Impairment

j. Proportions Urban/Rural and Male/Female "A Lot of Difficulty" Seeing
k. Proportions Urban/Rural and Male/Female "Cannot See at All"
l. Proportions Urban/Rural and Male/Female Any Difficulty Seeing

Hearing Impairment

m. Proportions Urban/Rural and Male/Female "A Lot of Difficulty" Hearing
n. Proportions Urban/Rural and Male/Female "Cannot Hear at All"
o. Proportions Urban/Rural and Male/Female Any Difficulty Hearing

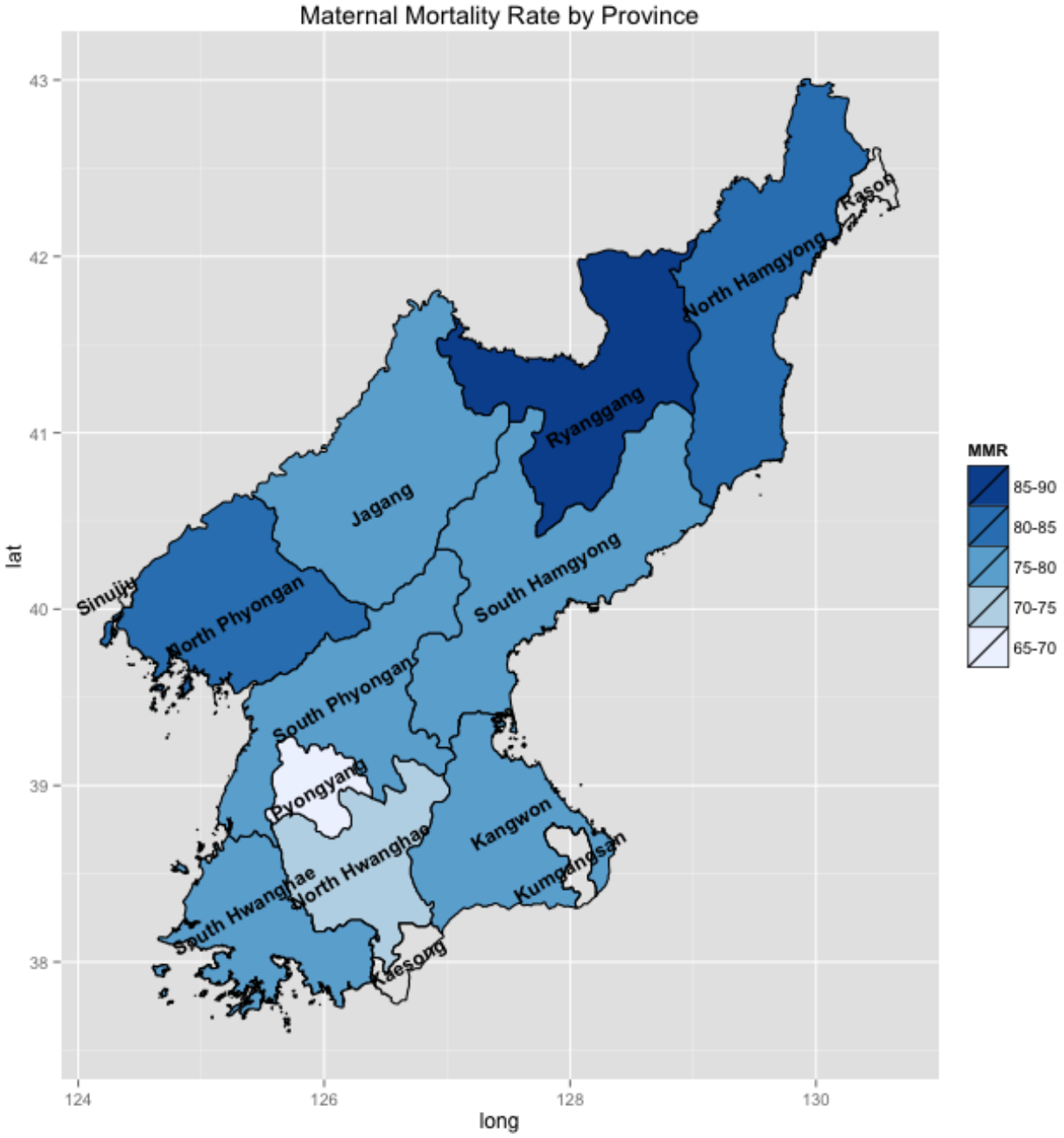
Walking and Climbing Difficulty

p. Proportions Urban/Rural and Male/Female "A Lot of Difficulty" Walking or Climbing
q. Proportions Urban/Rural and Male/Female "Cannot Walk at All"
r. Proportions Urban/Rural and Male/Female Any Difficulty Walking or Climbing

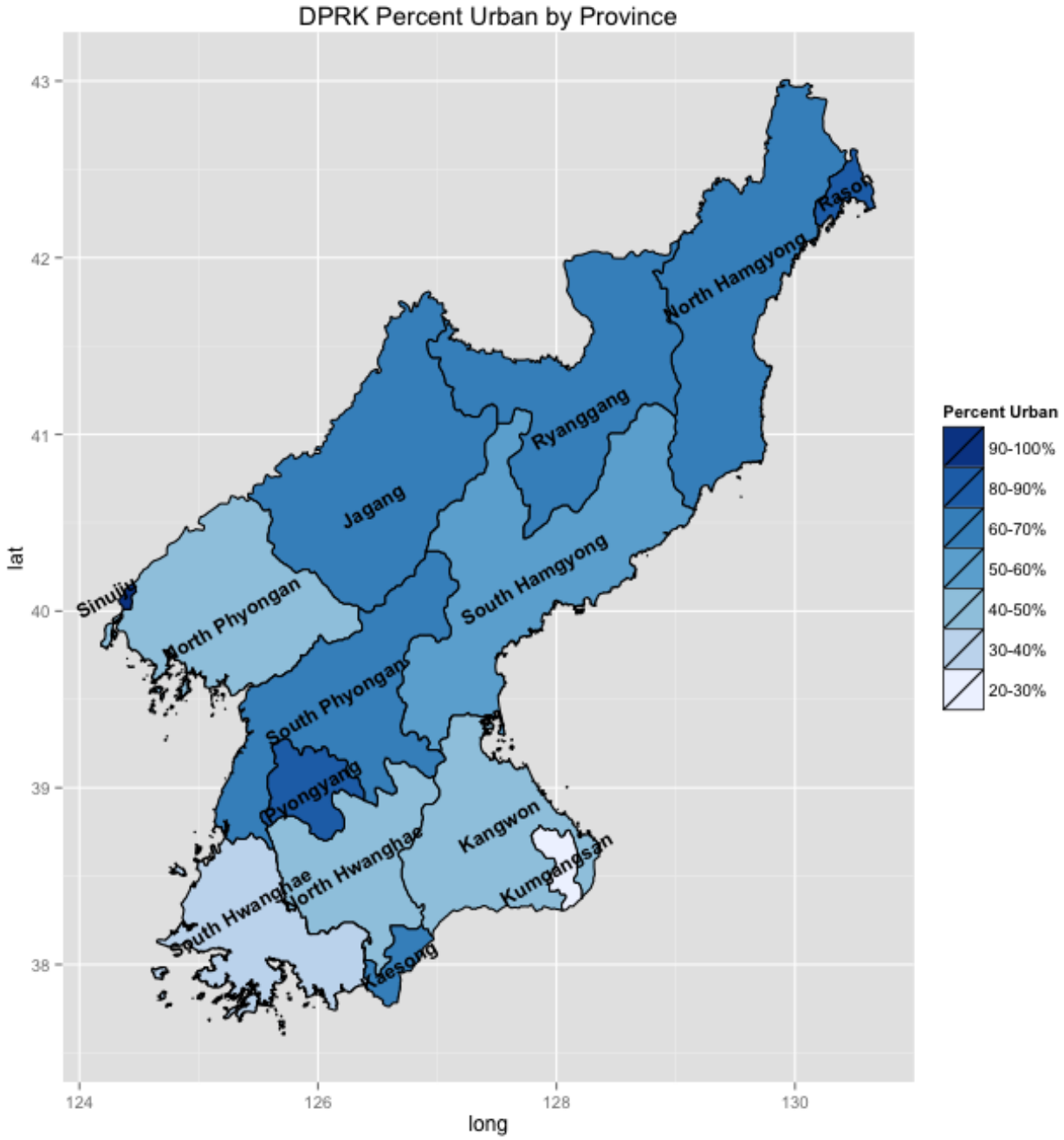
a. Maternal Mortality Ratio by Province Table

Province	MMR
Ryanggang	85.99274
North Hamgyong	83.56626
South Hamgyong	76.09026
Kangwon	79.19747
Jagang	78.76497
North Phyongan	83.24152
South Phyongan	78.12627
North Hwanghae	70.61240
South Hwanghae	76.61426
Pyongyang	68.46531

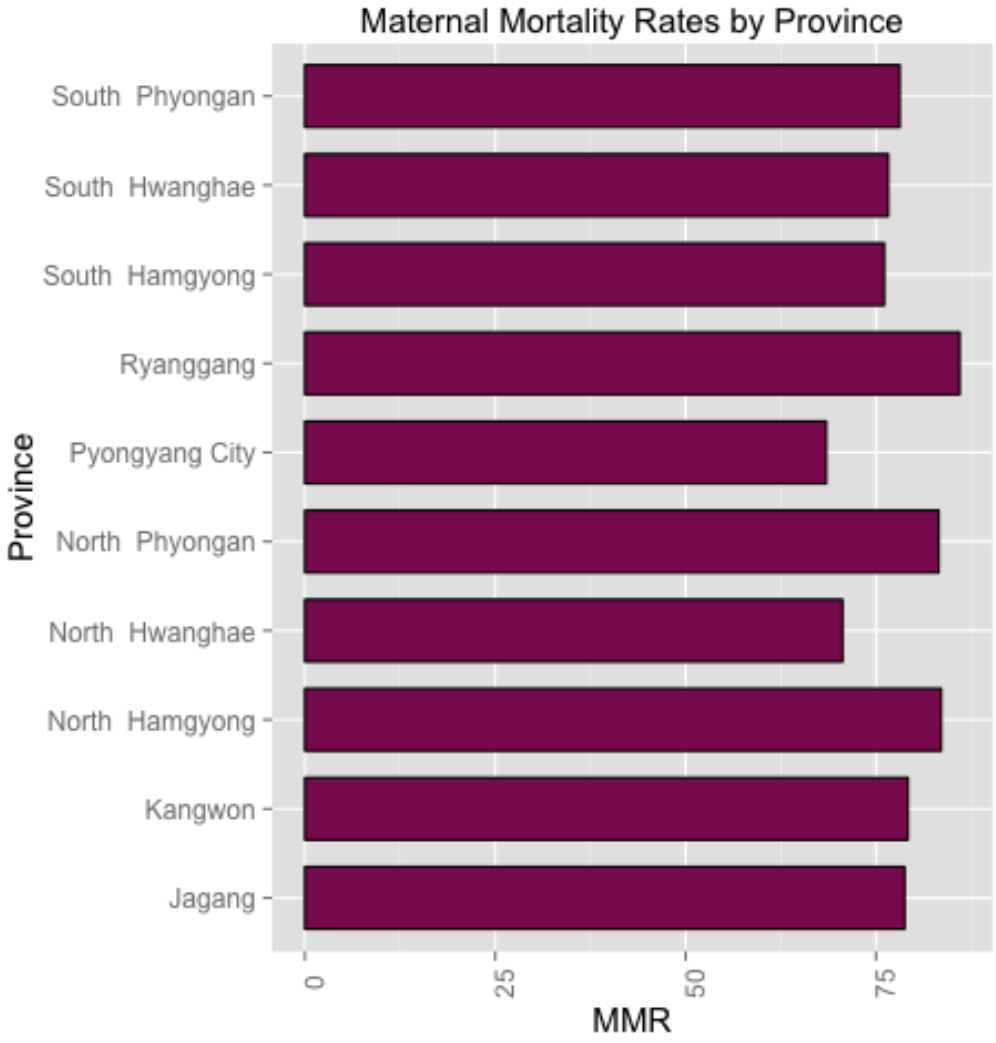
b. Maternal Mortality Ratio Political Map



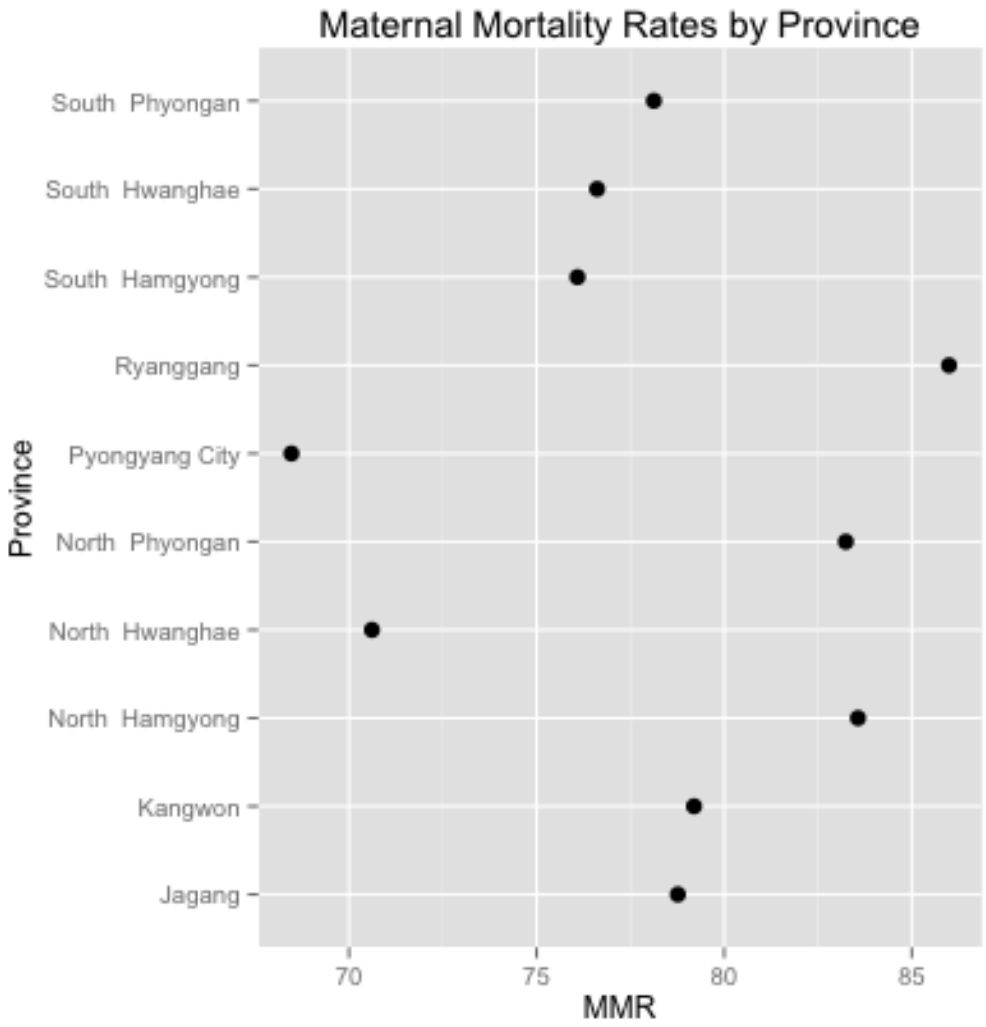
c. DPRK Percent Urban by Province Political Map



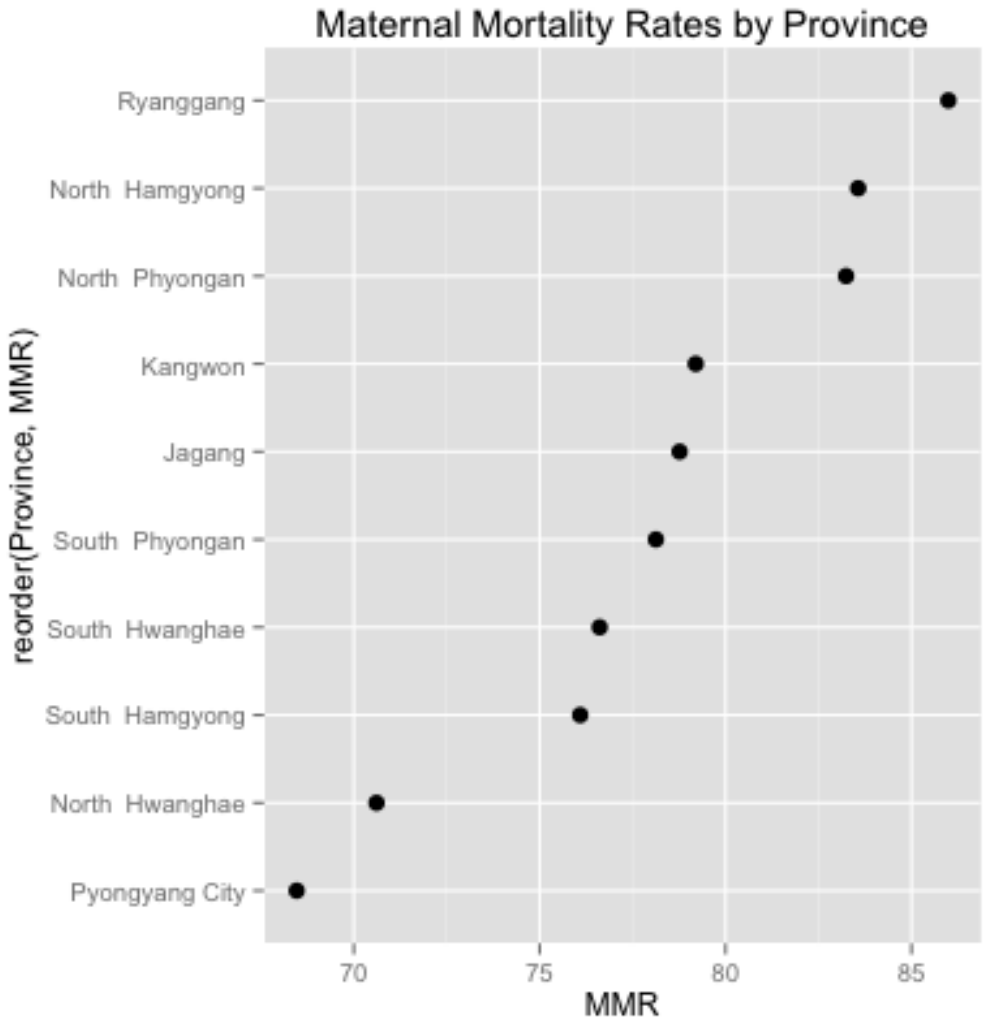
d. Maternal Mortality Ratio Horizontal Bar Graph



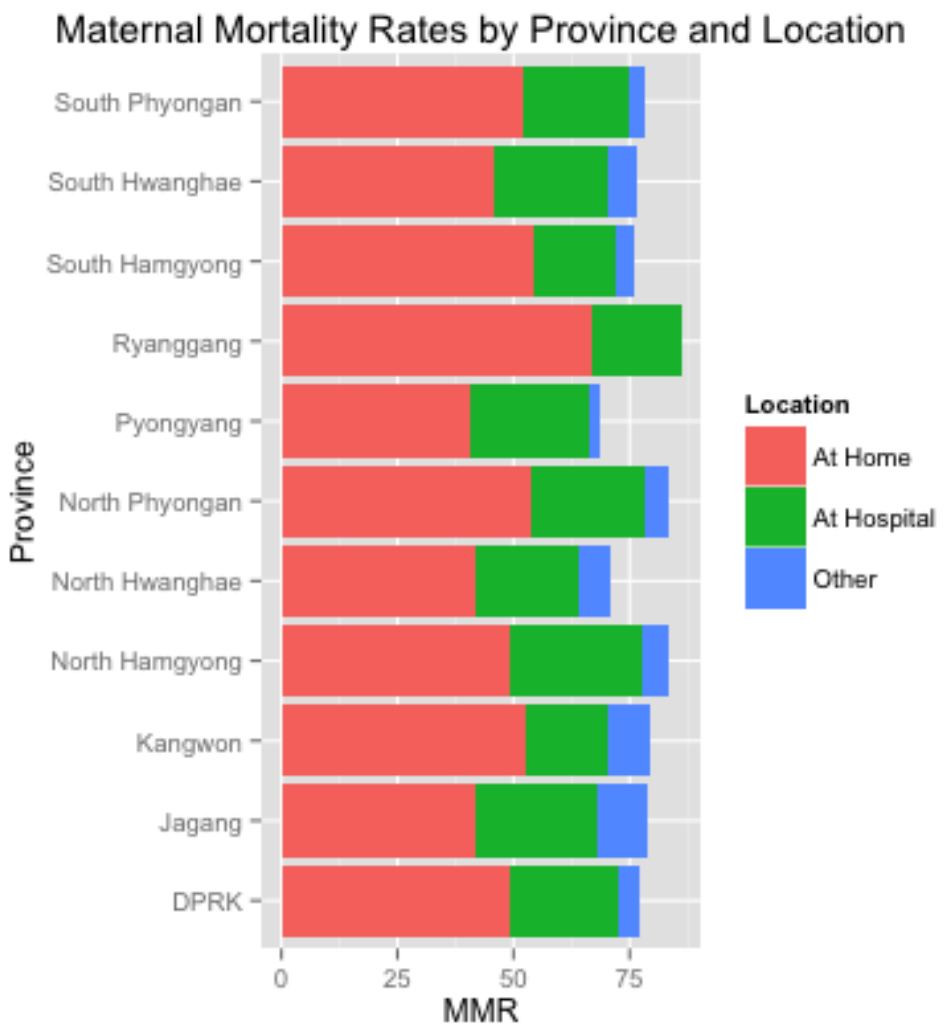
e. Maternal Mortality Rate Dot Plot



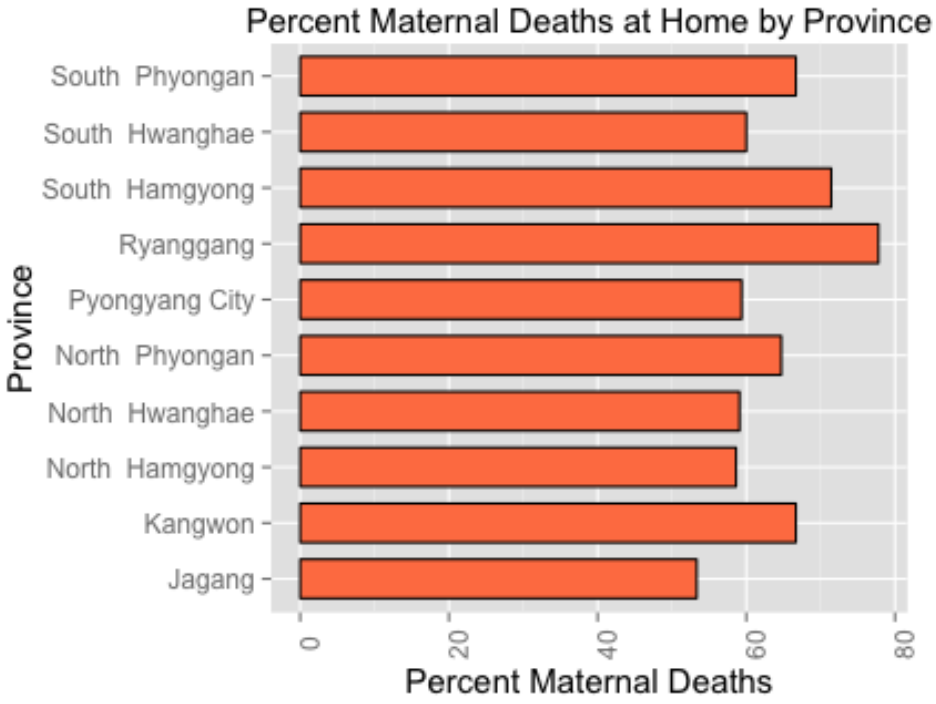
f. Reordered Maternal Mortality Rate Dot Plot

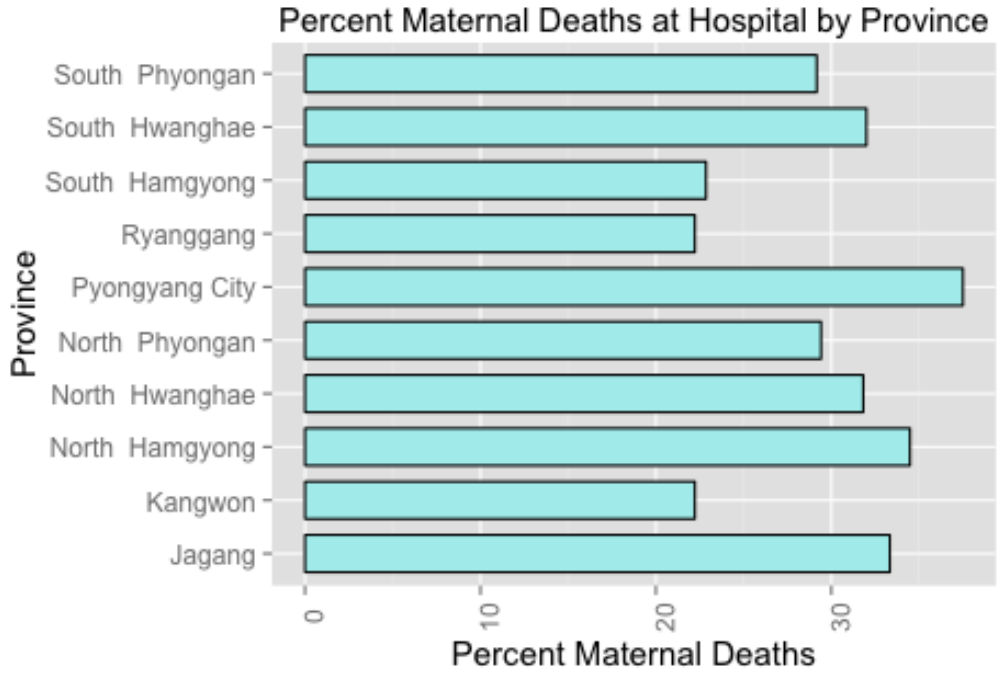


g. Stacked Maternal Mortality Horizontal Bar Graph by Death Location

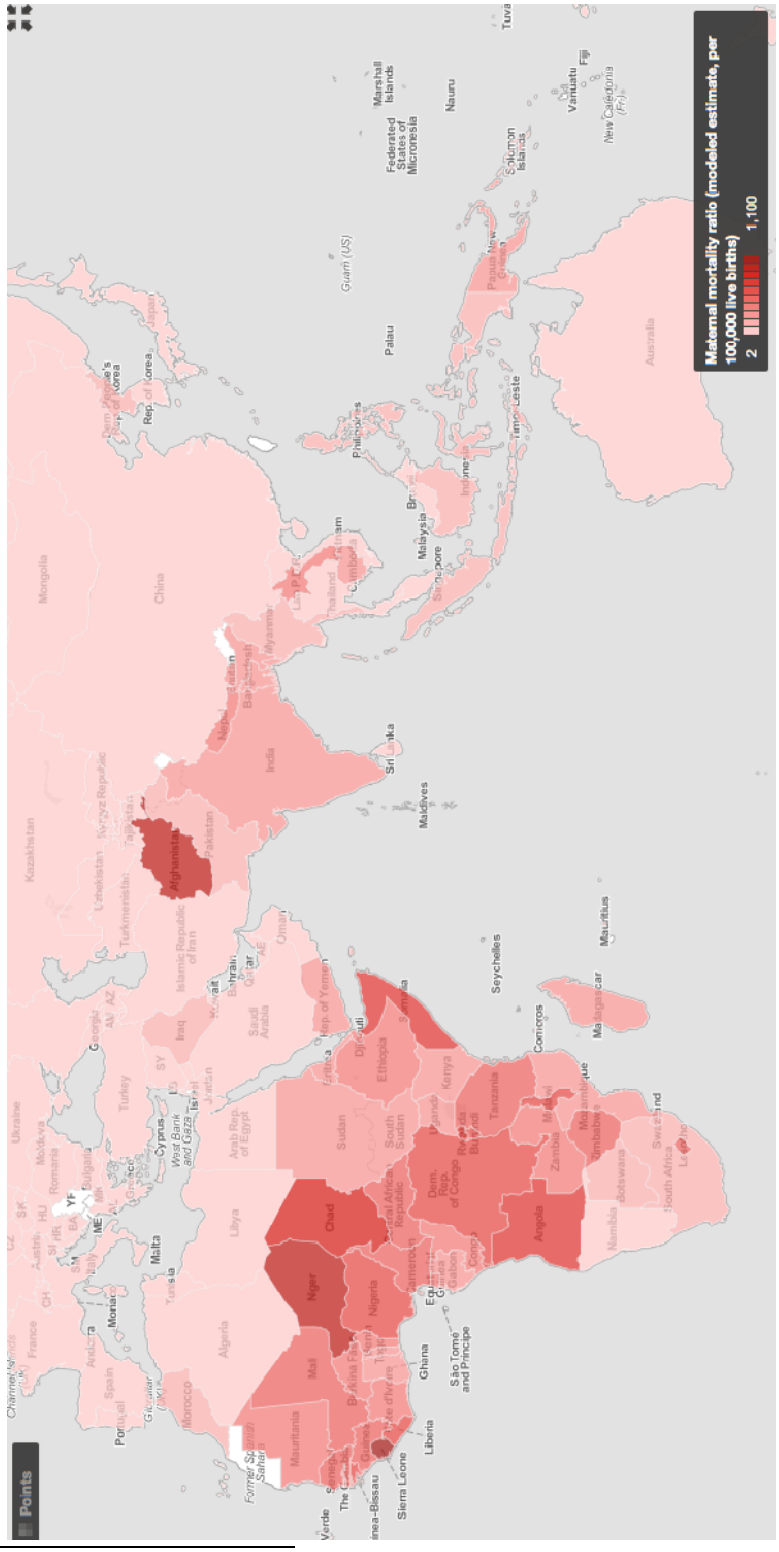


h. Maternal Mortality Ratio Bar Graphs by Death Location



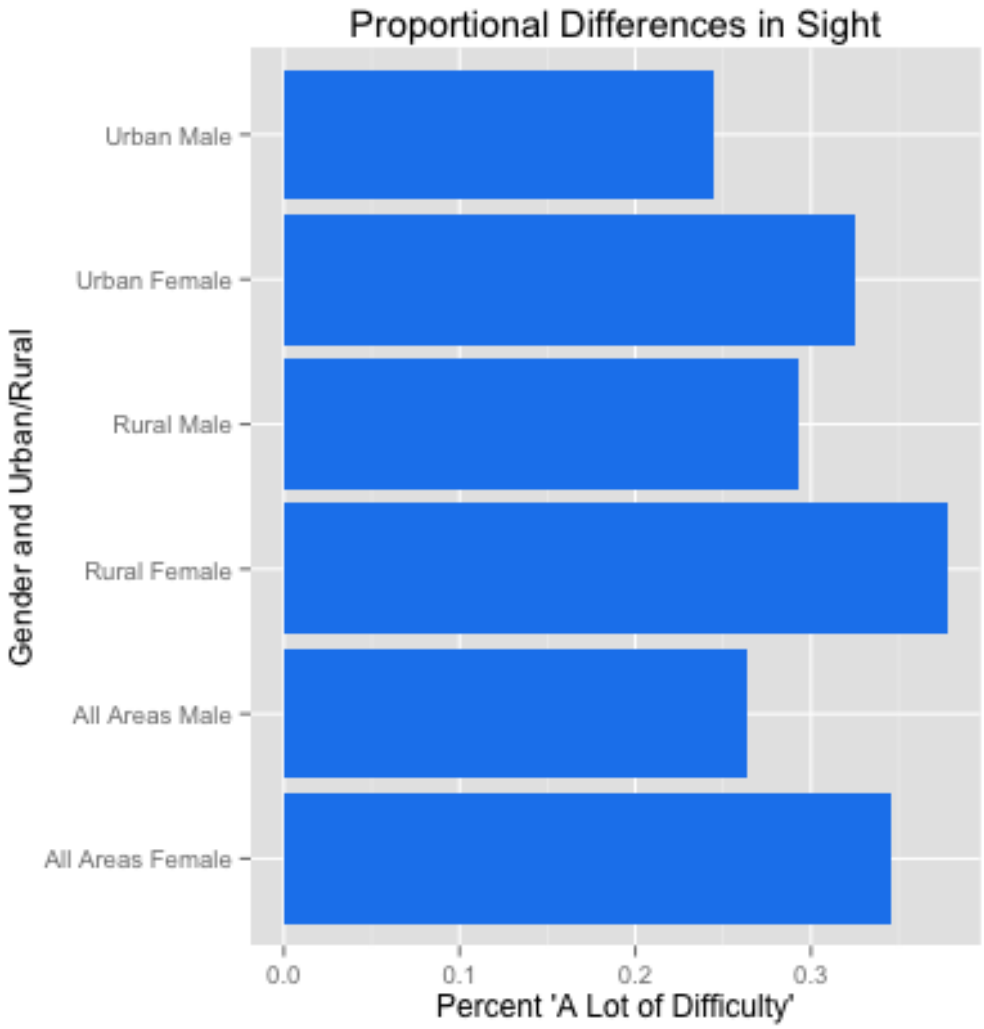


i. World Bank Maternal Mortality Visualization*

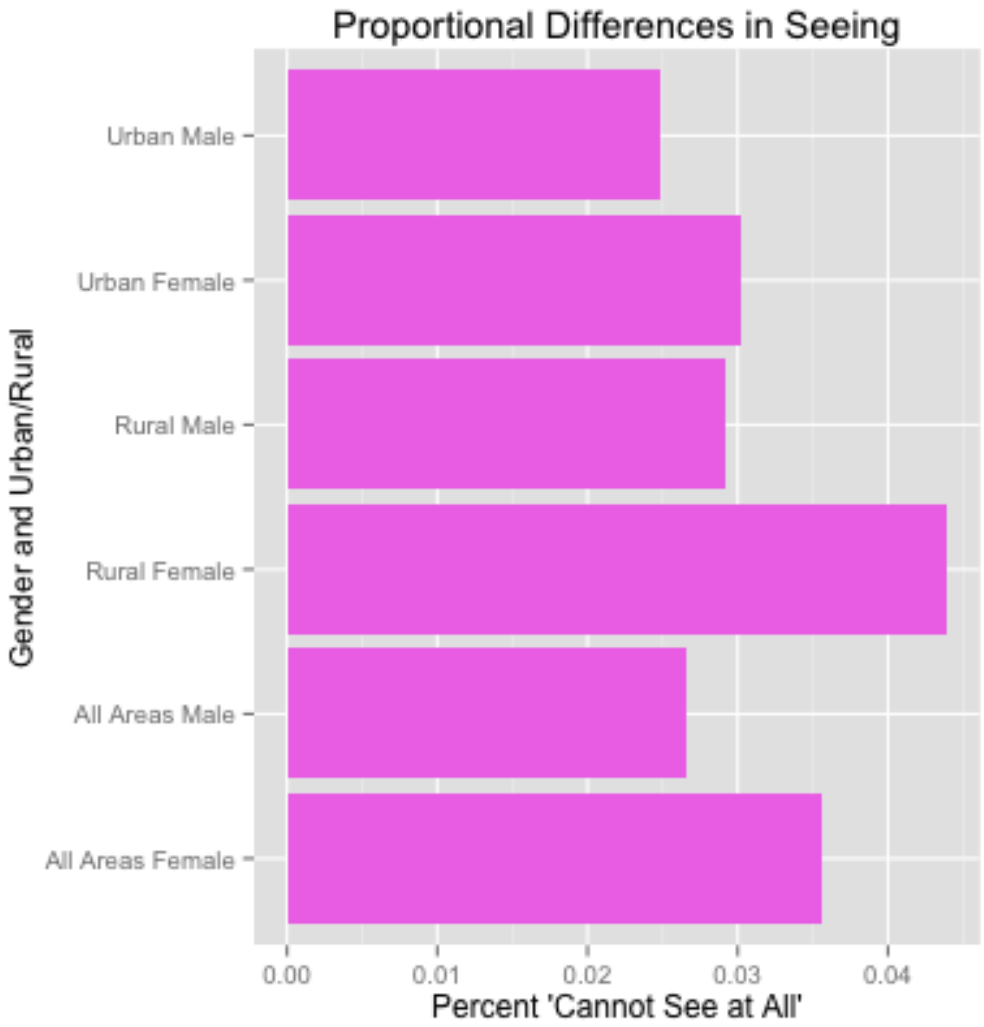


* The author purposely zoomed in on Asia and Africa to more clearly display the range in MMR between the continents and within East Asia.

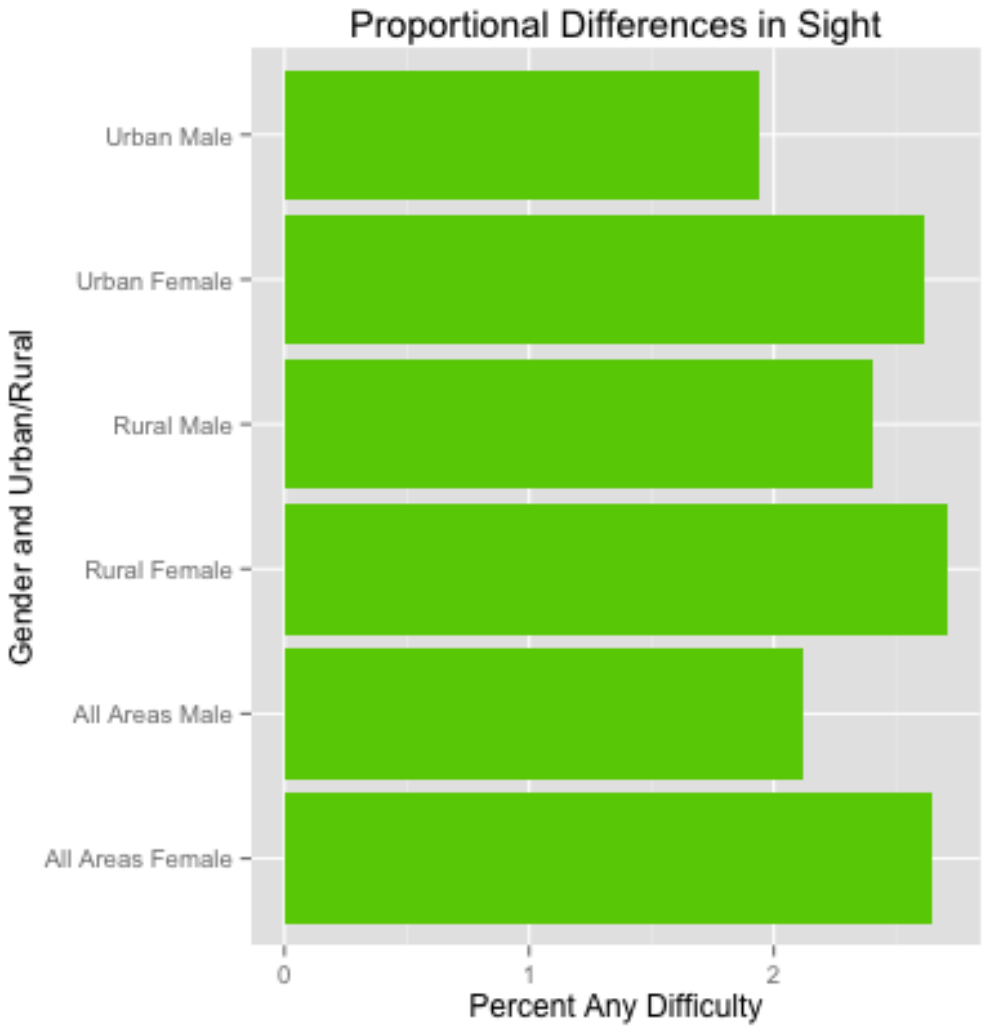
j. Proportions Urban/Rural and Male/Female "A Lot of Difficulty Seeing"



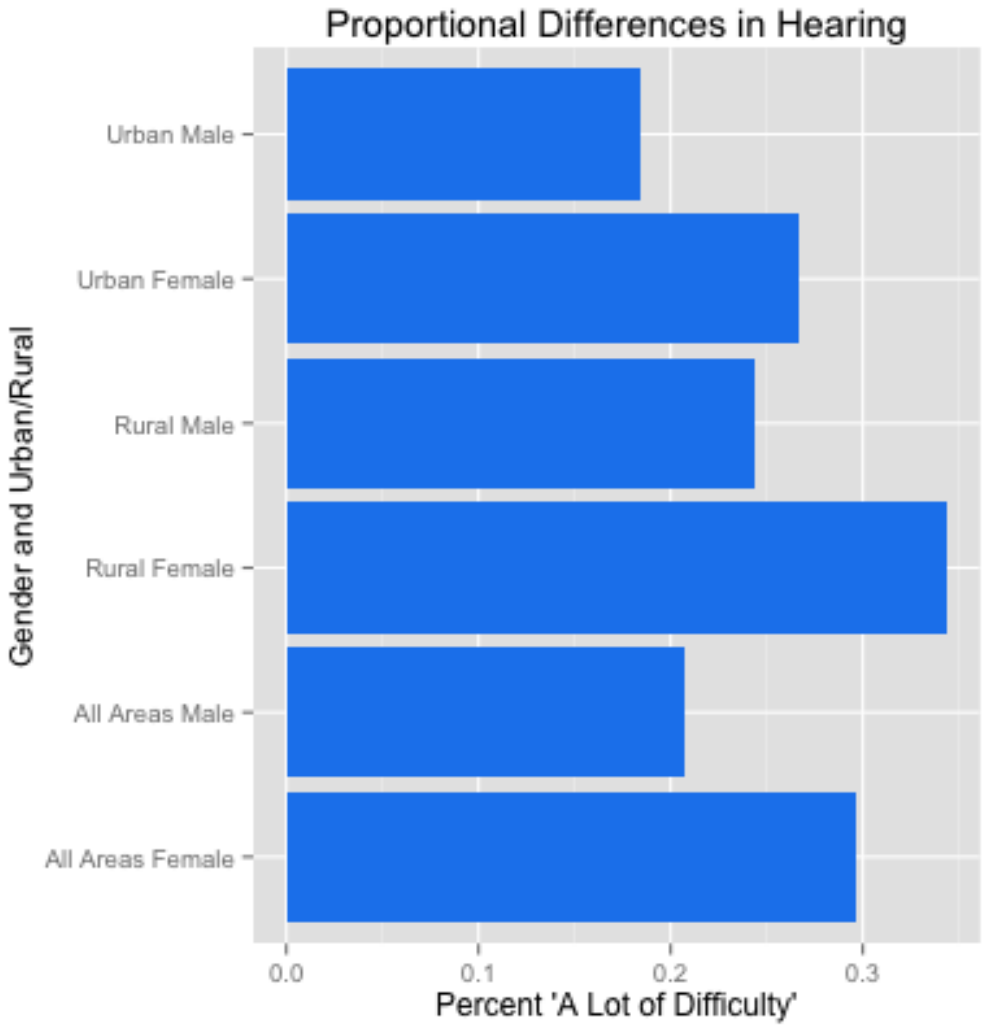
k. Proportions Urban/Rural and Male/Female "Cannot See at All"



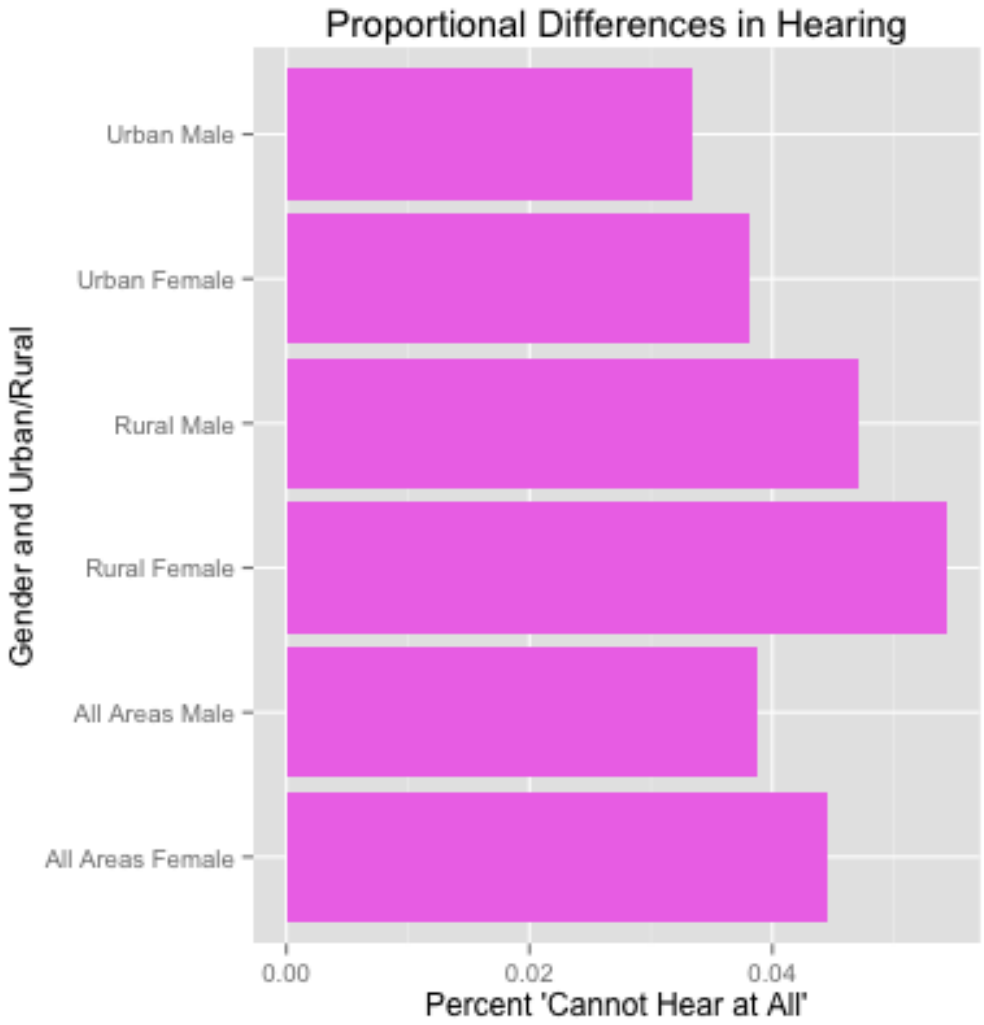
l. Proportions Urban/Rural and Male/Female Any Difficulty Seeing



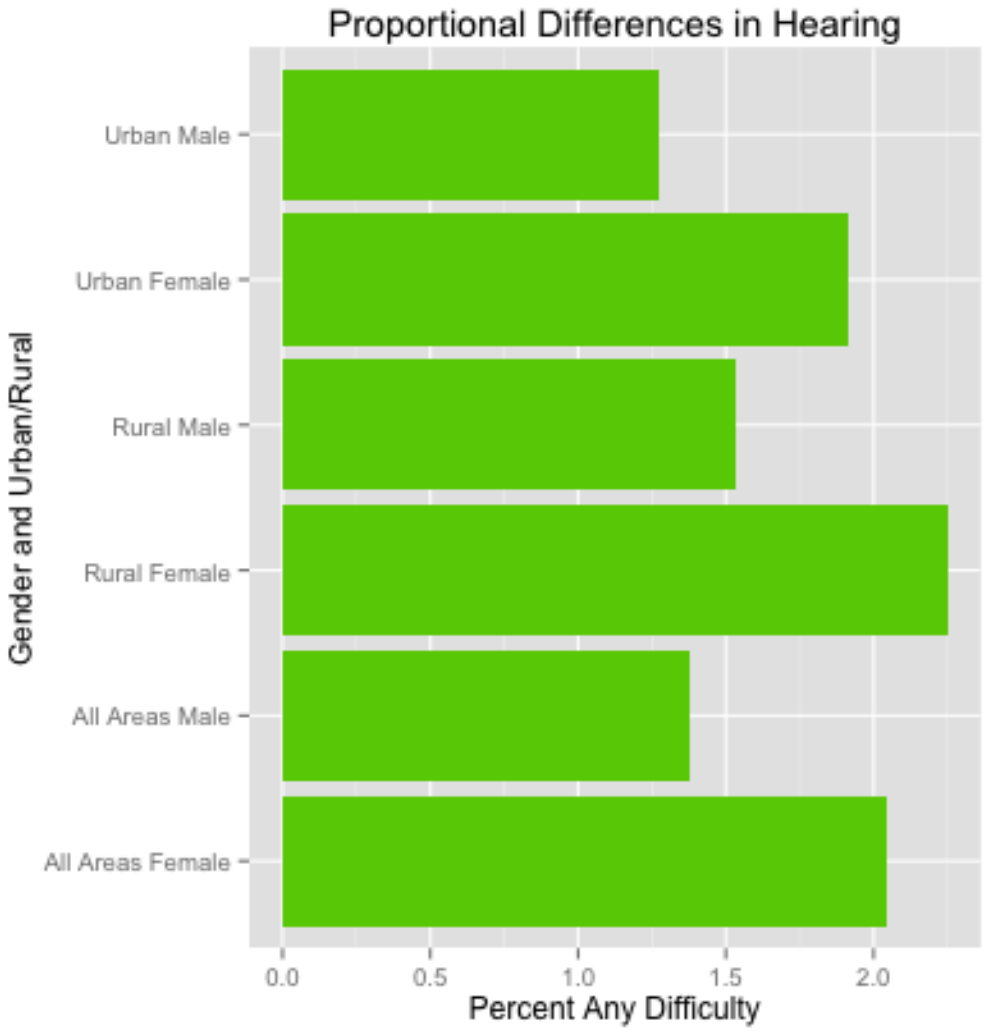
m. Proportions Urban/Rural and Male/Female "A Lot of Difficulty" Hearing



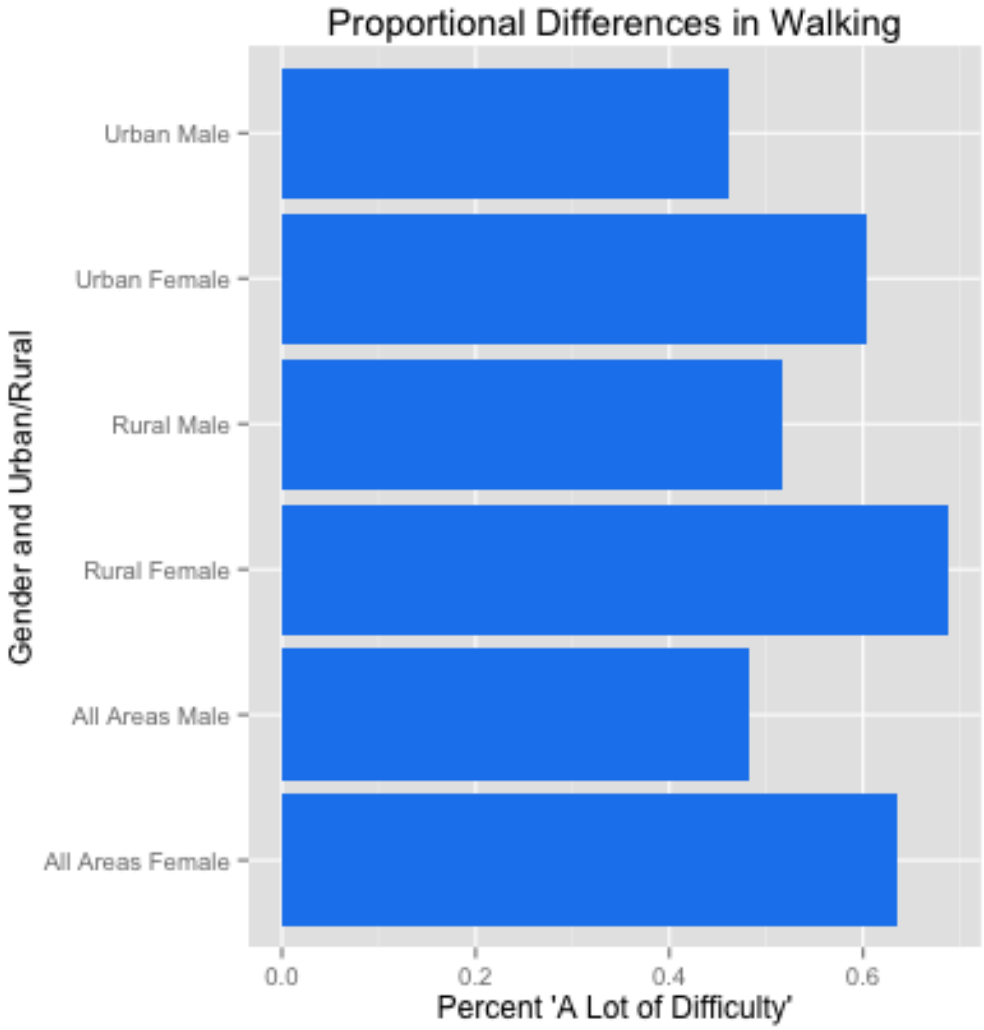
n. Proportion Urban/Rural and Male/Female "Cannot Hear at All"



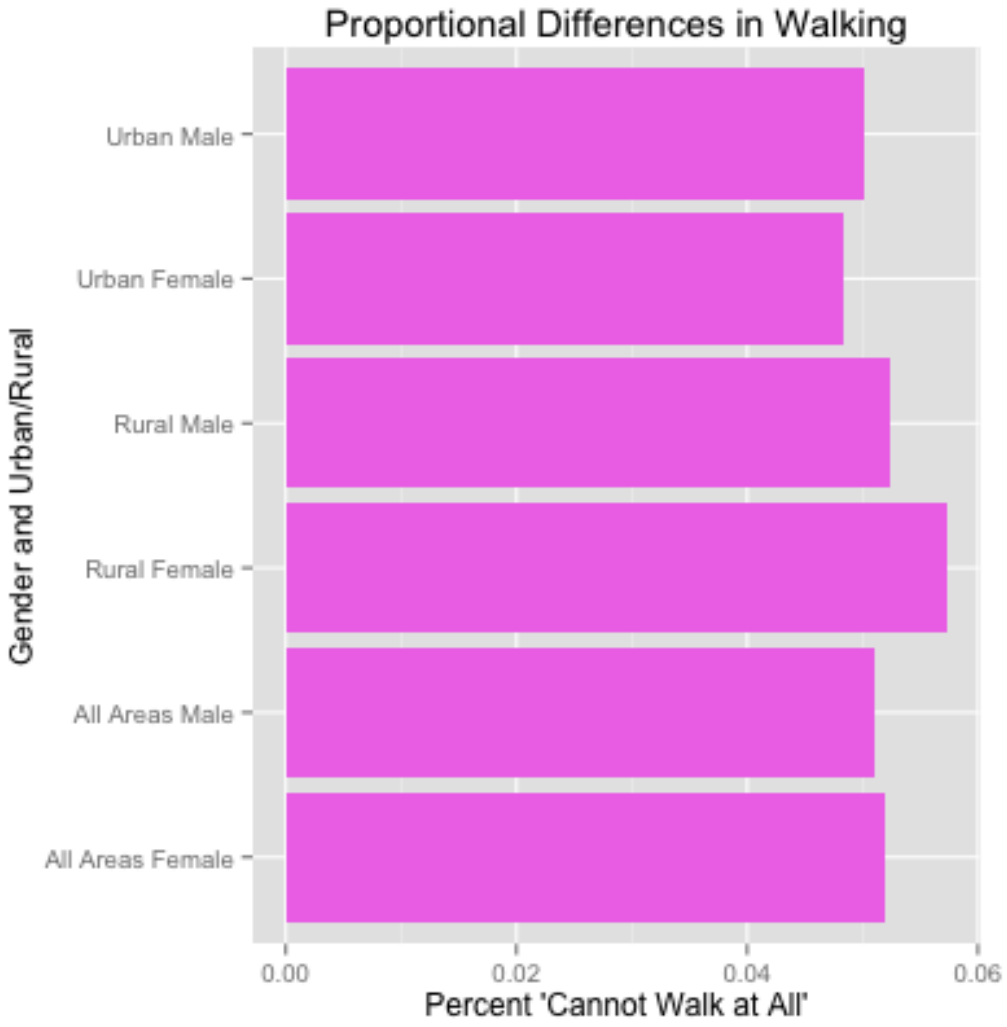
o. Proportion Urban/Rural and Male/Female Any Difficulty Hearing



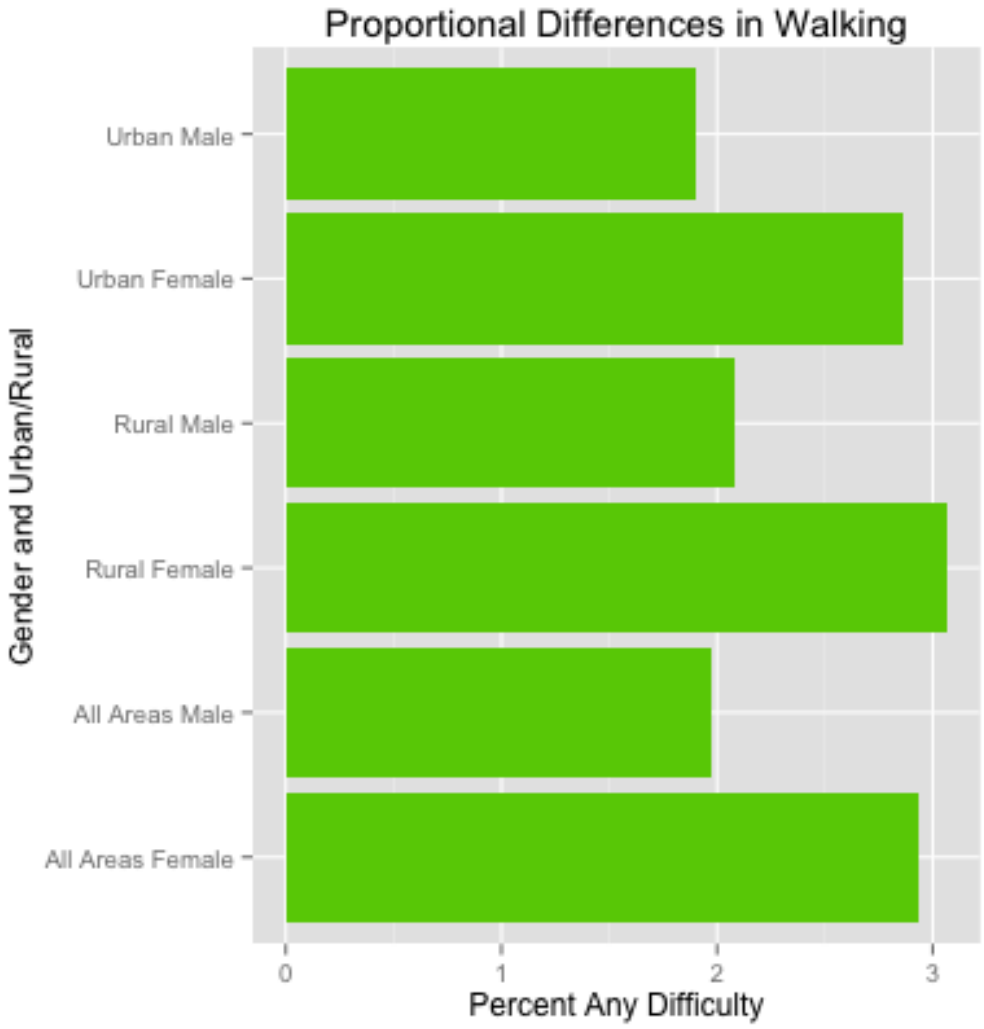
p. Proportions Urban/Rural and Male/Female "A Lot of Difficulty" Walking and Climbing



q. Proportions Urban/Rural and Male/Female "Cannot Walk at All"



r. Proportions Urban/Rural and Male/Female Any Difficulty Walking or Climbing



Appendix 2: R Coding Scripts

a. Maternal Mortality
b. Difficulty Seeing
c. Difficulty Hearing
d. Difficulty Walking/Climbing
e. Percentage of Province Determined to be “Urban”
f. How to use these scripts

a. Maternal Mortality

```
#Table 18 - Maternal Deaths
## Import Maternal Deaths File

#Run csv

MDeaths<-read.csv("Table 18 - Maternal Deaths.csv")


#assign variable names
LiveBirths = MDeaths$Number.of.Live.Births
MDeathsTotal = MDeaths$Total.Deaths
MDeathsHome = MDeaths$Home
MDeathsHospital = MDeaths$Hospital
MDeathsOther = MDeaths$Others

#Calculate province Maternal Mortality ratios
#add Maternal Mortality Rate variable to MaternalDeaths data frame as MMR
MDeaths$MMR = ((MDeathsTotal/LiveBirths)*100000.0)

#Calculate national MMR average
(sum(MDeathsTotal[1:10])/sum(LiveBirths[1:10]))*100000

#Barplots with MMR
# Set variables
MMR = ((MDeathsTotal/LiveBirths)*100000)
Provinces = MDeaths$Province

library(ggplot2)

## Horizontal bar plot
ggplot(MDeaths, aes(x=Province, y=MMR)) + geom_bar(state="identity",
fill="maroon 4", colour="black", width=0.7) + theme(text =
element_text(size=15), axis.text.x = element_text(angle=90, vjust=1)) +
coord_flip() + opts(title="Maternal Mortality Rates by Province") +
theme(plot.title=element_text(size=15))

#Plot same information into a dot plot
# Use variables from above
ggplot(MDeaths, aes(x=Province, y=MMR)) + geom_point(size=3) + coord_flip()
+ ggtitle("Maternal Mortality Rates by Province")
#Reordered dot plot so MMR is ascending
ggplot(MDeaths, aes(x=reorder(Province, MMR), y=MMR)) +
geom_point(size=3) + coord_flip() + ggtitle("Maternal Mortality Rates by
Province")
```

```
#Compare locations of deaths across provinces thorugh stacked horizontal bar
graph
#Constructed new csv to include Province, location of death, and location of
death's MMR through original table information
Stacked_MMR <- read.csv("~/Desktop/Capstone
copy/CSVs/Stacked_MMR.csv", header=TRUE)
View(Stacked_MMR)
#construct graph
ggplot(Stacked_MMR, aes(x=Province, y=MMR, fill=Location)) +
geom_bar(stat="identity") + ggtitle("Maternal Mortality Rates by Province and
Location") + coord_flip()

#Bar graphs showing proportions of maternal deaths at home, hospital, and other
by Province

#at home
ggplot(MDeaths, aes(x=MDeaths$Province,
y=((MDeathsHome/MDeathsTotal)*100))) + geom_bar(state="identity",
fill="coral", colour="black", width=0.7) + theme(text = element_text(size=15),
axis.text.x = element_text(angle=90, vjust=1)) + coord_flip() +
opts(title="Percent Maternal Deaths at Home by Province") +
theme(plot.title=element_text(size=15)) + ylab("Percent Maternal Deaths") +
xlab("Province")

#at hospital
ggplot(MDeaths, aes(x=MDeaths$Province,
y=((MDeathsHospital/MDeathsTotal)*100))) + geom_bar(state="identity",
fill="paleturquoise2", colour="black", width=0.7) + theme(text =
element_text(size=15), axis.text.x = element_text(angle=90, vjust=1)) +
coord_flip() + opts(title="Percent Maternal Deaths at Hospital by Province") +
theme(plot.title=element_text(size=15)) + ylab("Percent Maternal Deaths") +
xlab("Province")

#mapping MMR

library(ggplot2)
library(maptools)
library(maps)
library(mapdata)
library(mapproj)

#read csv that includes all 14 administrative regions (to correspond with map, 4
regions without census data will read "NA")
MaternalDeaths<-read.csv("deaths.csv", header=TRUE)
```

```
#To make map, begin with reading in shapefile, make sure working directory is
correct
dprk_shp<-readShapePoly("PRK_adm1.shp")

dprk_map<-fortify(dprk_shp)

Provinces<-read.csv("provinces.csv", header=TRUE) # province names and
locations for label

#read csv that shows all DPRK administrative regions (to be consistent with map,
which includes all 14, in contrast to the census which only shows 10)
#note that all administrative regions not included in census don't have data and
therefore have "NA" in vectors
deaths <- read.csv("deaths.csv")

#assign variable names
LiveBirths = deaths$Number.of.Live.Births
MDeathsTotal = deaths$Total.Deaths
MDeathsHome = deaths$Home
MDeathsHospital = deaths$Hospital
MDeathsOther = deaths$Others

#Maternal Mortality Rate (stat used by UN, uses to rank countries)
deaths$MMR = ((MDeathsTotal/LiveBirths)*100000.0)

#make breaks in MMR
deaths$MMRcut<-cut(deaths$MMR, breaks=c(65, 70, 75, 80, 85, 90),
include.lowest=TRUE, labels=c("65-70", "70-75", "75-80", "80-85", "85-90"))
#merge dataframes
MMRmap <- merge(dprk_map, deaths, by.x="id", by.y="ID_1")


### basic map
ggplot(dprk_map, aes(x=long, y=lat, group=group)) +
geom_polygon(fill="white", colour="black")

cnames <- aggregate(cbind(long, lat) ~ Province, data=MMRmap,
                    FUN=function(x)mean(range(x)))


theme_clean<-function(base_size= 12)
{
  require(grid)
  theme_grey(base_size) %+replace%
    theme(
      axis.title        = element_blank(),
```

```
    axis.text         = element_blank(),
    panel.background  = element_blank(),
    panel.grid        = element_blank(),
    axis.ticks.length = unit(0, "cm"),
    axis.ticks.margin = unit(0.01, "cm"),
    panel.margin      = unit(0, "lines"),
    plot.margin       = unit(c(0,0,0,0), "lines"),
    complete = TRUE
   )
}


ggplot(MMRmap, aes(long, lat)) +
  geom_polygon(aes(group=group), colour="black", fill="white") +
  geom_text(data=provinces, aes(long, lat, label = Province), size=2.0,
fontface="bold", angle=30) +
  coord_map()

## Add the data, make final map

ggplot(MMRmap, aes(x=long, y=lat, group=group)) +
  +   geom_polygon(aes(fill=MMRcut), colour="black") +
  +   labs(fill="MMR") + scale_fill_brewer() +
  +   ggtitle("Maternal Mortality Rate by Province" ) +
  +   guides(fill=guide_legend(reverse=TRUE)) +
  +   geom_text(data=provinces, aes(long, lat, group=NULL, label = Province),
size=4.0, fontface="bold", angle=30)
```

b. Difficulty Seeing

```
#Sight
#used slightly modified version of Table 42 (no age groups)
DifficultySeeing.df <- read.csv("~/Desktop/Capstone copy/CSVs/Table 42
corrected - Difficulty Seeing by age and urban-rural.csv")

##stat analysis

#test significance of all area gendered differences
# make vectors for tests, taking values from appropriate columns in
DifficultySeeing.df
# test gender by using "All Areas Male" and "All Areas Female" rows
# code those that cannot see as x and y, respectively, as cannot.see
# code totals for all areas male and female as x and y, respectively, as total.see
cannot.see <- c(2709, 4080)
total.see <- c(10187316, 11452504)

#prop.test to see significance of x (male) proportions consistently being lower
than y (female) proportions
prop.test(cannot.see, total.see, alt="less")

#repeat with different vectors/levels of difficulty seeing

difficulty.see <- c(26877, 39589)
prop.test(difficulty.see, total.see, alt="less")

some.difficulty.see <- c(186304, 260014)
prop.test(some.difficulty.see, total.see, alt="less")

# add all levels of difficulty to create new vector any.difficulty

any.difficulty <- c(DifficultySeeing.df$Some.difficulty +
DifficultySeeing.df$A.lot.of.difficulty + DifficultySeeing.df$Cannot.see.at.all)
any.difficulty #find values here to use for new vector for testing any difficulty
proportion of total
FM.any.difficulty <- c(215890, 303683)
prop.test(FM.any.difficulty, total.see, alt="less")

#test signifiance for urban/rural differences in proportions
# use "Urban Both Sexes" and "Rural Both Sexes" rows from DifficultySeeing.df

# make vector UR.cannot.see to include urban and rural, x and y respectively,
respondents who cannot see at all
UR.cannot.see <- c(3650, 3139)
```

```
# make vector UR.total to include urban and rural, x and y respectively, total
respondents
UR.total <- c(13161784, 8478036)
prop.test(UR.cannot.see, UR.total, alt="less")

#make vectors and tests for other levels of difficulty seeing
UR.difficulty.see <- c(37805, 28661)
prop.test(UR.difficulty.see, UR.total, alt="less")

some.difficulty.see <- c(260677, 185641)
prop.test(some.difficulty.see, UR.total, alt="less")

# add all levels of difficulty to create new vector any.difficulty

any.difficulty <- c(DifficultySeeing.df$Some.difficulty +
DifficultySeeing.df$A.lot.of.difficulty + DifficultySeeing.df$Cannot.see.at.all)
any.difficulty #find values here to use for new vector UR.any.difficulty for testing
any difficulty proportion of total
UR.any.difficulty <- c(302132, 217441)
prop.test(UR.any.difficulty, UR.total, alt="less")

# compare men and women across urban and rural areas
#begin with women across urban and rural areas, urban = x, rural = y
#use figures for any difficulty from any.difficulty vector (above), make new
vector with only urban and rural female figures of any.difficulty
F.any.difficulty <- c(181774, 121909)
#make F.total vector to include total females in urban and rural areas, respectively
F.total <- c(6952594, 4499910)
#prop test
prop.test(F.any.difficulty, F.total, alt="less")

#now same process with men
M.any.difficulty <- c(120358, 95532)
M.total <- c(6209190, 3978126)
prop.test(M.any.difficulty, M.total, alt="less")

#compare men and women within urban or rural areas, male = x, female = y
#begin with urban areas
U.any.difficulty <- c(120358, 181774)
U.total <- c(6209190, 6952594)
prop.test(U.any.difficulty, U.total, alt="less")

#now rural areas
R.any.difficulty <- c(95532, 121909)
R.total <- c(3978126, 4499910)
prop.test(R.any.difficulty, R.total, alt="less")
```

```
##visualizations

#remove rows you don't want
DifficultySeeingNew.df <- DifficultySeeing.df[-c(1, 4, 7), ]

library(ggplot2)

#example one category of difficulty (percentage)
ggplot(DifficultySeeingNew.df, aes(x=Age.and.Area,
y=(100*(A.lot.of.difficulty/Total)))) + geom_bar(position="dodge",
fill="dodgerblue2") +  ggtitle("Proportional Differences in Sight") +
ylab("Percent 'A Lot of Difficulty'") + xlab("Gender and Urban/Rural") +
coord_flip()

#graph blind
ggplot(DifficultySeeingNew.df, aes(x=Age.and.Area,
y=(100*(Cannot.see.at.all/Total)))) + geom_bar(position="dodge", fill="orchid2")
+  ggtitle("Proportional Differences in Seeing") + ylab("Percent 'Cannot See at
All'") + xlab("Gender and Urban/Rural") + coord_flip()

#add all difficulty to make "any difficulty" category
DifficultySeeingNew.df$Any.difficulty <-
DifficultySeeingNew.df$Some.difficulty +
DifficultySeeingNew.df$A.lot.of.difficulty +
DifficultySeeingNew.df$Cannot.see.at.all

#make "any difficulty" visualization
ggplot(DifficultySeeingNew.df, aes(x=Age.and.Area,
y=(100*(Any.difficulty/Total)))) + geom_bar(position="dodge",
fill="chartreuse3") +  ggtitle("Proportional Differences in Sight") + ylab("Percent
Any Difficulty") + xlab("Gender and Urban/Rural") + coord_flip()

#Does urban/rural make a difference?

### test for independence between urban/rural, male/female, and impairment
proportions
library(xtable)

## from your data, create the actual census dataframe with 21639820 observations
gender<-rep(c("male", "female"),c(10187316,11452504))
#add AnyDifficulty vector to DifficultySeeing.df
DifficultySeeing.df$AnyDifficulty <- c(DifficultySeeing.df$Some.difficulty +
DifficultySeeing.df$A.lot.of.difficulty + DifficultySeeing.df$Cannot.see.at.all)

#use data from DifficultySeeing.df to make vectors below
```

```
impaired<-rep(c("yes", "no", "yes", "no"), c(215890, 9971426, 303683,
11148821))
locm<-rep(c("urban", "rural", "urban", "rural"),
c(120358,95532,6088832,3882594))
locf<-rep(c("urban", "rural", "urban", "rural"),
c(181774,121909,6770820,4378001))
residence<-c(locm,locf)

#make new data frame with 3 variables above
Variables3.df<-data.frame(gender, impaired, residence)

## summarize it to check
summary(Variables3.df)

## make some simple tables

## simple table
table(Variables3.df$impaired)

## proportion table
prop.table(table(Variables3.df$impaired))

## now a three-way contingency table with residence
variables_cont_table<-xtabs(~impaired+gender+residence, data=Variables3.df)

## test for independence using Mantel-Haenzel test
mantelhaen.test(variables_cont_table)
```

c. Difficulty Hearing

```
#Hearing
#used slightly modified version of Table 43 (no age groups)
#import file
DifficultyHearing.df <- read.csv("~/Desktop/Capstone copy/CSVs/Table 43
modified - Difficulty Hearing by age, sex, urban-rural.csv")

##stat analysis

#test significance of all area gendered differences
# make vectors for tests, taking values from appropriate columns in
DifficultyHearing.df
# test gender by using "All Areas Male" and "All Areas Female" rows
# code those that cannot hear as x and y, respectively, as cannot.hear
# code totals for all areas male and female as x and y, respectively, as total.hear
cannot.hear <- c(3949, 5091)
total.hear <- c(10187316, 11452504)

#prop.test to see significance of x (male) proportions consistently being lower
than y (female) proportions
prop.test(cannot.hear, total.hear, alt="less")

#repeat with different vectors

difficulty.hear <- c(21138, 33995)
prop.test(difficulty.hear, total.hear, alt="less")

some.difficulty.hear <- c(115236, 195043)
prop.test(some.difficulty.hear, total.hear, alt="less")

# add all levels of difficulty to create new vector any.difficulty

DifficultyHearing.df$any.difficulty <- c(DifficultyHearing.df$Slight.Difficulty +
DifficultyHearing.df$A.lot.of.difficulty + DifficultyHearing.df$Cannot.do.at.all)
any.difficulty #find values in new column  to use for new vector for testing any
difficulty proportion of total
#make vector with values for male, female as x and y, respectively
FM.any.difficulty <- c(140323, 234129)
prop.test(FMany.difficulty, total.hear, alt="less")

#test signifiance for urban/rural
# use "Urban Both Sexes" and "Rural Both Sexes" columns from
DifficultyHearing.df
```

```
# make vector UR.cannot.see to include urban and rural, x and y respectively,
respondents who cannot hear at all
UR.cannot.hear <- c(4717, 4323)
UR.total <- c(13161784, 8478036)
prop.test(UR.cannot.hear, UR.total, alt="less")

#make vectors and tests for other levels of difficulty hearing
UR.difficulty.hear <- c(29941, 25192)
prop.test(UR.difficulty.hear, UR.total, alt="less")

UR.some.difficulty.hear <- c(177444, 132835)
prop.test(UR.some.difficulty.hear, UR.total, alt="less")

# add all levels of difficulty to create new vector any.difficulty

#find values to use for new vector UR.any.difficulty for testing any difficulty
proportion of total
UR.any.difficulty <- c(212102, 162350)
prop.test(UR.any.difficulty, UR.total, alt="less")

# compare men and women across urban and rural areas
#begin with women across urban and rural areas, urban = x, rural = y
#use figures for any difficulty from any.difficulty vector (above), make new
vector with only urban and rural female figures of any.difficulty
F.any.difficulty <- c(132861, 101268)
#make F.total vector to include total females in urban and rural areas, respectively
F.total <- c(6952594, 4499910)
#prop test
prop.test(F.any.difficulty, F.total, alt="less")

#now same process with men
M.any.difficulty <- c(79241, 61082)
M.total <- c(6209190, 3978126)
prop.test(M.any.difficulty, M.total, alt="less")

#compare men and women within urban or rural areas, male = x, female = y
#begin with urban areas
U.any.difficulty <- c(79241, 132861)
U.total <- c(6209190, 6952594)
prop.test(U.any.difficulty, U.total, alt="less")

#now rural areas
R.any.difficulty <- c(61082, 101268)
R.total <- c(3978126, 4499910)
prop.test(R.any.difficulty, R.total, alt="less")
```

```
##visualizations

library(ggplot2)

#remove rows you don't want
DifficultyHearingNew.df <- DifficultyHearing.df[-c(1, 4, 7), ]

#example one category of difficulty (percentage)
ggplot(DifficultyHearingNew.df, aes(x=X, y=(100*(A.lot.of.difficulty/Total)))) +
geom_bar(position="dodge", fill="dodgerblue2") +  ggtitle("Proportional
Differences in Hearing") + ylab("Percent 'A Lot of Difficulty'") + xlab("Gender
and Urban/Rural") + coord_flip()

#graph "cannot do at all"
ggplot(DifficultyHearingNew.df, aes(x=X, y=(100*(Cannot.do.at.all/Total)))) +
geom_bar(position="dodge", fill="orchid2") +  ggtitle("Proportional Differences
in Hearing") + ylab("Percent 'Cannot Hear at All'") + xlab("Gender and
Urban/Rural") + coord_flip()

#make "any difficulty" visualization
#first refine "any difficulty" from previous concatenation of same name
DifficultyHearingNew.df$any.difficulty <-
DifficultyHearingNew.df$Slight.Difficulty +
DifficultyHearingNew.df$A.lot.of.difficulty +
DifficultyHearingNew.df$Cannot.do.at.all
ggplot(DifficultyHearingNew.df, aes(x=X, y=(100*(any.difficulty/Total)))) +
geom_bar(position="dodge", fill="chartreuse3") +  ggtitle("Proportional
Differences in Hearing") + ylab("Percent Any Difficulty") + xlab("Gender and
Urban/Rural") + coord_flip()

#Does urban/rural make a difference?

### test for independence between urban/rural, male/female, and impairment
proportions
library(xtable)

## from your data, create the actual census dataframe with 21639820 observations
gender<-rep(c("male", "female"),c(10187316,11452504))

#use data from DifficultySeeing.df to make vectors below
impaired<-rep(c("yes", "no", "yes", "no"), c(140323, 10046993, 234129,
11218375))
locm<-rep(c("urban", "rural", "urban", "rural"),
c(79241,61082,6129949,3917044))
locf<-rep(c("urban", "rural", "urban", "rural"),
c(132861,101268,6819733,4398642))
```

```
residence<-c(locm,locf)

#make new data frame with 3 variables above
Variables3.df<-data.frame(gender, impaired, residence)

## summarize it to check
summary(Variables3.df)

## make some simple tables

## simple table
table(Variables3.df$impaired)

## proportion table
prop.table(table(Variables3.df$impaired))

## now a three-way contingency table with residence
variables_cont_table<-xtabs(~impaired+gender+residence, data=Variables3.df)

## test for independence using Mantel-Haenzel test
mantelhaen.test(variables_cont_table)
```

d. Difficulty Walking/Climbing

```
###Walking and climbing difficulty

DifficultyWalking.df <- read.csv("~/Desktop/Capstone copy/CSVs/Table 44 modified - Difficulty Walking by age, sex, rural-urban.csv")
>   View(DifficultyWalking.df)

##stat analysis

#test significance of all area gendered differences
# make vectors for tests, taking values from appropriate columns in DifficultyWalking.df
# test gender by using "All Areas Male" and "All Areas Female" rows
# code those that cannot see as x and y, respectively, as cannot.walk
# code totals for all areas male and female as x and y, respectively, as total.walk
cannot.walk <- c(5197, 5939)
total.walk <- c(10187316, 11452504)

#prop.test to see significance of x (male) proportions consistently being lower than y (female) proportions
prop.test(cannot.walk, total.walk, alt="less")

#repeat with different vectors/levels of difficulty seeing

difficulty.walk <- c(49237, 72825)
prop.test(difficulty.walk, total.walk, alt="less")

some.difficulty.walk <- c(146391, 257907)
prop.test(some.difficulty.walk, total.walk, alt="less")

# add all levels of difficulty to create new vector any.difficulty

any.difficulty <- c(DifficultyWalking.df$Slight.difficulty + DifficultyWalking.df$A.lot.of.difficulty + DifficultyWalking.df$Cannot.do.at.all)
any.difficulty #find values here to use for new vector for testing any difficulty proportion of total
FM.any.difficulty <- c(200825, 336671)
prop.test(FM.any.difficulty, total.walk, alt="less")

#test signifiance for urban/rural differences in proportions
# use "Urban Both Sexes" and "Rural Both Sexes" rows from DifficultyWalking.df

# make vector UR.cannot.walk to include urban and rural, x and y respectively, respondents who cannot walk/climb at all
```

```
UR.cannot.walk <- c(6469, 4667)
# make vector UR.total to include urban and rural, x and y respectively, total
respondents
UR.total <- c(13161784, 8478036)
prop.test(UR.cannot.walk, UR.total, alt="less")

#make vectors and tests for other levels of difficulty walking/climbing
UR.difficulty.walk <- c(70558, 51504)
prop.test(UR.difficulty.walk, UR.total, alt="less")

some.difficulty.walk <- c(239460, 164838)
prop.test(some.difficulty.walk, UR.total, alt="less")

# add all levels of difficulty to create new vector any.difficulty

any.difficulty <- c(DifficultyWalking.df$Slight.difficulty +
DifficultyWalking.df$A.lot.of.difficulty + DifficultyWalking.df$Cannot.do.at.all)
any.difficulty #find values here to use for new vector UR.any.difficulty for testing
any difficulty proportion of total
UR.any.difficulty <- c(316487, 221009)
prop.test(UR.any.difficulty, UR.total, alt="less")

# compare men and women across urban and rural areas
#begin with women across urban and rural areas, urban = x, rural = y
#use figures for any difficulty from any.difficulty vector (above), make new
vector with only urban and rural female figures of any.difficulty
F.any.difficulty <- c(198661, 138010)
#make F.total vector to include total females in urban and rural areas, respectively
F.total <- c(6952594, 4499910)
#prop test
prop.test(F.any.difficulty, F.total, alt="less")

#now same process with men
M.any.difficulty <- c(117826, 82999)
M.total <- c(6209190, 3978126)
prop.test(M.any.difficulty, M.total, alt="less")

#compare men and women within urban or rural areas, male = x, female = y
#begin with urban areas
U.any.difficulty <- c(117826, 198661)
U.total <- c(6209190, 6952594)
prop.test(U.any.difficulty, U.total, alt="less")

#now rural areas
R.any.difficulty <- c(82999, 138010)
R.total <- c(3978126, 4499910)
```

```
prop.test(R.any.difficulty, R.total, alt="less")


##visualizations

library(ggplot2)

#remove rows you don't want
DifficultyWalkingNew.df <- DifficultyWalking.df[-c(1, 4, 7), ]

#example one category of difficulty (percentage)
ggplot(DifficultyWalkingNew.df, aes(x=X, y=(100*(A.lot.of.difficulty/Total)))) +
geom_bar(position="dodge", fill="dodgerblue2") +  ggtitle("Proportional
Differences in Walking") + ylab("Percent 'A Lot of Difficulty'") + xlab("Gender
and Urban/Rural") + coord_flip()

#graph "cannot do at all"
ggplot(DifficultyWalkingNew.df, aes(x=X, y=(100*(Cannot.do.at.all/Total)))) +
geom_bar(position="dodge", fill="orchid2") +  ggtitle("Proportional Differences
in Walking") + ylab("Percent 'Cannot Walk at All'") + xlab("Gender and
Urban/Rural") + coord_flip()

#add all difficulty to make "any difficulty" category
DifficultyWalkingNew.df$Any.difficulty <-
DifficultyWalkingNew.df$Slight.difficulty +
DifficultyWalkingNew.df$A.lot.of.difficulty +
DifficultyWalkingNew.df$Cannot.do.at.all

#make "any difficulty" visualization
ggplot(DifficultyWalkingNew.df, aes(x=X, y=(100*(Any.difficulty/Total)))) +
geom_bar(position="dodge", fill="chartreuse3") +  ggtitle("Proportional
Differences in Walking") + ylab("Percent Any Difficulty") + xlab("Gender and
Urban/Rural") + coord_flip()

#Does urban/rural make a difference?

### test for independence between urban/rural, male/female, and impairment
proportions
library(xtable)

## from your data, create the actual census dataframe with 21639820 observations
gender<-rep(c("male", "female"),c(10187316,11452504))
#add AnyDifficulty vector to DifficultySeeing.df
DifficultyWalking.df$AnyDifficulty <- c(DifficultyWalking.df$Slight.difficulty +
DifficultyWalking.df$A.lot.of.difficulty + DifficultyWalking.df$Cannot.do.at.all)
```

```
#use data from DifficultySeeing.df to make vectors below
impaired<-rep(c("yes", "no", "yes", "no"), c(200825, 9986491, 336671,
11115833))
locm<-rep(c("urban", "rural", "urban", "rural"),
c(117826,82999,6091364,3895127))
locf<-rep(c("urban", "rural", "urban", "rural"),
c(198661,138010,6753933,4361900))
residence<-c(locm,locf)

#make new data frame with 3 variables above
Variables3.df<-data.frame(gender, impaired, residence)

## summarize it to check
summary(Variables3.df)

## make some simple tables

## simple table
table(Variables3.df$impaired)

## proportion table
prop.table(table(Variables3.df$impaired))

## now a three-way contingency table with residence
variables_cont_table<-xtabs(~impaired+gender+residence, data=Variables3.df)

## test for independence using Mantel-Haenzel test
mantelhaen.test(variables_cont_table)
```

e. Percentage of Provinces Determined to be "Urban"

```
#To map provinces' percentage of area urban

library(ggplot2)
library(maptools)
library(maps)
library(mapdata)
library(mapproj)

#import shapefile used as map (data to be overlayed)
dprk_shp<-readShapePoly("../shapefiles/PRK_adm1.shp")
#import map data
dprk_data<-read.csv("../data/PRK_adm1.csv", header=TRUE)
#import province names and locations for map labeling
provinces<-read.csv("../data/provinces.csv", header=TRUE)

#fortify map to turn it into data frame
dprk_map<-fortify(dprk_shp)

#import urban/rural data, make vectors
urbanrural<-read.csv("../data/table2.csv", header=TRUE)
urbanrural$pop<-urbanrural$Umale+urbanrural$Ufemale+urbanrural$Rmale+
urbanrural$Rfemale
urbanrural$urban<-urbanrural$Umale+urbanrural$Ufemale
urbanrural$propurban<-urbanrural$urban/urbanrural$pop

#make breaks in urban percentage
urbanrural$propurbancut<-cut(urbanrural$propurban, breaks=c(.1, .2, .3, .4, .5, .6,
.7, .8, .9, 1.0), include.lowest=TRUE,
                        labels=c("10-20%", "20-30%", "30-40%", "40-50%", "50-
60%", "60-70%", "70-80%", "80-90%", "90-100%"))


##merge data to allow later mapping
urban_map <- merge(dprk_map, urbanrural, by.x="id", by.y="ID_1")

### basic outline map
ggplot(dprk_map, aes(x=long, y=lat, group=group)) +
geom_polygon(fill="white", colour="black")

## Add province names
province_names<-urbanrural$Province

# Create location values for province names
cnames <- aggregate(cbind(long, lat) ~ Province, data=urban_map,
```

```
                FUN=function(x)mean(range(x)))


# Make a clean theme to eliminate background elements
theme_clean<-function(base_size= 12)
{
  require(grid)
  theme_grey(base_size) %+replace%
    theme(
      axis.title        = element_blank(),
      axis.text         = element_blank(),
      panel.background  = element_blank(),
      panel.grid        = element_blank(),
      axis.ticks.length = unit(0, "cm"),
      axis.ticks.margin = unit(0.01, "cm"),
      panel.margin      = unit(0, "lines"),
      plot.margin       = unit(c(0,0,0,0), "lines"),
      complete = TRUE
    )
}

## Outline with province names
ggplot(urban_map, aes(long, lat)) +
  geom_polygon(aes(group=group), colour="black", fill="white") +
  geom_text(data=provinces, aes(long, lat, label = Province), size=2.0,
fontface="bold", angle=30) +
  coord_map()

## Combine province names and data, make final map
ggplot(urban_map, aes(x=long, y=lat, group=group)) +
  +   geom_polygon(aes(fill=propurbancut), colour="black") +
  +   labs(fill="Percent Urban") + scale_fill_brewer() +
  +   ggtitle("DPRK Percent Urban by Province" ) +
  +   guides(fill=guide_legend(reverse=TRUE)) +
  +   geom_text(data=provinces, aes(long, lat, group=NULL, label = Province),
size=4.0, fontface="bold", angle=30)
```

f. How to use these scripts

The above scripts are the author's personal coding[40] that were used to create the visualizations and conduct the statistical analysis for this project. All visualizations and data analysis performed for this study, therefore, can be reproduced using the above scripts. All scripts were designed for R statistical analysis.[41] The author chose the names of all CSVs, variables, and vectors used in the above coding, and therefore researchers reproducing her work may choose to use different names without receiving fundamentally different visualizations and analysis. The basic structure remains functional regardless of differing names.

To access the machine-readable data used in the above scripts, visit: http://maxwell.syr.edu/moynihan/kpac/Datasets/2008_DPRK_Census_Data/.

To contribute improvements to the above codes, email your suggestions to the author at mklinhor@syr.edu.

[40] The author's mentor, Professor Stuart Thorson, assisted the author in writing many of the above scripts.

[41] R can be downloaded at: http://www.r-project.org/