Dear Executive Director Taylor,

In the spirit of the multidisciplinary nature of data science, my research proposal will combine the necessity of economic development and the significance of gender roles through a thorough analysis of the largest developing country in the world. By investigating the research question of what relationship does economic development share with women's wages and participation in the labor force, I hope to not only evaluate the country's unorthodox approach to human development, but also answer a question essential for the modern world's fight for social justice. Will the modern era's successes of universal continuous economic growth revolutionize gender equality, or does economic progress come at the cost of social progress?

To answer this essential question, my study suggests the construction of a low-cost gender disaggregated spatial model of women's advancement versus economic development made possible by utilizing both China's advanced infrastructure and cutting-edge data science methods. In today's modern age of governments prioritizing economic growth, it is imperative that we understand how current policies will affect gender equality. Instead of rejecting conventional economic statistics due to their inability to provide meaningful insight into overlooked but equally important aspects of human development, my research proposal chooses to integrate the methodology from previously established studies to answer a more wholistic question by combining economic data with relatively easy to generate gender distribution datasets.

By bridging a gap between financial and sociological research, I hope to obtain funding from the UN Women organization to help in your mission to ensure that women around the world have decent work and economic autonomy.

Thank you for your consideration,

Maxtense

Since the conclusion of the Second World War, the global community of the United Nations has led the transition of humanity's focus from the expansion of empires to improving the welfare of all peoples. This unprecedented shift in from inherently unequal multiethnic empires to self-ruled nation states has allowed the global community to adopt the technologies of the industrial revolution for development rather than foreign exploitation. The transition that followed has allowed humanity to exponentially progress towards the goal of ensuring that all people have the means to live fulfilling lives without the fear of disease, violence, or poverty. However, as independent developing nations continue to create the foundations needed for a modern society, the goal of maximizing general welfare is often at odds with economic efficiency, political expediency, and the contradictory priorities of different demographic groups. In our data driven world, information is key to understanding the essential factors to developing a society to 'developed' standards. Throughout post Second World War history, there have been many instances of economic 'miracles' in which formerly impoverished countries were quickly able to reach developed economic standards, including Singapore, South Korea, Japan, and many more. However, the People's Republic of China's rapid economic expansion stands out among the others due to many factors. Unlike other countries with similar success stories, modern China was not subject to direct economic influence from Western powers, was diplomatically isolated, and had an initially unimpressive education system. However, the country continues to experience unprecedented growth in gross domestic product today while also avoiding many of the pitfalls experienced by similar countries. Unlike countries that experienced high post war growth rates that have stagnated, China's financial sector has healthy interest rates and maintains mutual trust between its citizens and its central bank. These factors indicate that understanding China's successes may be the key to a blueprint to help developing nations create and maintain sustainable growth.

While issues such as public health and access to food are certainly factors that are essential to creating a developed society of United Nation's standards, there is a clear causal relationship between economic development and most of the other sustainable development goals. History has shown that while many countries that have experienced rapid economic development may not have adequate health facilities, there are conversely no countries that can maintain health facilities without adequate capital. In the spirit of finding factors that are as universal as possible to encouraging growth, I have decided to investigate the spread of Chinese urbanization. Urbanization provides an adequate balance between humanitarian and economic approaches to understanding human development since it can be spatially mapped, but still is directly correlated with growth.

Additionally, there are significant harms that a country without adequate urbanization faces. Cities create a concentration of resources and people, which is essential for human development due to the concept of economies of scale. As human and financial capital become more concentrated, returns on education and investment increase exponentially due to increased connections between different agents. In areas without cities, returns generated from both physical and knowledge-based infrastructure simply cannot be greater than the costs associated with their initial investment. Therefore, society remains in a stagnant state of agriculture where output cannot be improved after a certain threshold, even as population continues to increase.

This idea of human capital providing diminishing rather than increasing returns was represented of reality before the technologies of the industrial revolution and provided the basis of Thomas Malthus's economic model of growth. Without cities, a developing society will be trapped in this Malthusian model of stagnation and be unable to achieve the goals determined by the United Nations to be essential for human development.

Although many western audiences understand the nature of the People's Republic of China's planned economy through the lens of eastern European models, the Communist Party of China's economic policies had a much more profound impact on the country's human development than its counterparts. While the regulation imposed by other Communist states was largely economic, Chinese regulations also centrally managed many other aspects of life that are today considered an integral part of human development. Up until 1983, the Chinese state had replaced traditional levels of administration with a system known as 'people's communes. These organizational bodies were different from townships in that they centrally managed essentially all aspects of life, including even the movement of citizens and meal preparation. During this period of central planning, China's development continued to lag other Asian countries until the abolishment of the people's communes and the economic reforms of current Chinese president Deng Xiaoping. President Xiaoping is credited with moving China's economic system from a centrally planned to a socialist market economy, which has led to a profound transformation in the country's urbanization and social structure. While most cities and institutions throughout the developed world have been shaped by centuries of individual choices, China's unique history has created circumstances in which the largest population in human history will reshape itself within the span of a couple decades. This creates a unique opportunity for data scientists, who for the first time in history will be able to analyze emerging complexity of urbanization with modern technology.

Although the standards of living of Chinese citizens within coastal provinces have risen significantly over the last few decades, populations within inland rural provinces have remained stagnant in terms of education, wages, and even lifespans. This development has created a concerning amount of inequality between the two regions, which could result in social unrest in the future. There are a plethora of problems facing rural Chinese citizens, including a lack of access to credit, inability to receive proper education, and exposure to environmental harms such as air pollution. The inherent nature of the gap between coastal and urban populations can be attributed to the United Nation's sustainability goal of urban development. I have attempted to investigate the cause of a lack of urbanization within rural China with my research, as well successful measures that have worked on a smaller scale.

Although the goal of urbanization may initially seem straightforward, there are many agents, existing social services, and complex interconnected networks that the Chinese central government faces in encouraging urbanization. Most of China's inland provinces are made up of scattered small communities governed by village leaders who associate the existing rural lifestyle with their unique regional culture that is distinct from the largely ethnic Han dominated cities. Additionally, many farmers are resistant to diversifying because they lack money to take on the

risks associated with investing in unfamiliar industries. This is compounded by the stance of China's financial sector, which see it was a potential liability to give farmers who don't have a credit history access to loans. The Chinese government must also contend with the complex network of migration when making decisions: any given decision could cause millions of citizens to migrate to other provinces, and each citizen will make a unique decision based on their wealth and education level. An example of this problem can be seen in a previous government program to offer farmers government subsidies for buying new land or investing in more profitable cash crops. Much of the land that was intended to benefit farmers was instead bought out by citizens by eastern provinces who used their wealth and education to further expand into the communities of those the program was intended to help. Directly investing inland may also negatively impact the growth of already large eastern cities, whose rural immigration to more efficient healthcare systems reduced total healthcare costs by an estimated 31 billion dollars last year. Evidently, migration is an emerging social and economic system that must be considered before any decisions are made by the government to encourage urbanization.

The urbanization of China can be categorized as an emerging complex system due to the variety of people of variables. Although urbanization precipitated by the movement of people from lower to higher density areas, this movement is influenced by relative differences in wealth and education, as well as the interests of China's central government. In addition, movement is influenced by geographical and economic factors that are outside the direct control of any party. The complexity of this system makes it useful to identify sub research questions that break up the problem into more approachable parts. First, it would be useful to identify what factors cause migration. It is also necessary to find out what factors make a city appealing to potential migrants, since this may help predict future migration. Lastly, it is important to acknowledge that the economy of China is still more regulated than any other developed nation. Can we still use traditional macroeconomic statistics to assess development, or will the distortion of market forces make econometric models ineffective?

In order to properly assess the urbanization of regions throughout China, it was necessary to pick statistics that could be both quantitatively measured and spatially aggregated. This would allow not only a net assessment of the country at large, but also an investigation into if China's successes have been enjoyed by all agents of its diverse population. During my research, I chose to focus on analyzing the intertemporal changes in real estate prices and population flow to measure urbanization. This choice was made to address some of the weaknesses of analyzing human development through urban populations. In many developing countries, there are large cities that still are unable to provide adequate access to essential services due to a variety of reasons. *The City of God's* depiction of a favela within Rio de Janeiro is a perfect example of how city-based metrics may be misleading. Brazil's capital is a major urban area that still provides many of its citizens with insufficient protection and access to opportunities. While there are many reasons for the city's shortcomings, the film focuses on systemic issues such as corruption and racial discrimination. To avoid misidentifying a city's progress in achieving human development, I chose to focus on factors that would paint a picture of how a given city interacts within the country's larger complex system. If a city truly provided opportunities for its citizens, it would be a reasonable assume that it would be both productive and attractive to potential migrants. Therefore, population flow and real estate prices were chosen as reasonably accurate metrics for determining a city's progress in human development. Both studies that I investigated used a variety of different data science methods, but both relied on the principle of regression.

In order to derive needed data and verify results, both studies relied on a larger collection of other data science methods. Some of these methods were designed to provide random simulations, while others were used to find patterns in collected data. The complementary data science methods used in both papers include autoregressive integrated moving average models, granger causality tests, exponential random graph models, random walk models, rank size models, and Bayesian networks.

Autoregressive integrated moving average models are designed to forecast the future change of a specific variable. In these models, it is assumed that the variable being analyzed is primarily influenced by the trend of its prior values over time. Therefore, scientists can use a set of timeseries data of the variable from the past to predict future change and. Since this method can accurately model the growth of important economic statistics but requires large amounts of data, it is used extensively by econometric studies.

Granger causality tests are a variation of traditional hypothesis tests that can determine if one variable likely caused changes in another variable through the analysis of timeseries data. Although granger causality tests are liked by statisticians for their simplicity, researchers must be careful not to incorrectly associate a supposed causality with datasets that show an arbitrary relationship out of pure coincidence. Due to this method's versality and potential for 'false

positive' conclusions of causality, there has been debate among scholars if it is bad practice to use granger causality tests outside of STEM related fields of study.

Exponential random graph models are a family of methods used to analyze a network. These take the form of mathematical distributions that can predict a desired variable by looking at the relationship between the network's nodes and connections. Due to the complex nature of a network, exponential random graph models are especially useful since they can help a researcher find 'almost anything' given enough data. For instance, one can decide if they want to find the amount of connections shared by any number of nodes or find the relative popularity of any vertex in a network among its nodes. The large scale of exponential random graph models makes them popular in sociology studies.

Random walk models are simulations that model an object taking a set number of completely random actions. Since almost all aspects of life are at least somewhat inherently random, the method can be used to model almost anything across any number of dimensions, be that a one-dimensional number line or a three-dimensional landscape. Given that random walk models are an integral part of probability, they are used in almost any field of study from game theory to molecular biology.

Rank size models are visual representations of data that graph datapoints based on their relative relationships. Generally, this means graphing the magnitude of the statistic on the y axis while graphing the relative ranking of the datapoint on the x axis. This method is especially useful for analyzing datasets where there are a few very large samples, and many smaller samples. Examples include people's heights, city populations, or the gross domestic product of countries.

Bayesian networks are networks of conditions that are linked by the causal relationship shared by each node. For instance, a Bayesian network about public health could show that obesity and genetics may cause diabetes, and that diabetes may also contribute to obesity. However, the network would still show that it has not been proven that type 1 diabetes directly can influence genetics through inheritance. Bayesian networks are similar in utility to granger causality tests, but they can be used on a much wider scale both in terms of number of variables and size. In today's modern world of big data, machine learning algorithms use Bayesian networks frequently to analyze entire populations. Due to these developments, methods such as Bayesian networks have laid the foundations for interdisciplinary fields between computer science and the humanities.

As stated by United Nations' director Selim Jahan, human development can be defined in its most basic terms as the availability of choices available to citizens that can allow them to live free and fulfilling lives. While freedom may be easy can recognize on an individual level, its abstract nature makes it difficult to measure on a larger scale. Data science methods allow researchers to try to abstract ideas like welfare or freedom into concrete statistics, which then can be transformed models that can provide suggestions, causality, or even predictions of the future.

However, abstracting complex systems into easy to digest models comes with many limitations, even with modern technology. Models are compromises between accuracy and understandability, and researchers must forgo information that could create useful insights in order to reach meaningful conclusions. While this can be most clearly seen in regression models, even a researcher's choice of a statistic is an active decision prioritize analyzing one aspect of a system over another. Despite these obstacles, the studies I analyzed were still able to draw meaningful conclusions about their respective areas of study by narrowing their focus to a specific area of interest and acknowledge the limitations of their selected data science methods.

In the study involving real estate prices, cointegration was used to determine the relationship between real inflation and returns on real estate. Since real returns must account for losses due to inflation, researchers needed to model the relationship between inflation and returns. Cointegration is a method used to determine the relationship between two variables and does not need information regarding causality in order to be used. The equation used by the study is the following:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Y is the predicted returns, alpha and beta are constants created by the regression model, and x is a vector of the independent variables used to model returns. The study integrated a variety of independent variables for each region, an example of the variables for a few regions can be found here:

Exhibit 2
Descriptions of the Regression Variables

Variables	Description
RP	Total return of real estate
RRP	Return of residential property
ROP	Return of office property,
RCP	Return of commercial property
EIR	Expected inflation rate
UIR	Unexpected inflation rate
AIR	Actual inflation rate
GRATE	GDP growth rate
RRETURN	Real return of stock market

Exhibit 1
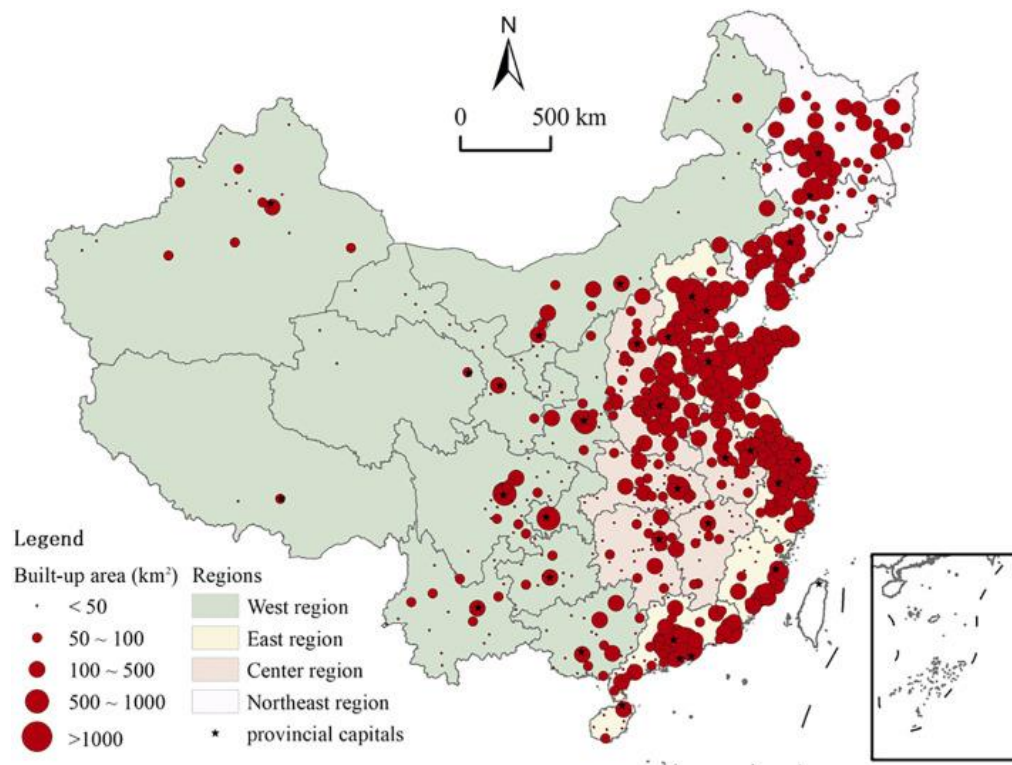### Inflation Rates in Beijing, Shanghai, Shenzhen and Chengdu

City	Beijing			Shanghai			Shenzhen			Chengdu		
Inflation	UIR	EIR	AIR	UIR	EIR	AIR	UIR	EIR	AIR	UIR	EIR	AIR
Mean	−0.0007	−0.0056	−0.0063	0.0000	−0.0052	−0.0053	0.0021	−0.0039	−0.0018	0.0001	−0.0051	−0.0049
Median	0.0009	−0.0042	−0.0034	−0.0018	−0.0041	−0.0014	0.0006	−0.0018	−0.0040	0.0004	−0.0053	−0.0029
Max.	0.0254	0.0113	0.0259	0.0728	0.0273	0.0688	0.0264	0.0353	0.0453	0.0175	0.0104	0.0248
Min.	−0.0202	−0.0223	−0.0387	−0.0336	−0.0311	−0.0503	−0.0220	−0.0363	−0.0428	−0.0140	−0.0255	−0.0396
Std. Dev.	0.0124	0.0083	0.0147	0.0205	0.0115	0.0238	0.0113	0.0173	0.0193	0.0100	0.0086	0.0133
Skewness	0.0262	−0.1161	−0.1280	1.3283	0.3621	0.5134	0.0597	0.1507	0.3595	0.1627	−0.1878	−0.5171
Kurtosis	2.2304	2.7896	2.8089	6.5335	4.0807	4.5259	2.9632	2.4672	3.0426	1.7761	2.5661	3.7090
Jarque−Bera	0.7437	0.1227	0.1275	2.4429	2.1156	4.2288	0.0175	0.4215	0.5839	1.8043	0.3705	1.7691
Probability	0.6894	0.9404	0.9381	0.0000	0.3472	0.1207	0.9912	0.8099	0.7468	0.4056	0.8308	0.4128
Sum	−0.0201	−0.1675	−0.1876	−0.0013	−0.1582	−0.1596	0.0580	−0.1079	−0.0498	0.0041	−0.1389	−0.1348
Sum Sq. Dev.	0.0044	0.0020	0.0063	0.0122	0.0038	0.0164	0.0033	0.0078	0.0096	0.0026	0.0019	0.0046

Additionally, epsilon represents the residuals, or distances between the datapoints inputted and trendline created by the regression model.

In the study involving population flow, logarithmic rank size models were used to model the popularity of cities relative to different statistics. The formula of each rank size model is the following:

$$lnB_i = lnB_1 - qlnR_i$$

B describes the relative size of any given area, while R is the ranking of that area relative to others. The variables i and 1 distinguish if the area is being measured relative to the city it is in, or a baseline major city that is used for all comparisons. The variable q models the current economic development of a given city. The rank size models were then used to generate a spatial distribution of China's urban areas:

Legend

Built-up area (km²)	Regions
· < 50	West region
• 50 ~ 100	East region
● 100 ~ 500	Center region
⬤ 500 ~ 1000	Northeast region
⬤ >1000	★ provincial capitals

The conclusions of both studies indicated that there was a significant and universal growth of urban areas throughout China, although there were some observations that indicate the unique nature of China's urbanization process. The real estate prices study concluded that investors in real estate suffered steep loses, which does indicate that the growth in real estate prices, or development of land was larger than the growth in inflation. However, the study also found that the real rate of inflation for China in a large section of the timeseries data was negative, which is usually an indicator of a severe recession. Due to this discrepancy between the study's findings and typical economic theory, I have concluded that the regulations of China's socialist economy make attempts to evaluate urbanization through macroeconomic statistics impractical.

In contrast to the real estate study, the data provided by the population flow study was shown to be effective by its researchers. The models created by the study were consistent with our general understanding of China's population distribution and created incredibly concise data that could help with both future studies and urban planning. However, researchers found that unlike almost all other studies, economic growth was not a significant factor in determining were migrants settled. Instead, it found that relief degree of land service and geographical region determined migration. Identifying the cause behind economic growth's lack of influence behind population flow is an important research gap that should be investigated in the future. Another research gap is a lack of data targeted at gender equality. Although there is a wide variety of data available concerning economic development and population mapping, data related to women's issues such as access to birth control or gender disaggregated data is notably lacking in Chinese studies.

Although there may be ample research on China's financial markets, urban development, and health, the country's economic importance to the global community has create a research gap in terms of women's issues. In contrast to other developing countries, studies on China lack sets of gender disaggregated data and fail to evaluate China's performance on equally important issues such as women's advancement. To address this research gap, my research proposal will focus on using China as a model to investigate if there is a relationship between economic development and the evolution of gender roles.

Although my literature review did show that the People's Republic of China was able to achieve unprecedented and sustained economic growth throughout the last few decades, the distribution of wealth between regions was incredibly unequal. This gap between the growth rates of different areas has increased cultural differences between regions, which can allow researchers to gain a better understanding of how economic development can change socioeconomic factors such as gender roles. Since there is a unified cultural identify throughout most of the country, there is a great opportunity to evaluate how advancements in other areas of human development can influence these socioeconomic factors over time.

In order to investigate the correlation between economic development and the evolution of gender roles within a rapidly changing socioeconomic system, my research proposal suggests creating a spatially disaggregated set of data that analyzes the gender gap between important economic statistics such as wages and labor force participation. In order derive these gender disaggregated statistics, traditional spatial data from GPS technology will be combined with information from social media, newly created census data, and official government records. This spatial distribution of gender disparities can then be compared to various potential causal factors such as access to education, marriage status, and access to birth control.

I believe that this research proposal is especially promising due to its low associated costs relative to its potential benefits, as well as the importance of its conclusions for the global community. China already has ample economic data in addition to the necessary technological infrastructure to collect data and gender distributions quickly and cheaply. The scale and diversity of the country can allow this easy collected data to be used to answer larger societal questions about humanity's future, such as if achieving gender equality is an inevitability in the face of economic progress or if it is an ideal that must be continually worked towards to maintain.

While previous studies have already generated spatially disaggregated data to measure population distribution and flow, these studies did not further categorize each sample based on different socioeconomic factors. With modern social media and cellphone devices, information such as gender is now stored on the same device as location. By integrating this extra data into a model, there is the potential to draw conclusions about social as well as economic progress. Additionally, cellular data can also be reinforced by more conventional means of data collection, including government and newly collected census data. Using multiple sources of data will improve the redundancy of the study, ensuring better accuracy and a more thorough verification process.

My current hypothesis for my study would be that increased economic development should directly lead to better gender equality. History has generally shown that gender roles follow a U-shaped curve. Hunter gatherer societies tend to necessitate equal gender roles in order to gather natural resources, while agrarian communities develop institutions and cultures that can create and enforce traditional gender roles. However, after adopting industrial technology, a society's

access increased education and freedom of thought can allow individuals to question the artificial nature of previously established customs. This evolution of gender roles can be seen almost universally throughout different human cultures, regardless of different geographical regions and cultural norms. As even the most rural parts of China continue to become more interconnected and technologically advanced, I expect this process to repeat itself.

There are two objectives to my research proposal: first, to evaluate how China's unique circumstances and actions have influenced gender roles on a macro level, and also to use the combination of China's unified cultural identity and diverse economic circumstances to investigate if there is an inherent relationship between economic and social development. In order to create a gender disaggregated spatially distributed model of gender equality versus economic growth, a variety of data science methods will be used. Bayesian networks will be used to draw connections between collected social and economic variables and will be supplemented by granger causality tests to ensure accuracy. Additionally, a rank size model will be created to investigate how the relative magnitude of any recognized trend changes at differing levels of wealth. The final product of the study will be map of China with each province assigned a value that represents the disparity of opportunities between genders relative to its general wealth compared to other provinces.

Although the process of electronically gathering information should be easy relative to other developing countries, there are many potential obstacles that are important to acknowledge. Throughout its modern history, China has often had a strained relationship with the west and its citizens are subject to much stricter rules for accessing information than other technologically advanced countries. For instance, the country is often said to be home to 'The Great Firewall of China', which is a series of protocols managed by the Cyberspace Administration of China to control and manage citizens' access to information. These circumstances have led to state affiliated domestic corporations filling the market of essential internet services such as social media and search engines. Given the often-controversial nature of social justice issues such as gender equality, it may be difficult to convince these companies to give the study access to their information. Moreover, this control of information extends to even the gathering of census data as well. Similar to many other authoritarian regimes, local authorities are often uncooperative with foreigners who try to question locals. Recently China has even expelled American journalists from the country in retaliation for American actions towards Chinese media outlets. These physical and legal barriers to obtaining information will likely influence the information gathering process of the study.

Compared to other potential studies, I believe that my research proposal may offer the best value due to its low relative costs and multidisciplinary nature. Even though China may have certain legal obstacles to obtaining data, the availability of electronic devices still will allow the implementation of a top down approach to research, which will inevitably generate models with better accuracy. My study also analyzes societal problems that are universal rather than country specific, which will allow the its conclusions to be used by a wider variety of researchers after its publication. Economic development is one of the few priorities that are sought after by every

government, and gender equality remains an issue in both the most and least developed countries. The economic significance of China has made the country a model for other developing countries, and its influence in regions such as sub Saharan Africa will make it likely that the country's strategies for economic development will be replicated my many others in the future. This makes it imperative for humanitarian organizations to understand how the Chinese approach to human development will impact the roles of millions of women in the developing world.

Relative to other potential studies, the advantages of analyzing a more technologically advanced country will require less time than others that must utilize a bottom up approach to gathering data. Additionally, the electronic component of the data collection will significantly reduce the overhead costs associated with digitizing and processing data from more traditional sources. However, as mentioned before, the rising tensions between China and the west necessitates a timeframe dedicated to handling potential unforeseen roadblocks. If necessary, a portion of the budget may need to be dedicated to paying for cell data, and census data could be supplemented by existing government census records if personal collection is deemed to be impractical on a larger scale due to legal concerns. Despite these obstacles, the digital nature of this study's data processing phase will minimize time and costs after the initial challenges of data collection are completed.

Bibliography

Zheng, Siqi. and Khan, Matthew, China's Bullet Trains Facilitate Market Integration and Mitigate the Cost of Megacity Growth (March 18, 2013). https://www.pnas.org/content/pnas/110/14/E1248.full.pdf

Démurger, Sylvie and Fournier, Martin and Yang, Weiyong, Diversification and Agrarian Change under Environmental Constraints in Rural China: Evidence from a Poor Township of Beijing Municipality (April 1, 2007). http://dx.doi.org/10.2139/ssrn.988057

Turvey, Calum G. and Kong, Rong, Farmers' Willingness to Purchase Weather Insurance in Rural China (May 6, 2010). https://ssrn.com/abstract=1601625

Aunan, Kristin, Shuxiao, Wang, Internal Migration and Urbanization in China: Impacts on Population Exposure and to Household Air Pollution (2016). https://www.sciencedirect.com/science/article/pii/S0048969714002472

Gaughan, A., Stevens, F., Huang, Z. et al. Spatiotemporal patterns of population in mainland China, 1990 to 2010. Sci Data 3, 160005 (2016). https://doi.org/10.1038/sdata.2016.5

Chu, Qingqiang. And Sing, Tien, Inflation Hedging Characteristics of the Chinese Real Estate Market (2004). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=538742

Deng, Yu and Qi, Wei, and Fu, Bojie, and Wang, Kevin, Geographical transformations of urban sprawl: Exploring the spatial heterogeneity across cities in China 1999-2015 (2015). https://www.sciencedirect.com/science/article/pii/S0264275119300307

Zhang, Weili and Chong, Zhaohui and Li, Xiaojian and Nie, Guibo, Spatial patterns and determinant factors of population flow networks in China: Analysis on Tencent Location Big Data (2019). https://www.sciencedirect.com/science/article/pii/S0264275119311862

Pan, Jinghu and Lai, Jianbo, Spatial pattern of population mobility across cities in China: Case study of the National Day plus Mid-Autumn Festival based on Tencent migration data. (2019). https://www.sciencedirect.com/science/article/pii/S0264275118311703?via%3Dihub

Flowminder, Mapping Indicators of Women's Welfare at High Spatial Resolution. https://web.flowminder.org/case-studies/mapping-indicators-of-womens-welfare-at-high-spatial-resolution