



Theses and Dissertations

2020-06-05

Identifying High-Potential Work Areas in Engineering for Global Development: Linking Industry Sectors to the Human Development Index

Daniel Oliver Smith
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Engineering Commons](#)

BYU ScholarsArchive Citation

Smith, Daniel Oliver, "Identifying High-Potential Work Areas in Engineering for Global Development: Linking Industry Sectors to the Human Development Index" (2020). *Theses and Dissertations*. 8447.
<https://scholarsarchive.byu.edu/etd/8447>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Identifying High-Potential Work Areas in Engineering for Global Development:
Linking Industry Sectors to the Human Development Index

Daniel Oliver Smith

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Science

Christopher A. Mattson, Chair
John L. Salmon
Spencer P. Magleby

Department of Mechanical Engineering
Brigham Young University

Copyright © 2020 Daniel Oliver Smith
All Rights Reserved

ABSTRACT

Identifying High-Potential Work Areas in Engineering for Global Development: Linking Industry Sectors to the Human Development Index

Daniel Oliver Smith
Department of Mechanical Engineering, BYU
Master of Science

Those working in Engineering for Global Development seek to improve the conditions in developing countries. A common metric for understanding the development state of a given country is the Human Development Index (HDI), which focuses on three dimensions: health, education, and income. An engineer's expertise does not always align with any of those dimensions directly, while they still hope to perform impactful work for human development. To discover other areas of expertise that are highly associated with the HDI, correlations and variable selection were performed between all World Development Indicators and the HDI. The resultant associations are presented according to industry sector for a straightforward connection to engineering expertise. The associated areas of expertise can be used during opportunity development as surrogates for focusing on the HDI dimensions themselves. The data analysis shows that work related to "Trade, Transportation, and Utilities", such as electricity distribution, and exports or imports, "Natural Resources and Mining", such as energy resources, agriculture or access to clean water, and "Manufacturing", in general, are most commonly associated with improvements in the HDI in developing countries. Also, because the associations were discovered at country-level, they direct where geographically particular areas of expertise have been historically associated with improving HDI.

Keywords: Engineering for Global Development (EGD), design for the developing world, human development index, opportunity development

ACKNOWLEDGMENTS

This research has been a collaboration of many individuals and I'd like to recognize a few who have contributed and encouraged me along the way.

First, thank you to my wife, Taryn, for her eternal encouragement, her listening ear, and her thoughtful feedback. Thank you for helping every aspect of my life be richer and more full of light.

I'd like to recognize the faith and investment of my graduate advisor, Dr. Mattson, in me. Thank you for allowing me a large amount of freedom in determining the focus of my research and my approach to the research, all the while offering candid and insightful suggestions. It has been a treat to be a pioneer in this research field with you, Dr. Mattson.

Thank you to my other graduate committee members, Drs. Salmon and Magleby. Your outside perspectives have been invaluable to considering how this research should translate to the engineering and academic community as a whole. Thank you for treating me as a colleague.

I would also like to recognize the efforts of my research lab mates who have listened to my challenges and offered very practical wisdom. Also, thank you for your many hours volunteered helping improve the credibility of this research by performing the categorization survey. And thanks for never taking me too seriously, but helping me laugh at difficult situations.

I would like to recognize the National Science Foundation for providing the Grants CMMI-1632740 and CMMI-1761505 that funded this research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Finally, thank you to my Heavenly Father for guiding me one step at a time to this point. I thank God for the physical support I have received during especially busy times. I recognize His hand in moments of spiritual enrichment while performing this research. I thank God for my love of learning and discovery that has deepened.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
NOMENCLATURE	viii
Chapter 1 Introduction	1
Chapter 2 Identifying high-potential work areas in Engineering for Global Development: Linking Industry Sectors to the Human Development Index	4
2.1 Introduction	4
2.1.1 Assumptions and limitations of research approach	7
2.2 Method	9
2.2.1 Acquire data	10
2.2.2 Rank-transform data, correlate WDI to HDI and filter	11
2.2.3 Perform stepwise variable selection on percent change of WDI and HDI	12
2.2.4 Serial correlation effects	14
2.2.5 Categorizing WDIs by NAICS Industry Sector	15
2.3 Results and discussion	17
2.3.1 Categorization agreement analysis	19
2.3.2 Areas of expertise associated with increasing HDI	19
2.3.3 Exploration of using historical associations to guide future work choices	23
2.4 Application scenarios	26
2.5 Conclusions and future work	29
Chapter 3 Discussion and Clarifications	34
3.1 Finding historical patterns between industry sectors and data of interest	34
3.1.1 Overview and general terms used	34
3.1.2 Acquire data	35
3.1.3 Keyword and correlation filter	37
3.1.4 Perform stepwise variable selection on percent change of WDI and HDI	39
3.1.5 Serial correlation effects	41
3.1.6 Categorizing WDIs by NAICS Industry Sector	42
3.1.7 Review	43
Chapter 4 Conclusions, Limitations, and Additional Future Work	44
REFERENCES	46
Appendix A Appendix	51
A.1 Human Development Index across the world	51
A.2 Keywords used to filter WDIs directly related to HDI dimensions	51

A.3	Categorization survey	53
A.4	Results table guide	53

LIST OF TABLES

2.1	Rubric for categorizing WDIs by NAICS Industry Sector.	15
2.2	Example of the WDIs selected for the multiple linear regression model during stepwise variable selection for Zambia. The WDIs from each country’s model were categorized according to industry sector, as shown in the right column, and the average for all developing countries summary is found in Table 2.4.	18
2.3	Intraclass correlation coefficient reliability chart.	19
2.4	Detail regarding areas of expertise that have been commonly related to increasing human development by reporting how frequently the WDI developing countries regressors were connected to NAICS Industry super-sectors, sectors, and sub-sectors. Notice that NAICS sector and sub-sector code is provided for reference.	20
2.5	The data analysis is summarized with how frequently the WDI developing countries regressors were connected to NAICS Industry super-sectors by region. If every WDI regressor in a particular region was connected to an industry super-sector, the super-sector would be connected 100% of the time. Regional patterns offer more detail in illustrating which areas of expertise have been historically associated with an increase in HDI in developing countries in different regions of the world. The regions were classified by the World Bank analytical group [1]. As a note, the number of developing countries in each region is included in parentheses next to the region’s name.	23
2.6	The data analysis from Section 2.2 was performed on data from 1990-2003 and then from 2004-2017 on 10 randomly selected countries. This table reports which industry sectors were associated with improvements in HDI for the first 14 years, and continued to be as commonly associated during the following 13 years, suggesting the likelihood that an EGD practitioner would have been led to work in an area that maintained its association with improving HDI.	24
2.7	Presented at country level, the more frequently a WDI regressor is related to an industry sector suggests that work in that industry sector has historically been a good predictor of an increase in HDI. Although there is a maximum of 8 WDI regressors per country (row), each WDI may be connected to none, one, or more than one industry sector, so the values in each row may not sum to 8.	30
A.1	Keywords used to filter WDIs that were directly related to HDI dimensions. Note that some words are repeated with a lower case first letter. This was because a case-sensitive function was used to find WDIs with these keywords.	52

LIST OF FIGURES

2.1	Opportunities with high-potential are at the intersection of an individual’s or group’s expertise, their passion, and what the market needs. In EGD, market needs can be broadly determined from HDI, but HDI is only focused on three areas: health, education, and income. Referencing HDI for market needs, opportunity is falsely limited to work in just health, education, and income.	6
2.2	This paper seeks to virtually expand the intersection of market needs and expertise by observing other areas of expertise commonly associated with improving HDI. Figure adapted from [2].	6
2.3	Summary of historical data analysis for discovering areas of expertise associated with improving HDI in each country.	9
3.1	Summary of historical data analysis for discovering areas of expertise associated with change in KMD in each country	35
3.2	Example of Kendall’s τ . Note that its calculation is not sensitive to exponential data or non-normal data, but simply captures how one dataset increases or decreases relative to the other dataset	39
A.1	Visualization of HDI across the world adapted from a figure by Alice Hunter [3]. Darker blue suggests more developed countries. Note that a high concentration of under-developed countries (light blue) are in the Sub-Saharan Africa region.	51
A.2	Screenshot of the Qualtrics survey used to categorize each WDI by NAICS Industry Sectors. If the WDI were “Moderately” or “Fundamentally” connected to any industry super-sectors shown on the left, the next page would allow the raters to categorize the WDI by industry sector and then sub-sector.	53
A.3	In general, the Industry Sector Breakdown Table (Table 2.4) is referenced first to determine <i>what</i> to work on and the Region and Country Tables (Tables 2.5 and 2.7) are most helpful in determining <i>where</i> , although these two questions are interdependent. Steps (A), (B), and (C) are not necessarily sequential. (A) After identifying an area of expertise that is commonly associated with increasing HDI across all developing countries, the Region Table is referenced to find if that area of expertise has been more commonly impactful in a particular region. (B) Understanding regionally where an area of expertise has been impactful is helpful, but the Country Table illustrates, with more focus, where that area of expertise has been most impactful. (C) When a country or region has already been identified as the desired place of work, the two right columns in the Industry Sector Breakdown Table give more detail as to which sectors and sub-sectors within the industry super-sectors have been associated with improving HDI.	54

NOMENCLATURE

δ_{I_t}	Percent change in the Human Development Index value from the past year to the current (or t -th) year
I_t	Human Development Index value for a given year
\bar{I}	Average Human Development Index value for available data between 1990-2017
δ_{w_t}	Percent change a World Development Indicator value from the past year to the current (or t -th) year
w_t	Value of one of the World Development Indicators for a given year
d	Durbin-Watson statistic
d_L	Durbin-Watson lower critical value
d_U	Durbin-Watson upper critical value
$[]_t$	$[]$ for a given year

CHAPTER 1. INTRODUCTION

Engineers are well versed in analyzing and predicting if a product will structurally withstand anticipated loads, or determining the torque needed from a motor to drive a mechanism. To some extent, engineers are even trained to anticipate the economic potential of a product. Unfortunately, engineers rarely have a basis for predicting or assessing how a product or project impacts the various facets of a potential user's day-to-day life [4]. Every product has some impact on its users, even if the engineer did not consider its potential impact during the product's design. That said, it is reasonable to assume that engineers seek to develop products that have a positive impact on people's lives. The research presented herein is a building block to better equip engineers to anticipate how their work impacts people.

Due to economic, healthcare, educational, and other dichotomies between developing and developed countries, a number of groups are working to improve circumstances in developing countries. Efforts to improve circumstances in developing countries include the United Nations Sustainable Development Goals [5], the proliferation of data by the World Bank [6], the proposal of Grand Challenges for Engineering from the National Academy of Engineering [7], and a myriad of other efforts from non-profit, non-governmental, academic, and religious organizations. Within the engineering community, the Engineering for Global Development (EGD) group has formed with the vision of improving the quality of life in developing countries [8,9].

EGD work is difficult for a number of reasons including discerning relevant user needs during opportunity development. Often, those involved in EGD are geographically removed from the location of their work, and thus are removed from the context of potential users. In addition, they are often required to work in different languages and cultures. Due to these challenges, EGD engineers can fall into the trap of assuming relevant user needs and implementing solutions that are discordant with the user's way of life. EGD engineers often struggle to effectively determine *what* to work on and *where* to implement their work. In terms of the product development process

proposed by Mattson and Sorensen, engineers in EGD are especially ill-prepared in the opportunity development phase [10].

After a rigorous study of 27 failure reports published by Engineers Without Borders Canada (EWB), Wood and Mattson determined that of these engineering projects in developing countries, 23.9% failed because of assuming the needs of the customers being served and another 28.9% failed due to lacking contextual knowledge needed for significant impact [11]. In short, over half of the reported failures were due to a lack of information about *what* their potential users' needs were and *where* a product or project would have the greatest impact, both of which are typically addressed in the opportunity development phase.

The purpose of this research is to aid engineers in the opportunity development phase by equipping them with information to determine and/or justify what to work on and where to work to positively impact human development. This objective aligns with the research Bryan Stringham and Chris Mabey have performed. Bryan's published research [12] demonstrates how sensor data can be collected remotely and continuously to provide engineers with frequent and timely user information. Chris Mabey's published research [13] provides a modern analysis of where geographically the greatest human needs and technology gaps are in addition to where there are the fewest political barriers to performing EGD work. This thesis reports patterns of how work in various areas of expertise have historically been associated with improvements in human development across developing countries. These three building blocks provide EGD practitioners with valuable information for systematically determining what to work and where to work for greatest positive impact.

Although not discussed in detail within this thesis, the first year of my graduate experience was working with Bryan Stringham on the research published in the *Journal of Mechanical Design* titled "Combining Direct and Indirect User Data for Calculating Social Impact Indicators of Products in Developing Countries" [12]. Performing this research left me with a few questions that directed the thesis work presented herein: (1) Are the "obvious" projects, such as helping people get clean water, actually helpful? (2) What other kinds of work can be helpful? (3) How can publicly available historical data be used to explore what is impactful?

Ultimately, I determined to use publicly available historical data and statistical tools to explore what areas of expertise have historically been associated with improvements in human devel-

opment. Chapter 2 of this thesis is the paper “Identifying High-potential Work Areas in Engineering for Global Development: Linking Industry Sectors to the Human Development Index” which has been accepted into the ASME 2020 International Design Engineering and Technology Conference and has been submitted for publication in the *Journal of Mechanical Design*. This paper details the results of a statistical analysis that shows which areas of expertise are most commonly associated with improvements in the Human Development Index metric in developing countries. For reference, a map of the Human Development Index scores in different countries across the world is provided in Figure A.1 in Section A.1. The results are presented at three levels of detail: (1) as an average of all developing countries, (2) by region, and (3) by country. In practice, understanding which areas of expertise have historically been associated with increasing human development helps EGD practitioners determine *what* types of products or projects have been impactful in various developing countries; when considered geographically, the associated areas of expertise also suggest *where* a type of product or project has been impactful; thus, the historical patterns can guide future choices of work.

In summary, this thesis informs engineers about how work in different areas of expertise has historically been related to improving human development in developing countries; thus, offering broad-scale perspective of how products have historically impacted people. The body of the thesis is intended to help EGD practitioners determine *what* to work on and identify *where* a project or product may have the highest-potential impact.

CHAPTER 2. IDENTIFYING HIGH-POTENTIAL WORK AREAS IN ENGINEERING FOR GLOBAL DEVELOPMENT: LINKING INDUSTRY SECTORS TO THE HUMAN DEVELOPMENT INDEX

2.1 Introduction

Although there has been significant improvement in general health, economic and other conditions throughout the world in the past few decades, there remains a striking difference in conditions between developed and developing countries. For example, the UN recently reported that 3 billion people lack clean cooking fuels and technology [5]. Manufacturing value added per capita in developing countries was 2.3% of that in Europe and Northern America in 2018 [5]. Regrettably, more people have access to cell phones than a functioning toilet [14]. These compelling circumstances and others have inspired the creation of the United Nations Development Goals, encouraged humanitarian efforts from non-governmental organizations and non-profit organizations, and rallied groups to causes such as Engineering for Global Development (EGD) [8, 15]. Although EGD applies to countries from low- to high-development, the focus of this paper is on low- and lower middle-income economies classified by the World Bank [16], which will be hereafter referred to as developing countries.

Individuals interested in EGD face unique challenges due to often being geographically removed from where their work is implemented, such as differences in language, culture, and context that can lead to incorrectly assumed user needs and lack of product adoption [11, 17]. At the same time, opportunity in EGD is replete and covers a number of disciplines, such as agriculture, sanitation, electricity, water, product design, manufacturing, among others [18–23]. Regarding EGD, two important and often difficult questions considered are *what* project or product should our organization work on? and *where* should we implement our work [24–26]?

Those working towards improving conditions in developing countries currently focus on a variety of areas, such as poverty alleviation, human development, gender equality, among oth-

ers [27–30]. The United Nations Development Programme (UNDP) has published measurements or indices in some of these areas, including the Human Development Index (HDI). The HDI summarizes the development state of a country using measures in three dimensions: “a long and healthy life, access to knowledge, and a decent standard of living” [31]. More succinctly, the HDI is quantified using data that captures the state of health, education, and income of individuals in a given country and its calculation is shown in Equation 2.1.

$$HDI = (I_{\text{Health}} \cdot I_{\text{Education}} \cdot I_{\text{Income}})^{1/3} \quad (2.1)$$

where I_{Health} represents the Health index, $I_{\text{Education}}$ the Education index, and I_{Income} the Income index. The indicators for each index are “Life expectancy (years)” for Health, “Expected years of schooling (years)” and “Mean years of schooling (years)” for Education, and “Gross national income per capita (2011 PPP \$)” for Income. Each dimension index is calculated according to Equation 2.2. Additional HDI details can be found in the UNDP Technical Notes for Calculating the HDI [31].

$$\text{Dimension index} = \frac{\text{actual value} - \text{minimum value}}{\text{maximum value} - \text{minimum value}} \quad (2.2)$$

The HDI has been referenced in a number of different disciplines and applications. For example, Ray et al. used the HDI to measure the impact of switching from kerosene lamps to solar lights [32]; Dufo-López et al. used the HDI as an optimization objective for off-grid electrical supply [33]; others have examined how electricity consumption [34, 35] and the sustainable water use index [36] are correlated with the HDI. The HDI provides a more well-rounded approach to quantifying the development of a country than solely focusing on gross national product growth [37]. This said, the HDI provides only a summary of human development in a given country, and as such, cannot communicate many of the details related to development [38].

While the HDI can provide a useful basis for what to work on and where to work, not every EGD practitioner has expertise in health, education, or income directly, yet they still hope to perform work that will improve human development. To put the research in this paper into perspective, consider the intersection of passion, expertise, and market needs presented in Figure 2.1. Focusing energy and resources on work that falls in the intersection of these three categories, Nucor Steel became a distinguished steel fabrication and Fortune 500 company [2]. Nucor’s Circles

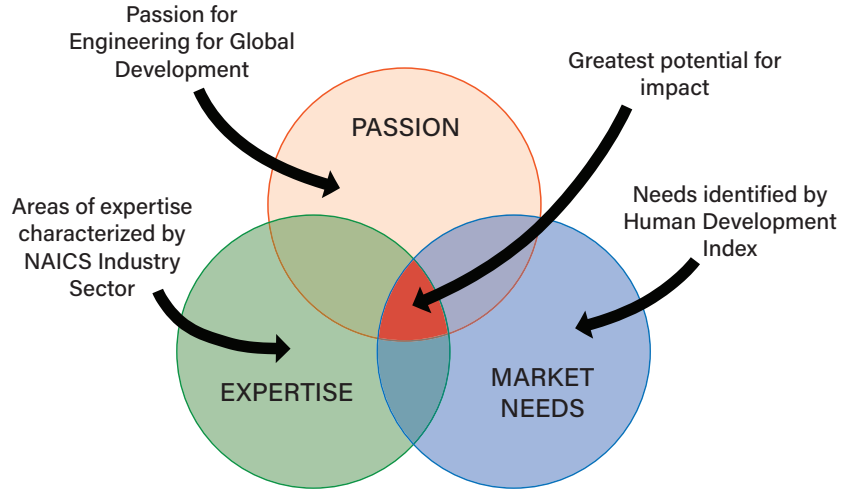


Figure 2.1: Opportunities with high-potential are at the intersection of an individual’s or group’s expertise, their passion, and what the market needs. In EGD, market needs can be broadly determined from HDI, but HDI is only focused on three areas: health, education, and income. Referencing HDI for market needs, opportunity is falsely limited to work in just health, education, and income.

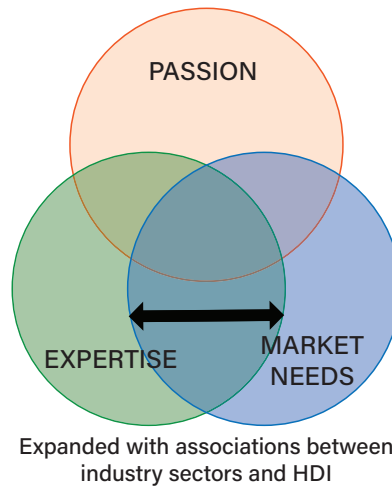


Figure 2.2: This paper seeks to virtually expand the intersection of market needs and expertise by observing other areas of expertise commonly associated with improving HDI. Figure adapted from [2].

is suggested by Mattson and Sorensen as a tool for thoughtfully converging on high-potential opportunities during the opportunity development stage of product development [10]. According to the Nucor’s Circles approach, high-potential opportunities fall in the intersection of an individual’s or group’s passion, expertise, and the needs of the market.

Under the assumption that EGD is the passion of the individual or group, two important considerations need to be made: what does the market need, and what are the individual's or group's expertise. This paper seeks to inform those considerations by (i) correlating HDI-related needs to all World Development Indicators (published by the World Bank), thus providing a clearer picture of what the market needs, and (ii) representing expertise by industry sector and linking them to the World Development Indicators (WDIs), thus helping to clarify engineering expertise with high potential for EGD impact. With a clearer picture of what the market needs and a clarification of engineering expertise with high potential for EGD impact, the intersection between market needs and expertise is virtually expanded, as shown in Figure 2.2.

In short, this research seeks to discover what areas of expertise—other than only health, education, and income—are commonly associated with improvements in HDI in developing countries.

We theorize that by examining historical correlations and performing variable selection between all WDI and the HDI for developing countries from 1990-2017, associated industry sectors will be found that can act as surrogates for focusing only on the three dimensions of the HDI (health, education, and income). Historical patterns from developing countries can help guide the choice of what to work on that has high-potential to improve human development in developing countries. For example, an engineer may seek to improve the Educational dimension of the HDI, but have no expertise related to educational products. Through the results provided in this paper, the engineer could recognize that work related to utilities, such as water distribution, is associated with improving HDI and infer that providing access to water can relieve water collection burdens children may face, allowing them to attend school more consistently.

2.1.1 Assumptions and limitations of research approach

In performing this worldwide data analysis, there were some inherent limitations in addition to some assumptions and approximations that were made to enable a tractable approach. The assumptions and limitations are:

1. **Research was performed with historical observational data.** Because these data are observational, and historical patterns do not always repeat in different areas of the world, the results presented herein cannot claim a causal relationship between WDIs and HDI, and re-

sults from one country cannot be inferred to another country. While historical patterns can be insightful, they will not flawlessly transfer to modern circumstances. The patterns discovered may carry forward for one year, five years, or more, but there is no basis for determining how long they may be applicable. Also, some patterns discovered in this data analysis may be coincidental; thus, the results are presented for all developing countries, by region, and by country, where patterns repeated regionally or worldwide may be more indicative of true relationships. Finally, historical patterns cannot fully represent current needs: current needs and where to work are discussed in more detail by Mabey et al. [13].

2. **Market needs are represented by HDI.** The HDI is an insightful metric that captures the general development state of a given country and thus can provide insight into market needs. Because it is reported at country level, and is a general measure, it does not capture the detailed needs of a community, family, or individual. Further investigation will need to be performed to validate more detailed and localized needs.
3. **World Development Indicators data are not comprehensive and are aggregated at a national level.** There are many advantages of using the WDI data published by the World Bank, such as the qualitative nature of many of the survey data, the yearly publications for many data for many countries, and the sheer quantity of publicly available raw data. It is important to recognize, however, that the breadth of WDI data is not comprehensive and thus cannot capture every aspect of what causes changes in HDI. More specifically, WDI data is not always engineering or industry specific, so the analysis and results presented herein focus on those WDIs that are apparently related to an industry sector.
4. **Limited number of terms were used in regression models.** A multiple linear regression model was fit for each country with a maximum of 8 terms. Too many terms can cause overfitting, but limiting the number of model terms could exclude some lesser, yet insightful, associations between WDIs and HDI. Therefore, the results focus on areas of expertise related to the 8 best fitting WDIs for each country; there are other less associated areas of expertise not captured by this analysis. A more detailed discussion of why 8 was chosen as the number of regressors is included in Section 2.2.3.

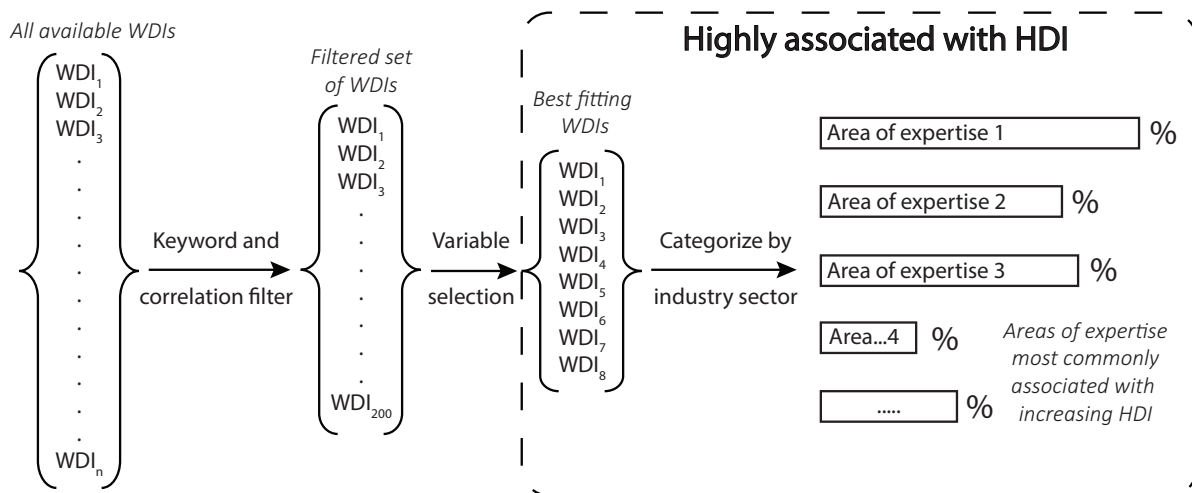


Figure 2.3: Summary of historical data analysis for discovering areas of expertise associated with improving HDI in each country.

The remainder of this paper is laid out as follows. In Section 2.2 the approach for discovering historical associations between WDIs and the HDI is detailed. Section 2.3 summarizes patterns of which industry sectors are related to increasing HDI in developing countries. Section 2.4 presents scenarios of how the results presented might be utilized by different individuals and groups. Finally, Section 2.5 discusses how the observed associations expand understanding of the intersection between expertise and market needs.

2.2 Method

The goal of this data analysis is to discover historical associations between WDIs and the HDI such that highly-associated WDIs point towards areas of expertise related to increasing HDI. The process followed to discover these associations was:

1. Acquire HDI and WDI data
2. Perform a keyword filter as well as a correlation filter on WDI data to reduce to a tractable dataset for variable selection
3. Perform variable selection on filtered and normalized WDI data to identify the WDIs that best fit the change in the HDI

4. Categorize the best fitting WDIs by NAICS Industry Sector for a more direct connection to areas of expertise

As illustrated in Figure 2.3, this approach offers the ability to filter through the many WDIs using correlation and variable selection to determine which WDIs best fit the change in the HDI in a given country [39]. When categorized by industry sector, the selected WDIs will suggest areas of expertise that have been commonly associated with increasing HDI in developing countries. Each step of this data analysis is discussed in detail below.

2.2.1 Acquire data

HDI data, published by the United Nations Development Programme, was available from 1990-2017 for 76 developing countries and was acquired for this study [31,40]. Additionally, the World Bank publishes WDI data that captures national development data in categories such as agriculture, education, gender, trade, population dynamics, etc. [6]. Typically collected through censuses and surveys, some examples of WDIs include “Electric power consumption (kWh per capita)”, “Manufacturing value added (current LCU)”, and “Rural population growth (annual %)”. Up to 1599 WDIs were available for the same 76 developing countries as HDI, and these data, also between 1990-2017, were acquired and included for this study.

Handling incomplete data

Many countries had incomplete HDI or WDI data between 1990-2017. For the HDI, it was typical that once the HDI began to be reported, it was reported for every year thereafter. For example, HDI data was not available from Ethiopia from 1990-1999, then was reported from 2000-2017. To avoid extrapolating into the past, the correlations and variable selection were only performed for years the HDI was reported.

For the WDIs, if only those WDIs that had complete data between 1990-2017 were used, this exclusion would reduce the number of WDIs in each country (depending on the country) from up to 1599 down to approximately 500, so the missing data was imputed. WDIs with missing data were often reported for a few years, and then failed to report for a few years. The strategy used to handle missing data was to impute the missing data by carrying the last observation forward [41].

For example, in El Salvador the “Government expenditure on education, total (% of GDP)” was reported in 1998-2000, 2002-2003, and 2005-2017. Because no data was reported until 1998, data was not extrapolated to 1990-1997; data from 2000 was carried forward into 2001, and data from 2003 was carried forward into 2004. Thus, percent change (discussed in detail in Section 2.2.3) calculated for 2000-2001 and 2003-2004 was zero. The change in a WDI could be positive or negative, so assuming there was no change can be conservative or nonconservative depending on the trend of the data.

2.2.2 Rank-transform data, correlate WDI to HDI and filter

As a first step in analyzing which WDIs are highly associated with HDI, a bivariate correlation that measures the strength and direction of association was calculated for each WDI and the HDI. Common correlation statistics, including the Pearson r correlation, assume that the data is normal [42]. After performing the Anderson-Darling test on a subset of WDI data, it was apparent that much of the data was not normally distributed [43]. Others, such as the Spearman and Kendall correlations, eliminate the importance of population distributions by accepting rank-transformed data [42, 44]. Therefore, the non-normal WDI and HDI data were transformed to ranks prior to calculating the correlation: the value of each WDI for each year was replaced by its rank relative to the values from other years of the same indicator [39].

For an initial screening of the many potentially associated WDIs, the WDI values from each country were correlated with the HDI values. The correlation was quantified using the Kendall’s τ correlation coefficient. Quantifying Kendall’s τ is a non-parametric test that calculates how many ranked values are concordant or discordant between two groups [44, 45]. For example, if the values of a WDI increase year-to-year over a given time period and the values of the HDI also increase year-to-year over the same time period (no matter if the relationship is linear or non-linear), Kendall’s τ will equal 1. Also, performed in MATLAB[®], the calculation of Kendall’s τ correlation coefficient included an adjustment for tied rank values [46]. Additional detail regarding the calculation and meaning of Kendall’s τ is provided in Section 3.1.3.

As mentioned in Section 2.2.1, the correlation coefficient was calculated only for years the HDI was reported in each country. Regarding the example of Ethiopia from Section 2.2.1, the HDI

was only reported from 2000-2017, so the HDI and WDI values were correlated for only those years in Ethiopia.

Filter the dataset

As the aim of this analysis was to expand the intersection between market needs and expertise (Figure 2.1), the WDIs that were directly related to the calculation of each HDI dimension—Health, Education, and Income—were filtered out of the dataset because they were assumed associated and would eliminate opportunities for associated indicators in other areas of expertise to be captured in each country’s linear regression model. For a complete list of the keywords used to filter the WDIs, see Table A.1 in Section A.2. Also, this was done because this research aimed to discover *other* areas of expertise that have been related to increasing HDI in developing countries. Finally, those filtered out were often colinear with the metrics used to calculate the HDI, and could thus adversely affect the regression model selected with stepwise variable selection [39].

After removing the directly related WDIs, the top 100 positively correlated WDIs and the top 100 negatively correlated WDIs were saved for variable selection. This reduced the likelihood of selecting coincidental associations during variable selection due to a large number of potential explanatory variables [47]. Also, it focused the variable selection on those WDIs that had strong correlations with the HDI (for most countries and WDIs, $|\text{Kendall's } \tau| > 0.6$), while maintaining a consistent number of WDIs for each country’s variable selection.

2.2.3 Perform stepwise variable selection on percent change of WDI and HDI

Following the initial filtering of WDIs according to Kendall’s τ , the WDIs that best fit change in the HDI in each country were identified using variable selection [39]. The values of the WDIs range in magnitude and vary in their configuration—some are percent values, others are raw counts, and some are ratios. Thus, to scale the WDIs in relation to the HDI, the percent change per year was calculated for each WDI and HDI specific to each country. Using the percent change also captured the direction of change in the indicator or index over time; thus allowing for simple interpretation.

Calculate percent change

The percent change in each HDI for each year from 1990-2017 was calculated as follows:

$$\delta_{I_t} = \frac{I_t - I_{t-1}}{\bar{I}} \quad (2.3)$$

where t represents the t -th year, δ_{I_t} represents the percent change in the index value from the past year to the current (or t -th) year, I_t represents the HDI value for the t -th year, and the average, \bar{I} , is calculated from available data between 1990-2017. This calculation was made for each country individually.

The calculation of percent change for each of the 1599 WDI values was of the same form as for the HDI data:

$$\delta_{w_t} = \frac{w_t - w_{t-1}}{\bar{w}} \quad (2.4)$$

where t represents the t -th year, δ_{w_t} represents the percent change in each WDI value from the past year to the current (or t -th) year, w_t represents the value of one of the WDIs for the t -th year, and the average, \bar{w} , is calculated from available data between 1990-2017, each country specific.

It is common to calculate percent change in relation to the past data point, but there were some WDI data whose change year to year was quite drastic. This led to influential outliers that biased the linear regression models that are not robust to outliers. Using the average of the data for each indicator helped attenuate the extremes in percent change per year.

Stepwise variable selection

The final step in discovering associations between WDIs and HDI was to perform a stepwise variable selection that created a multiple linear regression model for each country that denoted those WDIs that best fit change in the HDI. *To be clear, the purpose of the variable selection was to identify up to 8 WDIs in each country that have been highly associated with increasing HDI*, rather than to make quantitative predictions of the HDI given WDI input; thus, the results will not focus on the regression models (regression coefficients, p -values, etc.), rather the WDI terms included and their associated industry sectors. Given percent change per year (from Section 2.2.3) of the 200 most correlated WDIs (from Section 2.2.2), forward selection was performed using multiple

linear regression [48]. The criterion used for adding or excluding terms to each country's model—to find the best fitting WDIs—was the adjusted R-squared metric [49].

It was desirable to exclude interaction terms to increase the diversity of WDIs included in each country's regression model. When interaction terms are included, it is appropriate to also include each of the explanatory variables in the interaction as individual regressors [39]. With a limited number of regressors per model, interactions would produce repeated WDI regressors. To determine if excluding interactions would adversely affect the accuracy of which WDIs are most associated with increasing HDI, multiple linear regression models were created for each country with and without interaction terms, and the models without interaction terms performed equally as well as those with interaction terms. Therefore, the results presented herein are based on a variable selection of only first-order terms.

Regarding the number of regressors in the model, Peduzzi et al. suggest that a minimum of 10 years of data per regressor will help avoid overfitting and other issues, while Vittinghoff et al. and Austin and Steyerberg show that this may be too conservative and suggest a minimum of 2-3 years of data per regressor [50–52]. Specific to the analysis in this paper, each country reported the HDI and WDIs for up to 27 years, with an average of approximately 24 years for HDI and 20 years for WDIs. It was desirable to maximize the number of included WDI regressors to potentially increase their variety, while also avoiding overfitting so the models created would be informative of historical patterns that may be more applicable moving forward. In line with the goal of this research, the number of WDI regressors was maximized to 8 regressors for a liberal 2.5 years of data per regressor.

In summary, forward selection was performed (terms were added to the model) for the HDI in each country until either (1) the adjusted R-squared criteria stopped increasing, or (2) the model fit 8 WDI regressors.

2.2.4 Serial correlation effects

One potential issue when creating regression models using data that is reported year to year is serial correlation (also known as autocorrelation), where the residuals from year to year are correlated. The effect of serial correlations is that the regression coefficients will be imprecise, suggesting more or less statistical significance than is true [53, 54]. In other words, the p -values

Table 2.1: Rubric for categorizing WDIs by NAICS Industry Sector.

Strength of Connection	Description of the strength of connection
Fundamentally connected	The WDI is a direct measure of an aspect of the Industry Sector.
Moderately connected	The WDI is the result of work in the Industry Sector, is made possible by work in the Industry Sector, or offers motivation to work in the Industry Sector.
Insufficiently connected	The WDI pertains to measurements outside of the Industry Sector being considered or requires two or more logical steps to connect the WDI to the Industry Sector.
No specific connection	The WDI measures employment, profit, activities, goods, or other metrics of industry as a whole with no reference to a specific sector or sectors .
Unknown	Insufficient WDI detail to categorize by Industry Sector or researchers do not understand the WDI and thus cannot determine if it is connected to an Industry Sector.

for the regression coefficients will be estimated too high or too low and the confidence intervals will be too wide or too narrow.

A common test to discover the presence of serial correlations is the Durbin-Watson test that reports a statistic, d , between 0 and 4, where $d = 2$ suggests no serial correlation [55, 56]. Critical values, d_L and d_U , for the Durbin-Watson statistics tabulated by Savin and White report that for a model with 8 regressors and a sample size of 20, $d_L = 0.50$ and $d_U = 2.52$ [57]. Performing the Durbin-Watson test on the regression models created for each country showed that 0% of models predicted regression coefficients with exaggerated (p -value too low) statistical significance (when $d_L < 0.50$), and 18.3% of models predicted regression coefficients with underestimated (p -value too high) statistical significance (when $d_U > 2.52$). With the presence of some serial correlation effects, a Newey-West robust covariance matrix was used when fitting each country's regression that adjusted for or mitigated the serial correlations [58].

2.2.5 Categorizing WDIs by NAICS Industry Sector

The previous steps in this data analysis filtered through up to 1599 WDIs per country down to approximately 8 WDIs per country that best fit increasing HDI. With the goal of understanding which areas of expertise are commonly associated with increasing HDI, the WDIs selected for each country's model were categorized according the North American Industry Classification System

(NAICS) Industry Sectors [59] and the results are presented according to how frequently each industry sector was connected to a WDI that best fit the change in HDI.

To perform the categorizations, four researchers, working in teams of two, examined each remaining unique WDI to determine if there was a connection to one or more industry sectors. By working in teams of two, each researcher was forced to justify, to a teammate, the reasoning for connecting a WDI to an industry sector. Recognizing that we cannot fully eliminate research bias, a brief description of the individuals included as well as a motivation for including them in this part of the study are as follows. The four researchers included a professor and three graduate students all in mechanical engineering. All four researchers were males whose home country is the United States. The professor and one of the graduate students were chosen because of their 11 and 5 years, respectively, working with EGD groups and implementing projects in a variety of locations including Uganda, Tanzania, Rwanda, Peru, Brazil, Mongolia, India, Cambodia, among others. Their familiarity with EGD, various cultures, as well as exposure to WDI data provided relevant context to the WDI data that were categorized. The other graduate student and undergraduate student both became thoroughly familiar with the NAICS Industry Sectors by studying the descriptions of each super-sector, sector, and sub-sector and accompanying examples for over 8 hours each and their expertise guided the discussions of how each WDI was connected or not connected to industry sectors.

First, each researcher completed a 30 minute training to familiarize himself with the NAICS industry sectors. A Qualtrics online survey platform was used to rate each of the 329 unique WDIs according to its connection to industry sectors. An screenshot of the survey format is provided in Figure A.2 in Section A.3. The order in which the WDIs were presented to each team was randomized to reduce bias. The teams first rated each WDI's connected-ness to the NAICS industry super-sectors, then by industry sectors (if there was a "Moderate" or "Fundamental" super-sector connection), and lastly by industry sub-sectors (if there was a sector connection). A detailed description of each WDI was available during the survey for the teams to refer to as needed.

The level of connected-ness was determined according to the definitions in Table 2.1, similar to a Likert scale with no neutral option [60]. The results presented in this paper focus on those WDIs that are "Moderately" or "Fundamentally connected" to at least one industry sector. The "No specific connection" option was included because there are a number of WDIs that reflect a

measure of industry as a whole while not referring to any specific sector of industry (see Table 2.2 for examples).

In general, the WDIs are metrics in a number of areas, such as economics, climate change, gender, health, poverty, among others; thus, how they connect to industry sectors is not always apparent and it was expected that the teams would differ in some of the categorizations. After both teams categorized all the unique WDIs, they met to discuss disagreements within teams and discrepancies between the teams' categorizations. The teams presented their reasoning for establishing the connection or no connection to the industry sector, after which each researcher individually and anonymously voted in agreement or disagreement with the industry sector connections of each discrepancy. Those that received a majority vote are included in the results.

2.3 Results and discussion

The analysis detailed in Section 2.2 produced a regression model for each developing country with up to 8 WDIs that best fit the change in HDI data for each country according to the adjusted R-squared metric. The results of this paper focus on areas of expertise connected to the WDIs that best fit increasing HDI; thus, the WDIs selected for each country's regression model were categorized by NAICS Industry Sector. The results of the categorization and agreement between researchers is included in Section 2.3.1. Table 2.2 provides an example of the WDIs that best fit the change in the HDI in Zambia and their corresponding industry connections. The culmination of this research is discussed in Section 2.3.2 that primarily presents the frequency of industry sector connections to increasing HDI in developing countries. Resultant associations between industry sector and HDI are summarized across all developing countries in Table 2.4, by region in Table 2.5, and by country in Table 2.7. Because of the EGD motivation for this research, these results focus on areas of expertise commonly related to increasing HDI in developing countries. Added understanding of which industry sectors have historically been related to increasing HDI virtually expands the intersection between areas of expertise and market needs, as illustrated in Figure 2.2.

Table 2.2: Example of the WDIs selected for the multiple linear regression model during stepwise variable selection for Zambia. The WDIs from each country's model were categorized according to industry sector, as shown in the right column, and the average for all developing countries summary is found in Table 2.4.

Zambia		
World Development Indicator	Industry Super-Sector Connection(s)	Industry Sector and Sub-Sector Connections(s)
Demand for family planning satisfied by modern methods	Trade, Transportation, and Utilities; Education and Health Services; Information	Health and Personal Care Stores (NAICS 446); Health Care and Social Assistance (NAICS 62); Merchant Wholesalers, Nondurable Goods (NAICS 424); Publishing Industries (except Internet) (NAICS 511); Broadcasting (except Internet) (NAICS 515); Internet Publishing and Broadcasting (NAICS 516)
Access to clean fuels and technologies for cooking (% of population)	Natural Resources and Mining	Oil and Gas Extraction (NAICS 211); Utilities (NAICS 22)
GDP per unit of energy use	Trade, Transportation, and Utilities	Utilities (NAICS 22)
Information and communication technology service exports	Trade, Transportation, and Utilities; Information; Professional and Business Services	Information (NAICS 51); Professional, Scientific, and Technical Services (NAICS 54); Postal Service (NAICS 491); Couriers and Messengers (NAICS 492)
Population in urban agglomerations of more than 1 million	Construction; Trade, Transportation, and Utilities	Utilities (NAICS 22); Construction of Buildings (NAICS 236); Heavy and Civil Engineering Construction (NAICS 236)
Women who believe a husband is justified in beating his wife (any of five reasons)(%)	Information; Education and Health Services	Social Assistance (NAICS 624); Publishing Industries (except Internet) (NAICS 511); Broadcasting (except Internet) (NAICS 515); Internet Publishing and Broadcasting (NAICS 516)
Imports of goods and services	Trade, Transportation, and Utilities	No specific sector/sub-sector connection
Goods and services expense	No specific super-sector connection	No specific sector/sub-sector connection

Table 2.3: Intraclass correlation coefficient reliability chart.

ICC value	Reliability
0-0.5	Poor
0.5-0.75	Moderate
0.75-0.9	Good
0.9-1	Excellent

2.3.1 Categorization agreement analysis

The categorization results presented below were calculated after the two teams discussed discrepancies between teams and voted to decide which of the discrepant WDIs were or were not connected to industry sectors. Out of the 579 total WDIs selected for regression models in 76 developing countries, it was determined that 52 WDIs were insufficiently connected to industry, 68 had no specific industry sector connection, and the remaining 459 had a moderate to fundamental connection to at least one industry sector, and thereby an area of expertise. The majority of the categorized WDIs were connected to multiple industry sectors.

To assess the reliability of these categorizations, an intraclass correlation coefficient (ICC) was calculated that quantifies the agreement of responses between the two teams of researchers [61]. The ICC was calculated according to super-sector and sector connections, where each data point was represented by a 0 if the WDI was “Insufficiently connected”, had “No specific connection”, or was “Unknown” and by a 1 if the WDI was “Moderately” or “Fundamentally connected”. Calculated with the Pingouin statistical package in Python [62], the two-way mixed-effects ICC was found to be 0.91 with a significance value <0.0001 and a 95% confidence interval from 0.90 to 0.91 indicating an excellent level of agreement between the respondents [63]. The scale of ICC and the reliability of the results according to Koo et al. is found in Table 2.3 [63]. Because of the excellent level of agreement between teams, we are confident that the categorizations were performed thoroughly and reliably.

2.3.2 Areas of expertise associated with increasing HDI

In practice, the results presented in Tables 2.4, 2.5, and 2.7 could be used in two ways:

Table 2.4: Detail regarding areas of expertise that have been commonly related to increasing human development by reporting how frequently the WDI developing countries regressors were connected to NAICS Industry super-sectors, sectors, and sub-sectors. Notice that NAICS sector and sub-sector code is provided for reference.

NAICS Industry Super-sector	Frequency of Industry Sector Connection to Increasing HDI in Developing Countries (%)	NAICS Industry Sector or Sub-sectors	Frequency of Industry Sector Connection to Increasing HDI in Developing Countries (%)
Trade, Transportation and Utilities	36.4	Utilities (NAICS 22) Wholesale Trade (NAICS 42) Transportation and Warehousing (NAICS 48-49) Retail Trade (NAICS 44-45)	15.5 8.6 7.1 5.2
Education and Health Services	26.4	Health Care and Social Assistance (NAICS 62) Social Assistance (NAICS 624) Ambulatory Health Care Services (NAICS 621) Educational Services (NAICS 61)	20.0 13.3 8.8 6.0
Natural Resources and Mining	19.3	Agriculture, Forestry, Fishing and Hunting (NAICS 11) Mining, Quarrying, and Oil and Gas Extraction (NAICS 21) Crop Production (NAICS 111) Forestry and Logging (NAICS 113) Animal Production (NAICS 112)	8.3 3.8 2.9 3.1 1.2
Financial Activities	18.3	Finance and Insurance (NAICS 52) Monetary Authorities - Central Bank (NAICS 521)	16.4 11.4
Manufacturing	16.8	Manufacturing (no specific sub-sector)(NAICS 31-33) Computer and Electronic Product Manufacturing (NAICS 334) Chemical Manufacturing (NAICS 325) Machinery Manufacturing (NAICS 333)	14.2 1.9 1.7 1.2
Construction	15.7	Construction of Buildings (NAICS 236) Heavy and Civil Engineering Construction (NAICS 237)	8.5 6.2
Information	12.6	Broadcasting (except Internet) (NAICS 515) Internet Publishing and Broadcasting (NAICS 516) Publishing Industries (except Internet) (NAICS 511) Telecommunications (NAICS 517)	8.6 9.3 8.1 1.9
Professional and Business Services	9.5	Professional, Scientific, and Technical Services (NAICS 54) Management of Companies and Enterprises (NAICS 55) Administrative and Support and Waste Management and Remediation Services (NAICS 56)	5.7 1.0 0.9
Other Services (except Public Administration)	3.5	Repair and Maintenance (NAICS 811) Grantmaking, Civic, Professional, and Similar Organizations (NAICS 813)	1.2 1.0
Leisure and Hospitality	3.3	Arts, Entertainment, and Recreation (NAICS 71) Accommodation and Food Services (NAICS 72)	1.4 1.4

1. To determine *what* to work on by referencing areas of expertise strongly associated with improving HDI, or
2. To determine *where* geographically to implement a product or project for highest-potential impact.

and a general approach to utilizing the tables is illustrated in Figure A.3 in Section A.4. The areas of expertise presented are those associated with increasing HDI, on average, between 1990-

2017. To be consistent with the previous statement, adjustments for South Sudan and the Syrian Arab Republic were made. South Sudan has only been reporting HDI data since 2010 and its HDI has only improved twice since that time; thus there was not sufficient data to fit a model that relates WDI to increasing HDI. HDI in the Syrian Arab Republic decreased from 2010-2015, so the results for the Syrian Arab Republic are from an analysis between 1990-2009 where, on average, HDI increased year-to-year.

Regarding what to work on (point (1) above), engineers commonly begin the design process by developing a clear understanding of market and engineering requirements for a project or product; in other words, working through opportunity development [10]. While an individual or group is brainstorming potential opportunities, under the assumption that EGD is the passion of the individual or group, a high-potential opportunity may be identified by examining Table 2.4 and noting those areas of expertise that have been closely related to an increase in HDI.

The left two columns of Table 2.4 show how frequently an industry super-sector has historically been associated with increasing HDI across all developing countries. For example, out of all the WDIs selected for each of the developing countries' regression models, the WDI was connected to "Trade, Transportation, and Utilities" an impressive 36.4% of the time. Considering that 8 WDIs were selected for each country, 36.4% means that, on average, approximately 3 WDIs in each country's model were related to "Trade, Transportation, and Utilities". The right two columns offer more detail: of the industry sub-sectors related to "Trade, Transportation, and Utilities", an indicator related to "Utilities" has been connected to 15.5% of the WDIs, more than nearly any other industry sector or sub-sector. In short, those industry sectors with a higher relative % connection to WDIs have been more commonly associated with an increase in HDI and may prove to be high-potential areas for improving human development.

For reference, any value above 13% means that, on average, at least one WDI per developing country was connected to that industry sector, suggesting at least a moderate connection between that industry sector and increasing HDI in developing countries. That said, a percentage lower than 13% does not mean that area of expertise is inconsequential; rather, that area of expertise may not be impactful across *all* developing countries, but may be impactful in specific countries or regions, so Tables 2.5 and 2.7 should be referenced for additional detail.

It is insightful that in developing countries, industry sectors and sub-sectors, or areas of expertise, that are most commonly associated with increasing HDI include:

- “Trade, Transportation, and Utilities”, such as merchandise exports and imports, electricity distribution, access to clean fuels and technologies for cooking, and basic drinking water and sanitation services
- “Health Care and Social Assistance”, such as access to contraceptives, immunizations, and social assistance programs such as Social Security in the US
- “Natural Resources and Mining”, such as energy resources, agriculture or access to clean water
- “Financial Activities”, especially influencing monetary policies such as distribution of aid, international trade agreements and tariff rates
- “Manufacturing”, in general
- “Construction”, such as residential building construction and utility systems construction

Additional comments related to the application of these results are found in Section 2.4.

Regarding where to work (point (2) above), a group may have specialized expertise and be interested in where geographically their expertise would have highest-potential impact. The results relating areas of expertise to increasing HDI are presented by region in Table 2.5 and by country in Table 2.7. The regional summary results in Table 2.5 show how frequently an industry sector has historically been associated with increasing HDI in a particular region. Industry sectors with higher % appearance in a particular region may prove to be high-potential areas of expertise for improving human development in that region. It is worth noting that while one area of expertise may not have been historically related to increasing HDI in one region, it may have proven otherwise in another region. For example, while work in Information has only been moderately related to increasing HDI across all developing countries (12.6%), it has been commonly associated with increasing HDI in Europe and Central Asia (25.0%). As a note, the majority of developing countries included in these results are from the Sub-Saharan Africa region, and thus the worldwide results presented in Table 2.4 are quite similar to those included in the regional results for Sub-Saharan Africa.

Table 2.5: The data analysis is summarized with how frequently the WDI developing countries regressors were connected to NAICS Industry super-sectors by region. If every WDI regressor in a particular region was connected to an industry super-sector, the super-sector would be connected 100% of the time. Regional patterns offer more detail in illustrating which areas of expertise have been historically associated with an increase in HDI in developing countries in different regions of the world. The regions were classified by the World Bank analytical group [1]. As a note, the number of developing countries in each region is included in parentheses next to the region's name.

NAICS Industry Super-Sector	Frequency of Industry Sector Connection to Increasing HDI in Developing Countries by Region (%)						
	East and Pacific (12)	Asia (6)	Europe and Central Asia (2)	Latin America and Caribbean (7)	Middle East and North Africa (6)	South Asia (6)	Sub-Saharan Africa (43)
Trade, Transportation, and Utilities	38.5		25.0	39.6	37.8	39.1	35.4
Education and Health Services	26.4		31.3	20.8	4.4	17.4	31.4
Natural Resources and Mining	17.6		0.0	20.8	22.2	21.7	19.8
Financial Activities	23.1		12.5	13.2	26.7	26.1	15.9
Manufacturing	18.7		0.0	22.6	13.3	17.4	16.5
Construction	15.4		18.8	18.9	15.6	17.4	14.9
Professional and Business Services	4.4		0.0	13.2	8.9	6.5	11.3
Information	11.0		25.0	7.5	15.6	10.9	13.1
Leisure and Hospitality	3.3		6.3	1.9	4.4	6.5	2.7
Other Services (except Public Administration)	4.4		0.0	1.9	4.4	4.3	3.4

More detail can be found in Table 2.7 at the end of this chapter that captures how many times an industry sector was connected to WDIs in each country's regression model. Table 2.7 can be used as a step beyond the regional summary to guide an individual or group with particular expertise toward more specific geographical options for implementing their work. For example, while work in "Construction" has been moderately related to increasing HDI in East Asia and Pacific, we can see from Table 2.7 that of the 8 WDIs selected in Indonesia, 4 are connected to "Construction"—more than any of the other countries in the East Asia and Pacific region; thus, this can provide some justification for implementing a project or product in Indonesia that may improve capabilities or increase work in residential building construction or construction of utility systems or other facet of the "Construction" sector. Again, additional insights for how these results might be utilized by various individuals and groups are found in Section 2.4.

2.3.3 Exploration of using historical associations to guide future work choices

This research presents patterns describing areas of expertise that have historically been associated with increasing HDI in developing countries. We are suggesting that these patterns may

Table 2.6: The data analysis from Section 2.2 was performed on data from 1990-2003 and then from 2004-2017 on 10 randomly selected countries. This table reports which industry sectors were associated with improvements in HDI for the first 14 years, and continued to be as commonly associated during the following 13 years, suggesting the likelihood that an EGD practitioner would have been led to work in an area that maintained its association with improving HDI.

Country	Maintained associations, M_n , from 1990-2003 to 2004-2017 (%)
Combined 10 Country Sample	81.6
Bangladesh	33.3
Bolivia	66.7
Burundi	33.3
Ghana	75.0
Guinea	33.3
Kyrgyzstan	50.0
Mauritania	33.3
Morocco	100
Niger	11.1
Rwanda	71.4

give direction to an individual or organization working in EGD to determine and/or justify their future choice of work. We cannot extrapolate and claim that these associations will continue into the future, but the exploration in this section can offer confidence for using the historical patterns from this research to guide future work.

To explore the use of historical patterns to guide future work, the following was performed. The analysis from Section 2.2 was followed for 10 randomly selected developing countries during the years 1990-2003 that had at least 10 years of HDI data that, on average, increased between 1990-2003. The best fitting WDI regressors were again categorized by industry sector to determine those most commonly associated with improving HDI. Then, the analysis from Section 2.2 was followed for the same 10 countries between 2004 and 2017. 4 regressors were used in each country's model, to stay consistent with ratio of 2.5 years per regressor (see Section 2.2.3). Also, splitting the dataset at year 2003 was chosen to allow both the first and second set of years to fit the same number of regressors while following the said ratio of 2.5 years of data per regressor. The associated industry sectors from the second time period (2004-2017) were compared to the associated industry sectors from the first time period (1990-2003) to see how well the results of the

first time period would have informed work in the second time period. This comparison is shown in Table 2.6.

Essentially, if we had performed this same study for the first time period, and followed its results to determine what to work on, would that work have been associated with improving HDI for the second time period? The following was used to calculate this comparison for n included countries where $n = 10$ for the first row and $n = 1$ for all other rows in Table 2.6.

$$M_n = 100 * \left(\frac{\sum_{i=1}^{10} \begin{cases} \frac{C_{2i,n}}{C_{1i,n}}, & \text{if } C_{2i,n} - C_{1i,n} < 0 \\ 1, & \text{otherwise} \end{cases}}{10} \right)$$

where M_n represents the percent maintained associations for all industry super-sectors for the n included country or countries, $C_{1i,n} = \sum_{j=1}^n C_{1i,j}$ and represents the number of connections for the i -th industry super-sector for n country or countries for the first time period. So, for one country, $C_{1i,1}$, only includes the i -th industry super-sector connections for that country during the first time period. $C_{2i,n}$ for the second time period. In other words, this calculation identifies when the number of industry connections that were present during the first time period decreased during the second time period and finds the percentage of connections still present during the second time period as an average over the 10 industry super-sectors.

Table 2.6 shows that for 10 randomly selected developing countries, 81.7% of the associated industry sectors discovered from the first time period were as commonly associated for the following time period. Conversely, if an EGD practitioner were to have used these results after the first time period, there would have been an 18.3% likelihood they were misled to work in an area of expertise that would not have been as commonly associated with improving HDI for the second time period. In general, this exploration indicates that using historical patterns to guide future work is a meaningful approach to finding work that is associated with improvements in HDI.

Also shown in Table 2.6, out of the 10 randomly selected countries, 5 of the countries had half or more of their industry sector associations maintained from the first to the second time period. The other 5 countries had few maintained industry sector associations, which could be due

to a number of reasons. For example, Burundi was ravaged by civil war between 1993-2003 and then made some remarkable economic progress in addition to infrastructure improvements between 2003-2010 [64]; thus, it is logical that industry sectors associated with HDI improvements for the a time of war (the first time period) were quite different than those associated during a time of relative peace (the second time period).

This exploration shows that historical patterns may persist more commonly across all developing countries, but not necessarily within each country individually. Therefore, current and localized circumstances should also be considered when narrowing the work focus to a particular country or community. Also, because only 4 regressors were used for this exploration, it is expected that the historical associations presented as the results of this paper, with additional data and 8 regressors used in the analysis, are even more likely to maintain their associations into the future.

One limitation of this exploration is that the percentage of maintained associations is a function of the number of countries included in its calculation. If the “Combined Country Sample” were calculated with 3 countries or 15 countries we would expect a different value for maintained associations. That said, it is expected that as n increases, maintained associations will become asymptotic.

2.4 Application scenarios

To illustrate the application of results presented, scenarios are provided for individual as well as business cases. While these are not comprehensive of how these results might be utilized, they offer ideas of how the results might be applied by individuals and businesses working in EGD.

EWB Student Individual A is a member of a student chapter of Engineers Without Borders (EWB). His chapter has connections with locals in Cambodia, Myanmar, and a few other countries in East Asia. The EWB student chapter is in the early stages of project planning and they are trying to determine what project will have the most potential to improve the development state of the community where their project is implemented. In addition to interacting with the chapter’s contacts in East Asia, Individual A uses the results presented in this paper to inform their chapter’s project planning.

He first references Table 2.5 and finds that work related to “Trade, Transportation, and Utilities”, “Education and Health Services”, and “Financial Activities” have been strongly related to an increase in HDI over the past few decades in East Asia and Pacific. Then, referencing Table 2.7 he finds that increasing HDI in Cambodia has been mostly related to “Education and Health Services” and “Manufacturing”. Also, he notices Myanmar’s regression model has 4 WDIs related to “Trade, Transportation, and Utilities” and 3 WDIs related to “Natural Resources and Mining”, but none related to “Financial Activities”. Finally, he considers the breakdown of industry sectors into sub-sectors in the two right columns of Table 2.4 for greater detail about what areas of expertise have been related to increasing HDI. Utilizing this information, in addition to the expertise of the group, the EWB students can now focus their exploration of project opportunities on those related to trade, utilities, manufacturing, agriculture or mining that have historically been related to increasing HDI in Cambodia, Myanmar, and East Asia overall.

EGD Professional Individual B is a manager of a team whose professional work is in EGD. Her company has recently received funding for the next 5 years to aid refugees in Pakistan who have fled conflict in Afghanistan. With the new focus and funding, Individual B seeks to establish a team that possesses the expertise that will best serve to improve the development state in Pakistan. She first needs to identify the areas of expertise most in demand. In addition to working with local officials and visiting refugee camps to identify immediate needs, Individual B examines the results presented in this paper to identify long-term patterns of which areas of expertise have been related to increasing HDI in Pakistan and the surrounding area.

She first references Table 2.7 and finds that 4 WDIs related to increasing HDI in Pakistan are connected to “Natural Resources and Mining”. Examining Table 2.4 exhibits that sub-sectors of “Natural Resources and Mining” include work in agriculture, forestry, mining, oil and gas extraction, among others. For a perspective of impactful work in the surrounding region, she references Table 2.5 and is encouraged to find that work in the “Natural Resources and Mining” sector has been commonly related to an increase in HDI across the South Asia region when compared to the other industry sectors. Individual B can seek additional team members with skills related to agriculture, mining, etc. with an understanding that work in these areas of expertise has been related to increasing HDI in Pakistan and the South Asia region.

Limited Time EWB Volunteer Individual C is a practicing engineer who is seeking fulfillment and to share his expertise through volunteering with EWB. His background has been in designing the latest hardware for cellular devices, and is not certain how his skills might apply in a developing world setting. Noting from Table 2.4 that manufacturing work specific to the “Computer and Electronic Manufacturing” industry is related to an increase in HDI in few developing countries, he is motivated to find where it has been impactful and to find a EWB chapter that works in that region. Table 2.5 suggests that work related to the “Manufacturing” sector has been related to increasing HDI in multiple regions, so Individual C takes a closer look by examining Table 2.7. He finds that in Bhutan and El Salvador 4 out of 8 selected WDIs associated with increasing HDI fall in the “Manufacturing” industry sector; thus, as he seeks an EWB chapter with whom to work, he will first seek any who might have projects in Bhutan or El Salvador where he can help improve manufacturing capabilities or opportunities, or refer back to Table 2.5 for secondary options in other regions.

Large Businesses Company A is a large company with an established product line. Company A is seeking to expand its manufacturing facilities to other locations throughout the world, and is motivated to utilize inexpensive labor in addition to improving the quality of life of underserved communities. From the summary in left two columns of Table 2.4, Company A is encouraged that improving manufacturing has been commonly related to an increase in HDI. As they consider potential locations, starting regionally, Table 2.5 shows that work in the “Manufacturing” sector has been most closely related to increasing HDI in the Latin America region, second in the East Asia and Pacific region and Caribbean following. Company A can use this information, coupled with other factors, to focus its location decision. If multiple options within these regions are reasonable, Table 2.7 offers more detail regarding which countries have had multiple WDIs—suggesting a closer association—in the “Manufacturing” sector related to an increase in HDI.

Small Businesses Company B is a startup that leads humanitarian trips to build schools and orphanages in Sub-Saharan Africa. Company B is seeking to expand its team and type of projects offered; specifically, improving the communities’ access to electricity. The investors of Company B are asking for justification for this choice of work and want to know what will be most impactful in their target communities. The results presented in this paper offer some justification.

First, Company B references the right two columns in Table 2.4 that suggest that on average for developing countries, work related to “Utilities” has been commonly related to an increase in HDI. A quick reference to the NAICS Industry Sectors online page shows that “Electric Power Generation, Transmission and Distribution” is one of the three main Utilities included in the Utilities sector. Also, of the industry sectors related to an increase in HDI in Sub-Saharan Africa, work in “Trade, Transportation, and Utilities” is the most common (see Table 2.5). With these data, in addition to an analysis of current local needs, Company B can better justify to its investors its choice to expand its work to providing access to electricity.

2.5 Conclusions and future work

From an examination of historical correlations and using variable selection to identify those WDIs, and their corresponding industry sectors, that have been highly associated with increasing HDI in developing countries, areas of expertise are identified that have been related to increasing HDI. The added understanding of which areas of expertise are commonly associated with increasing HDI virtually expands the intersection of market needs and expertise, as shown in Figure 2.2. When exploring EGD opportunities, areas of expertise with strong historical associations with increasing HDI may be used as surrogates for focusing on the main categories of the HDI (health, education, and income) only. The results presented can aid EGD practitioners in identifying *what* to work on and *where* to work for high-potential impact. Details regarding what to work on is informed by patterns in developing countries summarized in Table 2.4, while where to work can be identified by referencing Tables 2.5 and 2.7 to examine regional and country specific patterns.

Due to the EGD motivation for this research, the results presented herein focus on associations between areas of expertise and HDI in developing countries. Rather than focusing on the three dimensions of the HDI only (health, education, and income), the data analysis shows that work related to “Trade, Transportation, and Utilities”, such as exports and imports of goods and services, electricity production and distribution, and basic sanitation services, “Natural Resources and Mining”, such as energy resources, agriculture or access to clean water, “Manufacturing”, in general, “Construction”, such as residential building construction as well as utilities construction, and “Professional and Business Services”, such as the engineering of communications technologies are commonly associated with improvements in the HDI in developing countries.

It is interesting to note that manufacturing work has been 1.3% *more* commonly associated with increasing HDI than work in “Utilities”. That said, it is encouraging that work in “Utilities” has been commonly associated with increasing HDI, especially for groups involved in electricity production and distribution such as providing micro- and off-grid electricity solutions. Also, work in “Financial Activities” especially work that influences monetary policies, including trade agreements and aid money decisions, has been commonly associated with increasing HDI. Finally, although improvements in agriculture have been commonly associated with human development improvements, this analysis shows that work related to improving “Trade, Transportation, and Utilities” is nearly 1.5 times more commonly associated with increasing HDI than work in “Natural Resources and Mining”.

This study determined WDI that have been associated with the HDI. It was not analyzed, however, which of these changed first and affected the other. For example, consider a country whose manufacturing capability is associated with increasing HDI: does the health, education, or income increase first, which in turn causes the manufacturing capability to increase, or visa versa? An investigative study into these interactions could be insightful future work to further inform practitioners regarding high-potential skill areas and where to implement their work.

Table 2.7: Presented at country level, the more frequently a WDI regressor is related to an industry sector suggests that work in that industry sector has historically been a good predictor of an increase in HDI. Although there is a maximum of 8 WDI regressors per country (row), each WDI may be connected to none, one, or more than one industry sector, so the values in each row may not sum to 8.

	Trade, Transportation, and Utilities	Education and Health Services	Natural Resources and Mining	Financial Activities	Manufacturing	Construction	Professional and Business Services	Information	Leisure and Hospitality	Other Services
Country or Territory	Number of Regression Model Indicators Related to Industry Sector									
East Asia and Pacific Region										
Cambodia	1	4	1	1	3	1	1	2	0	1
Indonesia	6	2	2	1	2	4	0	0	0	0

Continued on next page

Table 2.7 – continued from previous page

	Trade, Transportation, and Utilities	Education and Health Services	Natural Resources and Mining	Financial Activities	Manufacturing	Construction	Professional and Business Services	Information	Leisure and Hospitality	Other Services
Country or Territory	Number of Regression Model Indicators Related to Industry Sector									
Kiribati	5	4	0	2	1	1	1	1	1	1
Lao People's Dem. Republic	1	3	0	1	1	0	0	3	0	0
Micronesia	2	1	2	2	2	2	0	1	0	0
Myanmar	4	2	3	0	2	2	1	0	0	0
Papua New Guinea	4	0	2	2	2	0	0	0	0	0
Philippines	3	2	3	3	0	1	1	1	1	0
Solomon Islands	3	1	0	3	1	0	0	0	0	0
Timor-Leste	4	2	1	2	2	1	0	1	1	0
Vanuatu	2	3	0	2	1	0	0	0	0	0
Viet Nam	0	0	2	2	0	1	0	0	0	2
Europe and Central Asia Region										
Kyrgyzstan	3	3	0	1	0	2	0	2	1	0
Tajikistan	1	2	0	1	0	1	0	2	0	0
Latin America and Caribbean Region										
Bolivia	4	3	2	1	3	1	2	1	0	1
El Salvador	3	1	1	0	4	2	0	1	1	0
Guatemala	4	1	0	2	0	2	1	0	0	0
Guyana	2	1	2	0	0	0	0	0	0	0
Haiti	2	1	2	1	1	1	1	1	0	0
Honduras	4	2	3	3	3	3	2	1	0	0
Nicaragua	2	2	1	0	1	1	1	0	0	0
Middle East and North Africa Region										
Djibouti	1	1	0	2	1	2	0	2	0	0
Egypt	5	0	2	2	1	1	1	1	1	1
Iraq	3	0	5	1	2	1	0	1	0	0
Morocco	3	0	1	0	0	1	2	1	0	0
Syrian Arab Republic	3	0	2	2	1	0	1	1	1	1
Yemen	3	1	1	4	1	1	0	1	0	0
South Asia Region										
Afghanistan	1	1	2	3	2	2	0	0	0	1
Bangladesh	4	1	2	3	1	2	0	0	0	0

Continued on next page

Table 2.7 – continued from previous page

	Trade, Transportation, and Utilities	Education and Health Services	Natural Resources and Mining	Financial Activities	Manufacturing	Construction	Professional and Business Services	Information	Leisure and Hospitality	Other Services
Country or Territory	Number of Regression Model Indicators Related to Industry Sector									
Bhutan	5	0	1	1	4	1	1	3	1	1
India	6	3	0	3	0	2	1	1	1	0
Nepal	0	3	1	1	0	0	0	1	0	0
Pakistan	2	0	4	1	1	1	1	0	1	0
Sub-Saharan Africa Region										
Angola	3	2	1	0	2	3	0	1	0	0
Benin	3	2	1	1	2	2	2	1	0	1
Burkina Faso	4	5	1	0	2	2	0	2	0	0
Burundi	3	2	0	2	0	0	1	1	0	0
Cabo Verde	1	2	2	1	0	0	0	0	0	0
Cameroon	2	4	2	0	2	1	1	1	0	0
Central African Republic	4	2	2	1	3	2	0	1	0	0
Chad	4	1	2	1	2	1	1	1	0	0
Comoros	5	2	0	2	2	0	1	1	1	1
Dem. Rep. of the Congo	3	3	3	2	2	3	0	2	0	0
Congo	2	3	2	2	1	0	2	4	1	0
Côte d'Ivoire	2	3	1	0	1	2	0	1	0	0
Equatorial Guinea	3	5	0	3	1	1	2	2	1	1
Eritrea	2	4	0	3	0	0	3	1	0	0
Eswatini	4	2	1	1	2	1	1	0	1	0
Ethiopia	2	1	4	1	0	0	0	0	0	0
Gambia	0	4	0	0	0	0	0	2	0	0
Ghana	4	3	0	3	0	2	1	0	1	0
Guinea	0	3	2	2	0	0	0	0	0	0
Guinea-Bissau	1	1	1	4	0	0	1	0	1	0
Kenya	5	2	4	1	2	3	1	2	0	1
Lesotho	1	5	0	1	0	1	1	0	0	0
Liberia	2	2	2	1	2	1	0	0	0	0
Madagascar	4	1	3	1	2	1	1	1	0	0
Malawi	2	3	1	1	2	1	0	1	0	0
Mali	3	1	2	3	2	1	2	0	1	3

Continued on next page

Table 2.7 – continued from previous page

	Trade, Transportation, and Utilities	Education and Health Services	Natural Resources and Mining	Financial Activities	Manufacturing	Construction	Professional and Business Services	Information	Leisure and Hospitality	Other Services
Country or Territory	Number of Regression Model Indicators Related to Industry Sector									
Mauritania	3	4	3	1	1	1	3	1	0	0
Mozambique	3	3	4	0	1	1	0	0	0	0
Namibia	2	0	0	2	2	2	3	2	1	1
Niger	1	1	1	1	0	1	1	1	0	0
Nigeria	2	4	0	2	2	0	1	1	1	0
Rwanda	1	2	1	1	1	1	0	1	0	0
São Tomé and Príncipe	2	1	2	0	1	2	1	0	0	0
Senegal	4	2	1	0	1	2	1	1	0	0
Sierra Leone	3	2	2	0	2	1	0	1	0	0
South Africa	1	3	1	1	0	0	1	1	0	0
Sudan	2	2	2	4	2	1	1	0	0	1
Tanzania	5	1	3	1	2	3	1	1	0	1
Togo	2	4	2	0	2	2	0	0	0	0
Uganda	2	3	2	0	1	2	0	1	0	0
Zambia	6	2	1	0	0	1	1	3	0	0
Zimbabwe	4	1	3	1	3	2	1	1	0	1

CHAPTER 3. DISCUSSION AND CLARIFICATIONS

3.1 Finding historical patterns between industry sectors and data of interest

The data analysis from Section 2.2 was created to discover associations between the World Development Indicators (WDIs) and the Human Development Index (HDI) such that highly-associated WDIs point towards areas of expertise related to increasing HDI. This approach can be generalized to many situations where there is a metric that captures a key motivation or market need and we would like to discover what types of work are associated with its change. For example, instead of using the HDI as the key motivation, consider the United Nations Sustainable Development Goal #2: Zero Hunger. This data analysis could be performed to discover what areas of expertise, characterized by industry sector, have historically been associated with change in the metrics for the Zero Hunger goal. This chapter presents the data analysis in more general terms with additional detail and resources that may be helpful for the implementation of this data analysis.

3.1.1 Overview and general terms used

The goal of this data analysis is to discover historical associations between any Potentially Explanatory Data (PED) and any Key Motive Data (KMD) such that highly-associated PED point towards areas of expertise related to change in KMD. The KMD is any data that provides motivation for or direction to a group's choice of work. KMD may often characterize what the market needs. PED can be any set of historical data that is collected at a similar scale (such as an individual, community, or national scale) to the KMD and where it is reasonable to assume there is some association between some of the PED and the KMD. Some examples of each type of data are provided in Section 3.1.2. An overview of the process followed to discover associations between PED and KMD is as follows:

1. Acquire KMD and PED data

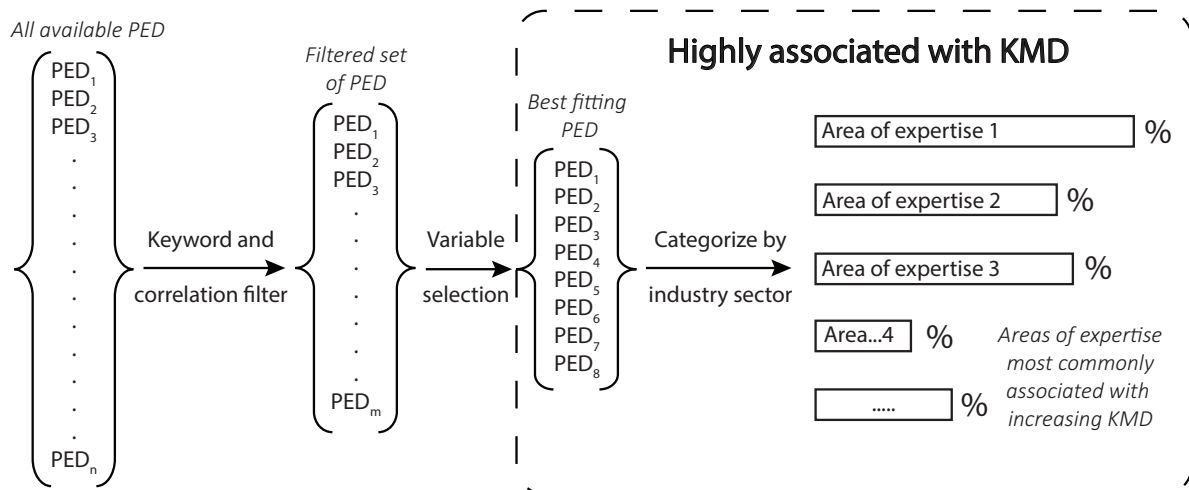


Figure 3.1: Summary of historical data analysis for discovering areas of expertise associated with change in KMD in each country

2. Perform a keyword filter as well as a correlation filter on PED data to reduce to a tractable dataset for variable selection
3. Perform variable selection on filtered and normalized PED data to identify the PED that best fit the change in the KMD
4. Categorize the best fitting PED by NAICS Industry Sector for a more direct connection to areas of expertise

As illustrated in Figure 3.1, this approach offers the ability to filter through the many PED using correlation and variable selection to determine which PED best fit the change in the KMD in a given country [39]. When categorized by industry sector, the selected PED will highlight areas of expertise that have been commonly associated with increasing KMD. Some limitations of this approach were discussed in Section 2.1, and will not be repeated or generalized in this chapter. Each step of this data analysis is discussed in detail below.

3.1.2 Acquire data

The KMD and PED could be from a variety of sources and reported at an individual, community, or national scale. Publicly available data are often published at a national or regional level,

and this analysis was created around data published at a national level, so this analysis may be more applicable to data published at a national level. In the context of EGD, some potential sources of KMD include:

1. Human Development Index [31]
2. A UN Sustainable Development Goal indicator or compilation of a few indicators [5, 65]
3. Market Potential Index [66]
4. World Happiness Index [67]
5. Gender Inequality Index [31]
6. Multidimensional Poverty Index [31]

Some potential sources of PED include:

1. World Development Indicators (1960-present) [6, 68, 69]
2. Statistical Capacity Indicators, Educational Statistics, Gender Statistics, or Health Nutrition and Population Statistics from the World Bank Databank (1960-present) [69]
3. US Bureau of Labor Statistics (United States data only)(1940-present) [70]
4. IBISWorld Global Industry Data (2000-present) [71]
5. UN Comtrade (international trade data)(1960-present) [72]
6. UN Sustainable Development Goals Indicators (2000-present) [5, 65]
7. International Telecommunication Union data [73]
8. Organisation for Economic Cooperation and Development data (1990-present) [74]

All the provided PED examples are worldwide unless otherwise noted.

Handling incomplete data

It is quite common for datasets to be incomplete. There are a number of principles and methods for dealing with incomplete data and a few approaches will be discussed here with references to helpful sources.

First, it is not typically acceptable to extrapolate outside a set of data. For example, if a KMD were available between 2000-2017, it would be inappropriate to extrapolate the data backwards into the 1990s or into the future past 2017.

Second, when there are missing data within a set of given data, there are a number of approaches to imputing the missing data and scenarios when it may be appropriate. Bartlett discusses in more detail when it is appropriate to impute missing data [75]. For example, if a PED were reported from 1995-2001, 2005-2018, there are accepted methods for imputing the data for the years 2002-2004. Some simple methods for imputing missing data include replacing the missing value with the mean or median of the data, a local mean or median, linear interpolation, or last observation carried forward [41, 75]. A more complex and often more robust method is called multiple imputation, and its application is described in detail by Kleinke et al. [76]. Additional detail for handling incomplete data can be found in [77].

Although this may be intuitive, it should be noted that the remainder of the analysis can only be performed on the years for which data was originally present or those years where data was imputed.

3.1.3 Keyword and correlation filter

Keyword filter

As the aim of this analysis is to find associations between PED and the KMD *other* than the indicators used to calculate the KMD, the PED that are directly related to the calculation of the KMD can be filtered out of the dataset. This offers additional opportunity for other associated PED to be captured in each country's linear regression model (discussed in more detail in 2.2.3). Also, the PED filtered out are often colinear with the metrics used to calculate the KMD, and could thus adversely affect the regression model selected with stepwise variable selection [39].

Removing PED that are directly related to the calculation of the KMD can be performed using a keyword filter. Keywords are words taken from the indicators used to calculate the KMD or other closely related words. Examples of keywords used to filter WDIs that were directly related to the HDI can be found in Table A.1 in Section A.2. A function such as MATLAB®'s “strfind” can be used to automatically check if any PED's titles contain the keywords and if they do, they can be flagged and eliminated from the dataset.

Correlation filter

As a first step in analyzing which PED are highly associated with KMD, a bivariate correlation that measures the strength and direction of association is calculated for each PED and the KMD. Common correlation statistics include the Pearson r correlation, Spearman correlation, Kendall's τ , among others. The Pearson r correlation assumes that the data is normal [42], and a test (such as the Anderson-Darling test which can be done in MATLAB® [43]) can be performed to determine if the PED and KMD are normal. Many datasets are non-normal, so the Spearman correlation and Kendall's τ are useful because they eliminate the importance of population distributions by accepting rank-transformed data: where the value of each PED or KMD for each year is replaced by its rank relative to the values from other years of the same PED or KMD [39,42,44].

Additional detail is now provided for the Kendall's τ statistic. Its uses and advantages are very similar to the Spearman correlation coefficient, but this discussion will focus on Kendall's τ . Kendall's τ test is a nonparametric test that calculates how many ranked values are concordant or discordant between two groups [44, 45]. In other words, it reports how one dataset increases or decreases relative to another dataset. This test is powerful because it does not care if one of the datasets is exponential and the other is linear. If Kendall's τ is 1 that means both datasets always increase or decrease together; -1 means one always increases when the other decreases. Less than one means there are periods of time when they move in opposite directions, and periods of time when they follow the same direction. Greater than 0.5 is considered a strong correlation. An illustration example of Kendall's τ can be found in Figure 3.2.

After calculating the correlation coefficient, where reasonable, the original set of PED should be reduced to those with strong correlations, such as those with a $|\text{Kendall's } \tau| > 0.5$.

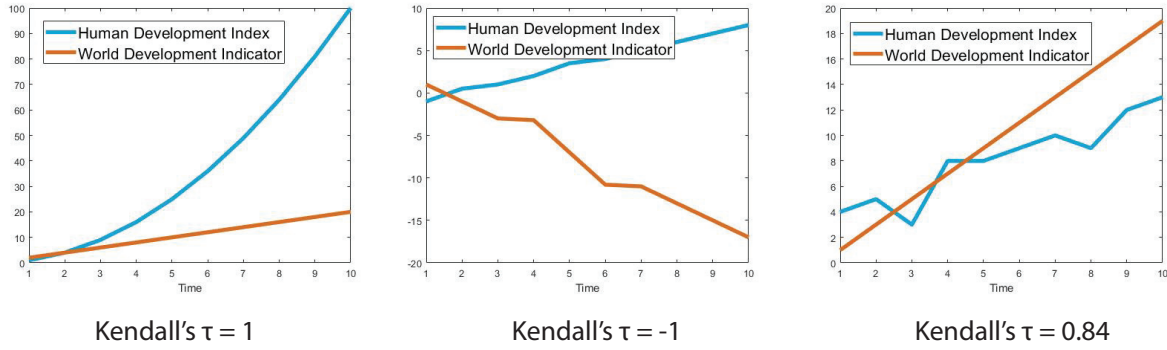


Figure 3.2: Example of Kendall's τ . Note that its calculation is not sensitive to exponential data or non-normal data, but simply captures how one dataset increases or decreases relative to the other dataset

By filtering to a smaller set of PED, the likelihood of selecting coincidental associations during variable selection due to a large number of potential explanatory variables is reduced [47].

3.1.4 Perform stepwise variable selection on percent change of WDI and HDI

Following the initial filtering of PED according to Kendall's τ , the PED that best fit change in the KMD can be identified using variable selection. Variable selection is a tool that systematically tests the combinations of the PED in regression models to determine which combination of PED creates the best fitting regression model for the KMD. This process can be performed on the raw PED and KMD data, but we are concerned with how the KMD *changes* relative to change in PED. Therefore, variable selection should be performed on percent change values of PED and KMD.

Calculate percent change

The percent change in each KMD or PED for each year can be calculated as follows:

$$\delta_t = \frac{I_t - I_{t-1}}{\bar{I}} \quad (3.1)$$

where t represents the t -th year, δ_t represents the percent change in the KMD or PED value from the past year to the current (or t -th) year, I_t represents the KMD or PED value for the t -th year,

and the average, \bar{I} , is calculated from available data. As a note, percent change is commonly calculated as $\delta_{I_t} = \frac{I_t - I_{t-1}}{I_{t-1}}$, but using the mean of the data in the denominator can attenuate drastic percent changes from year to year and thus avoid influential statistics during variable selection. This calculation is made for each country individually.

Stepwise variable selection

The final step in discovering associations between PED and KMD is to perform a stepwise variable selection that creates a multiple linear regression model for each country that denotes those PED that best fit change in the KMD. To be clear, the purpose of the variable selection is to identify the PED in each country that have been highly associated with increasing KMD, rather than to make quantitative predictions of the KMD given PED input. The PED terms will then be categorized by appropriate industry sectors and the associated industry sectors will be the focus of the results, illustrating areas of expertise most commonly associated with change in the KMD. There are multiple stepwise variable selection methods, including forward, backward, and mixed. There is not one best method, and using different stepwise variable selection methods may produce different models. A discussion of each method can be found in Chapter 12.3 of Ramsey's *The Statistical Sleuth* [39].

There are various metrics for evaluating the quality of the models created by the stepwise variable selection, such as R-squared, adjusted R-squared, Cp, BIC, and AIC. Essentially, each metric quantifies how well the PED model fits the KMD while adding a penalty for each additional regressor. The BIC generally applies the most severe penalty for additional regressors, which often leads to a model that is unlikely to overfit the current datasets. On the other hand, R-squared has no penalty for additional regressors and reports a calculation of the magnitude of the residuals. Again, additional discussion regarding the metrics for evaluating the quality of the models can be found in Chapter 12.4 of [39].

Another choice during this step is whether to include or not to include interaction terms in the models. Interaction terms can capture how a KMD is associated with two or more interacting PED. When PED interaction terms are included, it is also appropriate to include each of the PED in the interaction as individual regressors [39]. With a limited number of regressors per model, interactions would produce repeated PED regressors. To determine if excluding interactions would

adversely affect the accuracy of which PED are most associated with change in KMD, multiple linear regression models can be created for each country with and without interaction terms, and the performance of the models with and without interaction terms can be compared (using the adjusted R-squared or BIC or other metric of choice). The models that perform better should generally be utilized.

Determining how many regressors to include in each model is a balance. Including too few regressors limits the amount of exploration of which PED and their connected industry sectors are associated with change in KMD; too many regressors creates a model that is very specific to the years of data included and does not report general patterns that may be applicable for choosing future work. Regarding the number of regressors in the model, Peduzzi et al. suggest that a minimum of 10 years of data per regressor will help avoid overfitting and other issues, while Vittinghoff et al. and Austin and Steyerberg show that this may be too conservative and show that a minimum of 2-3 years of data per regressor is appropriate [50–52]. It is desirable to maximize the number of included WDI regressors to potentially increase their variety, while also avoiding overfitting so the models created will be informative of historical patterns that may be more applicable moving forward. In line with the goal of this research, the number of PED regressors should be maximized for a non-conservative 2-3 years of data per regressor.

In summary, stepwise variable selection is performed for the KMD in each country to optimize the adjusted R-squared or BIC or metric of choice as long as the model follows the 2-3 years of data per regressor criterion.

3.1.5 Serial correlation effects

One potential issue when creating regression models using data that is reported year to year is serial correlation (also known as autocorrelation), where the residuals from year to year are correlated. The effect of serial correlations is that the regression coefficients will be imprecise, suggesting more or less statistical significance than is true [53, 54]. In other words, the p -values for the regression coefficients will be estimated too high or too low and the confidence intervals will be too wide or too narrow.

A common test to discover the presence of serial correlations is the Durbin-Watson test that reports a statistic, d , between 0 and 4, where $d = 2$ suggests no serial correlation [55, 56]. Critical

values, d_L and d_U , of the Durbin-Watson statistic can be looked up in tables provided by Savin and White. d_L and d_U are dependent on the number of years of data and the number of regressors in each model. The Durbin-Watson test should be performed that reports d for each model and if $d < d_L$ then the regression coefficients are predicted with exaggerated significance (p -value too low) and if $d > d_U$ then the regression coefficients are predicted with underestimated statistical significance (p -value too high). Serial correlations should be adjusted for when $d < d_L$ or $d > d_U$ for any model.

There are multiple approaches to adjust for serial correlations, one of which is to use a Newey-West robust covariance matrix when fitting each model [58]. For example, I used the “statsmodel” Python package and could specify the covariance matrix type when fitting each model.

3.1.6 Categorizing WDIs by NAICS Industry Sector

The previous steps in this data analysis filter through many PED to just a handful of PED that best fit change in KMD in each country. With the goal of understanding which areas of expertise are commonly associated with change in the KMD, the PED selected for each country’s model are to be categorized according to area of expertise. Some PED sets may already include indicators that directly capture a succinct set of areas of expertise, while others do not. One option is to categorize the PED according the North American Industry Classification System (NAICS) Industry Sectors or the International Standard Industrial Classification of all Economic Activities (ISIC) providing a direct link to areas of expertise [78, 79].

To perform the categorizations, researchers working in teams of two, examine each remaining unique PED to determine if there is a connection to one or more industry sectors. By working in teams of two, each researcher is forced to justify, to a teammate, the reasoning for connecting a PED to an industry sector. It is desirable to include researchers from a variety of backgrounds—nationality, gender, expertise—to reduce research bias; although it is also helpful to include researchers that are experts in the topics covered by the PED and especially the KMD.

When completing the categorizations, first, each researcher completes a training to familiarize himself/herself with the NAICS Industry Sectors (or other categories of choice). Next, the researchers, working in teams of two, complete a categorization survey where they assign each

PED to none, one, or more industry sectors. A helpful platform for a categorization survey is Qualtrics and an example of the survey used to categorize WDIs by industry sector is shown in Figure A.2 in Section A.3. Each PED is rated according to how connected or not connected it is to one or more industry sector. A common scale is the Likert scale whose terminology varies depending on what is being rated [60]. Providing a rubric for the researchers involved in the survey provides more reliable categorizations and an example of a rubric is provided in Table 2.1 in Section 2.2.5.

The order in which the WDIs are presented to each team is randomized to reduce bias. The teams first rate each PED's connected-ness to the NAICS industry super-sectors, then by industry sectors (if there was a super-sector connection), and lastly by industry sub-sectors (if there was a sector connection). A detailed description of each PED is made available during the survey for the teams to refer to as needed.

The results can be presented according to the PED that are “Moderately” or “Fundamentally connected” to at least one industry sector. As a note, a “No specific connection” option may be helpful to include when there are PED that reflect a measure of industry as a whole while not referring to any specific sector of industry.

PED cover a number of areas, such as economics, climate change, gender, health, poverty, and many others; thus, how they connect to industry sectors is not always apparent and it was expected that the teams would differ in some of the categorizations. After the teams categorize all the unique PED, they may meet to discuss disagreements within teams and discrepancies between the teams' categorizations. The teams offer their reasoning for establishing the connection or no connection to the industry sector, after which each researcher individually and anonymously votes in agreement or disagreement with the industry sector connections of each discrepancy. Those that receive a majority vote are included in the results.

3.1.7 Review

This data analysis is intended to reduce a large set of PED to a set of PED that are highly associated with the KMD, and, when categorized by industry sector, they show which areas of expertise have historically been highly associated with change in KMD.

CHAPTER 4. CONCLUSIONS, LIMITATIONS, AND ADDITIONAL FUTURE WORK

This thesis presents patterns of how various areas of expertise have historically been associated with increasing HDI in developing countries. Examining historical patterns provides a basis for understanding how projects or products related to various areas of expertise have impacted people. It is intended that engineers working to improve people's opportunities and conditions in developing countries (EGD practitioners) can look to these results for direction when determining *what* to work on and *where* to implement their work.

The patterns discovered in this research suggest that, aside from working in the areas of health, education and income directly, products or projects related to the following have been historically associated with increasing HDI in developing countries: trade (exports and imports), access to electricity, basic sanitation facilities, manufacturing, construction of residential buildings as well as utility systems, agriculture, mining or oil and gas extraction, among others. Because of their close association with increasing HDI, products or projects that fall into these areas of expertise can be pursued as surrogates for focusing on the three dimensions of HDI alone (health, education, and income). Further research can be done to determine if increasing human development likely caused an improvement in these other areas, or if improving these areas likely caused an increase in human development. Although causation is difficult to claim in any study, understanding which improved first could add credibility to a justification for focusing resources on a particular area of expertise in a given developing country.

While this research focuses on the areas of expertise themselves that best fit the change in human development, further research could be performed that creates quantitative models that could predict the change in HDI given a unit increase in metrics for other areas of expertise. There are a number of challenges surrounding quantitative predictions, but with a number of simplifying approximations, quantitative relationships between an engineer's work and its impact on end user could be modeled using public or generated data. An initial prototype could be created with linear

or logistic regression, where higher-fidelity predictive models could incorporate machine learning approaches where the model can be extracted (non-black box methods).

It is intended that the presented areas of expertise commonly associated with increasing human development could be used by a diversity of groups and some examples have been provided as to how. In general, the patterns presented are useful during the opportunity development phase as an individual or group is exploring product or project options as well as where to implement their work. As a research group focused on exploring the relationships between products and people, further research could be performed in partnership with a company or non-governmental organization that shows how the information herein would be applied to identify and justify a high-potential opportunity, in addition to validating their choice through interviews, observational studies, or co-design with the individuals or community of interest.

Additionally, the method used in Chapter 2 to discover areas of expertise associated with increasing HDI was presented generally in Chapter 3 such that associations might be discovered between areas of expertise and any data of interest (KMD as discussed in Chapter 3). Other data sources were recommended for both the KMD and PED that could be analyzed using the general data analysis described in Chapter 3.

To build on this research, in addition to the current state of knowledge regarding product social impacts, patterns could be identified between areas of expertise and the 11 social impact categories. In order for this to be feasible, a set of indicators that characterize the social impact in each category would need to be established. An understanding of how areas of expertise relate to social impacts could make incorporating social impact into design and engineering products and projects much more tractable, especially because engineers traditionally consider the health and safety impact of their work while neglecting other social impact categories [80].

REFERENCES

- [1] Hoornweg, D. A., Bhada-Tata, P., and World Bank, 2012. What a waste : a global review of solid waste management Tech. rep. vi, 23
- [2] Collins, J., 2001. *Good to Great: Why some companies make the leap... and others don't*. HarperBusiness, New York, NY. vii, 5, 6
- [3] Hunter, A., 2018. UN Human Development Index - Wikimedia Commons. vii, 51
- [4] Ottosson, H. J., Mattson, C. A., and Dahlin, E. C., 2019. "Analysis of Perceived Social Impacts of Existing Products Designed for the Developing World, With Implications for New Product Development." *Journal of Mechanical Design*, **142**(5). 1
- [5] United Nations, 2019. The Sustainable Development Goals Report Tech. rep. 1, 4, 36
- [6] The World Bank WDI - Home. 1, 10, 36
- [7] National Academy of Engineering, 2008. National Academy of Engineering Grand Challenges For Engineers Tech. rep. 1
- [8] Mattson, C. A., and Winter, A. G., 2016. "Why the Developing World Needs Mechanical Design." *Journal of Mechanical Design*, **138**(7), 6, p. 070301. 1, 4
- [9] Goodier, R., and Engineering for Change, 2016. Introduction to Engineering for Global Development. 1
- [10] Mattson, C. A., and Sorensen, C. D., 2020. *Product Development: Principles and Tools for Creating Desirable and Transferable Designs*. Springer. 2, 6, 21
- [11] Wood, A. E., and Mattson, C. A., 2016. "Design for the Developing World: Common Pitfalls and How to Avoid Them." *Journal of Mechanical Design*, **138**(3), 1, p. 031101. 2, 4
- [12] Stringham, B., Smith, D. O., Mattson, C. A., and Dahlin, E. C., 2020. "Combining Direct and Indirect User Data for Calculating Social Impact Indicators of Products in Developing Countries." *Journal of Mechanical Design*. 2
- [13] Mabey, C., Mattson, C. A., and Dahlin, E. C., 2020. "Assessing Global Needs When Identifying Potential Engineering for Global Development Projects." In *Submitted to International Design Engineering Technical Conference*. 2, 8
- [14] Wang, Y., 2013. More People Have Cell Phones Than Toilets, U.N. Study Shows. 4
- [15] Engineering For Global Development (EGD) — ASME Engineering Network. 4

- [16] World Bank, 2017. World Bank Country and Lending Groups World Bank Data Help Desk. 4
- [17] Pakravan, M. H., and MacCarty, N. A., 2020. “Design for Clean Technology Adoption: Integration of Usage Context, User Behavior, and Technology Performance in Design.” *Journal of Mechanical Design*, **142**(9), 9. 4
- [18] Shamsbery, P., and Winter, A. G., 2018. “Shape and Form Optimization of On-Line Pressure-Compensating Drip Emitters to Achieve Lower Activation Pressure.” *Journal of Mechanical Design, Transactions of the ASME*, **140**(3), 3. 4
- [19] Diaz Lankenau, G. F., and Winter, V. A. G., 2020. “Design of a specialized tractor to replace draft animals in small farms.” *Journal of Mechanical Design*, 4, pp. 1–21. 4
- [20] Geneva: World Health Organization and UN_HABITAT, 2018. Progress on Wastewater Treatment 2018 Piloting the monitoring methodology and initial findings for SDG indicator 6.3.1 Tech. rep. 4
- [21] Njiru, C. W., and Letema, S. C., 2018. “Energy Poverty and Its Implication on Standard of Living in Kirinyaga, Kenya.” *Journal of Energy*. 4
- [22] Zinn, C., Bailey, R., Barkley, N., Walsh, M. R., Hynes, A., Coleman, T., Savic, G., Soltis, K., Primm, S., and Haque, U., 2018. “How are water treatment technologies used in developing countries and which are the most effective? An implication to improve global health.” *Journal of Public Health and Emergency*, **2**(0). 4
- [23] Mattson, C. A., and Wood, A. E., 2014. “Nine Principles for Design for the Developing World as Derived From the Engineering Literature.” *Journal of Mechanical Design*, **136**(12), 10. 4
- [24] Wood, A. E., and Mattson, C. A., 2014. “A Method for Determining Customer Needs in the Developing World.” *Proceedings of the ASME IDETC/CIE 2014*, 8, pp. 1–13. 4
- [25] Chambers, R., 1978. “Project selection for poverty-focused rural development: Simple is optimal.” *World Development*, **6**(2), 2, pp. 209–219. 4
- [26] Grady, C. A., He, X., and Peeta, S., 2015. “Integrating social network analysis with analytic network process for international development project selection.” *Expert Systems with Applications*, **42**(12), 7, pp. 5128–5138. 4
- [27] Judge, B. M., Hölttä-Otto, K., and Winter V, A. G., 2015. “Developing world users as lead users: A case study in engineering reverse innovation.” *Journal of Mechanical Design, Transactions of the ASME*, **137**(7), 7. 5
- [28] Wasley, N. S., Lewis, P. K., Mattson, C. A., and Ottosson, H. J., 2017. “Experimenting with concepts from modular product design and multi-objective optimization to benefit people living in poverty.” *Development Engineering*, **2**, 1, pp. 29–37. 5
- [29] Campos, F., Goldstein, M., McGorman, L., Munoz Boudet, A. M., and Pimhidzai, O., 2015. Breaking the Metal Ceiling Female Entrepreneurs Who Succeed in Male-Dominated Sectors. 5

- [30] Arif, S., and Siddique, Z., 2012. “Design and selection of safe water supply solutions for emerging regions - A demography based demand driven approach.” In *Proceedings of the ASME Design Engineering Technical Conference*, Vol. 3, American Society of Mechanical Engineers Digital Collection, pp. 443–453. 5
- [31] UNDP - United Nations Development Programme, 2018. Technical Notes of the 2018 Statistical Update - Calculating the HDI Tech. rep. 5, 10, 36
- [32] Ray, S., Ghosh, B., Bardhan, S., and Bhattacharyya, B., 2016. “Studies on the impact of energy quality on human development index.” *Renewable Energy*, **92**, 7, pp. 117–126. 5
- [33] Dufo-López, R., Cristóbal-Monreal, I. R., and Yusta, J. M., 2016. “Optimisation of PV-wind-diesel-battery stand-alone systems to minimise cost and maximise human development index and job creation.” *Renewable Energy*, **94**, 8, pp. 280–293. 5
- [34] Pasternak, A. D., and Lawrence Livermore National Laboratories, 2000. Global Energy Futures and Human Development: A Framework for Analysis Tech. rep. 5
- [35] Wu, Q., Clulow, V., and Maslyuk, S., 2010. “Energy consumption inequality and human development.” In *2010 International Conference on Management Science and Engineering, ICMSE 2010*, pp. 1398–1409. 5
- [36] Kumar, M. D., Saleth, R. M., Foster, J. D., Niranjana, V., and Sivamohan, M. V. K., 2016. “Chapter 2 - Water, Human Development, Inclusive Growth, and Poverty Alleviation: International Perspectives BT - Rural Water Systems for Multiple Uses and Livelihood Security.” In *Rural Water Systems for Multiple Uses and Livelihood Security*. Elsevier, pp. 17–47. 5
- [37] Programme, U. N. D., 1990. Human Development Report 1990 Tech. rep. 5
- [38] Ranis, G., Stewart, F., Samman, E., and Elizabeth, Q., 2006. “Human Development: Beyond the Human Development Index.” *Journal of Human Development and Capabilities*, **7**(3), pp. 323–358. 5
- [39] Ramsey, F. L., and Schafer, D. W., 2013. *The Statistical Sleuth*. Cengage Learning. 10, 11, 12, 14, 35, 37, 38, 40
- [40] United Nations Development Programme Human Development Data (1990-2018) — Human Development Reports. 10
- [41] Streiner, D. L., 2012. Last Observation Carried Forward In: *Encyclopedia of Research Design*. 10, 37
- [42] Schober, P., and Schwarte, L. A., 2018. “Correlation coefficients: Appropriate use and interpretation.” *Anesthesia and Analgesia*, **126**(5), pp. 1763–1768. 11, 38
- [43] Anderson, T. W., and Darling, D. A., 1952. “Asymptotic Theory of Certain ”Goodness of Fit” Criteria Based on Stochastic Processes.” *Annals of Mathematics and Statistics*, **23**(2), pp. 193–212. 11, 38
- [44] Kendall, M. G., and Gibbons, J. D., 1990. *Rank Correlation Methods*. Edward Arnold. 11, 38

- [45] Statistics Solutions Correlation (Pearson, Kendall, Spearman) - Statistics Solutions. 11, 38
- [46] Linear or rank correlation - MATLAB corr. 11
- [47] Smith, G., 2018. "Step away from stepwise." *Journal of Big Data*, **5**(1), p. 32. 12, 39
- [48] Efroymson, M., 1960. "Multiple Regression Analysis." In *Mathematical Models for Digital Computers*, A. Ralston and H. Wilf, eds. Wiley, New York, NY. 14
- [49] Theil, H., 1961. *Economic Forecasts and Policy*. Amsterdam. 14
- [50] Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., and Feinstein, A. R., 1996. "A simulation study of the number of events per variable in logistic regression analysis." *Journal of Clinical Epidemiology*, **49**(12), 12, pp. 1373–1379. 14, 41
- [51] Vittinghoff, E., and McCulloch, C. E., 2007. "Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression." *American Journal of Epidemiology*, **165**(6), 1, pp. 710–718. 14, 41
- [52] Austin, P. C., and Steyerberg, E. W., 2015. "The number of subjects per variable required in linear regression analyses." *Journal of Clinical Epidemiology*, **68**(6), 6, pp. 627–636. 14, 41
- [53] Pindyck, R., and Rubinfeld, D., 1976. *Econometric Models and Economic Forecasts*. McGraw-Hill Incorporated. 14, 41
- [54] Bence, J. R., 1995. "Analysis of short time series: Correcting for autocorrelation." *Ecology*, **76**(2), pp. 628–639. 14, 41
- [55] Durbin, J., and Watson, G. S., 1950. "Testing for Serial Correlation in Least Squares Regression: I." *Biometrika*, **37**(3/4), 12, p. 409. 15, 41
- [56] Chatterjee, S., and Simonoff, J. S., 2013. *Handbook of Regression Analysis*. John Wiley and Sons, 1. 15, 41
- [57] Savin, N. E., and White, K. J., 1977. "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many." *Econometrica*, **45**(8), pp. 1989–1996. 15
- [58] Newey, W., and West, K., 1987. "A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix." *Applied Econometrics*, **55**(3), pp. 703–708. 15, 42
- [59] Statistics Canada, Instituto Nacional de Estadística y Geografía, and Economic Classification Policy Committee, 2017. North American Industry Classification System Manual Tech. rep. 16
- [60] Rea, L. M., and Parker, R. A., 2005. *Designing and conducting survey research: a comprehensive guide (3rd ed.)*. 16, 43
- [61] Eye, A., and Mun, E.-Y., 2014. "Analyzing rater agreement: Manifest variable methods." *Analyzing Rater Agreement: Manifest Variable Methods*, pp. 1–190. 19

- [62] Vallat, R., 2018. “Pingouin: statistics in Python.” *Journal of Open Source Software*, **3**(31), p. 1026. 19
- [63] Koo, T. K., and Li, M. Y., 2016. “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.” *Journal of Chiropractic Medicine*, **15**(2), 6, pp. 155–163. 19
- [64] Eggers, E. K., and Lemarchand, R., 2020. Burundi. 26
- [65] Nations, U. Sustainable Development Goals Indicators. 36
- [66] globalEdge Market Potential Index (MPI) - 2019. 36
- [67] Layard, R., Sachs, J. D., Neve, J.-e. D., Huang, H., and Wang, S., 2020. World Happiness Report Tech. rep. 36
- [68] The World Bank WDI - Sources and Methods. 36
- [69] The World Bank DataBank — The World Bank. 36
- [70] US Bureau of Labor Statistics Databases, Tables & Calculators by Subject. 36
- [71] IBISWorld IBISWorld - Industry Market Research, Reports, & Statistics. 36
- [72] UN Comtrade International Trade Statistics Database. 36
- [73] International Telecommunications Union Statistics. 36
- [74] Organisation for Economic Cooperation and Development OECD Data. 36
- [75] Bartlett, R., 2010. Missing Data, Imputation of In: Encyclopedia of Research Design. 37
- [76] Kleinke, K., 2020. *Applied Multiple Imputation*. Statistics for Social and Behavioral Sciences. Springer International Publishing, Cham. 37
- [77] Kim, J. K., 2014. *Statistical Methods for Handling Incomplete Data*. CRC Press, Boca Raton, FL. 37
- [78] Statistics Canada, Instituto Nacional de Estadística y Geografía, and Economic Classification Policy Committee, 2017. NORTH AMERICAN INDUSTRY CLASSIFICATION SYSTEM EXECUTIVE OFFICE OF THE PRESIDENT OFFICE OF MANAGEMENT AND BUDGET Tech. rep. 42
- [79] United Nations, 2015. International Standard of Industrial Classification of All Economic Activities (ISIC), Rev. 4 Tech. rep. 42
- [80] Pack, A. T., Rose Phipps, E., Mattson, C. A., and Dahlin, E. C., 2020. “Social Impact in Product Design, An Exploration of Current Industry Practices.” *Journal of Mechanical Design*, **142**(7), 7. 45

APPENDIX A. APPENDIX

A.1 Human Development Index across the world

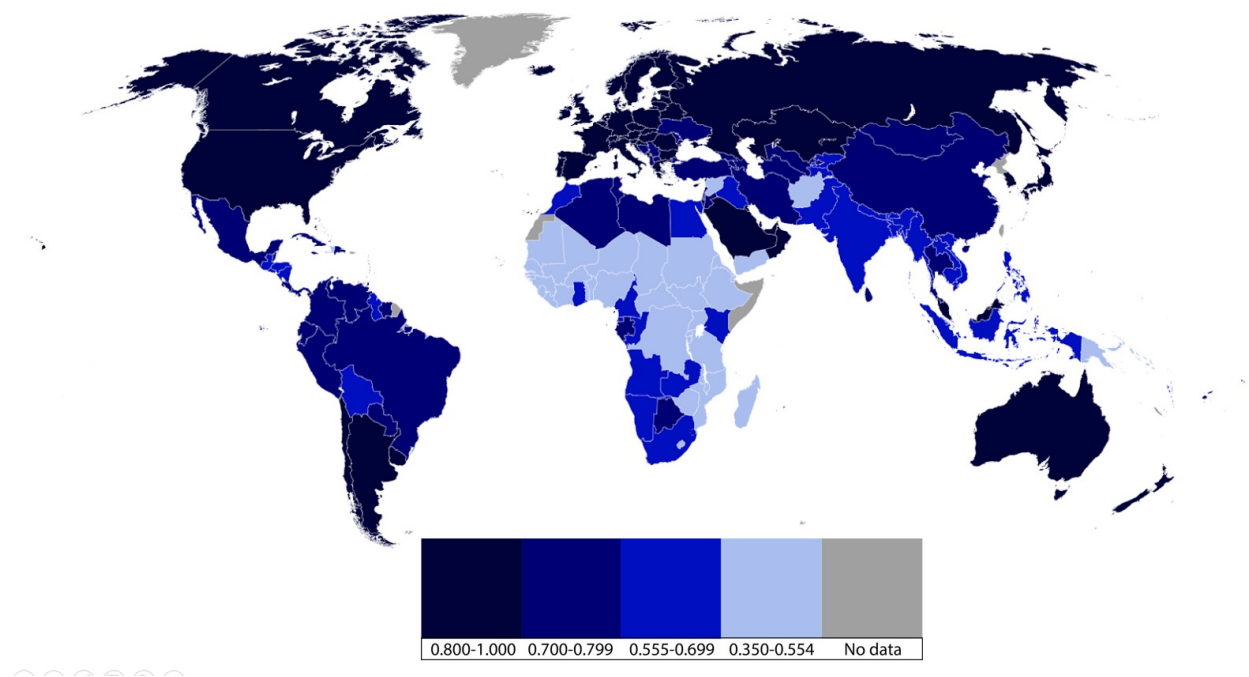


Figure A.1: Visualization of HDI across the world adapted from a figure by Alice Hunter [3]. Darker blue suggests more developed countries. Note that a high concentration of under-developed countries (light blue) are in the Sub-Saharan Africa region.

A.2 Keywords used to filter WDIs directly related to HDI dimensions

The keywords in Table A.1 were used in a computer program that identified any WDI with these keywords in their title and filtered them out of the list of WDIs used during variable selection.

Keywords		
Health	Education	Income
Mortality	School	GDP
Newborns	school	Income share
tetanus	enrollment	Gross national expenditure
infant deaths	teachers	Gross savings
healthcare	pupils	gross savings
medicine	Primary education	gross capital formation
hospitals	Secondary education	Gross capital formation
Death rate	Gross intake ratio	gross loans
Life expectancy	education expenditure	debt
birth	Educational attainment	GNI per capita (
Birth	Tertiary education	GNI per capita,
Survival	tertiary education	GDP, PPP
infected	Literacy rate	GDP per capita (
health	student	GDP per capita growth
disease		GDP per capita,
illness		PPG
virus		national income
doctor		GDP (
Population, total		GNI (
Population, female		capital formation
Population, male		GNI, PPP (
Population ages		GDP: linked series
deaths		DEC alternative
infant		GNI, Atlas method
Birth rate		Net secondary income
death		GDP per person employed (
vaccine		
antidote		
fertility		
fertility rate		
Fertility rate		
neonatal		
Mortality rate		

Table A.1: Keywords used to filter WDIs that were directly related to HDI dimensions. Note that some words are repeated with a lower case first letter. This was because a case-sensitive function was used to find WDIs with these keywords.

A.3 Categorization survey

Please consider and report how the following World Development Indicator (WDI) is connected or not connected to NAICS Industry Sector(s).

WDI: *Social contributions (current LCU)*

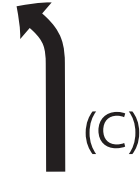
-- [more detail](#)

	Fundamentally connected	Moderately connected	Insufficiently connected	No specific connection	Unknown
Natural Resources and Mining	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Construction	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Manufacturing	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trade, Transportation, and Utilities	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Information	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Financial Activities	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional and Business Services	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Education and Health Services	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Leisure and Hospitality	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other Services (except Public Administration)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure A.2: Screenshot of the Qualtrics survey used to categorize each WDI by NAICS Industry Sectors. If the WDI were “Moderately” or “Fundamentally” connected to any industry super-sectors shown on the left, the next page would allow the raters to categorize the WDI by industry sector and then sub-sector.

A.4 Results table guide

NAICS Industry Super-sector	Frequency of Industry Sector Connection to Increasing HDI in Developing Countries (%)	NAICS Industry Sector or Sub-sectors	Frequency of Industry Sector Connection to Increasing HDI in Developing Countries (%)
Trade, Transportation and Utilities	36.4	Utilities (NAICS 22)	15.3
		Wholesale Trade (NAICS 42)	8.6
		Transportation and Warehousing (NAICS 43)	7.1
		Retail Trade (NAICS 44-45)	5.2
Education and Health Services	26.4	Health Care and Social Assistance (NAICS 62)	20.0
		Social Assistance (NAICS 624)	13.3
		Ambulatory Health Care Services (NAICS 621)	8.8
		Educational Services (NAICS 61)	6.0
Natural Resources and Mining	19.3	Agriculture, Forestry, Fishing and Hunting (NAICS 11)	8.3
		Mining, Quarrying, and Oil and Gas Extraction (NAICS 21)	3.8
		Crop Production (NAICS 111)	2.9
		Forestry and Logging (NAICS 113)	3.1
		Animal Production (NAICS 112)	1.2



NAICS Industry Super-Sector	Frequency of Industry Sector Connection to Increasing HDI in Developing Countries by Region (%)						
	East Asia and Pacific (12)	Europe and Central Asia (2)	Latin America and Caribbean (7)	Middle East and North Africa (6)	South Asia (6)	Sub-Saharan Africa (43)	
Trade, Transportation, and Utilities	38.5	25.0	32.2	33.8	39.1	33.4	
Education and Health Services	26.4	31.3	24.6	17.7	17.7	15.4	
Natural Resources and Mining	17.6	0.0	20.8	22.2	21.7	19.8	
Financial Activities	23.1	12.5	13.2	26.7	26.1	15.9	
						16.5	
						14.9	
						11.3	
						13.1	
						2.7	
						3.4	

Country or Territory	Number of Regression Model Indicators Related to Industry Sector										
	Trade, Transportation, and Utilities	Education and Health Services	Natural Resources and Mining	Financial Activities	Manufacturing	Construction	Professional and Business Services	Information	Leisure and Hospitality	Other Services	
East Asia and Pacific Region											
Cambodia	1	4	1	1	3	1	1	2	0	1	
Indonesia	6	2	2	1	2	4	0	0	0	0	
Kiribati	5	4	0	2	1	1	1	1	1	1	
Lao People's Dem. Republic	1	3	0	1	1	0	0	3	0	0	
Micronesia	2	1	2	2	2	2	0	1	0	0	
Myanmar	4	2	3	0	2	2	1	0	0	0	
Papua New Guinea	4	0	2	2	2	0	0	0	0	0	
Philippines	3	2	3	3	0	1	1	1	1	0	
Solomon Islands	3	1	0	3	1	0	0	0	0	0	
Timor-Leste	4	2	1	2	2	1	0	1	1	0	
Vanuatu	2	3	0	2	1	0	0	0	0	0	
Viet Nam	0	0	2	0	2	0	1	0	0	2	
Europe and Central Asia Region											
Kyrgyzstan	3	3	0	1	0	2	0	2	1	0	
Tajikistan	1	2	0	1	0	1	0	2	0	0	
Latin America and Caribbean Region											
Bolivia	4	3	2	1	3	1	2	1	0	1	
El Salvador	3	1	1	0	4	2	0	1	1	0	
Zimbabwe	4	1	0	3	0	1	0	0	0	0	

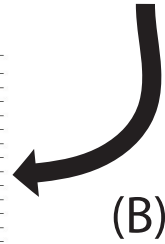


Figure A.3: In general, the Industry Sector Breakdown Table (Table 2.4) is referenced first to determine *what* to work on and the Region and Country Tables (Tables 2.5 and 2.7) are most helpful in determining *where*, although these two questions are interdependent. Steps (A), (B), and (C) are not necessarily sequential. (A) After identifying an area of expertise that is commonly associated with increasing HDI across all developing countries, the Region Table is referenced to find if that area of expertise has been more commonly impactful in a particular region. (B) Understanding regionally where an area of expertise has been impactful is helpful, but the Country Table illustrates, with more focus, where that area of expertise has been most impactful. (C) When a country or region has already been identified as the desired place of work, the two right columns in the Industry Sector Breakdown Table give more detail as to which sectors and sub-sectors within the industry super-sectors have been associated with improving HDI.