

2019

Data Literacy in Economic Development

Simon Halliday

Smith College, shalliday@smith.edu

Follow this and additional works at: https://scholarworks.smith.edu/eco_facpubs



Part of the [Economics Commons](#)

Recommended Citation

Halliday, Simon, "Data Literacy in Economic Development" (2019). Economics: Faculty Publications, Smith College, Northampton, MA.

https://scholarworks.smith.edu/eco_facpubs/12

This Article has been accepted for inclusion in Economics: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

DATA LITERACY IN ECONOMIC DEVELOPMENT

Simon D. Halliday

Department of Economics, Smith College, Northampton, MA, USA

Abstract: In economic development and other economics electives, students regularly encounter economic measures of absolute and relative deprivation, from poverty measures like the Foster-Greer-Thorbecke index to measures of distribution like the Gini index. By “doing economics,” students practice applying economic measurement to real-world data and develop more general data literacy. The author proposes a series of exercises starting with stylized 10-household economies, proceeding to nationally representative cross-sectional surveys using MS Excel or Google Spreadsheets, and culminating in students applying their acquired data literacy to a team project. The data sources are easily tailored to alternative household surveys in low- and middle-income countries that include the required variables. Students learn data literacy through recognizing the properties of rectangular data, visualizing data appropriately, and creating aggregate economic measures.

Keywords: data literacy, development, economic education, inequality, poverty

JEL codes: A22, I32

Contact: Simon D. Halliday; shalliday@smith.edu; Assistant Professor, Department of Economics, Smith College, 234 Wright Hall, Northampton, MA 01062, USA.

The author thanks Vis Taraz, Madeline Wettach, attendees at CTREE and LAC-DEV, and two anonymous referees for helpful comments on the article. Support was received from the TIER Faculty Fellowship.

A generation ago, Bartlett and King (1990, 182) argued that “we expect students to eventually learn to think like economists without providing them with any real opportunity to learn how economists go about ‘doing’ economics.” In economic development, a practitioner “doing economics” needs to use data in a variety of ways: by constructing measures of outcomes like income and poverty and by using the measures to understand both the status quo and the effects of new policies. The instruction of economic development—and similar courses—therefore requires a marriage of data to economic theory, policy analysis, and decision-making. In this article, I aim to address these ideas and fill a gap in the literature by providing a set of exercises for classes that teach topics on income, poverty and inequality, and which aspire to use real-world data and facilitate experiential learning while improving student engagement (Hawtrey 2007). I provide step-by-step instructions for the exercises in the supplementary material.

Although development economics and the study of poverty are featured in the economic education literature, few examples exist of deliberate and sustained analysis of the measurement issues in an undergraduate development economics or economics of poverty classroom. Diduch (2012) is the main exception. Considering a question like “How do you overcome poverty in a big and diverse country?,” Diduch demonstrates that the problem of poverty motivates students to learn the details of measurement and, as a result, they learn the tools of policy analysis in the United States. However, no similar demonstrations exist for a course in development economics or in the use of data from middle- and low-income countries.

A variety of papers have certainly attempted to address questions of content and research methods in economic development and related curricula. Singh and Russo (2013) provide insight into how the use of a “dream experiment” motivates students at a small liberal arts college to design a randomized controlled trial in a developing country while focusing on a policy-relevant

topic. In a behavioral economics class discussed in Castilla (2014), students engaged both with data literacy and experimental design by analyzing data they collected themselves to answer a research question. In discussing economic growth theories in the undergraduate curriculum, two articles provide a relevant perspective. Chen (2005) showcases the great diversity in theories that explain economic growth whereas Acemoglu (2013) makes the case that teaching growth and development in the undergraduate curriculum exposes students to facts about the world economy that are inherently interesting and that instructors should teach growth models in a nuanced and historically situated manner. In this article focusing on microeconomic development, I assume that an economic development course will cover topics of income, poverty and inequality, and I innovate in *how* those topics are taught.

To improve their understanding of measurement, students need to cultivate data literacy. Data literacy is separate from statistical literacy (MacGillivray and Pereira-Mendoza 2011) and economic literacy (Hansen, Salemi, and Siegfried 2002; Gilleskie and Salemi 2012). Researchers conceive of data literacy as the ability “to read, create and communicate data as information” (MacGillivray 2017, 6). Data literacy therefore requires basic skills in data management, data visualization, and the computation of quantitative results using data to create information (IISD 2017). However, research on data literacy in economics is scanty, and research outside economics tends not to explain the methods by which it can be taught in an economics classroom.

Notwithstanding this gap, a focus on data literacy in the context of economic education may help to meet Hansen’s (2001, 231) criterion that, in studying economics, students should have the ability to “interpret and manipulate economic data.” Furthermore, data collection and analysis fall under the ‘essential competencies’ outlined by Allgood and Bayer (2016, 106). I

argue that a focus on data literacy therefore helps to promote these competencies and that students will more readily achieve them through step-by-step encounters with data to hone their data literacy and facilitate experiential learning by getting them to measure development outcomes themselves.

The course in which I incorporate these exercises is a lower-level elective in economic development that does not require students to have taken calculus or statistics, but does require either introductory microeconomics or macroeconomics. The course is taught at a liberal arts college where the development economics elective enrolls anywhere from 25 to 55 students in a given semester (see online appendix A.1 for the syllabus; a list of the online resources is available in the appendix to this article). Although I have tailored the exercises to a course in economic development, the exercises could be incorporated into courses on poverty and inequality, public policy, urban economics, or economic policy analysis in which working with data or data literacy is an important course component. Students do not need competency with spreadsheet programs on entering the course, but the instructor will need to dedicate some class time to introducing basic spreadsheet functions such as IF(), COUNT() or SUM(). The practice of the course is similar to that used in higher-level courses where experiential learning (also known as learning-by-doing and hands-on learning) is used in capstone courses to teach economics research skills at the best-ranked liberal arts college and university economics departments in the United States (Hoyt and McGoldrick [2017] summarize these ideas and introduce a variety of case studies examining the practices more closely, such as Brunnermeier [2017] and Lima and Tsiang [2017]; with similar practices proposed in Ball and Medeiros [2012]).

I begin by outlining the main measures that are taught in an economic development classroom—income, poverty and inequality—and explain related theory and history from which these aggregate measures emerge. Second, I explain the initial exercise that I use to introduce students to the basics of data literacy for development: individuals with incomes who form an economy (a sample), the aggregate properties of which need to be measured (data analysis). The exercise employs stylized 10-person (or 10-household) economies for students to start learning data literacy and development theory (see online appendices B. 1 and B. 2). Third, the exercise proceeds to an advanced stage by providing students with publicly available household and cross-country data in Excel. Students become familiar with the constituent parts of data (observations, variables) and use the data to cultivate an understanding of visualization and aggregation by calculating measures of income, poverty and inequality. Students work on team projects on a low- or middle-income country that they choose and for which they apply their data literacy skills to a research question developed by the team. The measures and methods used in economic development are used in a variety of other core courses and electives where improving data literacy could be beneficial: from microeconomics to public policy, and from urban economics to the economics of poverty or inequality.

MEASURING INCOME, POVERTY AND INEQUALITY

Income

Most economic development courses introduce the measurements of income, poverty and inequality by starting with national measures of gross domestic product (GDP) and gross national product or income (GNP/GNI). Although artificial input-output tables provide the basis for teaching income calculation, instructors generally cannot get students to calculate GDP with

real data from an economy as the process is too complex and beyond the scope of even some national statistical agencies (see, e.g., Jerven 2013).

Furthermore, national income measures have a variety of shortcomings in terms of understanding poverty and inequality: First, if the per capita GDP of a country is, say, USD 10,000 in 2014, that does not imply that each person in the economy spends that amount of money, or that the *average* (median) consumer in the economy has an equivalent disposable income. Second, and importantly for economic development, per capita GDP measures say nothing about the distribution of income or the skewness of that distribution. For example, two economies may have similar per capita GDP, but different poverty headcounts or Gini index values. Third, GDP measures are compiled from data with which most students are unfamiliar and from which they are likely to remain disengaged. Many students, in their work and family life, though, will be familiar with the idea that the family as a unit has a certain amount of income that can be spent on the goods and services that the family wants and with which the family may compare itself to other families, or through which policymakers may compare families in an economy.

Such familiarity motivates the use of household datasets to understand incomes, poverty and inequality and to improve data literacy in the process, thinking about economies comprising households or individuals with certain levels of daily (or monthly) income that correspond more closely to how economists measure poverty and inequality, while also being the unit of analysis with which students themselves are more familiar.

Poverty

To understand poverty in an economy, it is standard practice to use the Foster, Greer and Thorbecke measures. Foster, Greer, and Thorbecke (1984) were engaged in the ongoing debate

about how to measure poverty in a way that was decomposable. That is, if we can separate all people who are poor into different groups (by gender, ethnicity, etc.), then a decrease (increase) in the poverty of one of the groups should result in a decrease (increase) in overall poverty (see Sen 1992).

The Foster-Greer-Thorbecke ($FGT(\alpha)$) formula provides a generalizable way to teach decomposable and relevant poverty measures. Poverty is indexed in the following way:

$$FGT(\alpha) = \frac{1}{N} \sum_{i=1}^q \left(\frac{z-y_i}{z} \right)^\alpha \quad (1)$$

Where N is the total population, z is the poverty line, y_i is the income of individual i , and q is the total number of individuals with incomes below the poverty line. $FGT(\alpha)$ provides three different measures of poverty depending on the value of α ¹: $\alpha = 0$ results in the poverty head count ratio; $\alpha = 1$ results in the poverty gap index, measuring the normalized distance of each poor individual's income from the poverty line; $\alpha = 2$ results in the poverty gap-squared index.

Notice, though, that equation 1 can be rewritten as equation 2 by factoring out z^α from the summation operator.

$$FGT(\alpha) = \frac{1}{Nz^\alpha} \sum_{i=1}^q (z - y_i)^\alpha \quad (2)$$

In my experience, rewriting $FGT(\alpha)$ as equation 2 allows for easier pedagogical outcomes for the following reasons. First, the instructor can separate the intuition of a gap (based on income less than the poverty line) from the practice of indexing for comparability across economies. Second, and importantly, when z^α is outside of the summation operator and with $\alpha = 1$, the summation provides the total poverty gap in local currency. This construction allows students to understand a first-order policy problem: how much it would cost to eliminate poverty.

The poverty headcount ratio counts the absolute number of poor individuals in an economy; that is, those whose incomes (y_i) are beneath the poverty line (z), and divides that total

by the size of the population to make it comparable across economies. FGT(0) falls between 0 and 1 and therefore measures the incidence of poverty or the proportion of the population that is poor. The poverty gap index sums all poor individuals' income gaps to find the total poverty gap and indexes it by the total population multiplied by the poverty line (Nz^α) to obtain a measure between 0 and 1 that policymakers can compare across economies. Students therefore learn an important aspect of data literacy and measurement: many aggregate measures must be indexed to permit comparison.

Students come to understand the extent of an individual's *depth* in poverty, the total poverty gap required to raise people out of poverty and therefore an economy's depth of poverty, and how to compare this depth across countries using FGT(1). For example, compare two countries, A and B, where A has a higher total poverty gap than B, which students may think implies that country A is "poorer" than country B. But if A has a much larger population than B, then A's FGT(1) will be lower than B's. A good example here would be comparing, say, the United States of America (A) and Botswana (B), for which this analogy would hold true. The total population of Botswana is smaller than the total number of poor people in the United States, but economists reasonably do not consider the United States to be poorer than Botswana. Similar insights apply to FGT(2) where students readily understand how it measures *the severity* of poverty by squaring an individual's poverty gap and thus weighting people further from the poverty line more heavily.

Inequality

Many ways to measure inequality exist in the economics literature. I focus on two measures: the Gini index and the Palma.² The Gini index is taught graphically as the difference between the Lorenz curve, or the cumulative distribution of income in an economy relative to cumulative

population, and the line of perfectly distributed income, divided by the total area beneath the curve of the line of perfect equality. Thus, the Gini (G) index is given by:

$$G = \frac{A}{A+B} \quad (3)$$

where A is the area between the Lorenz curve and the line of equality and A + B is the area beneath the line of equality, as illustrated in figure 1.

[Insert figure 1 about here]

The Gini index can also be measured with household and individual data (Pyatt, Chen, and Fei 1980) as follows:

$$G = \frac{cov(y, F(y))}{\bar{y}} \quad (4)$$

where *cov* is the covariance of income level *y*, the cumulative distribution of *y* given by *F(y)*, and \bar{y} is average income.

The relationship between the Lorenz curve and the Gini helps students to understand how the Gini depends on the extent to which those individuals or households at different parts of the cumulative income distribution deviate from equal income shares. But, the Gini index, by construction, is not particularly sensitive to changes at the tails of the income distribution.³ The Gini therefore fails to correspond to the intuitions that students have about the definition of inequality: that it should be *higher* for differences between the highest and lowest parts of the income distribution rather than differences within the middle of the distribution.

As an alternative, the Palma (P) measure (Palma 2011) is an intuitive measure of inequality in wealth or income:

$$P = \frac{Y_{10}}{\sum_{i=1}^4 Y_i}$$

where incomes shares (divided by decile) are ranked from lowest (Y_1) to highest (Y_{10}), the Palma takes the income share of the top 10 percent of the income distribution (Y_{10}) and divides it by the income share of the bottom 40 percent of the income distribution ($Y_1 + Y_2 + Y_3 + Y_4$). Cobham and Sumner (2013b) show that the Palma ranges from 0.757 (Belarus in 1988) to 15.081 (Namibia in 1993) using the World Bank's PovCal dataset. Cobham and Sumner argue that this measure captures the main insights of what people intuit when thinking about inequality and that it reflects the fact that the "middle 50%" (income shares 5 through 9 in 10 percent increments) has remained relatively stable over time (the Gini coefficient is sensitive to relatively small such changes, but the Palma is not). Cobham and Sumner's result shows how the Palma accords with a popular conceptualization of inequality: inequality is driven not by changes in income for those whose positions in the middle and upper-middle of the distribution are relatively stable, but by differences between those at the top and those at the bottom of the income distribution.

THE EXERCISE

The exercise unfolds in several steps. First, students are introduced to the idea of a 10-person artificial economy ranked by income from poorest to richest. Students use artificial economies to compute the measures of income, poverty and inequality. Students are then provided a 10-row (10 households) dataset extracted from a real-world economy and repeat the process of compiling income, poverty and inequality measures using MS Excel (see online appendix C.2). Students repeat the analysis with a full dataset of households. In this and subsequent sections, I refer liberally to the online appendix at

simondhalliday.com/data_literacy_in_economic_development/.

Table 1 presents the economies given to the students in the exercise (see online appendix B.2). Groups of 3 to 5 students are each given an economy corresponding to the rows in the

table. The economies (A through J) comprise 10 people with incomes given by the columns (P1 to P10). Each economy has a total daily income ranging from 35 monetary units to 100 monetary units. Ten people are used because each person corresponds to an income decile, which is necessary for later computation of the Palma and Gini measures.

[Insert table 1 about here]

Income

Students start by working in teams to compute daily total and per capita incomes for their economy (also converting them to annual measures). They then compare the economies across groups. The relevance of computing per capita income and *how* per capita income differs from individual income becomes immediately apparent, which is an insight students sometimes fail to grasp when first thinking about GDP in the abstract. Students should do this task *before* their class on poverty, typically at the end of a class on income and GDP.

Poverty

Having seen the income calculations in a previous class, students are then told that in each of the economies there are two poverty lines: \$2 per person per day and \$3 per person per day (corresponding roughly to the current \$1.90 and \$3.10 World Bank poverty lines). The purpose is to achieve four learning outcomes:

1. The choice of poverty lines is crucial to understanding *how much* poverty exists in an economy.
2. Two countries that have *an identical* income per capita (low or high) need not have identical rates of poverty.

3. Countries with *different* incomes per capita (low vs. high) need not have *different* rates of poverty. On the contrary, rates of poverty and income per capita may be unrelated.
4. Fiscal policy and redistribution of income can alleviate poverty.

Each group of 4 to 6 students receives the information about a given economy, computes the relevant poverty measures using the FGT(α) formula, and compares their values with members of other groups with different incomes. The results they obtain are summarized in table 2. They also consider the different scenarios for fiscal policy, altruism and aid outlined below.

[Insert table 2 about here]

Inequality

Students develop a deeper understanding of how to measure poverty by building up to each part of the Gini index formula. First, they compute equal shares of income and sketch a line of equality, which for 10 people will be 0.1, 0.2, and so on to 1. Second, they find each decile's income share. Third, they accumulate these shares and sketch them as a bar chart or line graph equivalent to a Lorenz curve to find the value of the area B (see online appendix C.4 for solutions). Fourth, they calculate the difference between the cumulative income and the line of equality to find the value of area A. Once they have calculated the values of areas A and B, they can use them to find the Gini index using equation 3. The values are shown in table 3 and illustrated in figure 2 using a bar chart to compute the Gini index for example economy D. Lastly, they use the income shares to compute the Palma. The process allows students to hone their data literacy through hands-on manipulation and visualization of their data.

[Insert table 3 about here]

[Insert figure 2 about here]

Individually, students may struggle with some of the steps in the exercise, but putting them in groups to discuss how to achieve the outcomes helps substantially. Most students are able to calculate equal shares of income and accumulate them. Finding the cumulative income shares for area B and using that to find area A challenges more students because it requires them to consider what a cumulative share—and what the Gini index itself—means *empirically*. Consider economy D as an example. The total daily income is 50. The bottom person’s income is 1, meaning their income share is $1/50 = 0.02$. If the poorest 10 percent were to have had an equal share of $5/50 = 0.1$, then they would require an additional $4/50 = 0.08$ of income. Therefore, the portion of area A (for the Gini index) for the lowest decile is $4/50$. The share of incomes needs to be computed for the 10 households *cumulatively*, after which students sum the 10 cumulative shares to find a total area of A. Area A must be divided by the total area beneath the line of equality resulting in the country’s Gini Index, as shown in table 4.⁴ Students also sketch Lorenz curves like those in figure 3 (see online appendix C.2). Most students do not have problems computing the Palma.

[Insert table 4 about here]

[Insert figure 3 about here]

Having completed these exercises, students are put in groups and re-paired with students from different country-groups to facilitate discussion of the following ideas: countries with the same/similar incomes can have *different* levels of inequality; countries with different total and per capita incomes can have *similar* levels of inequality; and, fiscal policy to redistribute income can address income inequality.

Extending the Exercise: Fiscal Policy, Altruism and Aid

Upon learning the preceding measures, students regularly want to think about how to “do something about it.” The artificial economies allow students to think about what role communities, governments and people could play in confronting poverty and inequality. Fiscal policy can take different paths. Using fiscal policy a government can provide transfers to people who have low incomes using tax revenue provided by those with higher incomes. Fiscal policy also could provide public goods from which people of different groups cannot be excluded (nonexcludable) and for which the benefits do not decrease as more people consume it (nonrival)—this provides a basis for talking about a government using progressive taxation and providing public goods with flat benefits. Similarly, students can discuss the effects of altruistic donations to others in the community or internationally through aid. Each of the examples creates different dynamics for considering the effectiveness of policy and the implementation of policy through local communities, state bureaucracies and international organizations with various levels of transactions costs.

PROCEEDING TO REAL-WORLD DATA

I use two kinds of data: aggregated national data from the World Bank Development Indicators (WDI) and a relevant household dataset. I have used three sets of household data so far: the Mexico Rural Household Survey (ENHRUM), the 2008/9 wave (wave 1) of the South African National Income Dynamics Study (NIDS) and the 2014/15 wave (wave 4) of NIDS as cross-sectional surveys. NIDS is a panel data survey that has been run consistently since 2008 (SALDRU 2016). The datasets are helpful for teaching the learning outcomes because of the variables (specified below) that allow analysis of incomes, poverty and inequality and how these measurements respond to government policy to redistribute income and alleviate poverty. Having encountered smaller datasets, students have the basic data literacy that is required to

move to larger datasets: understanding the components of data (observations and variables), using data to produce new variables, and using formulas to find aggregate properties of data and to visualize data.

In the dataset, students are given a spreadsheet with eight variables: a household identifier (hhid); household size (hhsz); household income net of remittances and government transfer (hhynet); government transfers received by the household (transfers); remittances received by the household (remittances); per capita net income (pcy), which is measured as net income divided by household size; the household's income rank from lowest to highest (rank(pcy)); and the distribution of the ranks (F(pcy)).⁵ The exercises apply equally well to other datasets that include these variables.

Prior to giving students the full dataset for problem sets and the midterm exam, students are first shown a 10-row sample dataset. The 10-row dataset, the idea of which is now familiar to students from the artificial economies, comprises a sub-sample of 10 households randomly selected from the full dataset where each household has the variables in the final dataset. We discuss how each of the measures we found for income, poverty, and inequality can be applied to the sub-sample dataset. Students can then re-apply these insights to the full dataset and think through the roles of government transfers and remittances.

Income

Students can easily learn the difference between a country's per capita GDP and the average incomes from a household survey. I ask students to identify how a country should be classified according to the World Bank Development Indicator (WDI) data in a given household survey year, e.g., 2008/9, and how that same country would be classified if the income classification were done based on the average incomes from a household survey. Students compute the local

country average annual incomes from the household data and convert the average income to PPP dollars for comparison with the WDI incomes. Students gain two skills from doing this: they recognize the difference between exchange rates without purchasing power parity, which is the number with which they are most familiar if they were to convert their own money to a foreign currency, and they realize that local prices and consumption must be catered for when comparing countries. Students develop an understanding of what it means to (linearly) transform data to enable comparison. Lastly, average incomes from household surveys often differ from the per capita income computed from GDP data, clarifying the difference between per capita GDP and household survey means. In-class discussion of how household surveys are conducted and how incomes are measured (under-reporting of top incomes, for example) also provides valuable lessons in data literacy: how social scientists think through how people respond in surveys, and the problems social scientists face in measurement.

Poverty

Having computed and compared incomes, students use the per capita income data to compute different values of the Foster-Greer-Thorbecke index to understand poverty. Students need familiarity with using mathematical formulas in Excel (addition, multiplication, etc.) and IF() statements. They need IF() statements to ensure that when they calculate the poverty measures, they count only individuals in households with income below the poverty line and they use mathematical formulas to compute the household's poverty gap and the household's poverty gap-squared (see online appendix C.5).⁶ Later, students repeat the analysis using transfers and remittances and see how anti-poverty policies can alleviate the incidence, depth and severity of poverty (see online appendices C.4 and E.2).

Inequality

Having examined poverty and the effects of transfers and remittances on it, students use the same data to compute a Gini index.⁷ Students use the covariance formula (equation 4) to compute the Gini index, which requires two pieces of information: first, the rank of each household in terms of the overall income distribution to form a distribution and, second, each household's per capita income. Using the formula for the NIDS 2008 data, for example, they would find a Gini index value of 0.6727, which is roughly consistent with more rigorous estimates of the Gini index using survey weights.⁸

Students can then proceed to analyze the effects of either remittances or government transfers on poverty and inequality (the steps are outlined in the supplementary materials). For example, using the 2008 NIDS data and including remittances and government transfers, the students find that the Gini decreases to 0.4787. This is an approximate decrease of 20 percentage points or a 40 percent decrease in inequality as a result of the transfers and remittances, showing the extent to which transfers and remittances alleviate inequality.

My students have generally navigated moving from the sample datasets to the real-world dataset with some struggles, but eventual ease. Moving from a 10-row dataset to a several-thousand-row dataset, students confront the challenge of moving around a large spreadsheet and realizing that it is easier to type in the values of the cells to which they are referring rather than highlighting cells when they write formulas. Furthermore, it is important for an instructor to show students shortcuts for navigating around a spreadsheet and how to apply formulas to many rows or columns.⁹

CONCLUSION AND EXTENSIONS

Using only five variables (household income, household size, government transfers, remittance receipts and a unique household identifier), students contrast the effects of government policy

and private remittances as anti-poverty and inequality-alleviating measures, as well as identifying the differences between national income measures (such as GDP and per capita GDP) with more granular measures of income, poverty and inequality.

The in-class exercises and the Excel-based problems provide the basis for data literacy as the architecture for team projects: thinking about observations and variables, producing data visualizations, and using the data to produce aggregate measures. Students glean the main understanding that, with a limited number of variables, they can deduce powerful insights about what the status quo is prior to a policy, such as government transfers to the poor, and what such a policy might achieve. Students went on to use data to tackle team projects on many topics, such as levels of awareness and knowledge about HIV/AIDS, malaria and TB in Nigeria, to the distribution of fertilizer investments in Malawi. Two projects highlight the skills that students gained, which they could transfer outside the class. In each case, the data literacy that they acquired enabled their investigations.

In the first project, a team of students used the China Health and Nutrition Survey to study Chinese health insurance, trying to understand the differences between rural and urban enrollment at two different time periods (2000 and 2009) and the mechanisms surrounding insurance in China. They began their investigation using spreadsheets where IF() statements allowed them to code dummy variables (1 for rural, 0 for not), to use a Pivot Table in MS Excel to evaluate the different choices to buy insurance (also a dummy variable they coded), and to produce corresponding bar charts. They therefore demonstrated basic data literacy skills in working with different observations, understanding variables, and creating aggregate properties of their data from their newly generated variables. They also explored mechanisms behind the differences they found, from changes in government expenditure to demographic trends.

In a second project, a team of students examined inequality in South Africa and they employed more directly the tools on poverty and inequality analysis. The students were interested both in the overall inequality in the country over time and, motivated by the country's history, the inequality within and among race groups in the country. They used two waves of the National Income Dynamics Study (NIDS) to assess inequality at two different time points in South Africa. Although they could not take account of the survey design in basic analysis in Excel (e.g., stratification and clustering), they could nonetheless draw conclusions about increasing within-race inequality in South Africa over time by comparing the Gini index values by race group using IF() statements, using the covariance formula for the Gini index, drawing Lorenz curves using new variables they created, and employing other economic analysis and data literacy learned in the course.

Filling a gap in the literature with respect to the teaching of data literacy in economic development, I hope to contribute to an ongoing discussion about experiential learning, transparency in research, knowledge and application of basic statistics, and the uses of empirics in undergraduate economics field courses. Future research should focus on the related ways in which activities such as these can impact learning directly, rather than on the construction and motivation of these activities such as I have provided here.

NOTES

¹ Although values of α larger than 2 can be used, their value in terms of instruction is limited.

Unless a student has taken political philosophy, for example, she would not understand that as α tends to infinity the value of the FGT index tends towards a Rawlsian Rule by which society judges poverty only on the basis of the poorest of the poor.

² Five axioms for inequality measurement are considered important in the literature: 1) The Pigou-Dalton transfer principle, 2) income scale independence, 3) Dalton's principle of population, 4) anonymity or symmetry, and 5) decomposability (see Cowell [2000] and Litchfield [1999]). Measures like the Atkinson index and the Theil Index (or the class of generalized entropy measures) are beyond this course.

³ See Palma (2011) and Cobham and Sumner (2013a, b) for a fuller explanation as to why the Gini index is insensitive to changes in the tails of the income distribution: the basic intuition is that substantial weight is given to deciles 2 through 5 of the income distribution and the tails are not given substantial weight in the Gini as each part of the distribution is equally weighted. This characteristic of the Gini is likely to be axiomatically true (Shorrocks 1980).

⁴ To give an accurate Gini coefficient, the cumulative income total should be divided by 5, which is the equivalent of the area of a smooth triangle. Note, though, that this produces a different Gini if a student cumulatively sums the heights of 10 columns, which would give a total area for A + B of 5.5. I accepted both answers in the class exercise, although the cumulative sum of columns results in a lower Gini coefficient.

⁵ Taylor and Lybbert (2015) include the Mexico Rural Household Survey in their supporting material for their textbook along with some exercises similar to those I ask students to perform

with these data. However, my exercises are grounded in initial exercises in-class as presented earlier, along with additional questions and ideas that Taylor and Lybbert do not include.

⁶ A variety of work has used the NIDS to understand South African poverty. See [Leibbrandt, Finn, Argent and Woolard](#) [2010], Leibbrandt and Levinsohn [2011], Finn and Leibbrandt [2013a], and Jansen et al. [2014].

⁷ Because each row corresponds to a household and we cannot easily manipulate this to create an individual-level dataset, we can compute only a household income-based Gini index. Although this is unfortunate from the perspective of consistency, in terms of pedagogy and learning outcomes the main concern is for students to think through what is required to find a Gini index and what the unit of observation implies about what gets computed.

⁸ I cannot provide the full NIDS data (for 2008 or 2014/15) as a spreadsheet, but this value is produced from the NIDS data in the spreadsheet that I can provide to those interested who request the data from the DataFirst at the University of Cape Town (www.datafirst.uct.ac.za). The Gini is found using the covariance formula as shown in the supplementary content. As with poverty, several authors have used the NIDS to examine inequality in South Africa and found results similar to those that I found, such as Leibbrandt, Finn and Woolard (2012), Finn and Leibbrandt (2013b), and Finn, Leibbrandt, and Levinsohn (2014). Admittedly, the measures should be weighted by their survey weights, but such a discussion is beyond the ambit of the course I teach.

⁹ Showing students shortcuts to highlight many cells at once, e.g., `CMD + SHIFT + DOWNARROW` on a Mac or `CTRL + SHIFT + DOWNARROW` on a PC with Windows, results in them saving time. I also recommend demonstrating to students how to double-click on the bottom corner of a cell for a formula to be applied to a whole column.

REFERENCES

- Acemoglu, D. 2013. Economic growth and development in the undergraduate curriculum. *Journal of Economic Education* 44 (2): 169–77.
- Allgood, S., and A. Bayer. 2016. Measuring college learning in economics. In *Improving quality in American higher education: Learning outcomes and assessments for the 21st century*, ed. R. Arum, J. Roksa, and A. Cook, 87–134. San Francisco: Jossey-Bass.
- Ball, R., and N. Medeiros. 2012. Teaching integrity in empirical research: A protocol for documenting data management and analysis. *Journal of Economic Education* 43 (2): 182–89.
- Bartlett, R. L., and P. G. King. 1990. Teaching economics as laboratory science. *Journal of Economic Education* 21 (2): 181–93.
- Brunnermeier, S. 2017. Learning by doing: The challenge of engaging undergraduates in economics research. *Journal of Economic Education* 48 (4): 290–94.
- Castilla, C. 2014. Field experiment in a course on behavioral economics: Nudging students around campus. *Journal of Economic Education* 45 (3): 211–24.
- Chen, E. K. 2005. Teaching and learning development economics: Retrospect and prospect. *Journal of Economic Education* 36 (3): 236–48.
- Cobham, A., and A. Sumner. 2013a. Is it all about the tails? The Palma measure of income inequality. CGD Working Paper 343. Washington DC: Center for Global Development.
- . 2013b. Putting the Gini back in the bottle? “The Palma” as a policy-relevant measure of inequality. Mimeograph. London: King’s College London.
- Cowell, F. A. 2000. Measurement of inequality. In *Handbook of income distribution*, ed. A. B. Atkinson and F. Bourguignon, 87–166. Amsterdam: North Holland.

- Diduch, A. M. 2012. Using the consumer expenditure survey to teach poverty measurement. *Journal of Economic Education* 43 (1): 99–106.
- Finn, A., and M. Leibbrandt. 2013a. The dynamics of poverty in the first three waves of NIDS. SALDRU Working Paper Series No. 119. Rondebosch, Cape Town, South Africa: Southern Africa Labour and Development Research Unit.
- . 2013b. Mobility and inequality in the first three waves of NIDS. SALDRU Working Paper Series No. 120. Rondebosch, Cape Town, South Africa: Southern Africa Labour and Development Research Unit.
- Finn, A., M. Leibbrandt, and J. Levinsohn. 2014. Income mobility in a high-inequality society: Evidence from the first two waves of the National Income Dynamics Study. *Development Southern Africa* 31 (1): 16–30.
- Gilleskie, D. B., and M. K. Salemi. 2012. The Cost of economic literacy: How well does a literacy-targeted principles of economics course prepare students for intermediate theory courses? *Journal of Economic Education* 43 (2): 111–32.
- Hansen, W. L. 2001. Expected proficiencies for undergraduate economics majors. *Journal of Economic Education* 32 (3): 231–42.
- Hansen, W. L., M. K. Salemi, and J. J. Siegfried. 2002. Use it or lose it: Teaching literacy in the economics principles course. *American Economic Review: Papers and Proceedings* 92 (2): 463–472.
- Hawtrey, K. 2007. Using experiential learning techniques. *Journal of Economic Education* 38 (2): 143–52.
- Hoyt, G. M., and K. McGoldrick. 2017. Models of undergraduate research in economics: Advice from eight exemplary programs, *Journal of Economic Education* 48 (4): 288–89.

- International Institute for Sustainable Development (IISD). 2017. *UN World Data Forum Bulletin*, January 21. 232 (1): 1–23.
- Jansen, A., M. Moses, S. Mujuta, and D. Yu. 2014. Measurements and determinants of multifaceted poverty in South Africa. *Development Southern Africa* 32 (2): 151–69.
- Jerven, M. 2013. *Poor numbers: How we are misled by African development statistics and what to do about it*. Ithaca, NY: Cornell University Press.
- Leibbrandt, M., A. Finn, J. Argent, and I. Woolard. 2010. Changes in income poverty over the post-apartheid period: An analysis based on data from the 1993 Project for Statistics on Living Standard and Development and the 2008 base wave of the National Income Dynamics Study. *Studies in Economics and Econometrics* 34 (3): 25–43.
- Leibbrandt, M., A. Finn, and I. Woolard. 2012. Describing and decomposing post-apartheid income inequality in South Africa. *Development Southern Africa* 29 (1): 19–34.
- Leibbrandt, M., and J. Levinsohn. 2011. Fifteen years on: Household incomes in South Africa. NBER Working Paper No. 16661. Cambridge, MA: National Bureau of Economic Research.
- Lima, V., and G. Tsiang. 2017. Training young researchers: Successful strategies from University of Chicago college economics. *Journal of Economic Education* 48 (4): 310–16.
- Litchfield, J. 1999. *Inequality: Methods and tools*. Text for World Bank's Web site on inequality, poverty, and socio-economic performance (no longer online). Washington DC: World Bank.
- MacGillivray, H. 2017. Data literacy: What, why and how? Data and statistics: The sciences, the literacies and collaboration. Presentation. UN World Data Forum, Cape Town, South Africa, January 16. <https://undataforum.org/WorldDataForum/wp->

content/uploads/2017/01/TA4.05_MacGillivray.UNWDF_MacGillivray_TA4_05.pdf
(accessed July 6, 2017).

MacGillivray, H., and L. Pereira-Mendoza. 2011. Teaching statistical thinking through investigative projects. In *Teaching statistics in school mathematics—Challenges for teaching and teacher education*. A joint ICMI/IASE study: The 18th ICMI study, ed. C. Batanero, G. Burrill, and C. Reading, 109–20. Dordrecht, The Netherlands: Springer.

Palma, G. 2011. Homogeneous middles vs. heterogeneous tails, and the end of the ‘Inverted-U’: The share of the rich is what it’s all about, *Development and Change* 42 (1): 87–153.

Pyatt, G., C-N. Chen, and J. Fei. 1980. The distribution of income by factor components. *Quarterly Journal of Economics* 95 (3): 451–73.

Sen, A. 1992. *Inequality Reexamined*. New York: Oxford University Press.

Shorrocks, A. F. 1980. The class of additively decomposable inequality measures. *Econometrica* 48 (3): 613–25.

Singh, P., and A. Russo. 2013. A dream experiment in development economics, *Journal of Economic Education* 44 (2): 158–68.

Southern Africa Labour and Development Research Unit (SALDRU). 2016. National Income Dynamics Study 2008 and 2014/15, Waves 1 and 4 [datasets]. Version 5.3. Cape Town, South Africa: SALDRU [producer]. Cape Town, South Africa: DataFirst [distributor].

Taylor, J. E., and T. J. Lybbert. 2015. *Essentials of development economics*. 2nd ed. Oakland, CA: University of California Press.

APPENDIX
List of Online Appendix Resources

- A. Class Context and Syllabus
 - 1. Syllabus for my class Economic Development (ECO211) in the Economics Department at Smith College [pdf](#)

- B. Example Economy Exercises
 - 1. Printout for students with example economies [pdf](#) and [docx](#)
 - 2. Income and poverty instructions for Students [pdf](#) and [docx](#)
 - 3. Inequality instructions for Students [pdf](#) and [docx](#)
 - 4. Guidelines for instructors for the exercises [pdf](#) and [docx](#)
 - 5. Solutions for instructors for example economies [xlsx](#)

- C. Excel Real-World Data Exercise Sample Documents
 - 1. Note: I cannot provide the data on a public site, a potential user can request the NIDS and upon that request I can share data and/or do files. The data are publicly available, but the user should preferably access the data herself at the DataFirst Research Unit.
 - 2. Sample sheet mirroring the data with 10 household rows [xls](#)
 - 3. Accompanying description in [docx](#) and [pdf](#)
 - 4. Solutions for instructors for sample data [xlsx](#)
 - 5. Additional sample economy with nested IF() statement as exercise [xls](#)

- D. Stata Files for Generating Spreadsheets
 - 1. Stata .do files for converting South African National Income Dynamics Study (NIDS) to a workable spreadsheet:
 - a. Working with 2008 NIDS: [nids_data_setup.do](#)
 - b. Working with 2014 NIDS: [nids2014.do](#)

- E. Additional Documents
 - 1. Example spreadsheet to teach students about **if statements** (the basics) [xls](#)
 - 2. Example spreadsheet for FGT(α) and understanding the poverty Head Count Ratio, Poverty Gap, etc. [xls](#)
 - 3. Example spreadsheet to teach students about **matching** and **indexing** [xls](#)
 - 4. Example spreadsheet to teach the **human development index** and using math in Excel: [xls](#)

- F. Problem Sets with Solutions and Examples
 - 1. Problem Set 1: [PS1](#), [PS1 Data xls](#), [PS 1 Data Solutions xls](#)
 - 2. Problem Set 2: [PS2](#), [PS 2 Data Solutions xls](#)
 - 3. Problem Set 3: [PS3](#), [PS3 Data xls](#), [PS 3 Data Solutions xls](#)

G. R Code for figures in the paper

I coded all of the figures in the paper in R. I provide links to the R code for the figures. This is intended for instructors or interested advanced students and should not be viewed as part of the exercise itself.

1. Figure 1: “The Lorenz Curve for Economy J is shown by the solid black line. The line of perfect equality is shown by the dashed line.” [Code](#)
2. Figure 2: “Bar chart to calculate Gini for economy D. Data are from table 3” [Code](#)
3. Figure 3: “Lorenz curves for artificial countries. Data are from table 3.” [Code](#)

TABLE 1: Individual Incomes in Each Artificial Economy with their Corresponding Daily Incomes (Y) and Per Capita Daily Incomes (PCY)

COUNTRY	PERSON										Y	PCY
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10		
A	1	1	1	2	2	3	4	5	6	10	35	3.5
B	2	2	2	2	2	3	3	5	6	8	35	3.5
C	1	1	1	1	2	2	2	10	12	18	50	5
D	1	2	2	3	4	4	4	8	10	12	50	5
E	1	1	1	1	6	6	9	10	15	20	70	7
F	2	3	3	4	4	4	5	6	9	30	70	7
G	1	2	3	3	4	4	6	12	20	30	85	8.5
H	1	1	5	5	8	9	10	12	16	18	85	8.5
I	2	2	3	5	7	9	13	15	18	26	100	10
J	1	1	2	2	8	8	14	15	19	30	100	10

TABLE 2: Poverty Measures for Each Economy at the Two Poverty Lines

COUNTRY	POVERTY LINE, $z = 2$						POVERTY LINE, $z = 3$					
	HC	PG	PG ²	FGT(0)	FGT(1)	FGT(2)	HC	PG	PG ²	FGT(0)	FGT(1)	FGT(2)
A	3	3	3	0.3	0.15	0.075	5	8	14	0.5	0.4	0.35
B	0	0	0	0	0	0	5	5	5	0.5	0.25	0.125
C	4	4	4	0.4	0.2	0.1	7	11	19	0.7	0.55	0.475
D	1	1	1	0.1	0.05	0.025	3	4	6	0.3	0.2	0.15
E	4	4	4	0.4	0.2	0.1	4	8	16	0.4	0.4	0.4
F	0	0	0	0	0	0	1	1	1	0.1	0.05	0.025
G	1	1	1	0.1	0.05	0.025	2	3	5	0.2	0.15	0.125
H	2	2	2	0.2	0.1	0.05	2	4	8	0.2	0.2	0.2
I	0	0	0	0	0	0	2	2	2	0.2	0.1	0.05
J	2	2	2	0.2	0.1	0.05	4	6	10	0.4	0.3	0.25

Note: HC = head count, PG = poverty gap, PG2 = poverty gap-squared, FGT(.) corresponds to the Foster-Greer-Thorbecke measure for each economy at the relevant poverty line ($z = 2, 3$) for each of $\alpha = 0, 1, 2$.

TABLE 3: Cumulative Income Shares for Each Economy Used to Compute the Gini Index

COUNTRY	CUMULATIVE SHARE OF INCOME									
	10	20	30	40	50	60	70	80	90	100
A	0.03	0.06	0.09	0.14	0.20	0.29	0.40	0.54	0.71	1.00
B	0.06	0.11	0.17	0.23	0.29	0.37	0.46	0.60	0.77	1.00
C	0.02	0.04	0.06	0.08	0.12	0.16	0.20	0.40	0.64	1.00
D	0.02	0.06	0.10	0.16	0.24	0.32	0.40	0.56	0.76	1.00
E	0.01	0.03	0.04	0.06	0.14	0.23	0.36	0.50	0.71	1.00
F	0.03	0.07	0.11	0.17	0.23	0.29	0.36	0.44	0.57	1.00
G	0.01	0.04	0.07	0.11	0.15	0.20	0.27	0.41	0.65	1.00
H	0.01	0.02	0.08	0.14	0.24	0.34	0.46	0.60	0.79	1.00
I	0.02	0.04	0.07	0.12	0.19	0.28	0.41	0.56	0.74	1.00
J	0.01	0.02	0.04	0.06	0.14	0.22	0.36	0.51	0.70	1.00
Equality	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.00

TABLE 4: Value of Area A at Each Decile, Providing the Total A Area for Each Economy and Therefore Each Economy's Gini Index

	AREA A AT EACH 10% SHARE										Total		
	10	20	30	40	50	60	70	80	90	100	Area A	Gini	Palma
A	0.07	0.14	0.21	0.26	0.30	0.31	0.30	0.26	0.19	0.00	2.04	0.41	2
B	0.04	0.09	0.13	0.17	0.21	0.23	0.24	0.20	0.13	0.00	1.44	0.29	1
C	0.08	0.16	0.24	0.32	0.38	0.44	0.50	0.40	0.26	0.00	2.78	0.56	4.5
D	0.08	0.14	0.20	0.24	0.26	0.28	0.30	0.24	0.14	0.00	1.88	0.38	1.5
E	0.09	0.17	0.26	0.34	0.36	0.37	0.34	0.30	0.19	0.00	2.41	0.48	5
F	0.07	0.13	0.19	0.23	0.27	0.31	0.34	0.36	0.33	0.00	2.23	0.46	2.5
G	0.09	0.16	0.23	0.29	0.35	0.40	0.43	0.39	0.25	0.00	2.59	0.52	3.33
H	0.09	0.18	0.22	0.26	0.26	0.26	0.24	0.20	0.11	0.00	1.82	0.36	1.5
I	0.08	0.16	0.23	0.28	0.31	0.32	0.29	0.24	0.16	0.00	2.07	0.41	2.17
J	0.09	0.18	0.26	0.34	0.36	0.38	0.34	0.29	0.20	0.00	2.44	0.49	5

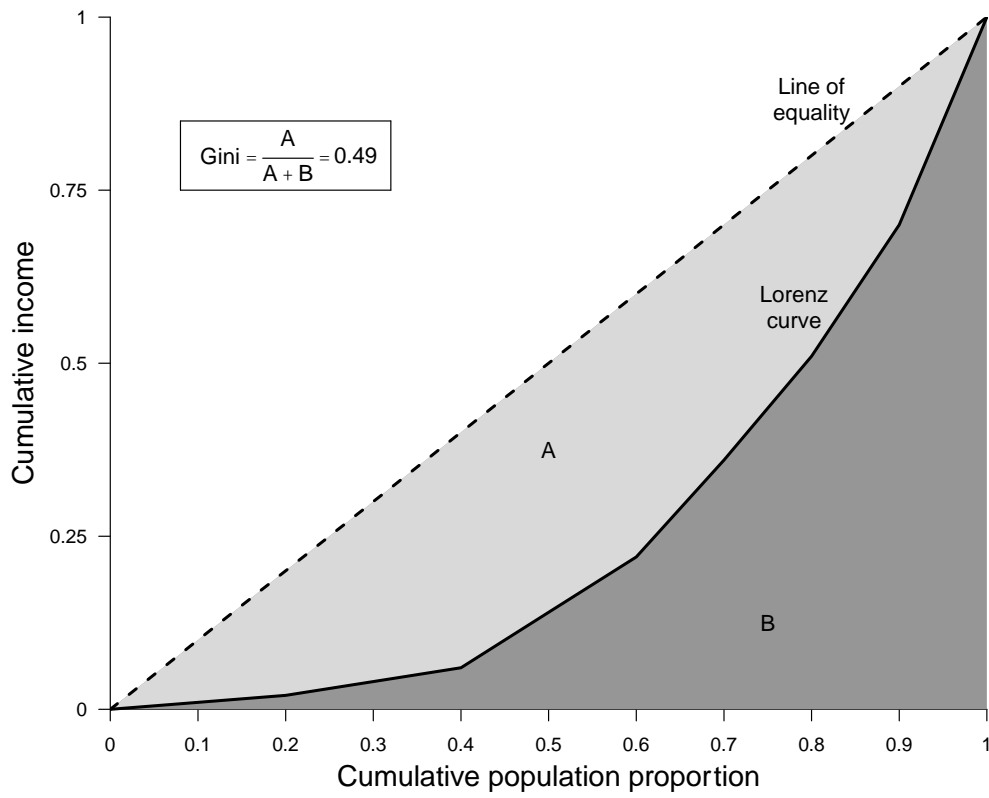


FIGURE 1: The Lorenz Curve for economy J is shown by the solid black line. The line of perfect equality is shown by the dashed line.

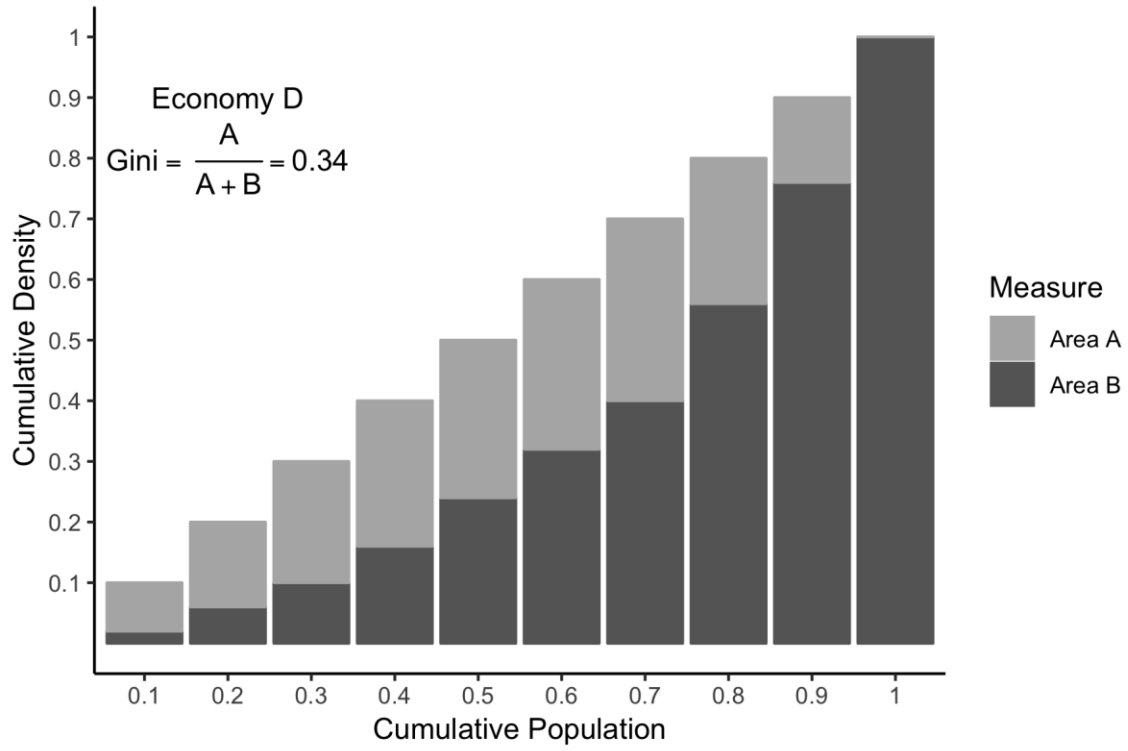


FIGURE 2: Bar chart to calculate Gini Index for economy D. Data are from table 3.

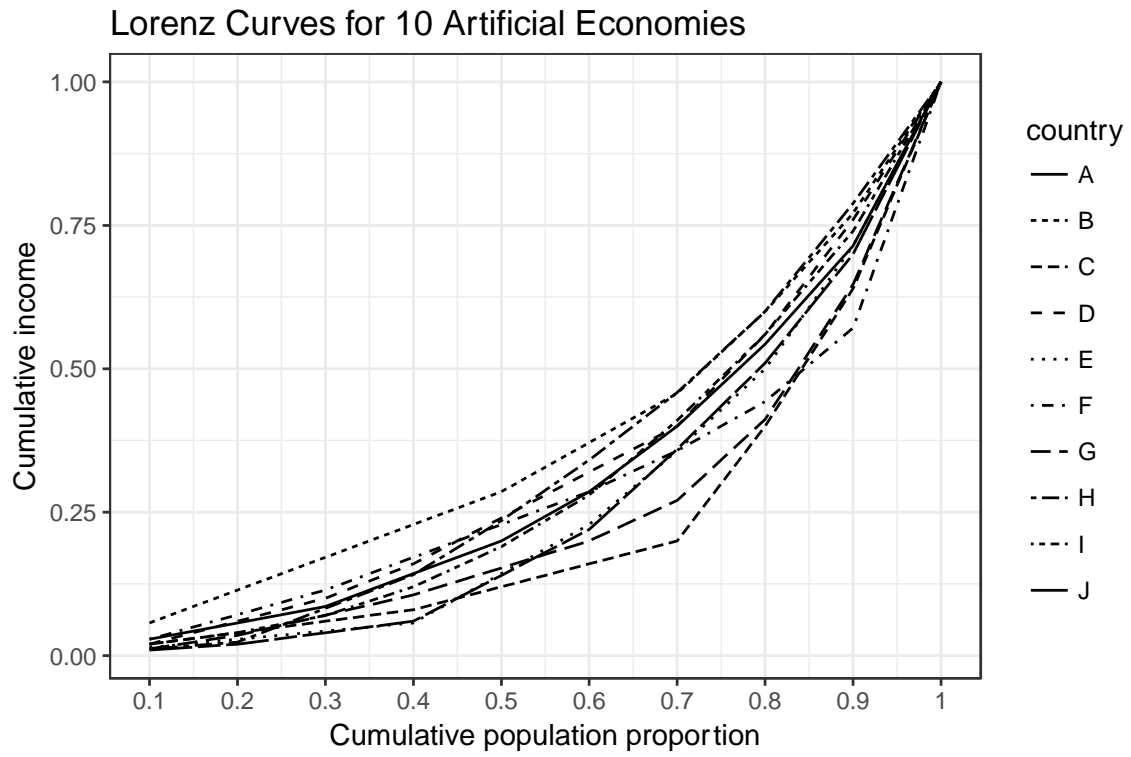


FIGURE 3: Lorenz curves for artificial countries. Data are from table 3.