Sub-Saharan Africa at a Crossroads: A Quantitative Analysis of Regional Development

Zachary C. Tyler and
Sucharita Gopal

AFRICA
2060

Sub-Saharan Africa at a Crossroads: A Quantitative Analysis of Regional Development

Zachary C. Tyler and Sucharita Gopal

Cover photo via iStockphoto.com.

Recent Publications

Issues in Brief series

Beyond the Resource Curse: Minerals and Global Development
Saleem Ali (No. 12), January 2010

The Changing Shape of Malnutrition: Obesity in sub-Saharan Africa
Arianna Fogelman (No. 11), December 2009

A Call for a Resilience Index for Health and Social Systems in Africa
Astier Almedom (No. 10), October 2009

Preferential Trade Agreements: "Free" Trade at What Cost?
Rachel Denae Thrasher (No. 9), September 2009

The Pardee Papers series

Sub-Saharan Africa at a Crossroads: A Quantitative Analysis of Regional Development
Zachary C. Tyler and Sucharita Gopal (No. 10), May 2010

Narcotics Trafficking in West Africa: A Governance Challenge
Peter L. McGuire (No. 9), March 2010

Community Targeting for Poverty Reduction: Lessons from Developing Countries
Moeed Yusuf (No. 8), February 2010

Linking Climate Knowledge and Decisions: Humanitarian Challenges
Pablo Suarez (No. 7), December 2009

Other Publications

Rio + 20: Another World Summit?
Miquel Muñoz and Adil Najam, *Sustainable Development Insights*
No. 2, November 2009

The Future of North American Trade Policy: Lessons from NAFTA
Pardee Center Task Force Report, November 2009

For a complete list and PDF versions of publications by the Frederick S. Pardee Center for the Study of the Longer-Range Future, visit www.bu.edu/pardee/publications.

Sub-Saharan Africa at a Crossroads:
A Quantitative Analysis of Regional Development

Zachary C. Tyler and Sucharita Gopal

Abstract

Sub-Saharan Africa is at a crossroads of development. Despite a quarter of a century of economic reforms propagated by national policies and international financial agencies and institutions, sub-Saharan Africa is still lagging in development. In this paper, we adopt two techniques using both qualitative (e.g., governance) and quantitative factors (e.g., GDP) to examine regional patterns of development in sub-Saharan Africa. More specifically, we examine and analyze similarities and differences among the countries in this region using a multivariate statistical technique, Principal Component Analysis (PCA), and an unsupervised neural network called Kohonen's Self-Organizing Map (SOM) to cluster levels of development. PCA serves as a tool for determining regional patterns while SOM is more useful for determining continental patterns in development. Both PCA and SOM results show a "developed" cluster in Southern Africa (South Africa, Namibia, Botswana, and Gabon). SOM exhibits a cluster of least developed countries in southern Western Africa and western Central Africa. The results demonstrate that the applied techniques are highly effective to compress multidimensional qualitative and quantitative data sets and to extract relevant information about development from a policy perspective. Our analysis indicates the significance of governance variables in some clusters while a combination of variables explains other regional clusters.
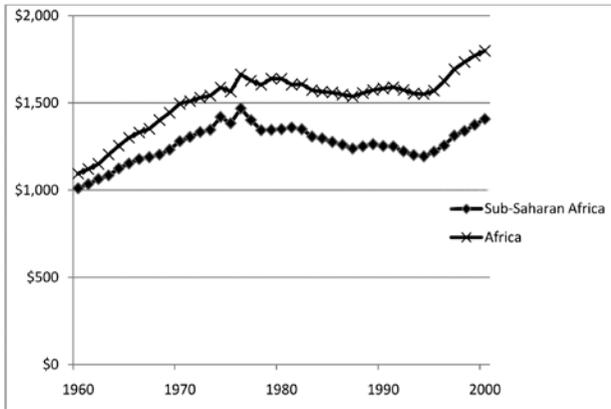
INTRODUCTION: AN OVERVIEW OF SUB-SAHARAN AFRICA'S DEVELOPMENT

Sub-Saharan Africa—that portion of the African continent south of the Sahara desert—includes 48 countries and one territory. It is historically referred to as "Black Africa," distinct from Arab North Africa. Before 1880, few areas of this region were under direct European power. By 1884, when the Berlin conference began, Africa was carved up and traded among various European powers including Britain, France, Belgium, Germany, Portugal, Spain, and Italy. By 1913, the entire continent, except Ethiopia, Liberia, and South Africa, was under European colonial rule. Decolonization began in 1957, and in the 1960s all countries except South Africa seemed to be set on a recovery stage. But development did not follow since many countries faced a difficult set of institutional and political problems as well as ethnic conflicts, secessionist movements, political corruption, and military governments. Once thought of as an area with huge potential for economic growth, sub-Saharan Africa is now at a crossroads, representing the poorest and least developed populations in the world and becoming the primary focus of international aid agencies. In 2001, sub-Saharan Africa had a poverty rate of 46.1 percent, the highest regional poverty rate in the world. This region represents 29 percent of the global population living on less than $1 per day (Go et al. 2007, 252). Figure 1 shows the trends of sub-Saharan Africa's GDP per capita compared with all of Africa from 1960 to 2000. Slight growth persisted from 1960 to the early 1970s, followed by an almost 20-year decline in economic growth. Figure 1 highlights the inability of sub-Saharan Africa to achieve significant economic growth over a 40-year period, leading many experts to question what the critical factors delaying or constraining development are.

> Once thought of as an area with huge potential for economic growth, sub-Saharan Africa is now at a crossroads, representing the poorest and least developed populations in the world and becoming the primary focus of international aid agencies.

Figure 1: GDP per capita for Africa 1960–2000

Source: Artadi and Sala-i-Martin (2003)

Prior studies have used a variety of factors to discuss development in sub-Saharan Africa including historical, governance, and economics. Some authors have examined the impact of colonial legacy (for example, Clague, Gleason, and Knack 2001) on democratization and development as well as political transitions. However, these factors alone cannot explain the lack of development. Countries such as Benin, Cape Verde, and Malawi have "undergone peaceful transitions from incumbent authoritarian regimes to democratic ones" (Lem 2005) without significant improvement in development. Economic factors, including GDP, debt, external aid, and other factors offer quantitative metrics for comparison (for example, Basu, Calamitsis, and Ghura 2000). The United Nations Development Programme (UNDP) recognized the importance of education and other social variables in development and therefore created the Human Development Index (HDI). HDI is an equally weighted index that includes economic growth (GDP per capita $US purchasing power parity or "PPP"), education (adult literacy rate and combined gross enrollment ratio), and health indicators (life expectancy at birth).

According to the United Nations, as of 2008 there were 26 countries in the world that qualified as countries with "low human development" (United Nations Development Programme 2008b). Low human development is defined as a country with an HDI value of less than 0.5. Of the 26 coun-

tries, all but one (Timor-Leste) are located in sub-Saharan Africa. Not one country in contiguous sub-Saharan Africa (in other words excluding island states) is considered to have "high human development." The country with the highest HDI rating located in contiguous sub-Saharan Africa is Gabon, with an HDI value of 0.729 (United Nations Development Programme 2008a). To put this HDI value in perspective, countries with similar HDI values are the Philippines, Paraguay, Sri Lanka, and Jamaica (United Nations Development Programme 2008a). It is clear that along with the highest rates of poverty in the world and stagnant economic growth, sub-Saharan Africa is also home to the least educated and least healthy populations in the world.

This paper examines regional patterns of development in sub-Saharan Africa. The data chosen for this research was carefully selected in consultation with experts familiar with African development. During this process, it became clear that experts use a wide range of variables, including economic, governance, and demographic variables, in their discussion of development. Therefore, it became necessary to assemble a dataset that included the above variables and select a methodology that would incorporate them in identifying regional classifications or clusters of development. Clusters are geographic concentrations of interconnected countries that share a similar set of development characteristics. Clustering offers development perspectives integrating a myriad of qualitative and quantitative variables. Clustering has been widely used in the literature, for example, to examine health outcomes (Hegyvary et al. 2008), foreign market potential (Cavusgil et al. 2004), and spatial differentiation (Scott and Garofoli 2007).

In looking at development patterns, we analyze similarities and differences amongst the countries in this region using two clustering techniques, Principal Component Analysis (PCA) and an unsupervised neural network called Kohonen's Self-Organizing Map (SOM), to cluster levels of development. PCA serves as a tool for determining regional patterns while SOM is more useful for determining continental patterns in development. Finally, integrating findings from both of these tools identifies features that are relevant in classification and in making the policy interpretations about development in sub-Saharan Africa.

DEVELOPMENT VARIABLES AND THEIR RELEVANCE

The World Governance Indicators, developed by the World Bank, are the first quantitative assessment of global governance. These indicators seek to capture very important qualitative measures of development, including perceptions, which are extremely difficult to quantify. According to Kaufmann, Kray, and Mastruzzi (2008, 1), "The indicators are based on several hundred individual variables measuring perceptions of governance, drawn from 35 separate data sources constructed by 32 organizations from around the world." The governance indicators are measured in units ranging from -2.5 to 2.5, with the higher positive values representing better governance. The values are constructed such that the governance indicators follow a normal distribution, with a mean of zero and a standard deviation of one in each period (Kaufmann, Kray, and Mastruzzi 2008, 16).

Without stable governance it is hard to maintain any significant economic growth or human development. This is particularly noticeable in the history of sub-Saharan Africa. Since the departure of colonialism, many sub-Saharan African states have been mired in corruption, overthrown governments, or a combination of both—such as Sudan, Nigeria, and the Republic of Congo.

World Governance Indicators are useful for broad cross-country comparisons and an analysis of trends over time. They cannot be used as a recommendation for particular governance reform in specific countries (Kaufmann, Kray, and Mastruzzi 2008, 9). Thus, for the purposes of our analysis, governance indicators are used mainly to identify the developmental status of nations.

Based on the 2004 World Governance Indicators, Bratton and Chang (2006, 1066), using an average of the five indicators, showed that 48 countries of sub-Saharan Africa consistently fell below the global mean. These authors did some supplementary survey analysis in conjunction with the governance indicators and found intriguing statistics with regards to perceptions of governance in sub-Saharan Africa. Using a survey known as the Afrobarometer, Bratton and Chang (2006, 1068) found that just 19 percent

of Africans believe that elected leaders look after their interests, or would listen to the average citizen[1].

The UNDP developed the HDI in the early 1990s, noting that there is more than economic growth that constitutes human development. The HDI is the most widely accepted index for human development around the world. The Human Development Report (United Nations Development Programme 2008b) highlights the following: "While economic growth is an important measurement of growth it is nonetheless limited in capturing how expanding income translates also into human development more broadly." The HDI incorporates education, through a mix of combined gross enrollment ratio and adult literacy rate; health, through life expectancy at birth; and economics, through GDP per capita, in order to determine a development index. It has been widely accepted that including education and health factors improves upon previous "development" projections based solely on economics.

Health indicators are extremely important to development globally, but are of particular interest in sub-Saharan Africa. The southern portion of sub-Saharan Africa, widely considered the most economically progressive, has been plagued with extremely high HIV rates since the early 1990s. In fact, South Africa, Botswana, and Namibia all have HIV rates of over 18 percent—among the highest on the continent. These three countries, along with Gabon, are the only countries in sub-Saharan Africa with an HDI of over 0.6. The HIV epidemic in this area of the continent has led to high mortality rates among productive age groups, which in turn leads to further impoverishment of these areas (Buve, Bishikwabo-Nsarhaza,

> Health indicators are extremely important to development globally, but are of particular interest in sub-Saharan Africa. The southern portion of sub-Saharan Africa, widely considered the most economically progressive, has been plagued with extremely high HIV rates since the early 1990s.

---

1. The survey results also explain the drastic variety of public perceptions of public officials in sub-Saharan Africa. The Afrobarometer shows that 77 percent of Africans believe that former President Nujoma of Namibia would rarely or never ignore the constitution. However, only 33 percent believe the same is true of former President Obasanjo of Nigeria.

and Mutangadora 2002, 2011). In other areas of the continent, where HIV prevalence is extremely low, life expectancy rates may be even worse than those areas with high HIV prevalence. In particular, the under five age cohort mortality rates (U5MR) are generally very high in countries such as Sierra Leone. Of the 30 million infants globally not receiving basic immunizations for diseases such as measles, tetanus, and malaria, 27 million live in countries with a GDP per capita of less than $1,200 (Jha et al. 2002, 2036). Both HIV and the U5MR statistics show the influence of health on development, and vice versa.

In an increasingly globalized world, access to information and communications technology (ICT) is becoming more and more a foundation of economic growth, and sub-Saharan Africa has limited access to various forms of technology. Access to telephones and cell phones is low in sub-Saharan Africa while access to basic electricity and the Internet is even worse. Buys et al. (2008) found that there is a significant digital divide within sub-Saharan Africa with regards to cell phone access. The average population with access to cell phones in an urban area is roughly 90 percent, whereas the average population with access in rural areas is only 20 percent (Buys et al. 2008, 1495). This dramatic difference highlights the challenges in addressing regional disparities both at national and sub-national scales.

The following variables are included in the present study:

- Governance Indicators 2007[2]
    - Voice and Accountability (V and A): Measures perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.
    - Political Stability and Absence of Violence (PS): Measures perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically motivated violence and terrorism.

---

2. The World Bank. (2009). *Aggregate Governance Indicators 1996–2008.* Retrieved June 5, 2009, from http://info.worldbank.org.ezproxy.bu.edu/governance/wgi/index.asp.

o Government Effectiveness (GE): Measures perceptions of the quality of public services, the quality of civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.

o Regulatory Quality (RQ): Measures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.

o Rule of Law (RoL): Measures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.

o Control of Corruption (CoC): Measures perceptions of the extent to which public power is used for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.

• Human Development Index

o GDP per capita (PPP $US) 2005[3]

o Combined Gross Enrollment Ratio (%) 2005[4]

o Adult Literacy Rate (% age 15 and older) 1995–2005[5]

o Life Expectancy at Birth 2005[6]

• Health

o HIV Prevalence (% of population aged 15–49) 2005[7]

o Under Five Mortality Rate (per 1000 live births) 2005[8]

o Access to Improved Water Source (%) 2006[9]

o Access to Improved Sanitation (%) 2006[10]

3. United Nations Development Programme. (2008). *Human Development Report 2007/2008.* Retrieved June 10, 2009, from http://hdr.undp.org/en/statistics/data/.
4. Ibid.
5. Ibid.
6. Ibid.
7. Ibid.
8. Ibid.
9. The World Bank. (2009). *Data and Statistics: Water and Sanitation.* Retrieved June 17, 2009, from http://go.worldbank.org.ezproxy.bu.edu/BQ5BEFBNX0.
10. Ibid.

- Technology
  - Mobile and Fixed Line Telephone Subscribers (per 100 people) 2007[11]
  - Electricity Consumption per capita (kWh)[12]
  - Internet Users (per 1,000 people) 2005[13]
- Trade
  - Imports of Goods and Services (% of GDP) 2005[14]
  - Exports of Goods and Services (% of GDP) 2005[15]
- Food
  - Food Consumption per capita (kcal/person/day) 2003–2005[16]
  - Food Production Index Numbers (PIN) base period 1991–2001 (2002–2004)[17]
  - Arable Land (1,000 Ha) 2007[18]
- Other
  - Freshwater Resources per capita (cumulative) 2007[19]
  - Foreign Direct Investment Net Inflows (% of GDP) 2007[20]
  - Population Density 2005[21]
  - Average Precipitation per year (mm/yr) 2003–2007[22]

---

11. The World Bank. (2009). *Data and Statistics: Information Technology.* Retrieved August 1, 2009, from http://go.worldbank.org.ezproxy.bu.edu/9Y5AECAM60.

12. United Nations Development Programme. (2008). *Human Development Report 2007/2008.* Retrieved June 10, 2009, from http://hdr.undp.org/en/statistics/data/.

13. Ibid.

14. Ibid.

15. Ibid.

16. Food and Agriculture Organization of the United Nations. (2009). *FAOSTAT.* Retrieved June 21, 2009, from http://faostat.fao.org/.

17. Ibid.

18. Ibid.

19. The World Bank. (2009). *Data and Statistics: Water and Sanitation.* Retrieved June 17, 2009, from http://go.worldbank.org.ezproxy.bu.edu/BQ5BEFBNX0.

20. United Nations Development Programme. (2008). *Human Development Report 2007/2008.* Retrieved June 10, 2009, from http://hdr.undp.org/en/statistics/data/.

21. Ibid.

22. Food and Agriculture Organization of the United Nations. (2009). *AQUASTAT.* Retrieved August 1, 2009, from http://www.fao.org/nr/water/aquastat/data/query/index.html.

METHODOLOGY

Rather than examining a multitude of country level indicators in isolation, clustering technique offers the opportunity to depict visually the similarities and differences between countries and explore the relationships between variables driving the composition of clusters. In this paper, we use two clustering techniques for reducing the complexity and exploring relationships between countries in sub-Saharan Africa. Both PCA and SOM are well suited for taxonomic problems like the clustering of countries based on a variety of development variables. The analysis was run in a Matlab programming environment. All of the island states (with the exception of Madagascar) considered to be part of sub-Saharan Africa were excluded, since they do not truly represent our study area. Somalia and Equatorial Guinea were excluded from our analysis due to a lack of data. The final database consisted of a total of 40 countries and 26 variables for each country.

Principal Component Analysis

Principal Component Analysis is a multivariate data reduction technique developed by Pearson (1901) and Hotelling (1903). The goal of PCA is to reduce the dimensionality of a large dataset while maintaining as much of the variance within the dataset as possible (Jolliffe 2002,1). PCA creates a reduced set of uncorrelated principal components that are linear combinations of the original set of variables. Basically, the extraction of principal components amounts to a variance maximizing (varimax) rotation of the original variable space. Note that as we extract consecutive components they account for less and less variability. The decision of when to stop extracting components basically depends on when there is only very little "random" variability left. The nature of this decision is arbitrary.

There are two major limitations in the PCA technique. First, several studies (Raftery et al. 2006; Fowlkes et al. 1988) have noted that the principal component direction is not designed to select the best directions for clustering and therefore may not provide directionality in interpreting variables in some contexts. The second limitation is the problem of treating high-dimensional data with low sample size, leading to difficulties in drawing inference (Baik et al. 2005; Johnstone 2001). These two limitations may

tend to bias our results. Hence our future work will address these limitations and compare PCA with better clustering techniques such as Structural Equation Models (SEM) to develop true explanatory models (Bollen 1989).

Principal Component Analysis can be visualized in two or three dimensions. Since we cannot visualize all of the components, it is important to visualize the principal component scores and weights to spot relationships among variables. This is generally most easily done in a two-dimensional visualization, such as a scatter plot of principal component one vs. principal component two, or any other combination of two principal components (Davies and Fearn 2004, 23). This type of visualization of principal components is useful in discussing feature selection in each cluster.

Self-Organizing Maps

In this paper, we have used unsupervised neural networks called Self-Organizing Maps (SOM) developed by Kohonen (1982, 2001) to find clusters and compare with PCA clusters. Biological neural networks of human learning inspire SOM. SOM is unique since it constructs a topology-preserving mapping from the high-dimensional input space onto a regular two-dimensional array of neurons or map units in such a way that relative distances between data points are preserved. The SOM can thus serve as a clustering tool of high-dimensional data. Because of its typical two-dimensional shape, it is also easy to visualize. Another important feature of SOM is its capability to generalize. In other words, it can "learn" from previously encountered inputs. Important parameters in the system are the learning rate, neighborhood radius, number of training steps, and number of neurons. SOM is widely used in statistics, signal processing, control theory, financial analyses, experimental physics, chemistry, and medicine. SOM has been used for finding patterns in diverse ecosystems (aquatic, forest, agriculture, etc.) (Lek and Guegan 2000; Recknagel 2003; Park et al. 2003). Kaski and Kohonen (1996) used a SOM to cluster countries based on 39 indicators for standard of living. There are distinct clusters; for example, almost all OECD nations are clustered together. Another example would be a cluster formed by primarily eastern European countries. The SOM offers a powerful method for exploring and visualizing geographic data and GIScience (Agarwal and Skupin 2008).

In this research, SOM is used to project similar vectors in the input space onto nearby neurons on the SOM map. Clustering of countries indicates similarities among them.

SOM can overcome some important limitations of conventional methods, including the two limitations of the PCA. Each variable is normalized using the data range of the variable and the data is presented using the parameters such as learning rate and learning steps/cycles. However, there are several challenges in running a SOM since it is difficult to determine the correct values of the parameters in the learning process, and the variations in learning. The "black box" structure of a neural network such as SOM is not conducive to statistical inference.

ANALYSIS AND DISCUSSION

PCA indicates that the top six principal components explained about 74 percent of the variance in the data. Results are shown in Table 1.

Table 1: Principal Component Weights

Principal Component 1	Weights	Principal Component 2	Weights	Principal Component 3	Weights
Government Effectiveness	0.2976	Exports	0.3557	Life Expectancy at Birth	-0.405
GDP/capita	0.2921	Adult Literacy Rate	0.3391	Internet Users	-0.3842
Rule of Law	0.2804	Regulatory Quality	-0.2759	Arable Land	-0.338
Telephone Subscribers	0.2778	HIV Prevalence	0.2569	Political Stability	0.286
Control of Corruption	0.2709	Rule of Law	-0.2525	Under 5 Mortality Rate	0.2781
Regulatory Quality	0.2625	Voice & Accountability	-0.248	Average Precipitation/year	0.2601
Kwh/capita	0.2597	Life Expectancy at Birth	-0.2401	HIV Prevalence	0.2226
Combined Gross Enrollment Ratio	0.2543	Freshwater Resources/capita	0.2325	Telephone Subscribers	-0.1972
Political Stability	0.2377	Access to Improved Sanitation	0.2262	Food Consumption/capita	-0.1942
Voice & Accountability	0.2298	Government Effectiveness	-0.2189	Voice & Accountability	0.1901
% Explained	**32.9688**	**% Explained**	**12.2532**	**% Explained**	**8.5851**
Principal Component 4	**Weights**	**Principal Component 5**	**Weights**	**Principal Component 6**	**Weights**
Life Expectancy at Birth	-0.4259	Freshwater Resources/capita	0.5085	Food PIN	-0.4112
Foreign Direct Investment	-0.3884	Average Precipitation/year	0.4677	Exports	-0.393
Imports	-0.3228	Imports	-0.436	Population Density	0.3673
Arable Land	0.3177	Food PIN	0.2518	Access to Improved H20	0.3461
Access to Improved Sanitation	0.2342	HIV Prevalence	-0.2088	Food Consumption/capita	-0.2894
Freshwater Resources/capita	-0.2308	Telephone Subscribers	0.1916	Arable Land	-0.2824
Exports	-0.2168	Under 5 Mortality Rate	0.1865	Average Precipitation/year	0.2398
HIV Prevalence	0.2144	Food Consumption/capita	0.1764	Freshwater Resources/capita	0.1944
Population Density	0.2103	Voice & Accountability	0.1571	Foreign Direct Investment	-0.1872
Food PIN	0.1837	Arable Land	0.1329	Under 5 Mortality Rate	-0.1799
% Explained	**7.6391**	**% Explained**	**6.7579**	**% Explained**	**6.0387**

Table 1 highlights the importance of all of the different variables chosen for this analysis. All variables exhibit some significance in at least one of the first six principal components, suggesting that all of the selected variables are important in defining development in sub-Saharan Africa. The use of PCA enables the integration of qualitative and quantitative factors in studying development. In addition, PCA enables comparison of countries with their cohorts in each cluster. Finally, PCA identifies prototypes or exemplars in each cluster that best represent the features of the cluster. Thus, PCA is useful in planning developmental strategies in sub-Saharan Africa and in identifying clusters of regional development as well as in identifying factors that best exemplify the cluster characteristics.

Next, we describe the first two components of the PCA and prototype of each component to shed more light on development in sub-Saharan Africa. The remaining clusters and weights are shown in Table 1.

Principal Component 1

The first principal component accounts for almost 33 percent of the variance in the dataset. The first principal component captures the following variables: Government Effectiveness, Regulatory Quality, Rule of Law, Combined Gross Enrollment Ratio, GDP per capita, electricity consumption per capita, and telephone access. Thus the first principal component is capturing governance, socioeconomic status, and access to technology—all critical to development in sub-Saharan Africa. Figure 2 shows that only a small set of countries score well on the first principal component; most countries score between 1 and -5. Thus, the first component indicates that much of sub-Saharan Africa scores low on many development indices, such as the HDI. South Africa has the highest score in terms of the first principal component.

The Case of South Africa

As shown in Figure 2, Bostwana (BWA) and Namibia (NAM) are clustered with South Africa (ZAF), albeit with significantly lower scores. In terms of governance, South Africa scores above zero in all of the governance indicators. It is the highest ranked country in three governance variables: Voice and Accountability, Government Effectiveness, and Regulatory Quality. Prior research (Sachs 2004, Kaufmann, Kray, and Mastruzzi 2005, 37) notes

Figure 2: 1st and 2nd Principal Components Plotted Country Scores
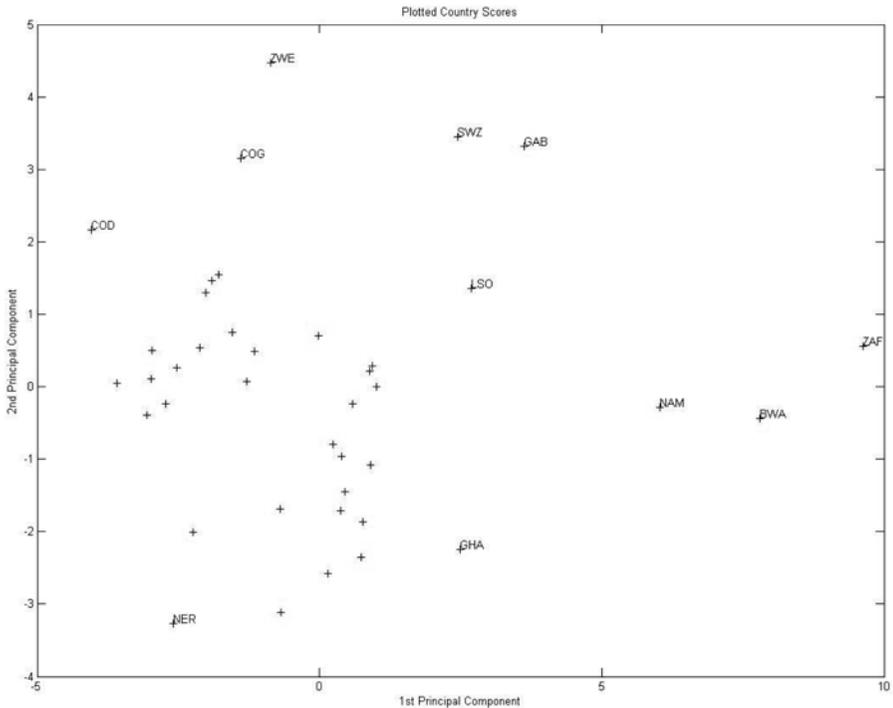
that there is a causal relationship between income and governance. Higher governance leads to higher incomes, with the largest impact being felt in the long run (Kaufmann, Kray, and Mastruzzi 2005, 37). South Africa had the second highest GDP per capita of $11,110 ($US PPP) in 2005, trailing only Botswana. Thus, in the case of South Africa there appears to be a correlation between governance and economic wellbeing.

In 1994 the African National Congress (ANC) of South Africa implemented a set of neoliberal economic policies that led to increased privatization of public enterprises, the liberalization of trade markets, and subsequently increased exports for the country (Habib and Padayachee, 2000). As it so happens, all of these economic reforms took place as South Africa was transitioning into a democracy and, compared to the previous years, improving governance.

Another significant portion of South Africa's development took place during the transition to a democracy. From 1990 to 2007, over five million households gained access to electricity in South Africa (Bekker et al. 2008, 3125).

South Africa is now set to have all of its citizens gain access to electricity by 2012, a stunning accomplishment considering the status of the country two decades ago.

Electrification was pushed to the forefront of the country's agenda beginning in 1992, and really took hold from 1994 to 1999. The primary reason for this push was that the ANC made electrification a priority along with access to water, education, and health (Bekker et al. 2008, 3128). The country is now set to have all of its citizens gain access to electricity by 2012—a stunning accomplishment considering the status of the country two decades ago.

The first component is a true measure of "development" on many levels. A combination of many attributes positions South Africa towards a high score in terms of this component. Thus, South Africa is a successful example of development, particularly in comparison to the rest of sub-Saharan Africa.

Principal Component 2

The second principal component represents roughly 12 percent of the variance in the dataset. Similar to the first, this component calls on a wide range of variables including Regulatory Quality, Rule of Law, Life Expectancy at Birth, Adult Literacy Rate, HIV Prevalence, and Exports. This component represents the formulation, implementation, and quality of policies with regards to governance. The second component also accounts for socioeconomic indicators (health and education). Finally, this component captures the percent of a country's GDP due to exports. Zimbabwe (ZWE) is the country that most strongly exemplifies the characteristics of the second principal component.

Zimbabwe

According to the World Bank's Governance Indicators, Zimbabwe has the third worst Regulatory Quality ranking in the world, trailing only Somalia and North Korea (The World Bank 2009a). As we can see from the Principal Component Weights in Table 1, Regulatory Quality has a negative weight; thus Zimbabwe's

extremely negative score for Regulatory Quality actually contributes to its high ranking within the second principal component. Of the countries analyzed here, Zimbabwe scores the lowest in the Rule of Law variable, also negatively correlated with the second principal component.

Zimbabwe is a true testament of the extent to which the AIDS crisis has impacted sub-Saharan Africa. In 2003, 19 percent of Zimbabwe's children were orphans, with four-fifths directly orphaned by this epidemic (Howard et al. 2006). The life expectancy in Zimbabwe has decreased from 63 years to 39 years in the past decade (Howard et al. 2006). This statistic alone highlights the compounding impact on human development of poor governance, an economic crisis, and a very high prevalence of HIV. As Howard et al. (2006) note:

> "Zimbabwe's AIDS epidemic feeds and is fed by an economic meltdown marked by 70% unemployment, triple-digit inflation, a shattered agriculture sector, drastic cuts in social spending, and political uncertainty and paralysis [5–7,12–15]. In 2002, 49% of Zimbabweans were in need of emergency food aid, ranking worst in southern Africa [7, 16]. The country's poverty rate has doubled since 1995 [17], and its 50 percent increase in under-5 mortality since 1990 is the largest in the world [18]. Isolated by Western donors critical of its government's human rights record, Zimbabwe receives a tiny fraction of foreign aid to the region [18]."

The statistics related to Zimbabwe's socioeconomics and health status speak for themselves. The country has truly been in tatters for the past decade, and in turn the country stands as a lone cluster in respect to the second principal component. The combination of poor governance, economic collapse, and a dominant AIDS epidemic is truly unmatched in the rest of sub-Saharan Africa.

DEVELOPMENT CLUSTERS IDENTIFIED USING SOM

SOM is a clustering algorithm that takes into account the entire dataset and clusters the countries based on varying weights and distances between the data points. SOM allows for broad country-wide analysis through

clustering, which can aid in formulating development decisions. A continent such as Africa often has so many different phenomena affecting development not just in a country, but in a region. Civil war refugees, HIV, malaria, and natural resource conflict are just some of the issues that affect regional development. SOM classification allows us to capture some of the regional development issues described below.

Sub-Saharan Africa Country Clusters

Figure 3 (see page 24) shows the country clusters produced from the SOM algorithm. The SOM was created using eight clusters with the entire dataset. The SOM parameters, including the shape, number of neurons or nodes, learning rate, and number of simulations for learning, are selected based on preliminary simulations. For instance, in this analysis 40 clusters could have been created, and each country could be in its own cluster. After some experimentation, eight clusters were chosen as they seemed to accurately represent the differences among countries within sub-Saharan Africa. Figure 3 highlights two findings. First, the clusters are geographically located. Southern Africa, Western Africa, eastern Central Africa, and western Central Africa all are geographically clustered, with a few exceptions. Second, Ghana is a sole member of one cluster. We examine the two findings in detail below.

The Upper Echelon of Development: South Africa, Botswana, Namibia, and Gabon

This cluster captures the highest development among the 40 countries that were included in this analysis. This SOM cluster is strong on the governance indicators, combined gross enrollment ratio, GDP per capita, food consumption per capita, under five mortality rate, and telephone access. South Africa, Botswana, and Namibia all have very good governance across all of the governance indicators. As a matter of fact, they are the only three countries that score above zero (on a scale of -2.5 to 2.5) on all six of the governance indicators and by far have the highest GDP per capita statistics in sub-Saharan Africa. By no coincidence, these three are the only countries in sub-Saharan Africa with an HDI above 0.6 (United Nations Development Programme 2008a). This is primarily due to their high educational

statistics and GDP per capita. Life expectancy, with the exception of Gabon, reflects the average that may be due to the prevalence of HIV. However, even with relatively average life expectancy at birth, all three countries have a relatively low U5MR. This can most likely be attributed to the fact that all the countries are economically wealthy and can allocate resources to the health sector. Finally, these countries all have high telephone access—once again most likely because of their economic success. South Africa has by far the highest telephone access in sub-Saharan Africa, with 98 out of 100 people subscribing to a mobile or land line telephone account (The World Bank 2009b).

One thing to note pertinent to this cluster is SOM's ability to capture non-linear relationships among the inputs that seem to lead to a balancing act of factors defining this cluster: a lack of one factor seems to be balanced by the presence of other factors in the country. Gabon is included in this cluster due to a balancing act of its variables. Gabon has a much lower HIV prevalence and a much higher life expectancy than South Africa, Botswana, and Namibia. However, Gabon also has scores much lower on the governance indicators than the other countries. Most of the countries' variables are extremely similar (within a cluster), and the countries make up for a lack of one variable with the prevalence of another.

The Low Road to Development: Angola, Central African Republic, Chad, Congo, Democratic Republic of the Congo, Côte d'Ivoire, Guinea, Guinea-Bissau, Liberia, and Sierra Leone

If the Southern cluster is at the highest rung of development in sub-Saharan Africa, then the Western and Central African cluster is at the bottom rung. All of the countries listed above, with the exception of Congo, are considered low development countries by the UNDP's Human Development Report (United Nations Development Programme 2008b). In fact, Liberia, the Democratic Republic of the Congo, Central African Republic, and Sierra Leone—in that order—are the bottom four of the HDI list. This cluster calls strongly on five of the six governance variables (no Voice and Accountability), all of the HDI variables, HIV prevalence, electricity consumption, and telephone access.

All of the countries in this cluster have extremely poor governance due to a combination of factors, including the prevalence of military coups. According to the *Encyclopedia of Earth*'s country profiles, the following countries within this cluster have experienced military coups: Guinea, Guinea-Bissau, Liberia, Democratic Republic of the Congo, and Central African Republic (Boukerrou, 2009). Along with military coups, this cluster has been mired in other forms of corruption and poor governance for decades. Another example of poor governance is the civil unrest created from conflict concerning diamonds. As Olsson (2007, 268) notes, "Fighting over diamond deposits is believed to have been an important reason for the initiation, maintenance, and prolonging of civil unrest in Angola, Sierra Leone, Liberia, and the Democratic Republic of the Congo."

This cluster is relatively poor, the richest country being Angola, with a GDP per capita of $2,335 in 2005. The poorest country in all of sub-Saharan Africa, Liberia, is also part of this cluster, with a 2005 GDP per capita of $400 (United Nations Development Programme 2008b). While this cluster shows relatively low HIV prevalence, it still exhibits the lowest life expectancies in all of sub-Saharan Africa; this can be attributed to the combination of poor governance and a lack of economic resources to combat health issues. Finally, this cluster has very limited access to technology, like the majority of sub-Saharan Africa.

The "Hope" Prescription: Benin, Burkina Faso, Gambia, Madagascar, Mali, Mauritania, and Senegal

This cluster includes some interesting variables in comparison to the others, including freshwater resources per capita, access to improved water, arable land, GDP per capita, and electricity consumption. This cluster clearly emphasizes environmental attributes and economic growth indicators as well as the relationship of the two.

This group has very few freshwater resources per capita, as well as limited arable land available for large scale agricultural production. Its population base has above average access to an improved water source. It has average governance, which could be a contributing factor to the

above average access to improved water and very high life expectancy in relation to the rest of sub-Saharan Africa. It also has an average GDP per capita for sub-Saharan Africa, and has relatively low access to technology, particularly electricity.

This cluster offers some hope for the future based on its governance, which is about the average. If governance improves, one could expect that its educational statistics may also improve. The life expectancy in this cluster is high relative to the rest of sub-Saharan Africa. Thus, its population structure is better than most countries in sub-Saharan Africa, making it well-positioned for a better future trajectory.

What Can Low Life Expectancy Do to the Educated?: Cameroon, Kenya, Malawi, Mozambique, Rwanda, Tanzania, Uganda, and Zambia

This group of countries, with the exception of Cameroon, is located on the eastern coast of central sub-Saharan Africa. The cluster calls on five of the six governance indicators (Control of Corruption not included), freshwater resources per capita, three of the four HDI variables (combined gross enrollment ratio not included), and exports.

This cluster has average governance overall, and is not nearly as poor as the western portion of Central Africa. It has low life expectancy, particularly in Zambia and Mozambique, which have a very high prevalence of HIV. The low life expectancy in other countries could be because there are very few freshwater resources per capita, as well as very little arable land available. This is important to note because most of the citizens in sub-Saharan Africa survive on subsistence farming.

This cluster is similar to others with poor governance. That is, the cluster is relatively poor, with a relatively low life expectancy, and very little access to technology. The thing that sets this cluster apart is that all of the countries have populations that are fairly well educated. The combination of being clustered towards the coast, having educated populations, and being geographically close to the upper echelon of development in sub-Saharan Africa shines some hope on this cluster.

The Landlocked Syndrome: Lesotho and Swaziland

These countries appear destined to be clustered with one another. Both Lesotho and Swaziland are small landlocked countries within South Africa. Many of their attributes are the same, and thus it is no surprise that they are clustered together. This cluster calls on all of the governance indicators, access to an improved water source, access to improved sanitation, GDP per capita, combined gross enrollment ratio, and telephone access.

Lesotho and Swaziland are very similar to the Southern African cluster, with a few very important exceptions. First, these countries have average governance, not the very good governance of the southern cluster. Second, the life expectancy of these countries is among the lowest in sub-Saharan Africa. The primary reason for this is HIV prevalence, with poor governance perhaps compounding the HIV issue. These countries have some of the economic resources necessary to combat this disease just as countries like South Africa and Namibia have done. The difference between Lesotho and Swaziland and the southern cluster is that Lesotho and Swaziland lack good governance. This seems to show a negative correlation between governance and life expectancy. If this correlation is significant, it is quite possibly a large reason for the low life expectancy in these countries in comparison to the Southern African cluster. These countries have good access to clean water sources, good access to sanitation, and educated populations. It appears the only thing that they lack to combat HIV is good governance.

Flying Solo: Ghana

Within our SOM clustering, Ghana is the only country that is in a cluster of its own. This cluster calls very strongly on all six of the governance indicators. These indicators are weighted almost twice as heavily as every other variable in this cluster. Ghana appears to be the only country in our analysis that is about average in each of the six governance indicators. That is, Ghana scores around zero for every indicator, in which the scores range from -2.5 to 2.5; it is near the average of all of the variables within our analysis, with the exception of a few. The exception variables include: life expectancy, food consumption per capita, the percentage of GDP which is created from imports, and the percentage of the population with access to improved sanitation. Ghana appears to be slightly more developed than

most sub-Saharan African countries. It is clustered by itself because of its unique combination of attributes. For example, of Ghana's population, 80 percent has access to an improved water source while only 10 percent of the population has access to improved sanitation (The World Bank 2009c). This is a level of disparity that is not displayed by any other country in sub-Saharan Africa with regards to these two variables.

The Deep Footprints of Unstable Governance: Sudan, Nigeria, Burundi, and Zimbabwe

This is one of the few clusters that is not necessarily geographically clustered. In other words, Sudan, Nigeria, Burundi, and Zimbabwe are all spatially separated. This cluster calls on the governance indicators more than any of the other variables. Similar to the Western and Central African cluster, these countries have been mired in corruption, civil war, and all around poor governance for decades, as exemplified by the Darfur conflict in Sudan. Nigeria is mired in corruption dealing with its natural resources; Zimbabwe has extremely poor governance and some of the worst food security in the world; and Burundi is a victim of spillover effects from the Rwandan genocides in the 1990s. Burundi's troubles exploded when its first democratically elected leader was assassinated after just 100 days in office in 1993 (Boukerrou 2009).

This cluster differs from the Western and Central African cluster in a couple of ways: it has access to technology and a high literacy rate that is not found in the Western and Central African cluster. While these two attributes make this cluster slightly better off than the Western and Central African cluster, the poor governance here will continue to hinder its development as long as it exists.

The Curious Case of Life Expectancy: Eritrea, Ethiopia, Niger, and Togo

This cluster is not geographically clustered like most of sub-Saharan Africa. Governance is an important factor in this cluster, weighted heavily like most of the others. Also weighted heavily in this cluster are access to improved sanitation, combined gross enrollment ratio, food consumption per capita, under five mortality rate, and telephone access.

Figure 3: SOM Classification—8 Clusters

SOM Classification: All Variables

Clusters

☐	The Low Road to Development
▤	The Curious Case of Life Expectancy
▨	The Hope Prescription
▦	Flying Solo: Ghana
▨	The Deep Footprints of Unstable Governance
vvv	What Can Low Life Expectancy do to the Educated?
xxx	The Landlocked Syndrome
▨	The Upper Echelon of Development

Figure 4: SOM Classification—8 Clusters with No Governance Indicators

SOM Classification: No Governance

Clusters

- ///// The Upper Echelon
- x x x A Cluster of Their Own: Congo and Gambia
- vvvv Health and Longevity Dominate
- \\\\ Uneducated and High U5MR
- ░░░ Economically Well Off with Low LE
- //// Educated with Low LE
- ≡≡≡ Low LE, FC/Capita, and High U5MR
- ☐ Very High U5MR, Low LE, and Lots of H20

0 205410 820 1,230 1,640 Miles

This cluster, like so many other clusters, has extremely poor governance. What is interesting here is that Eritrea, Ethiopia, Niger, and Togo all have among the highest life expectancies in sub-Saharan Africa, in contrast to the general trend, where poor governance is generally accompanied by a low life expectancy. This is particularly true when there are very few economic resources, as is the case in this cluster. Poor governance, poor economics, and an uneducated population described many of the SOM clusters in this analysis. This cluster also has by far the lowest telephone access. While not many sub-Saharan African countries have access to electricity or the Internet, many do have access to mobile phone subscriptions. These four countries are behind the curve in terms of any sort of ICT access.

The Dominance of Governance in SOM Clustering of sub-Saharan Africa

Figure 4 (see page 25) shows SOM clustering with all of the variables used in the previous analysis, with the exception of the six governance indicators. A visual comparison of Figures 3 and 4 indicate the significance of governance in SOM clusters.

A distinct feature of Figure 4 is the lack of spatial or geographic clustering. The cluster in Western Africa, consisting of Benin, Burkina Faso, Gambia, Madagascar, Mali, Mauritania, and Senegal, is now separated. The same is true for the cluster that was located in eastern Central Africa and composed of Cameroon, Kenya, Malawi, Mozambique, Rwanda, Tanzania, Uganda, and Zambia. The fact that these clusters are split up highlights the geographic importance of the governance indicators. There are a number of reasons why governance would be important regionally. Africa has a history of neighborhood conflicts, which would bring down the governance scores. War refugees often travel to neighboring countries when they have nowhere else to go. This generally leads to a very poor situation for the refugees, as neighboring governments are not prepared to deal with such issues. For example, by 1995 Rwanda had produced 1.7 million refugees, Liberia 750,000, Somalia 450,000, and Sudan 450,000 (Kalipeni and Oppong 1998, 1637). Kalipeni and Oppong (1998, 1642) note that there is an interesting phenomenon with African refugees, in which countries who are harboring neighboring countries' refugees are often sending their own refugees into the country whose citizens they are harboring. An example of

this trans-border refugee movement was in 1992 when Sudan was harboring 730,000 refugees, while Sudanese refugees sought safety in Ethiopia, Uganda, and Kenya. Situations like these would lead to governance problems that would result in large SOM spatial clusters of neighboring countries.

While there are clusters that are different between the SOM maps, there are also clusters that remain the same. Within these clusters, it appears that, while governance may be important, the similarities among all other variables will keep countries clustered together. For example, the Southern Africa cluster scores higher than the rest of sub-Saharan Africa in so many of the variables that the same four countries remain clustered even when eliminating the governance indicators. At the other end of the spectrum, the cluster displayed in white in both Figures 3 and 4 remains largely the same. This cluster scores so poorly on so many variables that most of the countries within this cluster stay the same even without the governance indicators. Thus, in some contexts SOM cluster shape and size can remain constant even when governance is eliminated.

CONCLUSION: HOW CAN CLUSTERING HELP IN FORMULATING POLICY IN SUB-SAHARAN AFRICA?

Both PCA and SOM are useful as broad-based policy tools. PCA allows us to view how countries score on a specific combination of variables. In this analysis, many of the principal components showed combinations of specific sets of variables, such as water or health indicators. In this context, PCA may be used to target specific sets of development variables that can move a country in a more positive trajectory. But results from PCA must be further improved using factor rotation technique and verified using Structural Equation Models (Bollen 1989 and Goodma 1974). Such an approach will lead to both confirmatory and exploratory modeling and may ultimately lead to theory development.

SOM may be considered a nonlinear generalization of PCA (Yin 2007). SOM describes a mapping from a higher dimensional input space to a lower dimensional map space. SOM thus allows for the analysis of clusters based on all of the variables in the dataset. However, SOM does create clusters

based on specific combinations of variables, which could be useful for improving development decisions. For instance, the Southern Africa cluster was created because the countries have a combination of economic well-being, good governance, and good education. SOM allows us to understand that the area most in need of assistance in this cluster is where the HIV pandemic hit hardest. If there is a need for developmental focus on civil conflict, access to sanitation, or access to technology then those efforts should be focused elsewhere in sub-Saharan Africa.

Much of sub-Saharan Africa is in desperate need of help. However, there are certain areas that need more immediate attention than others. The cluster of the most immediate concern is the cluster colored white in both Figure 3 and Figure 4. This cluster, which is composed of countries in southern Western Africa and western Central Africa, has many obstacles to overcome before any developmental milestones are achieved. This cluster has been mired in corruption and poor governance, life expectancies are extremely low, the under five mortality rates are the highest in sub-Saharan Africa, the populations are uneducated, and there is no access to technology. That list is far too long, and the hope is that this analysis highlights that one variable or need can be addressed to spark development. Other areas such as northern Western Africa and eastern Central Africa have issues of their own, but they have better building blocks. For example, northern Western Africa has average governance, high life expectancies, and an average combined gross enrollment ratio, which may allow them to overcome obstacles such as an illiterate adult population. The cluster in eastern Central Africa has well educated populations for sub-Saharan Africa, which could eventually help them combat the low life expectancies in the area. Thus SOM clustering shows how countries can compensate for the lack of one factor given the nature of relationships among other factors.

Governance appears to influence the middle rung of the development ladder in sub-Saharan Africa. In other words, governance affects all clusters other than the most and least developed. This is shown through the comparison of Figures 3 and 4. The Southern cluster and the western Central African cluster remain roughly the same regardless of whether governance is included or not in the SOM analysis. This is an indication of the impor-

tance of all other variables in these clusters. Whether it is GDP per capita, life expectancy at birth, or technological access, these clusters have drastic values in one direction or another that dominate the SOM clustering and diminish the importance of the governance indicators. However, many of the other clusters break up when governance is excluded from the SOM. This highlights the importance of governance in these countries.

This paper highlights the direct relationship between governance and economic well-being through the case study of South Africa, where improving governance scores through any combination of variables (democratization and opening trade barriers in the case of South Africa) could significantly increase the economic well-being of a country. Economic well-being has the obvious potential to improve the health of the population, access to sanitation and clean water, and access to technology. If sub-Saharan African countries can improve their governance, the chances of positive development will be much greater. A stable government in place leads to better implementation of economic and educational policies without disruption from civil war or corruption. Governance appears

> Economic well-being has the obvious potential to improve the health of the population, access to sanitation and clean water, and access to technology. If sub-Saharan African countries can improve their governance, the chances of positive development will be much greater.

to be the most important factor affecting the future of sub-Saharan Africa's development, and it would benefit all parties to take this into consideration while formulating economic, financial, trade, environmental, sustainability, or other developmental policies. Further analysis is needed to validate these findings.

BIBLIOGRAPHY

Agarwal, P. and Skupin, A. (2008). *Self-Organising Maps: Applications in Geographic Information Science*. England: John Wiley and Sons.

Artadi, E. and Sala-i-Martin, X. (2003). *The Economic Tragedy of the 20th Century: Growth in Africa.* NBER Working Paper W9865. Cambridge, MA: The National Bureau of Economic Research.

Baik, J., Ben Arous, G. and Peche, S. (2005). "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices." *Annals of Probability* 33: 1643–1697.

Basu, A., Calamitsis, E. A., and Ghura, D. (2000). "Promoting growth in sub-Saharan Africa: Learning what works." *Economic Issues* No. 23, International Monetary Fund.

Bekker, B., Eberhard, A., Gaunt, T., and Marquard, A. (2008). "South Africa's rapid electrification programme: Policy, institutional, planning, financing and technical innovations." *Energy Policy* 36: 3125–3137.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Boukerrou, L. (ed.) (2009). "Africa collection." *Encyclopedia of Earth* (online), ed. C.J. Cleveland. Washington, D.C.: Environmental Information Coalition, National Council for Science and the Environment. Available at http://www.eoearth.org/article/Africa_(collection)

Bratton, M. and Chang, E. (2006). "State building and democratization in sub-Saharan Africa." *Comparative Political Studies* 39(9): 1059–1083.

Buve, A., Bishikwabo-Nsarhaza, K., and Mutangadura, G. (2002). "The spread and effect of HIV-1 infection in sub-Saharan Africa." *The Lancet* 359: 2011–2017.

Buys, P., Dasgupta, S., Thomas, T., and Wheeler, D. (2008). *Determinants of a Digital Divide in Sub-Saharan Africa: A Spatial Econometric Analysis of Cell Phone Coverage.* Washington, D.C.: The World Bank Development Research Group, Sustainable Rural and Urban Development Team.

Cavusgil, S. T., Kiyak, T., and Yeniyurt, S. (2004). "Complementary approaches to preliminary foreign market opportunity assessment: country clustering and country ranking." *Industrial Marketing Management* 33 (7): 607–617.

Clague, C., Gleason, S., and Knack, S. (2001). "Determinants of lasting democracy in poor countries: culture, development, and institutions." *The ANNALS of the American Academy of Political Science* 573:16–41.

Davies, A. M. C., and Fearn, T. (2004). "Back to the basics: The principles of principal component analysis." *Spectroscopy Europe.* Available online at: www.spectroscopyeurope.com/columns/tony-davies-column/2006-back-to-basics-the-principles-of-principal-component-analysis

Fowlkes, E. B., Gnanadesikan, R., and Kettenring, J. R. (1988). "Variable selection in clustering." *Journal of Classification* 5: 205–288.

Food and Agriculture Organization of the United Nations. (2009a). *AQUASTAT.* Retrieved August 1, 2009 from http://www.fao.org/nr/water/aquastat/data/query/index.html

Food and Agriculture Organization of the United Nations. (2009b). *FAOSTAT.* Retrieved June 21, 2009 from http://faostat.fao.org/

Go, D., Nikitin, D., Wang, X., and Zou, H. (2007). "Poverty and inequality in sub-Saharan Africa: Literature survey and empirical assessment." *Annals of Economics and Finance* 8(2): 251–304.

Goodman, L. A. (1974). "Exploratory latent structure analysis using both identifiable and unidentifiable models." *Biometrika* 61: 215–231.

Habib, A., and Padayachee, V. (2000). "Economic policy and power relations in South Africa's transition to democracy." *World Development* 28(2): 245–263.

Hegyvary, S. T., Berry, D. M., and Murua, A. (2008). "Clustering countries to evaluate health outcomes globally." *Journal of Public Health Policy* 29(3): 319–339.

Hotelling, H. (1933), "Analysis of a complex of statistical variables into principal components." *Journal of Educational Psychology* 24: 417–441, 498–520.

Howard, B., Phillips, C., Matinhure, N., Goodman, K., McCurdy, S., and Johnson, C. (2006). "Barriers and incentives to orphan care in a time of AIDS and economic crisis: A cross-sectional survey of caregivers in rural Zimbabwe." *BMC Public Health* 6(27).

Jha, P. et. al. (2002). "Improving the health of the global poor." *Science* 295: 2036–2039.

Johnstone, I. M. (2001). "On the distribution of the largest eigenvalue in principal components analysis." *Annals of Statistics* 29: 295–327.

Jolliffe, I. T. (2002). *Principal Component Analysis: Second Edition.* New York: Springer Series in Statistics.

Kalipeni, E., and Oppong, J. (1998). "The refugee crisis in Africa and implications for health and disease: A political ecology approach." *Social Science and Medicine* 46(12): 1637–1653.

Kaski, S., and Kohonen, T. (1996). "Exploratory data analysis by the self-organizing map: Structures of welfare and poverty in the world." *Proceedings of the Third International Conference on Neural Networks in the Capital Markets,* London, England: 498–507.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2008). *Governance Matters VII: Aggregate and Individual Governance Indicators 1996–2007.* Washington D.C.: The World Bank.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2005). *Governance Matter IV: Governance Indicators for 1996–2004.* Washington D.C.: The World Bank.

Kohonen, T. (2001). *Self-Organizing Maps.* Springer Series in Information Sciences, Vol. 30.

Kohonen, T. (1982). "Self-organized formation of topologically correct feature maps." *Biological Cybernetics* 43:59–69.

Lem, S. (2005). *Reassessing the Impact of Colonialism on Democracy in Sub-Saharan Africa.* Paper presented at the annual meeting of the International Studies Association, Honolulu, Hawaii, March 5.

Lek, S. and Guegan, J.F., (2000). "Artificial neuronal networks: application to ecology and evolution. Lek, S., Guegan, J. F. (eds.), *Environmental Science.* Berlin: Springer.

Olsson, O. (2007). "Conflict diamonds." *Journal of Development Economics* 82: 267–286.

Park, Y. S., Céréghino, R., Compin, A., and Lek, S. (2003). "Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters." *Ecological Modeling* 160(3): 265–280.

Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space." *Philosophical Magazine* 2(6): 559–572.

Raftery, A. E. and Dean, N. (2006). "Variable selection for model-based clustering." *Journal of the American Statistical Association* 101(473):168–78.

Recknagel, F. (2003). *Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation.* New York: Springer.

Sachs, J., McArthur, J. W., Schmidt-Traub, G., Kruk, M., Bahadur, C., Faye, M., and G. McCord. (2004). *Ending Africa's Poverty Trap.* Available online at http://www.brookings.edu/es/commentary/journals/bpea_macro/papers/20040325_sachs.pdf.

Scott, A. J. and Garofoli, G. (eds) (2007). *Development on the Ground: Clusters, Networks and Regions in Emerging Economies.* London and New York: Routledge.

The World Bank Group. (2009a). *Aggregate Governance Indicators 1996–2008.* Retrieved June 5, 2009 from http://info.worldbank.org.ezproxy.bu.edu/governance/wgi/index.asp.

The World Bank Group. (2009b). *Data and Statistics: Information Technology.* Retrieved August 1, 2009 from http://go.worldbank.org.ezproxy.bu.edu/9Y5AECAM60.

The World Bank Group. (2009c). *Data and Statistics: Water and Sanitation.* Retrieved June 17, 2009 from http://go.worldbank.org.ezproxy.bu.edu/BQ5BEFBNX0.

United Nations Development Programme. (2008a). *Human Development Indices: A Statistical Update 2008.* New York: United Nations.

United Nations Development Programme. (2008b). *Human Development Report 2007/2008.* Retrieved June 10, 2009 from http://hdr.undp.org/en/statistics/data/.

Yin, H. (2007). "Learning nonlinear principal manifolds by self-organising maps." Gorban A. N. et al. (eds.), *Lecture Notes in Computational Science and Engineering* 58. New York: Springer-Verlag Inc.

The Frederick S. Pardee Center for the Study of the Longer-Range Future at Boston University convenes and conducts interdisciplinary, policy-relevant, and future-oriented research that can contribute to long-term improvements in the human condition. Through its programs of research, publications, and events, the Pardee Center seeks to identify, anticipate, and enhance the long-term potential for human progress, in all its various dimensions.

The Pardee Papers series features working papers by Pardee Center Fellows and other invited authors. Papers in this series explore current and future challenges by anticipating the pathways to human progress, human development, and human well-being. This series includes papers on a wide range of topics, with a special emphasis on interdisciplinary perspectives and a development orientation.

ABOUT THE AUTHORS

Zachary Tyler works for a consulting firm in Massachusetts that conducts program evaluations for energy efficiency programs, and he continues to work on statistical and geospatial analyses of human development issues. In 2010, he will receive a master's degree in energy and environmental analysis from Boston University.

Sucharita Gopal is Professor and Director of Graduate Studies in the Department of Geography & Environment and a member of the Cognitive & Neural Systems (CNS) Technology Lab at Boston University. She teaches and conducts research in geographical information systems (GIS), spatial analysis and modeling, and remote sensing for environmental and public health applications. Her recent research includes the development of a marine integrated decision analysis system (MIDAS) for Belize, Panama, and Massachusetts, and a post-disaster geospatial risk model for Haiti.

This paper is part of the **Africa 2060 Project,** a Pardee Center program of research, publications, and symposia exploring African futures in various aspects related to development on continental and regional scales. For more information, visit **www.bu.edu/pardee/research/.**

BU

Boston University The Frederick S. Pardee Center
for the Study of the Longer-Range Future