Building Statistics and Data Science Capacity for Development

Eric A. Vance[1] and Kim Love[2]
[1]Laboratory for Interdisciplinary Statistical Analysis, University of Colorado Boulder, USA, Eric.Vance@Colorado.EDU
[2]K. R. Love Quantitative Consulting and Collaboration, Athens, GA, USA

Abstract

Statistics, as a discipline and a practice, enables and accelerates all aspects of data-driven research, business, and policy and therefore development. Since local challenges require local solutions, those best positioned to analyze data are local statisticians and data scientists who understand the local context of how the data was produced. We propose a new model for building statistics and data science capacity, one that helps statisticians and data scientists in developing countries to build their own capacity to engage in data-driven development by working at the intersection of data producers, data analyzers, and data decision makers to transform evidence into action. Based on the collective experiences of more than 30 newly created stat labs in the LISA 2020 Network, we discuss nine lessons learned for building statistics and data science capacity for development.

Locally Powered, Data-Driven Development

We consider development, at a very broad level, to be sustainable actions that positively impact society. We believe that three components lead to such development: scientific and research innovations, creating jobs for sustainable activities, and implementing effective policies. Statistics, as a discipline and a practice, enables and accelerates all aspects of data-driven research, business, and policy. Data is key to knowledge, and statistics and data science are the bridge to understanding that data. Understanding data allows people to make scientifically sound decisions, thereby answering questions, solving problems, and generating development benefits, often at a deeply meaningful level. This is especially relevant for developing countries, which often have pressing development needs for which solutions can be guided through both the production and analysis of data.

We are mindful that local challenges require local solutions. Solutions provided by a consultant from a highly developed Western country may not apply in a developing country. The best people to collect or produce data are local experts who understand the context of the local data collection activity. The people best positioned to analyze that data are local statisticians and data scientists who understand the local context of how the data was produced. The best people to craft and implement policy innovations are locals with the lived experience of those to be affected by changes in policy. In other words, it is local researchers, businesses, and policy makers who actually solve local challenges.

In our experience, policy making and other types of decision making often ignores the need for rigorous statistical analyses. The simplistic perception is that there are only two actors necessary for data-based decision making: those who collect or produce the data and those

who review the data and make policy or other decisions. In many cases, the data producers and the data decision makers exist within the same organization or individual, such as a business that collects data on its customers and uses it to make a business decision or an academic researcher who conducts an experiment to answer a scientific question. As collaborative statisticians who have worked on hundreds of projects and supervised thousands more, we are keenly aware that there are, in fact, at least four important components required to make informed data-driven decisions: the domain expertise required to ask the right questions; high-quality, relevant data; appropriate, nuanced statistical analyses; and the power to make and implement a decision. Assuming that domain expertise exists within data producers and data decision makers, this requires the addition of a third actor--the statistician or data scientist--who possesses specialized skills in data analysis and interpretation. The intersections of these three actors are shown in Figure 1. Sustainable development requires all three actors operating at full capacity.

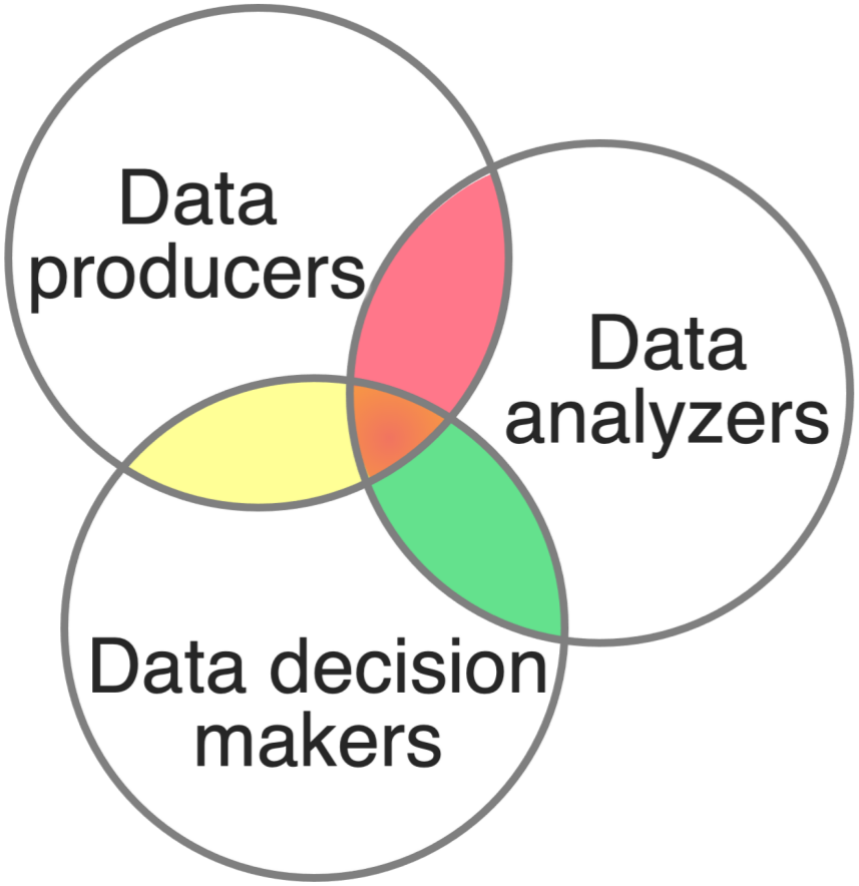

Figure 1. Venn Diagram of Data-Driven Development Actors

Consequently, our strategy for engaging in locally powered, data-driven development starts with acknowledging four potential gaps:

Gap 1: Local expertise to frame locally relevant development questions
Gap 2: Consistent production of high-quality data through carefully designed experiments, studies, or surveys at the international, national, and local levels

Gap 3: Technical ability in statistics and data science to design studies or experiments and appropriately model, analyze, and interpret data

Gap 4: Transforming evidence from data into action for development.

A New Model for Building Statistics and Data Science Capacity

One of the International Statistical Institute's four strategic priorities is building statistical capacity in developing countries. Historically, this has focused on building the capacity of developing countries to produce state-sponsored data, for example, to conduct accurate population censuses, consumer price index surveys, etc. We believe that while the state-sponsored production of data is foundationally important for development, it only addresses one of the potential gaps in data-driven development. Even though the data produced may be of high quality, the data might not provide the information or evidence needed to answer the questions raised by the data decision makers, the data may be interpreted incorrectly, or the data alone may be insufficient for taking action for development.

In this paper, we propose a new model for building statistics and data science capacity, one that helps statisticians and data scientists in developing countries to build their own capacity to engage in data-driven development.

It is important to keep in mind that working in academic isolation will not create positive development outcomes. To create real-world impact, statisticians must take steps to apply theories and methods to help transform academic evidence into action for the benefit of society. For example, at the University of Calcutta, India, statisticians and data scientists created a model to forecast the solar energy output of solar farms and then collaborated with decision makers to use the model to optimize the production of sustainable solar energy. Transforming technical statistical methods into positive action for society requires statisticians and data scientists to be skilled collaborators as well as skilled methodologists and analysts. In other words, these statisticians must have the skill to work in the intersections of Figure 1 in addition to the skills to analyze and interpret data. They must be able to understand the data and projects they are working with on both a deep and broad level and be able to communicate the results of statistical methods and analytical work in ways that provide actionable evidence to those who can use it to positively impact society.

Our model for building statistics and data science capacity is to create statistics and data science collaboration laboratories ("stat labs") that work in the intersections of data-driven development by collaborating with data producers and data decision makers to transform evidence into action. We believe that we have successfully built the capacity of our statistics and data science partners in developing countries by helping them engage in data-driven development. Increased collaborations between statisticians and development actors in developing countries can contribute to improved infrastructure, business development, education, agricultural growth, and human rights issues, among many other areas in urgent need of progress. Especially in developing countries, it is essential that *local* statisticians and data scientists interact with *local* researchers, businesses, and policy makers to develop sustainable solutions to local challenges.

In the remainder of this paper we describe the foundation of our capacity building model--the statistics and data science collaboration laboratory. Then we discuss the LISA 2020 Network

supporting these labs around the world, the results of our statistics and data science capacity building efforts, and lessons learned so far.

Stat Labs Are Engines for Development

Statistics and data science collaboration laboratories ("stat labs") provide a mechanism for increasing collaboration of statisticians with researchers, business professionals, and development policy actors. In our context, these labs are housed within research institutions in developing countries (generally universities), and these stat labs have three main objectives. Stat labs

- *train* statisticians to have a collaborative, evidence-to-action mindset;
- *teach* researchers, business professionals, and development policy actors to become more capable of using data and more aware of the power of statistical analysis to inform decisions; and
- *provide a collaborative space* for statisticians and data scientists to work with those individuals to create data-driven innovations and solutions leading to widespread development impacts and outcomes.

Stat labs are not rooms full of computers. Rather, a stat lab is a team of statisticians and data scientists empowered to collaborate with domain experts to ask relevant questions, produce high-quality data, analyze and interpret data to create knowledge and evidence, and transform that evidence into action for development.

We view these stat labs as "engines for development." Stat labs initially focus on training students, faculty, and staff to become effective interdisciplinary researchers, with the technical and collaborative skills necessary to accelerate research and transform it into solutions to development challenges. These stat labs will reach out to the local community of researchers, business leaders, government agencies, and non-governmental organizations, both to provide training and to provide tailored statistical support. As a result, the community becomes more aware of the ability (and indeed, the necessity) of the stat labs to provide this type of assistance.

As this awareness grows, the community provides more requests to the stat lab for this type of assistance, which in turn provides more opportunities for capacity development within the lab. Initial successes create a positive feedback loop in which more development actors want to collaborate with the stat labs and become "data-capable," and more statistics students, faculty, and staff desire work in the stat lab. Stat labs that particularly focus on using projects as opportunities to train students rise to the challenge of building capacity quickly to successfully complete a higher number of projects. When projects come with funding, senior students and faculty can be compensated to both work on projects and mentor junior students. Successful, high-profile projects also engender support for stat labs from administrators within their universities, potentially loosening restrictive rules and removing institutional barriers to success, and attract more students to study statistics and data science. This self-reinforcing cycle of using experience on projects to build capacity to work on more projects is key to the long-term success of stat labs. See Vance's 2015 paper on "Academic Statistical Consulting Centers" for more information on the history of stat labs.

Over time, these stat labs output a supply of experienced statisticians with a collaborative, evidence-to-action mindset and a cadre of more capable development actors who recognize the power of statistics and data science to help solve their development challenges. The stat labs

also provide a physical and intellectual space for statisticians/data scientists and development actors to collaborate on projects to transform evidence to action to produce data-driven innovations to solve development challenges. These data-driven innovations lead to widespread, substantial and sustainable development impacts because a well-trained collaborative statistician or data scientist can enable and accelerate 10 or more development projects per year, and those projects can positively impact thousands upon thousands of people or more. Ways for statisticians and data scientists to learn to become more collaborative and impactful are described in Vance and Smith's 2019 paper on "The ASCCR Frame for Learning Essential Collaboration Skills."

Through all of these mechanisms, stat labs not only fulfill the role of data analyzer as shown in Figure 1, but serve as a conduit that leads to development work in the intersections of the various actors. When collaborative statisticians are able to participate in and even initiate projects that involve policy and decision makers, they can translate the policy makers' questions and desired information into quantitative questions that can be answered through the production of data. When collaborative statisticians are trained directly to work with data producers, they can help to design experiments or studies that result in high-quality data, which they can then analyze in a way that can be relied on. When collaborative statisticians can do both of these (connect with both the data decision makers and the data producers), they can ensure that data production occurs in a way that is appropriate for answering the questions posed by the decision makers, and analyze and interpret the resulting high-quality data in a way that provides the information required by the decision makers to make recommendations and implement policies. This three-way intersection, for which stat labs can be the catalyst, provides the most desirable setting for data-driven development.

The LISA 2020 Network of Stat Labs

While Director of the Laboratory for Interdisciplinary Statistical Analysis (LISA) at Virginia Tech, USA, the first author recognized the power of the LISA stat lab to build the technical and collaborative capacity of his students while collaborating on projects to benefit researchers, businesses, and policy makers. The LISA students were in fact learning more statistics and learning how to translate their technical work into useful results by collaborating on these projects. The more projects they worked on, the more effective they became and the better they were at training and mentoring junior students in LISA. Based on his extensive travels through more than 70 countries, he recognized that the LISA model, when adapted to local conditions, could help statisticians in developing countries build their own capacity to do similar work to solve local challenges. And so, in 2012, he created the LISA 2020 program to build statistical analysis and data science capacity in developing countries, with a goal of creating a network of at least 20 stat labs by the year 2020.

In 2016, with five newly established stat labs as members of the LISA 2020 Network, the first author moved LISA from Virginia Tech to the University of Colorado Boulder. In collaboration with the International Statistical Institute, word of the benefits of the LISA 2020 Network spread widely, with particularly strong uptake and adoption in Africa, South Asia, and Brazil.

As of August 2021, the LISA 2020 Network consists of 34 stat labs located in 10 low- and middle-income countries. There are 14 additional stat labs that are in the process of becoming full members of the LISA 2020 Network. See Figure 2.

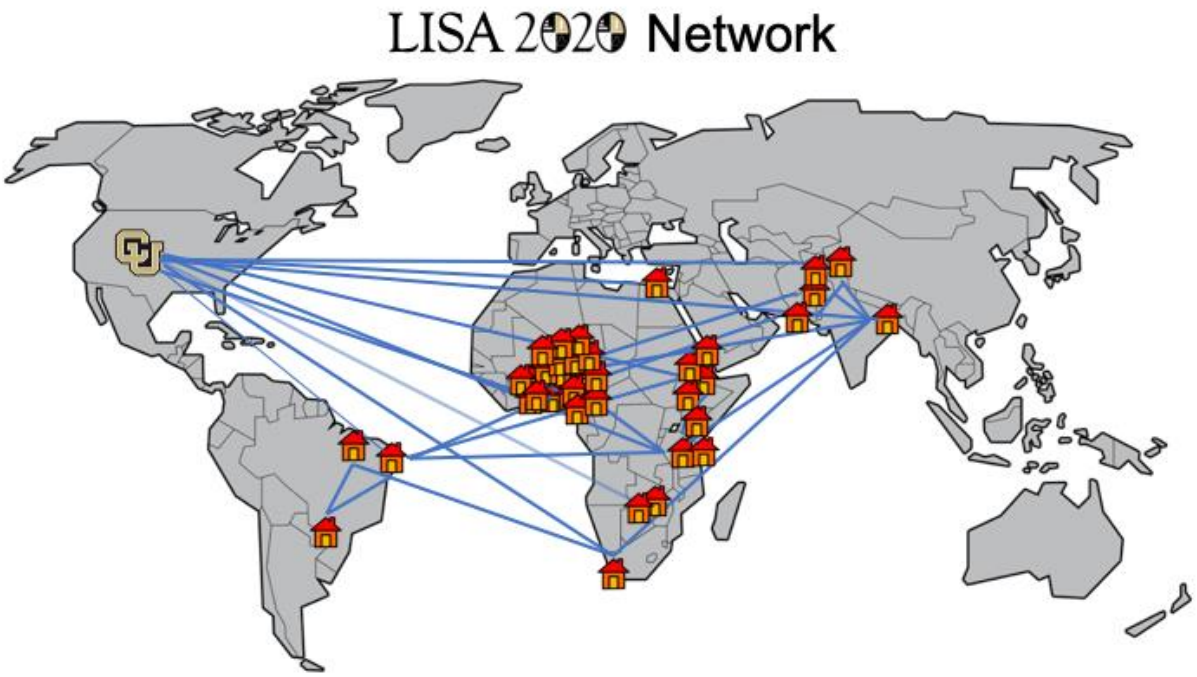

Figure 2. Map of current LISA 2020 Network stat labs. On World Statistics Day, October 20, 2020, there were 28 full member stat labs in the network.

The overall purpose of the network is to facilitate the process of individual stat labs becoming engines for development. This allows not only for faster capacity growth at each individual lab, but also increases the potential for a worldwide impact of the stat labs in the network through lab- and country-level collaborations. This begins with the creation of labs. Although two labs joined the network having been previously established, the vast majority of stat labs are created and grown through the LISA 2020 Network. Stat labs follow a seven-step process to become a member of the LISA 2020 Network:

1. Identify a potential Director or Coordinator of the stat lab and a mentor from within the LISA 2020 Network who will help the potential Director or Coordinator receive sufficient training and guidance in the non-technical skills needed to move between theory and practice to apply statistics and data science to solve real problems as well as to learn from other stat labs and share best practices.
2. Gather and document support from within the department and across the university for the stat lab in the form of letters from senior university officials supporting the official creation of such a stat lab.
3. Complete and submit via email the **Full Lab Plan/Proposal** (available at http://bit.ly/LISA2020FullProposal) to become a "Proposed Member" of the LISA 2020 Network. The proposal documents the purpose/mission of the stat lab, a name for it, and identifies a physical location with enough space to meet with domain experts. It also includes detailed statements about the lab's Context/Environment/Leadership; Mission, Goals, and Objectives; Activities; Personnel; Budget; Expected Outcomes (metrics); and Desired Impacts (metrics), as well as how the lab plans to stay connected with the LISA 2020 Network.

4. Respond to a review committee's feedback on the **Full Lab Plan/Proposal**. If the response to the feedback is satisfactory, the lab will be granted "Transitional Member" status in the LISA 2020 Network.
5. Open the stat lab: a) Train students and staff, b) Provide research infrastructure for local domain experts, c) Teach short courses/workshops to improve statistical skills and data literacy widely, d) Report on the stat lab's activities, outcomes, and impacts (metrics).
6. Stay connected with the network via semi-monthly Zoom meetings, annual symposia, quarterly reports of stat lab activities and numbers, and other channels.
7. Report a full quarter of metrics and present about the lab to the LISA 2020 Network.

Typically, labs will complete steps 1-4 before step 5, though some of the steps may be taken out of order. For example, a stat lab can open before they have completed step 3. We encourage labs to become connected with the LISA 2020 Network (step 6) at the early stages of the process. In general, labs that complete steps 1-3 are considered "Proposed Members". Labs successfully responding to the feedback on the Full Lab Plan/Proposal (step 4) will become "Transitional Members." The next step is to begin operation of the stat lab. The final steps are to report a full quarter of metrics and introduce the lab to the LISA 2020 Network at a Zoom meeting. Then, if accepted by the LISA 2020 Network by a two-thirds majority vote, they become "Full Members."

A further purpose of the LISA 2020 Network is to enable continued connections among the stat labs at various stages of development. This allows them to share progress and learn from one another and to collaborate on projects on an international scale. More established stat labs share their cultivated best practices with newer labs; the new labs then innovate to fit these practices to their particular circumstances, and in turn share new successes and challenges with the network. The LISA 2020 Network facilitates these exchanges of information through regular emails and newsletters, twice-monthly Zoom meetings (including more extended presentations by member labs), and annual symposia.

Finally, the network is a united organization that can act as a venue to provide funding and connect stat labs to decision and policy makers for data-driven development. As an example of the utility of the network, in 2018, the U.S. Agency for International Development signed a cooperative agreement with the University of Colorado Boulder for the LISA 2020 Network to provide funding for several stat labs to engage in pilot projects with data decision makers. This fund is called the Transforming Evidence to Action (TEA) fund and enables the selected stat labs to collaborate in the intersections with data producers and data decision makers. A selection of TEA fund projects currently underway include:

- The University of Ibadan in Nigeria partnered with the Independent National Electoral Commission (INEC) to assess the quality of the country's Continuous Voter Registration exercise, examine the effectiveness of the electoral process for voters, and make recommendations regarding the quality of the voter register and future election-related activities. See the article by Olubusoye et al. in this issue (2021, pp 13–23).
- Wolkite University in Ethiopia is partnering with the Gurage Zone Vital Events Registration Agency to improve the current vital events (e.g., births, deaths) registration system through design and analysis of resident surveys, database creation, and training agency workers in data management and analysis

- The Federal University of Rio Grande do Norte (UFRN) in Brazil is partnering with the Department of Public Policy at UFRN and União dos Dirigentes Municipais de Educação to address educational inequalities in the state of Rio Grande do Norte by analyzing data obtained from the Brazilian Ministry of Education and producing models to determine relationships across school infrastructure, students' social background, and students' performance in standardized tests
- The African Center for Education Development in Nigeria is partnering with the Nigerian National Bureau of Statistics to study the impact of COVID-19 on small scale business enterprises in northern Nigeria through the design and analysis of a survey of small business owners, and using the results to assist in the development of a road map for implementing economic intervention for small scale businesses
- The Kwame Nkrumah University of Science and Technology (KNUST) in Ghana conducted a two-part workshop for 26 female scientists in government positions on data analysis for decision making (part one) and methods for policy analysis, planning, evaluation, and leadership (part two).

Referring to Figure 1, these projects allow stat labs to work in the intersections between data analyzers, data producers, and policy makers (i.e., data decision makers). In several cases, the data producer and data decision maker is the same entity (e.g., INEC). The stat lab provides collaborative statistics and data science expertise to their partner to ask development-relevant questions, produce high quality data, analyze the data, and formulate policy recommendations, fully operating within the three-way intersection of Figure 1.

Statistics and Data Science Capacity Built

The LISA 2020 Network currently uses multiple indicators, collected quarterly from all labs in the network, to evaluate the accomplishments of the network as well as to promote learning and sharing of information among the network members. The indicators that are collected quarterly by the LISA 2020 Network are described in Table 1.

These metrics have been collected since January 2019, when there were 10 members of the LISA 2020 Network who were in a position to collect them; as of March 31, 2021 (the last day of the most recent quarter of data collection), with 34 labs reporting metrics, the labs have reported a total of 752 stat lab projects. They have also reported a total of 1,552 statisticians trained to be collaborative statisticians, 70 percent of whom are students (graduate and undergraduate) and 36 percent of whom are female. The stat labs have offered a total of 220 workshops, with 8,728 attendees, of whom 43 percent are female. Finally, the work of the stat labs has resulted in a total of 75 peer reviewed publications with a total of 226 authors including both stat lab collaborators and researchers (of whom 36 percent are female).

Indicator	Description	Disaggregation
Stat Lab Projects	Stat lab projects include any time a domain expert (researcher, business person, government worker, etc.) reaches out to the stat lab for statistical advice, analysis, or other statistical support.	None
Number of New Stat Lab Trainees	This includes all new stat lab trainees who were trained during the quarter. Individuals may be trained in several ways, including: (1) participating in stat lab projects, (2) enrolling in a consulting/collaboration course in the department, (3) attending a special training session, (4) participating in a video coaching/feedback session, (5) attending regular meetings where stat lab projects are discussed.	Faculty, staff, or student; male or female
Training Events	Training events are events planned for non-statisticians to learn more about using statistics and data science. This includes short courses (which are less than one day long), one-day workshops, and multi-day workshops that are taught, coordinated, or otherwise sponsored by the stat lab.	Short course, one-day workshop, or multi-day workshop
Training Event Attendees	Training event attendees are the non-statistician individuals who attend each training event. If the same individual attends multiple events during a quarter, they are counted once for each training event.	Internal faculty/staff, internal student, or external; male or female
Publications	This includes two general categories: peer-reviewed publications and non-peer-reviewed publications. A peer-reviewed publication generally refers to a publication in an academic journal. A non-peer-reviewed publication refers to a magazine-type article, publications produced by the stat lab, etc. These publications must be the result of lab projects to be counted in the LISA 2020 metrics.	Peer reviewed or non-peer reviewed
Publication Authors	This includes all individuals included as an author on a publication. If an individual contributed to more than one publication during a quarter, they are counted once for each publication.	Male or female

Table 1. Explanation of Indicators Collected by LISA 2020 Network

In addition to the indicators collected from all labs, there are indicators that we collect from labs that undertake funded development-oriented projects. These include the number of program and policy changes made by public sector, private sector, or other development actors that are influenced by lab-funded research results or related scientific activities; number of convenings held to disseminate research for use and/or develop policy recommendations; and publications specifically related to the results of these projects. The majority of these projects have only recently begun and we are still awaiting these results.

Stat labs are also encouraged to collect and report customized metrics unique to their lab, which may be determined by the lab's unique stakeholders. So far this has included metrics such as the number of events to promote the lab to potential collaborators, undergraduate lectures, and number of theses and dissertations assisted by the lab.

Nine Lessons Learned (So Far)

Through the process of creating and growing stat labs, measuring and evaluating the progress of those labs through metrics, and assisting with the implementation and administration of the TEA fund projects, the LISA 2020 Network has learned nine important lessons regarding the various ways in which stat labs can function productively in their role as data analyzers in Figure 1, while facilitating the intersections of all three actors in data-driven development.

Lesson 1. Stat labs should attain a stronger role in state-sponsored data production and analysis of that data. Stat labs are skilled at helping data producers produce high-quality data. For example, the LISA stat lab at the University of Ibadan was able to successfully assist the INEC in designing sampling plans and initiating sampling of the voting population and the voting register to answer questions regarding the quality of the voter register and the strength of Nigerian democracy. However, we have observed a divide between academic statisticians (including stat labs) and the national producers of data. Therefore, stat labs should work toward bridging this gap to attain a more prominent role in the data production process, as those who produce data at a national level are often not aware of the benefits of working with skilled statisticians and data scientists during the planning stage of data production. In addition, the production of state-sponsored data is often separate from the analysis or use of that data. Expert statistical analyses by stat labs of state-sponsored data could produce useful evidence leading to action for development. To facilitate future interactions and to strengthen the data production system, stat labs should focus on placing more of their graduating students in jobs in national statistics offices.

Lesson 2. Stat labs are skilled in training data decision makers to become statistically aware and data literate and should focus more effort on training policy makers. Training is a strong point of the LISA 2020 Network and opens the door for future collaborations. In just over two years, and with the COVID-19 pandemic making it more difficult, LISA 2020 stat labs have taught 220 workshops to 8,728 attendees, the vast majority of whom were research-focused data decision makers (i.e., university staff and students). An opportunity for stat labs to increase their impact is to focus training outside of their universities to business and policy decision makers as well as to state-sponsored data producers. An example of a stat lab expanding their scope for training was the TEA fund project at KNUST in Ghana to build data analysis and interpretation capacity for policy decision making and strategic planning, develop leadership capabilities, and provide a mentoring platform for mid-career female scientists in government positions. The impacts of these workshops have been nothing short of amazing and

include 11 job promotions and acceptances into funded Ph.D. programs, five research grants funded, five scientific publications relying on statistics learned in the workshop, and one participant (a virologist) has since become one of the leading voices in Ghana in the fight against COVID-19.

Lesson 3. Projects building capacity in the three-way intersection of data-driven development actors have the most potential for impact. When the data producer is also positioned to be a policy decision maker, a stat lab can provide statistical expertise in all phases of the project to frame development-relevant questions, produce high-quality data, analyze the data, and make policy recommendations. Combining state-sponsored data production and local-level data production from individual researchers, businesses, non-governmental organizations, or local policy makers with thorough analyses and interpretation of data will help development actors move beyond the common practice of thinking that data alone is sufficient for making decisions. Therefore, the LISA 2020 Network should focus its capacity building efforts on helping stat labs to work in this intersection, and by doing so, build their and their country's capacity for producing data at the international, national, and local levels; analyzing that data; and using that data for development. The LISA 2020 Network should redouble its efforts to provide training for statisticians to collaborate with the data producers and the data decision makers to transform evidence into action.

Lesson 4. Projects in the intersection of the research, business, and policy domains also have high potential for impact. Our labs, being primarily centered at universities, naturally focus on supporting researchers by helping them design experiments and studies to produce data and/or analyze the data to make decisions about scientific research questions. Disseminating the findings of a collaboration between a stat lab and an academic researcher through a journal, however, is often insufficient for influencing policy decisions. Therefore, if a project's goal is to influence development decisions, policy makers should become involved in the project and stat labs must deliberately reach out to them. Similarly, involving the local business community in academic research projects can increase the potential development impacts of those projects. By reaching out early as skilled collaborators, stat labs can ensure that they are helping to answer questions of interest to policy or business decision makers. The highest impact projects will involve all three of the research, business, and policy domains.

Lesson 5. Transforming evidence to action (TEA) requires a mindset shift uncommon in statisticians and data scientists. Even when statisticians collaborate with other development actors, the "traditional" end of the statistics or data science cycle (i.e., a timely, cogent, well-motivated, and contextualized analysis with easily digestible findings, conclusions, and recommendations) is still only part way toward development action. We have found that even when working directly with policy actors who can make data-based decisions, it is difficult to follow project outcomes until they result in verifiable policy changes. This is partly because policy change is often a long and complex process and our statisticians feel compelled to move on to the next project. Nevertheless, we believe that statisticians must adopt a TEA mindset to see a project all the way to its end if we are to transform evidence into action for the benefit of society.

Lesson 6. LISA 2020 stat labs are becoming engines for development. In a wide variety of contexts around the world in ten developing countries, our stat labs are successfully carrying out their missions. They are training their own staff and students and providing them with projects to further enable their growing capacity in collaborative statistics and data science. They are

conducting short courses and workshops to broad audiences, and they are engaging deeply in research projects, as evidenced by many co-authored publications. The stat labs are establishing themselves as local infrastructure to enable and accelerate data-driven development.

Lesson 7. LISA 2020 stat labs are adaptable in adverse circumstances. In 2020, navigating the COVID-19 pandemic required operational changes for individuals and organizations at a global scale. This unexpected crisis also forced the stat labs to adopt novel approaches to their training, teaching, and collaborative activities. As one example of this, the lab at UFRN in Brazil was able to transfer their student collaborator training activities to an online environment, providing lessons as well as group discussion; the initial semester provided the opportunity to record videos of the lessons, allowing for asynchronous lessons in future semesters where this approach is necessary. Another example is workshops offered by Afe Babalola University (ABUAD), which continued to be offered to internal participants at the university through online means throughout pandemic restrictions. The technology available at many of the stat labs is adequate to provide alternatives when in-person gatherings are not possible; the motivation of the lab personnel is sufficient to prioritize lab activities even when they cannot be done in the typical setting of the lab.

Lesson 8. LISA 2020 stat labs have many opportunities for improvement. We observe a lack of gender equality, with fewer women represented in every area in which gender is recorded--including collaborative statisticians trained, workshop attendees, and publication authors. Gender issues are not specific to stat labs, but rather stem from larger societal issues in the countries where the labs are located. Despite the current lack of gender equality, however, we are pleased to report that the stat lab directors are enthusiastic about achieving gender equality. Labs that are approved as full members of the network each provide a plan for including females in the administration of the lab and for working toward future gender equality in their activities.

Another area challenging stat labs is measuring longer-term impacts of their activities. For example, although we record the number of short courses, workshops, and attendees for all of the labs, only KNUST has implemented a longer-term follow up evaluation of their attendees to learn how they incorporated the training received into their work or research. KNUST allocated part of their TEA fund budget to document the impact of their workshop and the LISA 2020 Network as a whole can learn from this bright spot in documenting a stat lab’s impact.

Lesson 9. Sustainable funding for stat labs is a challenge. Some labs have obtained ongoing funding from their universities; most rely on volunteer efforts from their passionate statisticians and data scientists as well as support from individual researchers and workshop fees to continue their operations. Some labs have received initial funding to work with organizations and often continue working with them because the organizations recognize the value of skilled statistical collaborators and are willing to provide funding for statistical support on future projects. This helps to ensure the continued existence of these labs as they strengthen their statistics and data science capacity while enabling and accelerating data-driven development.

Conclusion

The LISA 2020 Network began almost a decade ago with the idea that collaborative statisticians in developing countries could create stat labs to build statistics and data science capacity. Based on the collective experiences of more than 30 newly created stat labs since then, our network is being transformed by the idea that we can build such statistics and data science capacity by focusing our research, education, and outreach efforts on the intersections of data-driven development. The current and near-future focus of our network is on improving the quality and sustainability of the individual stat labs and strengthening connections between them. We are excited to discover what the next decade holds for using data-driven development to build statistics and data science capacity.

References

Olubusoye, O.E., Akintande, O.J., and Vance, E.A. (2021) “Transforming Evidence to Action: The Case of Election Participation in Nigeria.” *CHANCE*, 34(3) 13–23.

Vance, E. (2012) “International Experiences in Statistics.” *Amstat News*, Issue 420, 17–19.

Vance, E.A. and Smith, H. (2019) “The ASCCR Frame for Learning Essential Collaboration Skills.” *Journal of Statistics Education*, 27:3, 265–274.

Vance, E.A. (2015) “Recent Developments and Their Implications for the Future of Academic Statistical Consulting Centers.” *The American Statistician*, 69:2, 127–137.