

NetApp Data Explorers

# Preparing youth for the data-filled future

Chad Dorsey, The Concord Consortium

Andee Rubin, TERC

Michelle Mann, NetApp



## We live in an era defined by data.

More than ever before, data plays a vital, expanding role across practically all aspects of society. Data has transformed industry, giving birth to entirely new sectors and becoming central to careers at all levels. Data has assumed new relevance in our personal lives as well, informing daily activities, major life decisions, and even democratic participation and engagement. As data's significance continues to grow in all quarters of life, it is more crucial than ever to empower youth to use data competently and productively in their future lives and careers. We must ensure that youth become fluent at working and communicating with data.

Fortunately, data provides many natural opportunities for developing new and engaging learning experiences. Using data to explore issues of societal importance can resonate particularly strongly with adolescents, who are exploring their place in the world and the impact they want to have. Such opportunities give youth the chance to experience for themselves the power data holds for illuminating and addressing real-world problems. Unfortunately, however, youth today have far too few opportunities to participate in powerful, data-centered learning experiences. Our kids—and their futures—deserve more.

NetApp's Data Explorers program represents a key example of such a future-focused learning opportunity. With a goal of empowering teens to discover and develop critical data science skills, Data Explorers makes use of flexible afterschool settings to involve youth with data starting at an early age. In the Data Explorers program, middle school aged youth examine data related to the United Nations Sustainable Development Goals (SDGs) to investigate pressing social and environmental challenges. As they do, they experience data use in the context of global issues, collaborate with adult volunteers to develop personal connections to the world of work, and experience firsthand the power data offers for identifying and solving problems they care about and want to be involved in.

Designed from the outset to embrace data-focused learning around issues of societal importance, Data Explorers provides a compelling illustration of the potential of out-of-school data education experiences. Engaging youth with social issues offers an accessible route to capturing their interests. Working with data-centered issues through personally relevant projects offers meaningful opportunities for youth learning and also creates chances for adult volunteers to share their perspectives and support youth in conducting investigations. And perhaps most importantly, learning opportunities focused on social topics provide youth an unparalleled window into issues of global concern, underscore the importance of global connectedness, and help them see how they can make a difference in a data-driven world.

# Learning about and with data

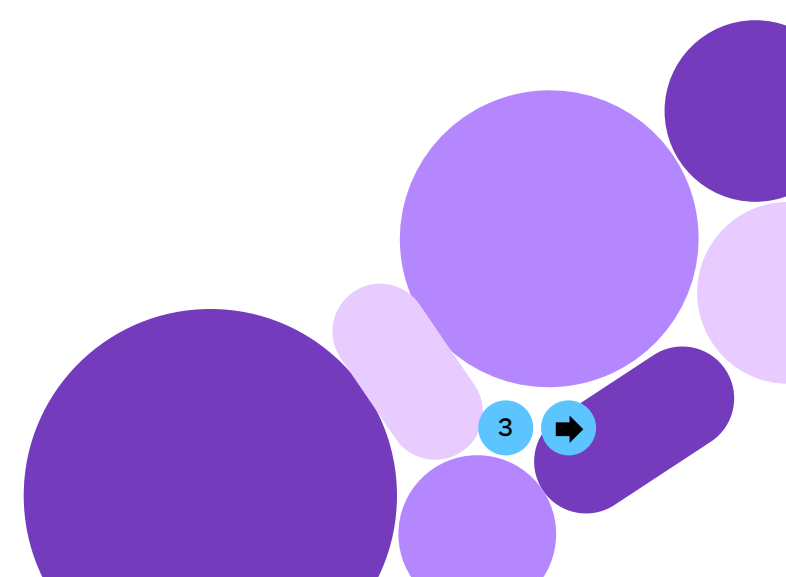
What are the keys to creating high-quality data-focused learning experiences? To identify the keys, we must first ask, What exactly is important for students to learn about data? The answer has expanded as data has grown ubiquitous and assumed greater consequence in our world.

To begin with, even the central issue of what we mean by “data” has changed in recent years. Certainly the term brings to mind large tables of numbers and associated charts. Yet data today encompasses a wider range of objects. In our data-focused world, practically anything—images, songs, words, websites, and much more—can, and has been, used as data. In the study of social issues, maps also play a key role as a source of data. As a critical part of their early encounters with data, today’s youth should recognize that “data is everywhere” and that anything they do can be made into data through a process of measurement. Understanding the process by which data is generated is key. Just like journalists encountering new information, young learners should learn to pose key questions about the source of any data they work with, asking, Who produced this data? For what purpose? When, where, and how? And throughout their data education, youth should become aware that even the best datasets still reflect only selected aspects of the situation they purport to reference.

It is important to note that learning about and with data differs in important ways from learning mathematics.<sup>1</sup> One difference is that data always refers to an idea within a greater setting—one famous description of data is “numbers with context.”<sup>2</sup> A second defining characteristic of data is variability. In simple terms, this means that data always carry some uncertainty. For example, even a measurement as simple as a person’s height will always be somewhat uncertain, whether because of the precision of the measuring instrument or even variation over the course of the day. Statements about “people’s heights” need to deal with a different kind of variability, because heights vary widely across a group of people, yet statistical reasoning provides the tools to make statements like “men are generally taller than women.” Data is best thought of as a distribution of values that contain both a “signal” and some “noise”; learning to distinguish the two is a long-term goal of data education.

Data comes in sets of numbers; rarely do we get one or two numbers and consider them data. Because of the large number of values in most datasets, visualization is a key strategy in finding meaningful patterns in data. When learning about data, youth need to be able to read complex data visualizations and to use appropriate technology to create their own. Once they find patterns, they need to know how to interpret them. Crucially, they need to understand to what degree it’s possible to generalize beyond the data they have to other related data or situations. A data-fluent learner must manage these aspects of working with data while pursuing the goal of data analysis: using data to generate knowledge about the world.

Working with data offers a rare opportunity for learning overall. Because understanding data is an inherently interdisciplinary endeavor, data-centered approaches to learning can find a place within an extremely wide range of topics and issues. When viewed through the right lens, there are few topics or issues in which data can’t find a suitable home. Further, both data and the topics it derives from work in tandem: analyzing data helps us learn about the world, and our questions about the world lead us to explore data more deeply.



# The importance of technology tools

Given the volume, ubiquity, and complexity of data in today's world, it's clear that anyone engaging with data—novice, expert, or in between—needs technological assistance in organizing, visualizing, and making meaning with it. Professionals have access to a multitude of powerful tools for working with data, including spreadsheets, databases, and programming languages especially adapted to data, such as R and Python. When helping youth learn, however, it's important to provide tools that are designed specifically to aid that learning.

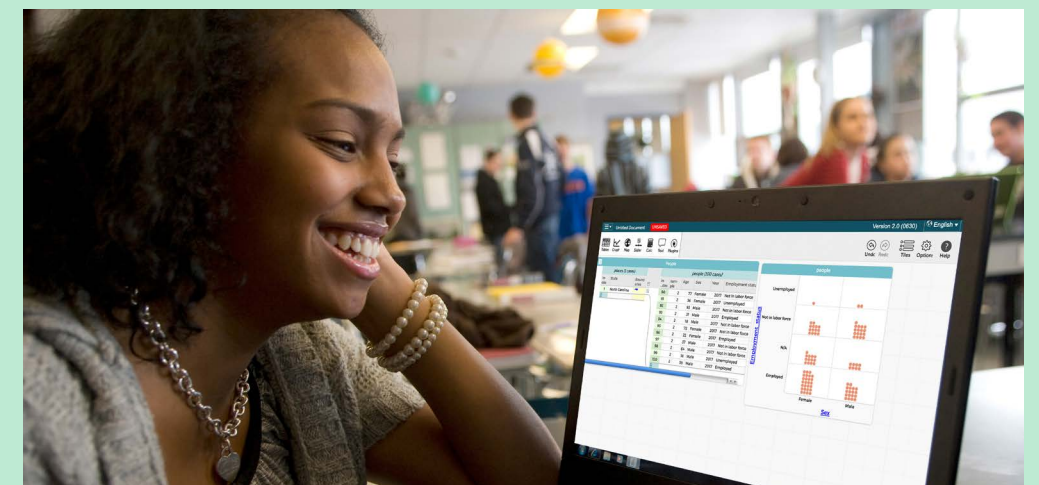
Creating effective learning tools requires applying research into the ways novices think about a topic, folding in understanding of learners' potential conceptual difficulties and bringing experience with flexible interfaces to bear. Fortunately, several decades of work have informed the design of tools to support youth in learning about data.<sup>3</sup> One of the most popular of these is the Common Online Data Analysis Platform (CODAP), a free, open-source tool designed to engage youth from grades 5 through 14 in learning with and about data. CODAP's intuitive interface allows youth to ask questions of a complex dataset quickly and easily, and to understand the answers, even if they have had no prior experience with data analysis tools.

By enabling learners to readily create multiple maps and graphs, filter viewed data, and reconfigure data tables, CODAP allows them to ask their own questions of a dataset and easily investigate a series of hunches and hypotheses. Freeing youth to approach data in this way transforms their approach to learning with data. Rather than searching for a single correct answer within a dataset, they instead begin to view datasets as objects they can explore broadly. Additionally, because CODAP's maps, graphs, and tables are all linked together, youth working with data such as those from the UN SDGs see connections between data points highlighted across multiple graphs and maps. These important features enable them to form a complex view of the interplay of variables and to build an understanding that different visualizations simply represent multiple ways of viewing a common dataset.<sup>4</sup>

NetApp's Data Explorers has incorporated CODAP as its primary data exploration tool. Its ability to help youth gain new insights into complex datasets and feel empowered to answer original questions has been critical to the program's success.



NetApp Data Explorers in Bangalore, India



Student using CODAP, source: Concord Consortium

# Opening the door to social change

Social topics are inherently compelling to young people. This is particularly true for members of the current generation, who are uniquely drawn to topics of equity, social concern, and global issues.<sup>5</sup> Such topics are often ideally suited to data-centered learning approaches.

When considering topics of social importance, bringing data to the table has multiple advantages. Working with social issue data that reflects a broad range of experiences can help youth see situations from points of view other than their own.<sup>6</sup> Although social issues are often presented in terms of anecdotes or individual cases, working with data helps youth recognize trends that go beyond individual cases. Exploring relationships among social variables can help them see broader patterns that underlie complex social realities. When working with social

issues data, even a question as simple as “How were these numbers determined?” can open the door to a nuanced discussion about social values and inequities or about the human decisions and ethical questions involved in producing the data itself.<sup>7</sup>

Examining data also invites thinking that crosses temporal and spatial scales. When looking at data, young people naturally work to “find themselves” in what they’re exploring. In the Data Explorers program, students work with data on multiple scales. They begin by generating data about themselves, but quickly move to look at data about different countries based on data from the United Nations. Soon they focus on a relatively local dataset from their city or state and dig deeper into the circumstances reflected in their immediate community. In this way, youth can see the spectrum of scales on which data can appear—personal, local, national, and international. The Data Explorers curriculum deliberately uses datasets at these different levels to help youth see the connections among their personal worlds and the wider world.

In working with youth around the world, we have seen many examples in which data education has provided relevant connections to topics of social concern based on the UN SDGs. One boy folded his interest in becoming a doctor into a detailed project that used data on the availability of doctors in different countries to imagine where a doctor could have the most impact. Another youth’s passion for women’s health drew her into an examination of the relationship between teen pregnancy rate and female educational status in different countries around the world. Over and over again, we have seen how working with data can leverage personalization and relevance in ways that deepen youth interests and fuel potentially lifelong passions for applying data to solve problems.



# Connecting personal, local, and global: How Data Explorers opens up the world

The NetApp Data Explorers program is designed to be delivered as an “afterschool club,” taking place during out-of-school time. The 12 weeks map to four stages: data awareness and inspiration, exploration and investigation, visualization and interpretation, and problem solving and taking action. This arc is part of an intentional structure designed to first ground youth in data around social issues and then provide opportunities for them to dig deeper into projects of personal interest.

The first six lessons introduce youth to “data moves,”<sup>4</sup> which they use to explore both an international and a local dataset of health and education indicators based on the United Nations Sustainable Development Goals. Students look at examples of how data is used in the real world as they prepare to do their own data investigations.

The next five lessons focus on the Data Explorers Project, in which students work in teams to investigate their chosen societal issue and question within a local dataset. Volunteers from NetApp offer a view into how data is used in various careers and give guidance throughout the project. In the final lesson teams give presentations, sharing their findings and recommendations via a Showcase event to a panel of volunteers and a broader audience, including other youth, parents, and NetApp employees.



# Complexities and caveats

As with any promising approach, data education is not without its complications. Working with data involves many different skills, most of which current educational opportunities leave unaddressed. As a result, when tackling data-centered learning, educators and youth must bridge many gaps.

Youth must know how to interpret data of different types, navigate uncertainty, and incorporate data visualizations effectively into their communication about complex ideas. They must be able to place technical data in context, folding data into the argumentation skills they are developing in middle and high school and beyond. Although youth of any age can acquire these skills to appropriate levels, doing so is far from straightforward or immediate. Learning about data takes time. Time in school for data-intensive learning may be limited, so out-of-school experiences serve as important additional opportunities for youth to encounter data.

Designing effective learning experiences involving data also requires starting with the right ingredients. Prime among these is the data itself. Effective data-centered learning begins with carefully selected datasets, chosen and organized in ways that make meaningful learning possible and accessible. Data should be curated in such a way that it invites explorations that bring youth face to face with both fundamental data concepts and interesting insights about the world. For example, Data Explorers datasets include many measures that are proportions or rates—for example, the number of doctors per 100,000 people in a country. Youth need to understand that, while there are many more doctors in China than in Costa Rica, for example, there are fewer doctors per person in China. This kind of reasoning is central both to understanding data and to understanding how resources are distributed nationally and internationally.

The kind of data that youth explore in Data Explorers represents a single point in time, and by the time it is made publicly available, it may already be somewhat out of date. In 2022, the international effects of the COVID pandemic magnify this issue. One of the indicators that youth explore is life expectancy. They will often ask when the data were collected, because they know that life expectancy has decreased around the world over the past two years. This experience raises their awareness of the importance of knowing when data was gathered and can help them temper conclusions they draw from the data until they can get more current information.

Rich data-centered learning experiences also require properly multivariable data. Tables with two columns may be sufficient for making a single graph, but that's not enough to engage youth in an extended inquiry. Each of the datasets used in Data Explorers has 10 to 15 variables, mostly health and education indicators associated with a particular country, county, or other local area. This kind of structure allows youth to see relationships among these indicators; it also provides enough options to allow them to choose an indicator they find particularly interesting to investigate in depth.

To spark rich data-centered learning, datasets must be big and broad enough that youth feel empowered to immerse themselves in explorations of their own making.

While it is important for datasets to be complex enough to offer multiple paths for exploration, it is important that youth don't become lost to the point of unproductive confusion. This fine balance can be difficult to achieve, but getting it right is critical. Part of engaging in data science, whether as a professional or as a young learner, involves becoming accustomed to occasionally feeling "awash in data." We want to provide experiences that allow youth to feel sufficiently "dunked in the pool" of a dataset's many possibilities, yet constrained enough to ensure that the problems they explore are sensical and tractable.

Experiences that prepare youth for a world filled with data also demand "properly messy" data. While data education experiences generally involve datasets that have been extracted and prepared explicitly for the learning experiences, overcurating datasets robs young learners of important opportunities for encounters with authentic qualities of data. For experiences focused on data exploration, datasets should be sufficiently cleaned and organized that youth can step in and explore them straight away. However, encountering and working with missing values, working through issues of mislabeled data, and tussling with complicated questions of how to reconcile data of differing types are all part of the toolkit that youth must develop if they are to become fluent with data. For example, in the United Nations dataset used in Data Explorers, several small countries are missing some data, possibly because they don't have the infrastructure needed to collect the necessary information, or perhaps because they missed the deadline to submit their results. To create a perfectly "clean" dataset, those countries could have been omitted, but that would have misrepresented the reality and complexity of such data.



NetApp Data Explorers Students in North Carolina

# Social issues topics: Pitfalls to recognize, but great rewards to reap

Engaging with socially important topics carries its own complexities as well. Although data that relates to these topics offers strong potential for engagement, settling on a properly suited topic is by no means straightforward. In fact, preparing data-centered learning experiences around such topics can be strikingly difficult. Although it's true that data is everywhere today, the data that you need can often be surprisingly elusive. In fact, the specific data needed to explore a chosen topic of interest may simply not exist. Alternatively, data that is available for a given topic may initially appear compelling but may ultimately not be usable. This can happen either because the data does not relate deeply enough to a selected social question, or simply because the data does not prove interesting enough to youth. When developing data-focused activities, being prepared to test multiple examples and to iterate frequently can guard against larger issues in the future. We have experienced this problem firsthand. In an early implementation of Data Explorers, we asked students to share where they saw a lack of fairness in their lives. We assumed that we would be able to connect issues they brought up with the data we had prepared. However, their responses covered such a wide swath of social topics that the overlap with the dataset we had prepared was small.

The nature of social topics themselves can raise additional issues. A main reason that topics of social interest are enduring is that they are inherently complex and difficult to grapple with. Such complexity often demands an equally complex base of supporting knowledge. As a result, youth may lack the necessary context or background to interpret social issue datasets effectively. Interpreting a dataset may require mathematical, scientific, or other specific understanding that transcends the existing skills and knowledge that youth bring to the table. This is the flip side of the interdisciplinary nature of data—engaging properly with practically any data-centered problem demands a certain level of expertise with the core subject area itself.

Exploration of data around social issues raises a final potential barrier, this one related to problem solving. Although the allure of identifying possible solutions to relevant social problems is an enticing aspect of data-centered learning, it is important to keep in mind that the scope of most social issues makes it impractical for youth to generate solutions they can truly implement. Designers of social issue data experiences must take this limited sphere of influence into account.

One way to do this is to consciously design opportunities that allow youth to connect global issues of concern to localized problem-solving opportunities. The Data Explorers curriculum addresses this by introducing youth to datasets that relate directly to the global issues they have been exploring. In Data Explorers, youth engage in a “Digging Deeper” phase of their projects in which they can delve into an issue by doing web research, finding additional data, doing further statistical analyses, interviewing family members or friends, taking pictures of their community, or exploring local maps.

**Data Explorers projects help youth see their place in using data as a tool for addressing social issues in meaningful ways.**





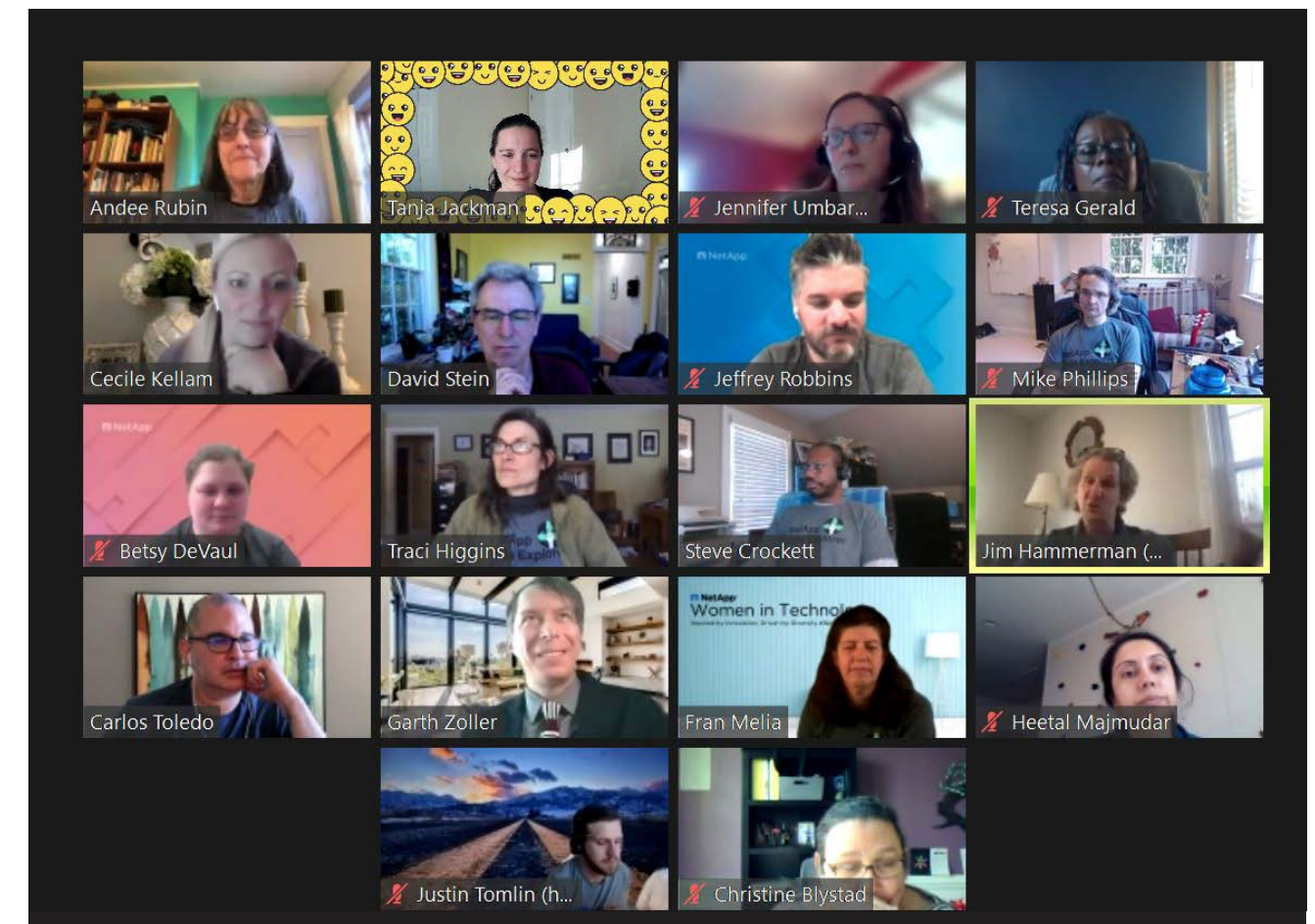
# Opportunities for adults

Because data is central to so many industries and workplaces today, many opportunities exist for introducing adult perspectives into data-centered learning experiences. In particular, involving adults from data-focused industries can help youth more concretely understand data's key role in real-world applications and see its relevance to the problems of daily life.

Additionally, data-centered learning opportunities can support learning at multiple levels. Many adults share an interest in gaining proficiency at working with data. Well-chosen datasets involving social issues can engage both adults and youth alike in data investigation and inquiry. Shared interest in a social topic investigation can stimulate dialogue and connection between adults and youth. In such cases, the different perspectives that each brings to the discussion enrich the overall conversation. The experience of investigating datasets together with adults can also empower youth, especially when an investigation provides opportunities for adults and youth to make discoveries together or allows youth to teach newly learned skills to their adult collaborators. Such cases also offer youth the powerful opportunity to see adults modeling lifelong learning.

Adults can assume a variety of roles when working with youth on data-centered learning experiences. They may provide context or perspective on the value of data within industry settings. They may help coach youth through a project or learning opportunity. They may draw upon a particular competency to support youth, providing expertise in a subject area such as social impact or data analytics. Or they may act as a seasoned, authentic audience for youths' presentation of artifacts or projects. Data Explorers involves adult volunteers from industry in all of these roles, making use of adults' unique perspectives as employees, mentors, and learners to help inspire and support young learners.

Of course, involving adults also introduces complexities into a learning experience. On the one hand, interested adults may not have firsthand experience working with data and data tools. On the other, professionals who are highly experienced in working with data may not necessarily know how to introduce youth to basic data concepts. Designers must take the potential diversity of adults' backgrounds, assumptions, and talents into careful consideration when designing experiences for a broad variety of volunteers. Programs should also make sure to allot sufficient time and resources to prepare and support adult volunteers.



# Important preparation for a rapidly arriving future

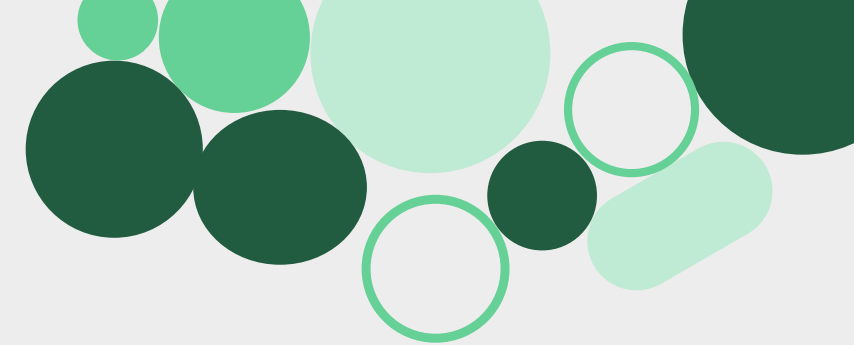
Work with data can pave meaningful pathways to engaged learning around social issues. Beginning these experiences for youth before or during adolescence helps ensure that they learn about the importance of data in life and work at an age when they are considering career options widely. Providing support for such experiences in out-of-school learning settings ensures useful leeway for addressing a wide variety of topics, including social issues. Out-of-school experiences also provide added freedom for programs to “go deep” in ways that allow youth to engage in broad exploration and experimentation in directions that are personally relevant to them.

NetApp Data Explorers stands as a useful model and an important source of new research findings to inform other youth development programs interested in engaging young learners with data around social issues. By developing a growing set of opportunities for youth to develop deep understandings of social issue data, NetApp is engaged in a dual mission. As a model for how to involve youth in learning with data, Data Explorers is planting the seeds for a global movement in data education focused on social issues. As a program, Data Explorers is focused on the most important mission of all: providing an expanding source of inspiration for youth around the world.

By preparing them to use data to understand and tackle the world’s most pressing issues, Data Explorers is helping to ensure that today’s youth gain the skills, knowledge, and experience they need to thrive as citizens and to enter relevant careers. We invite others to join this important cause, either by collaborating to help expand the Data Explorers model to new areas of the globe or by contributing new programs and approaches that prepare today’s youth for the future.

[> Learn more about NetApp’s Data Explorers program](#)

[> Learn more about the Concord Consortium’s free data exploration software CODAP](#)



## References

1. Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences*, 29, 154-164.
2. Moore, D., & Notz, W.I. (2020) *Statistics: Concepts and Controversies*. Macmillan.
3. Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In *International handbook of research in statistics education* (pp. 473-502). Springer, Cham.
4. Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*, 12(1).
5. <https://www.pewresearch.org/social-trends/2020/05/14on-the-cusp-of-adulthood-and-facing-an-uncertain-future-what-we-know-about-gen-z-so-far-2/>
6. Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, 29(1), 1-10.
7. Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A Call for a Humanistic Stance Toward K– 12 Data Science Education. *Educational Researcher*, 50(9), 664-672.