NATURAL HISTORY

9/20

IN DEFENSE OF BATS

NATURAL HISTORY

www.naturalhistorymag.com

SEPTEMBER 2020 VOLUME 128 NUMBER 8

FEATURES

16 PROFILING A PANDEMIC
Who were the victims of the Spanish flu?
BY SVENN-ERIK MAMELUND

20 FORBIDDEN FRUIT?
A plant chemist has second thoughts.
BY JORDAN L. HARTMAN

COVER STORY
22 GOING TO BAT FOR BATS
We lose something when we talk about animals as general groups.
BY ALYSON F. BROKAW

26 CATTAIL TALE
Why does a hybrid overrun some North American wetlands but not others?
BY VERENA SESIN

30 A VITAL BALANCE
Vitamin D helps combat immunological diseases, but it gets an "F" when faced with cancer.
BY ANYA M. NIKOLAI-YOGERST

33 VISUALIZING UNCERTAINTY
What can graphics convey about the character of data?
BY DERYA D. AKBABA

37 WHEN OUTREACH BECOMES A PRIORITY
Tracing the history of JunoCam
BY STEPHANIE HAMILTON

DEPARTMENTS

2 THE NATURAL MOMENT
Big Gulp
Photograph by Rathika Ramasamy

5 UPFRONT
Budding STEMS

6 SAMPLINGS

9 NATURALIST AT LARGE
A Fly's Wedding Party
Bernd Heinrich

12 MEDICAL EXAMINER
Our Year of Living Dangerously
Druin Burch

40 EXPLORING SCIENCE AND NATURE

42 THIS LAND
Iowa's Floating Mat
Robert H. Mohlenbrock

45 SKYLOG
A Planetary Grand Prix
Joe Rao

46 BOOKSHELF
Laurence A. Marschall

48 ENDPAPER
Impossibly Living in Possibility
Christa R. Avampato

ON THE COVER

*Egyptian fruit bat (*Rousettus aegyptiacus*), also known as the Egyptian rousette, ranges from Africa to India. Although farmers who grow fruits may regard it as a pest, the species serves various plants by pollinating them and dispersing seeds.*

Photograph by Marina Jay/Shutterstock

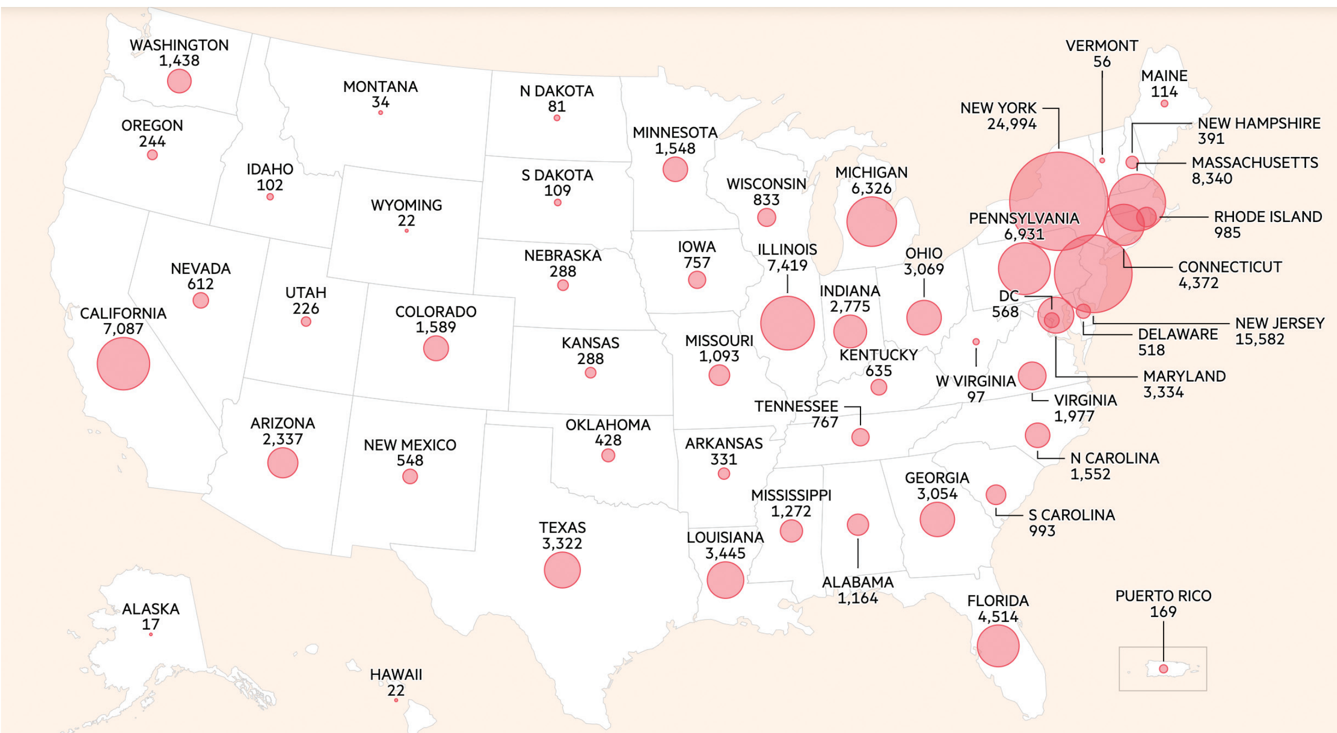
Visualizing Uncertainty

What can graphics convey about the character of data?

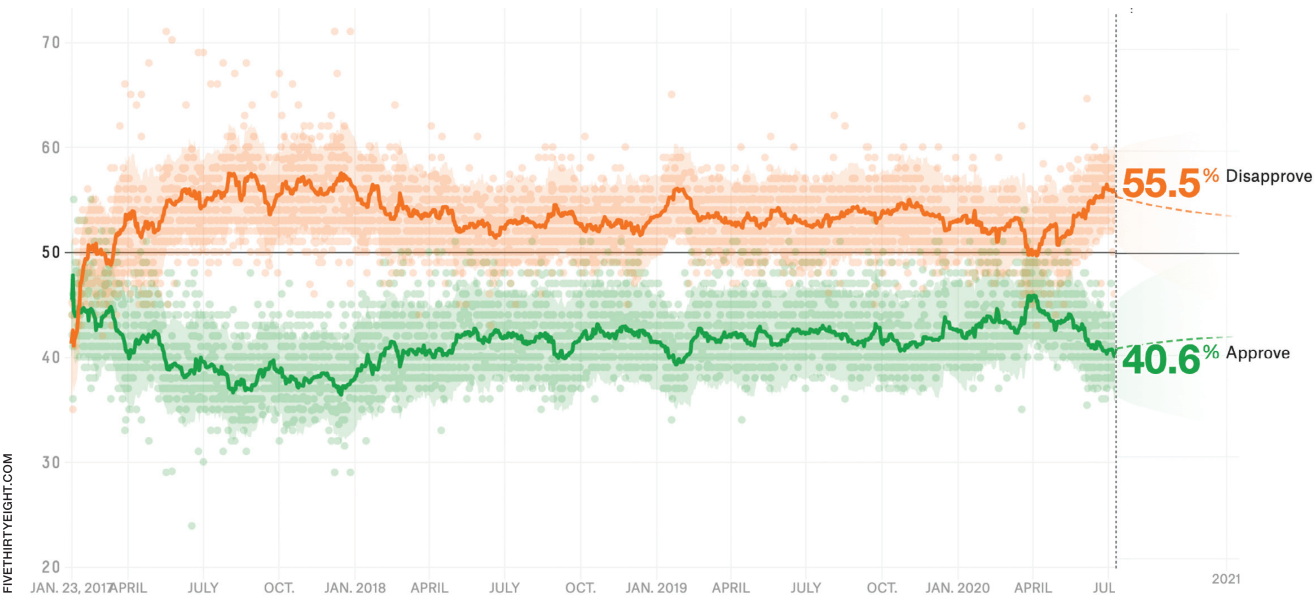
By Derya D. Akbaba

Every morning I make my coffee and check my Twitter feed for the latest COVID-19 maps and graphs. I examine the latest charts to understand where we are headed. And as a PhD candidate studying the ways in which numerical data can be represented visually, I also peruse the diagrams and the accompanying academic chatter and analyses to see how designers are handling the discrepancies and shortcomings in their data. When well designed, data visualization can often highlight important trends not easily conveyed by words and numbers alone. When dealing with a global pandemic, however, it is becoming more critical for designers to communicate the complexity and uncertainty inherent in the ever-changing data. And consumers of information need to understand what designers—possibly teams of artists, statisticians, journalists, and others—know about the data.

Take the United States map that was posted July 15th on the *Financial Times* website [*see illustration below*]. It visualizes the number of deaths in each state as different-size circles. Given that numbers accompanying the circles are not rounded (such as 767 for Tennessee and 22 for Wyoming), I would assume the authors feel confident in the precision of their data. But given the nuances of testing and verifying the infection in individuals, I become skeptical when I see exact numbers such



Circles of different size provide a visualization of the state-by-state COVID-19 death count in the U.S. as of July 15, 2020. The sizes do not reveal the proportion of deaths to the populations of the different states.



A graphic from a website dedicated to statistical prediction conveys President Donald J. Trump's changing approval ratings. Dots represent data from various polls taken from January 2017 to July 2020, while orange and green lines represent weighted averages of those polls—with greater weight given to polls that, based on experience and other factors, are considered more dependable guides. The bands in lighter colors provide 90 percent confidence intervals—meaning that nine times out of ten, any new poll taken at a corresponding time will fall within the ranges of the bands. The ratings end in July with 55.5 percent disapproval and 40.6 percent approval. Results for later in the year are projected at right, with increasing uncertainty.

as these. My initial reaction is to check the data source.

In this case, it turns out to be the COVID Tracking Project, a partly volunteer organization that grew out of efforts by journalists at *The Atlantic* and the founder of a medical venture firm, Related Sciences. After spending two minutes on their website (www.covidtracking.com), I discovered the following caveat:

> Remember that even death counts are uncertain. . . . There are some concerns that official death statistics may overcount COVID-19 fatalities by assuming any patient who tests positive for the virus [and dies] was killed by it. But Marc Lipsitch, a Harvard University epidemiologist, told FactCheck.org that "the number of such cases will be small." Undercounting is a bigger problem, he says. "A greater issue is errors in the other direction." If people die from COVID-19 before they are tested, their death might not be included in the official tally.

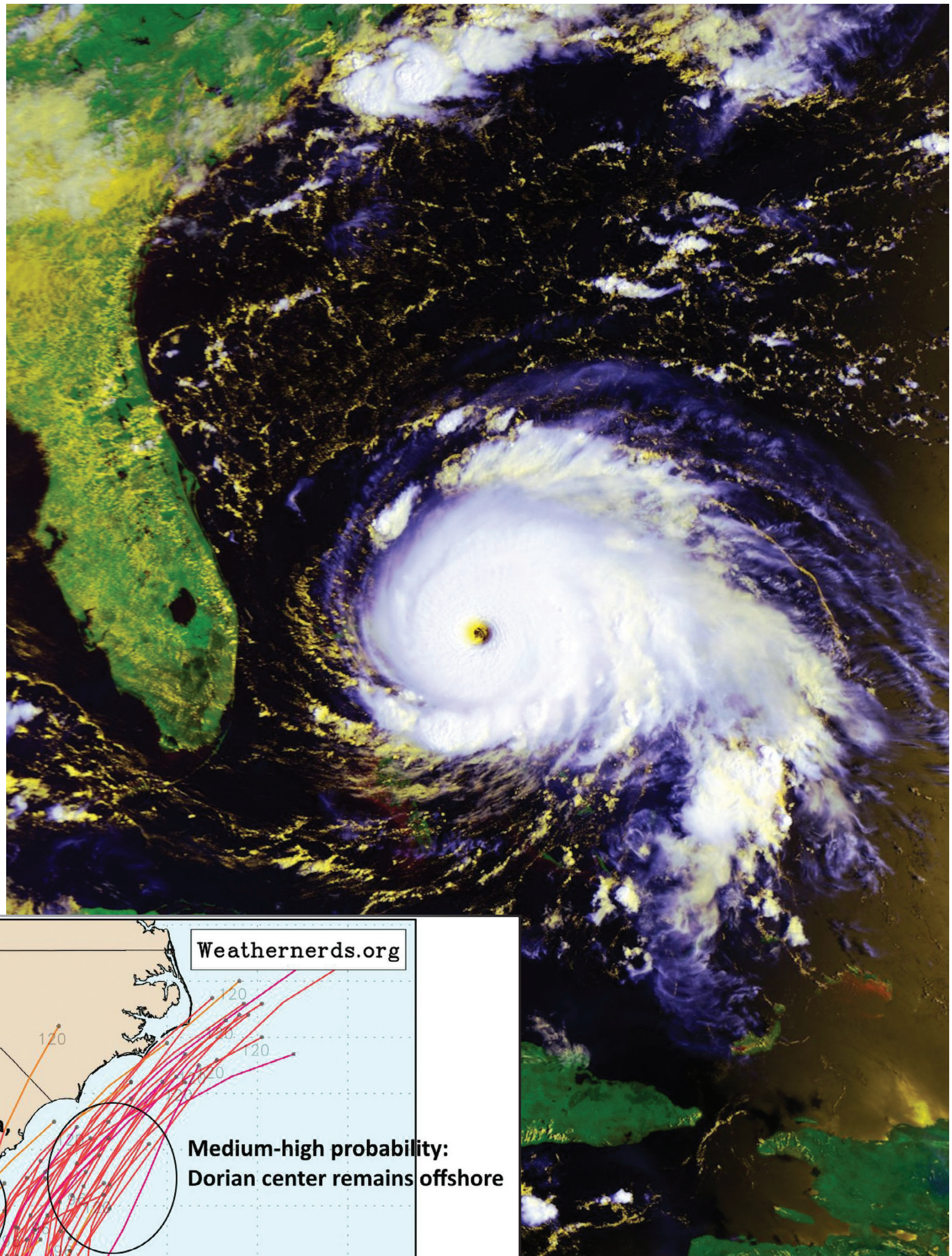
So, while it is not inaccurate to show precise numbers, which are a reflection of confirmed deaths, the questions remain: What are considered deaths by COVID-19? How thoroughly are they being tabulated?

As difficult as it is to collect data on deaths from COVID-19, data on the rate of infection is even more elusive. It is logistically impossible and prohibitively expensive to test everyone in the United States on a daily basis for COVID-19. We might get a more accurate idea of how many people are infected if we applied statistical analysis to a sample of the population. But sampling introduces a different kind of uncertainty.

For a survey to accurately reflect the United States population, it would have to meet certain standards, including (but not limited to) overall sample size and and proportionate representation of its constituent groups. Statisticians also apply mathematical principles of probability to assess the reliability of their conclusions. In a survey of voter preferences in a close political contest, for example, how confident can we be that the data match what would be obtained if the entire population could be surveyed? Numerically, the degree of discrepancy might be reported as a "margin of error." One way to visualize this is to show an interval of confidence [*see illustration above*].

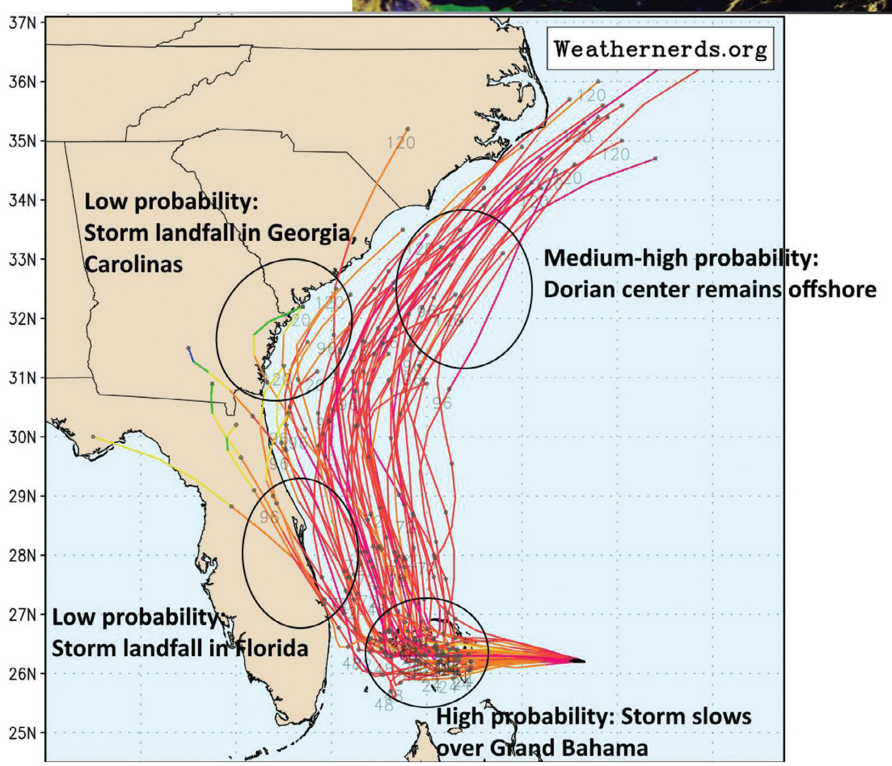
Not only do we use statistics to describe what we know about the data, but we also use it for modeling and making predictions for, among other things, weather forecasting, election outcomes, and sports events [*see illustration on opposite page*]. In any of these instances, as consumers we should be looking to the fine print about how these models are generated and visually represented.

Hurricane Dorian, at peak intensity, reaches Grand Bahama Island on September 1, 2019. The RGB Color image was taken by the Advanced Very High Resolution Radiometer on board the MetOp-A satellite.



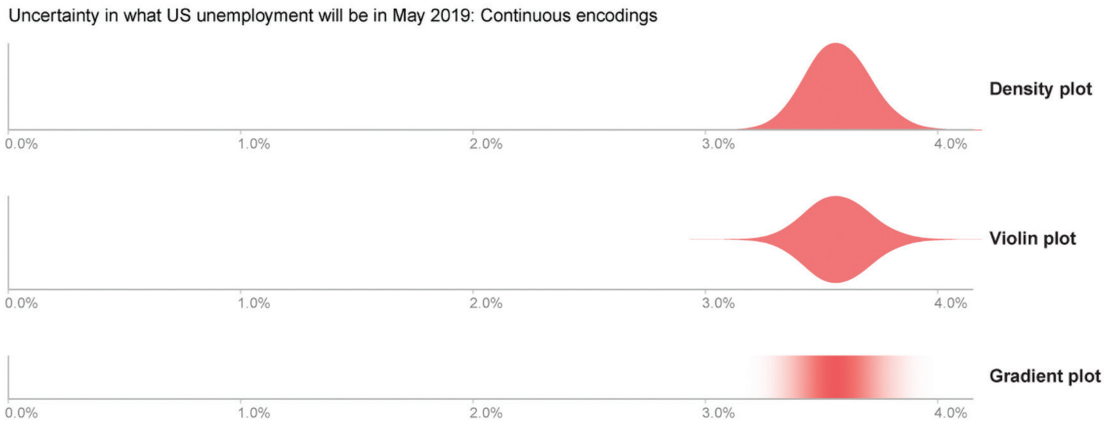
NOAA

Jessica Hullman, an assistant professor of Computer Science and Engineering at Northwestern University, and Matthew Kay, an assistant professor of Information at the University of Michigan, are among the researchers who specifically study how to visualize uncertainty, from whatever source it comes. The two, who direct the Midwest Uncertainty (MU) Collective, have posted an approachable introduction to their field on the Internet platform medium.com ("Uncertainty + Visualization, Explained") [*see illustrations on next page*].

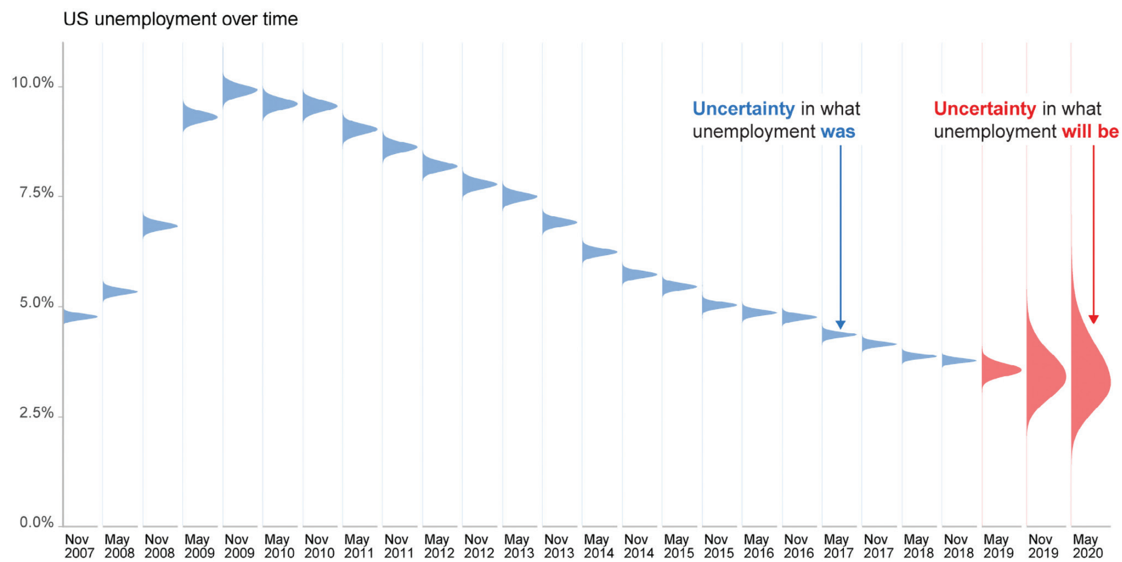


WEATHERNERDS.ORG

About September 1, 2019, multiple meteorological projections of the path of Hurricane Dorian, when taken together, suggest that the hurricane, traveling from its location lower right, will slow over Grand Bahama Island (bottom circle) and not cross into the U.S. mainland. Such a visualization, called an ensemble (or, for its appearance, a "spaghetti plot") may not reassure residents along the coast.



Uncertainty in predicting the U.S. unemployment rate for a particular month is portrayed in three ways, all suggesting the actual value will fall between 3 and 4 percent. The density plot (top), with a rising and falling curve, suggests that the value will most likely fall near the center than the extremes. The violin plot (middle) horizontally reflects the density plot and uses the symmetry in shape to convey the same thing, while the gradient (bottom) uses shading to display the likelihood of different values.



MIDWEST UNCERTAINTY COLLECTIVE

Density plots (drawn here, unfamiliarly, sideways) indicate uncertainty in the unemployment rates that were tabulated each November and May, from 2007 through 2018, as well as what were projected rates for 2019 and 2020. At a glance, the visualization portrays an overall trend while also emphasizing the limitation of prediction (which, because of the pandemic, is now undeniable).

While scientists may be trained to interpret technical presentations of data, user-friendly graphics are usually essential for public communication. A simple chart can evoke emotion, convey a strong message, and possibly prompt action. While designers are exploring and learning more about how data may be visualized, consumers must remain vigilant to visual cues—and fine print—about the uncertainty present and become skeptical when it is not openly acknowledged.

Derya D. Akbaba is a PhD student at the University of Utah's School of Computing. Her focus is on understanding the role that design and technology play in assisting people to find the information they seek, with the goal of critically examining and redefining these systems. Currently, mindful that COVID-19 has had an unequal impact within the United States, Akbaba is working on a grading system for state "dashboards," the collection of visual representations issued by each state for their constituents. Her hope is that such a system will make designers more aware of the impact their data visualizations can have on the public's comprehension of the pandemic.