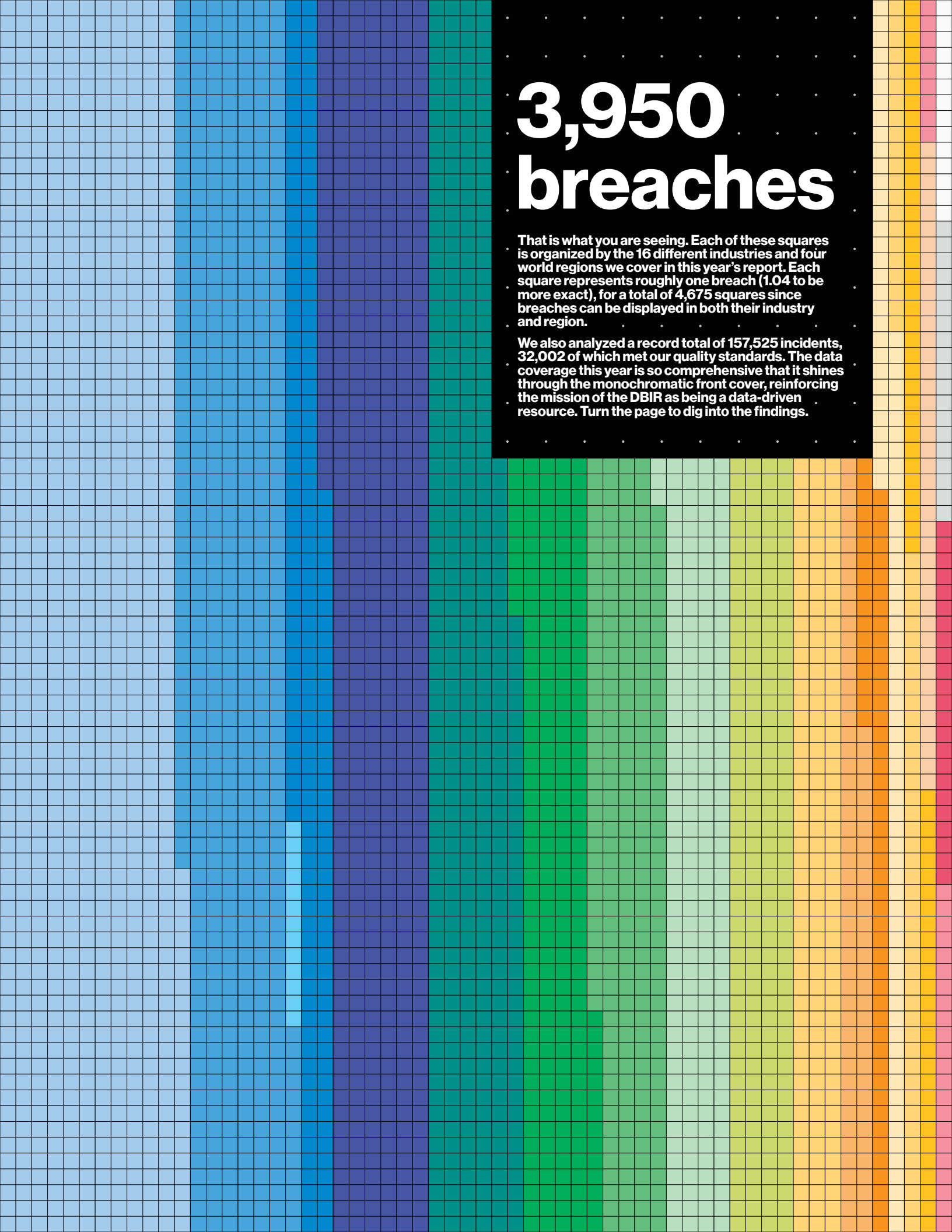
DBIR

2020 Data Breach Investigations Report

verizon



3,950 breaches

That is what you are seeing. Each of these squares is organized by the 16 different industries and four world regions we cover in this year's report. Each square represents roughly one breach (1.04 to be more exact), for a total of 4,675 squares since breaches can be displayed in both their industry and region.

We also analyzed a record total of 157,525 incidents, 32,002 of which met our quality standards. The data coverage this year is so comprehensive that it shines through the monochromatic front cover, reinforcing the mission of the DBIR as being a data-driven resource. Turn the page to dig into the findings.

Table of contents

DBIR Cheat sheet

Hello, and welcome to the *2020 Data Breach Investigations Report* (DBIR)! We have been doing this report for a while now, and we appreciate that all the verbiage we use can be a bit obtuse at times. We use very deliberate naming conventions, terms and definitions and spend a lot of time making sure we are consistent throughout the report. Hopefully, this section will help make all of those more familiar.

VERIS resources

The terms "threat actions," "threat actors" and "varieties" will be referenced a lot. These are part of the Vocabulary for Event Recording and Incident Sharing (VERIS), a framework designed to allow for a consistent, unequivocal collection of security incident details. Here is how they should be interpreted:

Threat actor: Who is behind the event? This could be the external "bad guy" that launches a phishing campaign or an employee who leaves sensitive documents in their seat-back pocket.

Threat action: What tactics (actions) were used to affect an asset? VERIS uses seven primary categories of threat actions: Malware, Hacking, Social, Misuse, Physical, Error and Environmental. Examples at a high level are hacking a server, installing malware and influencing human behavior through a social attack.

Variety: More specific enumerations of higher-level categories, e.g., classifying the external "bad guy" as an organized criminal group or recording a hacking action as SQL injection or brute force.

Learn more here:

- **github.com/vz-risk/dbir/tree/gh-pages/2020** includes DBIR facts, figures and figure data.
- **veriscommunity.net** features information on the framework with examples and enumeration listings.
- **github.com/vz-risk/veris** features the full VERIS schema.
- **github.com/vz-risk/vcdb** provides access to our database on publicly disclosed breaches, the VERIS Community Database.
- **http://veriscommunity.net/veris_webapp_min.html** allows you to record your own incidents and breaches. Don't fret, it saves any data locally and you only share what you want.

Incident vs breach

We talk a lot about incidents and breaches and we use the following definitions:

Incident: A security event that compromises the integrity, confidentiality or availability of an information asset.

Breach: An incident that results in the confirmed disclosure—not just potential exposure—of data to an unauthorized party.

Industry labels

We align with the North American Industry Classification System (NAICS) standard to categorize the victim organizations in our corpus. The standard uses two- to six-digit codes to classify businesses and organizations. Our analysis is typically done at the two-digit level. We will specify NAICS codes along with an industry label. For example, a chart with a label of Financial (52) is not indicative of 52 as a value. "52" is the NAICS code for the Finance and Insurance sector. The overall label of "Financial" is used for brevity within the figures. Detailed information on the codes and classification system is available here:

https://www.census.gov/cgi-bin/sssd/naics/naicsrch?chart=2012

Dotting the charts and crossing the confidence

Last year, we introduced our now (in)famous slanted bar charts to show the uncertainty due to sampling bias.[1] One tweak we added this year was to roll up an "Other" aggregation of all the items that do not make the cut on our "Top (whatever)" charts. This will give you a better sense of the things we left out.

Not to be outdone this year, our incredible team of data scientists decided to try dot plots[2] to provide a better way to show how values are distributed.

The trick to understanding this chart is that the dots represent organizations. So if there are 100 dots (like in each chart in Figure 1), each dot represents 1% of organizations.

1 Check "New chart, who dis?" in the "A couple of tidbits" section on the inside cover of the 2019 DBIR if you need a refresher on the slanted bar charts.
2 To find out more about dot plots, check out Matthew Kay's paper: http://www.mjskay.com/papers/chi2018-uncertain-bus-decisions.pdf

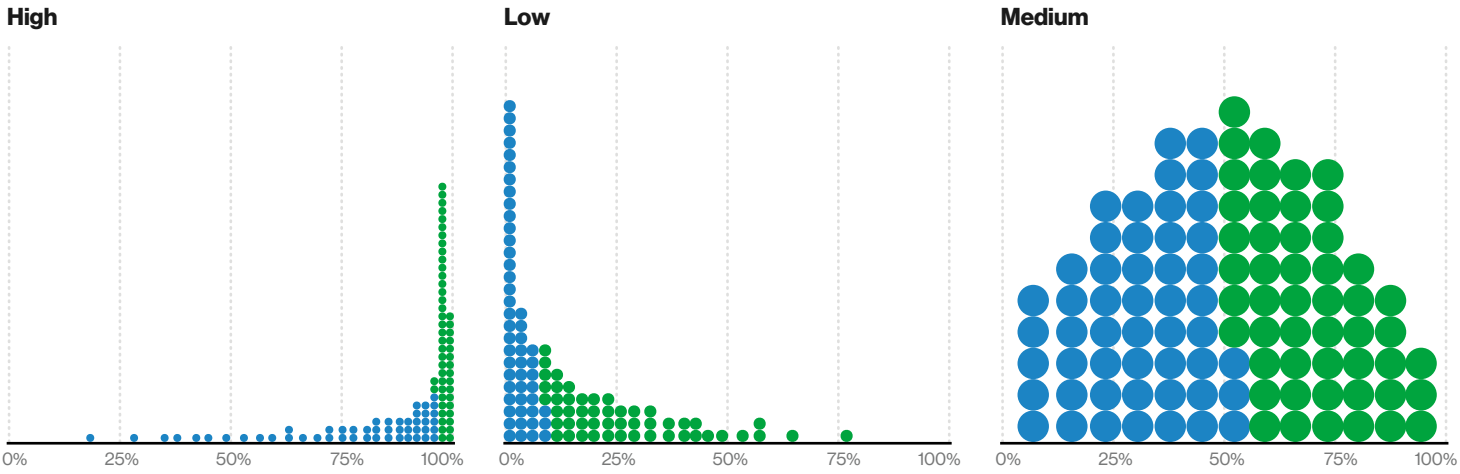


Figure 1. Example dot plots

In Figure 1, we have three different charts, each representing common distributions you may find in this report. For convenience, we have colored the first half and the second half differently so it's easier to locate the median.

In the first chart (High), you see that a lot of companies had a very large value[3] associated with them. The opposite is true for the second one (Low), where a large number of the companies had zero or a low value. On the third chart (Medium), we got stuck in the middle of the road and all we can say is that most companies have that middle value. Using the Medium chart, we could probably report an average or a median value. For the High and Low ones, an average is statistically undefined and the median would be a bit misleading. We wouldn't want to do you like that.

Questions? Comments? Still mad because VERIS uses the term "Hacking"?

Let us know! Drop us a line at dbir@verizon.com, find us on LinkedIn, tweet @VerizonBusiness with the #dbir. Got a data question? Tweet @VZDBIR!

3 Don't worry about what the value is here. We made it up to make the charts pretty. And don't worry later either, we'll use a real value for the rest of the dot plots.