Discover stories within data using SandDance, a new Microsoft Research project

March 23 2016

Data can be daunting. But within those numbers and spreadsheets is a wealth of information. There are also stories that the data can tell, if you're able to see them. SandDance, a new Microsoft Garage project from Microsoft Research, helps you visually explore data sets to find stories and extract insights. It uses a free Web and touch-based interface to help users dynamically navigate through complex data they upload into the tool.

While data science experts will find that SandDance is a powerful tool, its ease of use can help people who aren't experts in data science or programming the ability to analyze information – and present it – in a way that is accessible to a wider audience.

"We had this notion that a lot of visualization summarized data, and that summary is great, but sometimes you need the individual elements of your data set too," says Steven Drucker, a principal researcher who's focused on information visualization and data collections. "We don't want to lose sight of the trees because of the forest, but we also want to see the forest and the overall shape of the data. With this, you'll see information about individuals and how they're relative to each other. Most tools show one thing or the other. With SandDance, you can look at data from many different angles."

Today, SandDance is available in two parallel versions: a standalone

Web-based tool and a custom Power BI visual.

"Using the Microsoft Garage as the release platform gives us the freedom to run experiments with the more accessible standalone version, and as we learn what you like and what works, we can add the right parts to the [Power BI](link) visual," says Drucker. SandDance will be announced as part of Power BI at the [Data Insights Summit](link) on March 22.

The standalone SandDance experience provides a way to organize all these elements on the screen, show individual content, but also access how it looks overall, too. Moving particles represent data, and they can "dance" from screen to screen as you select and filter data and show it the way you want to, through 3D scatterplot, maps, charts, histograms and many other options. Other features include guided tours (tutorials), shareable insights, scripts and themes.

The SandDance Power BI visual is different from most other visuals that already exist for Power BI in that it shows all the data organized into aggregations, allowing users to see both overall patterns as well as individual outliers.

"SandDance simply rocks! It represents not only fresh, ground-breaking data visualization but also an inspiring partnership with Microsoft Research that has produced amazing results," says Nick Caldwell, general manager for Power BI. "Most importantly, our customers will love SandDance and the possibilities it unlocks for exploring their data."

"When people are looking at a bar chart, they're probably wondering what the bars represent, and are they averages? Or sums?" Drucker says. "With SandDance, people can have a good understanding of what the data means. And then they see these animated transitions, and it means the series of visualizations they're seeing are linked together in their minds and how they relate to each other."

And while there are many other ways and tools to present data, they're mostly linear. With SandDance, you can see where tangents take you, and follow it to a new discovery.

"Flipping around to different views makes the story you're telling easier to understand," Drucker says.

As an example, you can take any kind of results – like those of the primaries – from thousands of counties. You'd load that information as an Excel data file, and then you could see how every row represents a county. You can assign colors to easily see who won those counties, how well they did in different parts of the country and drill down the demographics on the voters in those areas. Swing counties start to emerge, as do other outliers, and other nuances that add depth to analysis.

"This experience is a lot about exploration, but it's also about storytelling," Drucker says.

In demos, he uses a data set from the Titanic to illustrate who survived and who didn't. You can break it down by gender, and see that more men died than women, but that doesn't tell the whole story. Most who perished were crew, and amongst the passengers, ticket class determined survival rates – for the most part. Going through SandDance, you find out some didn't pay as much for their tickets, but still ended up in first class. What's the story behind that? That's what SandDance helps reveal.

Each passenger is a data point in SandDance, so you can click on one and go to Bing to find out more information about individuals.

Being able to actually touch the screen and spin around the imagery in 3D also adds to the compelling nature of the tool, and why the researchers think it's going to be popular with non-datageeks.

"We think of it as an immersive, intuitive way to explore your data and also to tell stories about your data," says Roland Fernandez, a software engineer who's the developer on the team. Early on, they experimented with particle animation in HTML, C# and WebGL. He works on the animations and data visualizations. He and Drucker are on the Visualizations in Business and Entertainment team within Microsoft Research. "People are attracted to this because it's so visual."

Drucker and Fernandez can see all sorts of uses for the tool, such as diving deeper into casualties from cars (and other vehicles), driving scores and insurance rates. You could even use it in planning a trip, pinpointing places to go based on several variables, or figure out where your money is going and use it to budget better.

"This can help you connect data with people," Drucker says.

SandDance is releasing through the [Garage](link), the official outlet for experimental projects from teams across Microsoft, to continue testing new and exciting data experiments.

Provided by Microsoft