

## Unit 10 DAVID MCCANDLESS

# The Beauty of Data Visualization

### Part 1

... So, I've been working as a data journalist for about a year, and I keep hearing a phrase all the time, which is this: "Data is the new oil." Data is the kind of **ubiquitous**<sup>1</sup> resource that we can shape to provide new innovations and new insights, and it's all around us, and it can be **mined**<sup>2</sup> very easily. It's not a particularly great metaphor in these times, **especially if you live around the Gulf of Mexico**<sup>3</sup>, but I would, perhaps, adapt this metaphor slightly, and I would say that data is the new soil. Because for me, it feels like a fertile, creative medium. Over the years, online, we've laid down a huge amount of information and data, and we irrigate it with networks and connectivity, and it's been worked and tilled by unpaid workers and governments. And, all right, I'm kind of milking the metaphor a little bit. But it's a really fertile medium, and it feels like visualizations, infographics, data visualizations, they feel like flowers blooming from this medium. But if you look at it directly, it's just a lot of numbers and disconnected facts. But if you start working with it and playing with it in a certain way, interesting things can appear and different patterns can be revealed.

Let me show you this. Can you guess what this data set is? What rises twice a year, once in Easter and then two weeks before Christmas, has a mini peak every Monday, and then flattens out over the summer? I'll take answers. [Audience: Chocolate.] David McCandless: Chocolate. You might want to get some chocolate in. Any other guesses? [Audience: Shopping.] DM: Shopping. Yeah, retail therapy might help. [Audience: Sick leave.] DM: Sick leave. Yeah, you'll definitely want to take some time off. Shall we see?

So, the information **guru**<sup>4</sup> Lee Byron and myself, we **scraped**<sup>5</sup> 10,000 **status Facebook updates**<sup>6</sup> for the phrase "break-up" and "broken-up" and this is the pattern we found—people clearing out for Spring Break, [Laughter] coming out of very bad weekends on a Monday, being single over the summer, and then the lowest day of the year, of course: Christmas Day. Who would do that? So there's a titanic amount of data out there now, unprecedented. But if you ask the right kind of question, or you work it in the right kind of way, interesting things can emerge. . . .

### Part 2

We need relative figures that are connected to other data so that we can see a **fuller picture**<sup>7</sup>, and then that can lead to us changing our perspective. As **Hans Rosling**<sup>8</sup>, the master, my master, said, "Let the dataset change your mindset." And if it can do that, maybe it can also change your behavior.

Take a look at this one. I'm a bit of a health nut. I love taking supplements and being fit, but I can never understand what's going on in terms of evidence. There's always conflicting evidence. Should I take vitamin C? Should I be taking wheatgrass? This is a visualization of all the evidence for nutritional supplements. This kind of diagram is called a balloon race. So the higher up the image, the more evidence there is for each supplement. And the bubbles correspond to popularity as regards to **Google hits**<sup>9</sup>. So you can immediately apprehend the relationship between efficacy and popularity, but you can also, if you grade the evidence, do a "worth it" line. So supplements above this line are worth investigating, but only for the conditions listed below, and

1- Something that is "ubiquitous" seems to be everywhere.

2- In this case, "mine" means to exploit a resource.

3- When McCandless mentions the Gulf of Mexico, he is referring to the oil spill that occurred there in the spring of 2010. McCandless's talk was given in July of the same year.

4- A "guru" is an expert on and teacher of a specific topic.

5- In computer terminology, to "scrape" means to copy data using a program.

6- The "status Facebook updates" that McCandless used in his research are posts that individuals made on their social networking profile pages.

7- The expression "fuller picture" refers to seeing the whole situation, instead of only a specific part of it. A similar expression is "the big picture."

8- Hans Rosling is a statistics expert from Sweden and a TED speaker. Like McCandless, he is well known for using infographics in his work.

9- "Google hits" refers to the list of results when a term is put into the search engine Google.

then the supplements below the line are perhaps not worth investigating.

Now this image constitutes a huge amount of work. We scraped like 1,000 studies from PubMed, the biomedical database, and we compiled them and graded them all. And it was incredibly frustrating for me because I had a book of 250 visualizations to do for **my book**<sup>10</sup>, and I spent a month doing this, and I only filled two pages. But what it points to is that visualizing information like this is a form of knowledge compression. It's a way of squeezing an enormous amount of information and understanding into a small space. And once you've **curated**<sup>11</sup> that data, and once you've cleaned that data, and once it's there, you can do cool stuff like this.

So I converted this into an interactive app, so I can now generate this application online—this is the visualization online—and I can say, “Yeah, brilliant.” So it **spawns**<sup>12</sup> itself. And then I can say, “Well, just show me the stuff that affects heart health.” So let's filter that out. So heart is filtered out, so I can see if I'm curious about that. I think, “No, no. I don't want

to take any synthetics, I just want to see plants and—just show me herbs and plants. I've got all the natural ingredients.” Now this app is spawning itself from the data. The data is all stored in a Google Doc, and it's literally generating itself from that data. So the data is now alive; this is a living image, and I can update it in a second. New evidence comes out. I just change a row on a spreadsheet. **Doosh!**<sup>13</sup> Again, the image recreates itself. So it's cool. It's kind of living. . . .

So, just to **wrap up**<sup>14</sup>, I wanted to say that it feels to me that design is about solving problems and providing elegant solutions, and information design is about solving information problems. It feels like we have a lot of information problems in our society at the moment, from the overload and the saturation to the breakdown of trust and reliability and runaway skepticism and lack of transparency, or even just interestingness. I mean, I find information just too interesting. It has a magnetic quality that draws me in. . . .

*This is an edited version of McCandless's 2010 TED Talk. To watch the full talk, visit TED.com.*

<sup>10</sup> When he talks about “my book,” McCandless is likely referring to a book called *The Visual Miscellaneum* which McCandless authored in 2009. He has since published two other books of his infographics, *Information is Beautiful* and *Knowledge is Beautiful*.

<sup>11</sup> To “curate” data means to choose it and organize it.

<sup>12</sup> To “spawn” means to produce or generate something. McCandless uses this term twice in his talk to explain how an interactive infographic re-creates itself when a user specifies which data to include.

<sup>13</sup> The word “Doosh!” that McCandless uses here works as a kind of sound effect. While it is not a widely used expression, McCandless uses “Doosh!” a few times in his full-length TED talk to create a sense of drama when showing his infographics.

<sup>14</sup> Note that McCandless uses the same term that McGonigal did in Unit 5 when finishing the talk: “wrap up.”