

When you ask people what data is, most reply with a vague description of something that resembles a spreadsheet or a bucket of numbers. The more technically savvy might mention databases or warehouses. However, this is just the format that the data comes in and how it is stored, and it doesn't say anything about what data is or what any particular dataset represents. It's an easy trap to fall in because when you ask for data, you usually get a computer file, and it's hard to think of computer output as anything but just that. Look beyond the file though, and you get something more meaningful.

WHAT DATA REPRESENTS

Data is more than numbers, and to visualize it, you must know what it represents. Data represents real life. It's a snapshot of the world in the same way that a photograph captures a small moment in time.

Look at Figure 1-1. If you were to come across this photo, isolated from everything else, and I told you nothing about it, you wouldn't get much out of it. It's just another wedding photo. For me though, it's a happy moment during one of the best days of my life. That's my wife on the left, all dolled up, and me on the right, wearing something other than jeans and a T-shirt for a change. The

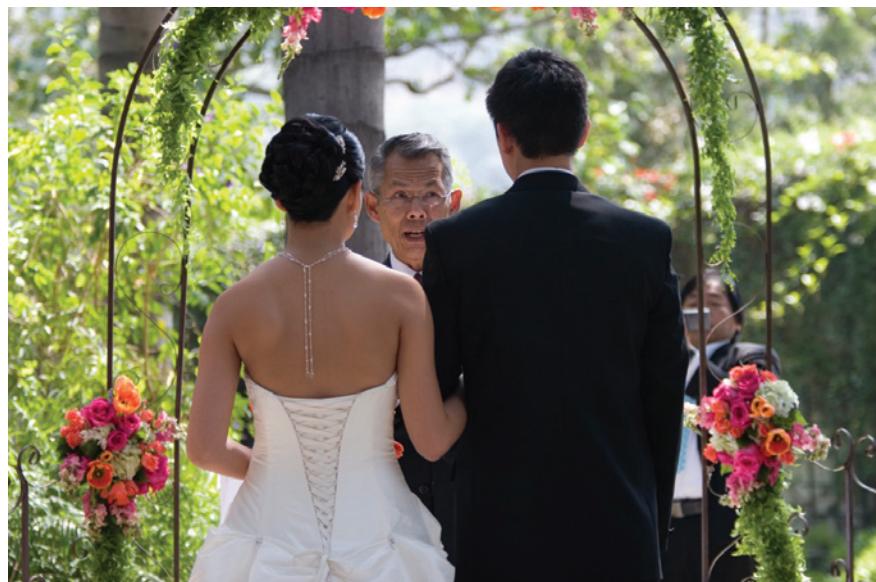


FIGURE 1-1 A single photo, a single data point

pastor who is marrying us is my wife's uncle, who added a personal touch to the ceremony, and the guy in the back is a family friend who took it upon himself to record as much as possible, even though we hired a photographer. The flowers and archway came from a local florist about an hour away from the venue, and the wedding took place during early summer in Los Angeles, California.

That's a lot of information from just one picture, and it works the same with data. (For some, me included, pictures are data, too.) A single data point can have a who, what, when, where, and why attached to it, so it's easy for a digit to become more than a toss in a bucket. Extracting information from a data point isn't as easy as looking at a photo, though. You can guess what's going on in the photo, but when you make assumptions about data, such as how accurate it is or how it relates to its surroundings, you can end up with a skewed view of what your data actually represents. You need to look at everything around, find context, and see what your dataset looks like as a whole. When you see the full picture, it's much easier to make better judgments about individual points.

Imagine that I didn't tell you those things about my wedding photo. How could you find out more? What if you could see pictures that were taken before and after?



FIGURE 1-2 Grid of photos

Now you have more than just a moment in time. You have several moments, and together they represent the part of the wedding when my wife first walked out, the vows, and the tea drinking ceremony with the parents and my grandma, which is customary for Chinese weddings. Like the first photo, each of these has its own story, such as my father-in-law welling up as he gave away his daughter or how happy I felt when I walked down the aisle with my bride. Many of the photos captured moments that I didn't see from my point of view during the wedding, so I almost feel like an outsider looking in, which is probably how you feel. But the more I tell you about that day, the less obscure each point becomes.

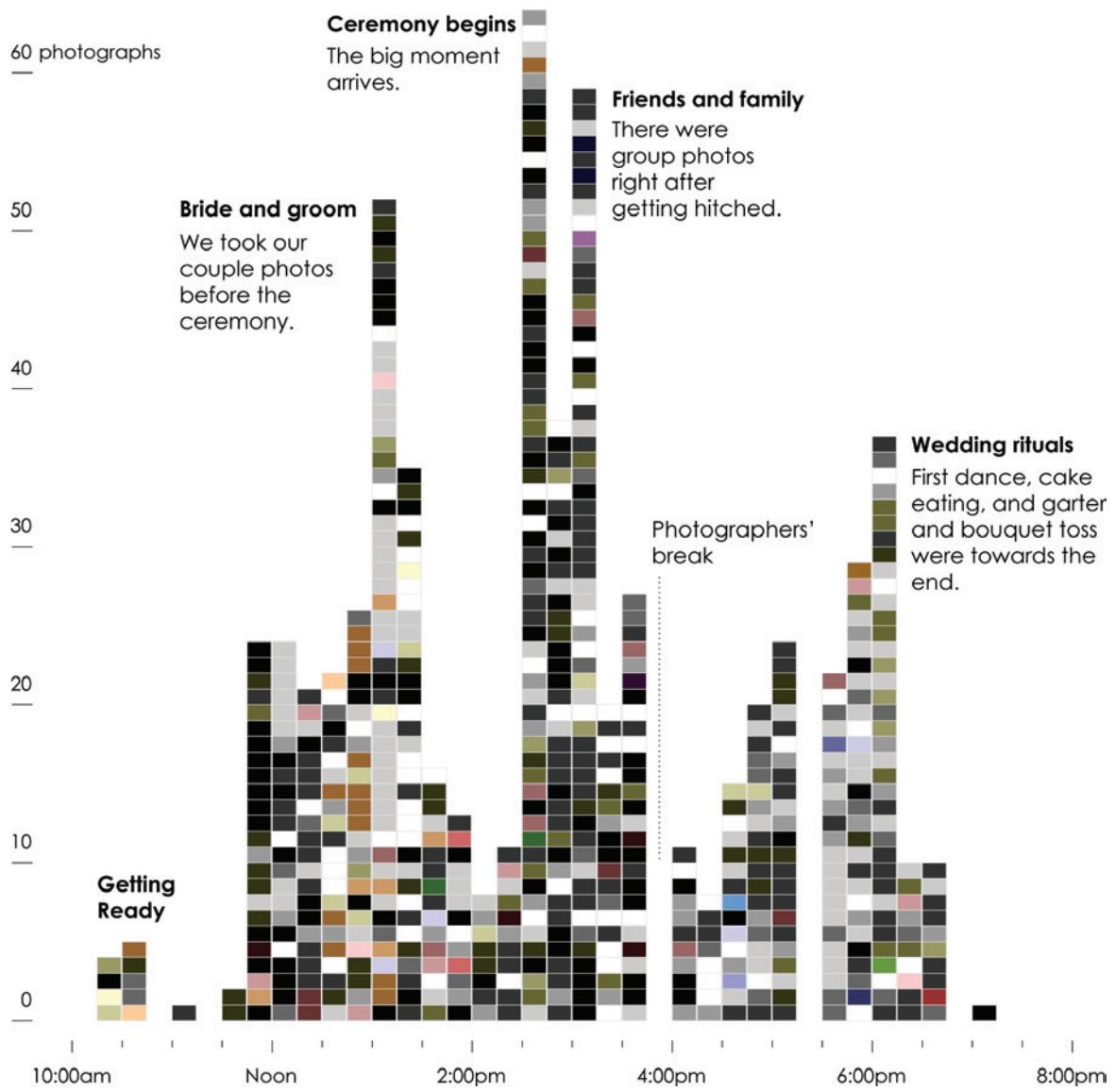
Still though, these are snapshots, and you don't know what happened in between each photo. (Although you could guess.) For the complete story, you'd either need to be there or watch a video. Even with that, you'd still see only the ceremony from a certain number of angles because it's often not feasible to record every single thing. For example, there was about five minutes of confusion during the ceremony when we tried to light a candle but the wind kept blowing it out. We eventually ran out of matches, and the wedding planner went on a scramble to find something, but luckily one of our guests was a smoker, so he busted out his lighter. This set of photos doesn't capture that, though, because again, it's an abstraction of the real thing.

This is where sampling comes in. It's often not possible to count or record everything because of cost or lack of manpower (or both), so you take bits and pieces, and then you look for patterns and connections to make an educated guess about what your data represents. The data is a simplification—an abstraction—of the real world. So when you visualize data, you visualize an abstraction of the world, or at least some tiny facet of it. Visualization is an abstraction of data, so in the end, you end up with an abstraction of an abstraction, which creates an interesting challenge.

However, this is not to say that visualization obscures your view—far from it. Visualization can help detach your focus from the individual data points and explore them from a different angle—to see the forest for the trees, so to speak. To keep running with this wedding photo example, Figure 1-3 uses the full wedding dataset, of which Figure 1-1 and Figure 1-2 were subsets of. Each rectangle represents a photo from our wedding album, and they are colored by the most common shade in each photo and organized by time.

Wedding colors

Each rectangle represents a photograph during my wedding, and each is filled with the most common color in the picture.



With a time series layout, you can see the high points of the wedding, when our photographers snapped more shots, and the lulls, when only a few photos were taken. The peaks in the chart, of course, occur when there is something to take pictures of, such as when I first saw my wife in her dress or when the ceremony began. After the ceremony, we took the usual group photos with friends and family, so there was another spike at that point. Then there was food, and activity died down, especially when the photographers took a break a little before 4 o'clock. Things picked up again with typical wedding fanfare, and the day came to an end around 7 in the evening. My wife and I rode off into the sunset.

In the grid layout, you might not see this pattern because of the linear presentation. Everything seems to happen with equal spacing, when actually most pictures were taken during the exciting parts. You also get a sense of the colors in the wedding at a glance: black for the suits, white for the wedding dress, coral for the flowers and bridesmaids, and green for the trees surrounding the outdoor wedding and reception. Do you get the detail that you would from the actual photos? No. But sometimes that level isn't necessary at first. Sometimes you need to see the overall patterns before you zoom in on the details. Sometimes, you don't know that a single data point is worth a look until you see everything else and how it relates to the population.

You don't need to stop here, though. Zoom out another level to focus only on the picture-taking volumes, and disregard the colors and individual photos, as shown in Figure 1-4.

You've probably seen this layout before. It's a bar chart that shows the same highs and lows as in Figure 1-3, but it has a different feel and provides a different message. The simple bar chart emphasizes picture-taking volumes over time via 15-minute windows, whereas Figure 1-3 still carries some of the photo album's sentiment.

The main thing to note is that all four of these views show the same data, or rather, they all represent my wedding day. Each graphic just represents the day differently, focusing on various facets of the wedding. Interpretation of the data changes based on the visual form it takes on. With traditional data, you typically examine and explore from the bar chart side of the spectrum, but that doesn't mean you have to lose the sentiment of the individual data point—that single photo. Sometimes that means adding meaningful annotation that enables readers to interpret the data better, and other times the message in the numbers is clear, gleaned from the visualization itself.

Photographs over time

Our wedding photographers snapped more pictures during the significant events with a peak of 63 during a 15-minute span.

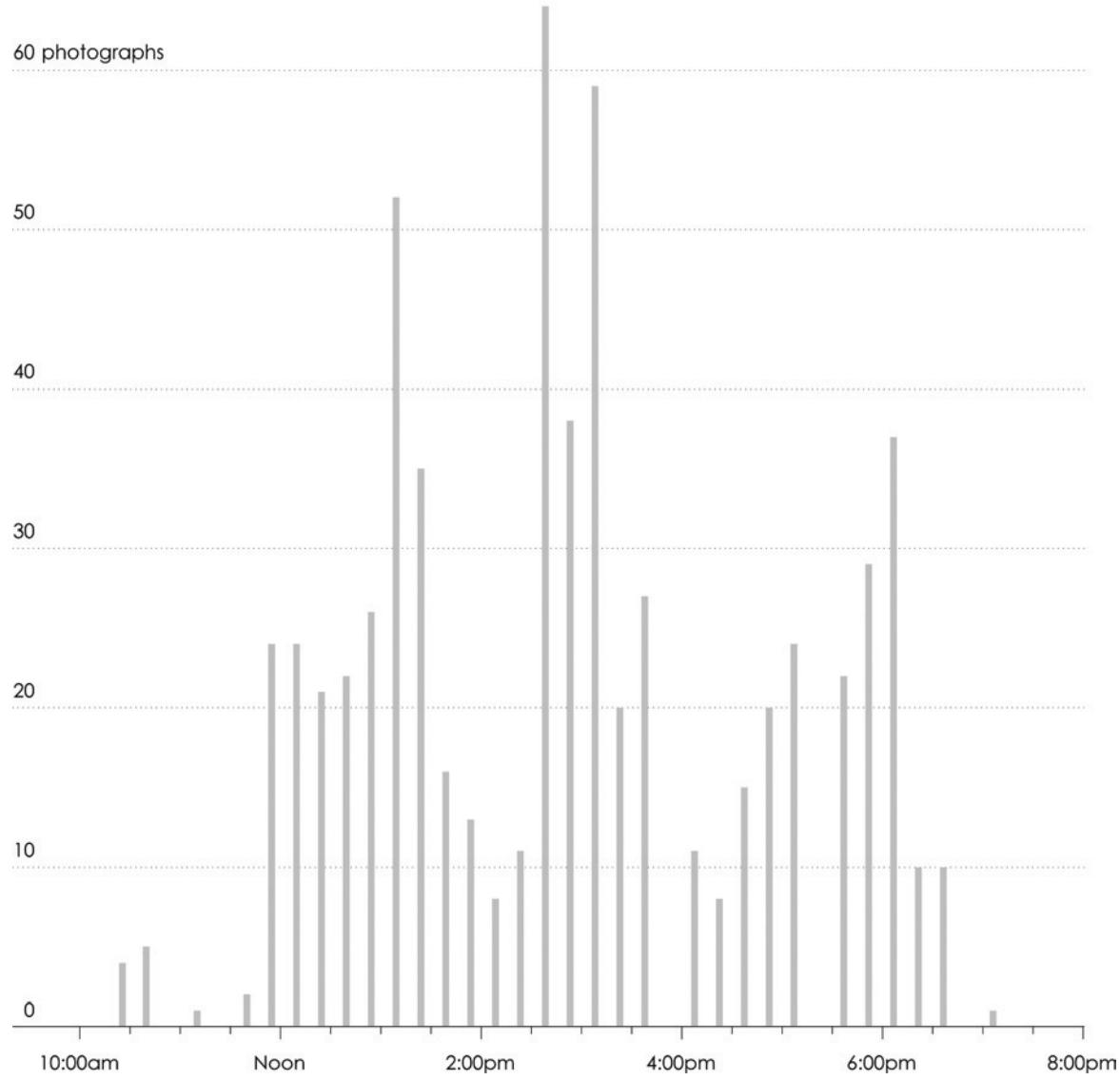


FIGURE 1-4 Photos over time

The connection between data and what it represents is key to visualization that means something. It is key to thoughtful data analysis. It is key to a deeper understanding of your data. Computers do a bulk of the work to turn numbers into shapes and colors, but you must make the connection between data and real life, so that you or the people you make graphics for extract something of value.

This connection is sometimes hard to see when you look at data on a large scale for thousands of strangers, but it's more obvious when you look at data for an individual. You can almost relate to that person, even if you've never met him or her. For example, Portland-based developer Aaron Parecki used his phone to collect 2.5 million GPS points over 3½ years between 2008 and 2012, about one point every 2 to 6 seconds. Figure 1-5 is a map of these points, colored by year.

As you'd expect, the map shows a grid of roads and areas where Parecki frequented that are colored more brightly than others. His housing changed a few times, and you can see his travel patterns change over the years. Between 2008 and 2010, shown in blue, travel appears more dispersed, and by 2012, in yellow, Parecki seems to stay in a couple of tighter pockets. Without more context it is hard to say anything more because all you see is location, but to Parecki the data is more personal (like the single wedding photo is to me). It's the footprint of more than 3 years in a city, and because he has access to the raw logs, which have time attached to them, he could also make better decisions based on data, like when he should leave for work.

What if there were more information attached to personal time and location data, though? What if along with where you were, you also took notes during or after about what was going on at some given time? This is what artist Tim Clark did between 2010 and 2011 for his project *Atlas of the Habitual*. Like Parecki, Clark recorded his location for 200 days with a GPS-enabled device, which spanned approximately 2,000 miles in Bennington, Vermont. Clark then looked back on his location data and labeled specific trips, people he spent time with, and broke it down by time of year.

As shown in Figure 1-6, the atlas, with clickable categorizations and time frames, shows a 200-day footprint that reads like a personal journal. Select "Running errands" and the note reads, "Doing the everyday things from running to the grocery store all the way to driving 30 miles to the only bike shop in southern Vermont opened on Sundays." The traces stay around town, with the exception of two long ones that venture out.

There is one entry titled "Reliving the breakup," and Clark writes, "A long-term girlfriend and I broke up immediately before I moved. These are the times that I had a real difficult time coming to terms that I had to move on." Two small paths, one within the city limits and one outside, appear, and the data suddenly feels incredibly personal.

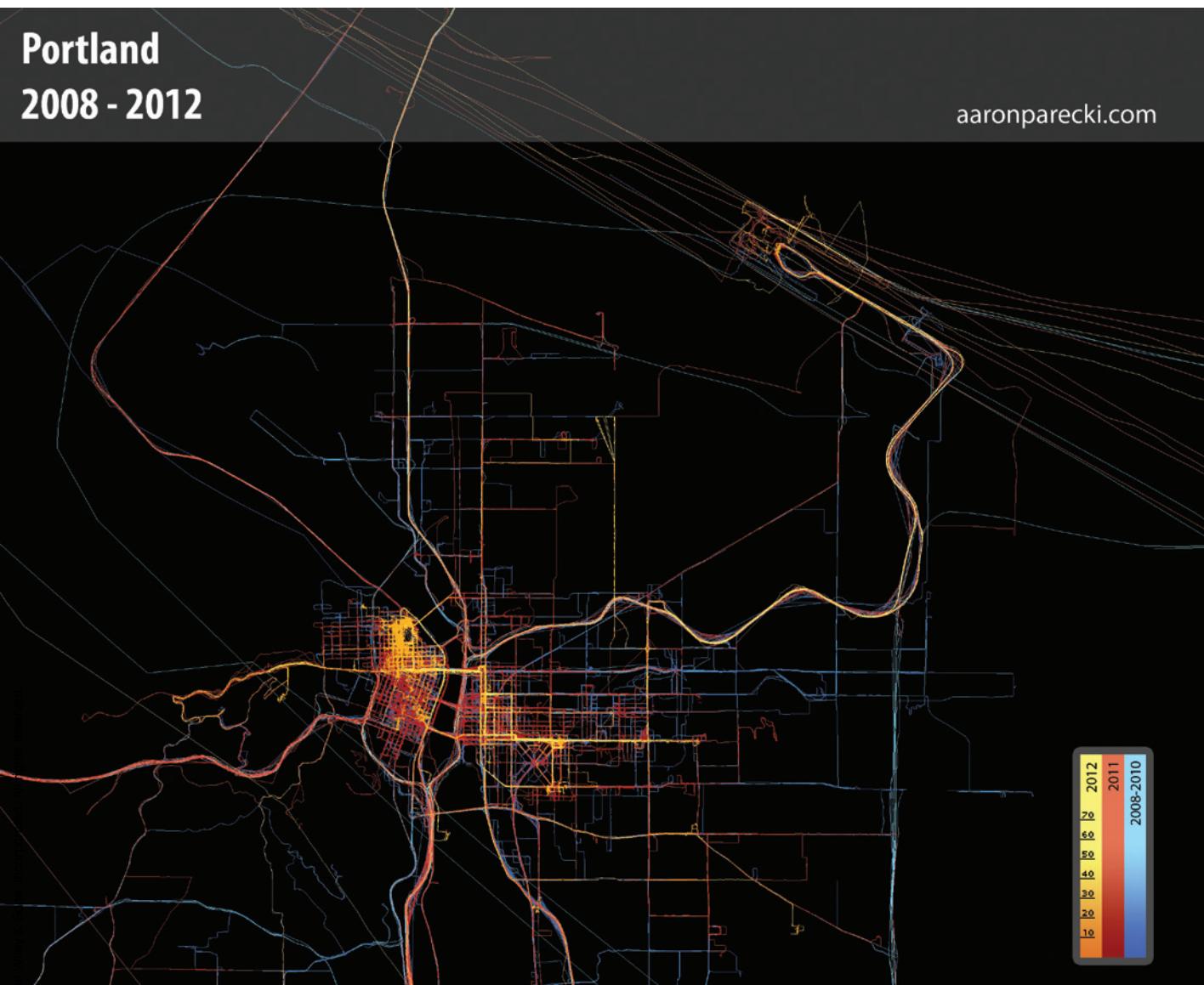


FIGURE 1-5 GPS traces collected by Aaron Parecki, <http://aaronparecki.com>

This is perhaps the appeal behind the Quantified Self movement, which aims to incorporate technology to collect data about one's own activity and habits. Some people track their weight, what they eat and drink, and when they go to bed; their goal is usually to live healthier and longer. Others track a wider variety of metrics purely as a way to look in on themselves beyond what they see in the mirror; personal data collection becomes something like a journal for self-reflection at the end of the day.

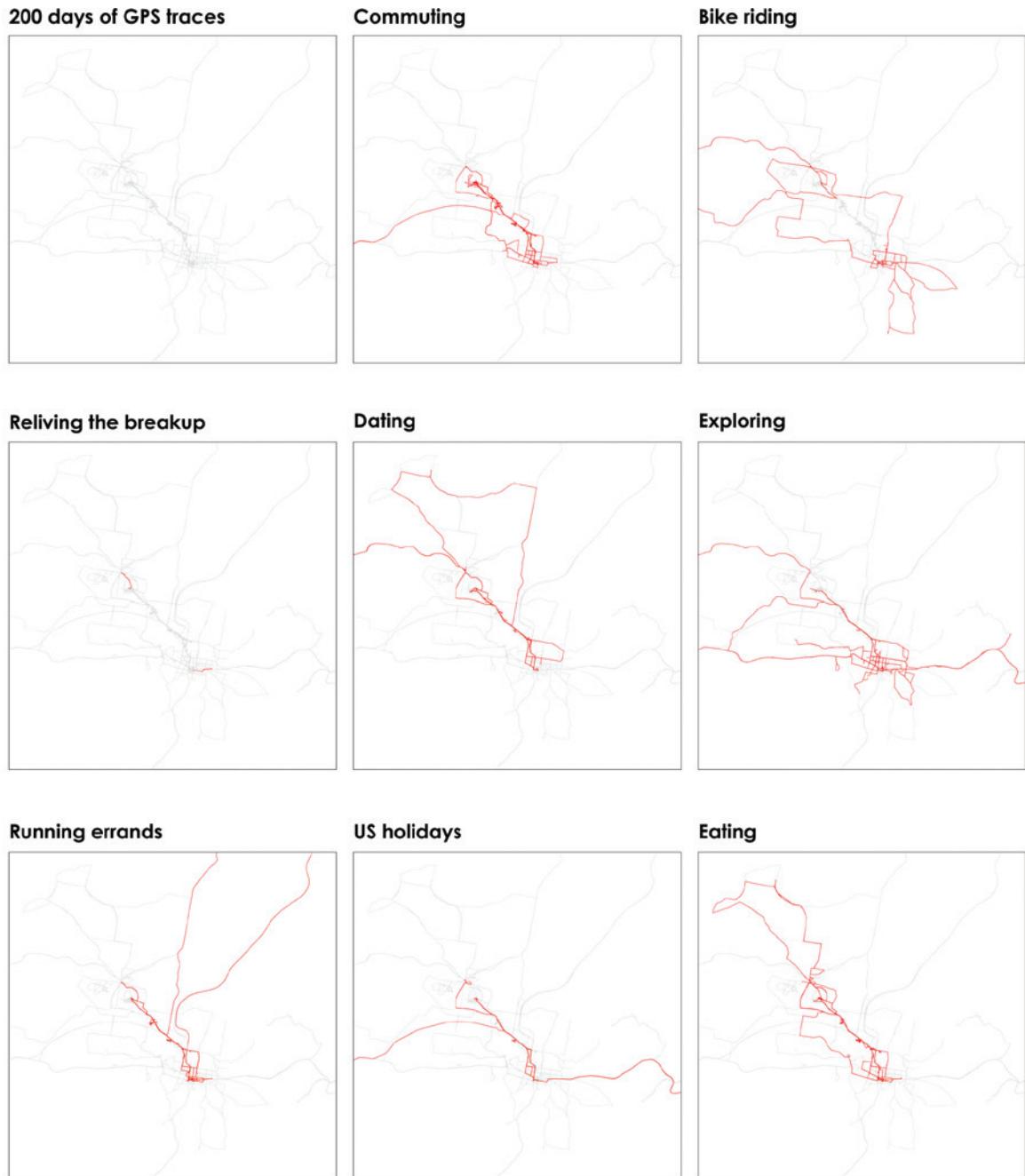


FIGURE 1-6 Selected maps from *Atlas of the Habitual* by Tim Clark, <http://www.tlclark.com/atlasonthehabitual/>

Nicholas Felton is one of the more well-known people in this area for his annual reports on himself, which highlight both his design skills and disciplined personal data collection. He keeps track of not just his location, but also who he spends time with, restaurants he eats at, movies he watches, books he reads, and an array of other things that he reveals each year. Figure 1-7 is a page out of Felton's 2010/2011 report.

Felton designed his first annual report in 2005 and has done one every year since. Individually, they are beautiful to look at and hold and satisfy an odd craving for looking in on a stranger's life. What I find most interesting, though, is the evolution of his reports into something personal and the expanding richness of data. Looking at his first report, as shown in Figure 1-8, you notice that it feels a lot like a design exercise in which there are touches of Felton's personality embedded, but it is for the most part strictly about the numbers. Each year though, the data feels less like a report and more like a diary.

This is most obvious in the *2010 Annual Report*. Felton's father passed away at the age of 81. Instead of summarizing his own year, Felton designed an annual report, as shown in Figure 1-9, that cataloged his father's life, based on calendars, slides, postcards, and other personal items. Again, although the person of focus might be a stranger, it's easy to find sentiment in the numbers.

When you see work like this, it's easy to understand the value of personal data to an individual, and maybe, just maybe, it's not so crazy to collect tidbits about yourself. The data might not be useful to you right away, but it could be a decade from now, in the same way it's useful to stumble upon an old diary from when you were just a young one. There's value in remembering. In many ways you log bits of your life already if you use social sites like Twitter, Facebook, and foursquare. A status update or a tweet is like a mini-snapshot of what you're doing at any given moment; a shared photo with a timestamp can mean a lot decades from now; and a check-in firmly places your digital bits in the physical world.

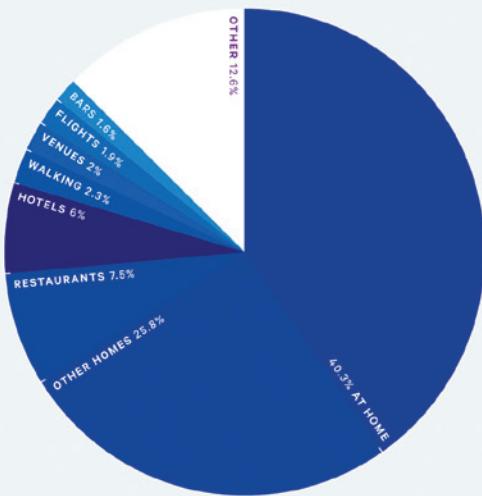
You've seen how that data can be valuable to an individual. What if you look at the data from many individuals in aggregate?

The United States Census Bureau collects the official counts of people living in the country every 10 years. The data is a valuable resource to help officials allocate funds, and from census to census, the fluctuations in population help you see how people move in the country, changing the neighborhood

FIGURE 1-7 (following page)
A page from 2010/2011 Annual Report by Nicholas Felton,
<http://feltron.com>

With Olga

EVERYWHERE



DAYS TOGETHER

191½

315 different encounters

MOST TIME SPENT TOGETHER

MANHATTAN — 63½ DAYS

BROOKLYN — 51½ DAYS

MILL VALLEY — 9 DAYS

ANCHORAGE — 7½ DAYS

SYDNEY — 4½ DAYS

MOST VISITED PLACE TOGETHER

Old Apartment

194 visits

DIFFERENT CITIES VISITED TOGETHER

56

In 3 countries, 9 states and Washington D.C.

FAVORITE BEVERAGES WITH OLGA

FILTER COFFEE — 111 SERVINGS

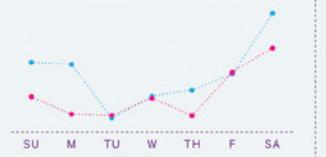
RED WINE — 78 SERVINGS

DALE'S PALE ALE — 35 SERVINGS

CHAMPAGNE — 30 SERVINGS

LATTE — 26 SERVINGS

TIME TOGETHER



BRIEDEST MONTH TOGETHER

June 2011

40 hours

MOST CONSECUTIVE HOURS TOGETHER

247

Australia trip — February 2010

TIME SPENT WITH OLGA AND...

SARAH — 6½ DAYS

MOM — 6½ DAYS

BRIAN — 5½ DAYS

OLGA'S MOM — 5 DAYS

RYAN — 4½ DAYS

WEDDINGS ATTENDED TOGETHER

Seven

Aaron & Jessica, Charlie & Bret, Glenn & Mariana, Lewis & Ange, Randy & Allison, Rob & Elise and Toby & Harriet

With Olga

IN THE BAY AREA



DAYS TOGETHER IN THE BAY AREA

13½

Approximately 7% of total time together

BAY AREA PLACES VISITED TOGETHER

77

18 stores, 13 restaurants, 10 homes, 6 outdoor places, 3 coffee shops, 3 grocery stores, 2 airport terminals, 2 bars, 2 gas stations, 2 hospitals, 2 hotels, 2 liquor stores, 2 parking garages, 2 parking lots, a cinema, a deli, a drug store, a laundromat, a library, a museum, a park and work

FAVORITE BAY AREA BOTTLESHOP

Vintage Wine & Spirits

Visited twice

FAVORITE BAY AREA BEER WITH OLGA

Lagunitas IPA

5 servings

BAY AREA MUSEUMS VISITED TOGETHER

The Exploratorium

With Marina — July 9, 2011

MOST PLAYED ARTIST TOGETHER

The Beach Boys

25 songs listened to from *Christmas with the Beach Boys*

TIME TOGETHER IN THE BAY AREA

Mill Valley

68% of time in the Bay Area

MOST VISITED BAY AREA PLACES

MOM'S HOUSE — 35 VISITS

MARIN GENERAL HOSPITAL — 6 VISITS

CHEVRON MILL VALLEY — 5 VISITS

SFO INTERNATIONAL TERMINAL — 4 VISITS

DAD'S HOUSE — 3 VISITS

CRISES INVOLVING A TICK

One

Spotted by Olga, removed by Mom

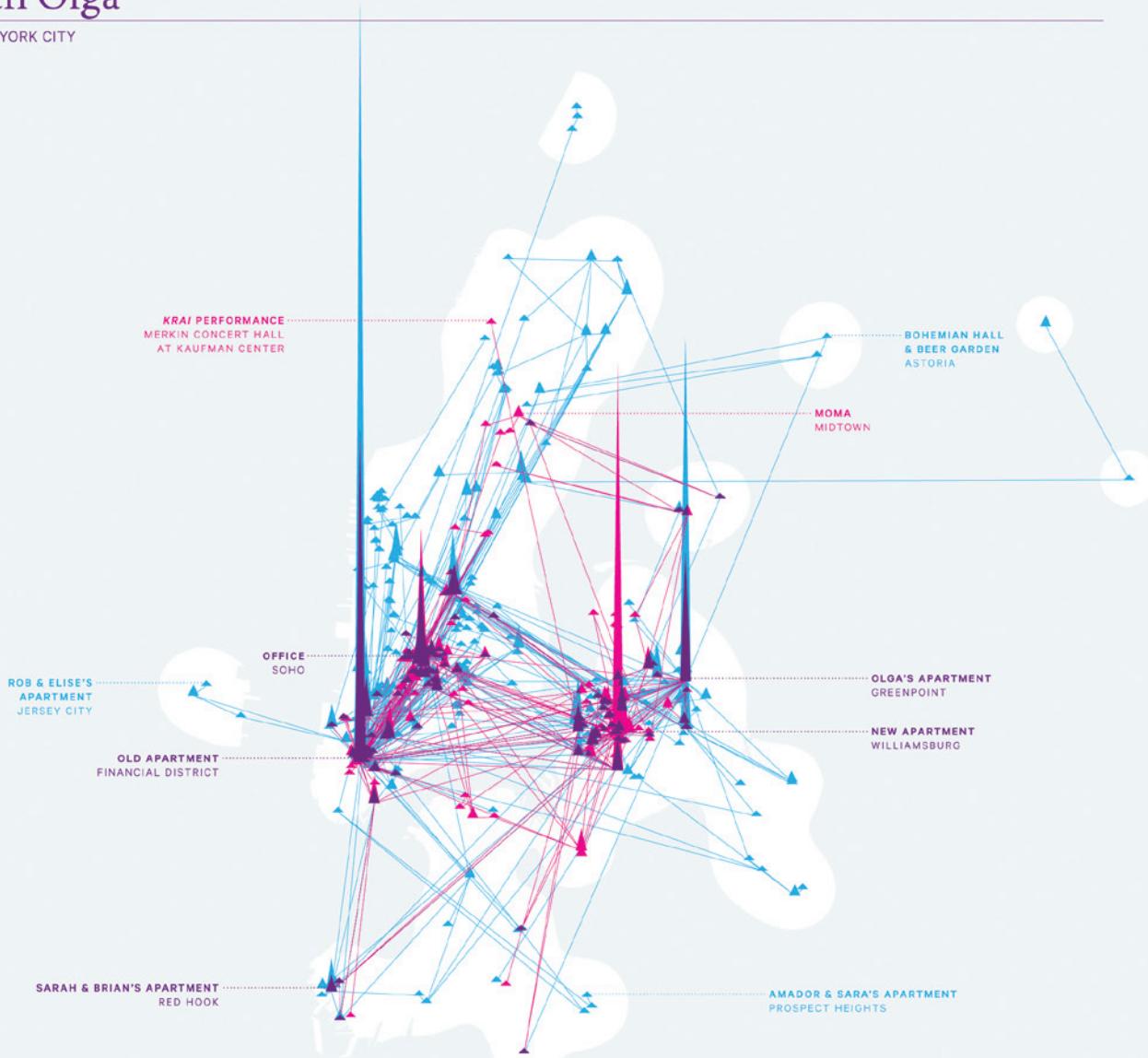
MOST VISITED RESTAURANTS TOGETHER

Le Garage, Picante and Sushi Ran

Each visited twice

With Olga

IN NEW YORK CITY



DAYS TOGETHER IN NEW YORK CITY

136¾

Approximately 72% of total time together

MOST VISITED NYC PLACES

OLD APARTMENT — 194 VISITS

OLGA'S APARTMENT — 84 VISITS

NEW APARTMENT — 87 VISITS

THE OFFICE — 35 VISITS

TAKAHACHI TRIBECA — 21 VISITS

TIME TOGETHER IN NEW YORK CITY



TIME IN NEW YORK SPENT WITH OLGA

31%

5% of time together spent in transit

MOST VISITED NYC RESTAURANTS

TAKAHACHI TRIBECA — 21 VISITS

LES HALLES ON JOHN STREET — 9 VISITS

DINER / ENID'S — 7 VISITS

MILLER'S TAVERN / FIVE LEAVES — 6 VISITS

RABBIT HOLE — 5 VISITS

FAVORITE NYC COCKTAIL WITH OLGA

Bloody Mary

22 servings

NYC PERFORMANCES WITH OLGA

Twenty-Eight

Bell (11), Bear in Heaven (3), Baths + How to Dress Well + Zola Jesus, Blonde Redhead + Pantha du Prince, Dexter Lake Club Band, Jason Nazary, Knights on Earth, Olga Bell Kral, Little Women, Nathan Fake + Four Tet, Now Ensemble + Matmos, Owen Pallett, Panda Bear, Pierre-Laurent Aimard, Sleigh Bells and *The Nose*

SIGNIFICANT NYC MISHAPS

Five

Abandoned keyboard stand, muddled dinner invitation date, missed ferry, shattered martini glass and smashed iPhone

composition, and how areas grow and shrink. In short, the data paints a picture of who lives in America. However, the data, collected and maintained by the government, can show only so much about the individuals, and it's hard to grasp who the people actually are.

What are their likes and dislikes? What kind of personality do they have? Are there major differences between neighboring cities and towns?

Media artist Roger Luke DuBois took a different kind of census, via 19 million online dating profiles in *A More Perfect Union*. When you join an online dating



FIGURE 1-8 Selected pages from 2005 Annual Report by Nicholas Felton, <http://feltron.com>

site, you first describe yourself: who you are, where you're from, and what you're interested in. After you uncomfortably fill out that information, and perhaps choose not to share a thing or two, you describe what your ideal mate is like. In the words of DuBois, in the latter, you tell the complete truth, and in the former, you lie. So when you aggregate people's online dating profiles, you get some combination of how people see themselves and how they want to be seen.

In *A More Perfect Union*, DuBois categorized online dating profiles, digital encapsulations of hopes and dreams, by postal code, and then looked for the word that was most unique to each area. Using a tracing of a Rand McNally map, DuBois replaced each city name with the city's unique word and painted a different picture of the United States: a more recognizable and personal one.

In Figure 1-10, around southern California, where they make the talkies, words such as *acting*, *writer*, and *entertainment* appear; on the other hand, in Washington, DC, shown in Figure 1-11, words like *bureaucrat*, *partisan*, and *democratic* appear. These mostly pertain to professions, but in some areas the words describe personal attributes, favorite things, and major events.

In Louisiana, shown in Figure 1-12, *Cajun* and *curvy* pop out at you, as does *crawfish*, *bourbon*, and *gumbo*, but in New Orleans, the most unique word is *flood*, a reflection of the effects of Hurricane Katrina in 2005.

People are defined by common demographic data such as race, age, and gender, but they also identify themselves with what they like to do in their spare time, what has happened to them, and who they hang around with. The great thing about *A More Perfect Union* is that you can see that in the data on a countrywide scale.

The same sentiment—where data points are recollections and reports are portraits and diaries—is seen in Felton's reports, Clark's atlas, and Parecki's GPS traces. Statisticians and developers call this analysis. Artists and designers call this storytelling. For extracting information from data, though—to understand what's in the numbers—analysis and storytelling are one and the same.

Just like what it represents, data can be complex with variability and uncertainty, but consider it all in the right context, and it starts to make sense.

FIGURE 1-9 (following page)
Selected pages from 2010 Annual Report by Nicholas Felton,
<http://feltron.com>

2000-2010
SAN RAFAEL

THE 21ST CENTURY

138 LOCATIONS

SOCIALIZING

FOREST KNOLLS

SAN DOMENICO

CHALET BASQUE

BOLINAS

STRAWBERRY

CHINA BEACH

ALCATRAZ

BERKELEY

OAKLAND AIRPORT



JAN 5, 2001
71 YEARS,
6 MONTHS
AND 1 DAY

PERSON SEEN
THE MOST

MARINA
117 TIMES

BLACK
PANTHERS
MET

ONE
BOBBY SEALE

WALKS
RECORDED

THIRTY-FIVE
AND 1 HIKE

2009-2010
GOLDEN GATE
TOLL BOOTH
PREFERENCE

LANE
6
14 VISITS

ENTERTAINMENT

123 MOVIES

45 MUSIC

38 LECTURES

29 DANCES

29 POKER

22 PLAYS

19 TELEVISION

8 SLIDE SHOWS

5 ACROBATS

3 MIMES

MOST
WATCHED
TV SHOW

THE OSCARS
8 TIMES

LAST DAY

SEP 12, 2010
81 YEARS, 2 MONTHS
AND 8 DAYS OLD

WEATHER
SEP 12, 2010
3:20 PM

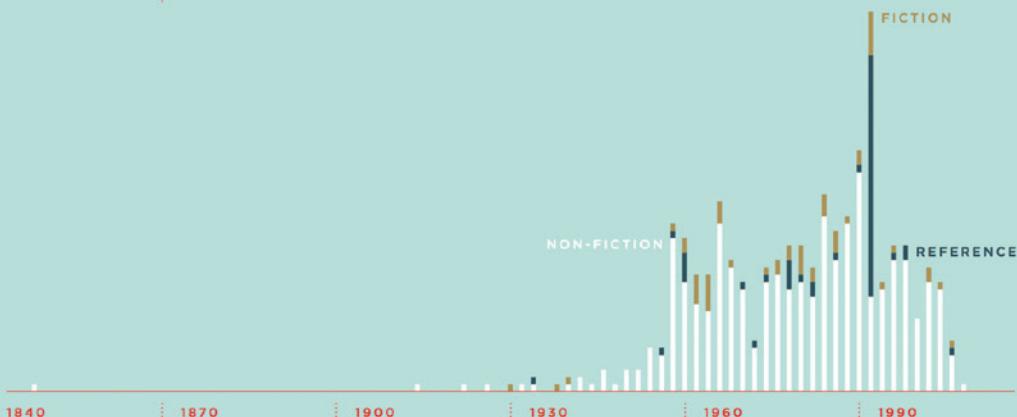
49.8° F AND OVERCAST
LARKSPUR, CALIFORNIA

1848-2009

BOOKS

536 BOOKS

DATE PUBLISHED



BOOKS

561
SPANNING 161 YEARS

MEDIAN
PUBLISHING
DATE

1983
11 BOOKS

REGION
WITH MOST
TRAVEL BOOKS

RUSSIA
6 BOOKS

TRAVEL
BOOKS FOR
UNVISITED
PLACES

SIX
AUSTRALIA, ICELAND, GREENLAND,
IRAN, PAKISTAN AND VENEZUELA

MOST
REPRESENTED
AUTHOR

MARTIN GILBERT
5 BOOKS

TYPES OF
BOOKS

88 TRAVEL
77 HISTORY
42 MACHINES
37 GEOGRAPHY
32 ENCYCLOPEDIA
26 RELIGION
23 HEALTH
23 NOVEL
22 SCIENCE
17 HOW-TO

COOKBOOKS

FIVE

WAR-RELATED
BOOKS

51
35 BOOKS
ABOUT
WORLD WAR 2

ELEVATOR
BOOKS

TWELVE
1941-1991

HOW-TO
TOPICS

FOURTEEN

BICYCLES, CLEANING, CROSS
COUNTRY SKIING, DOING
EVERYTHING RIGHT, HANDICRAFT,
HOME REPAIR, PEST CONTROL,
PHOTOGRAPHY, PREVENTING
AND SURVIVING FIRES, SAILING,
SURVIVAL AND TAI CHI

BIOGRAPHIES

8
FROM
ARMSTRONG
TO STALIN



FIGURE 1-10 California map from *A More Perfect Union* (2011) by R. Luke DuBois, courtesy of the artist and bitforms gallery, New York City, <http://perfect.lukedubois.com>



Washington, District of Columbia

Single Population: 66,174
Words: 676,266

FIGURE 1-11 Washington, DC map from A More Perfect Union (2011)