Luggage_MinnPost_UplinkStory

Notebook: CAR
Created: 6/11/2012 3:16 PM **Updated:** 6/14/2012 9:24 AM
URL: http://ire.org/blog/uplink/2011/06/13/first-venture-pivot-tables-help-show-aiport-los/

First Venture: Pivot tables help show aiport loss claims

By Jeff Severns Guntzel, Minn Post | 06.13.2011

Uplink

TAGGED
transportation, spreadsheet, First Ventures, federal

Leave a comment

Somewhere in the tilting tower of boxes that passes for my personal archive are the files from my first attempt at data journalism—an investigative piece about the Minneapolis Police Department's Sex Crimes Unit. You'll find no CDs with database files in there. No printed spreadsheets either; just a small stack of hand-drawn grids with case counts, closure stats, and other vitals. I don't even know if I had Microsoft Excel installed on my work computer at the time. It never occurred to me that there might be a better way to store and process the information I was gathering. I was still doing the hand-drawn grid trick on the eve of my departure for CAR Boot Camp at the Missouri School of Journalism in January of this year. I was crunching the U.S. Environmental Protection Agency's Toxic Release Inventory numbers for Minnesota and it took me hours. The data came in a spreadsheet, sure, but the rankings and big calculations all happened on paper. Old habits are a funny thing. Boot Camp is a fog to me now—I remember it all as flashes: Pivot tables in Excel changed everything. Structured Query Language (SQL) in Access database manager changed everything some more. Back in my hotel one night I decided to experiment with my new powers. I opened up my EPA data and set about recreating the findings I had scrawled out by hand a few days earlier. I recreated the work in fewer than 15 minutes. For another 15 minutes I went further with the data—going in directions that had not occurred to me the first time through. My first real world venture was the result of lost luggage. I

flew back from Boot Camp but my suitcase did not. In line at the airport to inquire about the bag, I got to wondering: How many people go through this? What kinds of things do they lose? If the bag is lost or damaged (mine came back with a busted handle), what are my odds and getting a little something for my inconvenience? Back home I poked around online and found U.S. Transportation Security Administration claims data on lost or damaged property covering roughly eight-and-a-half years. A pivot table party ensued. What is a pivot table and why party with it? Good question. Simply put, pivot tables allow you to creat subtotals from individual data records quickly. The TSA spreadsheet I was working with had 12 columns and more than 55,000 rows. In a matter of seconds I had a list showing me total number of claims by U.S. airport. Another pivot table gave me a list of Minnesota airports ranked by the amount they paid out in a year. Is this a party you would come to? That's what I thought. Good news for those of you who don't have access to Excel, Google Docs just added pivot tables to its spreadsheets. Here's some of what I found: Just over $215,000 was paid out in response to complaints where a Minnesota airport was listed as the "claim site." That's 1.5 percent of the $13.6 million nationwide total. Of the 2,700 claims involving Minnesota airports filed between January 2002 and May 2010, 45 percent were denied straightaway. Claims for broken luggage (the thing that got me digging through this data in the first place) did about as well. It was a CAR project with a small "c" (can you do this to an acronym?) but it was a victory all the same. Every fruitful sunshine request, each successful enterprise join—hell every productive WHERE statement in SQL is like a toning exercise. I'll be ready when that project or dataset that really matters comes along.

It's thrilling to open up a massive collection of data and know there's a story in there somewhere. It's more thrilling still to know exactly how to proceed. I started requesting data while I was still at Boot Camp. Don't worry about me, I told state agency data people, whatever form the data is in, I can probably handle it. I was cocksure and maybe even a little power-drunk. Pretty soon the files started landing in my lap. One day I was up to my eyeballs in lobbyist data and the next it was state employee salaries. Then there were all the lesser databases. It was like data Christmas (it's okay to wince when I say things like this). I'd hardly be finished cleaning one data set and I'd be tearing into the next. A few weeks out of Boot Camp I had to step back and create some sort of order. This aspiring CAR reporter needed some sort of process. Here's what has helped me so far—maybe it will help you too: Create a data inbox I keep a folder labeled "DATA INBOX" right on my desktop. Every piece of data

I request or download goes into this folder—even those pesky PDF files that I am yet to work my conversion magic on. The data stays in that folder until I'm ready to work with it. Request, repeat Like I said, I started requesting data while I was still at Boot Camp and I haven't stopped. It pays off. In just a few short months I've already lost count of how many times my editor has asked me: "You got anything for next week?" just as a Data Practices Act request is granted. Keep a data journal Every time I open a spreadsheet or database I also open a simple text file. I record every step, from data cleaning through my final SQL queries. I do this because I am an idiot. No matter how careful I am, I will always pull some stunt and screw everything up, forcing me to retrace my steps. The data journal makes it painless. Create an annotated draft Before I hand in a story that relies even a little bit on SQL queries I go through the draft and locate every sentence that has a corresponding query. Then I copy and paste those queries just below the related sentence and run each query one more time. Sometimes I write the queries over from scratch. It's humbling to discover an error just before filing, but it's more humbling by far to have somebody else discover that error after publishing. There it is. A very special thanks to the data enthusiasts who turned my life around in Missouri: David Herzog, Jaimi Dowdell, Coulter Jones and Mark Horvit. Onward!