

Database Design and Data Visualization for Demographic Data

As an Applied Demographer and website developer I am maintaining many data tables and show many charts on my web site.

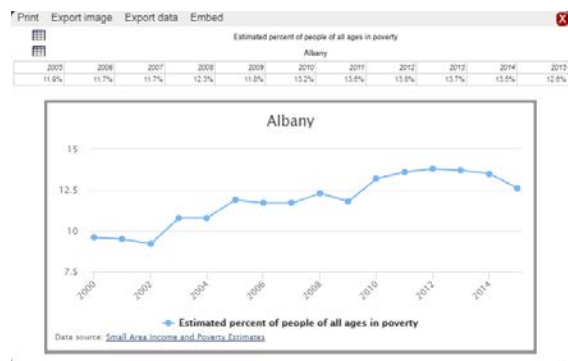
By organizing my data in a systematic way, I was able to make it a lot easier to create data driven visualizations on my web site.

I realized that almost all my data contained a field identifying location (Where), a timestamp (When) and then a value for a metric (What), specific for that place and that time.

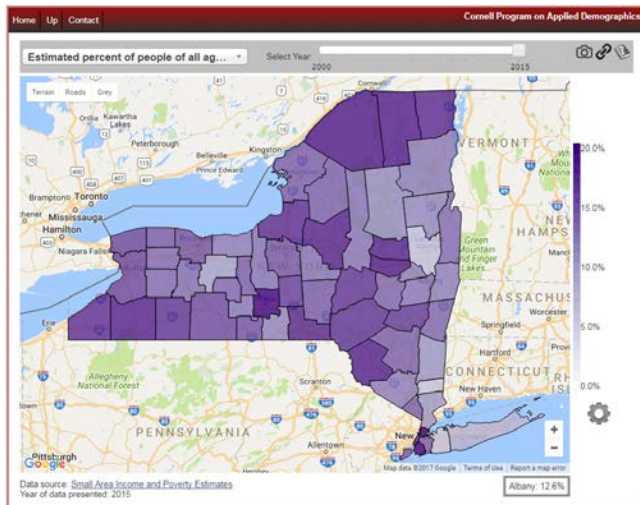
The different visualizations depend on which of the columns What/Where/When are fixed and variable.

If I have a fixed [What], a fixed [Where] and a Fixed [When], I get a value for a single metric and a single point in time and at a specific location.

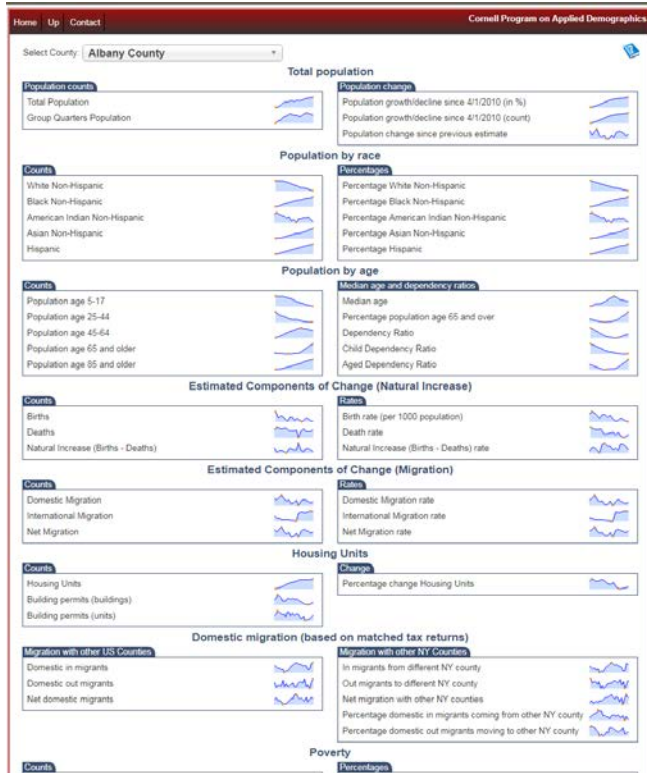
When I only start varying the [When], I can create a timeseries of this metric and display a single line chart.



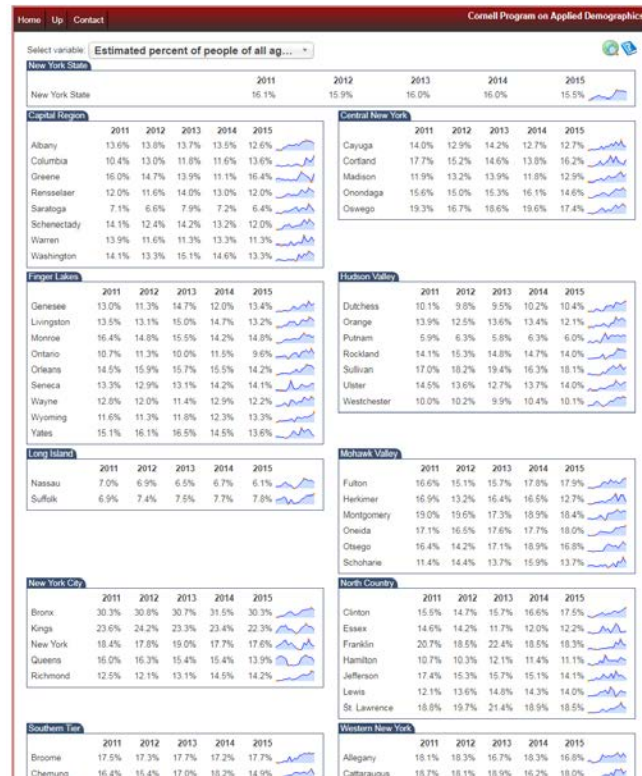
If I keep the [What] and the [When] fixed, but vary the [Where], I have the information I need to create a choropleth map.



If I only keep the [Where] fixed and vary the [What] and the [When], I can create a dashboard with a variety of timeseries, so I get a quick impression of what is going on the specified area.



Similarly, if I only keep the [What] fixed and vary the [Where] and the [When], I can create a dashboard with a variety of timeseries that enables comparing trends between areas.



Creating one big data table, would be harder to maintain because different metrics get updated at different times.

I chose to create metadata for each of the metrics. This metadata contains a shorthand identifier, a longer name that is used in the displays and an explanation that can pop-up on for example a mouse-over. Also in this metadata is the name of the table, or tableview in which this data can be found. There is also information on the formatting of this variable, e.g. does it have a %-postfix, or one digit after a decimal point.

Putting this information in the metatable eliminated the need for my webpages to contain formatting instructions, information on labels, etc.

A second metatable contains records for each of the tables. In my setup, all of my individual datatables contain data from a single source, and this metastable on the tables contains a label and a link to that source. It also contains field names for the [Where] and the [When], allowing me to use the field names used in the data source (which makes data maintenance easier) and still be able to create my queries.

In the near future I would like to create a set of API's that allows third parties to retrieve data from my databases and use that in their own applications.

For example:

Root-url/api/counties?what=totpop&when=2000-2017&where=36001

Would return data on the total population in Albany county from 2000 through 2017.

Presenter:

Jan Vink

Cornell Program on Applied Demographics