

# De-crosstabbing

Ben Klemens

16 August 2013

Max Richman<sup>1</sup> linked to a post from the World Bank on pivot tables<sup>2</sup>. My impression is that outside of the spreadsheet world these are referred to as *crosstabs*. Last time, I talked about the standard format for data, in which each row is one observation and each column is a different variable. As the WB blog notes, this is what most of the world's data analysis software expects.

I'd already shown you an example of coming into a crosstab from standard form data, in the entry on Fisher exact tests (entry #158). The example here comes out of a crosstab into the standard form. Notice how the output of the function gets written to the database, so you can subset it, join it with other data, and so on.

I also throw in a moving average, because why not. Doing it showed me that for the small slice of data here, the 2006 data is the average of the 2005 and 2007 data, implying that it might be a linear interpolation anyway.

---

<sup>1</sup><https://twitter.com/richmanmax>

<sup>2</sup><http://blogs.worldbank.org/opendata/unpivoting-data-excel-open-refine-and-python>

---

```
#include <apop.h>

int main(){
    apop_data *d = apop_text_to_data("163-xtab",.delimiters=" ",.has_row_names='y',.
        has_col_names='y');
    apop_data_print(d);

    //de-crosstab into the database. Arguments are names for the db.
    apop_crosstab_to_db(d,"wbdata","country","year","expectancy");

    //pull from the database; write to file.
    sprintf(apop_opts.db_name_column,"country");
    apop_data_print(apop_query_to_data("select * from wbdata"),.output_file="unxtabbed");

    /* oops: getting ahead of myself. Apop_matrix_row_t will be added shortly.
    printf("\nWhile we're here, a moving average for Algeria:\n");
    Apop_matrix_row_t(d->matrix,"Algeria",af);
    apop_vector_show(apop_vector_moving_average(af,2));
    */
}
```

---