Economists write a lot about a lot of things, but not nearly as much about themselves. Data is scraped from abstracts based on IDEAS at RePEc. I’ve included only the top 100 economists according to this list. Sorry Paul Romer, but someone has to be number 101.

For all code and methodology, check it out on GitHub.

Economics is a big, broad field. So what differentiates one economist from the other? Principal components has a pretty good answer, actually. It’s not too far a stretch to say that economists are split into two camps: the ones who talk about autoregression and heteroscedasticity and moments, and the ones who talk about democracy and globalization and infrastructure. Here are the leading components (i.e. the thirty words with the most positive and most negative projections on the first principal component):

<List of top and bottom 30 words on PC1>

Below is a plot of all 100 economists along the first two principal components.

<Scatterplot with all economists on PC1 and PC2>

The group that weights high on PC1 and around neutral on PC2 is probably a group that you’d associate with one another: Newey, Bollerslev, White, Watson, Hansen, Engle, and Granger. They’re the ones that write about econometrics and time series. On the other hand, those that weight low on PC2 and at least negative on PC1 are also often mentioned in the same breath: Heckman, Angrist, Kreuger, Card, List. None of this is particularly surprising – I’m just using principal components as a fancy way to compare relative word frequencies.

In a similar vein, I looked at cosines between economists’ writings to see the closest pairs. The closest pairs are below. This is mostly just a measure of most common co-authors, so nothing is too surprising here:

<Table of most common pairs, with cosine angle>

Lastly, I made a Markov model to general a sample abstract:

<Markov-model generated abstract>