Week 11: Geography as Data

Ben Schmidt

4/8/2015

These are going to be some very broad questions.

Choose **either** number one or number two, and **definitely** do number three. (I want you to practice picking up someone else's piece of software and using it.)

1. Geography

Find and map some data of interest to you on a map.

One possibility: pull out the names of states from the State of the Union addresses we've been working with, and map their changing uses. (This should be a simple `inner_join` once you have a data.frame of state names. You might find this around, or might have to create it by cutting and pasting by hand!)

A second: find a shapefile online of an urban space. (These usually come from city or state GIS websites, and are pretty easily Googleable in at least the United States). (If you can't find a shapefile, you could also read the ggmap docs for creating a base layer that looks just like Google Maps.)

Then find a text that contains street names. Write a regular expression to pull out the street names (this will look much like ones we've already used, but search for "street" or "avenue" as well), and geocode the top 100 using the function `geocode("Boylston St Boston")`. Note that you might have to change the name of the vector to include the street names you have.

A third: look at publication locations of the books in our textual metadata corpus.

2. Texts

Run a topic model on a corpus of interest, and visualize the topics across existing metadata.

3. Next Week

Find and install Matt Jockers' `syuzhet` package: use it to analyze the sentimental trajectory of two or three different texts in whatever way seems best.