

Table of contents (1 of 1)

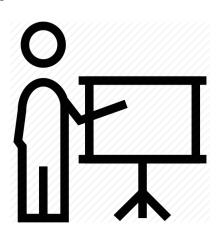
- Introduction
- 2 Data sources
 - All from the US Census Bureau
- 3 Data manipulation
- 4 Results

- Conclusion
- 6 Files

A quick overview

Questions that were/are interesting:

- What are the relative population centers
- What are the absolute population centers
- What places are changing the most



All these questions are based on time.



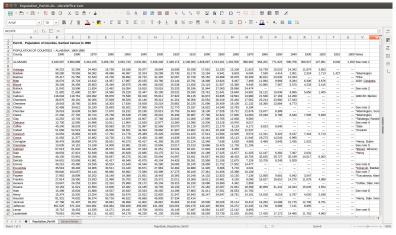
Census data pre-2000

All from the US Census Bureau

The file contains a lot of interesting, but not useful information. Things like: a Notes column to the right, when the state was not part of the Union, triple dashes when no data was available (because no one was there, or the county didn't exist), a textual section after the state that addresses the notes.

"Messy" data in the worksheet.

Same image.



"Messy" data in the worksheet.

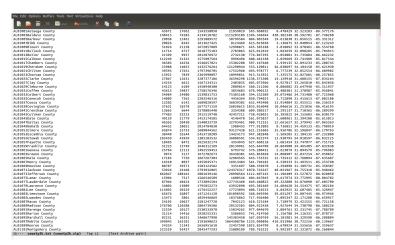
Census data 2000

The file has one line per county, and is column oriented (i.e., there are no delimiters between data fields).



Just need to pick the right columns.

Same image.



Just need to pick the right columns.

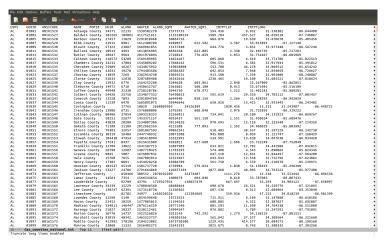
Census data 2010

The file has one line per county with tab characters between data fields.



Sometimes tabs are treated "funny" by the browser.

Same image.



Sometimes tabs are treated "funny" by the browser.



Only the simplest of manipulations

- 1 Data was "wrangled" and stored in a Postgres database
- 2 Each row in the database had:
 - County centroid (latitude and longitude)
 - Year (1790 to 2010 by 10, even for those counties and states that didn't exist)
 - Opulation (0 for those that don't exist)
- Ourrent names used, even when census was taken using old names

Performance could be improved using a data frame vice database. Performance not bad.

Static images don't do the question justice

- Click on the GIFs in the Files section
- Each image has latitude, longitude, and a Z component
- The Z component varies based on the type of image



What have we covered?

- Population and geographic data available from US Census Bureau
- Lots of long term demographic data is available
- Counties change over time
- Lots of interesting questions could be asked and answered

Next time: see where the wind blows



Files of interest

- pop.gif percentage population
 - growth by county 1790 2010
- 2 pop.gif normalized population by
- county 1790 2010 pop.gif - absolute population
- growth by county 1790 2010 viewRGL.R - an R script to view
- RGL data of the last year in the gif

- glData.dat an R data file with normalized data to be viewed with viewRGL.F
- rglData.dat an R data file with raw data to be viewed with
 - viewRGL.R
- rglData.dat an R data file with growth data to be viewed with viewRGL.F