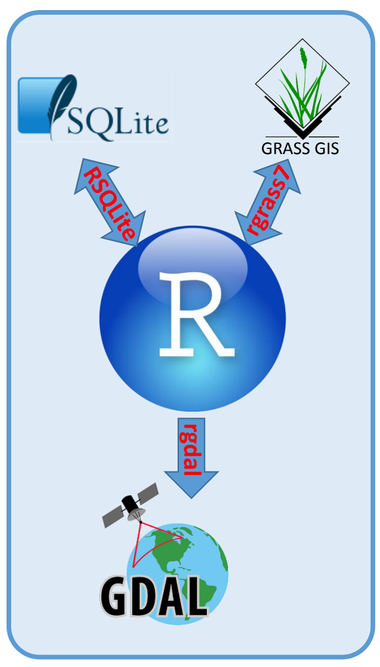
R AS A TOOL FOR GEOSPATIAL MODELING IN LARGE DATASET

Dasymetric modeling example on the continental scale

Anna Dmowska, Faculty of Geographical and Geological Sciences, Adam Mickiewicz University in Poznań

email: dmowska@amu.edu.pl

R as geospatial tool

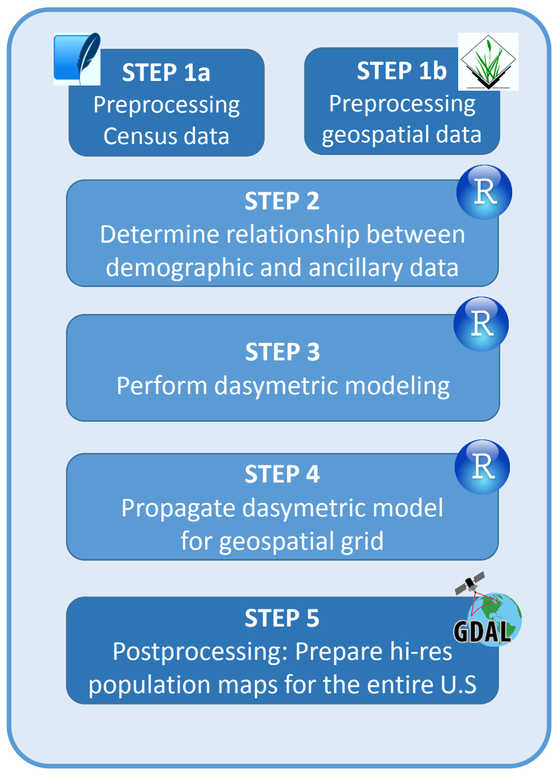


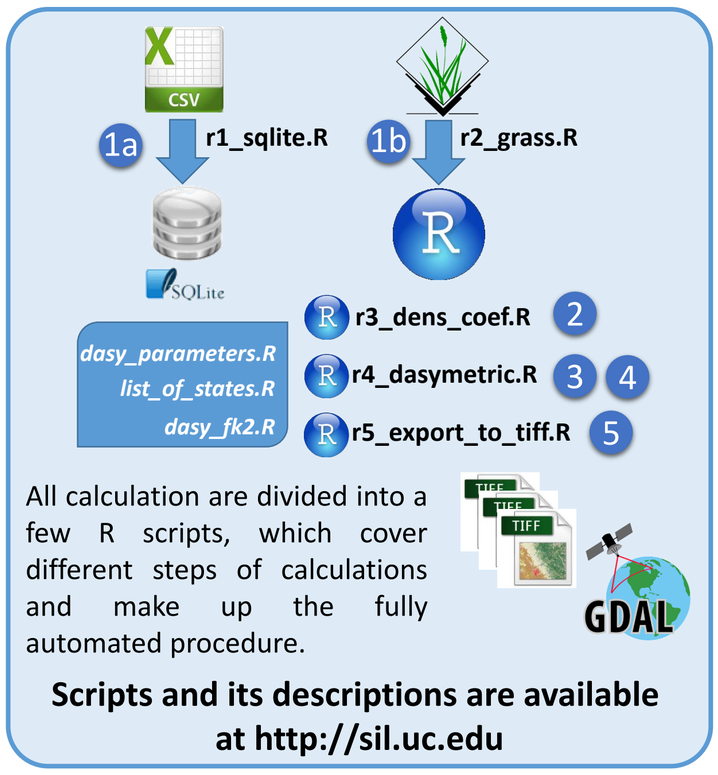
R language offers extensive amount of tools designed to work with geospatial data like ***sp*** library as well as bindings to external data source (***rgrass7, rgdal, RSQLite***).

Here we present **efficient**, **flexible** and **fully automated** computational environment which was designed to work over continental scale high resolution datasets (11 millions of records in tabular data and over 8 billions of cell grids). Algorithm was designed to **perform dasymetric modeling**.

All calculation was implemented in **R**. In addition, ***GRASS GIS 7.0, SQLite*** database and ***GDAL*** library have been used in the pre-processing and post-processing steps.

How algorithm works?





All calculation are divided into a few R scripts, which cover different steps of calculations and make up the fully automated procedure.

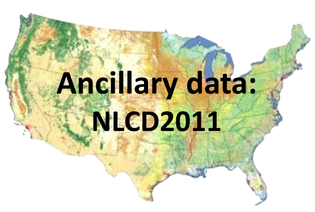
Scripts and its descriptions are available at http://sil.uc.edu

What is dasymetric modeling?

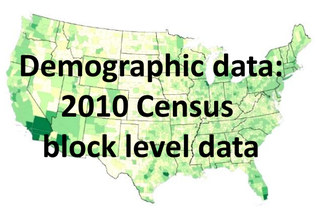
Dasymetric modeling refers to a process of **disaggregating spatial data to a finer unit** of analysis, **using additional** (or **ancillary**) **data** to help refine locations of population or other phenomena (Mennis 2003).

Dataset in numbers

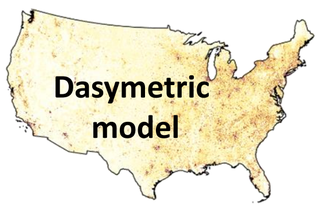
Dasymetric modeling was applied to U.S.-wide data



8 651 173 750 cells

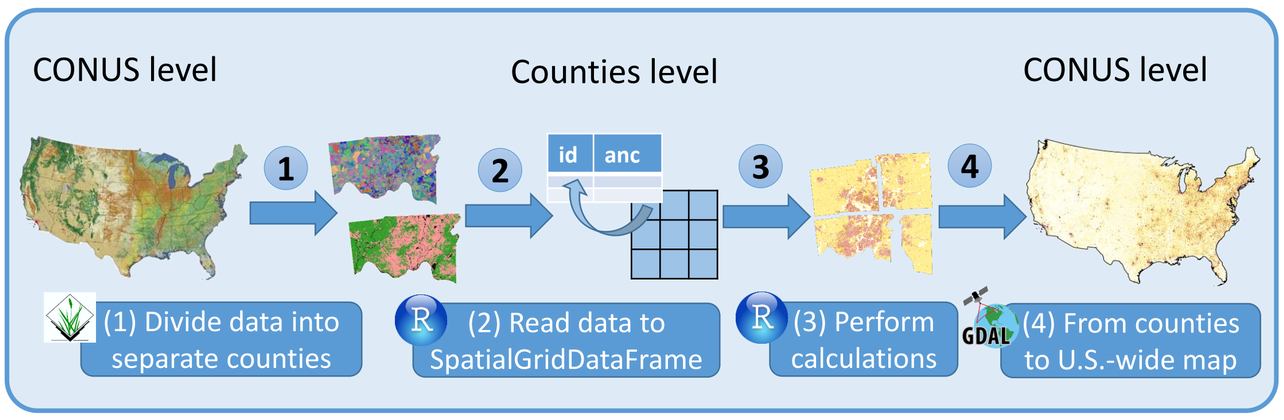


#blocks: 11,15 milions

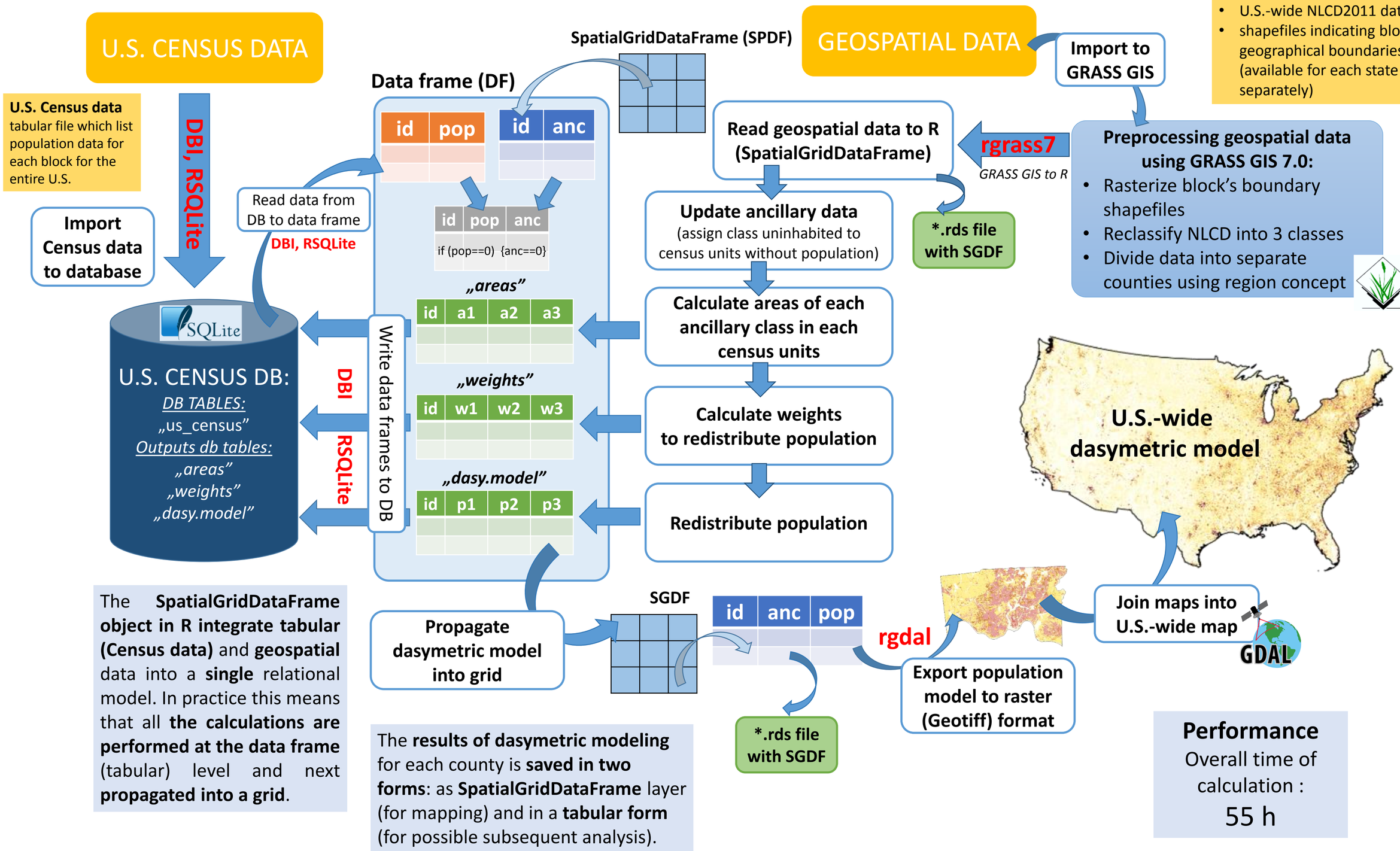


8 651 173 750 cells

Handle large dataset in R



How it works?



The **SpatialGridDataFrame object in R integrate tabular (Census data)** and **geospatial** data into a **single** relational model. In practice this means that all **the calculations are performed at the data frame** (tabular) level and next **propagated into a grid**.

The **results of dasymetric modeling** for each county is **saved in two forms**: as **SpatialGridDataFrame** layer (for mapping) and in a **tabular form** (for possible subsequent analysis).

Performance
Overall time of calculation :
55 h

The U.S.-wide dasymetric model is available online for interactive exploration and data download using GeoWeb application SocScape:

http://sil.uc.edu/webapps/socscape_usa