This position paper outlines my thoughts regarding the current state and ideal future directions of spatial data and methods used to support demographic research. My own experience suggests that general awareness of the value that spatial data and analytical tools bring to the research process is increasing. Awareness of techniques for dealing with spatial correlation in statistical analysis has lagged (pun intended) behind data availability but is still much greater than it was 10 years ago. However, in many cases both spatial data and tools are being misused. I believe that to make progress, it is incumbent on geographers to put the genie back in the bottle by ensuring that existing tools and data we promote are applied correctly. I also feel that readily available theory governing spatial demographic processes has not kept pace with the evolution of the tools and data, and that we should direct more efforts toward linking spatial demographic theory with the statistical models we use to evaluate them.

Future Directions for Data
In the last 10 years we have seen a surge in useful spatial demographic data products. An increase in the availability of geostatistical methods (kriging and its cousins) has made it possible to produce high-resolution continuous surfaces of virtually any phenomena. This is not necessarily a good thing. High-resolution gridded surfaces of demographic data are useful for mapping and display purposes but these datasets are often used as input into both stochastic and deterministic models. Geostatistical methods generally rest on the assumption that the underlying process is some function of spatial proximity and it’s not always clear that this assumption holds with demographic processes (just because they exhibit spatial correlation does not mean a spatial process is at work). Furthermore these datasets are generally compiled from a combination of survey and census data collected from different spatial supports, at different times, with varying degrees of accuracy. Despite their flaws though, integrated global datasets are useful and the demand is sufficient enough that they are probably here to stay. Our future direction then, at a minimum, should be to ensure that accompanying these datasets are error estimates that account not only for spatial uncertainty (how much error was introduced when changing spatial supports) but also temporal uncertainty (how long has it been since the source data was published). Ideally, despite their convenience, I think we should move away from publishing these datasets in grid form, and keeping them in whatever spatial support they were originally collected at (usually some set of sub-national administrative boundaries). For those that must have data below the spatial unit at which they were collected, the direction should be toward context specific small area estimation and simulation techniques rather than broad global interpolations.
Future Directions for Theory and Methods

A researcher faced with statistically analyzing a spatially correlated cross-section will usually proceed with one of four approaches:

1) Ignore the spatial component;
2) Use a cluster robust variance estimator;
3) Fit some form of spatial econometric model (toss a W matrix in there); or
4) Fit a hierarchical model.

Depending on the context one or more of these approaches is appropriate. It seems a lot of progress has been made in developing both the asymptotic theory and computational tools to implement approaches 2–4. Where we seem to lack however, is in clarifying the motivations for these approaches and connecting them to specific theoretical processes. I believe we need to do a better job of distinguishing between when spatial correlation is a nuisance to be corrected for, or a spillover process that you want to study in and of itself. In the former case it’s not clear that simply fitting a spatial error model (the most common prescription) is the best cure. The biggest weakness in this approach are the assumptions that must be made when specifying the nature of correlation via the W matrix. There are several promising areas where W becomes almost unnecessary; specifically SHAC (Spatial Heteroskedastic and Auto-Correlated) and newer cluster robust variance estimators. I believe future directions should focus on testing, implementing, and communicating these techniques.

However, arguably the more interesting elements of spatial demography and analysis are when there is some specific spatial process that we wish to capture and analyze. In these cases we are stuck with W or some other representation of a pre-specified spatial processes. Here is where the most difficult and interesting work lies ahead. To my knowledge (and I’m not a demographer) there are a paucity of concrete theoretical models defining spatial demographic processes that can be readily translated into statistical specifications. It would be an extremely valuable endeavour to compile a basic catalog of W specifications grounded in spatial demographic phenomena. Of particular use would be guidelines for specifying and estimating models where the nature of spatial correlation changes over time (dynamic panels and state space models). I am sure there is a lot existing work in network theory and peer effects that could be brought to bear on this process, but nothing yet that collects it into one compendium intended for use by the applied researcher. Our objective should be to assemble a diverse collection of spatial demographic theory with guidelines for translating that theory into testable models. This goal seems both attainable and fundamental toward making forward progress in the discipline.