Evaluating Community Analyst for Use in School Demography Studies

Richard Lycan  
*Portland State University*

Charles Rynerson  
*Portland State University*, rynerson@pdx.edu

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/prc_pub](https://pdxscholar.library.pdx.edu/prc_pub)

Part of the Urban Studies and Planning Commons

Let us know how access to this document benefits you.

Citation Details

[https://pdxscholar.library.pdx.edu/prc_pub/21](https://pdxscholar.library.pdx.edu/prc_pub/21)

This Presentation is brought to you for free and open access. It has been accepted for inclusion in Publications, Reports and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
Evaluating Community Analyst for Use in School Demography Studies

Richard Lycan and Charles Rynerson
Population Research Center
Portland State University

ESRI User Group Meeting
San Diego, July 2013
School demography studies

• The Population Research Center (PRC) at Portland State University provides demographic services, to many Oregon School Districts.
  – Main product is the enrollment forecast
  – Work paid for by the school district
  – We attempt to provide good value for the cost.

• These projects depend on data from a variety of sources including:
  – Student record data from the district
  – Data from the Census and ACS
  – GIS data for boundaries
  – School data from National Center for Educational Statistics

• Assembling the data is labor intensive.
  – Now done by research staff
  – but has educational benefits for graduate students.
Purpose of paper

• Would using ESRI’s Community Analyst improve our efforts?
  – Does CA provide most of the needed information?
  – Would we save labor costs?
  – Would the labor cost savings justify the licensing costs?
  – Are there issues with use of CA that could be resolved?

• We present a case study based on the North Clackamas School District in the Portland metro area showing how CA might be used and which identifies benefits and issues.
  – We uploaded boundary files for the District’s elementary school attendance areas and
  – for comparison a group of nearby school districts.
Two levels of analysis for our work for school districts:

- **Exploration** – In the first stage of work on a new school district we begin by exploring the data to see what this district is like and how it compares to others in the region.
  - Initially we typically look at the district as a whole
  - Later we examine the variations across the district, from school to school looking at population and housing in the attendance areas
  - Much of this work is based on viewing tables and maps to become familiar with the district.

- **Enrollment forecasting** – In the second stage we assemble the data required to support and enrollment forecasting model.
  - We require specific data at a high degree of precision and need to be assured of the quality of the data.
  - Examples are census age/sex data for attendance areas and geo-coded student record data.
  - Most of this work is done using forecasting models in MS Excel
What is Community Analyst (CA)

- CA is an ESRI product that provides easy access to data from a variety of sources: the Census; American Community Survey; and data from other sources about topics such as business, education, and health care.
- It provides tools for accessing and visualizing these data. It can, for example, summarize data for user provided geographies.
- Use of CA entails licensing costs.
- Provides tabular data, graphs, and maps
- Provides access to Census, ACS, and proprietary ESRI data
What kinds of data are available in CA?

• Data available from other sources, but facilitated through CA interface
  – Census 2010, 2000, some 1990
  – American Community Survey 2005-2009
  – Health, welfare, housing data, from various sources

• ESRI proprietary data
  – Consumer expenditure
  – Income
  – Short term forecasts of population, housing, income
  – Tapestry data

• User supplied geographies
  – Geocoded user supplied point data
  – Import of point and polygon data
  – Distance rings around points
What types of tools are in CA?

- CA has a limited set of geoprocessing tools.
  - It can make limited use of user supplied point and polygon data. We illustrate that with school district and attendance area polygons and student and school point locations.
  - It can geocode data with a valid street address.
  - It can generate tabular reports for standard areas such as census tracts but also for user supplied polygons and distance or travel time zones around points.
  - It can produce thematic maps of most of the data in its included databases.
  - The tabular data can be downloaded as PDF or Excel tables.
  - Other than as noted above it does not have tools for analyzing user supplied data, such as student locations.

- A related ESRI product, Business Analyst has a more robust set of tools.

- Data can be extracted using CA and analyzed in other software, such as ArcGIS Desktop.
Learning about a new school district

• Assume that we are about to begin our first project for the North Clackamas School District.

• Some of the things that we would want to know include:
  – Age/Sex data, current and historical
  – Births and fertility rate trends
  – Household size and type, current and historical
  – Income and poverty levels
  – Where the students live
  – Competition for students from private schools
  – How North Clackamas compares to nearby districts
Example of household type

- CA provides tables from the 2010 Census showing household and family type.
- The data for family type households was selected and generated using the “create comparison report” option.
- Loaded into Excel and a table and graph created.
- North Clackamas appears to be in the middle of the pack with respect to what percent of the households are “married couple households.”
- The equivalent data are not available as a “comparison report” for 2000.
We normally want to look at how the age-sex distribution is changing over time, particularly the school age population and their parents.

The “comparison reports” included a table that provided age-sex for five year age groups for 2010 and a forecast for 2017 which we tabulated for the Portland area school districts.

The combined table showed North Clackamas to have a growing group of aging baby boomers and a slowly growing population of pre-school age children.

It was similar to nearby Oregon City but dissimilar to Portland.

Issues:

- 1990 and 2000 data easily accessible in comparison report. 2010 data more difficult to access in standard reports.
- 2017 forecast will be dealt with later
Standard reports

- Example for 2005-2009 ACS Profile for Ardenwald elementary attendance area
- Includes MOE for user supplied geographies
- Could be done using ArcMap GIS tools, but would be labor intensive
- More appropriate for rates and percentages than for magnitudes.
- Comments later on accuracy of allocations to user supplied geographies.
Explore with maps

- Select data to map
- View maps
  - ESRI Income Block group
  - ESRI Income tract
  - ACS Income
    - Estimate
    - MOE

Issues:

Creating maps is easy but one cannot combine narrow classes, e.g. age 0-4 and 5-9.

Data from ACS and 2000 Census are handled differently. No % for ACS and no MOE for Census.
Find our where the students live

- CA provides the capability of geo-coding (address matching) data with a valid address or X/Y coordinates.
- We uploaded 3,482 KG-02 student records with a street address to CA and geo-coded them.
- Here is a map showing the points. All but three records matched.
- However, CA does not provide many tools to carry out further analysis of user supplied data, for example counting the number by attendance area.
Locate the schools

- CA provides locational data for schools, hospitals, and other types of public facilities.
- Here is a plot of school locations, with a label for Happy Valley elementary. You might find more comprehensive data on the National Center for Educational Statistics website.
- It offers us the opportunity to create drive time rings around the school and can create reports for the drive time areas.
- You can count the students in the drive time areas, but at this time CA can not.
Using CA to support enrollment forecasting

- Enrollment forecasts are important tools for school capital planning.
- One of the first steps in developing an enrollment is to develop a demographic database for the school district and the attendance areas.
  - GIS tools are used to develop this database, including geo-coding of student record and birth data and various geoprocessing tools to organize the data by school geographies.
  - This work is done with care and the process is well documented.
  - The work is time consuming and costly.
- Could CA help us do this work more efficiently?
- We begin with a simple example of organizing the age data for five year age groups for the elementary attendance areas for North Clackamas School District and then look at using the population forecasts in CA as a basis for enrollment forecasts.
Allocation of data to user polygons

- One of the most useful features of CA is the ability to summarize Census and other data to user supplied polygons.
- For use in forecasts the allocation of data to school districts and attendance areas must be as accurate as possible and the methods used must be understood.
- We used census block level data from the 2010 census for the North Clackamas School District to compare. We used the age data for 5 year age groups.
- We compared the use of a simple point in polygon (PIP) allocation where the data associated with each block centroid was associated with the school attendance area in which it was located.

<table>
<thead>
<tr>
<th>From ESAA Polygons and SF1 Block Data for 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESAA</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Ardenwald</td>
</tr>
<tr>
<td>Biquist</td>
</tr>
<tr>
<td>Concord</td>
</tr>
<tr>
<td>Duncan</td>
</tr>
<tr>
<td>Happy Valley</td>
</tr>
<tr>
<td>Lewelling</td>
</tr>
<tr>
<td>Milwaukee</td>
</tr>
<tr>
<td>Mount Scott</td>
</tr>
<tr>
<td>Oak Grove</td>
</tr>
<tr>
<td>Oregon Trail</td>
</tr>
<tr>
<td>Riverside</td>
</tr>
<tr>
<td>Scooters Mountain</td>
</tr>
<tr>
<td>Spring Mountain</td>
</tr>
<tr>
<td>Sunnyside</td>
</tr>
<tr>
<td>View Acres</td>
</tr>
<tr>
<td>Whitcomb</td>
</tr>
<tr>
<td><strong>District</strong></td>
</tr>
</tbody>
</table>
Allocation of data to user polygons

- We then created and downloaded a comparative report in CA for the same 2010 age data using the same attendance area polygons, uploaded to CA.
- Next we compared the two reports, subtracting the CA values from those we created with a point in polygon approach.
- We found that the district and some attendance areas were similar using the two methods, but that some varied greatly.
- Look at the example for Linwood and Whitcomb, where the differences mirror each other, suggesting that a block, or more than one, with a large population has been placed differently by the two approaches.

<table>
<thead>
<tr>
<th>SF1 Block minus CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESAA</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Aldenwald</td>
</tr>
<tr>
<td>Concord</td>
</tr>
<tr>
<td>Duncan</td>
</tr>
<tr>
<td>Happy Valley</td>
</tr>
<tr>
<td>Lewelling</td>
</tr>
<tr>
<td>Milwaukee</td>
</tr>
<tr>
<td>Mount Scott</td>
</tr>
<tr>
<td>Oak Grove</td>
</tr>
<tr>
<td>Oregon Trail</td>
</tr>
<tr>
<td>Riverside</td>
</tr>
<tr>
<td>Spring Mountain</td>
</tr>
<tr>
<td>Sunny side</td>
</tr>
<tr>
<td>View Acres</td>
</tr>
<tr>
<td>Whitcomb</td>
</tr>
<tr>
<td>District</td>
</tr>
</tbody>
</table>
• When we examine the map in detail we see that the Whitcomb-Linwood border splits a block with 1,497 population.
• We add detail showing building footprints and students and then counts of housing units and students.
• 614 of 890 of the housing units are located in Whitcomb thus we can allocate 69% of the population to Whitcomb and 31% to Linwood.

Both PIP and CA got it wrong. PIP put the whole block in Whitcomb, and CA put it in Linwood. It perhaps should have been split 31% / 69% between them.

But wait, there’s more. 350 of the 614 housing units in Whitcomb were senior housing. Should we count these in allocating the student level population? What about allocating the seniors?
Calculating the public school capture rate

- **Capture rate** – enrolled by grade level / age eligible
  - Example: KG-02 enrolled 3,610, from school district
  - Age 5 – 7 4,245, from census
  - Capture rate = 3,610/4,245 = 0.850 preferred method

- **Other approaches**
  - American community survey, enrolled public and private
  - Private school enrollment data, NCES or local sources

- A key variable for converting a population forecast to an enrollment forecast.
- CA can provide needed ACS and Census data
- Example for North Clackamas SD
- Most data are consistent but note major discrepancy between the Enrolled/Census and Census SF3, perhaps due to sampling error in the Census SF3 sample data, or non-enrolled. Also note large MOE for ACS

### Capture Rate for North Clackamas SD

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KG-08</td>
<td>12,023</td>
<td>14,683</td>
<td>1,109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>10,558</td>
<td>13,047</td>
<td>775</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>1,465</td>
<td>1,636</td>
<td>1,130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop Public</td>
<td>0.878</td>
<td>0.889</td>
<td></td>
<td>0.863</td>
<td>0.869</td>
</tr>
</tbody>
</table>

|         | 0.911           | 0.890                  |      | 0.826                | 0.912 |

- 09-12  | 4,980           | 6,056                  | 622  |                      |      |
| Public  | 4,335           | 5,391                  | 602  |                      |      |
| Private | 445             | 665                    | 435  |                      |      |
| Prop Public | 0.911          | 0.890                  |      | 0.826                | 0.912 |
A simple enrollment forecast: Using CA single year of age data and capture rates

- The single year of age data were downloaded from CA for the North Clackamas District boundary.

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1,397</td>
</tr>
<tr>
<td>6</td>
<td>1,541</td>
</tr>
<tr>
<td>7</td>
<td>1,469</td>
</tr>
<tr>
<td>8</td>
<td>1,486</td>
</tr>
<tr>
<td>9</td>
<td>1,561</td>
</tr>
<tr>
<td>10</td>
<td>1,524</td>
</tr>
<tr>
<td>11</td>
<td>1,622</td>
</tr>
<tr>
<td>12</td>
<td>1,591</td>
</tr>
<tr>
<td>13</td>
<td>1,610</td>
</tr>
<tr>
<td>14</td>
<td>1,472</td>
</tr>
<tr>
<td>15</td>
<td>1,491</td>
</tr>
<tr>
<td>16</td>
<td>1,539</td>
</tr>
</tbody>
</table>

- The population data were grouped into age classes that correspond to grade levels and then multiplied by the capture rate (using the enrolled/census calculation)

<table>
<thead>
<tr>
<th>Based on CA by Age</th>
<th>Rate Est</th>
<th>2017 Actual</th>
<th>CA - PRC 2012</th>
<th>PRC - 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>KG-02</td>
<td>4,407</td>
<td>0.850</td>
<td>3,748</td>
<td>3,582</td>
</tr>
<tr>
<td>03-05</td>
<td>4,533</td>
<td>0.873</td>
<td>3,956</td>
<td>3,799</td>
</tr>
<tr>
<td>06-08</td>
<td>4,737</td>
<td>0.883</td>
<td>4,181</td>
<td>3,921</td>
</tr>
<tr>
<td>09-12</td>
<td>6,112</td>
<td>0.912</td>
<td>5,573</td>
<td>5,688</td>
</tr>
<tr>
<td>KG-12</td>
<td>19,789</td>
<td>0.883</td>
<td>17,469</td>
<td>16,990</td>
</tr>
</tbody>
</table>

- The changes in enrollment forecast by the CA based forecast and one developed by PRC in 2012 are significantly different. CA forecasts much more growth and the grade level composition varies. Which should be believed? It could impact the school’s capital planning.
CA provides current population estimates and a five year forecast, here forecast for 2017.

PRC recently prepared a population forecast as part of an enrollment forecasting contract.

The results of the two forecasts are quite different with PRC forecasting nearly twice the growth in population.

The age distribution of the changes also are quite different.
Conclusions regarding exploration

• Benefits of using Community Analyst for exploring a school district
  – Provides easy access to a wide range of relevant data from the Decennial Census and the American Community Survey. Provides MOE for ACS, but not 2000 SF3 data.
  – Provides an easy method of summarizing data for user supplied geographies such as school attendance areas.
  – The mapping tools in CA allow the user to view a wide variety of maps with little investment of time or technical expertise.

• Some limitations
  – Compiling an education related profile involves extracting data from a variety of standard reports.
  – The comparison reports are limited in scope and provide little time comparison data, such as 2000 and 2010 Census data.
  – Extracting data for further analysis from the Excel and PDF versions of the standard reports is time consuming and difficult.
  – No method to combine narrow ranges, e.g. household income, for maps
  – We have some reservations about the accuracy of the data allocations for user supplied school demographies.
Conclusions regarding *forecasting*

- **Benefits of using Community Analyst for enrollment forecasting**
  - Provides access to a variety of Census and ACS data that are needed for enrollment forecasting purposes.
  - Provides access to data not easily available elsewhere such as income estimates.
  - Provides a limited set of geoprocessing tools. Geocoding appears to work well. Allocation of data to user supplied polygons works well in most cases.

- **Some limitations**
  - The 2000 – 2010 data for comparisons is limited. Hopefully when the 2008-2012 ACS data become available this will improve.
  - We question the accuracy of the allocation of data to school attendance area level geographies where data needs to be split at the block level.
  - Our one case example for North Clackamas School District suggests that the five year age forecasts of population may not be realistic at the school district or attendance area scale and are not a substitute for more comprehensive forecasting methods.
Cost – Benefit Analysis

- The cost ranges from $999/year for a single copy of a Basic version to $3,995/year for a Standard Plus version.
- For school demography applications we likely would not use many of the features of the Standard Plus Version and could make use of the Standard version at $2,495/year.
- There are discounts for multiple users and educational pricing is available.
- CA is available on the ESRI University of site license for teaching, but not for research or commercial uses.

<table>
<thead>
<tr>
<th>Licensing</th>
<th>Standard Plus</th>
<th>Standard</th>
<th>Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single User License</td>
<td>$3,995/yr</td>
<td>$2,495/yr</td>
<td>$995/yr</td>
</tr>
<tr>
<td>3 User License—Save 50% per User</td>
<td>$5,995/yr</td>
<td>$3,795/yr</td>
<td>$1,495/yr</td>
</tr>
<tr>
<td>5 User License—Save 60% per User</td>
<td>$7,995/yr</td>
<td>$4,995/yr</td>
<td>$1,995/yr</td>
</tr>
<tr>
<td>10 User License—Save 70% per User</td>
<td>$11,095/yr</td>
<td>$7,495/yr</td>
<td>$2,995/yr</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Options</th>
<th>50</th>
<th>36</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report types included with subscription</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data variables available for creating custom reports and maps</td>
<td>5,800</td>
<td>3,400</td>
<td>2,200</td>
</tr>
<tr>
<td>Access to web maps from ArcGIS.com</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Smart Map Search to find places that match multiple variables</td>
<td>✓</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Add your logo to reports and custom maps</td>
<td>✓</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Advanced facility and business search</td>
<td>✓</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Ability to export data from color-coded maps</td>
<td>✓</td>
<td>✓</td>
<td>—</td>
</tr>
<tr>
<td>Mapping of crime data</td>
<td>✓</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mapping and reporting of behavior and preference data</td>
<td>✓</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

* The purchase of a Community Analyst Subscription also gives you access to the **Community Analyst Add-in** for ArcGIS for Desktop.

Are You an Educator or Nonprofit?
Special pricing is available. Contact us to discuss your needs.
Alternatives

- The National Center for Educational Statistics provides a wide range of free data related to school and education including the School District Demographic System
- The SDDS provides tools for data download and mapping and access to special tabulation of the Census and ACS that identifies special universes such as students enrolled in public schools.
- It has profile tables that combine student and school administrative data.
The US Census Bureau also provides free access to a wide range of Census and other data through their American FactFinder (AFF) tool for searching and downloading data.


American FactFinder also has tools for thematic mapping of Census data.

Census also makes data profiles available on the Geography area of their website in shapefile and geodatabase format.


Census/TIGER indicates that they in the future will provide a tool for aggregating data for user supplied geographies.
Evaluating Community Analyst for Use in School Demography Studies
ESRI User Group Meeting
San Diego, July 2013

Richard Lycan
Professor Emeritus of Geography and Urban Studies
lycand@pdx.edu
503-880-3230

Charles Rynerson
Research Associate
rynerson@pdx.edu
503-725-5157

Center for Population Research and Census
College of Urban and Public Affairs
Portland State University
Portland, Oregon 97207-0751
http://www.pdx.edu/prc/about-prc