

---

# How to Generate a Lorenz Curve and Gini Coefficient

Center for Urban Simulation and Policy Analysis  
University of Washington

January 12, 2008

Daniel J. Evans School of Public Affairs  
University of Washington  
Seattle, Washington 98195  
USA  
Email: [info@urbansim.org](mailto:info@urbansim.org)  
Web Site: [www.urbansim.org](http://www.urbansim.org)

## Introduction

This tutorial covers creating an indicator with a Lorenz Curve selected as the visual output.

This is intended to assist new and more experienced users in taking advantage of the functionality specific to generating a Lorenz Curve and Gini Coefficient.

## Tutorial Prerequisites

If you have not already done so, install the necessary software and data for this tutorial by following the directions at [http://www.urbansim.org/wiki/external/index.php/Nightly:Install\\_Eugene\\_Tutorial](http://www.urbansim.org/wiki/external/index.php/Nightly:Install_Eugene_Tutorial). This tutorial also assumes that the user is familiar with the OPUS framework and generating indicators, or has completed the Eugene tutorial linked above.

## Formulas used in computing the Lorenz Curve and Gini Coefficient

The Lorenz curve can be represented by a function  $L(F)$ , where  $F$  is the horizontal axis, and  $L$  is the vertical axis.

For a population of size  $n$ , with a sequence of values  $y_i$   $i = 1$  to  $n$  that are indexed in non-decreasing order ( $y_i \leq y_{i+1}$ ) the Lorenz curve is the continuous piecewise linear function connecting the points  $(F_i, L_i)$ ,  $i = 0$  to  $n$ , where  $F_0 = 0, L_0 = 0$ , and for  $i = 1$  to  $n$ :

$$\begin{aligned}F_i &= i/n \\S_i &= \sum_{j=1}^i y_j \\L_i &= S_i/S_n\end{aligned}$$

The amount of inequality in two societies, or in two scenarios can be compared based on their Lorenz curves. If the curve in one case is farther away from the line of perfect equality for every value along the horizontal axis, then that case is considered to have less equality than a case with a curve nearer to the equality line.

The line of perfect equality in the Lorenz curve is  $L(F) = F$  and represents a uniform distribution, or equality. In cases where the curve does not lie on the line of perfect equality, every point on the curve represents a statement like: "the bottom 20% of households has 10% of the total income".

The Gini coefficient is defined as a ratio of the areas on the Lorenz curve diagram. It is a measure of the inequality of a distribution. If the area between the line of perfect equality and Lorenz curve is A, and the area under the Lorenz curve is B, then the Gini coefficient is  $A/(A + B)$ . Since  $A + B = 0.5$ , the Gini coefficient,  $G = A/(.5) = 2A = 1 - 2B$ . If the Lorenz curve is represented by the function  $Y = L(X)$ , the value of B can be found with integration and:

$$G = 1 - 2 \int_0^1 L(X)dX$$

The higher the Gini coefficient, the greater the inequality.

## Meaningful Inputs

The types of inputs that are commonly used to generate a Lorenz Curve and Gini Coefficient are non-aggregate variables.

The most common input for the Lorenz curve and Gini coefficient is individual or household income, because income is an important measure of well-being. Other variables could also be the basis for distributional comparisons. These should be non-categorical quantities, such as the amount of water consumed, carbon emitted, or taxes paid. At a higher level of aggregation it also makes sense to consider percentages, such as a fraction of a certain area whose residents are ethnic minorities, voted Independent, or have incomes below poverty. Finally, it may be helpful to consider that the normative implication of a Gini coefficient is that perfect equality is preferred. Certain variables which can be analyzed for their distribution, such as miles of light rail track, may not necessarily be best when equally spread but rather more valuable when grouped together in certain places.

The typical unit of analysis for a Gini coefficient is the individual. UrbanSim allows for the use of the Gini in assessments of equity across units such as gridcells and zones. Parcel level analysis will be supported in the future. It is valid to measure distributions across these units, however the user should keep the units in mind when interpreting the results.

## Computing an Indicator

UrbanSim "indicators" are useful representations of dataset attributes. How these dataset attributes are displayed is specified by the user. In this example a Lorenz Curve is generated.

1. In your command window, make sure you are in the 'eugene\tools' directory. Then type the following command to open a tool that will allow us to specify and then view an indicator:

```
python create_indicator.py
```

2. In the "Cache directory" field, enter the location of the cache directory created by the simulation, e.g. 'C:\urbansim\_cache\run\_2006\_11\_15\_15\_12'. This will be a directory in the output directory you specified when starting the simulation.
3. Leave the "Compare to another cache directory" unchecked. This option lets you compare results from two different simulations.

4. Select “Lorenz Curve” from Type drop down menu. This automatically generates the Gini Coefficient as well.
5. In the “Attribute” field, enter:

```
urbansim.gridcell.average_income
```

This entry contains the “Opus path” for a variable that shows the average income within each gridcell. More specifically, this Opus path specifies the location of a Python file defining this gridcell : it is in the *urbansim* Opus package, is defined for the *gridcell* dataset, and is located in the Python file named ‘average\_income.py’. You can find the file containing the *urbansim* package inside your workspace directory.

6. Leave the optional “Name” entry empty. If entered, it is used as the name for the created indicator (and also determines the filename of the resulting indicator). We’ll let the program use the default name (“average\_income” in this case).
7. Type in “gridcell” in the “Dataset” entry. This is consistent with the attribute we’re going to create indicator for.
8. Type in 1982 in the “Year(s)” entry. Leave the optional Scale entry as it is.
9. Press “Run Request” to generate this indicator.
10. The tool will then compute the indicators. When the “View results” button becomes active, it means the computation has finished. Press this button. A web browser will be launched and page loaded with a link to the requested indicator.
11. Click on the “1982” link to see the Lorenz curve. This simple plot was produced by matplotlib, a Python graphing and mapping package.  
See <http://www.urbansim.org/wiki/external/index.php/Indicators> for a partial list of indicators that are known to work on the Eugene-Springfield data. Interpreting the Lorenz Curve and Gini Coefficient is defined for non-aggregate variables.
12. Press “Close” to exit this tool.

## Closing Comments

At this point, you have generated a Lorenz Curve and Gini Coefficient produced by an indicator from the simulated results. For more information, please see the accompanying Reference Manual and User Guide.

If you have any particular requests for documentation, let us know by emailing the [users@urbansim.org](mailto:users@urbansim.org) email distribution list.

## Acknowledgments

This tutorial was put together by students from the 2007 UrbanSim Capstone class of the Computer Science and Engineering department at the University of Washington and from the Daniel J. Evans School of Public Affairs to accompany the integration of Lorenz Curve output into the OPUS core.