Application of Python Scripting for Major Work Destination Analysis

Yunkyung Choi, Jung Seo, and Tom Vo

Research & Analysis
Southern California Association of Governments

2014 ESRI User Conference | July 14-18, 2014 | San Diego Convention Center
Contents

- Introduction of SCAG
- Objectives
- Methodology
- Results
- Conclusions
Overview of SCAG
Quick Facts of SCAG

- Nation’s largest Metropolitan Planning Organization (MPO)
- 6 counties and 191 cities
- 15 sub-regions
- 18.4 million people (2012)
- 38,000 square miles
- 16th largest economy in the world (GRP: $924 Billion in 2012)
Objectives

- Identify the major work destinations for each jurisdiction in the SCAG region
  - 191 cities and 6 counties

- Visualize the spatial patterns of the major work destinations for each jurisdiction
  - To understand where residents of each jurisdiction are employed

- Provide informational data resources to local jurisdictions for planning purposes
  - E.g. 2013 Local Profiles: Planning data reports
Methodology for identifying work destination

- Using *OnTheMap* application

OnTheMap application

- Web-based mapping and reporting application
- Based on 2002-2011 LODES data
Programming scripts with *LODES* raw datasets

- State-based annual data from 2002-2011
- 3 types of CSV data files
  - Origin-Destination (OD)
  - Residence Area Characteristics (RAC)
  - Workplace Area Characteristics (WAC)
- Enumerated by 2010 census blocks
# Pros and Cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>OnTheMap</th>
<th>Programming Scripts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td>Easy-to-use interface and able to visualize data instantly</td>
<td>Able to manipulate data for multiple areas at once</td>
</tr>
<tr>
<td></td>
<td>Support analyses at multiple levels of geography</td>
<td>Improve efficiency of managing and processing big data</td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td>Not able to search/download for multiple locations</td>
<td>Incorporate with ArcGIS to visualize the data</td>
</tr>
</tbody>
</table>
LODES raw data manipulation using SAS

**LODES raw dataset description**

- **LODES CA OD dataset (Version 7, 2011)**
- **3 Variables:** w_geocode, h_geocode, S000

<table>
<thead>
<tr>
<th>Pos</th>
<th>Variable</th>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>w_geocode</td>
<td>Char15</td>
<td>Workplace Census Block Code</td>
</tr>
<tr>
<td>2</td>
<td>h_geocode</td>
<td>Char15</td>
<td>Residence Census Block Code</td>
</tr>
<tr>
<td>3</td>
<td>S000</td>
<td>Num</td>
<td>Total number of jobs</td>
</tr>
<tr>
<td>4</td>
<td>SA01</td>
<td>Num</td>
<td>Number of jobs of workers age 29 or younger</td>
</tr>
<tr>
<td>5</td>
<td>SA02</td>
<td>Num</td>
<td>Number of jobs of workers age 30 to 54</td>
</tr>
<tr>
<td>6</td>
<td>SA03</td>
<td>Num</td>
<td>Number of jobs of workers age 55 or older</td>
</tr>
<tr>
<td>7</td>
<td>SE01</td>
<td>Num</td>
<td>Number of jobs with earnings $1250/month or less</td>
</tr>
<tr>
<td>8</td>
<td>SE02</td>
<td>Num</td>
<td>Number of jobs with earnings $1251/month to $3333/month</td>
</tr>
<tr>
<td>9</td>
<td>SE03</td>
<td>Num</td>
<td>Number of jobs with earnings greater than $3333/month</td>
</tr>
<tr>
<td>10</td>
<td>SI01</td>
<td>Num</td>
<td>Number of jobs in Goods Producing industry sectors</td>
</tr>
<tr>
<td>11</td>
<td>SI02</td>
<td>Num</td>
<td>Number of jobs in Trade, Transportation, and Utilities industry sectors</td>
</tr>
<tr>
<td>12</td>
<td>SI03</td>
<td>Num</td>
<td>Number of jobs in All Other Services industry sectors</td>
</tr>
<tr>
<td>13</td>
<td>createdate</td>
<td>Char</td>
<td>Date on which data was created, formatted as YYYYMMDD</td>
</tr>
</tbody>
</table>
**LODES** raw data manipulation using SAS (cont.)

**O-D Analysis Processing Steps**

1. **Correspondence between Census Block codes and Place codes**
2. **Merging O-D dataset by Census Block codes**
3. **Calculating job number for residents of each O-D at place level**
   - Ranking top 10 work destination
   - Sorting dataset in descending order
4. **Exporting CSV dataset to dbf file for GIS**
LODES raw data manipulation using SAS (cont.)
Visualizing Major Work Destinations

- Using *ESRI ArcGIS* application
  - *Data Driven Pages* – To create a multi-page map series from a single map document

- Using *Python* programming language
  - *Python* – Interpreted, object-oriented, high-level general-purpose programming language
Arcpy Module

Python site package for performing GIS functions available in ArcGIS

- Encompassing the arcgisscripting module
- E.g. arcpy.mapping (mapping module)

Useful and productive way to perform geographic data analysis, data conversion, data management, and map automation with Python
Using ArcGIS and Python to create a series of work destination maps for 191 cities and 6 counties.
Python Script to Visualize Top 10 Work Destinations

- Importing ArcPy and ArcPy.mapping modules in Python to automatically create work destination maps
  - Data Driven Page, Definition Query, and Exporting functions in ArcGIS
    - mxd.dataDrivenPages.currentPageID
    - lyr.definitionQuery = sqlExp
    - arcpy.mapping.ExportToJPEG

```python
# Import ArcPy and os modules
import arcpy, os
import arcpy.mapping

# Set the current workspace
arcpy.env.workspace = "P:/LEHD/esri_uc_2014/shapes/local_profiles_2013.gdb"

# Set the current map document
mxd = arcpy.mapping.MapDocument("P:/LEHD/esri_uc_2014/mxs/appendix/appendix_D.mxd")

# Define data driven page variables
totalPage = mxd.dataDrivenPages.pageCount
curPage = mxd.dataDrivenPages.currentPageID
```
Identified top 10 work destinations where residents of each jurisdictions commute to work (191 cities and 6 counties)
Visualization of Top 10 Work Destinations

- Created a series of maps that depict top 10 work destinations for 191 cities and 6 counties, based on automated workflow using ArcGIS and Python.
Incorporation of Python and MS Excel

Utilize modules in Python to create Top 10 Destination tables for each jurisdiction

- E.g. xlrd and xlwt module
  
  - citynames = sheet.cell_value(i,2)
  
  - s.write(0,0,citynames,header_style1)
Conclusions

- Significantly efficient in processing O-D analysis with SAS for numerous jurisdictions

- Some inconsistencies issue in the total number of job at place level between SAS programming and OnTheMap application

- Data limitation of LODES datasets

- Incorporation of Python and ArcGIS improves processing time and accuracy of map production
Future Studies

- Work destinations at different geographical level
  - City to City † Sub-region to Sub-region
  - Top 10 work destinations † Top 20 work destinations

- Produce major work destination data using Census Transportation Planning Products (CTPP)

- Desktop application or ArcGIS toolbox to generate maps based on a selection of jurisdictions
Thank you!

Yunkyung Choi (ChoiY@scag.ca.gov)
Tom M. Vo (Vo@scag.ca.gov)