US Census Spatial and Demographic Data in R: The UScensus2000-suite

Zack W Almquist

Department of Sociology
University of California, Irvine
email: almquist@uci.edu

useR! 2010
July 22nd 2010

---

1This work was supported in part by an ONR award #N00014-08-1-1015 and a National Science Foundation (NSF) award BCS-0827027.
Overview

Why R for Spatial Analysis

Preliminaries
   The sp and maptools Packages

The UScensus2000-suite of Packages

Examples

Future Directions

References
Why R for Spatial Analysis

R now has a number of contributed packages

| Classes for spatial data: | sp, maptools, rgdal (Bivand et al., 2008) |
| Access to spatial data:   | spsurvey, rwoldmap, maps, UScensus       |
| R/W spatial data:         | rgdal, maptools, RgoogleMaps             |
| Spatial statistics:       | PBSmapping, spatial, spatstat, spdep, spgwr, splancs |
| For more information see: | CRAN Task View: Analysis of Spatial Data |
Why R for Spatial Analysis

R now has a number of contributed packages

- **Classes for spatial data**: sp, maptools, rgdal (Bivand et al., 2008)
- **Access to spatial data**: spsurvey, rwoldmap, maps, UScensus
- **R/W spatial data**: rgdal, maptools, RgoogleMaps
- **Spatial statistics**: PBSmapping, spatial, spatstat, spdep, spgwr, splancs
- For more information see: CRAN Task View: Analysis of Spatial Data
## Why R for Spatial Analysis

R now has a number of contributed packages

<table>
<thead>
<tr>
<th>Classes for spatial data:</th>
<th>sp, maptools, rgdal (Bivand et al., 2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to spatial data:</td>
<td>spsurvey, rwoldmap, maps, UScensus</td>
</tr>
<tr>
<td>R/W spatial data:</td>
<td>rgdal, maptools, RgoogleMaps</td>
</tr>
<tr>
<td>Spatial statistics:</td>
<td>PBSmapping, spatial, spatstat, spdep, spgwr, splancs</td>
</tr>
<tr>
<td>For more information see:</td>
<td>CRAN Task View: Analysis of Spatial Data</td>
</tr>
</tbody>
</table>
Why R for Spatial Analysis

R now has a number of contributed packages

- **Classes for spatial data**: sp, maptools, rgdal (Bivand et al., 2008)
- **Access to spatial data**: spsurvey, rwoldmap, maps, UScensus
- **R/W spatial data**: rgdal, maptools, RgoogleMaps
- **Spatial statistics**: PBSmapping, spatial, spatstat, spdep, spgwr, splancs
- For more information see: CRAN Task View: Analysis of Spatial Data
Why R for Spatial Analysis

R now has a number of contributed packages

- **Classes for spatial data**: `sp`, `maptools`, `rgdal` (Bivand et al., 2008)
- **Access to spatial data**: `spsurvey`, `rwoldmap`, `maps`, `UScensus`
- **R/W spatial data**: `rgdal`, `maptools`, `RgoogleMaps`
- **Spatial statistics**: `PBSmapping`, `spatial`, `spatstat`, `spdep`, `spgwr`, `splancs`

- For more information see: [CRAN Task View: Analysis of Spatial Data](https://CRAN.R-project.org/web/views/Spatial.html)
The sp and maptools Packages

- Bivand et al.'s book *Applied Spatial Data Analysis with R*
- Contain tools for handling many (most?) of the different spatial data formats
- Contain tools for managing standard GIS activities such as plotting and overlays
- Inter-operate with a number of packages for statistical spatial analysis
UScensus2000-suite of packages

- 6 packages
  - UScensus2000
  - UScensus2000add
  - UScensus2000cdp
  - UScensus2000tract
  - UScensus2000blkgrp
  - UScensus2000blk
- 2 packages of helper functions
- 4 packages of polygon/shapefiles and demographic data
- All data from US Census Bureau’s SF1 files and TigerLine Shapefiles
UScensus2000-suite of packages

- 6 packages
  - UScensus2000
  - UScensus2000add
  - UScensus2000cdp
  - UScensus2000tract
  - UScensus2000blkgrp
  - UScensus2000blk

- 2 packages of helper functions

- 4 packages of polygon/shapefiles and demographic data

- All data from US Census Bureau’s SF1 files and TigerLine Shapefiles
UScensus2000-suite of packages

- 6 packages
  - UScensus2000
  - UScensus2000add
  - UScensus2000cdp
  - UScensus2000tract
  - UScensus2000blkgrp
  - UScensus2000blk
- 2 packages of helper functions
- 4 packages of polygon/shapefiles and demographic data
- All data from US Census Bureau’s SF1 files and TigerLine Shapefiles
UScensus2000-suite of packages

- 6 packages
  - UScensus2000
  - UScensus2000add
  - UScensus2000cdp
  - UScensus2000tract
  - UScensus2000blkgrp
  - UScensus2000blk

- 2 packages of helper functions

- 4 packages of polygon/shapefiles and demographic data

- All data from US Census Bureau’s SF1 files and TigerLine Shapefiles
Structure of the UScensus2000 Packages
Organization of the US Census

Diagram:

- County
  - Tract
    - Block Group
      - Block
Organization of the US Census
Available Data

Via The Comprehensive R Archive Network (CRAN)
http://cran.r-project.org/

- Block Group (UScensus2000blkgrp)
- Tract (UScensus2000tract)
- Census Designated Place (UScensus2000cdp)
- Helper functions (UScensus2000 and UScensus2000add)

Via NCASD Lab
http://www.ncasd.org/census2000/

- Block (UScensus2000blk)
Installing and Loading Packages

> install.packages("UScensus2000", + dependencies=T)
> install.packages("UScensus2000add" + dependencies=T)
> library(UScensus2000)
> install.blk("osx")
The Data!
Structure of the UScensus2000 Data-Packages

- Package (e.g., UScensus2000tract)
- State (e.g., california.tract)
- data and polygons (e.g., california.tract@data or california.tract@polygons)

▶ All data is stored as `SpatialPolygonsDataFrame` object
▶ data is a data.frame object with ID (factors) and demographic (numeric) values
▶ polygons is a list of the spatial data
Examples!

- Slide 1: Command
  
  >

- Slide 2: Output
Loading the Data

> library(UScensus2000)
> data(california.tract)
> summary(as(california.tract,"SpatialPolygons"))

Object of class SpatialPolygons
Coordinates:

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>-124.40959</td>
<td>-114.13443</td>
</tr>
<tr>
<td>r2</td>
<td>32.53416</td>
<td>42.00952</td>
</tr>
</tbody>
</table>
Is projected: FALSE
proj4string:

[+proj=longlat +datum=NAD83 +ellps=GRS80 +towgs84=0,0,0]

> names(california.tract)
Loading the Data

Load/display/etc

[21] "h.white"[22] "h.black"[23] "h.american.es"[24] "h.asian"
[25] "h.hawn.pi"[26] "h.other"  [27] "males"    [28] "females"
[29] "age.under5"[30] "age.5.17"[31] "age.18.21"[32] "age.22.29"
[33] "age.30.39"[34] "age.40.49"[35] "age.50.64"[36] "age.65.up"
[37] "med.age"  [38] "hsehld.1.m" [39] "hsehld.1.f"[40] "fhh.child"
[49] "hh.urban"[50] "hh.4person"[51] "hh.5person"[52] "hh.6person"
[53] "hh.owner"[54] "hh.7person"[55] "hh.7person"[56] "hh.7person"
[57] "hh.3person"[58] "hh.4person"[59] "hh.4person"[60] "hh.4person"
[61] "hh.owner"[62] "hh.owners"[63] "hh.owners"[64] "hh.owners"
[65] "hh.nh.white.1p"[66] "hh.nh.white.2p"[67] "hh.nh.white.3p"[68] "hh.nh.white.4p"
[69] "hh.nh.white.5p"[70] "hh.nh.white.6p"[71] "hh.nh.white.7p"[72] "hh.nh.white.8p"
[73] "hh.nh.black.1p"[74] "hh.nh.black.2p"[75] "hh.nh.black.3p"[76] "hh.nh.black.4p"
[77] "hh.nh.black.5p"[78] "hh.nh.black.6p"[79] "hh.nh.black.7p"[80] "hh.nh.black.8p"
[81] "hh.nh.asian.1p"[82] "hh.nh.asian.2p"[83] "hh.nh.asian.3p"[84] "hh.nh.asian.4p"
[85] "hh.nh.asian.5p"[86] "hh.nh.asian.6p"[87] "hh.nh.asian.7p"[88] "hh.nh.asian.8p"
Help!

```r
help()

> help(california.tract)
```
Help!

help()

california.tract {UScensus2000tract}

california.tract

Description

california.tract is a SpatialPolygonsDataFrame with polygons made from the 2000 US Census tiger/line boundary files (http://www.census.gov/geo/www/tiger/) for Census Tracts. It also contains 86 variables from the Summary File 1 (SF 1) which contains the 100-percent data (http://www.census.gov/prod/2000doc/sf1.pdf).

All polygons are projected in CRS("+proj=longlat +datum=NAD83")

Usage
data(california.tract)

Details

ID Variables

data field name Full Description
state State FIPS code
county County FIPS code
tract Tract FIPS code

Census Variables

<table>
<thead>
<tr>
<th>Census SF1 Field Name</th>
<th>data field name</th>
<th>Full Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P007001</td>
<td>pop2000</td>
<td>population 2000</td>
</tr>
<tr>
<td>P007002</td>
<td>white</td>
<td>white alone</td>
</tr>
<tr>
<td>P007003</td>
<td>black</td>
<td>black or african american alone</td>
</tr>
<tr>
<td>P007004</td>
<td>amer.ind</td>
<td>american indian and alaska native alone</td>
</tr>
<tr>
<td>P007005</td>
<td>asian</td>
<td>asian alone</td>
</tr>
<tr>
<td>P007006</td>
<td>hawan.pi</td>
<td>native hawaiian and other pacific islander alone</td>
</tr>
<tr>
<td>P007007</td>
<td>other</td>
<td>some other race alone</td>
</tr>
</tbody>
</table>
Useful Functions in the UScensus2000 Package
# UScensus2000

## Functions

- `choropleth()`
- `county()`
- `MSA()`
- `city()`
- `poly.clipper()`
- `demographics()`
choropleth()
choropleth()

2000 US Census Tracts
California

Population Count
(0,3399]
(3399,4546]
(4546,5932]
(5932,36146]

Quantiles (equal frequency)
UScensus2000

county() – Output: SpatialPolygonsDataframe

> la.county <- county(name="los angeles", + state="ca", level="tract")
> plot(la.county)
UScensus2000

MSA() – Output: SpatialPolygonsDataFrame

> losangeles.msa<-MSA(msaname="Los Angeles", + state="CA", level="tract")
> plot(losangeles.msa)
UScensus2000

MSA()
city() – Output: SpatialPolygonsDataFrame

```r
> losangeles <- city(name = "los angeles", 
+ state = "ca")
> plot(losangeles)
```
UScensus2000
city()
poly.clipper() – Output: SpatialPolygonsDataframe

> losangeles.tract <- poly.clipper(  
  + name="Los Angeles", state="ca", level="tract")  
> plot(losangeles.tract)
poly.clipper()
uscensus2000

demographics() – Output: matrix

> laMSAarea <- demographics(
+   dem=c("pop2000","white","black"),
+   "CA", level="msa", msaname="Los Angeles")
> laMSAarea
<table>
<thead>
<tr>
<th>County</th>
<th>pop2000</th>
<th>white</th>
<th>black</th>
</tr>
</thead>
<tbody>
<tr>
<td>san bernardino county</td>
<td>1709434</td>
<td>1006960</td>
<td>155348</td>
</tr>
<tr>
<td>ventura county</td>
<td>753197</td>
<td>526721</td>
<td>14664</td>
</tr>
<tr>
<td>los angeles county</td>
<td>9519338</td>
<td>4637062</td>
<td>930957</td>
</tr>
<tr>
<td>riverside county</td>
<td>1545387</td>
<td>1013478</td>
<td>96421</td>
</tr>
<tr>
<td>orange county</td>
<td>2846289</td>
<td>1844652</td>
<td>47649</td>
</tr>
</tbody>
</table>
UScensus2000

demographics() – Output: matrix

> ca.cdp<-demographics(
+  dem=c("pop2000","white","black",
+  "hh.units","hh.vacant"),
+  "CA",level="cdp")
> ##Alphabetic order the first 10 CDPs
> ca.cdp[order(rownames(ca.cdp))[1:10],]
### demographies() – Output: matrix

<table>
<thead>
<tr>
<th>City</th>
<th>pop2000</th>
<th>white</th>
<th>black</th>
<th>hh.units</th>
<th>hh.vacant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acton</td>
<td>2390</td>
<td>2130</td>
<td>17</td>
<td>873</td>
<td>76</td>
</tr>
<tr>
<td>Adelanto</td>
<td>18130</td>
<td>9147</td>
<td>2377</td>
<td>5547</td>
<td>833</td>
</tr>
<tr>
<td>Agoura Hills</td>
<td>20537</td>
<td>17858</td>
<td>272</td>
<td>6993</td>
<td>119</td>
</tr>
<tr>
<td>Alameda</td>
<td>72259</td>
<td>41148</td>
<td>4488</td>
<td>31644</td>
<td>1418</td>
</tr>
<tr>
<td>Alamo</td>
<td>15626</td>
<td>14119</td>
<td>74</td>
<td>5497</td>
<td>91</td>
</tr>
<tr>
<td>Albany</td>
<td>16444</td>
<td>10078</td>
<td>675</td>
<td>7248</td>
<td>237</td>
</tr>
<tr>
<td>Alhambra</td>
<td>85804</td>
<td>25758</td>
<td>1437</td>
<td>30069</td>
<td>958</td>
</tr>
<tr>
<td>Aliso Viejo</td>
<td>40166</td>
<td>31395</td>
<td>828</td>
<td>16608</td>
<td>461</td>
</tr>
<tr>
<td>Almanor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>Alondra Park</td>
<td>8622</td>
<td>3584</td>
<td>1088</td>
<td>2933</td>
<td>103</td>
</tr>
</tbody>
</table>
What if we want other SF1 demographics

For example:

1. College dormitories (PCT016033)
2. Military quarters (PCT016034)
3. Population of two or more races (P005010)
demographics() – Output: SpatialPolygonsDataFrame

> library(UScensus2000add)
> rhode_island<-demographics.add(dem=
+  c("PCT016033","PCT016034","P005010")
+ ,state="ri",level="tract")

WARNING requires internet access – depending on state and a few other things – and may require downloading very large files!
Future Directions

- Add access to SF3 data (economic data)
- Expand to other countries (Europe, South America, etc)
Thanks!