US Census Spatial and Demographic Data in R: The UScensus2000-suite[1]

Zack W Almquist

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

useR! 2010
July 22nd 2010

[1] This 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

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

- 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

- 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

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

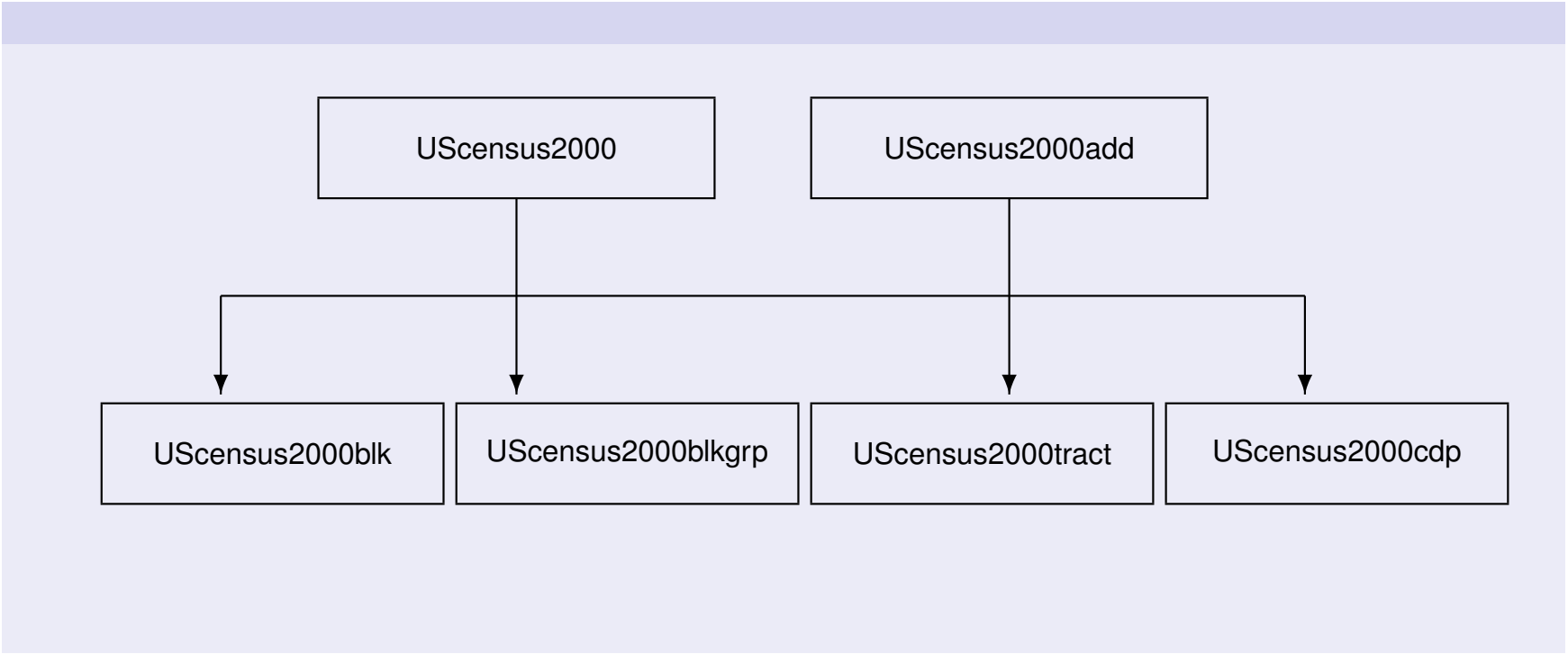
UScensus2000-suite of packages

- 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

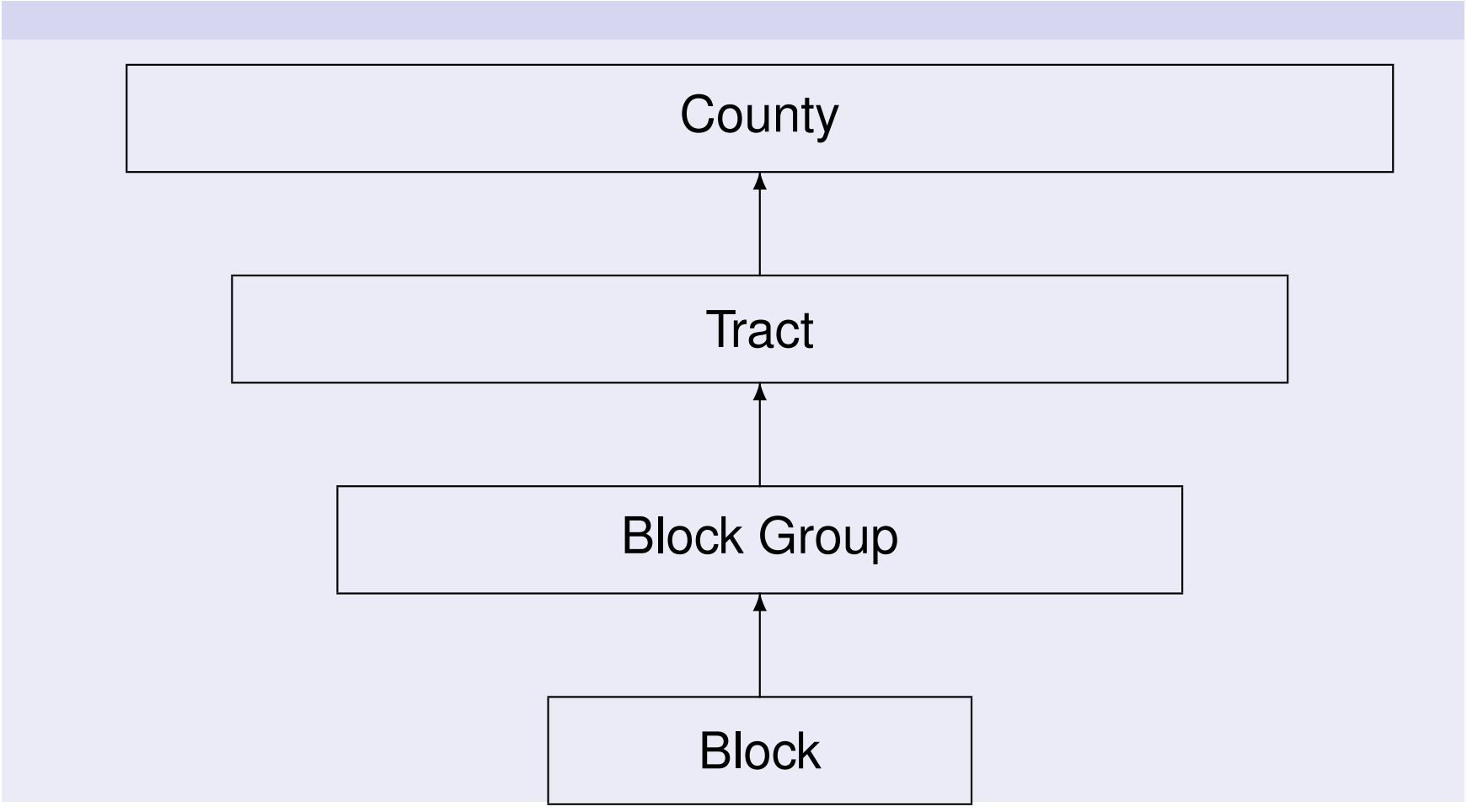
UScensus2000-suite of packages

- 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

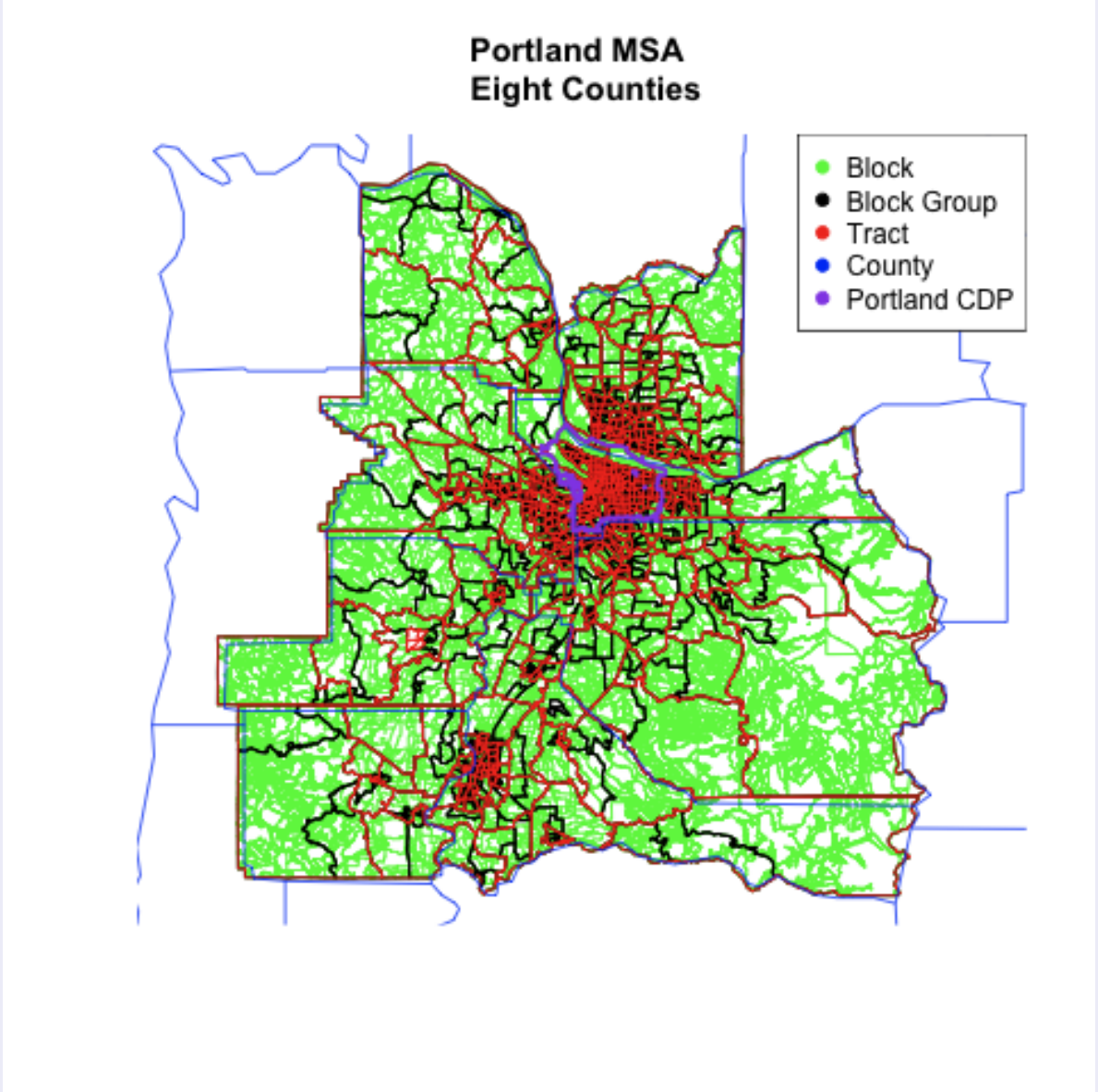
Structure of the UScensus2000 Packages



Organization of the US Census



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/

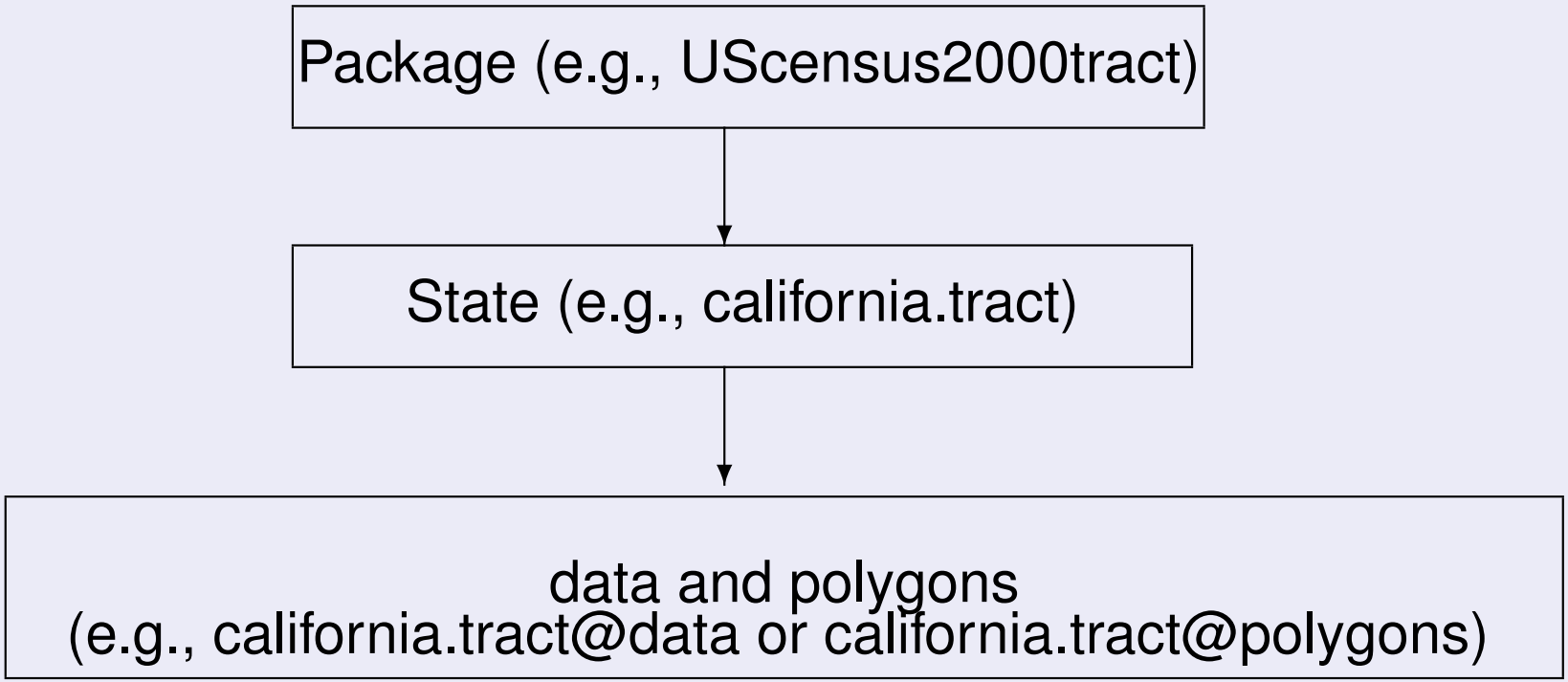
- Block (UScensus2000blk)

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



- 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

Load/display/etc

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

Object of class SpatialPolygons
Coordinates:
          min         max
r1 -124.40959 -114.13443
r2   32.53416   42.00952
Is projected: FALSE
proj4string :
[+proj=longlat +datum=NAD83 +ellps=GRS80 +towgs84=0,0,0]
>

> names(california.tract)
```

Loading the Data

Load/display/etc

```
 [1] "state"          "county"         "tract"          "pop2000"
 [5] "white"          "black"          "ameri.es"       "asian"
 [9] "hawn.pi"        "other"          "mult.race"      "hispanic"
[13] "not.hispanic.t" "nh.white"       "nh.black"       "nh.ameri.es"
[17] "nh.asian"       "nh.hawn.pi"     "nh.other"       "hispanic.t"
[21] "h.white"        "h.black"        "h.american.es"  "h.asian"
[25] "h.hawn.pi"      "h.other"        "males"          "females"
[29] "age.under5"     "age.5.17"       "age.18.21"      "age.22.29"
[33] "age.30.39"      "age.40.49"      "age.50.64"      "age.65.up"
[37] "med.age"        "med.age.m"      "med.age.f"      "households"
[41] "ave.hh.sz"      "hsehld.1.m"     "hsehld.1.f"     "marhh.chd"
[45] "marhh.no.c"     "mhh.child"      "fhh.child"      "hh.units"
[49] "hh.urban"       "hh.rural"       "hh.occupied"    "hh.vacant"
[53] "hh.owner"       "hh.renter"      "hh.1person"     "hh.2person"
[57] "hh.3person"     "hh.4person"     "hh.5person"     "hh.6person"
[61] "hh.7person"     "hh.nh.white.1p" "hh.nh.white.2p" "hh.nh.white.3p"
[65] "hh.nh.white.4p" "hh.nh.white.5p" "hh.nh.white.6p" "hh.nh.white.7p"
[69] "hh.hisp.1p"     "hh.hisp.2p"     "hh.hisp.3p"     "hh.hisp.4p"
[73] "hh.hisp.5p"     "hh.hisp.6p"     "hh.hisp.7p"     "hh.black.1p"
[77] "hh.black.2p"    "hh.black.3p"    "hh.black.4p"    "hh.black.5p"
[81] "hh.black.6p"    "hh.black.7p"    "hh.asian.1p"    "hh.asian.2p"
[85] "hh.asian.3p"    "hh.asian.4p"    "hh.asian.5p"    "hh.asian.6p"
[89] "hh.asian.7p"
```

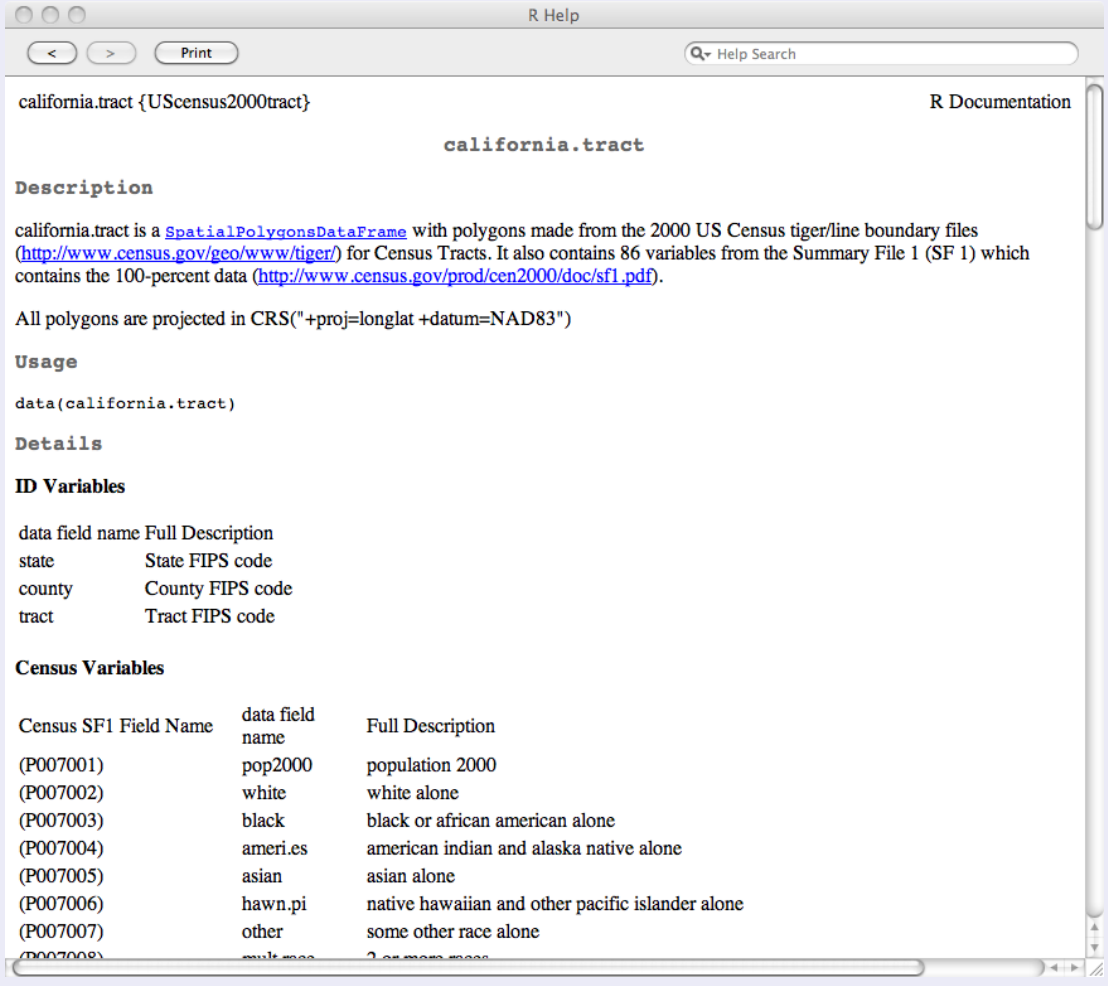
Help!

help()

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

Help!

help()


R Help
Print
Help Search
california.tract {UScensus2000tract}
R Documentation
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/cen2000/doc/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
Census SF1 Field Name data field name Full Description
(P007001) pop2000 population 2000
(P007002) white white alone
(P007003) black black or african american alone
(P007004) ameri.es american indian and alaska native alone
(P007005) asian asian alone
(P007006) hawn.pi native hawaiian and other pacific islander alone
(P007007) other some other race alone

Useful Functions in the UScensus2000 Package

UScensus2000

Functions

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

choropleth()

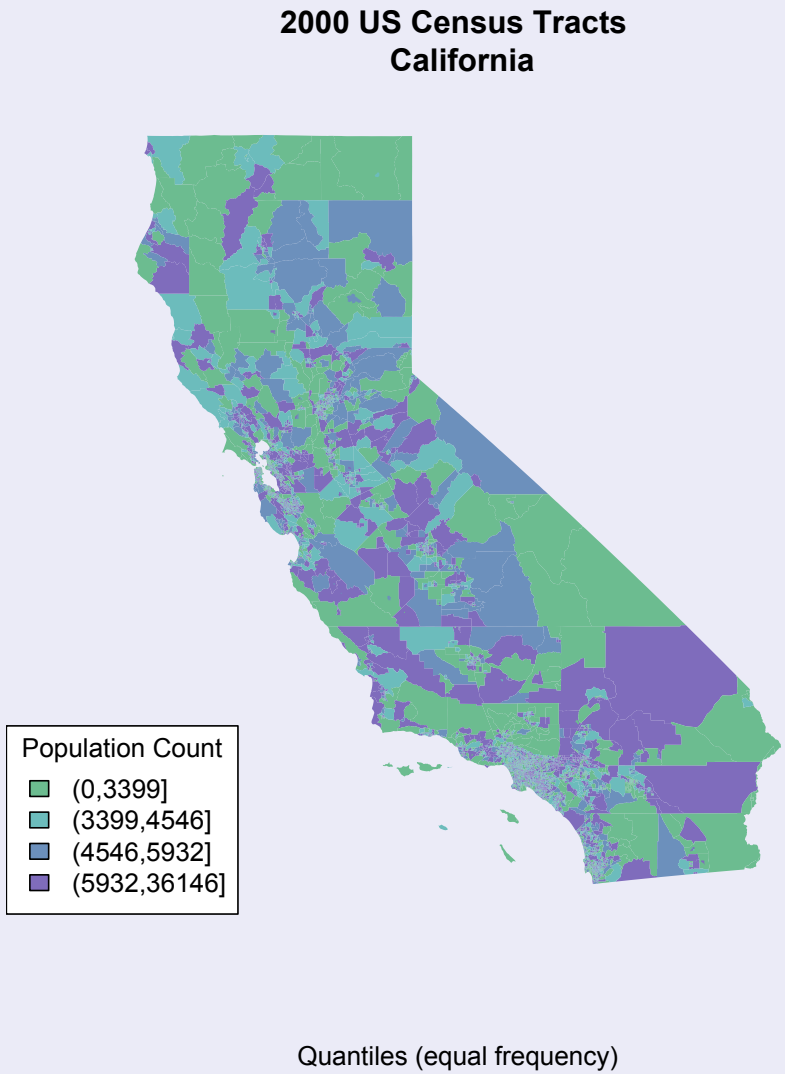
choropleth map based on plot()

```
> choropleth(california.tract,
+  main="2000 US Census Tracts \n California",
+ border="transparent")
```

Note:

choropleth(*,type="spplot") produces a quantile choropleth map and legend of population counts based on spplot().

choropleth()


2000 US Census Tracts
California
Population Count
(0,3399]
(3399,4546]
(4546,5932]
(5932,36146]
Quantiles (equal frequency)

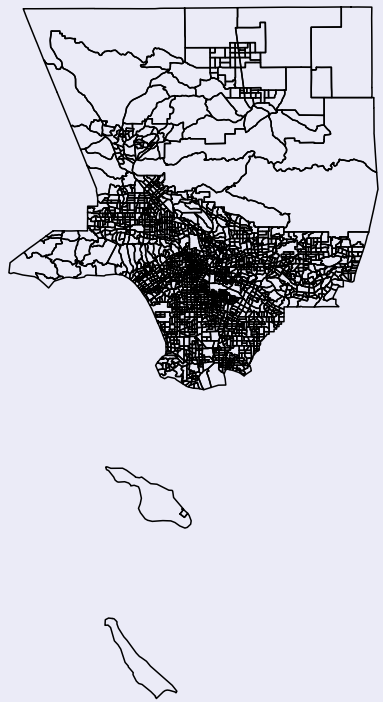
UScensus2000

county() – Output: SpatialPolygonsDataframe

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

UScensus2000

county()



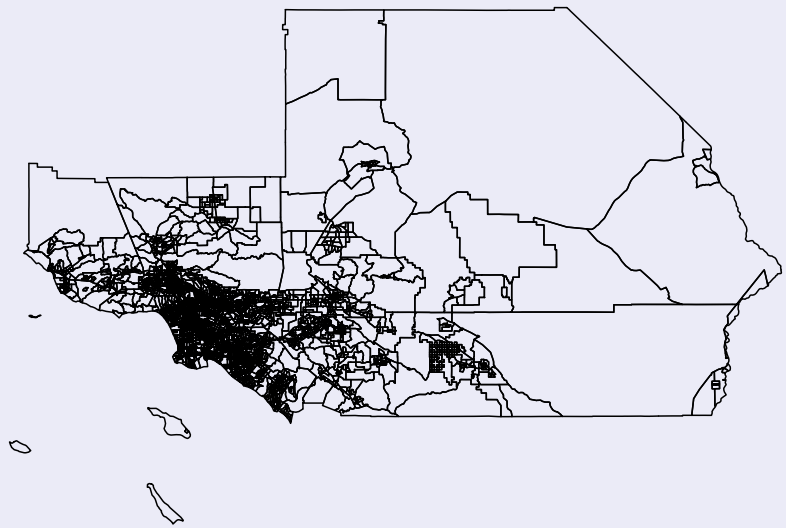
UScensus2000

MSA() – Output: SpatialPolygonsDataframe

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

UScensus2000

MSA()



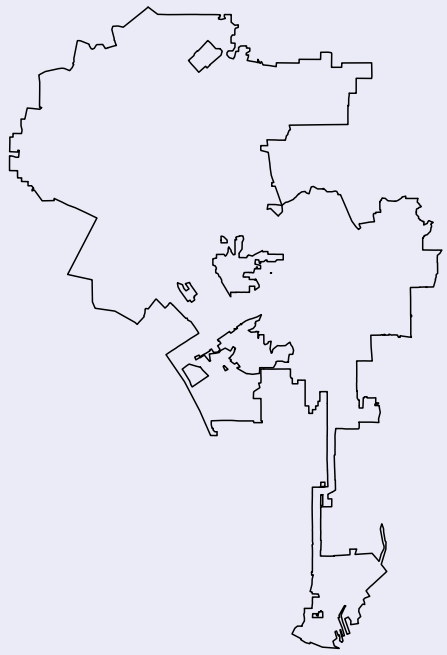
UScensus2000

city() – Output: SpatialPolygonsDataframe

```
> losangeles<-city(name="los angeles",
+   state="ca")
> plot(losangeles)
```

UScensus2000

city()



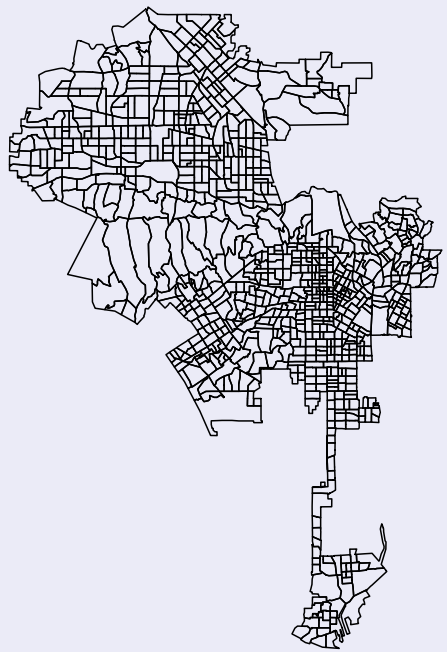
UScensus2000

poly.clipper() – Output: SpatialPolygonsDataframe

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

UScensus2000

poly.clipper()



UScensus2000

demographics() – Output: matrix

```
>   laMSAarea<-demographics(
+   dem=c("pop2000","white","black"),
+   "CA",level="msa",msaname="Los Angeles")
>   laMSAarea
```

UScensus2000

demographics() – Output: matrix

	pop2000	white	black
san bernardino county	1709434	1006960	155348
ventura county	753197	526721	14664
los angeles county	9519338	4637062	930957
riverside county	1545387	1013478	96421
orange county	2846289	1844652	47649

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],]
```

UScensus2000

demographics() – Output: matrix

	pop2000	white	black	hh.units	hh.vacant
Acton	2390	2130	17	873	76
Adelanto	18130	9147	2377	5547	833
Agoura Hills	20537	17858	272	6993	119
Alameda	72259	41148	4488	31644	1418
Alamo	15626	14119	74	5497	91
Albany	16444	10078	675	7248	237
Alhambra	85804	25758	1437	30069	958
Aliso Viejo	40166	31395	828	16608	461
Almanor	0	0	0	74	74
Alondra Park	8622	3584	1088	2933	103

UScensus2000add

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)

UScensus2000

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 US Census's (1970, 1980, 1990)
- Expand to other countries (Europe, South America, etc)

- Thanks!

References I

Bivand, Roger S., Edzer J. Pebesma, and Virgilio Gómez-Rubio. 2008. *Applied Spatial Data Analysis with R*. New York, NY: Springer.