<http://franklincenterhq.org/2544/computing-distance-matrix-between-missouri-sex-offenders-and-child-daycare-facilities/>

\FranklinCenter\Projects\2011\KSHB-Sex-Offenders

**Computing distance matrix between Missouri sex offenders and child daycare facilities**

By Earl Glynn Thursday, July 28th, 2011 –

This is the third of four articles about analyzing distances between sex offenders and child daycare centers in Missouri as part of a [*joint project with KSHB NBC Action News in Kansas City*](http://www.nbcactionnews.com/dpp/news/local_news/investigations/loophole-in-law-allows-hundreds-of-missouri-sex-offenders-to-live-near-church-daycares).

The [previous article explained how to use Google geocoding](http://www.franklincenterhq.org/2541/geocoding-addresses-from-missouri-sex-offender-registry/) to find the longitude-latitude coordinates from street addresses in a file.  That article focused on finding the coordinates of registered Missouri sex offenders.

The same geocoding approach was also applied to obtain the coordinates of child daycare providers from addresses, and in particular, the church daycare centers in Missouri.

**Longitude-Latitude Coordinates**

In this example we use geolocation coordinates of selected Missouri sex offenders and church child day care providers  from Independence, Missouri.  We label offenders and providers only with numbers to focus on the algorithm here.

Let’s use R to define the coordinates of  6 offenders and 4 providers:

> offender <- data.frame(  
+ longitude=c(-94.4232950, -94.4201300, -94.4202860,  
+ -94.3775998, -94.4536700, -94.4190120),  
+ latitude =c( 39.0944540, 39.0954390, 39.0956020,  
+ 39.1147769, 39.0908940, 39.0954190) )  
> rownames(offender) <- paste("offender", 1:nrow(offender), sep="")

**> offender  
longitude latitude  
offender1 -94.42329 39.09445  
offender2 -94.42013 39.09544  
offender3 -94.42029 39.09560  
offender4 -94.37760 39.11478  
offender5 -94.45367 39.09089  
offender6 -94.41901 39.09542**

> provider <- data.frame(  
+ longitude=c(-94.4522520, -94.4210226, -94.4202310,  
+ -94.3745936),  
+ latitude =c( 39.0890650, 39.0941899, 39.0933280, + 39.1153204) )  
> rownames(provider) <- paste("provider", 1:nrow(provider), sep="")

**> provider  
longitude latitude  
provider1 -94.45225 39.08906  
provider2 -94.42102 39.09419  
provider3 -94.42023 39.09333  
provider4 -94.37459 39.11532**

Longitude is measured in degrees north, and latitude is measured in degrees east.

The negative longitude values reflect they are to the west of the [prime meridian](http://en.wikipedia.org/wiki/Prime_Meridian) (i.e., “negative east” is west).  If given as positive values, the longitude numbers are degrees west.

**Great Circle Distance Computation**

In high school geometry one learns the [Euclidean distance between two points (x1,y1) and (x2, y2)](http://en.wikipedia.org/wiki/Euclidean_distance) is computed as SQRT( (x2-x1)^2 + (y2-y1)^2 ), where SQRT is square root. But that formula is not relevant to computing distances using longitude and latitude coordinates.

The equator is about 25,000 miles in 360 degrees of longitude (0 to 180 degrees east and 0 to 180 degrees west).  If the Earth were a perfect sphere, each degree of longitude at the equator would be about 69.4 miles (25,000 miles/360 degrees).

At least near the equator, we could get distance approximations by assuming each degree of latitude or longitude was about 69.4 miles and then applying the Euclidean distance formula.

But as we go north or south of the equator the length of one degree of longitude decreases and is zero at the north or south poles.  How do we compute a distance between two points on a sphere?

The shortest distance between two points on the surface of a sphere is known as the [great-circle distance](http://en.wikipedia.org/wiki/Great-circle_distance).  Since the Earth is roughly a sphere, we can get good distance approximations using the great circle distance formula.  This distance is sometimes call the “crow fly” distance and does not correct for hills or buildings.

The formula to compute great circle distances uses a number of functions from trigonometry and is a bit “messy.”  [Please see this wiki page for details](http://en.wikipedia.org/wiki/Great-circle_distance).

The R package “fields” provides a function **rdist.earth** for great circle distance computations.

For example, the distance in miles between offender1 and provider1 can be computed with this R statement.

> rdist.earth(data.frame(-94.4232950,39.0944540),  
data.frame(-94.4522520, 39.0890650))  
[,1]  
[1,] **1.598711**

If we want distances in feet instead of miles, we can multiply by 5280 feet/mile and use R’s “round” function to ignore fractional feet:

> round(5280\*rdist.earth(data.frame(-94.4232950,39.0944540),  
data.frame(-94.4522520, 39.0890650))     )  
[,1]  
[1,] **8441**

The result says provider1 is about 8441 feet from offender1, which is about 1.6 miles.

**Computation of Distance Matrix**

The **rdist.earth** function normally is not used to compute a single distance but a whole matrix of distances between two sets of geolocation coordinates:

> distance.matrix.ft <- round(5280\*rdist.earth(offender, provider, miles=TRUE))  
> colnames(distance.matrix.ft) <- rownames(provider)  
> rownames(distance.matrix.ft) <- rownames(offender)

**> distance.matrix.ft  
provider1 provider2 provider3 provider4  
offender1 8441 651 961 15767  
offender2 9399 522 772 14808  
offender3 9371 556 831 14818  
offender4 23149 14422 14400 875  
offender5 780 9333 9521 24123  
offender6 9704 725 838 14537**

Sadly, WordPress does not  maintain the fixed-width spacing well in the display above.

Here’s the same matrix in a more readable format (best viewed with Firefox):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Distance[ft]** | provider1 | provider2 | provider3 | provider4 |
| offender1 | 8,441 | 651 | 961 | 15,767 |
| offender2 | 9,399 | 522 | 772 | 14,808 |
| offender3 | 9,371 | 556 | 831 | 14,818 |
| offender4 | 23,149 | 14,422 | 14,400 | 875 |
| offender5 | 780 | 9,333 | 9,521 | 24,123 |
| offender6 | 9,704 | 725 | 838 | 14,537 |

Caution:  These are only approximate distances and they should be verified with “ground truth” of actual measurements.

**Problem Distances**

Since our concern is about distances less than 1000 feet, let’s ignore longer distances to focus on potential problems:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Distance[ft]** | provider1 | provider2 | provider3 | provider4 |
| offender1 | - | 651 | 961 | - |
| offender2 | - | 522 | 772 | - |
| offender3 | - | 556 | 831 | - |
| offender4 | - | - | - | 875 |
| offender5 | 780 | - | - | - |
| offender6 | - | 725 | 838 | - |

Depending on the desired perspective, we can read across the rows of this matrix to find providers who are too close to an offender, or read down the columns to find offenders too close to a provider.  For example:

By Row:

* Offender1, Offender2, Offender3 and Offender6 are all too close to Provider2 and Provider3.
* Offender4 is too close to Provider4.
* Offender5 is too close to Provider1.

By Columns:

* Provider1 is too close to Offender5
* Provider2 and Provider3 are too close to Offender1, Offener2, Offender3, and Offender6.
* Provider4 is too close to Offender4.

The R **rdist.earth** function can be applied to fairly large datasets.

In the Missouri example investigated with KSHB, we looked at about 11,000 Missouri registered sex offenders and about 500 church daycare centers.

Earlier this year we developed this methodology in looking at [about 11,000 Colorado registered sex offenders and about 4,000 child care center](http://investigates.i2i.org/2011/02/04/databases-show-sex-offenders-registered-at-state-subsidized-child-care-facilities/)s.

**Rough Map**

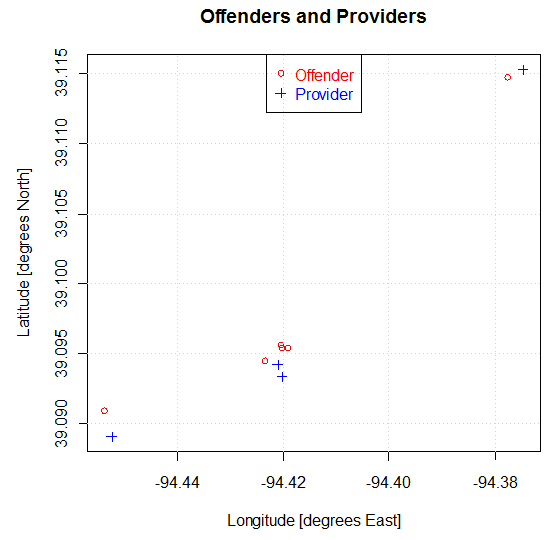
R can be used to plot a quick map to see the locations of the offenders and providers

# Rough map  
plot(offender, col="red", pch=1,  
xlim=range(c(offender$longitude, provider$longitude)),  
ylim=range(c(offender$latitude, provider$latitude )),  
xlab="Longitude [degrees East]", ylab="Latitude [degrees North]",  
main="Offenders and Providers")

points(provider, col="blue", pch=3)

grid()

legend("top", c("Offender", "Provider"), pch=c(1,3),  
col=c("red", "blue"), text.col=c("red","blue"))

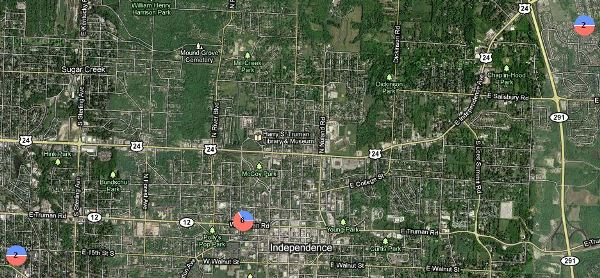


*Example of nearby sex offenders and church child care centers in Missouri*

This map has slightly different scaling horizontally and vertically.

**A Better Map**

An interactive Google map is a better way to display the data.



Next:  [Displaying Missouri sex offender/child day care facility proximity map using batchgeo.com](http://www.franklincenterhq.org/2563/displaying-missouri-sex-offenderchild-day-care-facility-proximity-map-using-batchgeo-com/)

**Program and Files**

[Create-Distance-Matrix-Offenders-Providers.R](http://cdn.watchdogmedia.org/national/computer-assisted-reporting/project/geocoding-and-distances/missouri-sex-offenders/distance-matrix/Create-Distance-Matrix-Offenders-Providers.R) reads longitude-latitude coordinates for sex offenders and child daycare providers, computes a distance matrix, and creates output files with various summaries.  (You will want to modify the basedir directory in line 9 to be your working directory.)

Input Files (output from geocoding process):

* geocode-MO-offender-out.txt ([extract from zip file](http://cdn.watchdogmedia.org/national/computer-assisted-reporting/project/geocoding-and-distances/missouri-sex-offenders/distance-matrix/geocode-MO-offender-out.ZIP))
* [geocode-ReligiousOrganizations-out.txt](http://cdn.watchdogmedia.org/national/computer-assisted-reporting/project/geocoding-and-distances/missouri-sex-offenders/distance-matrix/geocode-ReligiousOrganizations-out.txt)

Output Files:

* [ChurchDaycare-min-dist-histograms.pdf](http://cdn.watchdogmedia.org/national/computer-assisted-reporting/project/geocoding-and-distances/missouri-sex-offenders/distance-matrix/ChurchDaycare-min-dist-histograms.pdf)
* [ChurchDaycare-min-distance-from-provider-to-any-offender.txt](http://cdn.watchdogmedia.org/national/computer-assisted-reporting/project/geocoding-and-distances/missouri-sex-offenders/distance-matrix/ChurchDaycare-min-distance-from-provider-to-any-offender.txt)
* ChurchDaycare-min-distance-from-offender-to-any-provider.txt ([extract from zip file](http://cdn.watchdogmedia.org/national/computer-assisted-reporting/project/geocoding-and-distances/missouri-sex-offenders/distance-matrix/ChurchDaycare-min-distance-from-offender-to-any-provider.ZIP))
* [ChurchDaycare-provider-offender-connections-threshold-1000ft.txt](http://cdn.watchdogmedia.org/national/computer-assisted-reporting/project/geocoding-and-distances/missouri-sex-offenders/distance-matrix/ChurchDaycare-provider-offender-connections-threshold-1000ft.txt)

Related:

* [Loophole in law allows hundreds of Missouri sex offenders to live near church daycares](http://www.nbcactionnews.com/dpp/news/local_news/investigations/loophole-in-law-allows-hundreds-of-missouri-sex-offenders-to-live-near-church-daycares), KSHB NBC Action News, Kansas City, July 18, 2011.
* [Analysis of Missouri Sex Offender Registry Data](http://www.franklincenterhq.org/2536/analysis-of-missouri-sex-offender-registry-data/), Franklin Center CAR Blog, July 18, 2011.
* [Geocoding addresses from Missouri Sex Offender Registry](http://www.franklincenterhq.org/2541/geocoding-addresses-from-missouri-sex-offender-registry/), Franklin Center CAR Blog, July 19, 2011.
* [Displaying Missouri sex offender/child day care facility proximity map using batchgeo.com](http://www.franklincenterhq.org/2563/displaying-missouri-sex-offenderchild-day-care-facility-proximity-map-using-batchgeo-com/), Franklin Center CAR Blog, July 28, 2011.
* [Databases show sex offenders registered at State of Colorado subsidized child care facilities](http://investigates.i2i.org/2011/02/04/databases-show-sex-offenders-registered-at-state-subsidized-child-care-facilities/), Independence Institute, Feb. 4, 2011.
* [Great-circle distance calculations in R](http://www.r-bloggers.com/great-circle-distance-calculations-in-r/), R-bloggers, Nov. 23, 2010.
* [Using Google’s geocoding service and the Great Circle Distance-formula to make distance-based queries against a MySQL database](http://maurus.net/resources/distance-queries/), maurus.net, visited July 27, 2011.

[**Earl F Glynn**](mailto:Earl.Glynn@FranklinCenterHq.org) **•** [**KansasWatchdog.org**](http://kansaswatchdog.org/) **• Franklin Center for Government and Public Integrity** - See more at: http://franklincenterhq.org/2544/computing-distance-matrix-between-missouri-sex-offenders-and-child-daycare-facilities/#sthash.GAlaBmdg.dpuf