

clustermx

Stata command to create clusters from point coordinates

Francisco Barba

International Food Policy Research Institute

October 18, 2018

Overview

- 1 Examples
- 2 The Algorithm
- 3 Further Improvements

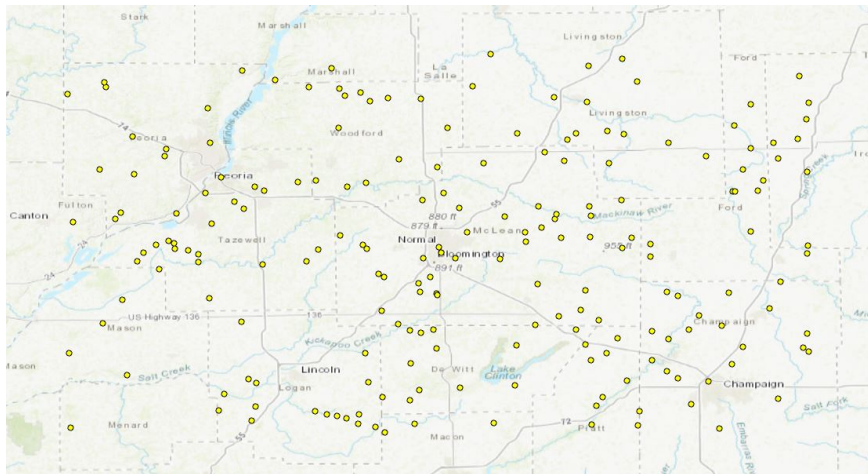
Overview

- 1 Examples
- 2 The Algorithm
- 3 Further Improvements

Generate 200 random points in Illinois, USA

```
. set seed 1  
. set obs 200  
. gen lat = 40 + runiform()  
. gen lon = -90 + runiform()*2
```

Examples



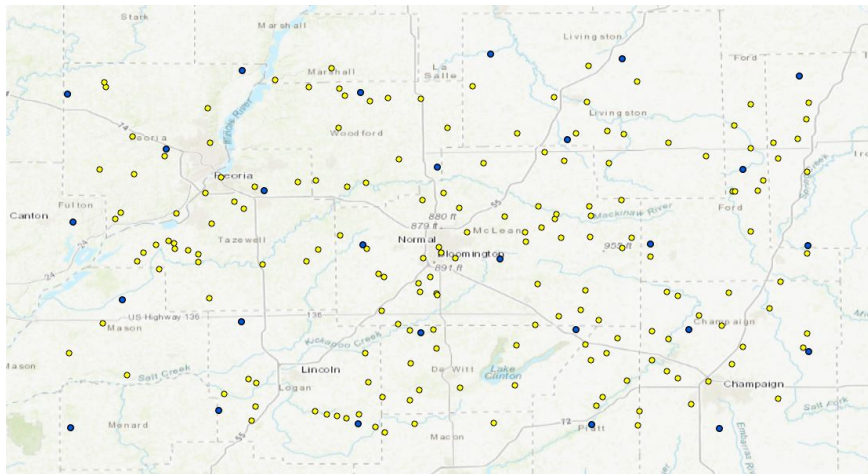
Question:

How many points with a minimum mutual distance of 25km can be drawn from this map?

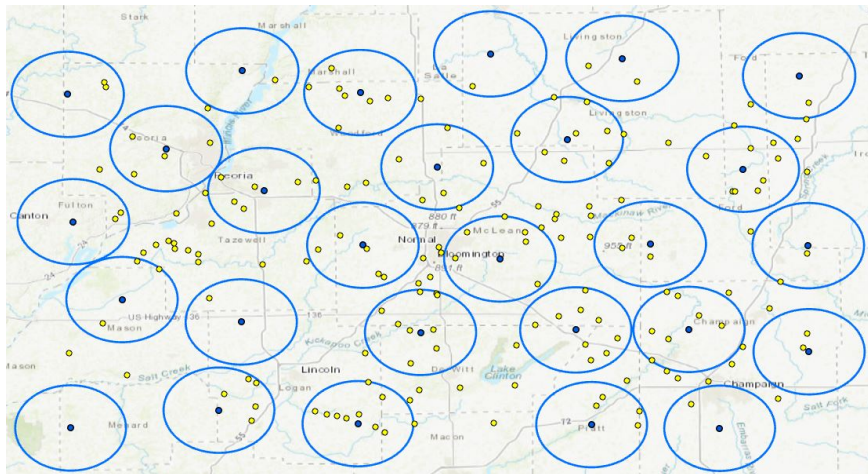
Run the command:

```
. clustermax lat lon, gen(cid) n(1) between(25)  
seed(1)
```

Examples



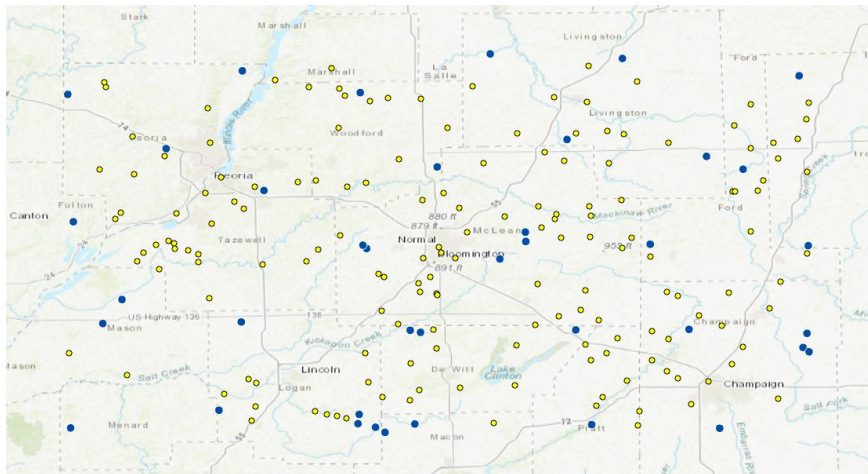
Examples



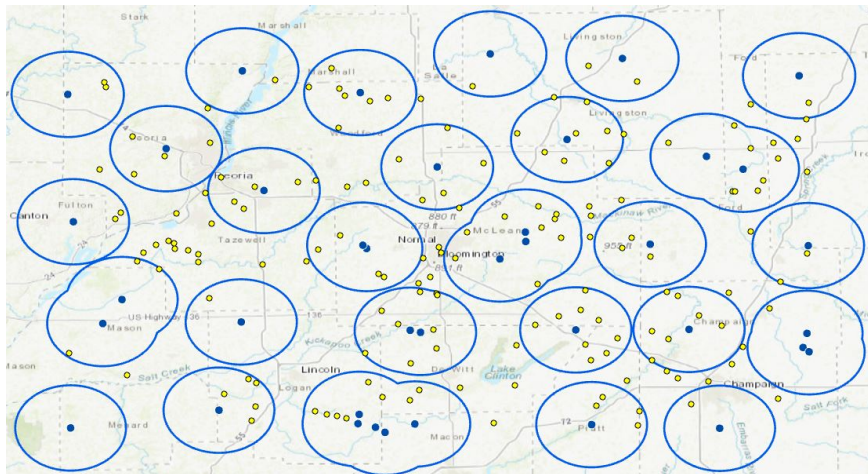
Can more points be added, given they satisfy the distance constraints?

```
. clustermax lat lon, gen(cid) n(1) between(25)  
within(15) seed(1)
```

Examples



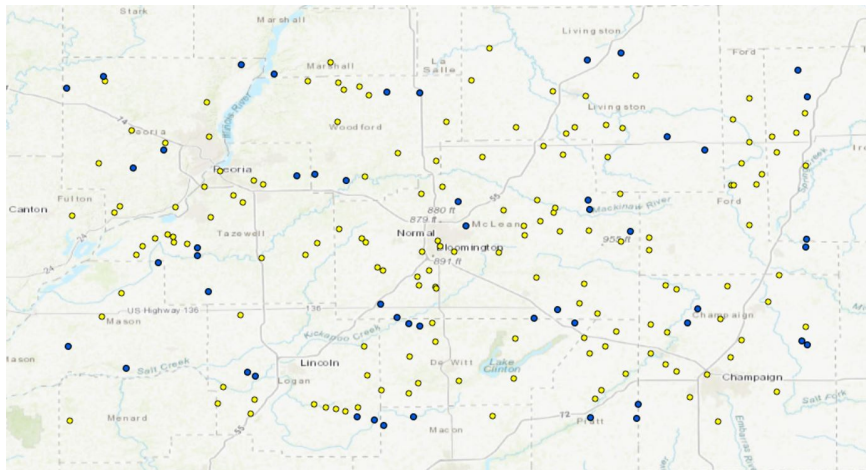
Examples



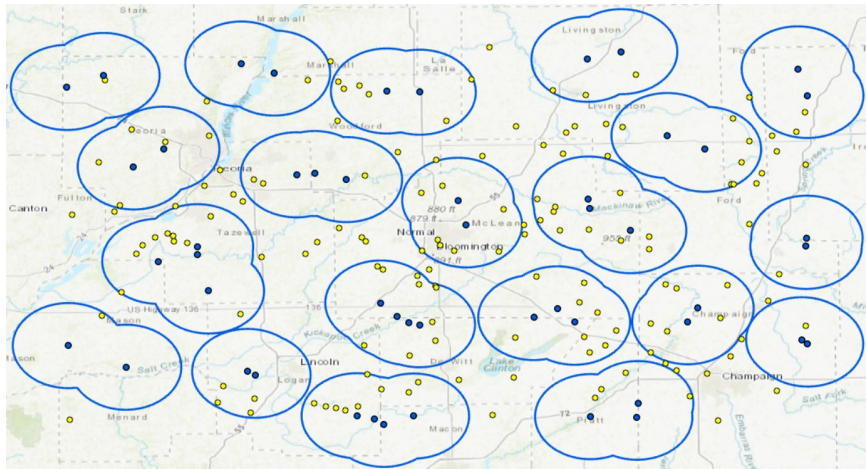
Set minimum of two points per cluster:

```
. clustermax lat lon, gen(cid) n(2) between(25)  
within(15) seed(1)
```

Examples



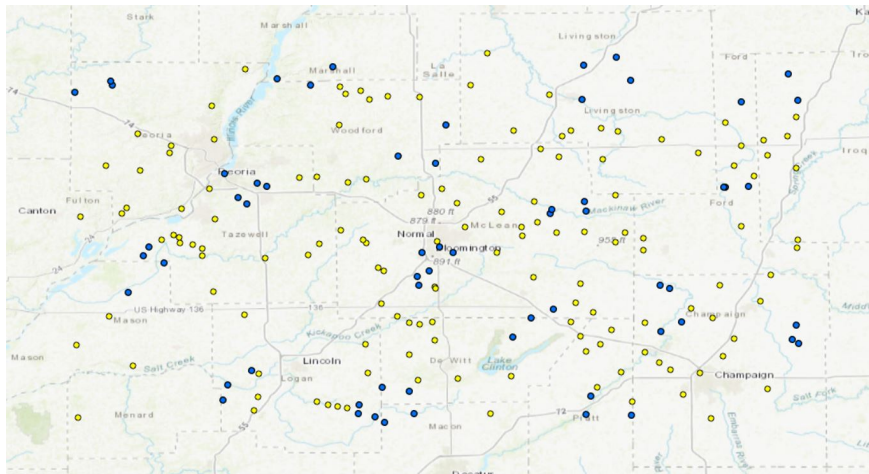
Examples



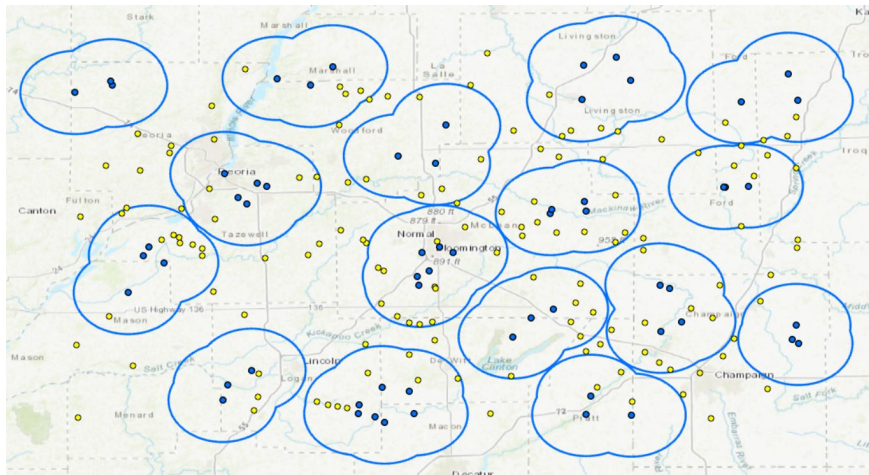
Set minimum of three points per cluster:

```
. clustermax lat lon, gen(cid) n(3) between(25)  
within(15) seed(1)
```

Examples



Examples



Overview

- 1 Examples
- 2 The Algorithm
- 3 Further Improvements

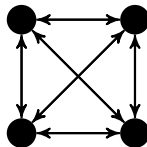
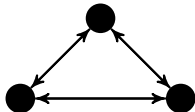
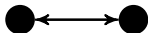
The Algorithm

1. Find groups of $n()$ points mutually in *within()* distance
2. Resolve *between()* cluster conflicts
3. Join additional points

The Algorithm

1. Find groups of $n()$ points mutually in *within()* distance
2. Resolve *between()* cluster conflicts
3. Join additional points

The Algorithm



The Algorithm



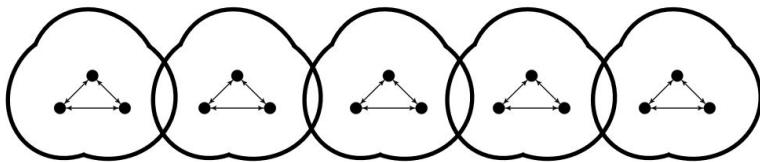
The Algorithm



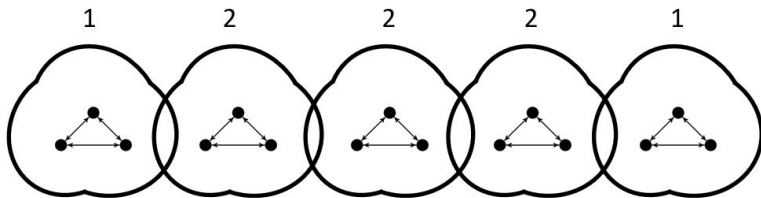
The Algorithm

1. Find groups of $n()$ points mutually in *within()* distance
2. Resolve *between()* cluster conflicts
3. Join additional points

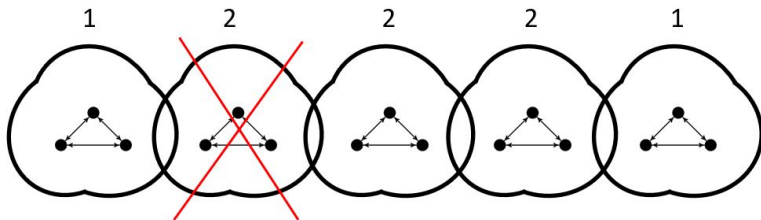
The Algorithm



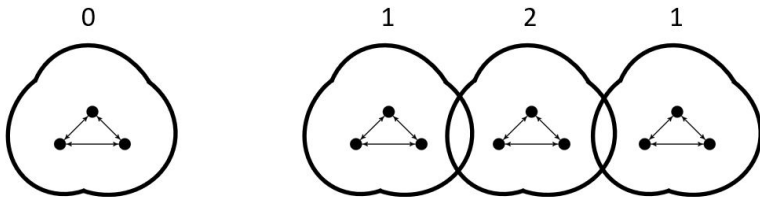
The Algorithm



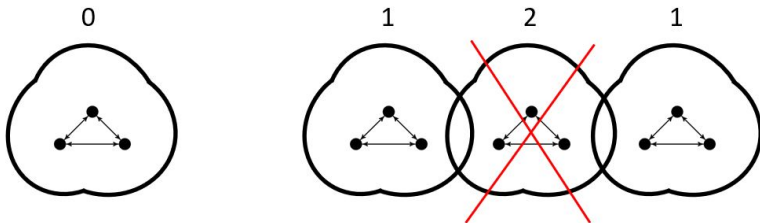
The Algorithm



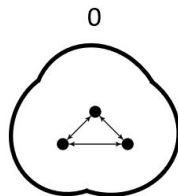
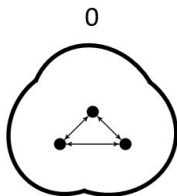
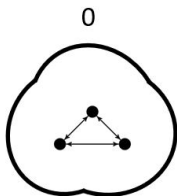
The Algorithm



The Algorithm



The Algorithm



The Algorithm

1. Find groups of $n()$ points mutually in *within()* distance
2. Resolve *between()* cluster conflicts
3. Join additional points

Overview

- 1 Examples
- 2 The Algorithm
- 3 Further Improvements**

Additional options

- Add option to import own distance matrix (travel time, social distance, etc.)
- Add within-cluster stratification conditions (at least 1 male and 1 female, etc.)

Stata ado-file and helpfile:

`https://github.com/fmbarba`

Related Literature:



Jade Benjamin-Chung, Benjamin F Arnold, David Berger, Stephen P Luby, Edward Miguel, John M Colford Jr, Alan E Hubbard (2018)

Spillover effects in epidemiology: parameters, study designs and methodological considerations

International Journal of Epidemiology, 2018, 332-347