



Comparing methods of safely plotting variables on a map

Joint UNECE/EUROSTAT work session on Statistical Data Confidentiality

Sapphire Yu Han*, Edwin de Jonge*, and Peter-Paul de Wolf* (*Statistics Netherlands)
31-October-2019, Statistics Netherlands, the Hague, the Netherlands

Content

- Home-taking messages
- Introduction
- Disclosure & Risk
- SDC for maps
- Data & Methods
- Results
- Conclusion



Home-taking messages

1. When disseminating data on maps, smoothed raster maps (as a SDC method) show great potential compared to traditional postal code maps.
 - Advantages of raster maps: spatial pattern, less risky units, and flexible risk management
2. Try our [R package sdcSpatial by de Jong & de Wolf \(2019\)](#)

Introduction

Disseminating data on cartographic maps is useful

- Topics: population, land use, geographical distribution of characteristics.
- Merits: Information, spatial pattern, policy, and communication
- Examples: [Eurostat](#), [World Bank](#), [CBS](#), [INSEE](#), [Statistics Finland](#), etc.



Introduction

Traditional maps

- Aggregated data plotted
- e.g. mean electricity consumption of enterprises in 2014 using administrative regions and zooming

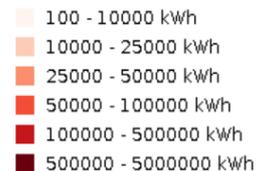
Introduction

Traditional maps

- Aggregated data plotted
- e.g. mean electricity consumption of enterprises in 2014 using administrative regions and zooming



Electricity supply per company (kWh/year)



Introduction

Traditional maps

- Aggregated data plotted
- e.g. mean electricity consumption of enterprises in 2014 using administrative regions and zooming



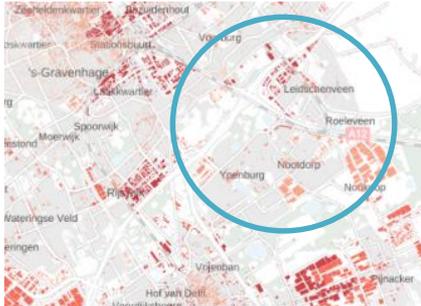
Electricity supply per company (kWh/year)



Introduction

Traditional maps

- Aggregated data plotted
- e.g. mean electricity consumption of enterprises in 2014 using administrative regions and zooming



Electricity supply per company (kWh/year)



Introduction

Traditional maps

- Aggregated data plotted
- e.g. mean electricity consumption of enterprises in 2014 using administrative regions and zooming



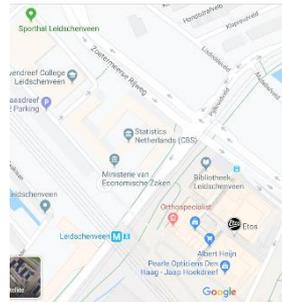
Electricity supply per company (kWh/year)



Introduction

Traditional maps

- Aggregated data plotted
- e.g. mean electricity consumption of enterprises in 2014 using administrative regions and zooming



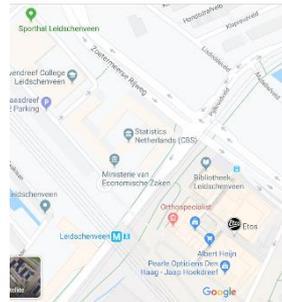
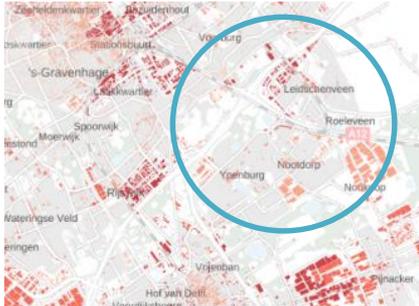
Electricity supply per company (kWh/year)



Introduction

Traditional maps

- Aggregated data plotted
- e.g. mean electricity consumption of enterprises in 2014 using administrative regions and zooming



Electricity supply per company (kWh/year)



Disclosure & Risk

1. Disclosure scenario

- Location of population unit is very identifying
- Zooming leads to pinpointing exact location of population unit
- Spatial distribution of sensitive variable
- Population distribution in space can be used to identify units

2. Risk measures

- Dominance-rule and minimal-frequency rule



Dealing with unsafe regions

1. Suppress/don't show unsafe regions
2. Combine information of surrounding regions
 - Quad-trees (local clustering regions)
 - Kernel density estimator

Data and Methods

1. Data

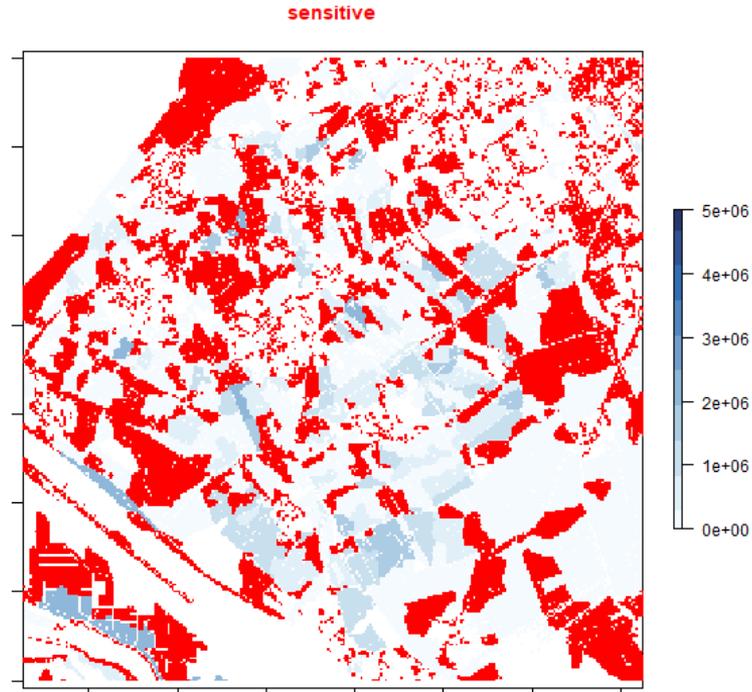
- Actual energy consumption of enterprises in “*Westland*” (Province South Holland) in the year of 2014.
- 7,016 enterprises
- 2,019 unique postal code six digits (regions)

2. Methods

- [R package sdcSpatial by de Jong & de Wolf \(2019\)](#)



Postal code map



Postal code 6

Postal code map

min \ k%	80	85	90	95
3	68%	65%	64%	62%
4	75%	74%	73%	72%
5	81%	80%	79%	79%

Percentage of unsafe postal code regions (tabular SDC)

Postal code map

min \ k%	80	85	90	95
3	68%	65%	64%	62%
4	75%	74%	73%	72%
5	81%	80%	79%	79%

Percentage of unsafe postal code regions (tabular SDC)

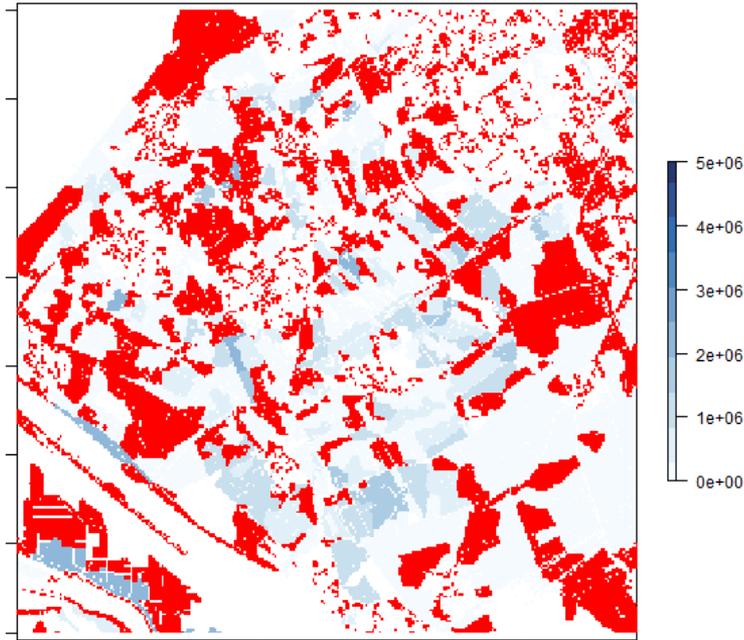
min \ k%	80	85	90	95
3	30%	27%	25%	23%
4	36%	34%	32%	31%
5	42%	41%	40%	39%

Percentage of enterprises in unsafe postal codes (tabular SDC)

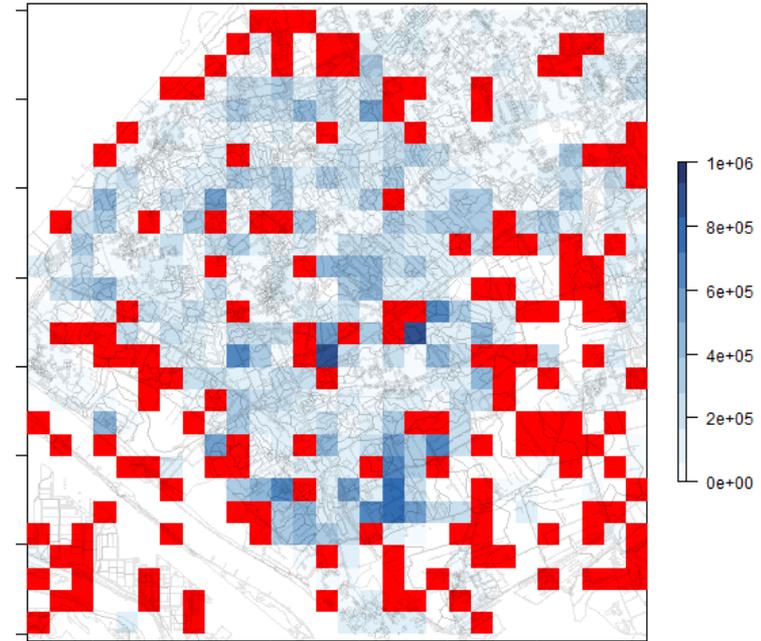


Raster maps

sensitive



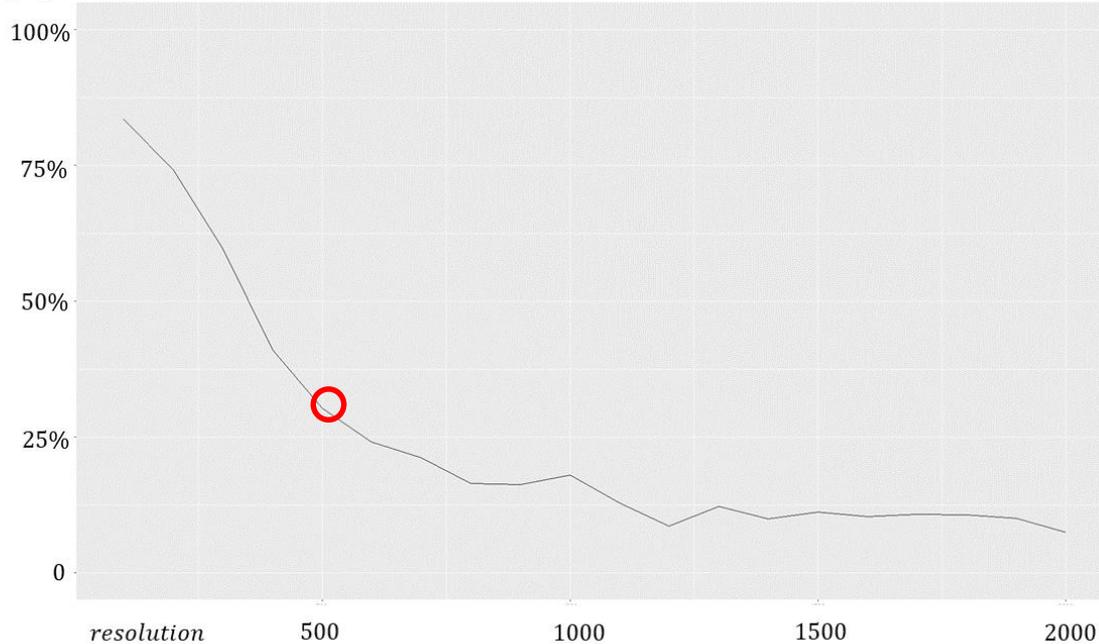
Postal code 6



Raster map (res = 500 m)

Raster maps

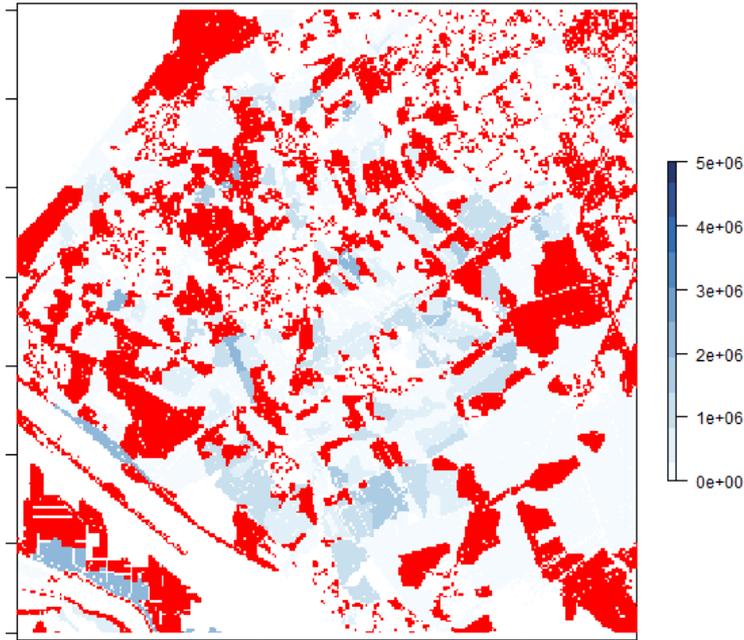
unsafe grids



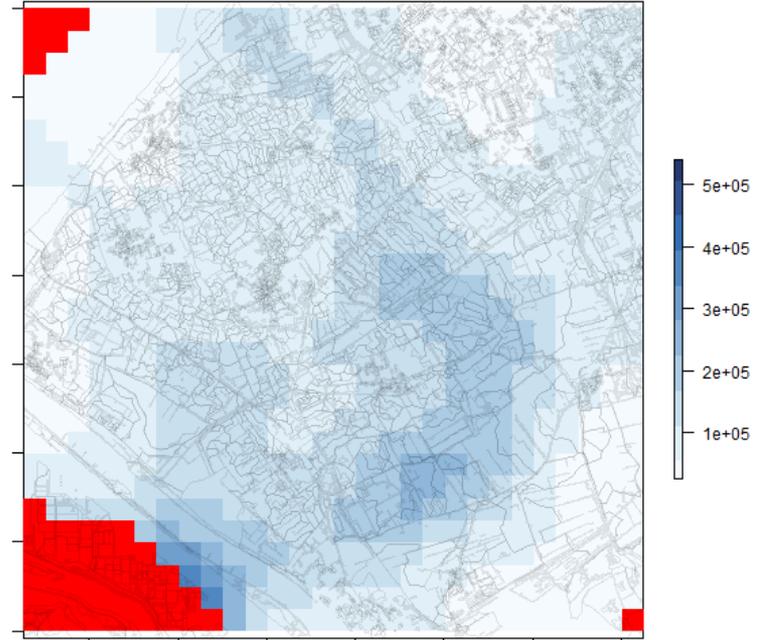
Percentage of unsafe grid cells (min $n = 3$, $k = 90$)

KDE smoothed maps

sensitive

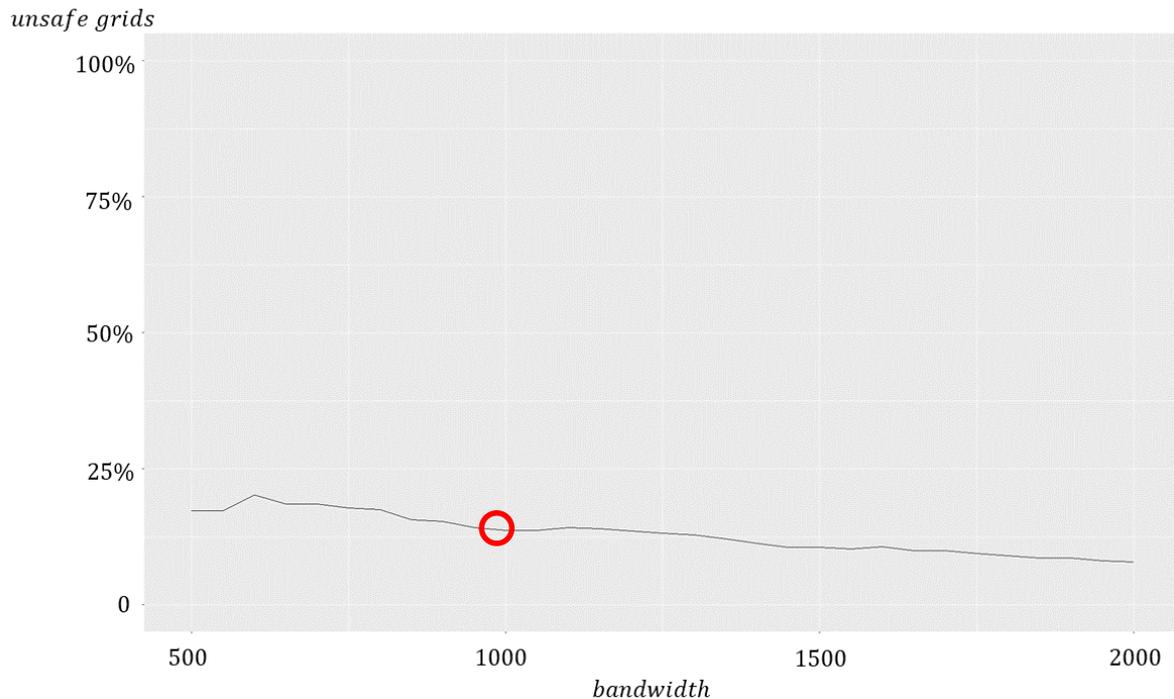


Postal code 6



Smoothed raster map

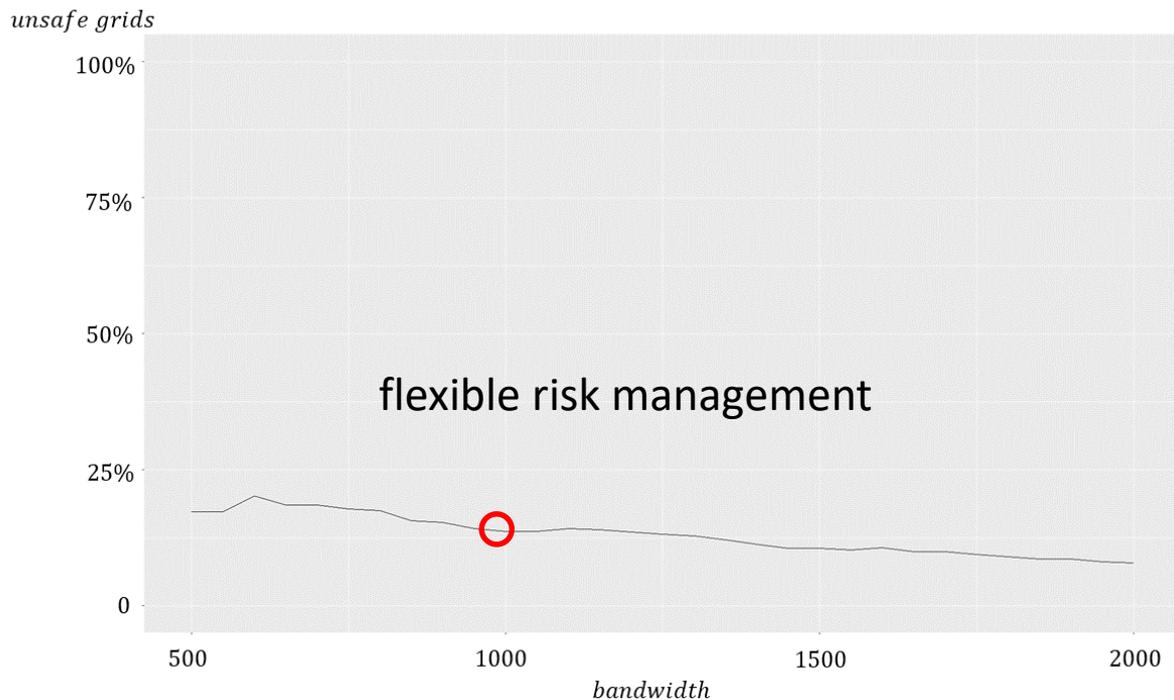
KDE smoothed maps



Percentage of unsafe grid cells (min $n = 3$, $k = 90$, res = 500)



KDE smoothed maps



Percentage of unsafe grid cells (min $n = 3$, $k = 90$, res = 500)



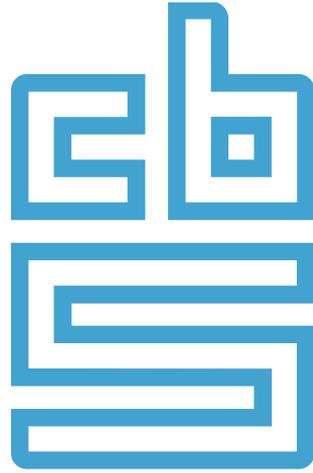
Comparison

KDE \ PC6	Safe	Unsafe
Safe	75%	25%
Unsafe	0	0

Cross tabulation: safe and unsafe enterprises KDE raster maps vs. postal code 6 (min $n = 3$, $k = 90$, res = 500, bandwidth = 1000). Values $< 0.5\%$ rounded to 0.

Conclusions

1. KDE smoother keeps spatial patterns, PC6 not (suppress regions)
2. The percentage of unsafe enterprises is lower using KDE smoother compared to PC6
3. KDE smoother is flexible in managing the disclosure risk by tuning parameters (resolution and bandwidth)
4. Attention on overlapping between KDE smoothed maps and PC6 maps.



Facts that matter