Merging statistics and geospatial information, 2013 projects - Slovenia

Statistics Explained

This article forms part of Eurostat 's statistical report on *Merging statistics and geospatial information: 2019 edition*
.

Final report 2 March 2016

Problem

Although the Statistical Office of the Republic of Slovenia (SURS) had considerable experience in using and disseminating various register-based geo-referenced statistics, this expertise did not extend to temporal analyses.

Objectives

This project looked at a specific challenge for official statistics, namely, to make use of already existing data managed by mobile network operators to open-up new fields of analyses. The main objectives of the project were to:

- examine the legal framework necessary to permit optimal/secure acquiring, handling and dissemination of data from mobile network operators;

- improve the integration of geo-information and geo-referencing into the statistical production process;

- illustrate how links between geo-information and statistical information could be used to potentially provide additional value and new information.

Method

SURS obtained data on mobile telecommunications from the second largest mobile network operator in Slovenia. The data set covered the period between 1 May and 31 October 2014 and contained information on all events that were initiated by mobile users, as recorded through geo-referenced x,y coordinates of mobile base stations (antennae). The project explored how such large volumes of geo-referenced information might be handled using standard geographic information system (GIS) processes to open-up new forms of temporal analyses for population mobility (as measured by changes in the location of mobile phone users).

The data set that was obtained for the project included the following information:

- the x,y coordinates of base stations used to send/receive radio signals to each mobile phone user;

- a unique user_ID that was anonymised;

- the exact time of initiated events (for example, when making a call, or sending an SMS).

The distribution of mobile base stations tends to be denser in urban and commercial areas or along major transport arteries and consequently scarcer in more rural areas. This issue might, at least to some degree, be rectified if the project were to be extended so it covered more than one mobile network operator.

Through the processing of mobile data, SURS acquired experience in various activities — from data transmission, secure and auditable processing, through to the development of GIS applications and data visualisations. The data were stored in tables (one for each month) with an auditing system centred on an ORACLE database containing in excess of a billion records. The main challenge was in relation to the storage of data (an issue that may prove to be of even greater relevance if the project is, one day, extended to cover the regular transmission of data from all mobile operators in Slovenia).

As SURS wished to compare the data on the position of mobile phones with register-based data, it was necessary to find a common territorial division whereby both data sources could be joined. To do this, the areas between individual base stations were analysed to determine a set of Voronoi polygons providing an estimation of the boundaries in coverage between base stations.

The Slovenian information commissioner dissuaded SURS from directly linking data on an individual's location/the position of their mobile phone with register-based data, so as to avoid any potential disclosure of user identity. Therefore, SURS developed a model based on a 500 m^2 grid, estimating temporal population densities for each cell. Point-based locations of the base stations were merged with information for municipalities to generate new municipality-like divisions based on the Voronoi polygons around each base station. This solution allowed SURS to develop analyses combining register-based data (such as that for employees or actual residences) with mobile phone data detailing an individual's position during the day. The temporal dimension of the data set placed these statistics in a new perspective as the data on mobile phone locations were captured in real-time, thereby highlighting temporal patterns which could be used for a variety of applications, for example, an analysis of the average time taken to get to work or preferred routes taken by commuters.

Results

On average 9.9 % of mobile phone users covered by the project generated at least three quarters of their events through a single base station and 41.1% generated at least half of their events through a single base station. Aggregating this information to the level of municipalities, almost half (47.0 %) of all users generated at least three quarters of their events in a single municipality, while 82.6 % of users generated at least half of their events in a single municipality.

The results also showed that there was a noticeable influx of commuters into some of Slovenia's major cities each morning, which led to a considerable change in their populations. Figure 1 contrasts day-time and night-time population densities for the territorial boundaries of Ljubljana (the capital of Slovenia) during the period 1 May–31 October 2014; the day-time population having been measured at lunchtime (12:00h-13:00h), while the night-time population was measured after most people had gone to bed (00:00h-01:00h). The information presented is based on administrative data (for the day-time population) and a central population register (for the night-time population) and was calibrated using hourly patterns observed in GPS tracking from mobile phone data. The general pattern of considerably more people being in the city centre during the day-time is clearly apparent when contrasting the two maps, whereas the night-time population grew in some sub-city districts away from the centre, as well as surrounding suburban areas.

Day time | Night time
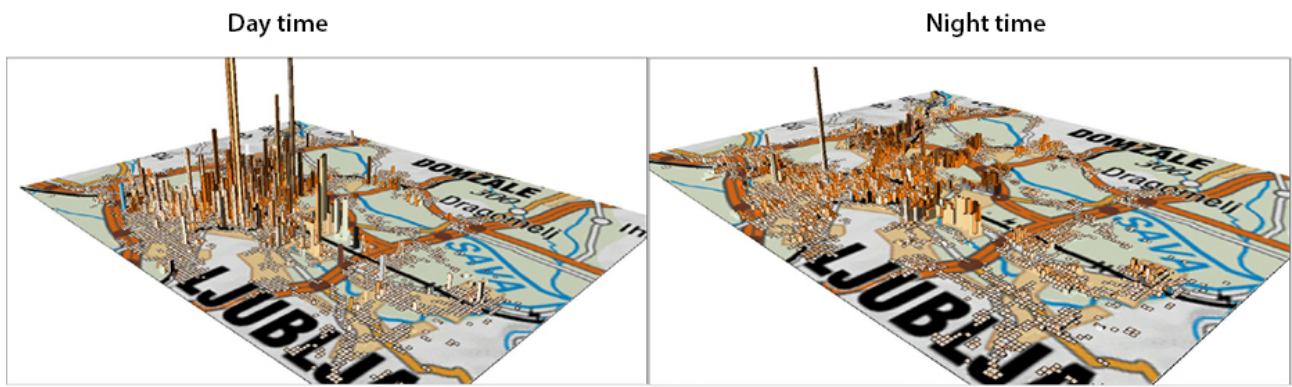
Source: SURS

Figure 1: Day and night time populations, 500 m² grid, Ljubljana, 2015

The results of tracking mobile phone users were also used to estimate the population of Ljubljana, at different points of time during the day (each hour). Figure 2 shows information for four different points in time (00:00, 06:00, 12:00 and 18:00), highlighting how the population of the central business district expands during the working day.
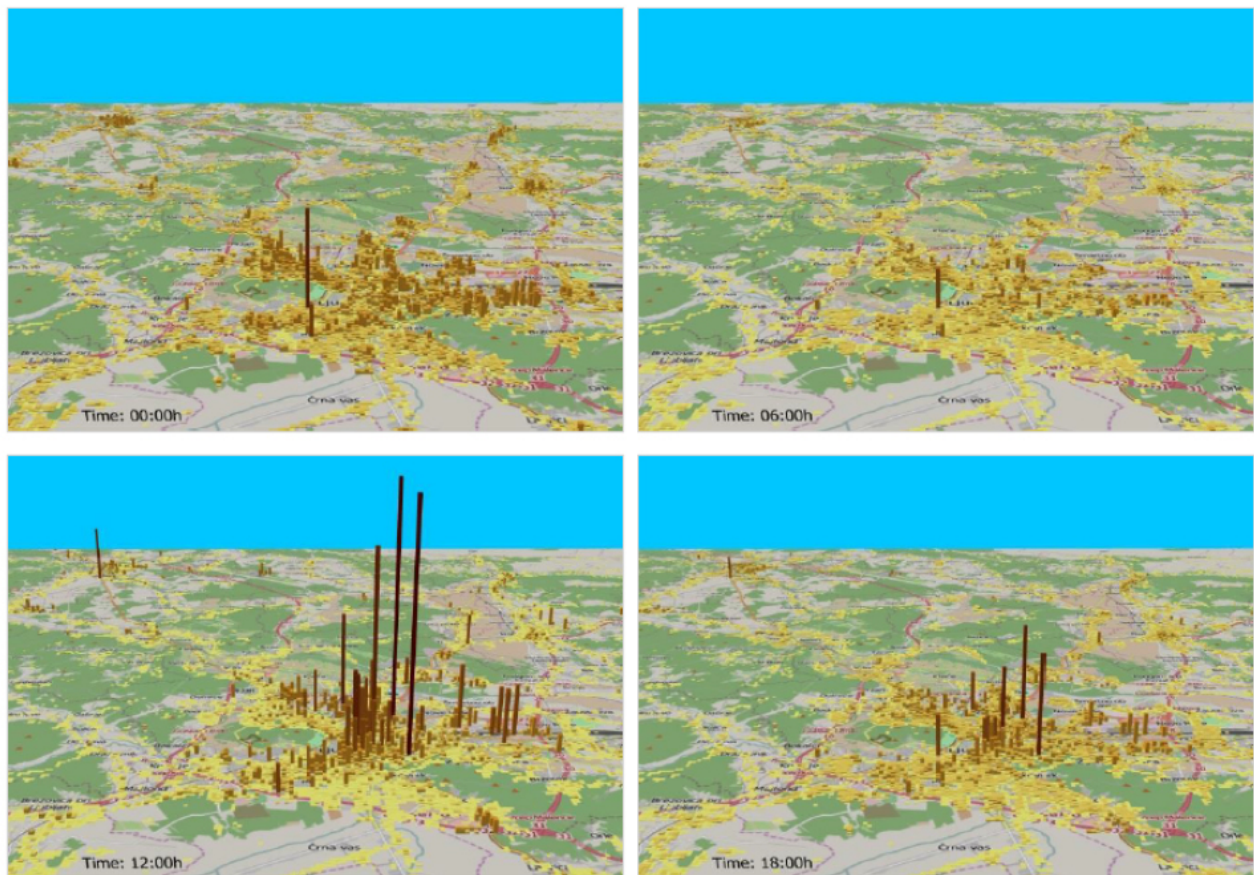
Figure 2: Estimated hourly distribution of the population, Ljubljana, 2015

Information on the mobility of mobile phone users was also used to estimate the attractiveness of various municipalities/cities. Figure 3 presents information on the attractiveness of Ljubljana, Maribor, Novo mesto and Koper, based on the number of commuter inflows into these cities during the working week (Monday-Friday).

A final example demonstrates how information for people usually resident in Ljubljana may be analysed to ascertain

if they leave the city at weekends (perhaps to visit family or to make use of a secondary dwelling in the countryside or on the coast). Figure 4 shows information on weekend locations that are favoured by Ljubljana residents.
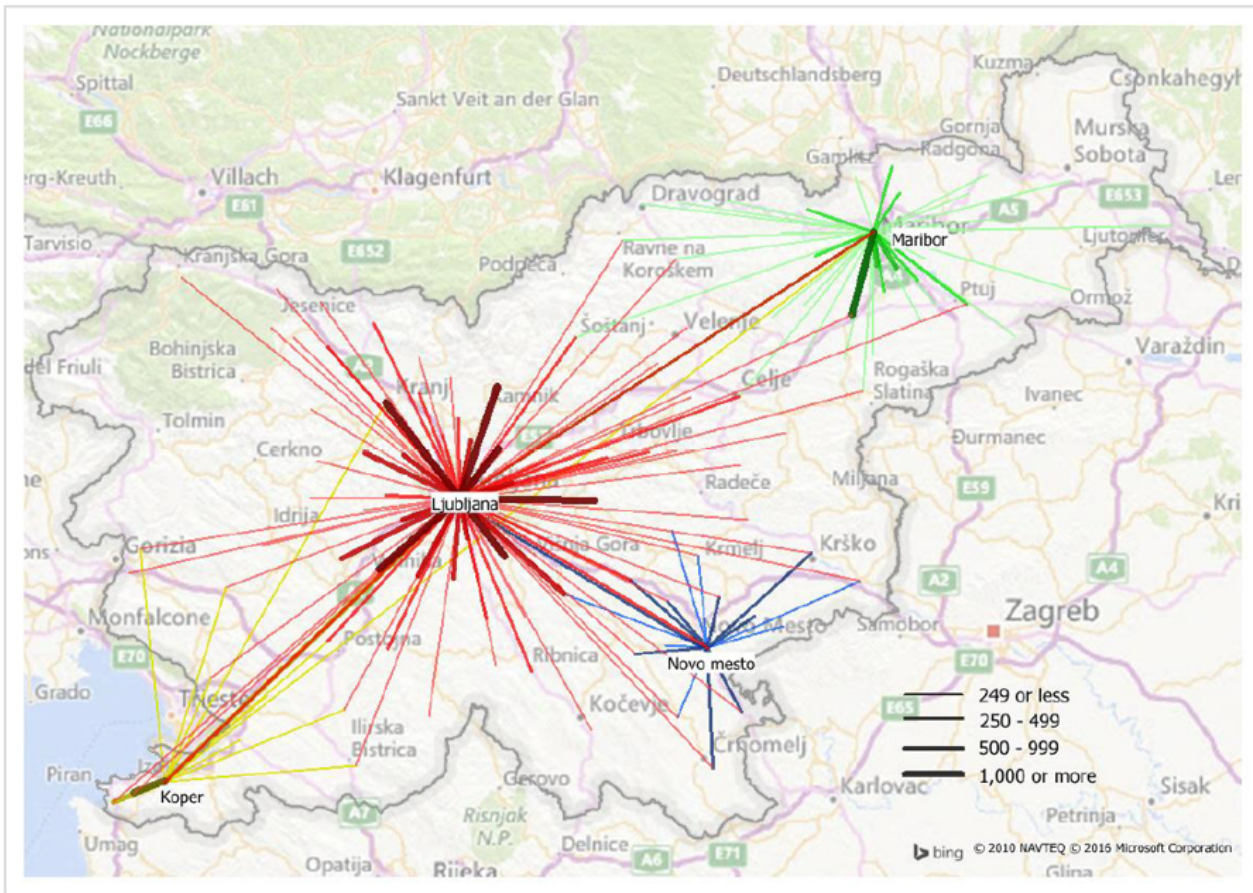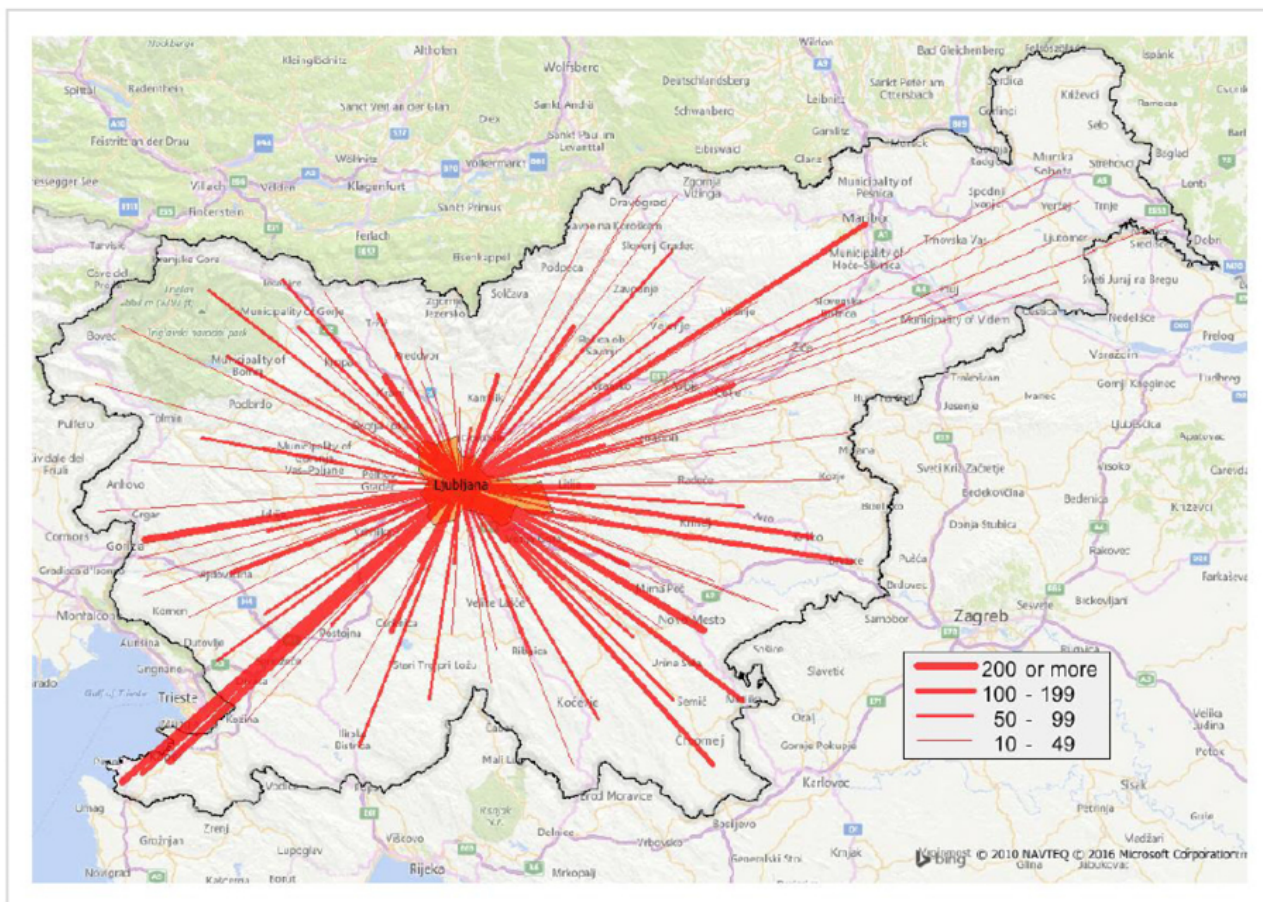

Figure 3: Attractiveness in terms of daily commuter inflows, 2015

Figure 4: Usual location of Ljubljana residents on Saturdays and Sundays, 2015

Given that data access was a considerable issue when setting-up the project — with lengthy discussions over the potential use that might be made of this big data source — SURS devoted a considerable amount of time following the study to promote the activities that were undertaken, with the goal of expanding the project to a range of new applications including, for example, studies of commuter, tourism or transport patterns, or applications relating to civil protection and/or disaster relief.

Methodology

- Final report

- Presentation

- Slovenian statistical regions and municipalities in numbers

External links

- Republic of Slovenia Statistical Office