

Paper 224-2011

Integrating SAS® and ArcMap: A Case Study of European Internet Use and Population

Hillary Huffer, East Carolina University, Greenville, NC, USA

ABSTRACT

This paper demonstrates how to effectively navigate between the SAS® and ArcMap environments. European internet use is utilized as an example to illustrate various techniques, including: how to use a variety of data formats, how to import shapefile data to SAS®, and how to export SAS® output to an ArcMap map layout. This paper concludes that, as expected, European internet use and population are positively correlated and that SAS® and ArcMap can work together seamlessly.

INTRODUCTION

Data comes in a variety of formats and today there is an ever increasing amount of data available as shapefiles. Shapefiles are a type of geospatial data that are used in conjunction with GIS software, such as ESRI's ArcMap. This case study of European internet use and population began with a shapefile of the world's countries. This shapefile was then joined to a European internet use text file. The resulting attribute table was then exported as a .dbf file and imported into SAS® for statistical analysis. The SAS® results were inserted into the final map layout to increase map user's understanding. Lastly, a spatial autocorrelation analysis was conducted in ArcMap. This paper will discuss the methods, the results, and the conclusions reached.

METHODS

This case study utilized data from ESRI on European internet use in the form of a text file and country data in the form of a World Geodetic Survey 1984 datum shapefile (See Figure 1). Part one of this case study was conducted using ArcMap 9.3. Part two of the case study was conducted using SAS Enterprise Guide 4.2®. Part three was performed in ArcMap.

<Paper title>, continued

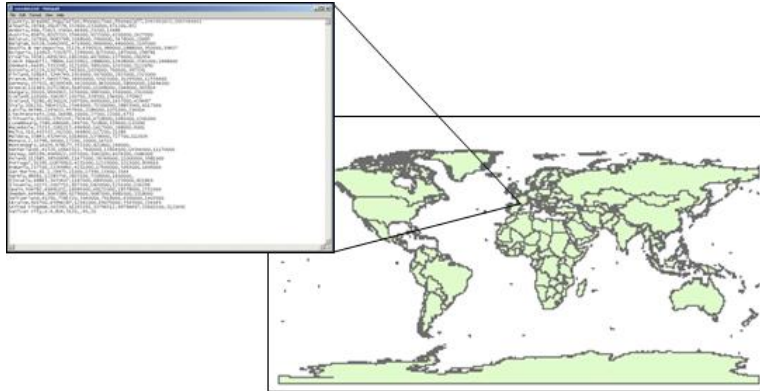


Figure 1. Data Sources

PART ONE

ArcMap was opened and a new empty map created. Next, the countries shapefile was added. The countries shapefile depicted all of the countries in the world. Since the case study was on European countries, not all of the countries needed to be displayed. One could have ran a Select by Attributes query where "Region" = 'Europe' and then created a layer by selection, which would have yielded a shapefile with only the European countries. However, this is not the most efficient method. Instead, one should add the European internet use text file into ArcMap. It is important to note that the text file data is not able to be mapped without manipulation. In order to display the internet use data, a join must be performed. In the join options, the type of join was specified as "join attributes from table" and country name was selected as the field to perform the join on. It is important to note that in order to use this type of join there must exist a unique identifier in both the shapefile and the text file that correspond to one another. Next, the "keep only matching records" option was selected. This option allows one to bypass creating a new layer with only European countries. Because the internet use text file only contains data about European countries, all other countries will be removed from the resulting table. The result of this join will be a shapefile with only European countries and it will now contain information about each countries cell phone use. This information can now be mapped showing country population in a graduated color format and internet use in a graduated symbol format (See Figure 2).

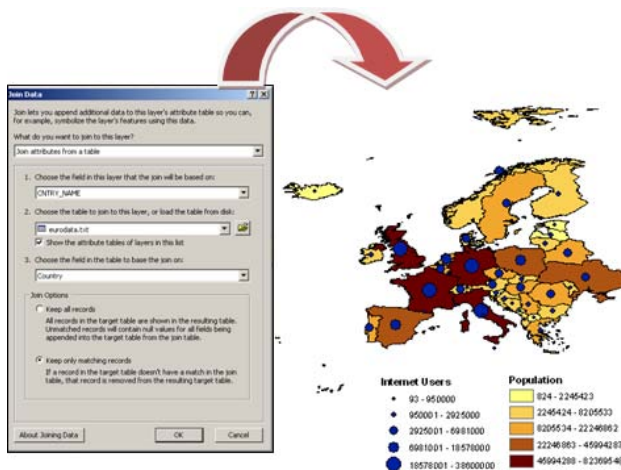


Figure 2. Using a Join to Create Internet Use and Population Map

<Paper title>, continued

PART TWO

The second part of this case study was to export the data to SAS® and run a correlation and construct a scatter plot. In order to achieve this, one must open the attribute table for the joined shapefile. Next, one must click options and export all records in a .dbf format (See Figure 3). It is important to note that one cannot simply click on the shapefile and click export data, as this would result in an exported shapefile and not a .dbf file.

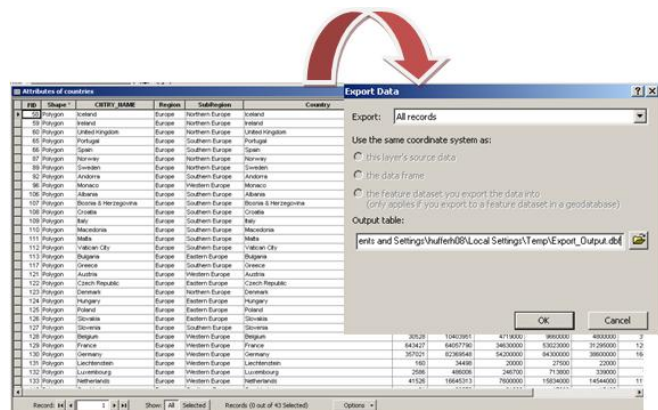


Figure 3. Exporting Data to .dbf Format

The .dbf file was then imported into SAS® Enterprise Guide 4.2® using File > Import data and by clicking through the prompts. Next a Pearson Correlation was conducted using Analyze> Multivariate> Correlation and selecting population and internet users as the variables. The scatter plot was constructed using Graph> Scatter plot and selecting population and internet users as the variables. Both the correlation output and scatter plot had a large range of display option such as colors and titles to make the output aesthetically pleasing. Although graphs can be constructed in ArcMap, more options are available in SAS® and it is recommended that graphs used in final layouts be constructed in that environment.

PART THREE

The resulting SAS® output was transferred into the ArcMap layout by copying and pasting. Although the SAS® output could be imported by saving the output as a PDF and inserting it as an object into ArcMap, it was more efficient to simply copy and paste. Lastly, a spatial autocorrelation test was conducted within ArcMap using Moran's I. This was accomplished by clicking on the Spatial Analyst toolbox and selecting Analyzing Patterns. To report these results on the final map layout, a text box was created which contained the resulting Moran's I (See Figure 4). It is important to note that while ArcMap has a wide array of spatial statistics tools, it is best to utilize SAS® for other statistical purposes, such as correlations and regressions.

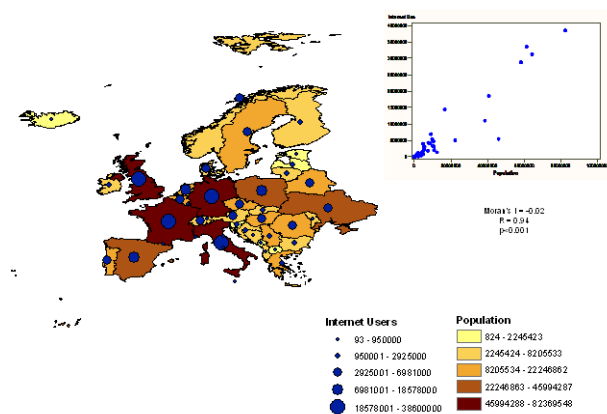


Figure 4. Final Map Layout

<Paper title>, continued

RESULTS

This case study of European internet use and population found a highly statically positive correlation between population and internet users ($r= 0.94$, $p <0.0001$). Also, this case study found there was no clustering of population or number of internet users at the country level (Moran's $I = -0.02$). It is speculated that if a smaller scale analysis were conducted, that there would exist some degree of clustering around major cities. All in all the results confirmed what was originally suspected.

CONCLUSION

This case study found that European internet use and population is positively correlated and that no spatial autocorrelation exists in internet use. It is suggested that future research examine internet use at a smaller scale in order to better capture spatial patterns. In terms of navigating between ArcMap and SAS®, this task was accomplished with ease by utilizing the .dbf format. Although it is easy to convert the shapefile data and work between the two software environments, it could be a daunting task for a SAS® user wanting to utilize data stored in a shapefile if they are unfamiliar with the appropriate methods. Hopefully, this paper will demonstrate to SAS® users that navigating between the two environments is easier than they think and might encourage them to use a wider array of available data.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Hillary Blair Huffer
University: East Carolina University
Address: Brewster A-427, Tenth Street
City, State ZIP: Greenville, NC 27834
Work Phone: 540-448-2099
E-mail: hufferh08@students.ecu.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.