A FOSS Web Tool for Spatial Regression Techniques and its Application to Explore Bike Sharing Usage Patterns

MGIS Capstone Proposal
Author: Spencer Bell
Presentation Outline

• Background
  - Bike Share
  - Spatial Regression

• Goals
  - Build a Web Tool to Explore Spatial Regression
  - Use Visual Analytics Techniques to present, explore, and disseminate results of Spatial Regression Analysis
  - Investigate and Explore Bike Share Hypothesis

• Approach & Timeline
Brief Overview of Bike Sharing

• Bike sharing has become increasingly popular in many large and medium sized urban areas
• This popularity is driven by the benefits of both the city and population
Bike Sharing Hypothesis

- The more roads with bike lanes around a given bike station, the greater the chance the station will have high bike usage
- Use the web tool to explore and prove/disprove hypothesis
Bike Data Sources

- Data sources for Indego Bike Share can be found on their website
- Variable Data is retrieved from ESRI’s Business Analyst extension

THE DATA

Each .csv file contains data for one quarter of the year. Each file contains the following data points:
- Trip ID
- Duration (sec) – Trip times listed are calculated by taking the check-out and check-in times and rounding down to the minute. For example: checkout time = 4:09:14 PM, return time = 4:15:49 PM (6 minute and 35 second trip). The dataset records the trip time as 6 minutes or 360 seconds.
- Start Date, Time
- End Date, Time
- Start Station ID, Lat/Lon – The station name corresponding to the station to the station ID can be found in the Station Table.
- End Station ID, Lat/Lon
- Bike ID
- Plan Duration – This shows the type of pass by number of days (i.e., thirty-day pass shows ‘30’)
- Trip Route Category – Round Trip or One Way
- Passholder Type – Walkup, Indego Flex or Indego30

2016 Esri Business Analyst Desktop Variable and Report List Summary

Data and report listing for over 10,000 demographic and business variables.
Overview of Spatial Regression Analysis

- Spatial regression analysis helps you answer the question WHY
- Allows the user to model, examine, and explore spatial relationships
- Assists in predictions and forecasts to help make decisions
Ordinary Least Square (OLS)

• Models a dependent variable in terms of its relationship to a set of explanatory variables
• Is a global regression model
Geographically Weighted Regression (GWR)

- Explore spatial non-stationarity
- Is a local regression model
- The equations incorporate the dependent and explanatory variables of features falling within the bandwidth of each target feature

\[
\hat{y}_i = \beta_0 + \sum_k \beta_k x_{ik} + \epsilon_i \\
\rightarrow \\
\hat{y}_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \epsilon_i
\]
Choosing Bandwidths for Localized Regression Models

- GWR constructs a separate equation for each target feature
- The features that fall within each target feature's separate equation is dependent upon the bandwidth
- Two types of bandwidth: Fixed distance or number of neighbors
Weight Matrix, Moran's I, and Spatial Autocorrelation
Alternative Exploration Techniques

• Spatial Lag Models
• Spatial Error Models
• Structured Equation Models [5]

Figure 1a BikeSEM Model
Introduction of R and Spatial Regression Packages
Presentation Outline

• Background
  - Bike Share
  - Spatial Regression

• Goals
  - Build a Web Tool to Explore Spatial Regression
  - Use Visual Analytics Techniques to present, explore, and disseminate results of Spatial Regression Analysis
  - Investigate and Explore Bike Share Hypothesis

• Approach & Timeline
Project Goals

- Create a web tool that can run spatial regression analysis on a wide range of datasets
- Use visual analytics techniques to present, explore, and disseminate results of spatial regression analysis
- Investigate the bike share hypothesis using the web tool
Web Tool Design

- Load data into web tool
- Ability to data wrangle
- Choose spatial regression
- Choose variables for analysis
- Select visualization method

- Explore data via webmap
- Add supporting charts
- Print webmap for presentations
- Combine multiple layers for improved visualizations
- Layering and brushing
Simplifying Code

```
install.packages("rgdal")
install.packages("maptools")
install.packages("spdep")
library(rgdal)
library(maptools)
library(spdep)

boston<-readOGR(dsn = "F:/RShortcourse", layer = "boston")
class(boston)
boston$LOGMEDV<-log(boston$MEDV)
coors<-coordinates(boston)
IDs<-row.names(as(boston, "data.frame"))
bost_kn1<-knn2nb(knearneigh(coors, k=1), row.names=IDs)
dist<-unlist(ndbdist(dist), coords)
summary(dist)
bost_kd1<-dbnearneigh(coors, d1=0, d2=3.973, row.names=IDs)
plot(boston)
plot(bost_kd1, coords, add=T)
bost_kd1_w<- nb2listw(bost_kd1)
moran.test(boston$LOGMEDV, listw= bost_kd1_w)
moran.plot(boston$LOGMEDV, bost_kd1_w, labels=as.character(boston$ID), xlab="Log of Median Home Value", ylab="Spatially Lagged Median Home Value")
title("Moran scatterplot")
bostlm<-lm(LOGMEDV~RM + LSTAT + CRIM +ZN + CHAS + DIS, data=boston)
summary(bostlm)
bostonlmresid<-residuals(bostlm)
lm.morantest(bostlm, bost_kd1_w)
moran.plot(bostlm$resid, bost_kd1_w)
lm.tests(bostlm, bost_kd1_w, test="all")
library(lmtest)
bptest(bostlm)
bostlag<-lagsarlm(LOGMEDV~RM + LSTAT + CRIM +ZN + CHAS + DIS, data=boston, listw=bost_kd1_w)
summary(bostlag)
bptest.sarlm(bostlag)
bosterr<-errrorsarlm(LOGMEDV~RM + LSTAT + CRIM +ZN + CHAS + DIS, data=boston, listw=bost_kd1_w)
summary(bosterr)
```

Packages can be pre installed with shiny. Libraries can be ready without the need for the user to turn them on.

This equation that is repeated multiple times can be assigned as "Model 1" in the web tool.

Figure 3. Choropleth maps of α parameter estimates and t-values by census tract for the GWR of median home value using an equal step data classification and a sequential no-lux colour scheme for each map.

Figure 4. Choropleth maps of α parameter estimates and t-values by census tract for the GWR of median home value. In the parameter estimate map, a modified standard deviation data classification and a diverging colour scheme is used whereas in the t-value map, an eigenvector data classification based on commonly accepted significance thresholds and a sequential no-lux colour scheme is used.
Matthews and Yang [3] Example of Isolines
Presentation Outline

• Background
  - Bike Share
  - Spatial Regression

• Goals
  - Build a Web Tool to Explore Spatial Regression
  - Use Visual Analytics Techniques to present, explore, and disseminate results of Spatial Regression Analysis
  - Investigate and Explore Bike Share Hypothesis

• Approach & Timeline
Displaying Results on an Interactive Map

Static Map in Base R vs Interactive Map in R + Leaflet

- Interactive map allows for layering of regression outputs
- The ability to tab between models will help determine which model is the most accurate
Explanation of Shiny as a Web Framework

- Makes building interactive web applications with R possible through “reactivity”
- Htmlwidgets package allows for HTML, CSS, Javascript to be added
- Shinyapps.io is a server where shiny applications can be hosted
Advantages of Web Tool over Existing Software Solutions

- ArcGIS: Very expensive, not as much customization as other GWR tools listed
- Using base R: Steep learning curve to understanding the language. Without shiny it can not be shared easily on the web
- Geoda: Unable to share results on the web
- GWR4: Great tool for setting up GWR model but there is no visual output and it can not be run on the web
Who Should Use this Web Tool?

• Students, Teachers, and Researchers

• Allow decision makers in low budget areas to make informed decisions based on solid data analysis
Project Timeline

January-February
- Build First Prototype of Web Tool in R/Shiny

March
- Have Researchers Test Web Tool and Offer Feedback

April-May
- Test Hypothesis using Web Tool

July 4-7
- Disseminate Web Tool using Shinyapps.io

August 14-19
- Finalize Web Tool Based on Feedback
Literature List


Thank You