Mapping Student Arrivals and Departures: Using Graphical Information Software (GIS) to Examine Admission, Enrollment, and Retention Trends

Christopher Neasbitt,

Programmer Analyst Associate (cjneasbi@valdosta.edu)

Kristina M. Cragg, Ph.D.,

Assistant to the President for Strategic Research & Analysis
(kmcragg@valdosta.edu)

Amir Atabekov,

Student Assistant

Valdosta State University – Strategic Research & Analysis

Issues for Institutional Researchers

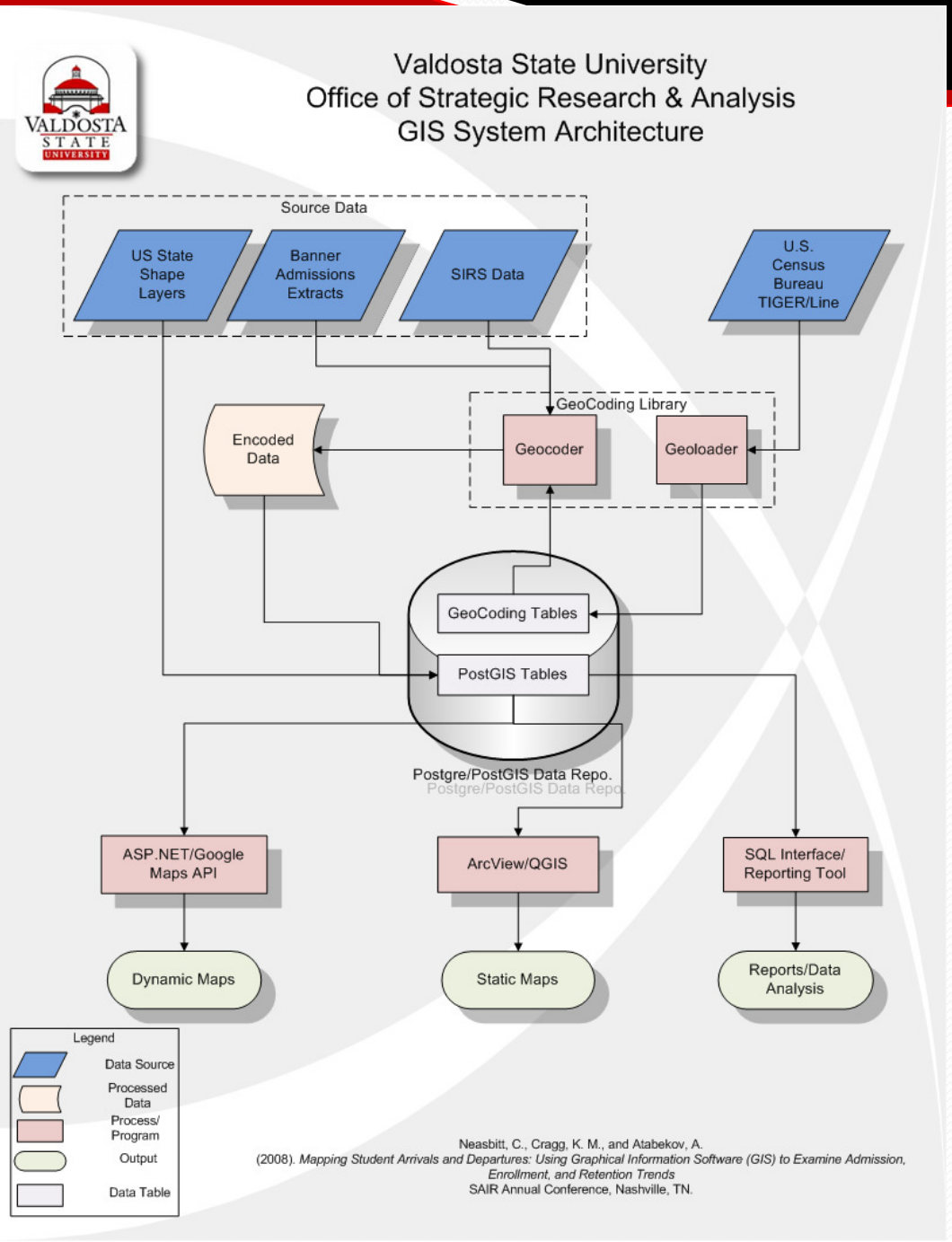
- You are interesting in learning about how GIS can be used in IR offices.
- You are looking for simple and visual ways of showing data patterns using different techniques.
- You are getting bored in your office and GIS sounds like something you'd like to learn more about.
- This was the best session during this timeslot.

By the End of this Presentation …

- … We Will Have Discussed:
  - Motivation for the Study & Research Questions
  - Research Questions
  - Data & Sample
  - Variables
  - Methods
  - Limitations
  - Results
  - Possible Implications
  - Questions & Comments

Motivation for the Study & Research Questions

- Most problems have a geographic element. How can we use that element in data analysis?
- GIS has been used in industry as a tool to analyze problems with geographic elements.
  - Ex. ATM Placement in urban areas
- GIS can be used in the university setting to better understand and analyze issues relating to student admissions, enrollment, and retention.
  - How have student origins (admissions, enrollment, retention) changes over time?



Neasbitt, C., Cragg, K. M., and Atabekov, A.
(2008). *Mapping Student Arrivals and Departures: Using Graphical Information Software (GIS) to Examine Admission, Enrollment, and Retention Trends*
SAIR Annual Conference, Nashville, TN.

Data

- SIRS (Student Information Record System)
  - Student enrollment information dataset
- Banner
  - Admissions data extracts
- U.S. Census Bureau
  - TIGER/Line
    - Fixed position ascii files
    - shapefiles
- U.S. State shape layers

Variables

- Demographic (SIRS)
  - Ethnicity
  - Gender
  - County
- Academic
  - GPA
  - SAT
- Admissions
  - Enrolled
  - Accepted
  - Denied
- Geographic (U.S. Census)
  - Addresses (includes Zip codes)
  - Other data elements are available, but not used at this time.
- Reference Data (U.S. State Shape Layers)
  - State boundaries
  - County boundaries

Methods

- Architecture Overview
  - Source Data
    - Individual Student Records
    - City, County, State Boundaries
  - Data Repository
    - PostgreSQL (http://www.postgresql.org/)
    - PostGIS (http://postgis.refractions.net/)

Methods

- Architecture Overview (cont.)
  - Data Rendering Layer
    - Google Maps API (http://code.google.com/apis/maps/documentation/)
    - ArcDesktop (http://www.esri.com/software/arcgis/arcims/index.html)
    - SQL Interface (Toad, RazorSql, Aqua Data Studio, etc.)
  - Output Layer
    - Static Maps
    - Dynamic Maps
    - Text Reports

Geocoding Library

- Created our own geocoding library to transform street addresses into long, lat coordinates
- Motivation
  - Commercial geocoding services charge by the record, budget constraints make these services an issue
  - Free geocoding services have limitations on the number of records that can be encoded in a period,
    - Google: 15,000 records per day
    - Yahoo: 5,000 records per day

Geocoding Library

- Operation
  - Geoloading
    - Converts US Census Bureau TIGER/Line data into a searchable, vendor neutral, table schema
  - Geocoding
    - Parses a text address into pieces
    - Searches the database for those corresponding pieces
    - On a direct address match the library returns the lat, long value
    - On a indirect address match the library calculates an approximate lat, long based on the available data points using linear interpolation
    - On no match if a zip code is available the lat, long of the zip code centroid is returned
  - Operational accuracy: 80%

Spatial Data Repository

- Many vendors offer spatial data storage
  - Oracle Spatial
  - Microsoft SQL Server Spatial
  - ESRI ArcSDE
  - MySQL Spatial Extensions
  - PostgreSQL PostGIS

Spatial Data Repository

- We chose to store our spatial data with in a PostgreSQL DB using PostGIS for several reasons.
  - Cost effective
  - Availability of documentation
  - Interoperability with other software suites
  - Open Source
  - Standards Compliance
    - OGC Compliant

PostGIS Spatial Queries

- PostGIS provides data types for storing spatial data and operations for manipulating that data
  - Operation Types (http://postgis.refractions.net/documentation/manual-1.3/ch06.html)
    - Relationship Functions
      - Distance, intersection, contains, etc.
    - Processing Functions
      - Area, Length, Centroid, etc.

PostGIS Spatial Queries

- Query Example
  - Find the names of all students from Lowndes county

```
Select
  all_sirs.last, all_sirs.first
From
  all_sirs, gacounty04
where
   ST_Within(all_sirs.the_geom,
  gacounty04.the_geom)
   and gacounty04.county = 'Lowndes'
```

Static Mapping

- Numerous tools available for visually displaying PostGIS layers, free and commercial.
  - QGIS (http://www.qgis.org/)
  - uDig (http://udig.refractions.net/)
  - zigGIS (http://pub.obtusesoft.com/)
    - ArcView plugin
- We have used both QGIS and ArcView with zigGIS v1.2.

Dynamic Mapping

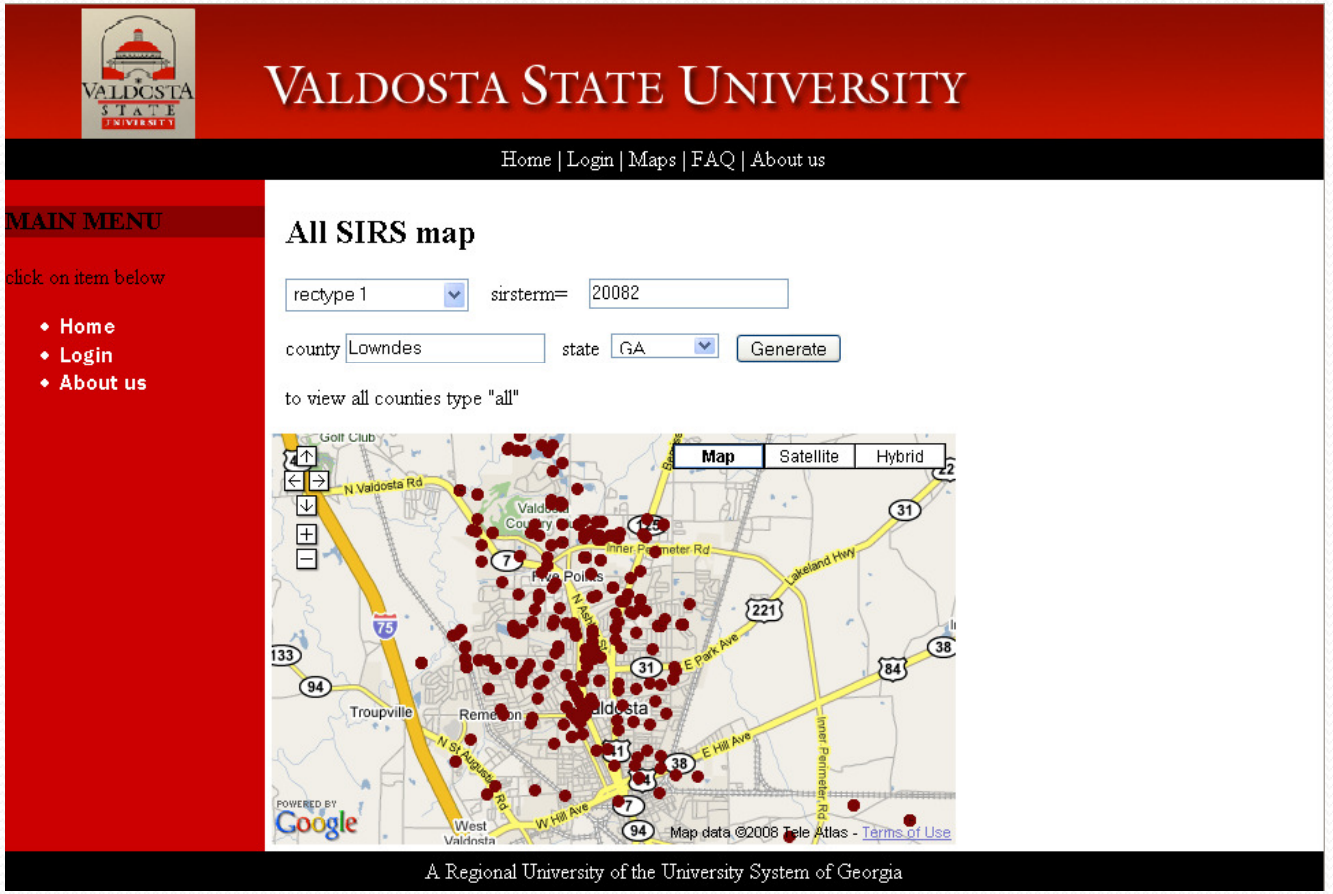
- Pitfall
  - ArcIMS, ArcSDE and PostgreSQL/PostGIS
    - As of version 9.3 ArcSDE has supported PostgreSQL as database repository
    - Non-ESRI research provided techniques that might allow ArcSDE 9.3 to use the PostGIS spatial format
    - ArcIMS would not render the PostGIS layers through ArcSDE
    - Abandoned ArcIMS and ArcSDE for dynamic mapping purposes.

Dynamic Mapping

- Second Attempt
  - Google Maps API
    - Use ASP.NET to generate xml docs from PostGIS data.
    - Loaded the xml docs using the Google Maps API to render the data over the Google Maps Viewer
  - Pros
    - Easy to generate XML
    - Simple, Free API
  - Cons
    - Rendering Performance

Dynamic Mapping

- By providing dynamically queried maps on the web you can increase the utility of the data to all stakeholders



Limitations

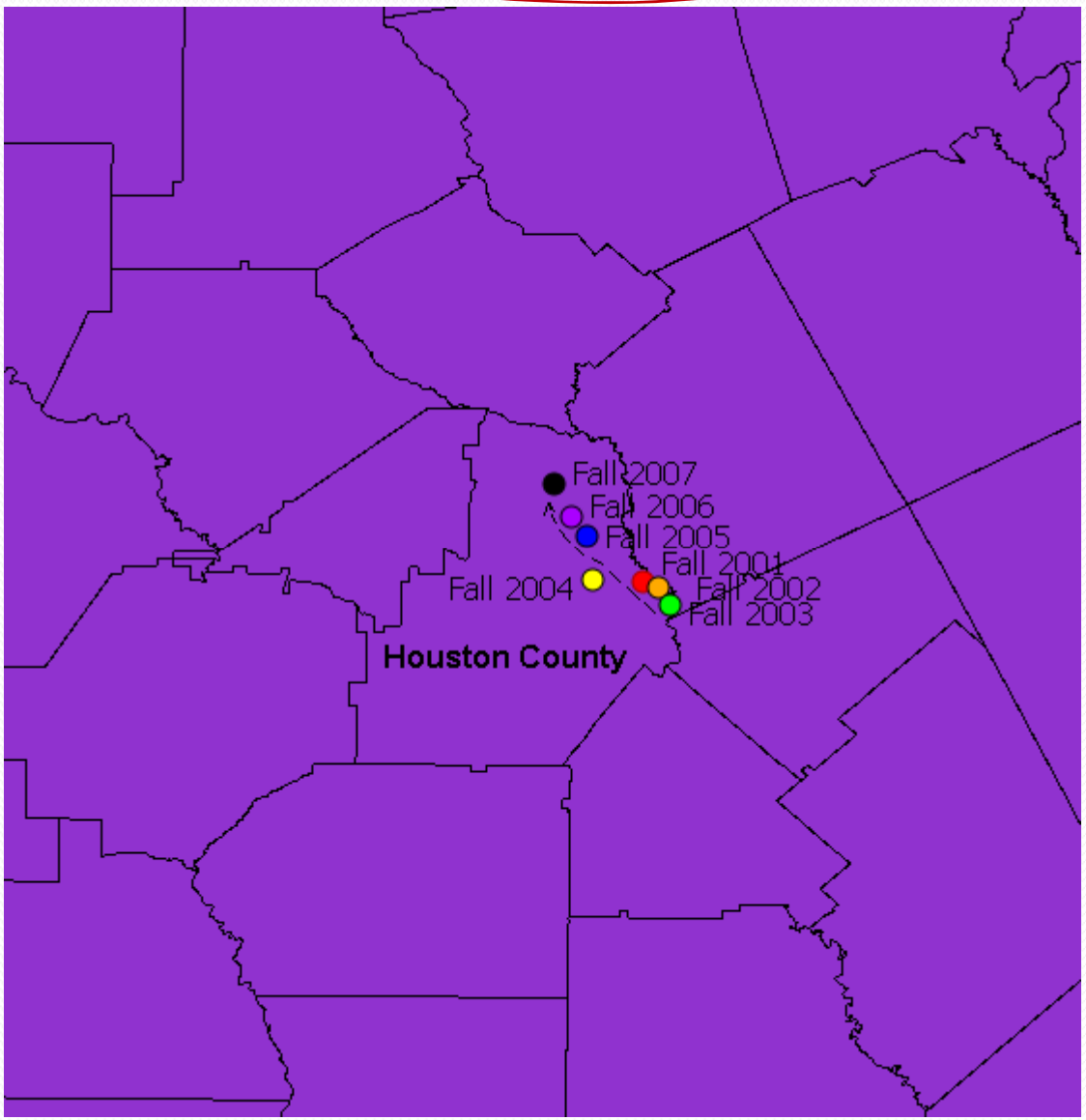
- Funds
  - Commercial software suites can be pricey
    - Open source alternatives can provide some relief
- Time
  - Time required to develop the geocoding library or learning to use commercial alternatives.
  - Time and effort in integrating multiple different datasets and software suites
- Accuracy
  - Accuracy of encoded addresses.
  - Accuracy in transforming addresses into lat, long coordinates.

Limitations

- Relevance
  - To analyze and question using GIS the problem set must have a geographic component
  - The spatial relationship to within a problem set might not be statistically significant.
- Skilled expertise in this area is not common
  - Spatial statistics is a separate field

Results

- How have student origins changed over time?
  - Calculating the spatial mean student origins for consecutive Fall semesters helps illustrate student origin shifts.



Conclusion for IR Practitioners

- Visual representation of complex data to senior leadership or key stakeholders.
- Increase the value of data by looking at same data in a different way.
- Patterns may be apparent only when looking at data using GIS.
- IR practitioners could do GIS analysis … there are "simple" tools available
  - ESRI Arcdesktop suite software applicable to all skill levels.

Thank You

Questions & Comments