Dynamic and Interactive Statistical Graphics for Spatially Referenced Data

Jürgen Symanzik
George Mason University, Fairfax, VA

*e-mail:* symanzik@galaxy.gmu.edu
WWW: http://www.galaxy.gmu.edu/~symanzik
Contents

- The ArcView/XGobi/XploRe Environment
- EPA’s Cumulative Exposure Web Page
ArcView/XGobi/XploRe: Collaborators

- Di Cook, Nicholas Lewin-Koh, Jim Majure: Iowa State University, Ames, IA

- Sigbert Klinke, Swetlana Schmelzer, Thomas Koetter: Humboldt University, Berlin, Germany

- http://www.public.iastate.edu/~arcview-xgobi/homepage.html
ArcView/XGobi/XploRe: Main Idea

- Link three kinds of software packages:
  - ArcView: Geographic Information System (GIS)
  - XGobi: dynamic statistical graphics program
  - XploRe: statistical computing environment
Remote Procedure Calls (RPCs):

- Process on the local system (client) invokes a procedure on a remote system (server)
- Request = client’s desire to execute a particular remote procedure
- Response = result produced by the remote procedure

ArcView, XGobi, XploRe: server & client
Features of ArcView/XGobi-Link

- Multivariate Link
- CDF Link
- Variogram-cloud Link
- Spatially Lagged Scatterplot Link
- Multivariate Variogram-Cloud Link
Multivariate Link

XGobi

ArcView
CDF Link

XGobi

ArcView
Variogram-Cloud Link

XGobi

ArcView
Spatially Lagged Scatterplot Link

XGobi
Multivariate Variogram-Cloud Link

XGobi
Addition of XploRe

- Exchange of Data and Commands
- External Smoother
- Dynamic Linked Brushing
- Capabilities for Spatial Data Analysis

XGobi Menus

- Clone XGobi
  - Smooth ...
  - Subset Data ...
  - Jitter ...
  - Parallel Coord Plot ...
  - Variable List ...
  - Case List ...
- Start XploRe ...
- Stop XploRe ...
- Launch Missing Data XGobi ...
- Impute Missing Values ...

Read ...
- Save (extend current file set) ...
- Save (create new file set) ...
- Print ...
- XploRe (pass variables) ...
- XploRe (pass projection) ...
- Quit (Q)

- Mean
- Median
- Nadaraya–Watson
- Spline
- LINEAR
- SYMMETRIZED KNN
- KNN
- LOWESS
- LOCAL POLYNOMIAL
- NEURAL NETWORK
- ISOTONIC
Dynamic Linked Brushing

XploRe

XGobi
Spatial Data Analysis in XploRe

Mean and Mean Summaries

Variogram
Full Scenario
Current Research

- Continuation ArcView <-> XploRe Link
- Spatial Statistics in XploRe
- Inclusion of the Virtual Reality GIS ViRGiS
- Point <-> Area Brushing
Collaborators

- David Wong, Dan Carr, Jingfang Wang: George Mason University, Fairfax, VA

- Dan Axelrad, Tracey Woodruff: EPA, Office of Policy, Washington, D.C.

- http://www.epa.gov/CumulativeExposure
- http://www.galaxy.gmu.edu/~symanzik/gpl/CEPstart
Cumulative Exposure Project: Idea

- Conducted by EPA’s Office of Policy
- Collection of analyses, addressing multiple pollutants from multiple sites
- National analyses of
  - Air Toxics (Outdoor Concentrations)
  - Food Contaminants (Exposures)
  - Drinking Water Contaminants (Exposures)
- Pilot Community-level study: Greenpoint/Williamsburg, NYC
Air Toxics

- 188 Hazardous Air Pollutants (HAPs) in Clean Air Act
- Limited availability of air toxics monitoring data
- Atmospheric dispersion modeling provides understanding of concentrations of toxics in outdoor air
Scope of Modeling

- Modeled concentrations for each census tract in continental US (> 60,000)
- 148 HAPs modeled
- Includes stationary and mobile sources of air toxics emissions
- Uncertainty bounds derived from model-monitor comparisons
WWW-based Access of HAP Data

**Goals:**
- Concise display
- Easy access
- Understandable to nonstatistical audience

**Solution:**
- WWW
- GPL
- Micromaps
Graphics Production Library (GPL)

- Tool for creation and modification of statistical graphics on the WWW
- Follows guidelines of modern statistical graphics
- JAVA-based
- Works with Netscape and Internet Explorer
History of the GPL

- Main developer: Dan Rope
- New commercial product based on GPL currently developed by Dan Rope
- Upgrade and new features added to GPL through BLS and contractors

The Barley Data

From Visualizing Data by W.S. Cleveland

<table>
<thead>
<tr>
<th>Variety</th>
<th>1931</th>
<th>1932</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cookston:</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>Manchuria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glabron</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Svansota</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velvet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trebi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 462</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peatland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 475</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin No. 38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duluth:</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>Manchuria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glabron</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Svansota</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velvet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trebi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 462</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peatland</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Yield (bushels/acre)

Show Data

Pan And Zoom
Micromaps

- Link of row-labeled univariate (or multivariate) statistical summaries to corresponding geographical region
- Focus on statistical display and not on maps
- S-Plus functions (available at ftp://galaxy.gmu.edu/pub/dcarr/newsletter/micromap/)
History of Micromaps

- First presented at 1995 American Statistical Association’s annual meeting (Olsen, Carr, Courbois, Pierson)

- Main references:
  - Carr, Olsen, Courbois, Pierson, Carr (1998) Linked Micromap Plots …., SCSG, Vol. 9, No.1
Bringing All Together

- Hierarchical Clickable Micromaps in the GPL Environment for the Display of Hazardous Air Pollutants Data
- GPL extended with micromaps
- Selection of HAPs
- Clickable maps
- Hierarchy of maps
Realization

- Automatic one-time creation of micromaps
- C code used to create follow-up data sets and Web pages
- Political vs. scientific design and layout
- Availability: hopefully February 1999
EPA's Cumulative Exposure Project (CEP) is examining how much toxic contamination Americans are exposed to cumulatively through air, food, and drinking water. The study will estimate exposure levels for different communities and demographic groups nationwide, and identify which types of communities and demographic groups appear to have the highest exposures. The Cumulative Exposure Project is being conducted by EPA's Office of Policy.

About the Project

What's New?

- Modeled 1990 air toxics concentrations are now available.

[Cumulative Exposure Project Home | EPA Home | Search | What's new]

http://www.epa.gov/CumulativeExposure/index.htm

Last updated 12/21/98

Please send comments about this page to excelrad.daniel@epa.gov
### United States – Summary

1990 Modeled Concentration of HAP benzene in μg/m³

<table>
<thead>
<tr>
<th>State</th>
<th>Tracts</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>P25</th>
<th>P75</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>60803</td>
<td>2.07</td>
<td>1.61</td>
<td>0.48</td>
<td>0.50</td>
<td>2.69</td>
<td>78.80</td>
</tr>
<tr>
<td>Alabama</td>
<td>1062</td>
<td>1.40</td>
<td>1.02</td>
<td>0.55</td>
<td>0.72</td>
<td>1.77</td>
<td>14.38</td>
</tr>
<tr>
<td>Arizona</td>
<td>810</td>
<td>1.90</td>
<td>1.84</td>
<td>0.48</td>
<td>0.97</td>
<td>2.69</td>
<td>12.58</td>
</tr>
<tr>
<td>Arkansas</td>
<td>593</td>
<td>1.03</td>
<td>0.73</td>
<td>0.50</td>
<td>0.59</td>
<td>1.10</td>
<td>12.26</td>
</tr>
<tr>
<td>California</td>
<td>5858</td>
<td>3.03</td>
<td>2.73</td>
<td>0.49</td>
<td>1.72</td>
<td>4.16</td>
<td>14.62</td>
</tr>
<tr>
<td>Colorado</td>
<td>979</td>
<td>2.02</td>
<td>1.66</td>
<td>0.48</td>
<td>0.80</td>
<td>2.62</td>
<td>17.61</td>
</tr>
<tr>
<td>Connecticut</td>
<td>834</td>
<td>2.12</td>
<td>1.80</td>
<td>0.69</td>
<td>1.31</td>
<td>2.64</td>
<td>6.69</td>
</tr>
<tr>
<td>Delaware</td>
<td>175</td>
<td>2.87</td>
<td>2.50</td>
<td>0.75</td>
<td>1.28</td>
<td>3.82</td>
<td>20.61</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>192</td>
<td>4.06</td>
<td>3.51</td>
<td>2.01</td>
<td>2.87</td>
<td>4.41</td>
<td>19.50</td>
</tr>
<tr>
<td>Florida</td>
<td>2448</td>
<td>1.61</td>
<td>1.43</td>
<td>0.49</td>
<td>0.68</td>
<td>2.10</td>
<td>8.71</td>
</tr>
<tr>
<td>Georgia</td>
<td>1470</td>
<td>1.50</td>
<td>1.13</td>
<td>0.56</td>
<td>0.79</td>
<td>1.97</td>
<td>6.56</td>
</tr>
<tr>
<td>Idaho</td>
<td>269</td>
<td>1.36</td>
<td>1.08</td>
<td>0.56</td>
<td>0.75</td>
<td>1.75</td>
<td>5.01</td>
</tr>
<tr>
<td>Illinois</td>
<td>2841</td>
<td>1.94</td>
<td>1.76</td>
<td>0.53</td>
<td>1.62</td>
<td>2.52</td>
<td>16.95</td>
</tr>
<tr>
<td>Indiana</td>
<td>1383</td>
<td>1.52</td>
<td>1.27</td>
<td>0.56</td>
<td>0.78</td>
<td>2.01</td>
<td>6.80</td>
</tr>
<tr>
<td>Iowa</td>
<td>784</td>
<td>0.94</td>
<td>0.74</td>
<td>0.50</td>
<td>0.55</td>
<td>1.21</td>
<td>4.50</td>
</tr>
<tr>
<td>Kansas</td>
<td>884</td>
<td>1.05</td>
<td>0.85</td>
<td>0.48</td>
<td>0.56</td>
<td>1.47</td>
<td>4.15</td>
</tr>
<tr>
<td>Kentucky</td>
<td>997</td>
<td>1.25</td>
<td>0.87</td>
<td>0.56</td>
<td>0.67</td>
<td>1.69</td>
<td>11.56</td>
</tr>
<tr>
<td>Louisiana</td>
<td>1105</td>
<td>2.26</td>
<td>1.87</td>
<td>0.50</td>
<td>0.92</td>
<td>3.02</td>
<td>35.86</td>
</tr>
<tr>
<td>Maine</td>
<td>384</td>
<td>1.06</td>
<td>0.74</td>
<td>0.49</td>
<td>0.61</td>
<td>1.01</td>
<td>6.55</td>
</tr>
<tr>
<td>Maryland</td>
<td>1151</td>
<td>2.41</td>
<td>2.03</td>
<td>0.56</td>
<td>1.37</td>
<td>2.92</td>
<td>16.95</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>1331</td>
<td>2.02</td>
<td>1.74</td>
<td>0.59</td>
<td>1.18</td>
<td>2.51</td>
<td>21.43</td>
</tr>
</tbody>
</table>
Rhode Island – Summary

1990 Modeled Concentration of HAP benzene in μg/m³

<table>
<thead>
<tr>
<th>County</th>
<th>Tracts</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>P25</th>
<th>P75</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bristol County</td>
<td>12</td>
<td>1.60</td>
<td>1.24</td>
<td>1.11</td>
<td>1.16</td>
<td>1.82</td>
<td>3.50</td>
</tr>
<tr>
<td>Kent County</td>
<td>36</td>
<td>1.46</td>
<td>1.33</td>
<td>0.94</td>
<td>1.17</td>
<td>1.79</td>
<td>2.56</td>
</tr>
<tr>
<td>Newport County</td>
<td>23</td>
<td>1.92</td>
<td>1.37</td>
<td>0.97</td>
<td>1.18</td>
<td>2.42</td>
<td>5.98</td>
</tr>
<tr>
<td>Providence County</td>
<td>138</td>
<td>2.20</td>
<td>2.13</td>
<td>0.85</td>
<td>1.65</td>
<td>2.63</td>
<td>4.83</td>
</tr>
<tr>
<td>Washington County</td>
<td>25</td>
<td>1.18</td>
<td>1.00</td>
<td>0.73</td>
<td>0.84</td>
<td>1.17</td>
<td>3.52</td>
</tr>
</tbody>
</table>

Tracts represents the number of census tracts in 1990 for each county.

Mean represents the arithmetic mean of all 1990 Modeled Concentrations for each county and Median the median. Min represents the minimum 1990 Modeled Concentration for each county. P25 the 25th percentile, P75 the 75th percentile, and Max the maximum. All concentrations are in micrograms per cubic meter (μg/m³).

There are 5 counties in this state.

[Cumulative Exposure Project Home] [EPA Home] [Search] [What's new]

The Cumulative Exposure Project – Data Table

Please send comments about this page to ace@epa.gov
### Rhode Island – Bristol_County

**1990 Modeled Concentration of HAP benzene in μg/m³**

<table>
<thead>
<tr>
<th>Tract Code</th>
<th>Concentration</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>30100</td>
<td>1.2</td>
<td>0.8</td>
<td>4.6</td>
</tr>
<tr>
<td>30200</td>
<td>1.4</td>
<td>0.7</td>
<td>5.1</td>
</tr>
<tr>
<td>30300</td>
<td>1.2</td>
<td>0.6</td>
<td>4.5</td>
</tr>
<tr>
<td>30400</td>
<td>1.1</td>
<td>0.6</td>
<td>4.2</td>
</tr>
<tr>
<td>30500</td>
<td>2.5</td>
<td>1.2</td>
<td>9.4</td>
</tr>
<tr>
<td>30600</td>
<td>1.3</td>
<td>0.7</td>
<td>5.1</td>
</tr>
<tr>
<td>30601</td>
<td>2.3</td>
<td>1.1</td>
<td>8.7</td>
</tr>
<tr>
<td>30700</td>
<td>3.5</td>
<td>1.7</td>
<td>13.2</td>
</tr>
<tr>
<td>30800</td>
<td>1.2</td>
<td>0.6</td>
<td>4.7</td>
</tr>
<tr>
<td>30801</td>
<td>1.1</td>
<td>0.6</td>
<td>4.2</td>
</tr>
<tr>
<td>30900</td>
<td>1.2</td>
<td>0.6</td>
<td>4.5</td>
</tr>
<tr>
<td>30901</td>
<td>1.4</td>
<td>0.6</td>
<td>4.3</td>
</tr>
</tbody>
</table>

Concentration represents the 1990 Modeled Concentration. The interval (Lower Bound, Upper Bound) represents the 90% Confidence Interval. All concentrations are in micrograms per cubic meter [μg/m³].

There are 12 census tracts in this county.

---

*Cumulative Exposure Project Home | EPA Home | Search | What's new*
*The Cumulative Exposure Project – Data Table*
*Please send comments about this page to eared.daniel@epa.gov*