

Visualizing Federal Statistics (enabling geo-visual analytics)

Dr. Alan M. MacEachren
GeoVISTA Center
Geography, Penn State

collaborators at PSU: Dr. Cynthia Brewer, Stephen Crawford, Dr. Mark Gahegan, Brenda Kluhsman, Dr. Gene Lengerich, Dr. James Macgill, Ann Ward,

collaborators elsewhere: Dr. Luc Anselin (IL), Dr. Dan Carr (George Mason), Dr. David Scott (Rice)


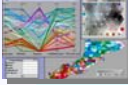
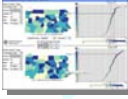
maceachren@psu.edu

Visual Analytics


- ♦ *Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces.*
- ♦ People use visual analytics tools and techniques to
 - ♦ **synthesize information** and **derive insight** from massive, dynamic, ambiguous, and often conflicting data;
 - ♦ **detect the expected** and **discover the unexpected**;
 - ♦ **provide timely, defensible, understandable assessments**; and
 - ♦ **communicate assessment effectively for action.**

From Illuminating the Path: Research and Development Agenda for Visual Analytics, National Visualization and Analytics Center – 2005 – <http://nvac.pnl.gov/agenda.stm>

Outline: highlights of 3 projects

- ♦ NSF: *Digital Government* – Quality Graphics for Federal Statistical Summaries – with Census, NCI and FedStats financial support, then an NCI follow up contract to develop the Exploratory Spatial-Temporal Analysis Toolkit (ESTAT)
- ♦ NCI: grant (Geovisualization and Spatial Analysis of Cancer Data)
- ♦ CDC: GIS/Atlas for Comprehensive Cancer Control



NSF Collaborative Research: Quality graphics for federal statistical summaries

Alan M. MacEachren
GeoVISTA Center & Dept. of Geography, Penn State University
www.geovista.psu.edu; maceachren@psu.edu

Daniel B. Carr
Dept. of Statistics, George Mason University
dcarr@gmu.edu


David Scott
Dept. of Statistics, Rice University
scotttdw@stat.rice.edu

NSF- EIA-9983451, 9983459, 9983461



Objective and Approach

- ♦ **Objective:** to develop strategies and methods that support: (a) exploration of data by federal agency users; (b) decision making by policy analysts; and (c) communication of statistical summaries to the public.
- ♦ **Approach:** development innovative statistical graphics grounded in perceptual and cognitive principles **and** iteratively assesses representation usability.



Selected accomplishments

- ♦ Developing Exploratory Spatial Data Analysis (ESDA) methods:
 - ♦ GeoVISTA *Studio* extensions
 - ♦ ESTAT, v 1.0

GeoVISTA Studio: Alan MacEachren, PI, EIA-9983451
 additional support from: NSF (DG 9978052, 9983445), NCI-CA95949, USGS, ARDA, Penn State

A Java, component-based, visual programming environment for development of applications and applets that integrate visual, statistical, & computational methods for (geospatial) data exploration, analysis & knowledge construction

side bar: multivariate EDA with Parallel Coordinate Plots (PCP)

breast cancer mortality rate (MD/100,000) and cervical cancer mortality rate (MD/100,000) for McKean, Lycoming, and Mifflin counties.

Example: Cancer mortality and covariates

white female lung cancer mortality rates purple = very high, green = very low

health service areas

per capita income 93

ESTAT: Exploratory Spatial-Temporal Analysis Tool
 developed under contract for NCI
 lead developer: Jin Chen

GEOVISUALIZATION AND SPATIAL ANALYSIS OF CANCER DATA
 NCI PROGRAM FOR GEOGRAPHIC-BASED RESEARCH IN CANCER CONTROL AND EPIDEMIOLOGY

Issues:

- Few forms of cancer are uniformly distributed across a population in either space or time.
- Geographic variation in cancer mortality/incidence is known to be associated with geographic variation in: risk factors, screening behaviors, health care access/use, genetic predisposition, and occupational hazards.
- However, current analytical methods/tools available to researchers limit and bias understanding of the geographic observations that prompt investigations.

93-97 cervical cancer rate

93-97 breast cancer rate

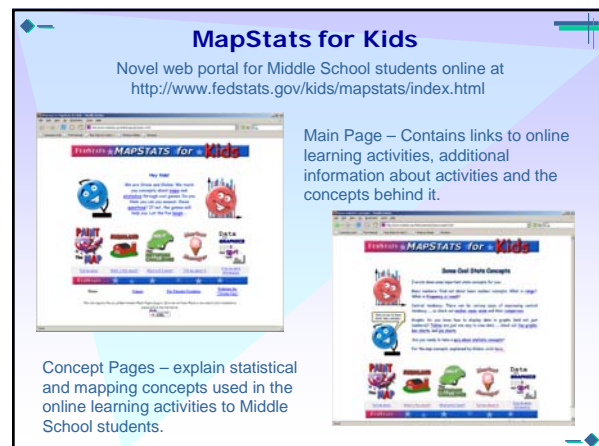
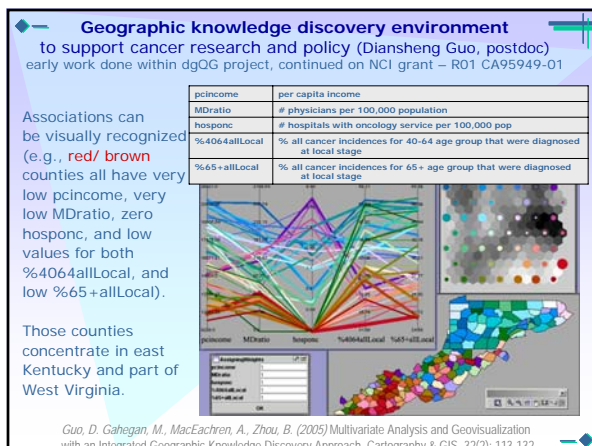
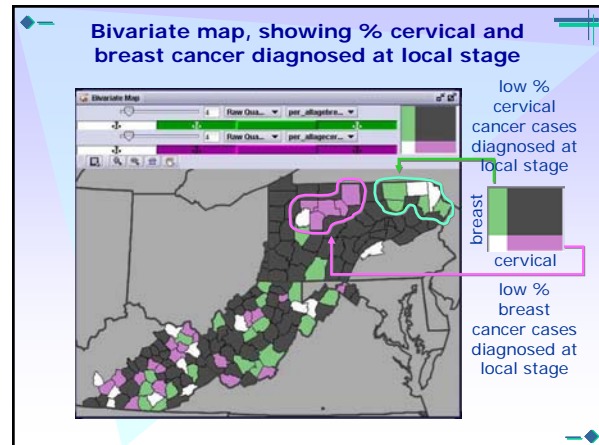
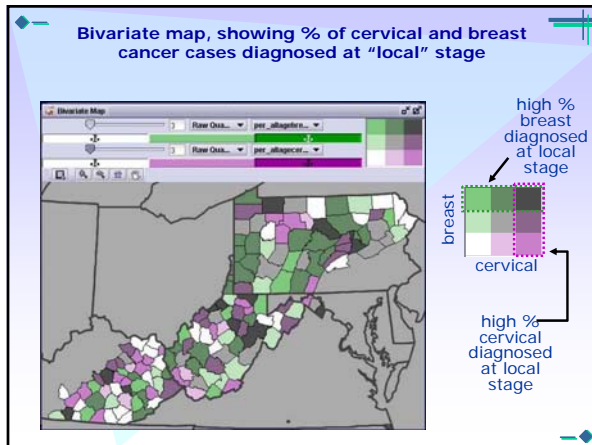
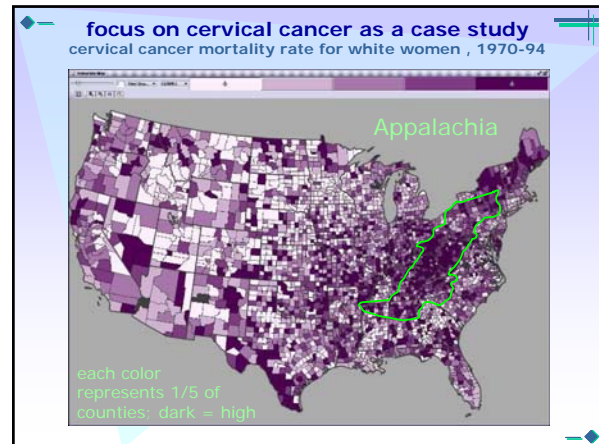
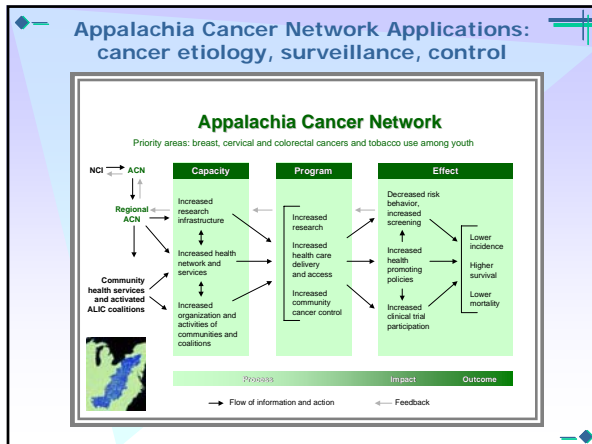
GEOVISUALIZATION AND SPATIAL ANALYSIS OF CANCER DATA
 NCI PROGRAM FOR GEOGRAPHIC-BASED RESEARCH IN CANCER CONTROL AND EPIDEMIOLOGY
www.geovista.psu.edu/grants/nci-esda

Objectives:

- to design and implement an integrated suite of visual-statistical-computational methods/tools for application to cancer research, surveillance, and control
- through application of these methods and tools – to improve methodologies for exploratory research in cancer epidemiology and policy decision-making
- methods/tools that (a) are usable by, and meaningful to, cancer professionals and (b) can be used to address important research and policy questions

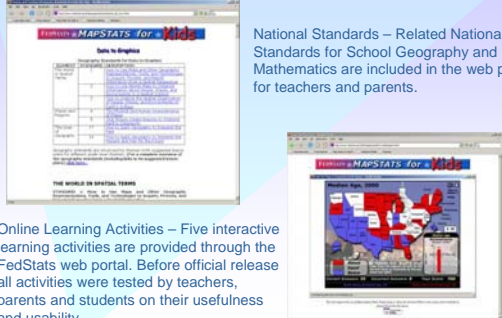
we approach these conceptually and technically →

PI: Dr. Alan MacEachren, with Dr. Eugene Lengerich, Dr. Mark Gahegan, Dr. James MacCall, Dr. Luc Anselin (IL), and Ann Ward



MapStats for Kids

Novel web portal for Middle School students online at <http://www.fedstats.gov/kids/mapstats/index.html>



National Standards – Related National Standards for School Geography and Mathematics are included in the web portal for teachers and parents.

Online Learning Activities – Five interactive learning activities are provided through the FedStats web portal. Before official release all activities were tested by teachers, parents and students on their usefulness and usability.

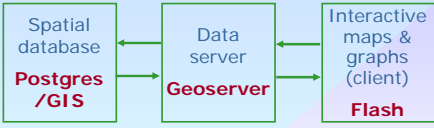
GIS/Atlas Cancer Research

Penn State Cancer Institute Department of Health Evaluation Sciences

- ♦ Goal: Develop, use and disseminate a “model” GIS/Atlas for State Comprehensive Cancer Control – thus, to create an geo-information technology “template” that states can adopt and modify
- ♦ Focus: PA Cancer registry data by county
- ♦ Kind of “GIS/Atlas”: interactive, web map service
- ♦ Target audience(s): Cancer control directors and staff, public health practitioners, interested public

PI: Gene Lengetch, with Dr. Alan MacEachren, Dr. James Macgill, Dr. Roxanne Parrott, Dr. Gary Chase + Stephen Crawford, and Brenda Klusman

Basic GIS-Atlas architecture



```

    graph LR
      A[Spatial database  
Postgres /GIS] <--> B[Data server  
Geoserver]
      B <--> C[Interactive maps & graphs  
(client)  
Flash]
  
```

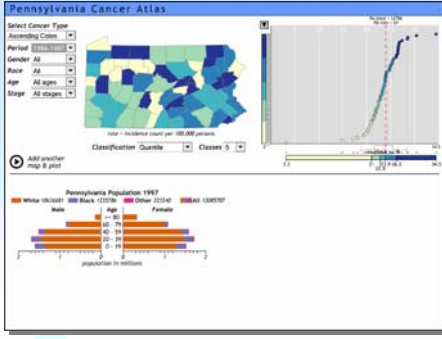
User at the client end requests maps or analyses on maps

Server converts these into queries to the spatial database and retrieves required data

Data are returned to the client for display

GIS/Atlas “Client” web-map features

<http://www.hmc.psu.edu/gisatlas/>



<http://www.geovista.psu.edu/grants/CDC/>

Summary

- ♦ emphasis is on
 - ♦ integrating an understanding of the geographic aspects of cancer into research and policy
 - ♦ basic and applied GIScience research that supports visual analytics applied to cancer surveillance, control, and prevention
- ♦ research addresses all aspects of the scientific research process, from exploring and processing data, through development of hypotheses, through analysis to policy formulation and communication to the public
- ♦ beyond visual analytics, research in the GeoVISTA Center also focuses on development of scientific collaboratories, support for group work more generally, and scientific knowledge management

thank you

questions?

in addition to faculty collaborators listed for each project, the following current and former grad students/postdocs have contributed to this work:

Tanuka Bhowmick, Xiping Dai, Robert Edsall, Diansheng Guo, Amy Griffin, Frank Hardisty, Matthew Mulbrandon, Soyhun Park, Anthony Robinson, Erik Steiner, Billang Zhou