# Data Visualization Techniques and Practices

## Introduction to GIS Technology

Michael Greene
Advanced Analytics & Modeling, Deloitte Consulting LLP
March 16th, 2010

# Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

# Agenda

Introduction

Geometry

Spatial Quantification of Data

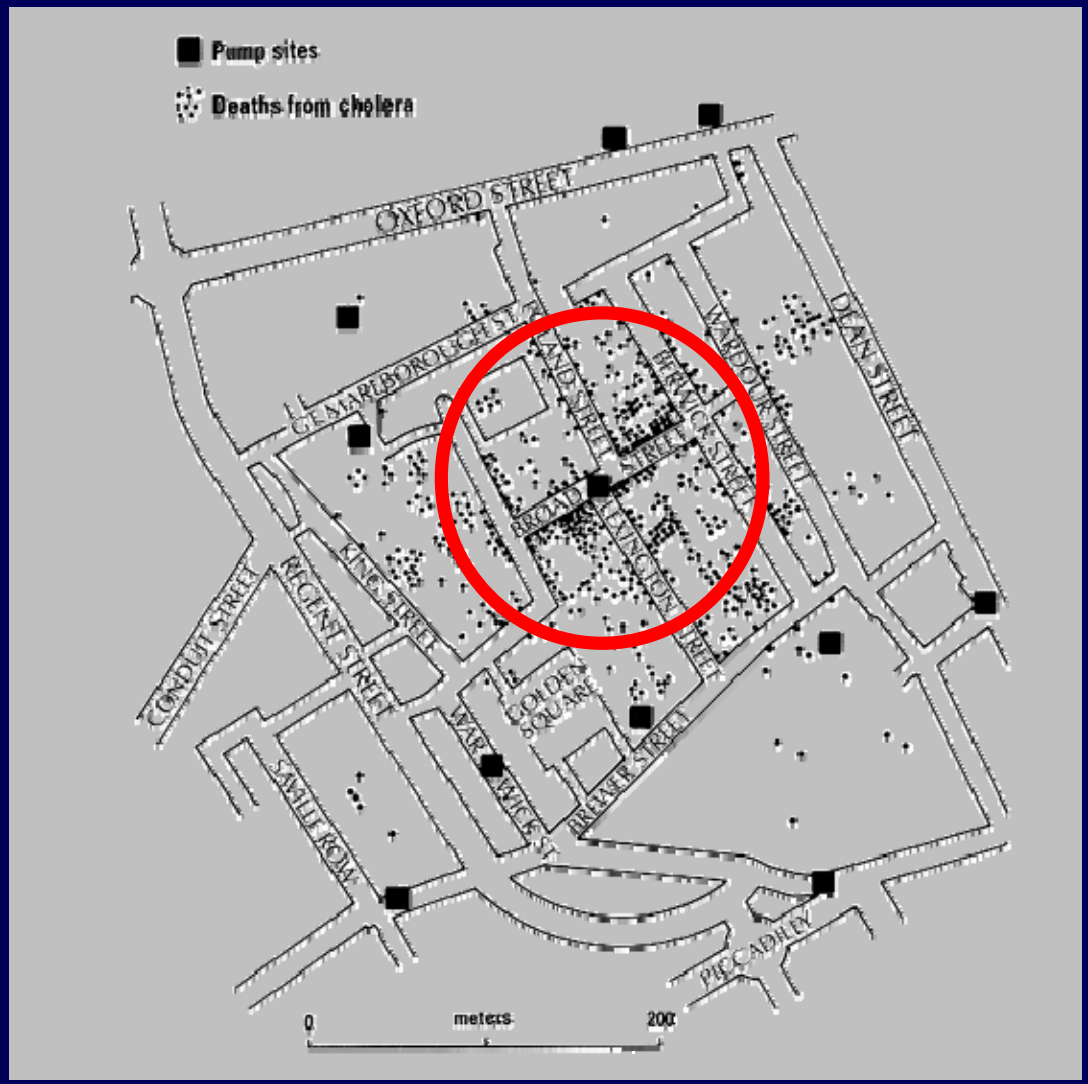
Case Study – Snow's Cholera Map

Conclusion

# Introduction

# Why is Geographic Data Important?

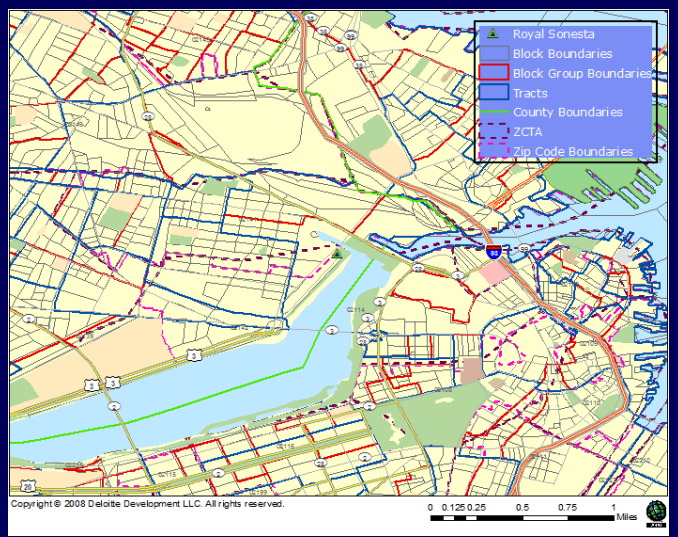
Mapping and geographic techniques allow complex patterns to be represented visually – revealing hidden patterns



# Geographic Information Systems - GIS

Geographic Information Systems are technology suites that allow analysts to quantitatively represent data in a spatial plane

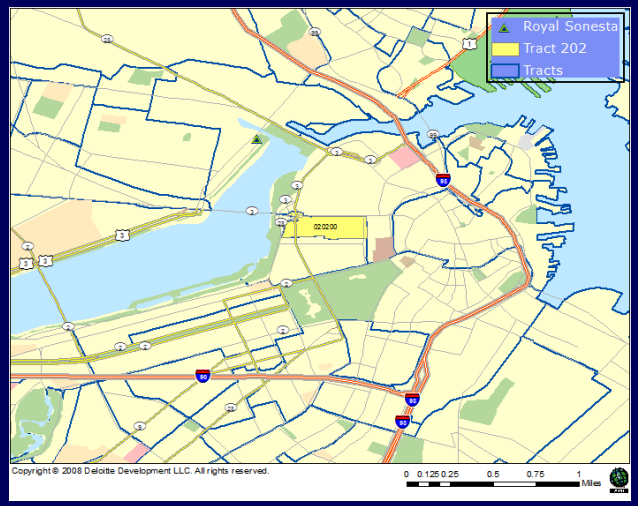
- Coordinate systems – i.e., longitude, latitude
- Representations of spatial concepts through geometry
  - Points, Lines, Polygons
- Allow for multiple “layers” of data to be represented on a single plane

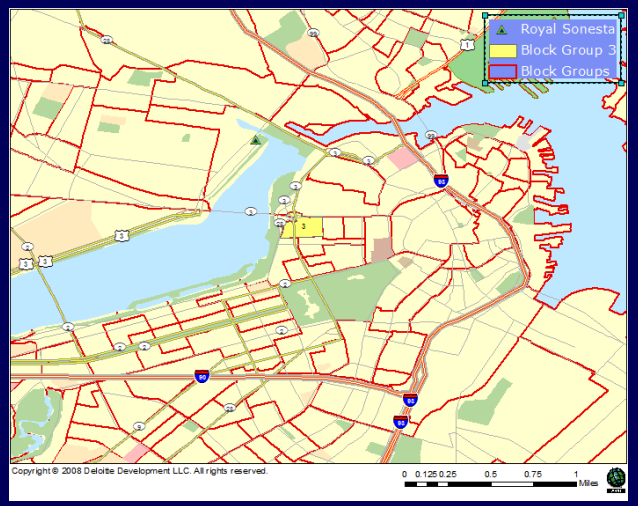


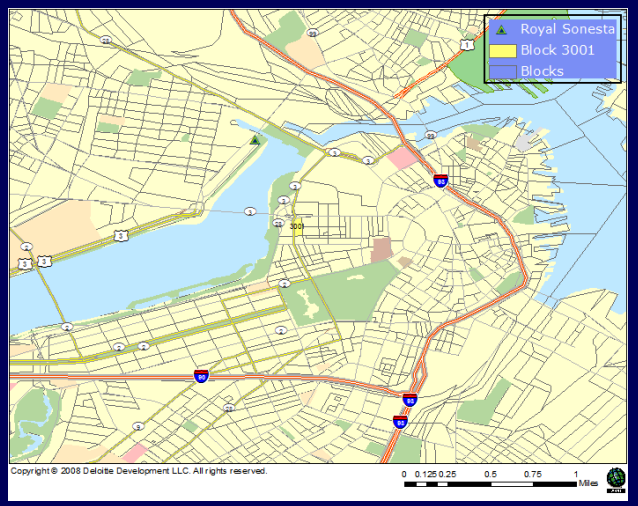
# Background and History

GIS technology has its roots in Urban Planning and areas that require layered information presented on maps

- Many sources are public from public databases
  - Roads
  - Land parcels
  - Boundaries, such as counties, Census areas, ZIP codes

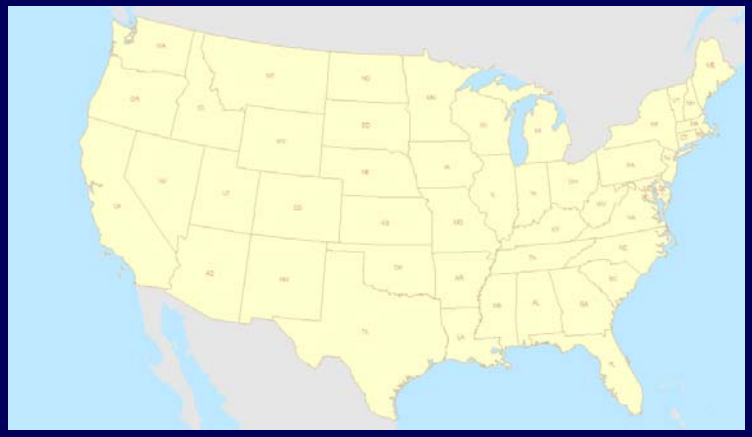


# Geometry

# Basis of GIS is Geometry



Polygons

- State boundaries
- ZIP Codes


Points

- Cities (on a large map)
- Addresses (Longitude/Latitude)

Lines

- Roads/Highways
- Rivers, natural boundaries

# Coordinate Systems

**Measurement Framework**

- Geographic: spherical coordinates
- Planimetric: projected coordinates onto a 2-dimensional surface

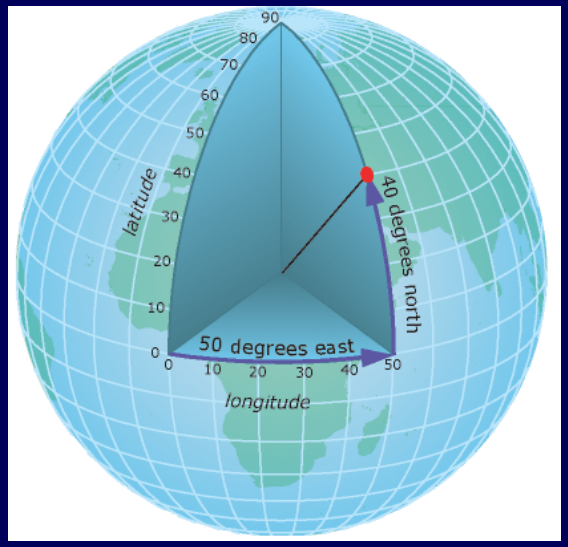
**Unit of Measurement**

- Miles, feet, meters, kilometers
- Decimal degrees

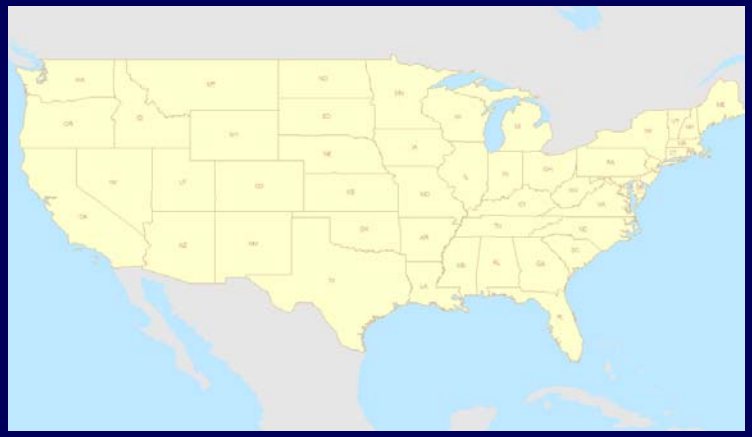
**Other Properties**

- Projection definition
- Spheroid of reference
- Datum
- Standard parallels, central meridian etc

# Geographic Coordinate Systems

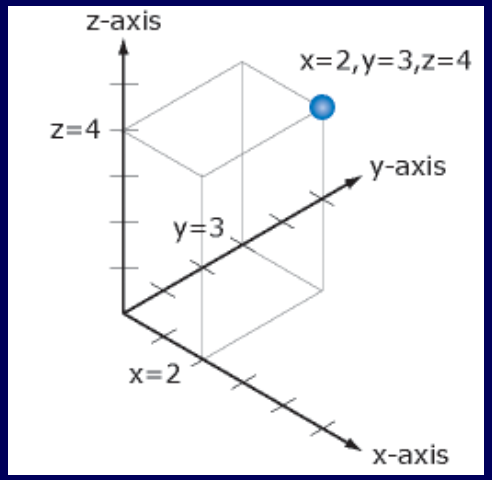


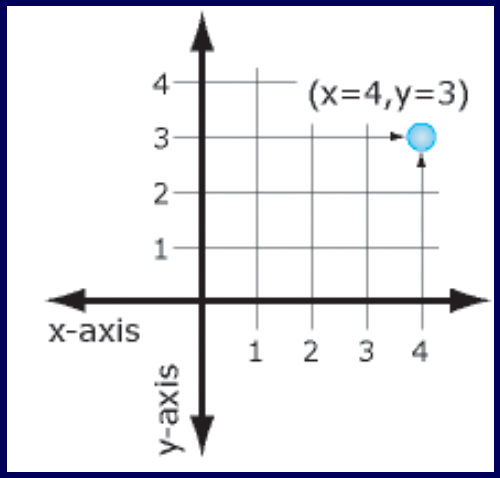

Source: ESRI ArcView

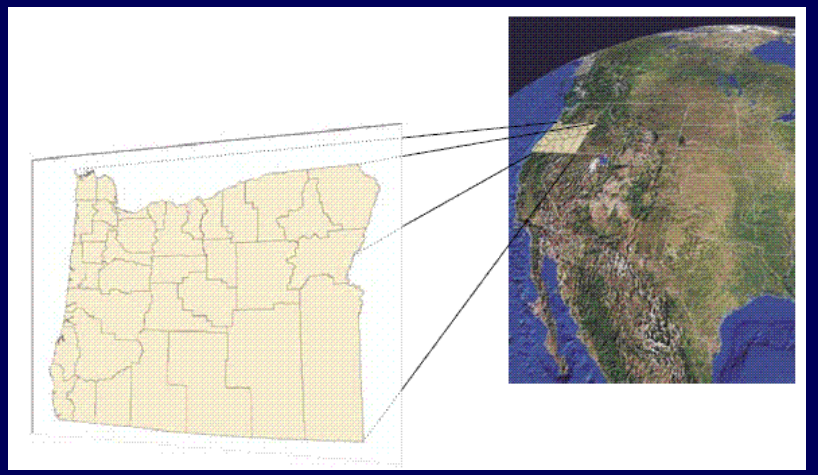


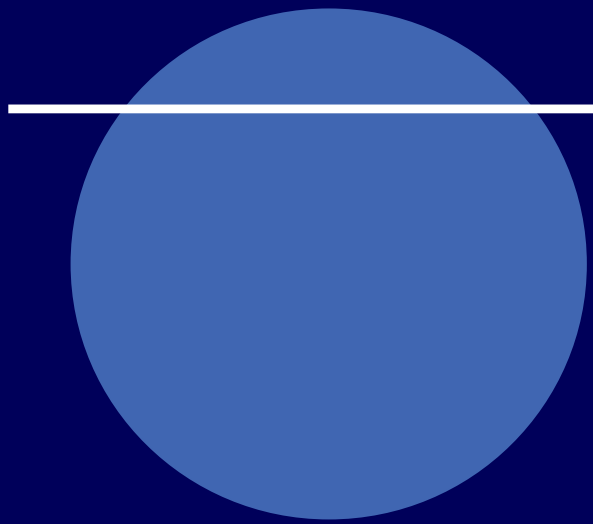
- Spherical measure of position in longitude and latitude (angles such as -180º - +180º)
- Northern Hemisphere and Western Hemisphere often have positive values
- Often more applicable to storage and usage of global data
- Local views plotted using a spherical system can appear distorted

# Projected Coordinate Systems

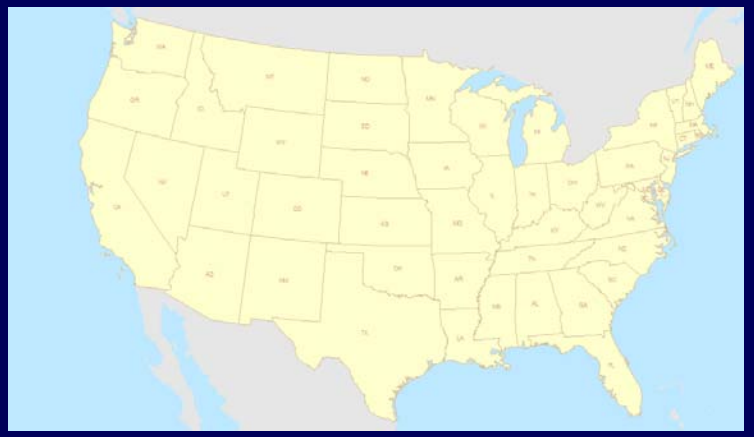




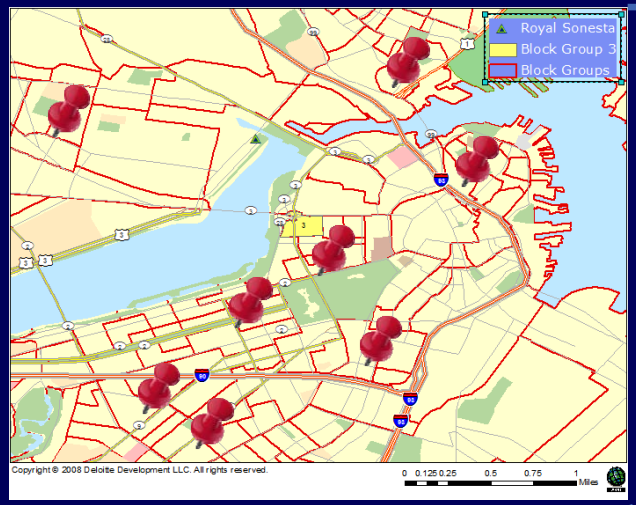
Projection Plane



Source: ESRI ArcView

# Spatial Quantification of Data

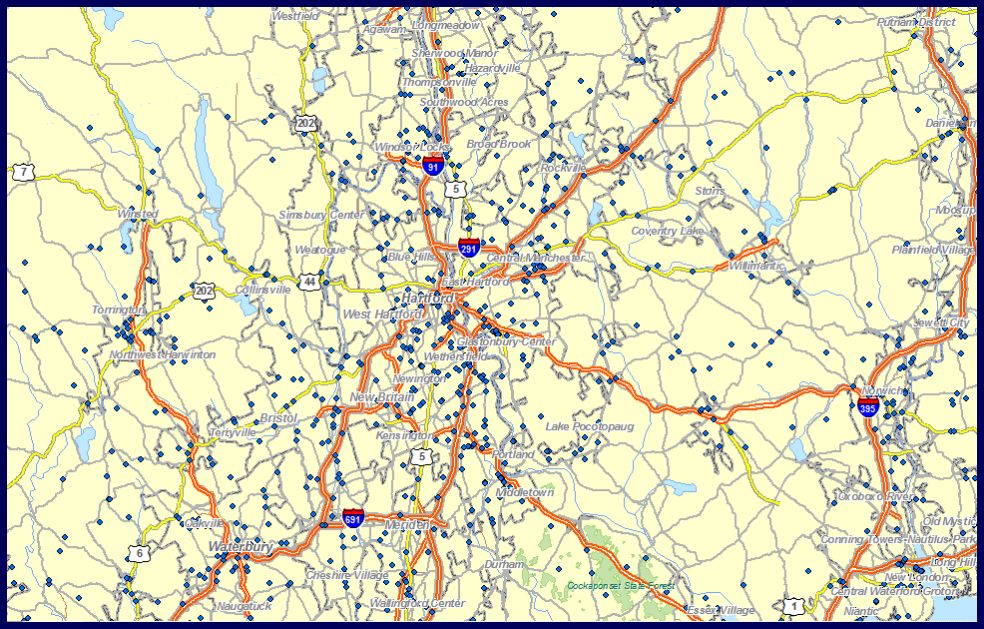
# What makes GIS so powerful?




| ZIP Code | Median Age | Median Income | Loss Freq. | Relative Loss Severity |
|---|---|---|---|---|
| 02115 | 30 | 60,000 | 0.155 | -10% |
| 02116 | 35 | 80,000 | 0.13 | +1% |
| 02114 | 45 | 120,000 | 0.08 | +5% |
| 02118 | 40 | 50,000 | 0.25 | +5% |

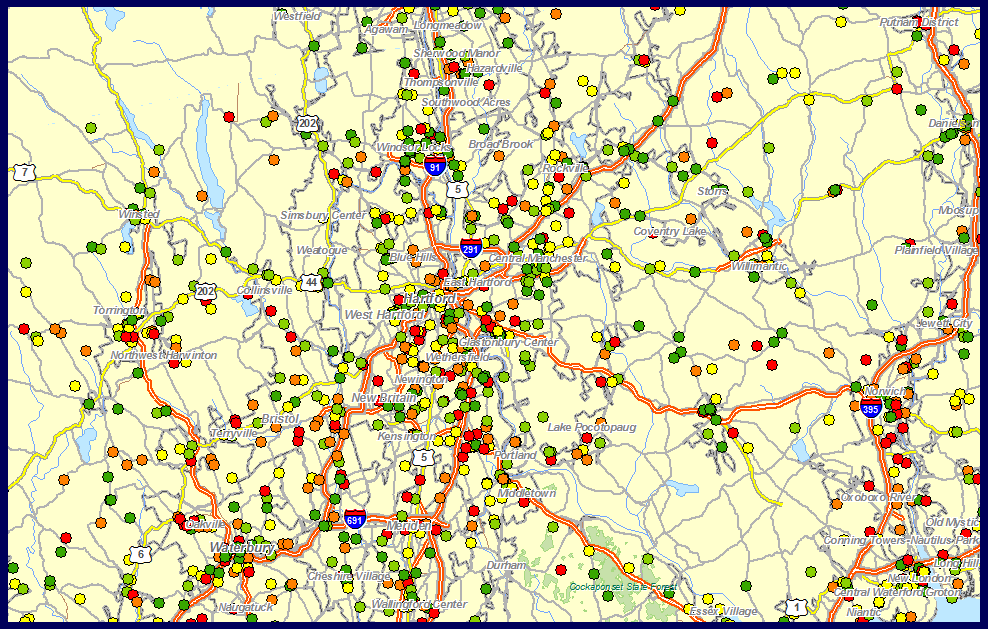
**Merging data associated with geographic areas and individual points can lead to powerful results**

# Integration of 2-D Space in Analysis




- Association of quantitative data and spatial techniques
- Spatial calculations take 2 dimensions into consideration
- Application of clustering, smoothing, and other mathematical methods

# Clustering Techniques

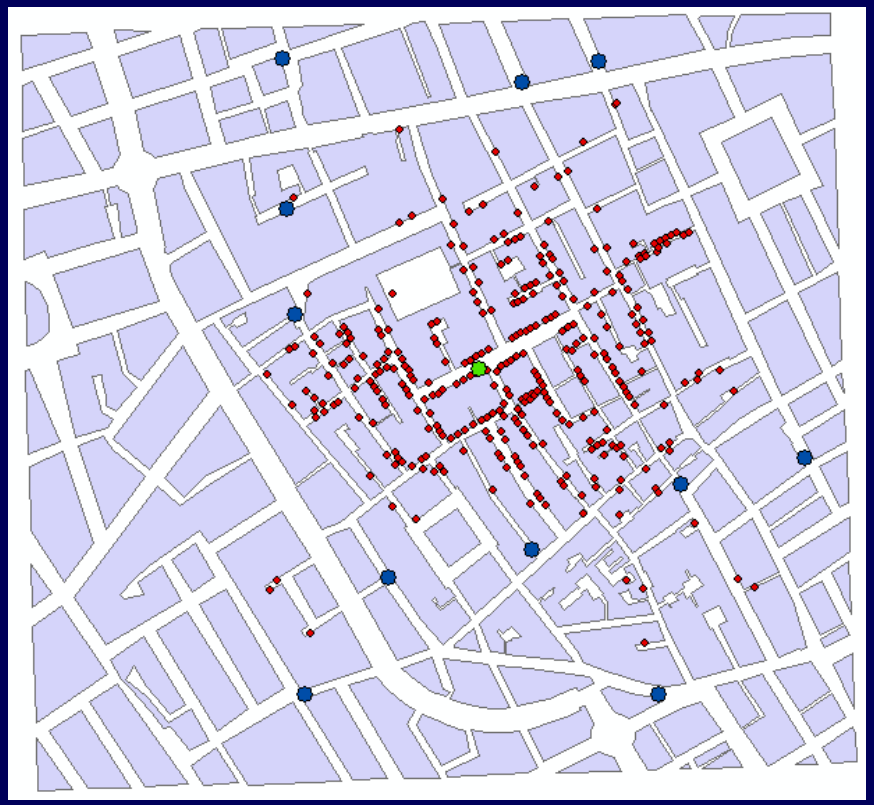



- Hotspot Analysis can be used to uncover hidden clusters
  - Getis-Ord G* - Local averaging of a quantity within a distance radius
- Cluster and Outlier Analysis
  - Moran's I – Spatial Autocorrelation
- Developed during analysis of crime statistics

# Case Study – Snow's Cholera Map

# Snow’s Cholera Map

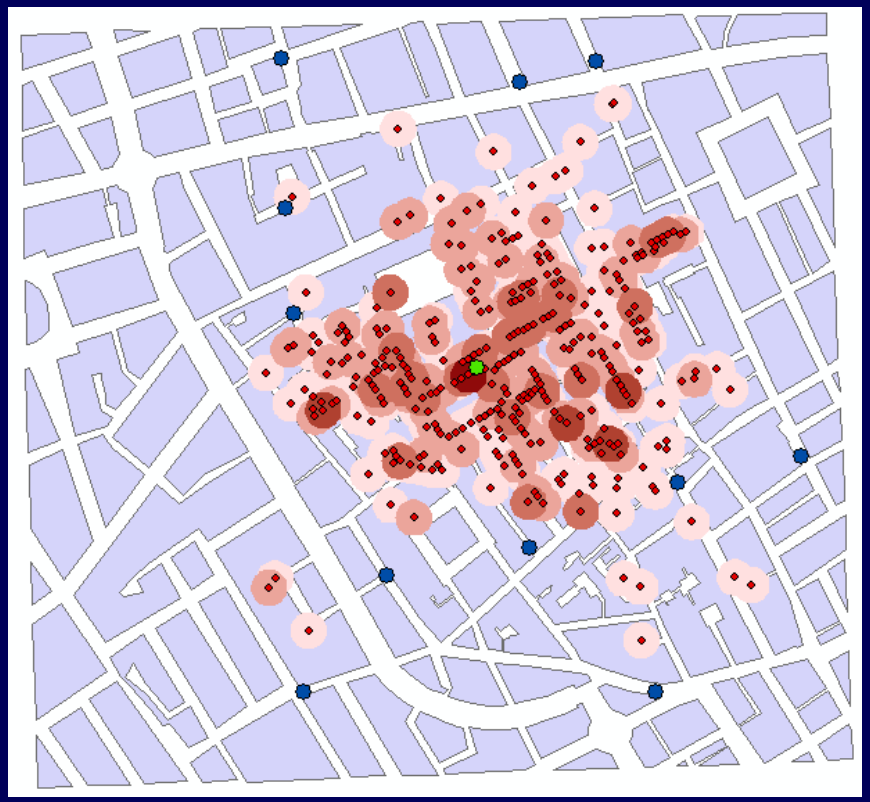
# Spatial Autocorrelation



Moran's I: Spatial Autocorrelation

- Based on spatial distance and a feature (number of deaths)
  - Range -1 to 1 indicates clustering
  - Statistical test: Z score
- Cholera Outbreak:
  - Moran's I: 0.0103
  - Z score: 2.45 (p value < .05)

# Hot Spot Analysis



Getis-Ord G* - Hot Spot Analysis

- Draw buffers around each point
- Calculate sum of value intersecting points within each buffer
- Compare to expected, calculate Z-score

# Conclusion

# Applications to Insurance

- Improved territory analysis
  - GAMS/GLM
  - Clustering
- Fraud detection
  - Clustering: Moran's I
  - Hot Spot Analysis: Getis-Ord G*
- Agent analysis and insights
  - Source and dispersion of customer base
  - Distribution of quotes versus current book

# Tools and References

- ESRI ArcGIS Product
  - http://www.esri.com
- R
  - http://www.r-project.org
  - R GIS package
- John Snow's Cholera Map
  - Prof. Waldo Tobler
    National Center for Geographic Information and Analysis
    Geography Department
    University of California Santa Barbara CA 93106-4060
    email: tobler@geog.ucsb.ed
  - http://www.ncgia.ucsb.edu/pubs/snow/snow.html
  - http://www.asdar-book.org/RC1

Deloitte.



Member of
Deloitte Touche Tohmatsu