OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

INF FAKULTÄT FÜR INFORMATIK

Visual Analytics in HealthCare

Potentials of visual analytics in epidemiology

Supervisor

Dr. Prof. Bernhard Preim

Team Members

Sneha Videkar

Niha Mohanty

Outline

- Epidemiology
- Evidence based decision making
- Visual Analytics
- Tool : Decision Support Environment
- Tool : IVA framework
- Summary

Epidemiology

What is

- Scientific discipline
- Reliable knowledge for clinical medicine

Focus on

- Prevention diseases
- Diagnosis diseases
- Treatment of diseases

Aims at

- Identify risk factors for the outbreak of diseases
- Evaluation of certain treatment strategies

Epidemiology

Uses

- Hypothesis and statistical Analysis
- Population-based or sample of the population
- Representative portions of a population (without known diseases) are examined
- From a city, a region or a country

Result

- The protection of people from passive smoking
- Recommendations for various vaccinations
- The introduction of early cancer detection strategies
- Official guidelines for any disease

Outline

- Epidemiology
- Evidence based decision making
  -by Brownson, Ross C. et al.
- Visual Analytics
- Tool : Decision Support Environment
- Tool : IVA framework
- Summary

Evidence based decision making in public health

Why
- Official guidelines for the treatment of widespread diseases such as diabetes
- Based on evidence from epidemiological studies

Who
- Public health practitioners always incorporate scientific evidence

What
- In making management decisions, developing policies, and implementing programs

Issues
- Decisions often are based on short-term demands rather than long-term study
- Policies and programs are developed frequently around unreliable evidence

Solution
- Need evidence-based approach in decision making

Evidence based medicine & public health

Evidence based medicine

- **Task:** Delivery of optimal individual patient care
- **How:** Uses current best evidence on pathophysiological knowledge, cost effectiveness, and patient preferences
- **Result :** applied in clinical practice
- **Skills:** medical experts
- **Data collected:** Randomized control trials
- **Example :** Medical studies of pharmaceuticals

Evidence based medicine & public health

Evidence based public health

- **Task:** Development, implementation, and evaluation of effective programs and policies in public health
- **How :** Makes systematic uses of data and information systems and appropriate use of program planning models.
- **Result:** The most viable approach to a public health problem is chosen from among a set of rational alternatives.
- **Skills :** epidemiology, biostatistics, behavioral sciences, health economics, and health care management.
- **Data collected:** cross sectional study and quasi experiments
- **Example:** Tobacco control

Case study

Tobacco control

- Decades of research and thousands of epidemiologic studies
- Cigarette smoking as the leading cause of preventable premature death.
- Increased tobacco taxes are an important tool for decreasing tobacco consumption
- Excise tax and media campaign
- Drop in both sales of cigarettes and in smoking.

Evaluating Evidence

Analytic Tools and Processes

- Meta-analysis
- Risk assessment
- Economic evaluation
- Public health surveillance
- Expert panels and consensus conferences

Is Evidence Sufficient?

Issue

- Determining when evidence is sufficient for public health action
- What the specific action should entail.

Solution

- Level 1 : Some-thing should be done
  - Analytic data that shows the importance of a particular health condition
  - Its link with some preventable risk factor
- Level 2: Specifically, this should be done
  - Evidence focuses on the relative effectiveness of specific interventions
  - To address a particular health condition.

Sequential framework for health practitioners

Stage one

- Develop an initial, concise, operational statement of the issue
- Problem definition
- Describe the mission, internal strengths and weaknesses, external opportunities and threats
- Also the vision for the future.

Stage two

- Determine what is known through the scientific literature
- A systematic approach to identify, retrieve, and evaluate relevant reports.
- MEDLARS, MEDLINE, PubMed, Current Contents, HealthSTAR, and CancerLit

Stage three

- Quantify the issue
- Cross-sectional data can provide data for use in the design of analytic studies
- Also can be used as baseline data to compare the effectiveness of public health interventions

Sequential framework for health practitioners

Stage Four

- Develop program or policy options
- The list of options can be developed from a variety of sources.

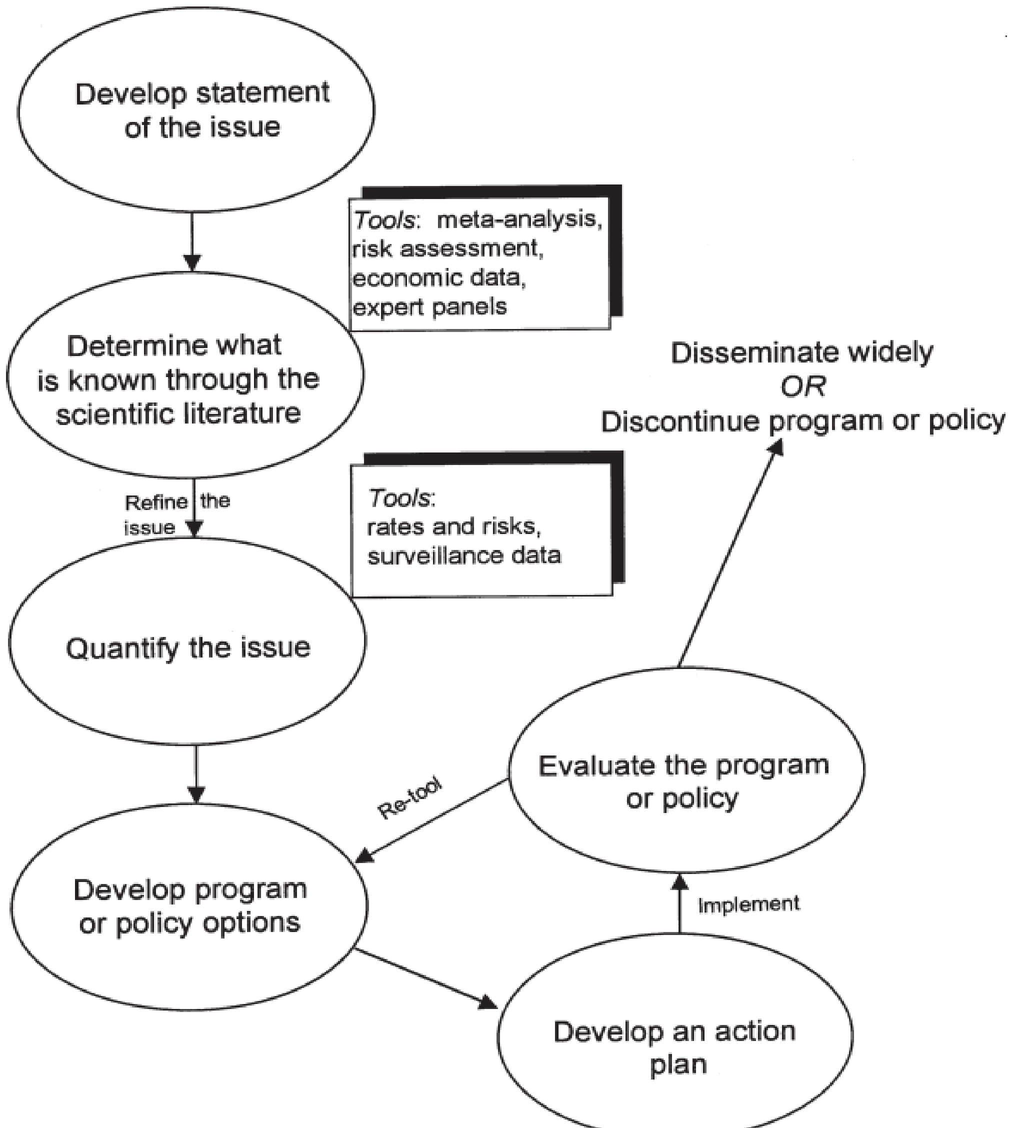
Stage Five

- Develop an action plan for the program or policy
- A goal is a long-term, desired change in the status of a priority health need
- And objective is a short-term, measurable, specific activity that leads toward achievement of a goal.
- How the goals and objectives will be achieved, what resources are required, and how responsibility of achieving objectives will be assigned.

Stage Six

- Evaluate the program or policy
- Evaluation is the determination of the degree to which program goals and objectives are met.
- Quasi-experimental designs

Sequential framework for health practitioners



Outline

- Epidemiology
- Evidence based decision making
- Visual Analytics
  - By Kenneth K. H. Chui et al.
- Tool : Decision Support Environment
- Tool : IVA framework
- Summary

What is Visual Analytics?

Visual Analytics

- Science of analytical reasoning facilitated by interactive visual interfaces – *by Jim Thomas*
- A tool that allows better understanding of the context of large, complex system
- Helps in decision making

Visual Analytics Relationship with Public Health

Visual Analytics & Public Health

- Greatly benefits Public Health professionals mainly through disease monitoring and bio-surveillance
- Especially useful in epidemiological investigations

Visual Analytics Potentials

Visual Analytics Applications in Epidemiology

- **Flow maps** to display population mobility and person-to-person contact across wide geographic space
- **Dynamic mapping** of disease occurrences with simultaneous depiction of spatio-temporal changes in environmental factor
- **Detection** of spatio-temporal hotspots
- **Displaying** genotype data

Methods

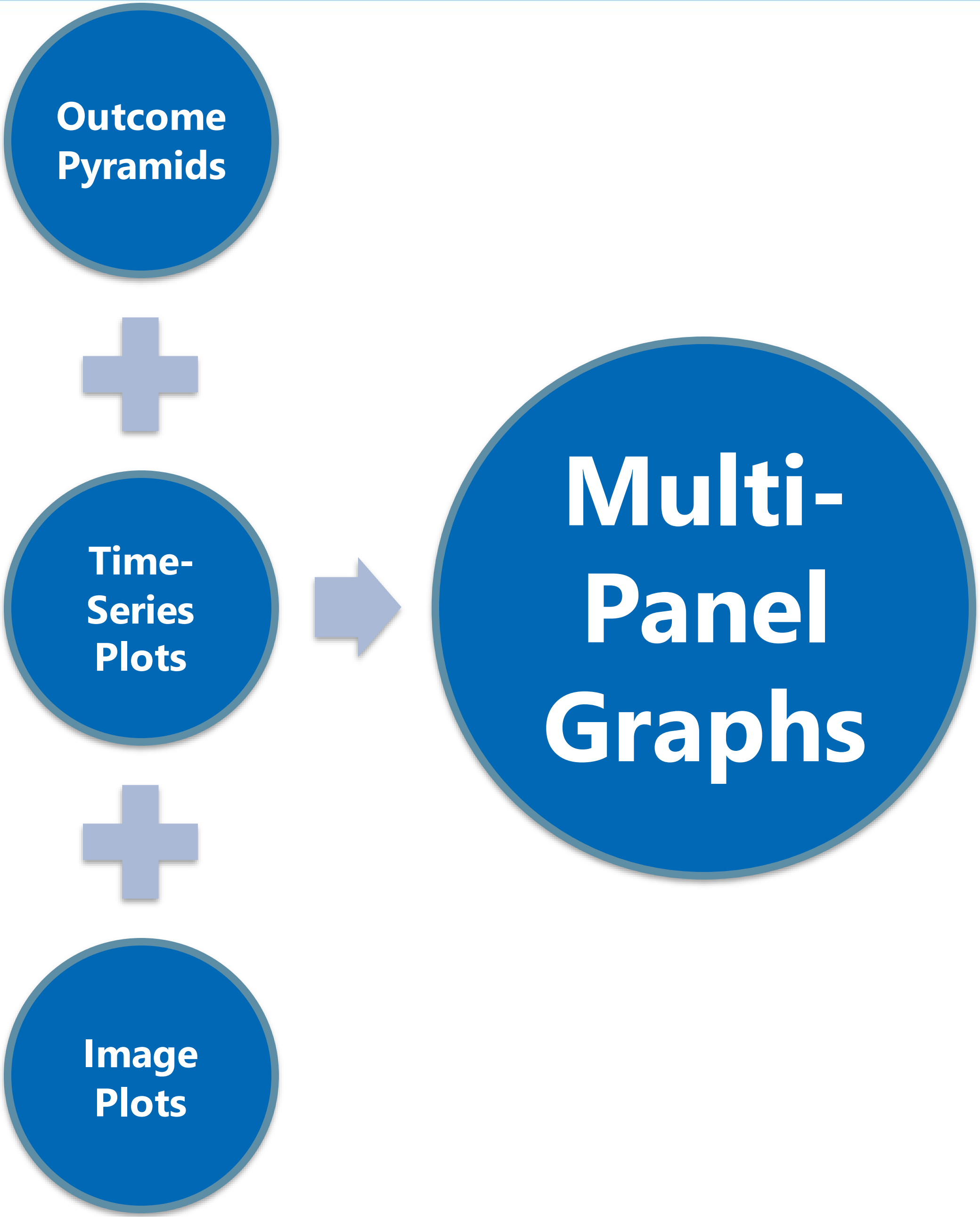
Data Sources

- US National **Influenza-associated** and **Salmonellosis-associated** hospitalization among the older adult population (>= 65 years old) between January 1991 and December 2004 by **Centers for Medicare & Medicare Services (CMS)**
- Confirmed **Salmonellosis** cases reported at **Massachusetts Department of Public Health** for the general population, 2004 – 2005
- Children aged from 0-8 visiting **Milwaukee Children's Hospital of Wisconsin** regarding **Asthma**, 1997-2006

Methods

Visualization & Analysis

- **Outcome Pyramids** – Based on population pyramid (type of graph commonly used to describe the combination of age and sex of a population)
- **Time-series Plots** – Visualizes temporal trends and seasonal patterns in disease rates
- **Image Plots** – Capable of displaying information for at least three variables


Outcome Pyramids
Time-Series Plots
Image Plots
Multi-Panel Graphs

Major Focus – Tool

Multi-Panel Graphs

- New graphical design involving the strategic positioning of two or more graphs sharing at least one common axis on a single canvas
- Offers a novel way to synthesize complex visual data
- Potentially useful tool for illustrating data with complex, multifaceted structure

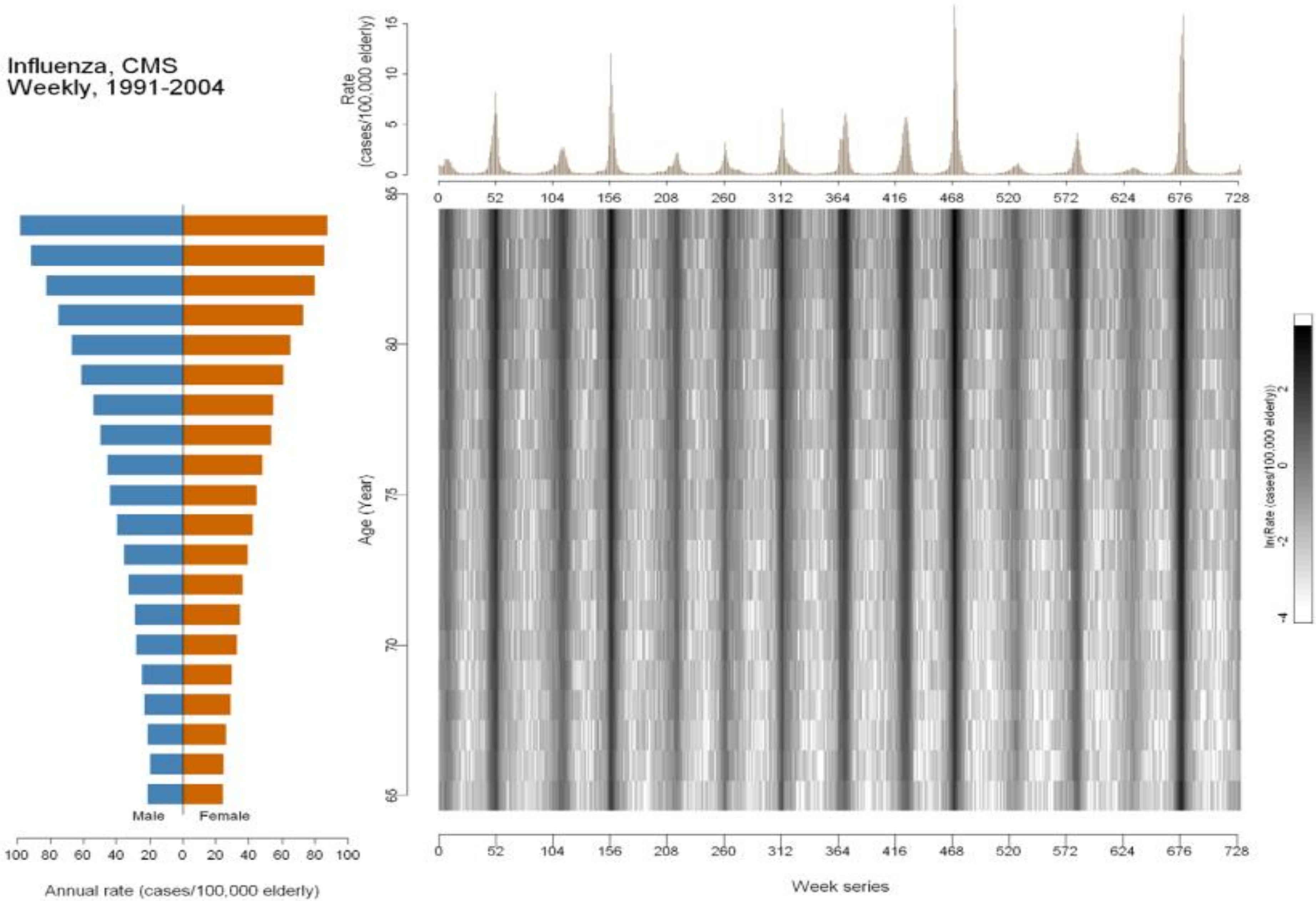
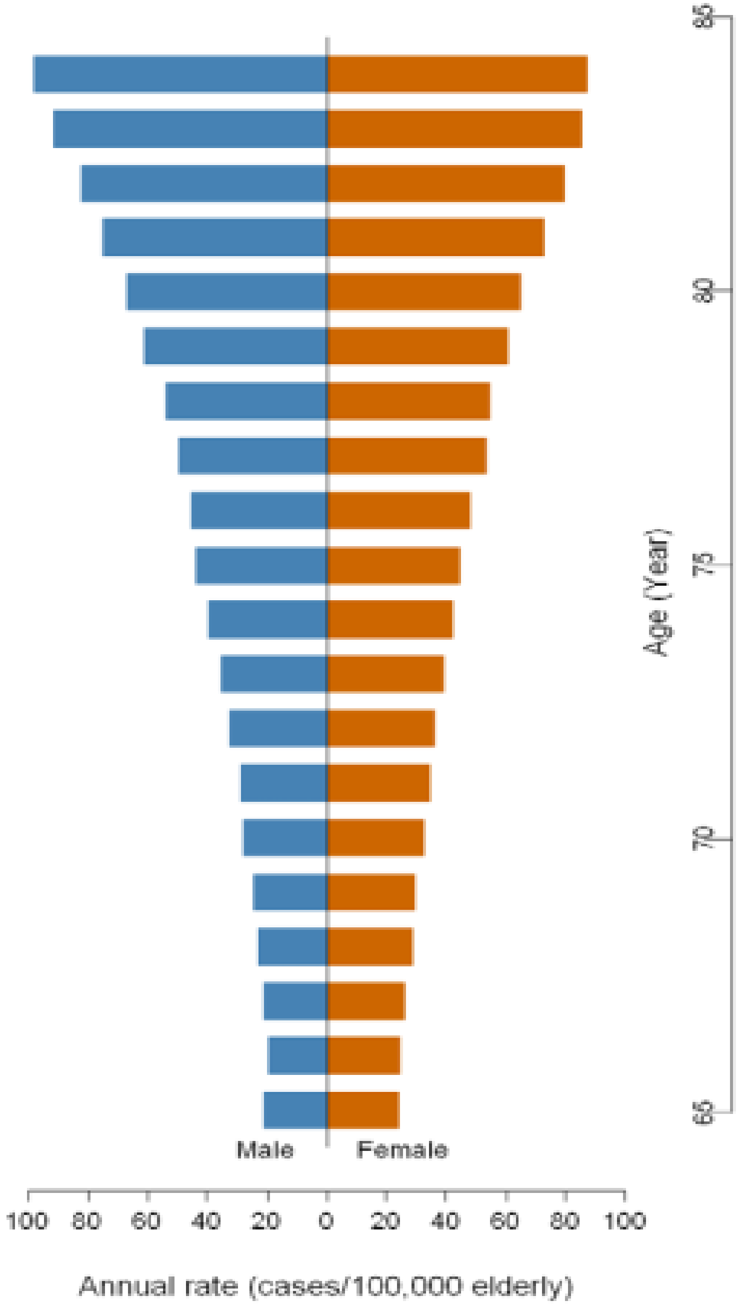
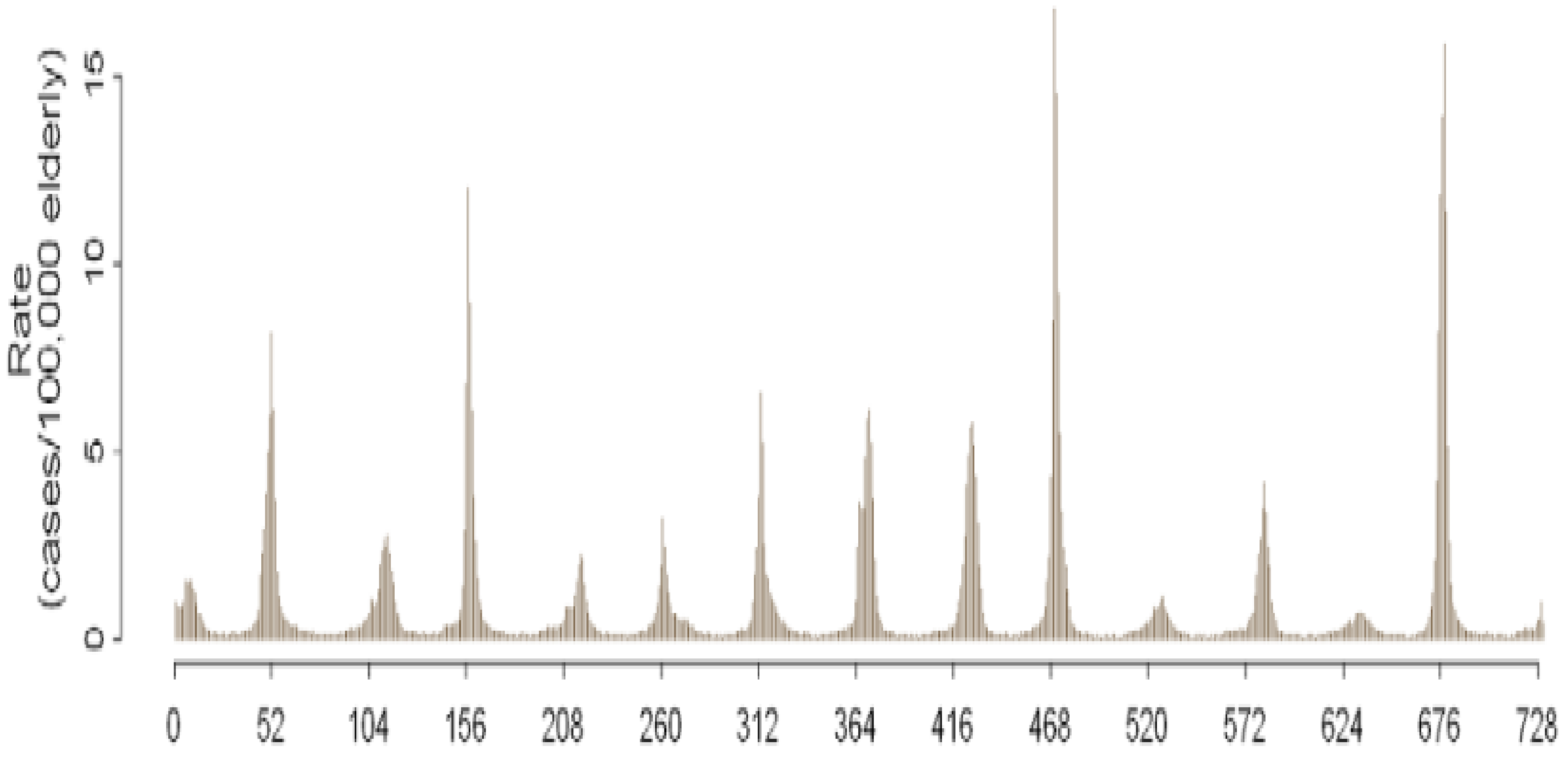


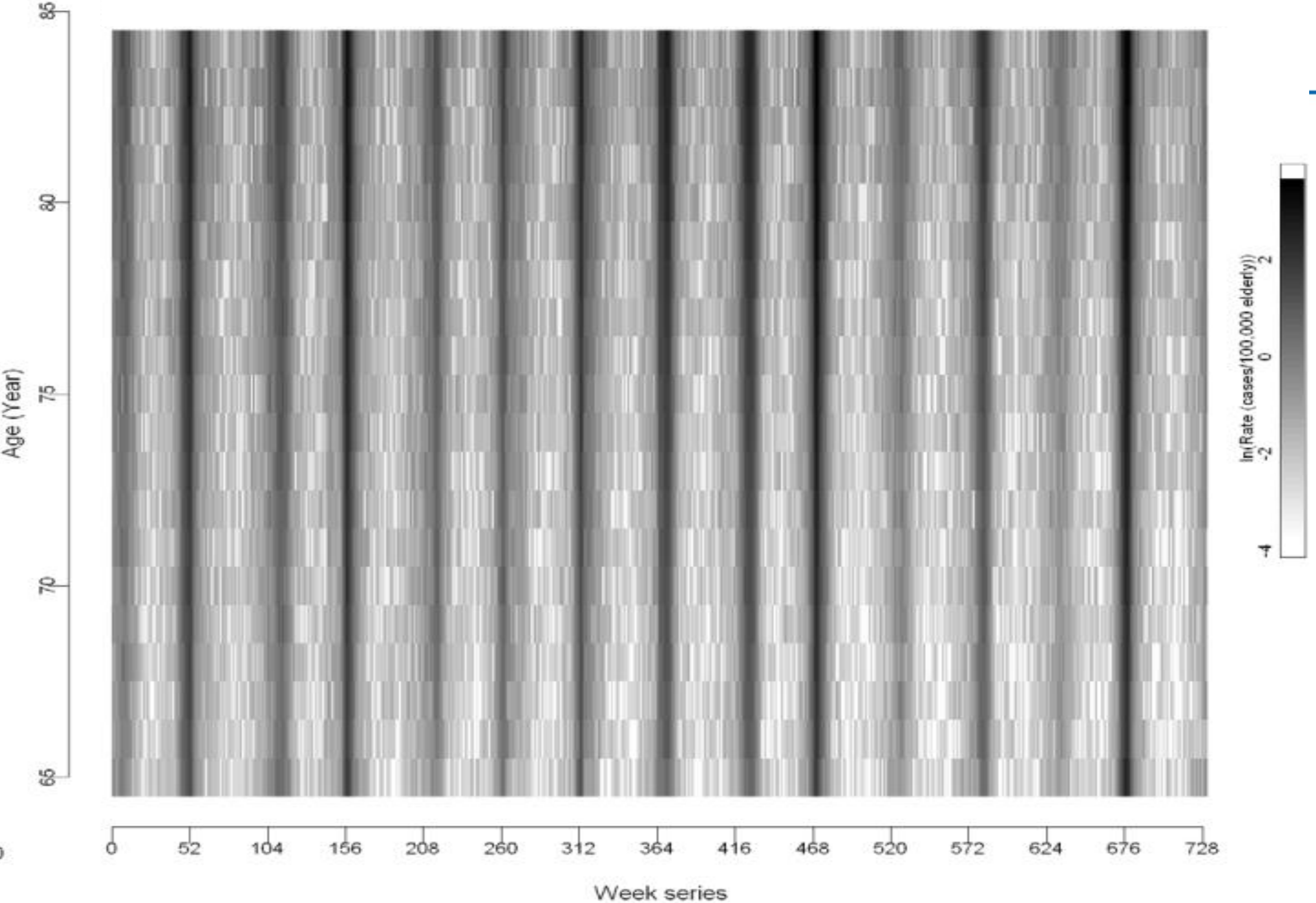
Fig. 1. Multi-Panel Graph of influenza in the US older adult population (aged 65 and above) 1991 - 2004



US Influenza – Outcome Pyramid

US Influenza – Time-series





US Influenza – Image Plot

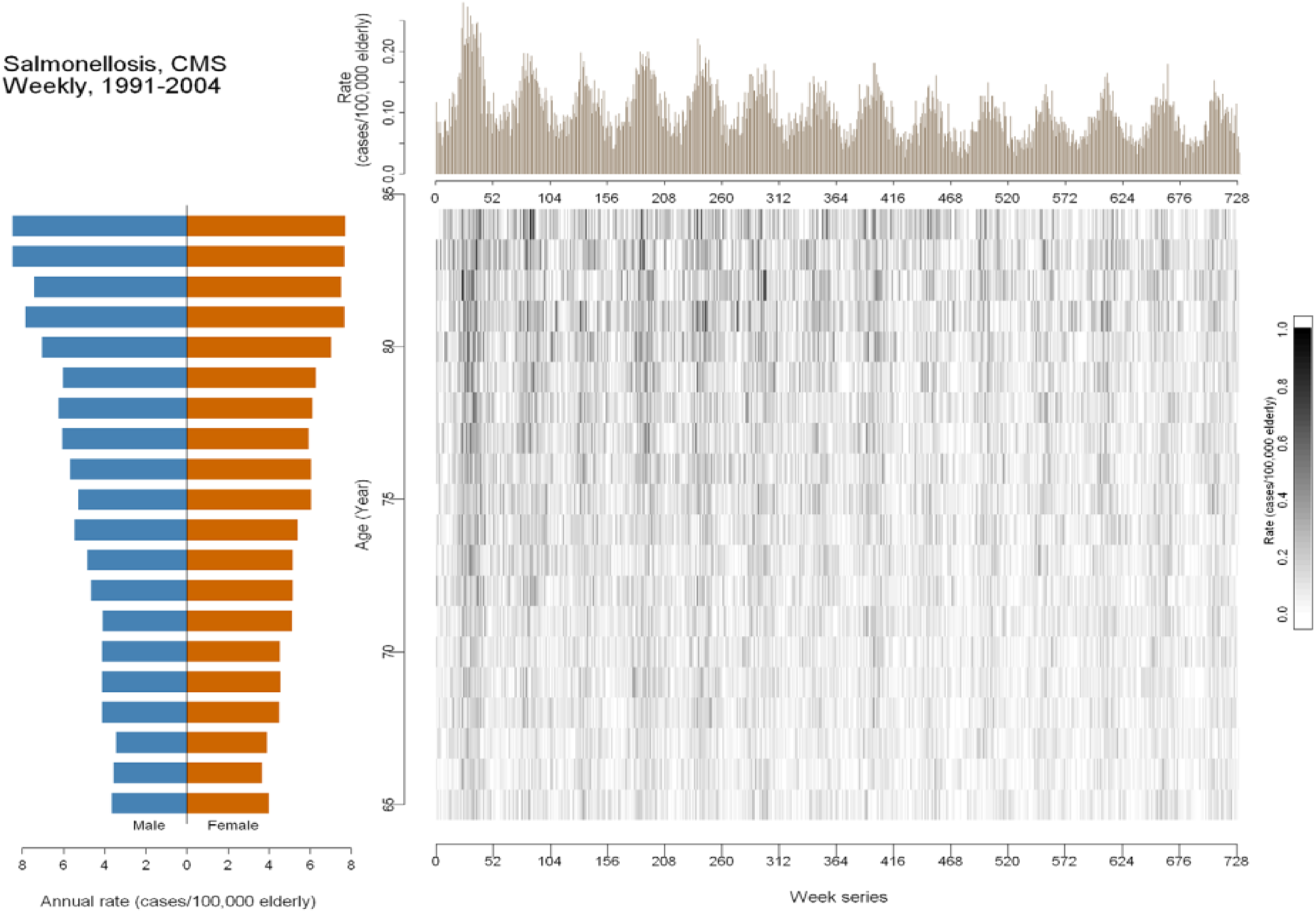


Fig. 2. Multi-Panel Graph of Salmonellosis in the US older adult population (aged 65 and above) 1991 - 2004

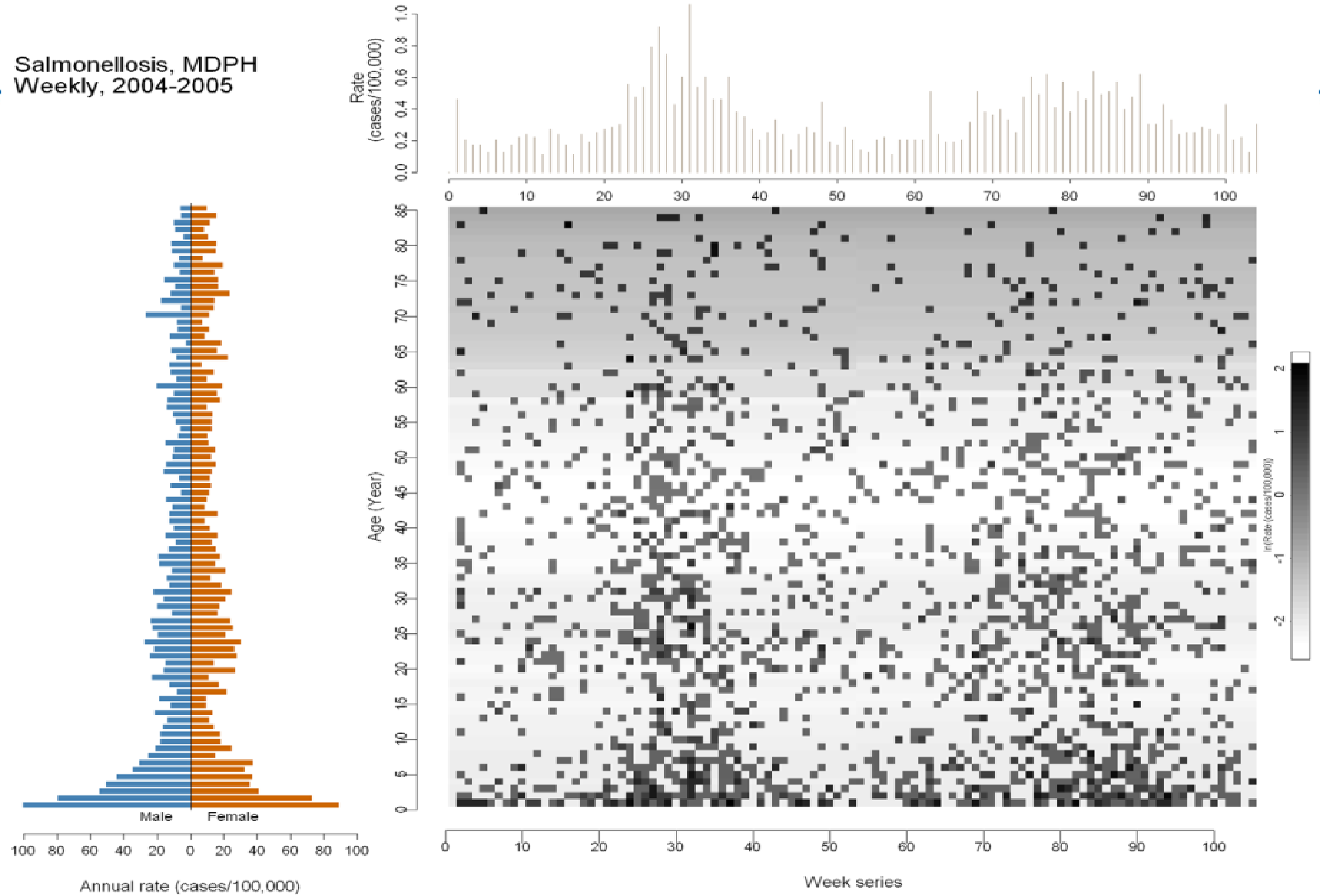


Fig. 3. Multi-Panel Graph of Salmonellosis in the general Massachusetts population 2004-2005

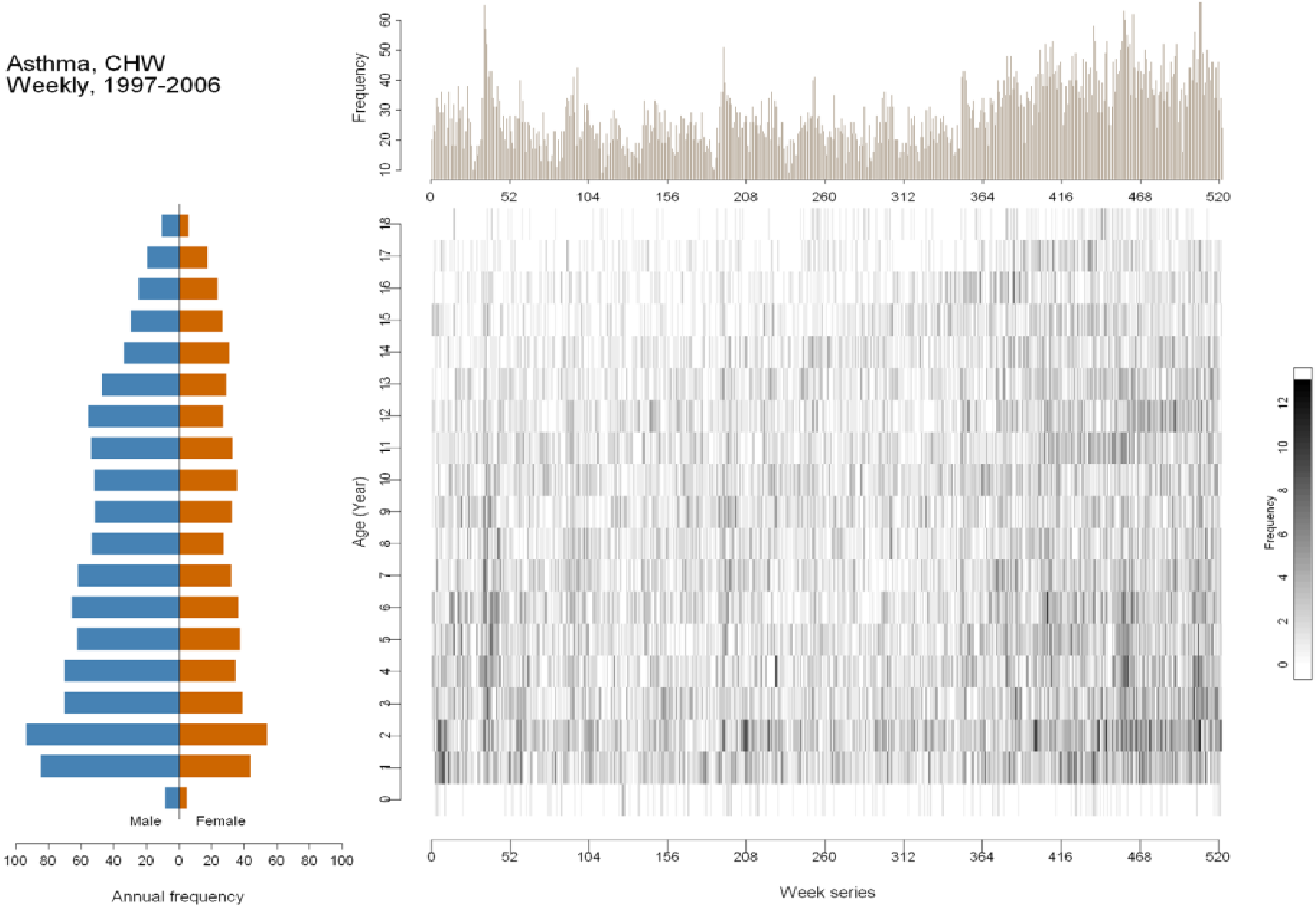


Fig. 4. Multi-Panel Graph of Asthma in children aged 0-18 in Milwaukee 1997-2006

What are the strengths of Multi–Panel Graphs?

Strengths

- Facilitate the detection of increased risk of disease in vulnerable subpopulations at a given time period
- Depicts the projected change in demographic structures
- Expected disease outcomes based on local forecasts

What are the precautionary steps need to be handled while using Multi-Panel Graphs?

Precautions

- Precision of age valuation in describing population structure
- Temporal scale and resolution
- Amplitudes of temporal variations in both population at risk and disease

Outline

- Epidemiology
- Evidence based decision making
- Visual Analytics
- Tool : Decision Support Environment
  – By Afzal, S. et al.
- Tool : IVA framework
- Summary

Decision Support Environment

Epidemiology Study

- Scientists are studying the mechanisms by which diseases spread
- Predicting the future course of the outbreak
- Evaluating strategies applied to control an epidemic

Problems faced

- By researchers, analysts and public health officials
- Accurate diseases modelling but lack of interactive tools
- How to evaluate future course of ourbreak
- How to evaluate response measures within outbreak scenarios

A priori knowledge

- Public health officials must prepare and exercise complex plans to deal with a variety of potential mass casualty events
- Utilize knowledge gained on summary details
- Based on information and trends provided via very complex modeling.
- Ignored underlying traits and actual characteristics while considering a solution

Decision Support Environment

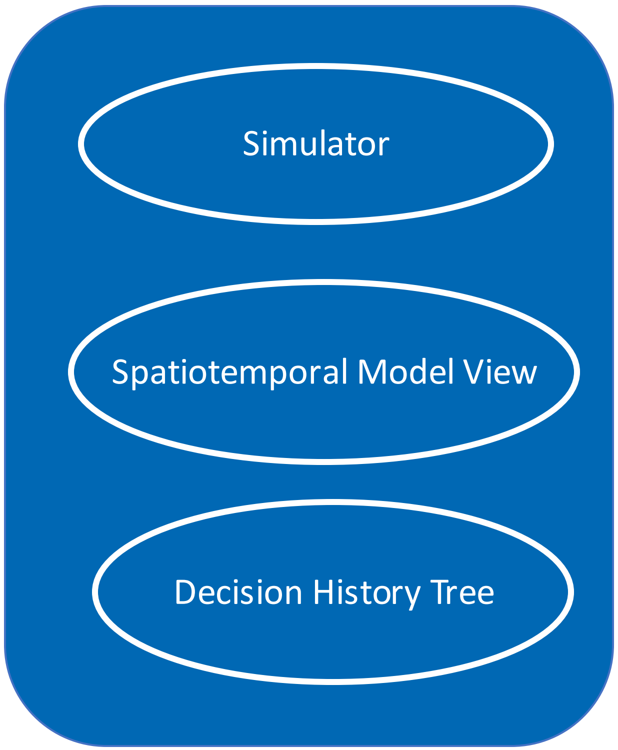
Solution

- Analysts and decision makers now incorporates computer based simulations to model potential disease
- Explore epidemic models and their impact
- Interactive decision support tools for analyzing the future course of the outbreak
- Evaluating potential disease mitigation strategies

Decision support environment

- Variety of parameters are employed in models
- Output complex multivariate data
- Flexible : for changes in data
- Allows comparison of results across space and time
- Gives best decision plan and also
- Evaluate various decision measures and understand the impact

System overview



Overview

- Two main views
- Spatiotemporal model view : to adjust model parameters and explore the effect of decision measures over space and time.
- To employ decision measures
- Decision history tree view : record users' decisions and allow them to compare, modify and insert new decisions
- Views are driven by an epidemic spread simulator
- Model input parameters are fed into the simulator
- The results of the simulation are modeled based on user-defined decisions

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG | INF | FAKULTÄT FÜR INFORMATIK

System overview

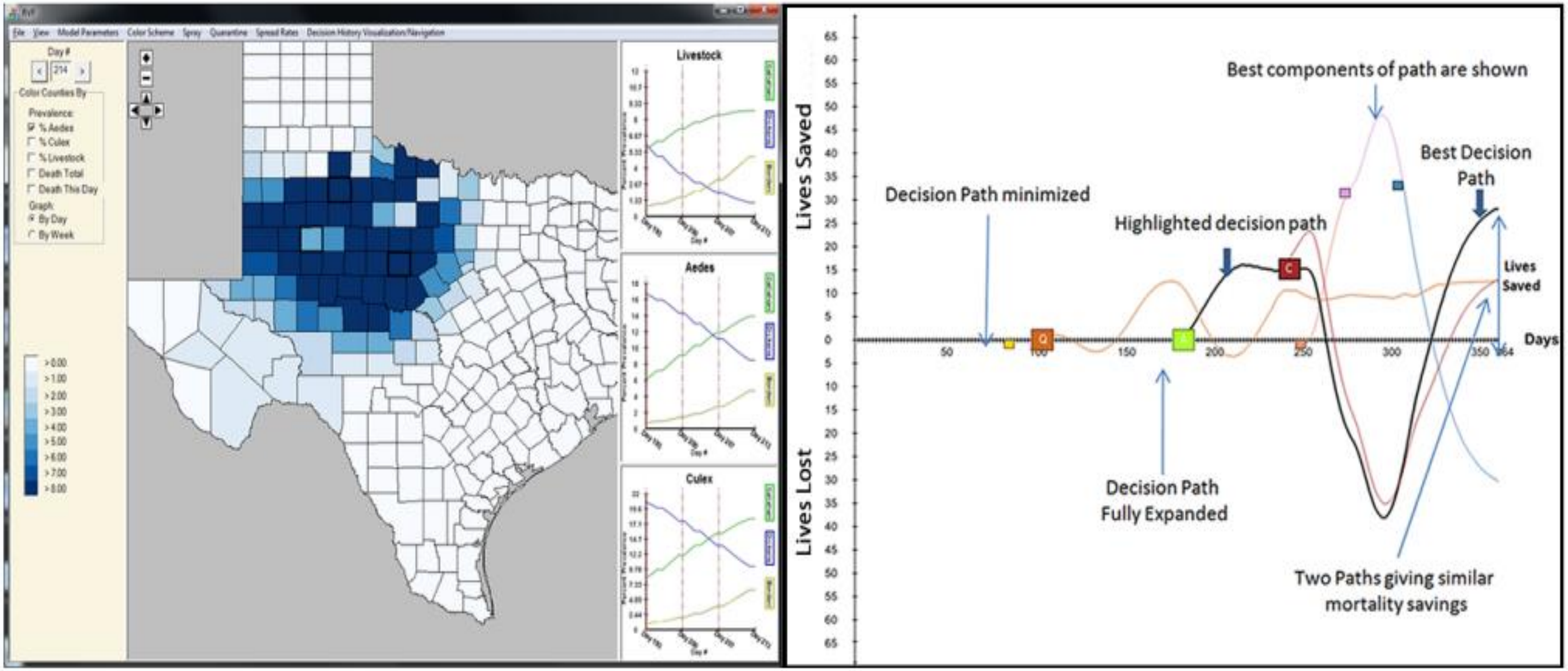


Figure 2: Visual analytics decision support environment. (Left) The spatiotemporal model view display. In this view, users can watch the spread of the model over space and time and introduce changes to the simulation as well as incorporate mitigative response measures to try and slow the disease spread. (Right) The decision history tree view. As users interact in the model view display, the different paths the simulation can take are calculated and visualized. The decision paths are plotted over time on the x-axis, with the y-axis representing the cumulative deviation from the baseline simulation.

Epidemiological Spread Simulator

Details

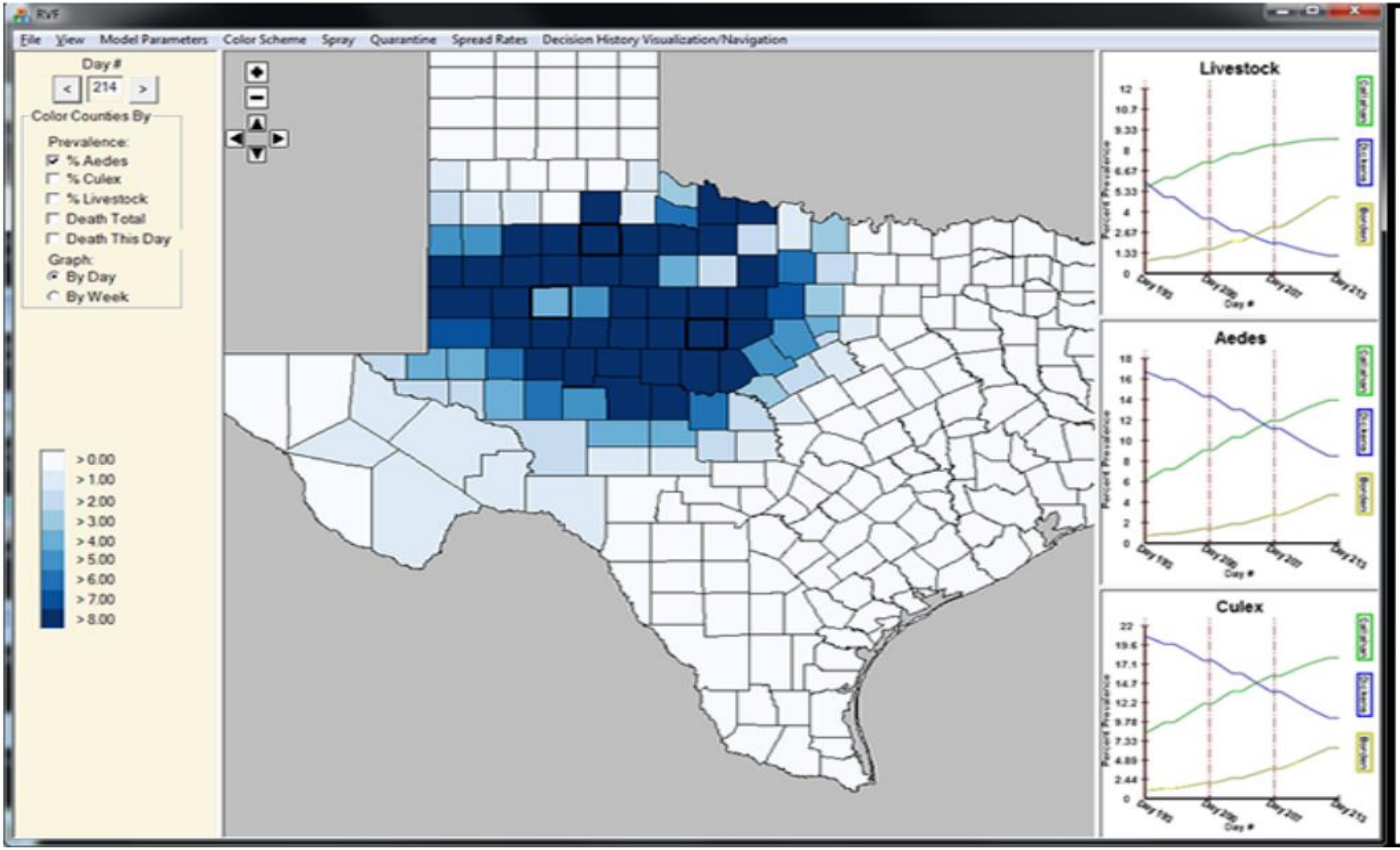
- Core component
- Makes sure entire system is fully functional and adaptable to other models
- Both views are driven by an epidemic spread simulator
- A given model is integrated into the system
- **Input :**
  - Population and demographic data is provided
  - Model input parameters
- **Output:**
  - A large scale spatial simulations
  - Spread data for specified number of days
  - Output information on the number of sick and dead within a given population by areal unit (e.g., county, zip code)
  - Provides color coded geographical representations of the data

Spatiotemporal Model View

Details

- Users can interact with their current scenario
- The spread data is then mapped into an interactive spatiotemporal view
- Exploration of the disease spread through an interactive time spinner control
- Introduce mitigative response measures
- Updation in the epidemiological spread model

Spatiotemporal Model View



Decision history tree

Details

- Modifications in selected scenario are captured in the second view
- Users can explore the different decision paths
- Compare the cumulative outcomes over time
- This view tracks all the mitigative response measures performed
- The corresponding mortality/sickness rate in a form of single visualization
- Also shows the consequences of each decision in terms of net gain or loss over time

Decision history tree

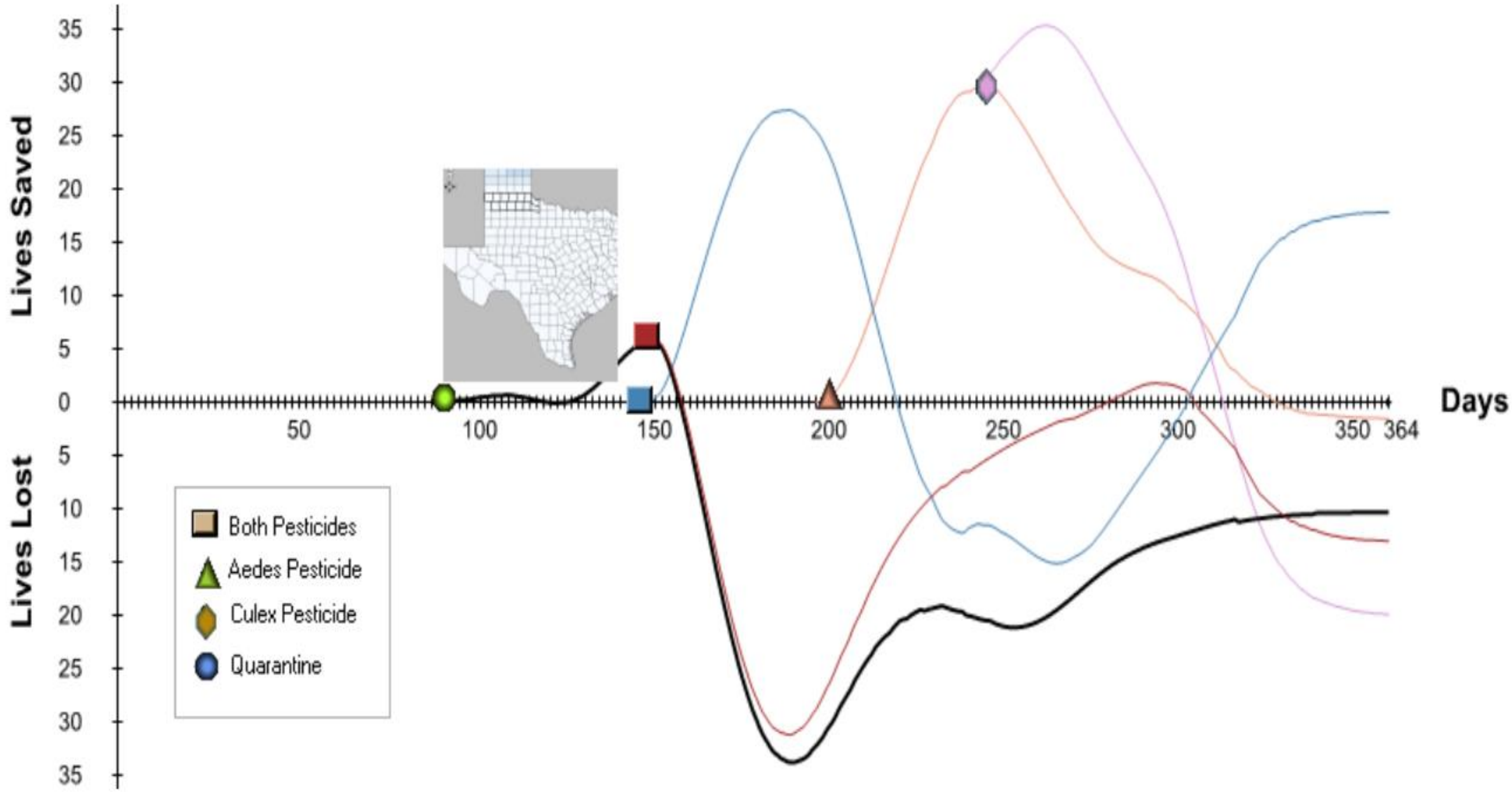


Figure 1: The decision history tree view. As users interact in the model view, the decisions made generate a history tree. Paths of the tree are plotted over time on the x-axis, with the y-axis representing the cumulative deviation from the baseline simulation. Mousing over on a node brings up a thumbnail view of the decision measures implemented at that point in the simulation.

Decision history tree

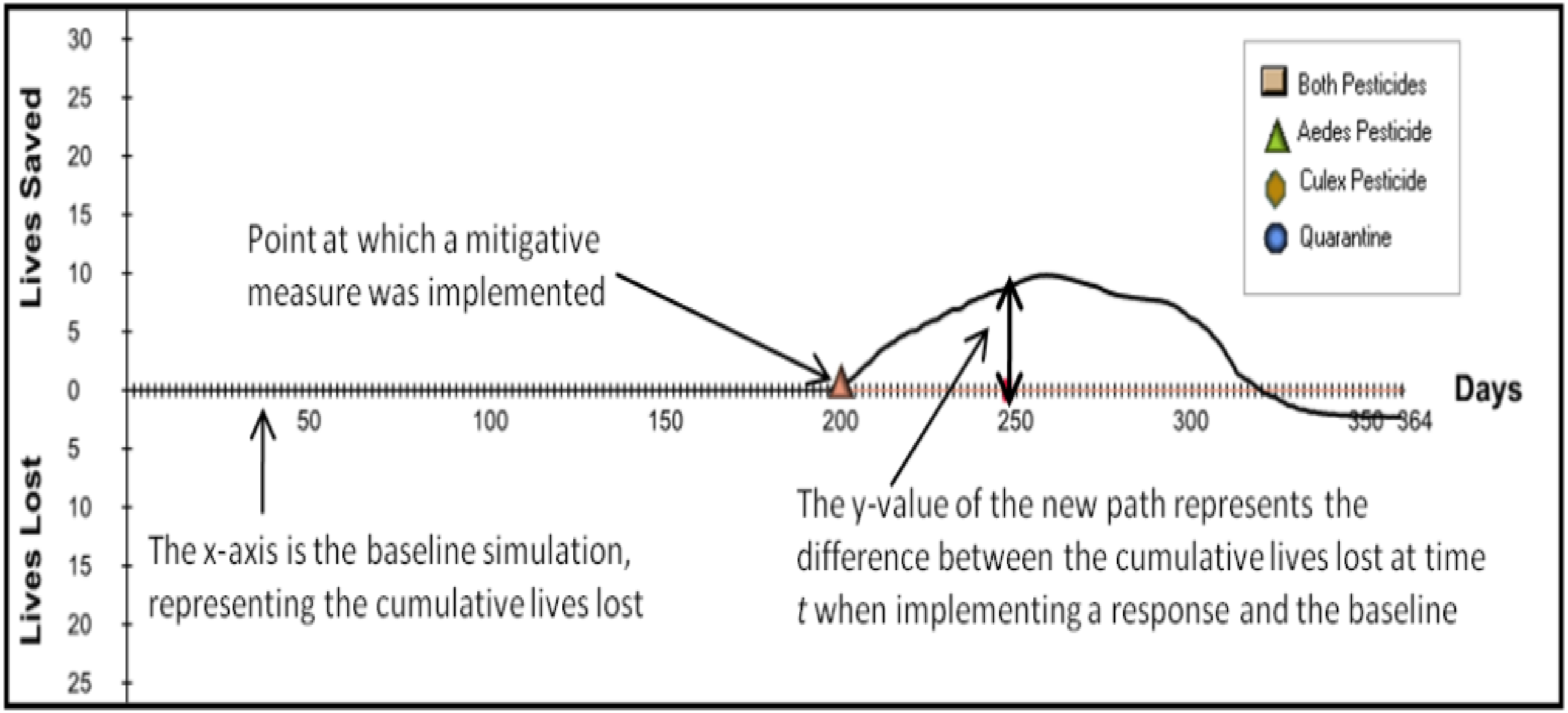


Figure 3: Creating a decision history tree. In this figure, we show the insertion of a decision point on day 200 of a rift valley fever simulation. The insertion of the decision points adds a node, and the resulting colored line shows the cumulative effects of this decision (as compared to the baseline) over time. If a path is above the x-axis, that decision has cumulatively performed better than the baseline up to that point in time. In this manner, users may track the magnitude of the disease spread with respect to the global impact.

Case studies

Pandemic Influenza

- Simulates the spread of a pandemic influenza across the United States starting from a user defined point source location and incorporating airport traffic model
- **Model:** designed to determine the number of influenza outbreak infections, hospitalizations, and deaths on a daily basis.
- **Global decision measures:**
  - School closures
  - Media alerts
  - Strategic national stockpile deployment (SNS).

Pandemic Influenza

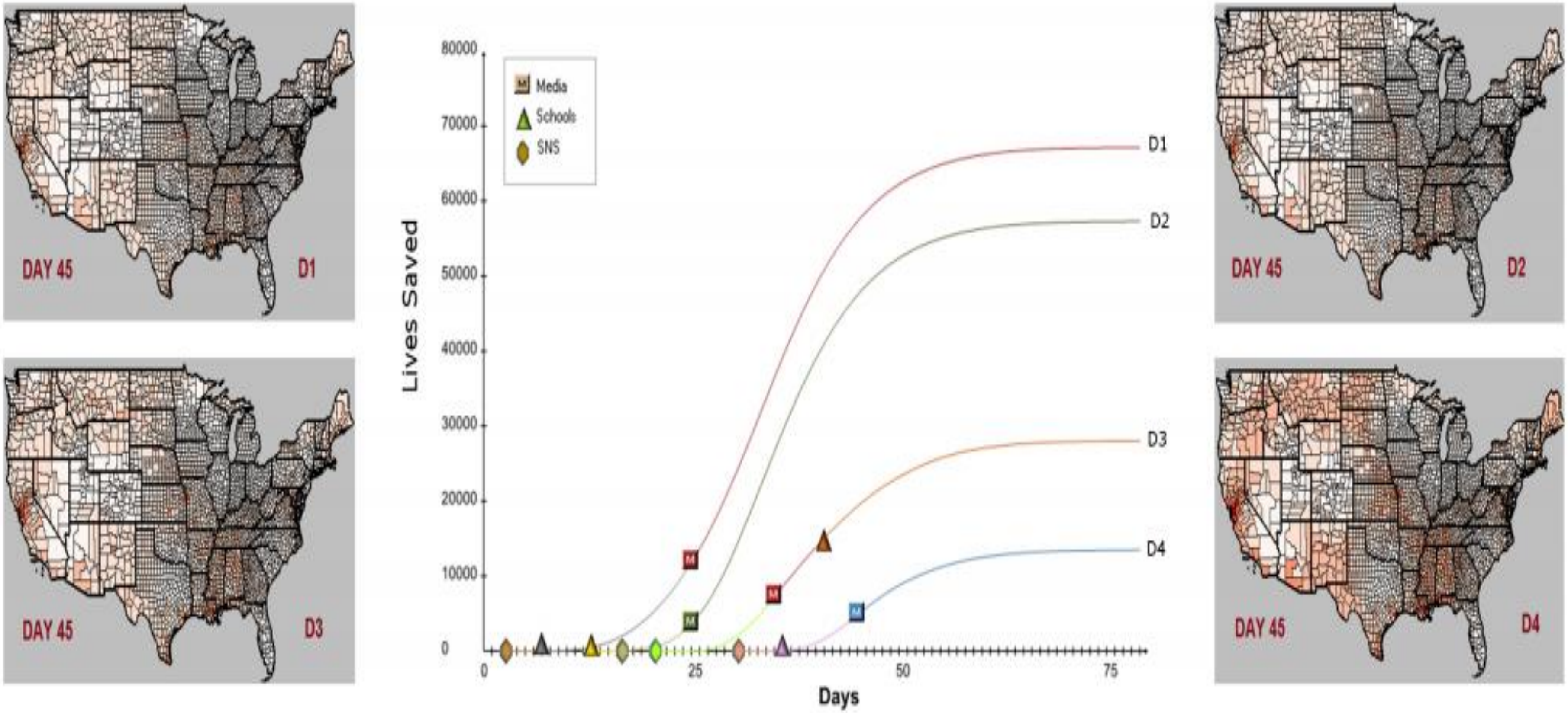


Figure 5: Pandemic Influenza Case Study. Here the user has introduced a variety of different decision measures at various points in time and in different combinatorial order. We explore the resultant simulation spaces in the geographical space with the maps surrounding the central image. Each map corresponds to a different decision tree branch as denoted by the corresponding label.

Case studies

Rift Valley Fever

- Simulates the spread of Rift Valley Fever (RVF) through a simulated mosquito and cattle population in Texas
- **Model:** then accounts for the transmission of the disease both through mosquito to cattle infections and cattle to cattle infections.
- **Local decision measures:**
  - Pesticides
  - Quarantine

Rift Valley Fever

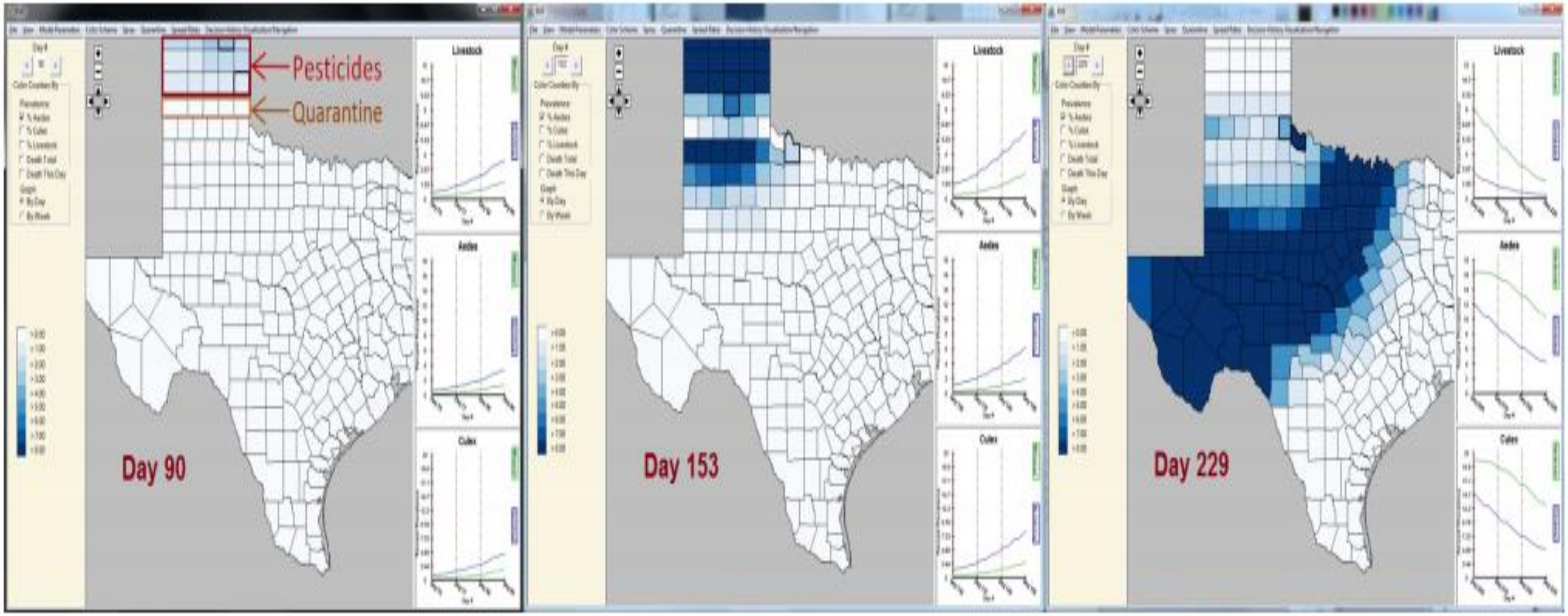


Figure 6: Here we illustrate the effects of utilizing decision measures within the confines of the rift valley fever simulation. In the left image, the analyst has employed both the use of quarantine and pesticide spray to try and reduce the disease spread. However, as infected mosquito eggs have already propagated to neighboring counties, they find that the decision measures taken have less impact then expected. The spread of the disease after the application of pesticides and quarantine is seen in the middle and right figures.

Rift Valley Fever

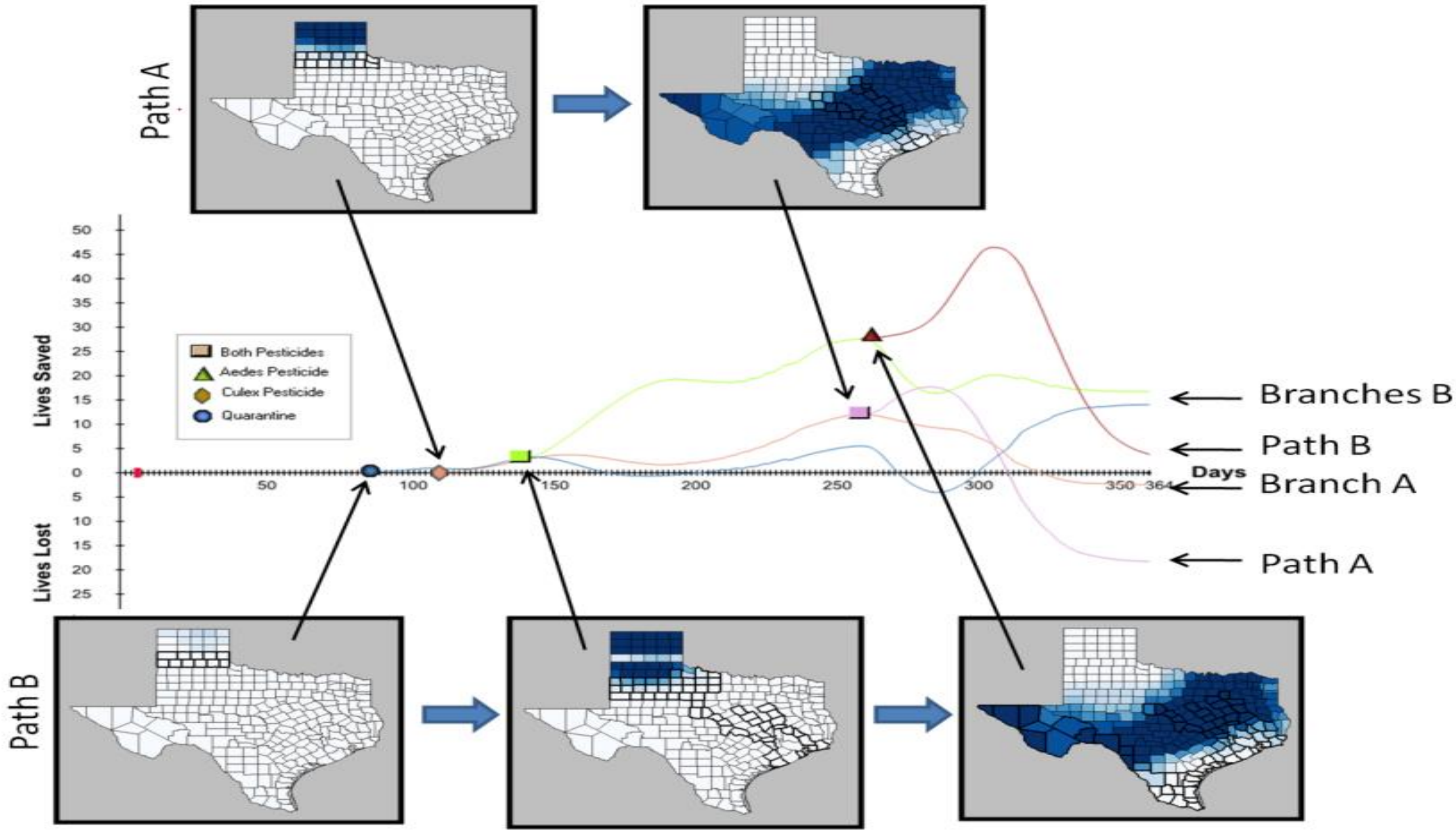


Figure 7: Rift Valley Fever Case Study. Here the user has introduced a variety of different decision measures at various points in time and in different combinatorial order. We explore the resultant simulation spaces in the geographical space with the maps surrounding the central image. Each map corresponds to a different decision tree branch as denoted by the corresponding label.

Outline

- Epidemiology
- Evidence based decision making
- Visual Analytics
- Tool : Decision Support Environment
- Tool : IVA framework
  - By *Paul Klemm et al.*
- Summary

What are the Prior & Related Works?

Prior & Related Work

- A relational database to organize cohort study data for Visual Analysis – by *Steenwijk et al.*
- Hypothesis generation based on descriptive statistic of the data dimensions – by *Turkey et al.*
- Data organization for an interactive visual analysis of heterogeneous cohort study data – by *Angelelli et al.*

What is the proposed?

Interactive Visual Analysis (IVA)

- IVA approach that enable epidemiologists to rapidly investigate the entire data pool for hypothesis validation and generation
- IVA approach helps in combined analysis of image and non-image data

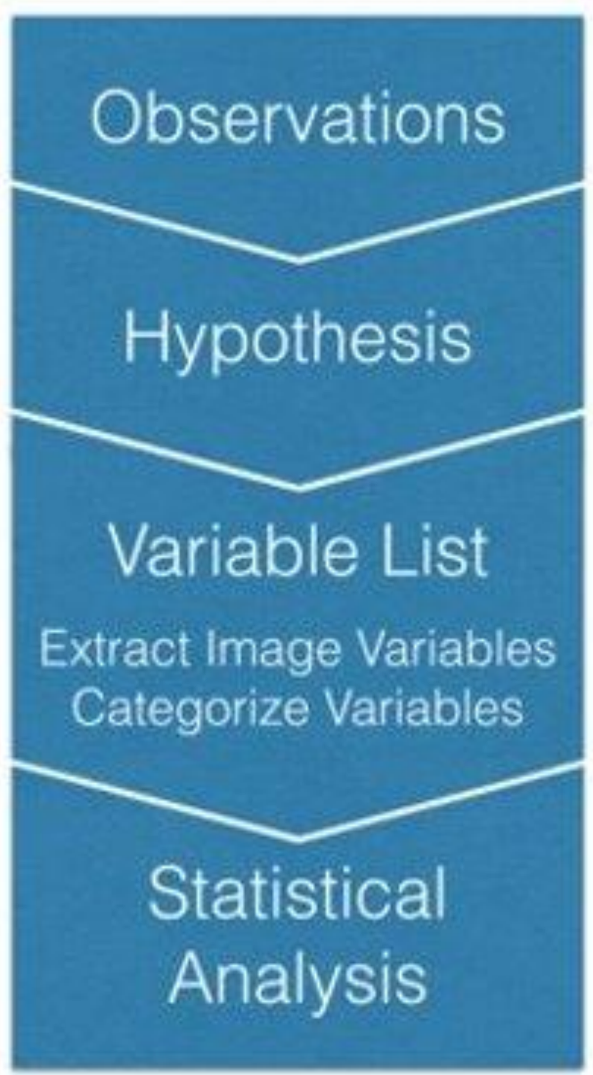
Why Interactive Visual Analysis?

Uniqueness

- Incorporating 3D models
- Hypothesis generation
- Visualization

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG | INF | FAKULTÄT FÜR INFORMATIK

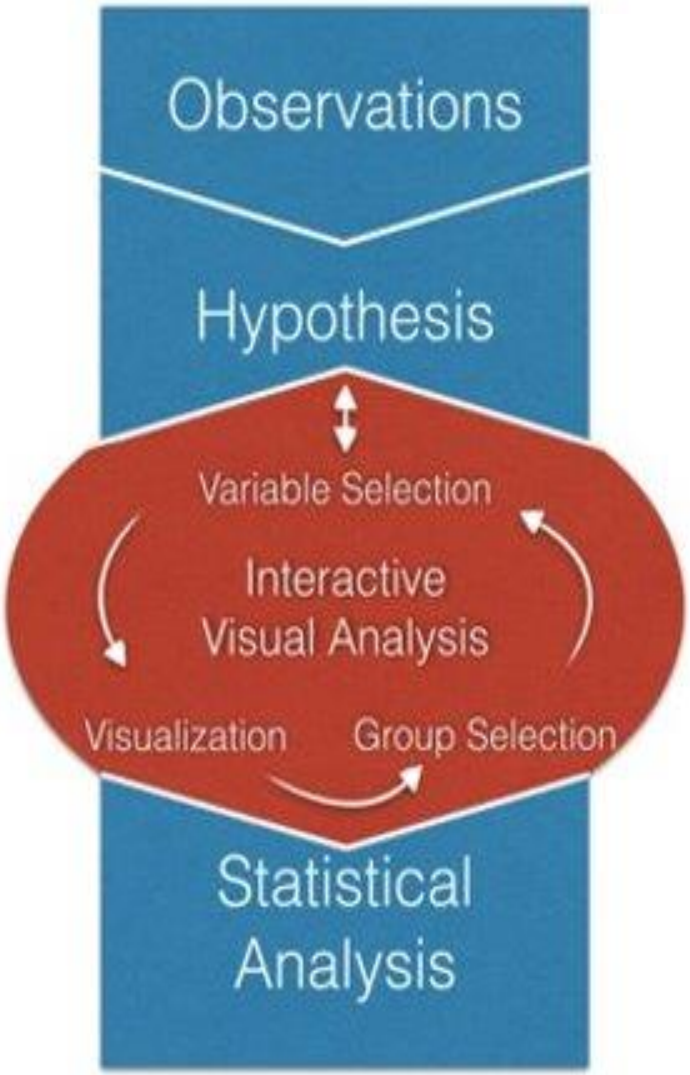
Standard Epidemiology Workflow



Workflow

- A hypotheses is derived from observations made by clinicians in their daily routine
- A set of variables depicting conditions affected by the hypothesis is compiled accordingly
- Confounding variables are identified and taken into an account
- Statistical methods such as regression analysis, assess the association of selected variables with the investigated disease

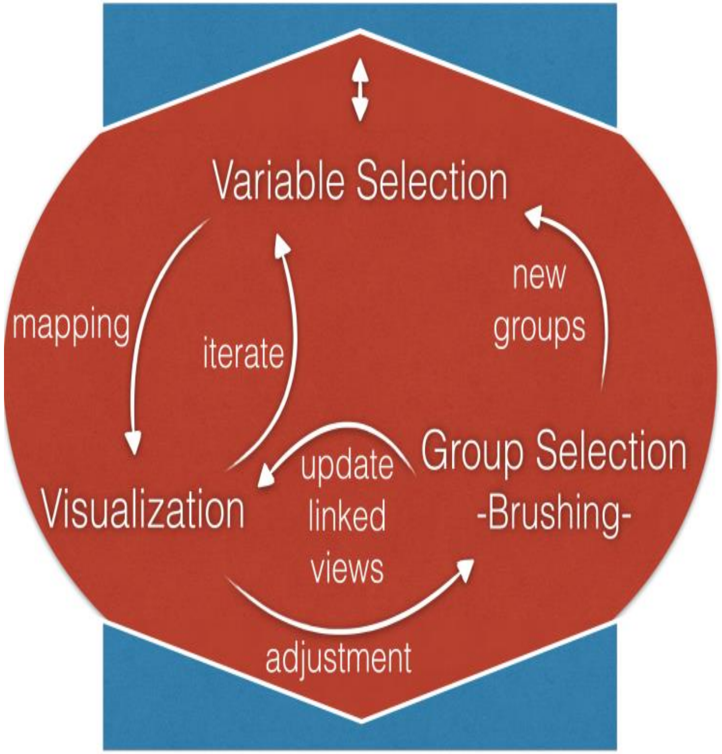
IVA Tools in Epidemiology Workflow



Details

- The IVA tools complement parts of this workflow instead of replacing them
- The combination of statistical and interactive analysis shows promising potentials to unveil information in the data
- The iterative red highlighted part is called as the IVA Loop

IVA Loop – Analysis Workflow



Workflow

- Starts with a **variable selection** of interest (user-driven or via data mining techniques)
- Data are mapped using **visualization** techniques which are appropriate for the selected data types
- The data are then visualized in range and domain space, which can be **brushed**, yielding new groups to be investigated using further variables

These all adjacent steps are directly connected via feedback loops, allowing for an iterative refinement and giving as much freedom to the users as possible

System Design

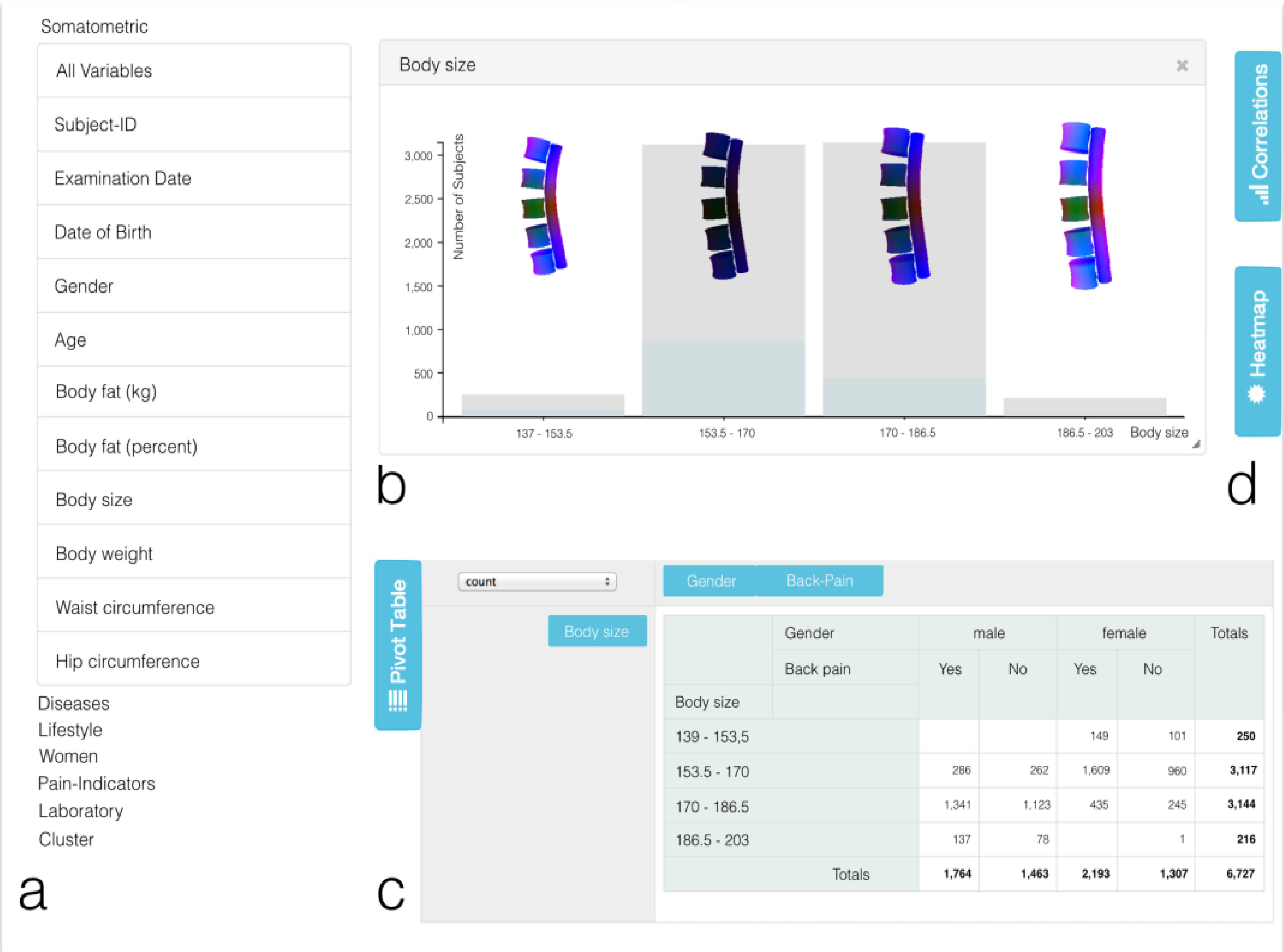
Design & Visualization Techniques

- Due to large spatial distance toward each other, online communication is required, hence system is built using web technologies
- Since IVA Workflow allows various different ways to analyze the data, the interface is designed as minimalistic as possible, treating the resulting space as the canvas for the data

System Design

IVA Framework

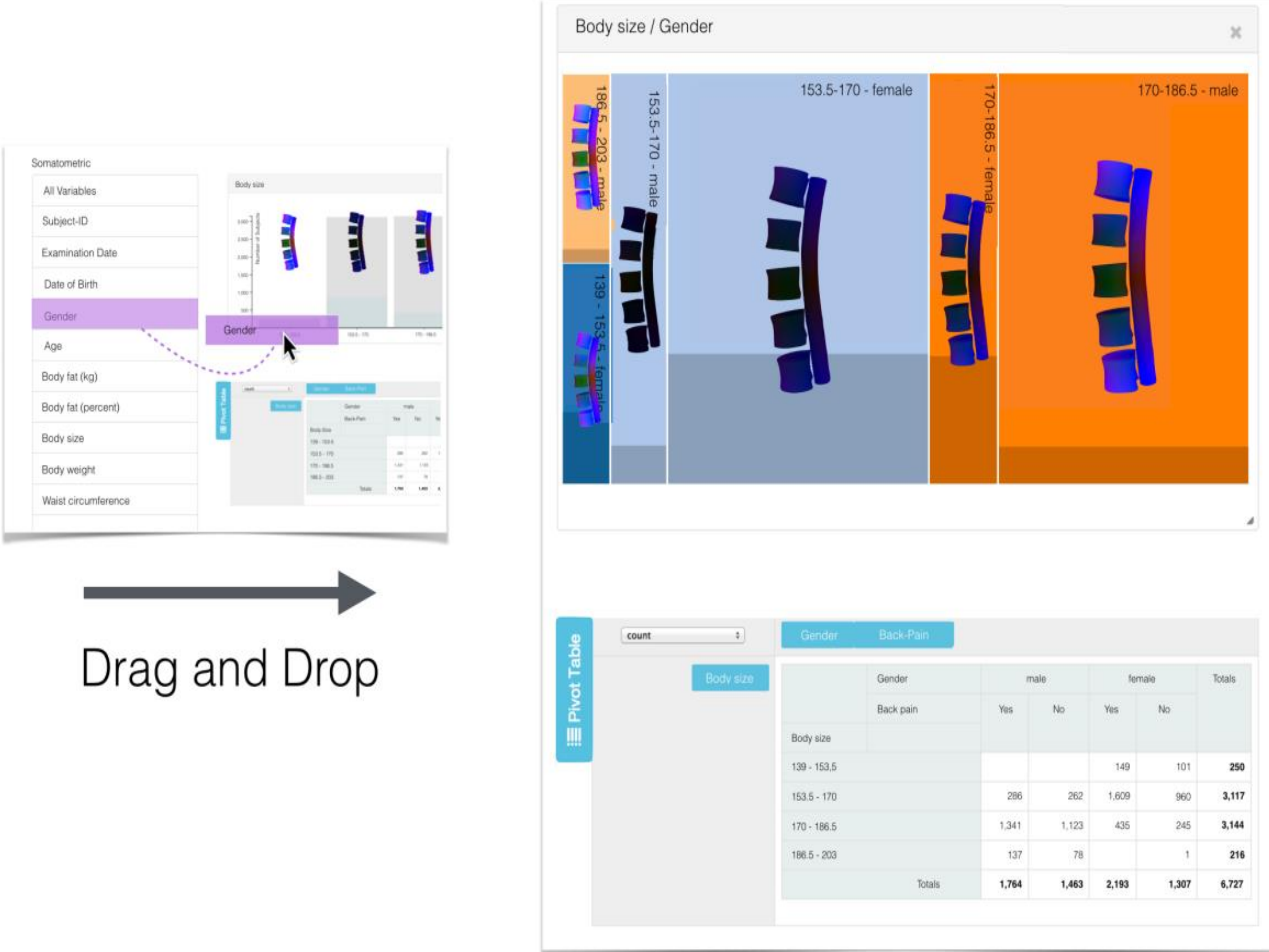
- The **sidebar** contains all epidemiological variables. The cluster results group variables like categorical variables and are part of the side bar.
- The **canvas** holding the visualizations. Elements can be added, arranged, resized and removed freely
- The interactive **pivot table** gives detailed numerical information of the variables in the canvas view. This view on data is familiar to the epidemiologists
- The **contingency view** depicts relations of variables in the canvas in a contingency matrix and a contingency list


Somatometric
All Variables
Subject-ID
Examination Date
Date of Birth
Gender
Age
Body fat (kg)
Body fat (percent)
Body size
Body weight
Waist circumference
Hip circumference
Diseases
Lifestyle
Women
Pain-Indicators
Laboratory
Cluster
a
Body size
Number of Subjects
3,000
2,500
2,000
1,500
1,000
500
0
137 - 153.5
153.5 - 170
170 - 186.5
186.5 - 203
Body size
b
Correlations
Heatmap
d
Pivot Table
count
Gender
Back-Pain
Body size
Gender
male
female
Totals
Back pain
Yes
No
Yes
No
Body size
139 - 153,5
149
101
250
153.5 - 170
286
262
1,609
960
3,117
170 - 186.5
1,341
1,123
435
245
3,144
186.5 - 203
137
78
1
216
Totals
1,764
1,463
2,193
1,307
6,727
c

IVA User Interface

IVA Framework User Interface

- (a) - The sidebar containing all variables as well as the groups defined in the analysis process
- (b) - The canvas area where variables can be added via drag and drop and the visualization is chosen automatically according to the data type
- (c) - The interactive pivot table showing the exact numbers for each displayed variable combination
- (d) - The buttons to open panes containing the contingency matrix, contingency pane and pivot table.



Somatometric
All Variables
Subject-ID
Examination Date
Date of Birth
Gender
Age
Body fat (kg)
Body fat (percent)
Body size
Body weight
Waist circumference
Body size
Gender
Drag and Drop
Body size / Gender
186.5 - 203 - male
153.5-170 - male
153.5-170 - female
170-186.5 - female
170-186.5 - male
139 - 153.5 - female
Pivot Table
count
Gender
Back-Pain
Body size
Gender
male
female
Totals
Back pain
Yes
No
Yes
No
Body size
139 - 153,5
149
101
250
153.5 - 170
286
262
1,609
960
3,117
170 - 186.5
1,341
1,123
435
245
3,144
186.5 - 203
137
78
1
216
Totals
1,764
1,463
2,193
1,307
6,727

Applications

Details

- **Lumbar Spine Data Set** – one of the most common disease in Western Civilization
- Epidemiological analysis of lumbar back pain is largely focused on non-image information
- Two task were then performed on this data set by the 2 experts - One is a **specialist in internal medicine** and the other is a **radiologist**

Tasks

- Hypothesis-free Analysis
- Hypothesis-driven Analysis

Hypothesis–free Analysis

Analysis By Radiologist

- **Lumbar Spine Data Set** – one of the most common disease in Western Civilization
- Epidemiological analysis of lumbar back pain is largely focused on non-image information
- Two task were then performed on this data set by the 2 experts - One is a **specialist in internal medicine** and the other is a **radiologist**

Case 1: Hypothesis-free Analysis

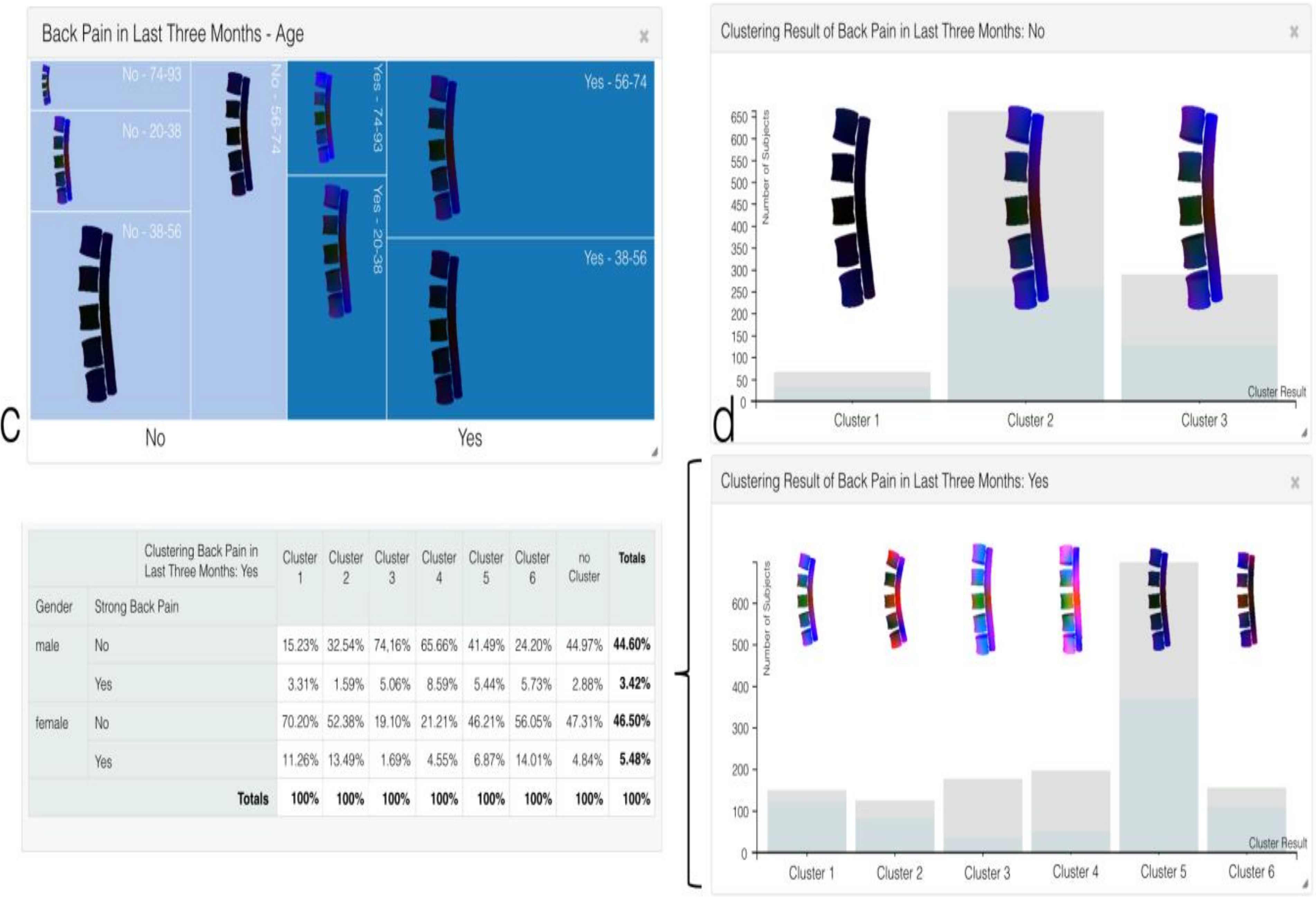
Figure Depictions

Figure (a)

- Mean curvature of lumbar spine canal plotted against the mean shape of 58-74 years old female subjects (light-blue bars)
- The high amount of this subject group relative to the total count in the third group
- The last group contains four outliers

Figure (b)

- Clustering of all subjects yields seven groups, whereas Cluster 4 assembles the mean
- The light blue bars indicate the share of females in the group



	Clustering Back Pain in Last Three Months: Yes	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	no Cluster	**Totals**
Gender	Strong Back Pain								
male	No	15.23%	32.54%	74,16%	65.66%	41.49%	24.20%	44.97%	**44.60%**
	Yes	3.31%	1.59%	5.06%	8.59%	5.44%	5.73%	2.88%	**3.42%**
female	No	70.20%	52.38%	19.10%	21.21%	46.21%	56.05%	47.31%	**46.50%**
	Yes	11.26%	13.49%	1.69%	4.55%	6.87%	14.01%	4.84%	**5.48%**
	Totals	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Case 2: Hypothesis-based Analysis

Figure Depictions

Figure (c)

- A mosaic plot mapping age against the dichotomous questionnaire answer to “Did you experience back pain in the past three months?”.

Figure (d)

- Clustering result of “Did you experience back pain in the past three months?” Yes/no with female share in each group.
- Cluster 1 and 6 for answer “Yes” contain mostly women.
- The pivot table shows how many subjects with strong back pain are in each cluster for answer “Yes”
- Subjects in Cluster 1, 2 and 6 report strong back pain more often than subjects in other clusters.

Outline

- Epidemiology
- Evidence based decision making
- Visual Analytics
- Tool : Decision Support Environment
- Tool : IVA framework
- Summary

Summary

Epidemiology

- Reliable knowledge for clinical medicine
- Focuses on prevention, diagnosis and treatment of diseases

Evidence Based decision making in public health

- Resources in public health always are limited
- Evidence for numerous public health interventions are growing
- Choices increasingly complex
- Rationally choose among alternatives and make the most prudent use of resources, stronger skills in EBPH are needed
- These are a blend of art and science.

Summary

Visual Analytics

- A sophisticated graphical representation for complex data set
- Multi-Panel Graphs is a powerful tool for public health and disease bio-surveillance efforts

Tool : Decision Support Environment

- A suite of predictive visual analytic tools
- Provides insight into the effectiveness of a decision
- For the investigation of multiple courses of responses
- Comparisons of the effectiveness of each component of a response plan

Tool : IVA framework

- Focuses on the closing the gap between data complexity and analyzability
- Allowing the domain experts to dig deep into the data and potentially obtain unexpected findings

References

[1] Afzal, S. et al. "Visual analytics decision support environment for epidemic modeling and response evaluation." IEEE Visual Analytics Science and Technology (VAST), 2011.

[2] Brownson, Ross C., James G. Gurney, and Garland H. Land. "Evidence-based decision making in public health." Journal of Public Health Management and Practice 5 (1999): 86-97.

[3] Chui, Kenneth KH, et al. "Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs." PloS one 6.2 (2011): e14683.

[4] Klemm, P. et al. (2014). Interactive visual analysis of image-centric cohort study data. IEEE transactions on visualization and computer graphics, 20(12), 1673-1682.

[5] Preim, B., et al. (2016). Visual analytics of image-centric cohort studies in epidemiology. In Visualization in Medicine and Life Sciences III (pp. 221-248). Springer

Thank you!

Additional Content

Randomized Control Trials	Cross Sectional Study	Quasi Experiment
Study	Observational study	Empirical Interventional Study
People are allocated at random	Analyses data from population or representative subset	Target population
To receive one of the several clinical intervention	At specific point in time	Used to estimate casual impact of an intervention without random assignment
Sugar pill tested on certain group of people		