W3C Using Open Data Workshop
Visual Analytics for Policy-Making
Opportunities and Research Challenges

Francesco Mureddu, Tech4i2.com

www.crossover-project.eu

#pmod

BACKGROUND

- CROSSOVER project "Bridging Communities for the Next Generation Policy Making" aims to create a new research Roadmap on ICT Tools for Governance and Policy Making
- The roadmap will be built on the CROSSROAD roadmap project, to be updated and expanded by adding a global perspective as well as more practical cases and examples
- Roadmap to be demand driven and calibrated on users' and practitioners' needs, improved and validated with input and feedback collected in online and offline activities
- CROSSOVER roadmap entails:
  - GC1 - Model-based Collaborative Governance
  - GC2 - Data-powered Collective Intelligence and Action

Demand Driven Approach

Research Challenge	Benefits Stemming from Policy Making 2.0 Tools	Phase in the Policy Cycle
	• Anticipate unpredictable, long-term and indirect policy impacts • Identify unknown-unknowns • Include human emotions and network effects in assessing impacts • Get fast real-time impact on what is happening to prevent policy delays • Manage and anticipate crisis • Include in debates also people not interested in politics • Generate new ideas from new participants • Raise awareness on the long-term impact of behaviours • Understand the impact of policy	• Agenda setting (participatory sensing, modeling, opinion mining,visual analytics) • Policy design (collaborative policy-making, simulation) • Policy implementation (persuasive technologies, collaborative action) • Policy evaluation (open data, opinion mining, visual analytics)

GC2 Data-powered Collective Intelligence and Action

DEFINITION

- The explosion in computing techniques led to the generation of a tremendous amount of data stored in the internet and processed in the IT infrastructures
- Visualisation is a way to interpret and translate data from computer understandable formats to human ones by employing graphical models, charts, graphs, etc…
- Visualisation is the analytical process output, while visual analytics is a dynamic tool that aims at integrating human capabilities of visual information exploration and the enormous processing power of computers
- Visual analytics combines human perception and computing power in order to solve the information overload problem
- Visual analytics is then useful for tackling the increasing amount of data available, and for using in the best way the information contained for the policy making process
- Visual analytics can be seen as an integral approach to decision-making, combining visualization, human factors and data analysis

Challenges in Governance

- Make sense of the increasing amount of data available (open data, big data) which information can be used for the policy making process
- Dig into data in order to extract patterns and validate models
- Have a more evidence-based policy debate
- Raise awareness and engage non specialist in policy making
- Raise awareness of invisible issues (e.g. to avoid short-termism) and trigger action
- Detect the unknown's unknown – answers to questions you didn't ask
- Personal behaviour visualisation leading to potential action and behavioural change
- Tackle complex and non-predictable patterns to assess and anticipate policy impact

Possible Solutions for Policy Making (1)

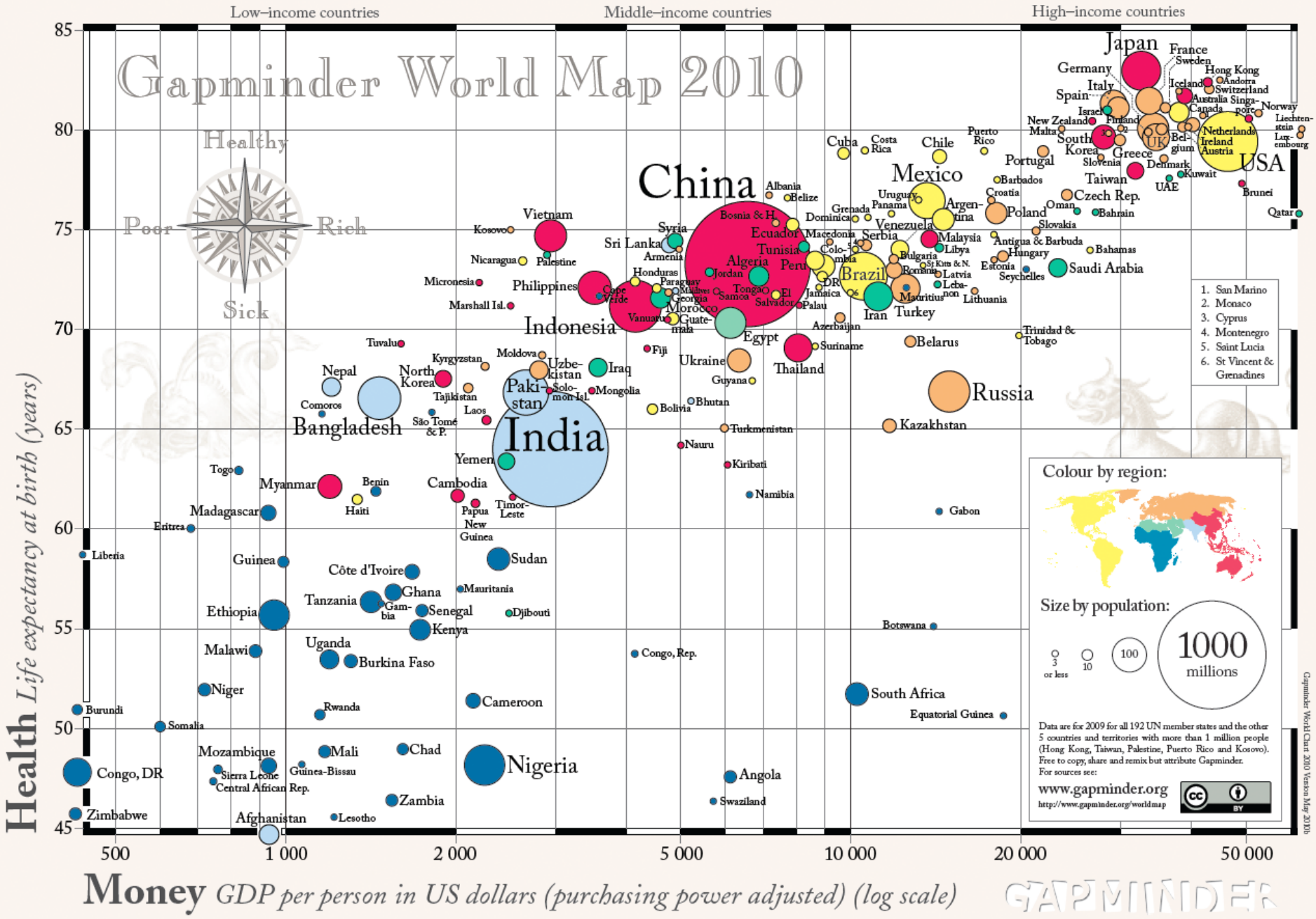
- **Economic data visualization** for visually exploring, assessing and forecasting trends, as well as for observing the localization of economic activities, especially for regional economic analysis
- **Visualization of health data**, which allows to display displaying historical trends, incidence rates, rate/trend comparison, screening, risk factors
- **Demographics visualizations**, allowing stakeholders and decision makers to have a clear picture of data and trends over time
  - Easier design and evaluation of various policies, as advanced algorithms are able to create figures and illustrations
- **Legal Arguments visualisation**: text analysis, argumentation mappings and visualisation algorithms can be applied to legal documents
  - Legislation accessible and comprehensible to the public
  - Possibility to visually represent corroborative evidence

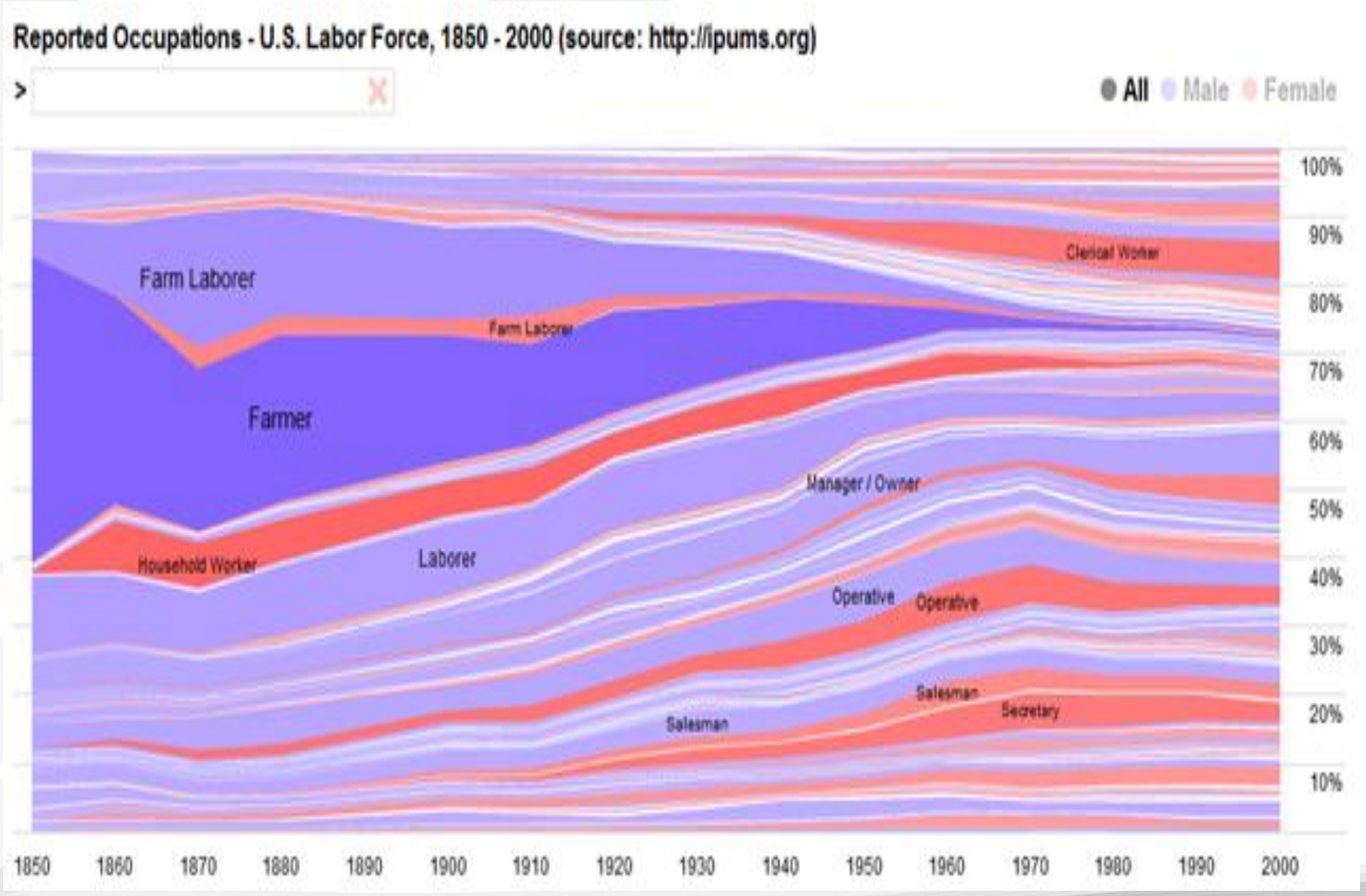
POSSIBLE SOLUTIONS FOR POLICY MAKING (2)

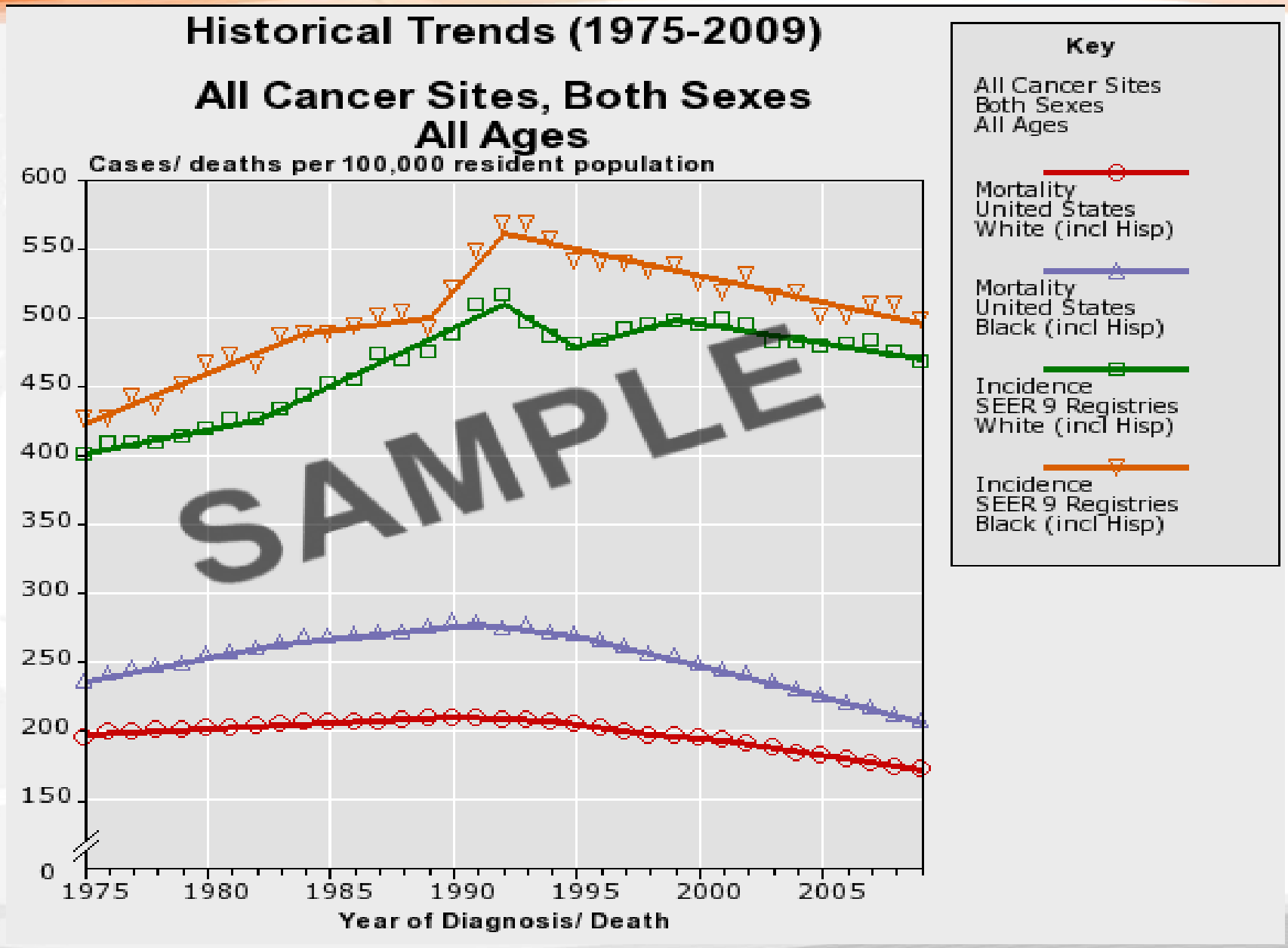
- **Discussion Arguments** visualisation of the flow of a discussion in order to instantly get awareness of the topics discussed, as well as of the arguments and the support such arguments gain
  - Stakeholders understand the flow of a discussion, which is presented to them in a structured and interactive format, avoiding numerous discussion threads
- **Geovisualization** based on the provision of tools and methods for visual analysis, exploration and representation of geo data/info
  - Possible to derive problem specific models and design task specific maps for incorporating geographical knowledge into planning and decision making
- Advanced visualization applications used for **security and national defense**, developed both on the military and on the corporate front
  - Business organizations also have urgent information visualization requirements that support their intelligence and situational awareness capability, data mining and reporting requirements
  - Software innovations are also applicable to the defense domain due to common data mining and information visualization challenges

INSPIRING CASES (1)

- **GapMinder**, an online free tool for visualizing global data through dynamic graphs along a time series
  - Built on the Trendalyzer platform uses data from global institutions like the OECD, World Bank, and the International Labor Organization
- **US Labour Force Visualization** is a on online tool able to decompose workforce according to genre and occupation's over time (1950- 2000)
  - Its charts clearly confirm the stylized fact according to which there has been a structural change in US economy in the last century
- **State Cancer Profiles** is an online free tool providing dynamic cancer statistics boosting cancel control efforts in the US states and counties
  - Provides interactive maps displaying historical trends, death/incidence rates, rate/trend comparison, 5 years rate changes, screening, risk factor table and demographic data table
- **Instant Atlas** is a data visualization and presentation software solution used for displaying location-based statistical data contained in spreadsheets and desktop Geographic Information System software
  - Interactive dynamic and profile reports combining statistics and map data


Low–income countries
Middle–income countries
High–income countries
Gapminder World Map 2010
Healthy
Poor
Rich
Sick
China
India
Indonesia
Bangladesh
Pakistan
Nigeria
Brazil
Russia
USA
Japan
Mexico
Health Life expectancy at birth (years)
85
80
75
70
65
60
55
50
45
500
1 000
2 000
5 000
10 000
20 000
50 000
Money GDP per person in US dollars (purchasing power adjusted) (log scale)
1. San Marino
2. Monaco
3. Cyprus
4. Montenegro
5. Saint Lucia
6. St Vincent & Grenadines
Colour by region:
Size by population:
3 or less
10
100
1000 millions
Data are for 2009 for all 192 UN member states and the other 5 countries and territories with more than 1 million people (Hong Kong, Taiwan, Palestine, Puerto Rico and Kosovo). Free to copy, share and remix but attribute Gapminder.
For sources see:
www.gapminder.org
http://www.gapminder.org/worldmap
Gapminder World Chart 2010 Version May 2010b
GAPMINDER


Reported Occupations - U.S. Labor Force, 1850 - 2000 (source: http://ipums.org)
All
Male
Female
Farm Laborer
Farmer
Farm Laborer
Clerical Worker
Household Worker
Laborer
Manager / Owner
Operative
Operative
Salesman
Salesman
Secretary
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000


Historical Trends (1975-2009)
All Cancer Sites, Both Sexes
All Ages
Cases/ deaths per 100,000 resident population
600
550
500
450
400
350
300
250
200
150
0
1975
1980
1985
1990
1995
2000
2005
Year of Diagnosis/ Death
SAMPLE
Key
All Cancer Sites
Both Sexes
All Ages
Mortality
United States
White (incl Hisp)
Mortality
United States
Black (incl Hisp)
Incidence
SEER 9 Registries
White (incl Hisp)
Incidence
SEER 9 Registries
Black (incl Hisp)

Cornwall CrimeExplorer

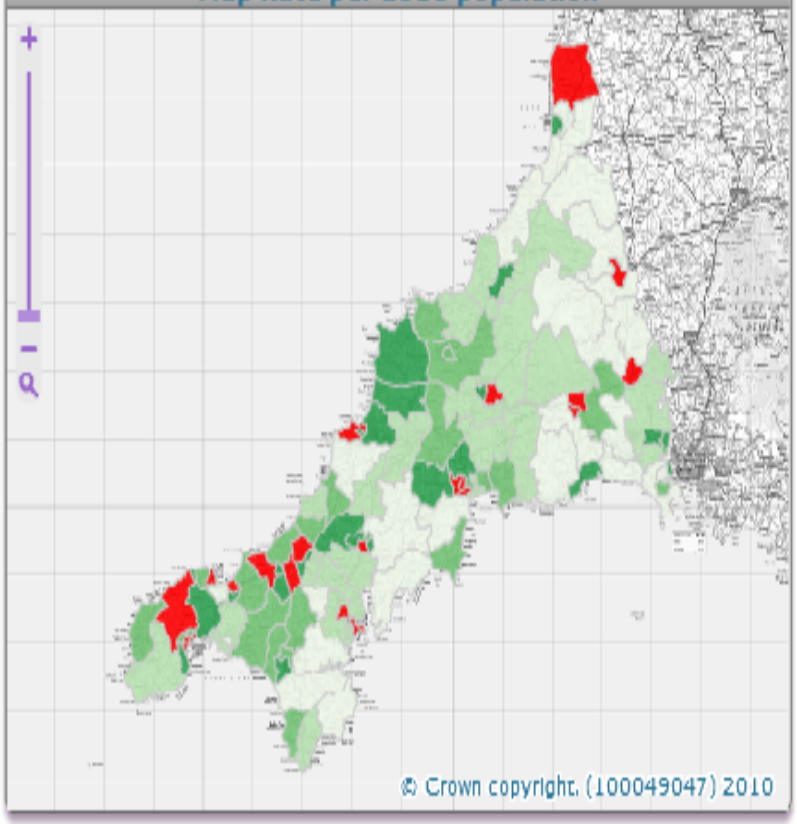
All Crime (rate per 1000) Jan-Mar 11

Filter view:

Amethyst Intelligence

Data | Map / Table | Filter | Notes | Time / Bar Ch... | Exp... | Print | Help

Map Rate per 1000 population



Legend

Wards

0.5 - 4.2

4.3 - 5.9

6.0 - 8.5

8.6 - 14.3

14.4 - 35.0

Comparison

Area	Rate

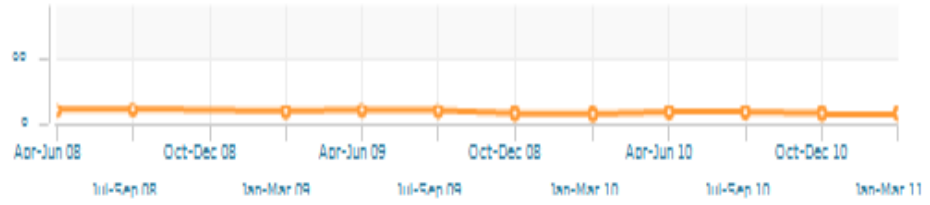
Putting the data into context

ALL CRIME

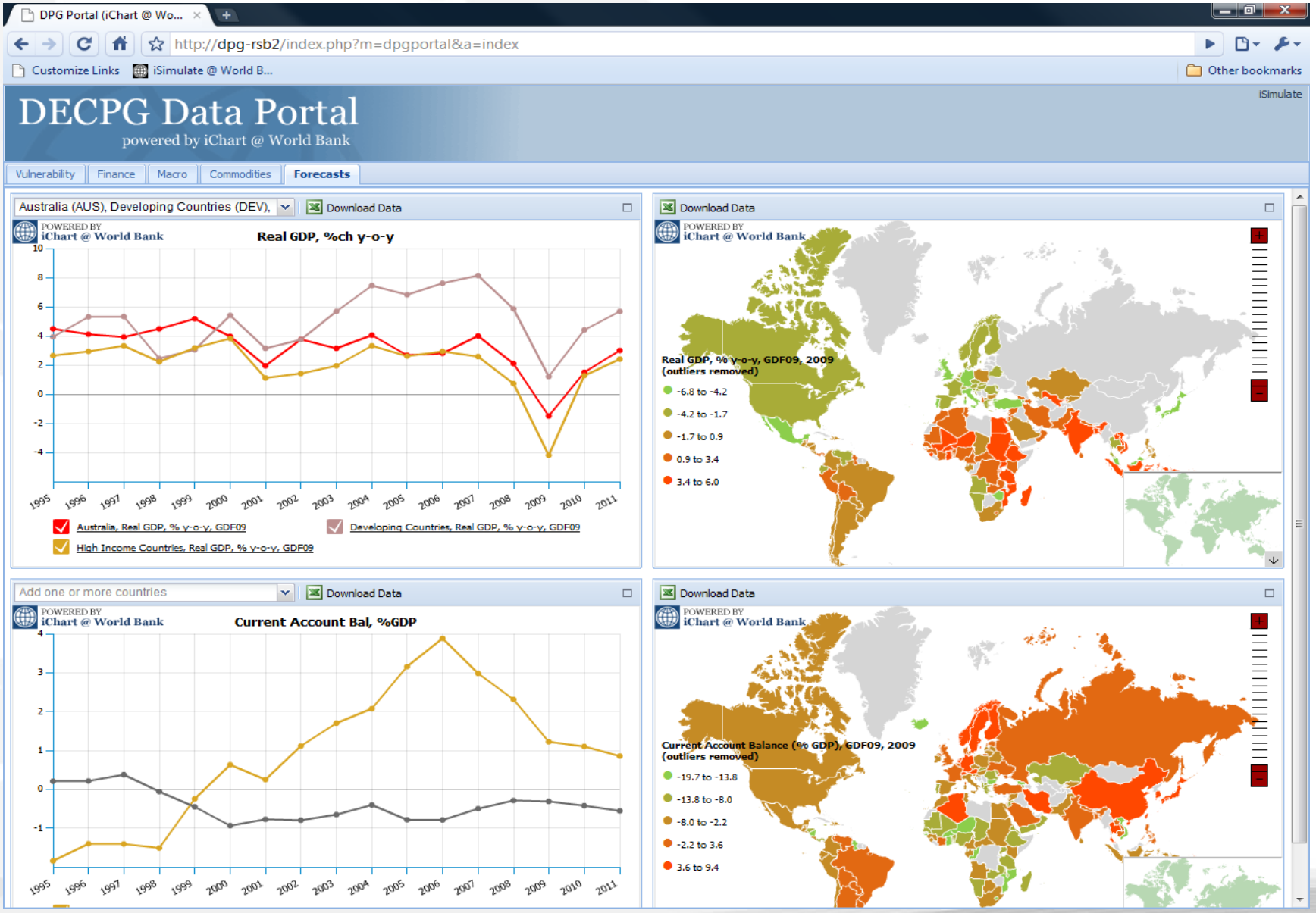
Comments:

- Cornwall has the lowest levels of 'all-crime' compared to its family group of similar policing areas nationally.
- There were 27,962 recorded crimes in Cornwall in 2008/09, a rate of 52.8 per 1000 population.
- The all-crime rate in Cornwall fell by 5% in 2008/09 compared to the previous year.

Time - Rate per 1000 populations



Statistics


DPG Portal (iChart @ Wo...
http://dpg-rsb2/index.php?m=dpgportal&a=index
Customize Links
iSimulate @ World B...
Other bookmarks
iSimulate
DECPG Data Portal
powered by iChart @ World Bank
Vulnerability
Finance
Macro
Commodities
Forecasts
Australia (AUS), Developing Countries (DEV),
Download Data
POWERED BY iChart @ World Bank
Real GDP, %ch y-o-y
Australia, Real GDP, % y-o-y, GDF09
Developing Countries, Real GDP, % y-o-y, GDF09
High Income Countries, Real GDP, % y-o-y, GDF09
Download Data
POWERED BY iChart @ World Bank
Real GDP, % y-o-y, GDF09, 2009
(outliers removed)
-6.8 to -4.2
-4.2 to -1.7
-1.7 to 0.9
0.9 to 3.4
3.4 to 6.0
Add one or more countries
Download Data
POWERED BY iChart @ World Bank
Current Account Bal, %GDP
Download Data
POWERED BY iChart @ World Bank
Current Account Balance (% GDP), GDF09, 2009
(outliers removed)
-19.7 to -13.8
-13.8 to -8.0
-8.0 to -2.2
-2.2 to 3.6
3.6 to 9.4

Current Research

- Simultaneous multiple visualisation
- Integration of visualisation with comments / wiki / blogs
- Interaction between visualisation , models and simulations
- Mobile visual analytics tools
- Geo-visualisation of government data
- Integration with opinion mining and participatory sensing
- Evaluation framework for visualisation effectiveness
- Visualisation infrastructures for policy modelling issues
- Close the loop of information selection, preparation and visualisation

KEY CHALLENGES AND GAPS

- **Users acceptability**: visualisation tools are still largely design for analyst and are not accessible to non-experts: intuitive interfaces/devices are needed
- Visualisation is largely a demand- and design-driven research area, so a big challenge is to ensure the **multidisciplinary collaboration** of engineering, statistics, computer science and graphic design
- Better **integration** between new opportunities for data collection, i.e. open data and participatory sensing, policy modelling and visual analytics tools
- **Integration** of visual analytics and econometric/ statistic techniques
- Existing visualization/visual analytics tech. to be **adapted** to policy modeling
  - Most applications related to visual analytics of public data remain at the level of visualisation only, with limited analytical functionalities
  - Geo-visualisation is a fast growing application area in the government context, but little integration with other related areas (parti. sensing)
  - Interactive data visualisation tools to foster the dialogue between data analysts and policy makers

VISUAL ANALYTICS AND ECONOMETRICS

- Inspiried by H. Rosling's Speech (2006) Saaibi got the idea to use Google motion charts to visualise R output
- Gesmann and Castillo (2012) developed googleVis, a free package provides an interface between R and the Google Visualisation API interactive charts
- The functions of the package allow the user to visualise R output and data with the Google Visualisation API
- googleVis provides output under the form of html which is then visible directly in the browser
- **GeoDaCenter** project to enhance criminal justice support:
  - Integration of techniques for geospatial visual analytics and spatial econometrics with state of the art geocomputation technologies

Future Research: Short Term Issues

- Reusability of mashup tools (web application combining data from one or more sources into a single integrated tool or application) for visual analytics
- Integration between automatic computation and interactive visualisation, i.e. availability of complex and powerful algorithms for manipulating the data under analysis, transformed to feed suitable visualizations
- Techniques and algorithms to create visualization tools based on perceptual psychology (dealing with process by which the physical energy received by sense organs forms the basis of perceptual experience), cognitive science (focusing on how information is represented, processed, and transformed) and graphical principles
- Interactive exploration techniques (e.g. focus & context), for the viewers to be able to see the object of primary interest in full detail while at the same time getting a overview/impression of surrounding info/context available
- Bias identification and signalling in visualisation
- Impact evaluation of visual analytics on policy choices

Future Research: Long Term Issues

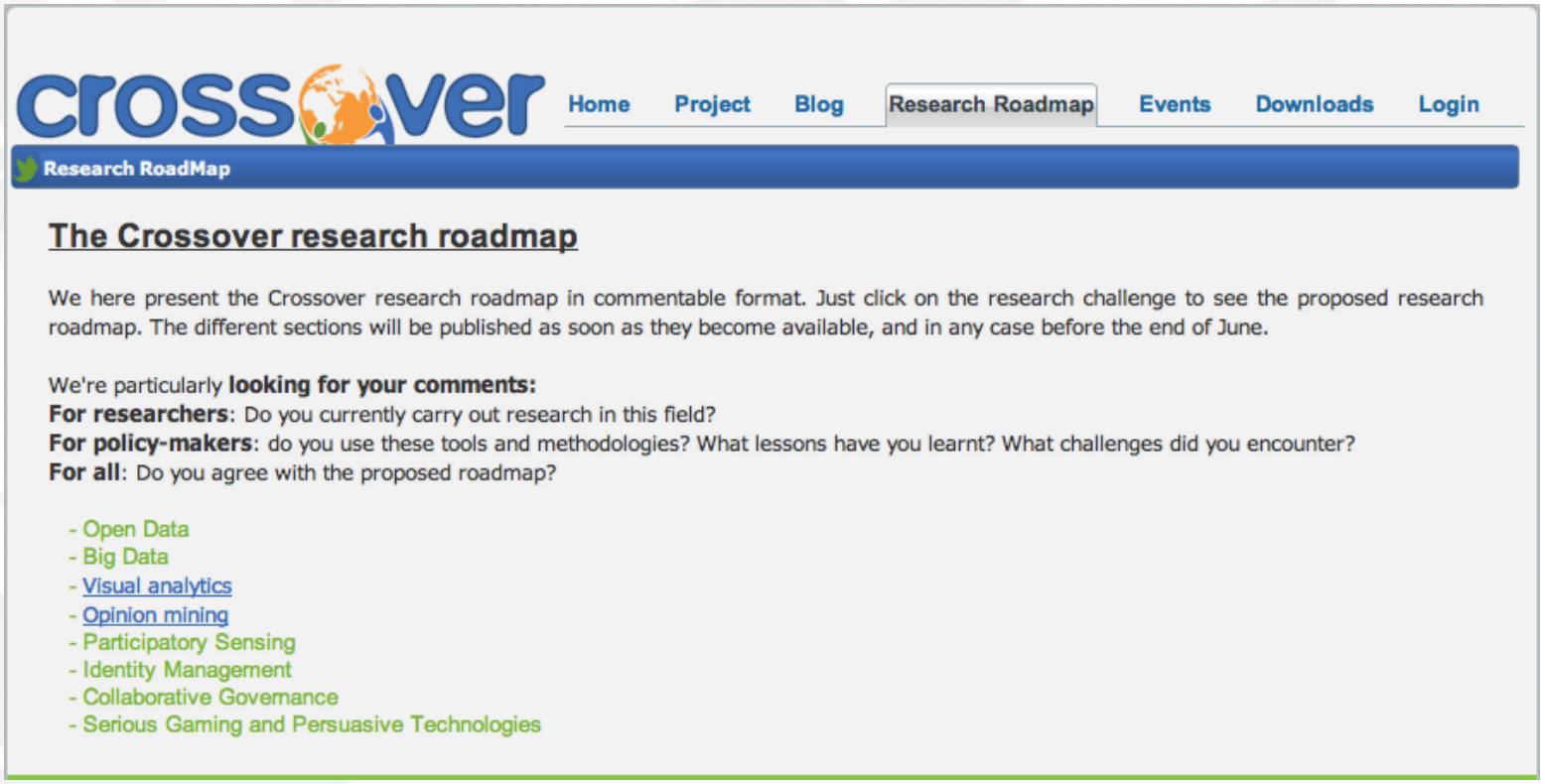
- Learning adaptive algorithm for users intent
  - capable to automatically change behaviour based on its execution context (data handled by the algorithm, configuration parameters of the runtime environment, resources used) to obtain optimal performances
- Advanced visual analytics interfaces, in which analytics and visualization don't need to be advanced in itself but synergy between automation and visualization is in fact advanced
- Development of novel interaction algorithms incorporating machine recognition of the actual user intent and appropriate adaptation of main display parameters such as the level of detail, data selection, etc. by which the data is presented
- Intuitive and affordable visual analytics interface for citizens

SUMMARY OVERVIEW

Market availability	Challenges and Gaps	Current Research	Short Term Future Research	Long Term Future Research
-Information visualization requirements for business intelligence and situational awareness -enterprise knowledge visualization linking -Online analytical processing and data mining -Advanced social network analysis and visualization -Data mining and interactive visualization communication of location-based statistical data -Information visualization tools for high dimensional non-linear data -Visual analysis of data in spreadsheet format	- Demographics visualizations, allowing stakeholders and decision makers to have a clear picture of the data and of their trends over time - Legal Arguments visualisation: text analysis, argumentation mappings and visualisation algorithms - Discussion Arguments visualisation, making use of visualisation techniques for visualizing a discussion's flow -Geographic visualization tools -Financial markets monitoring and visualizing in real time -Advanced applications for security and defense	-Close the loop of information selection, preparation and visualisation -Simultaneous multiple visualisation -Integration of visualisation with comments / wiki / blogs -Collaborative platform display Interaction between visualisation and models -Mobile visual analytics tools -Geo-visualisation of government data -Integration with opinion mining and participatory sensing -Evaluation framework for visualisation effectiveness -Visualisation infrastructures for policy modelling issues	-Re-usable, mashable tools for visual analytics -Tighter integration between automatic computation and interactive visualisation -Bias identification and signalling in visualisation -Perceptual, cognitive and graphical principles -Efficiency of the visualisation techniques to enable interactive exploration interaction techniques such as focus & context -Impact evaluation of visual analytics on policy choices	-Learning adaptive algorithm for users intent -Advanced visual analytics interfaces -Intuitive affordable visual analytics interface for citizens -Development of novel interaction algorithms incorporating machine recognition of the actual user intent and appropriate adaptation of main display parameters such as the level of detail, data selection, etc. by which the data is presented

COMMENT THE ROADMAP

http://www.crossover-project.eu/ResearchRoadmap.aspx



THANK YOU!

Francesco.Mureddu@tech4i2.com

@muredduf