CSCE 590-1: From Data to Decisions with Open Data: A Practical Introduction to AI

Lecture 3: A Systematic AI Approach to Solutions

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

19TH JAN 2021

Carolinian Creed: "I will practice personal and academic integrity."

Organization of Lecture 3

- Introduction Segment
  - Recap of Lecture 2
- Main Segment
  - Real-World Problems – Smart City
  - A Systematic Approach
    - Value of decision: before and after
    - Data-needed
    - Method
    - Evaluation
    - Integrating with overall process
  - Illustrate with **Health**, Water *(next class)*
- Concluding Segment
  - About Next Lecture – Lecture 4
  - Ask me anything

Introduction Segment

Recap of Lecture 2

- We looked at data of different types: by structure, content type and source
- Open data is useful
- Accessing open data is quite easy, Open 311 helps
- Annotations are important for discovery
- Publishing good data is important for useful AI

Main Segment

Real World Problems

- Domains
  - Health
  - Water
  - Traffic
  - Food
  - Energy
  - …

- What can AI do here?

Basics: Smart City

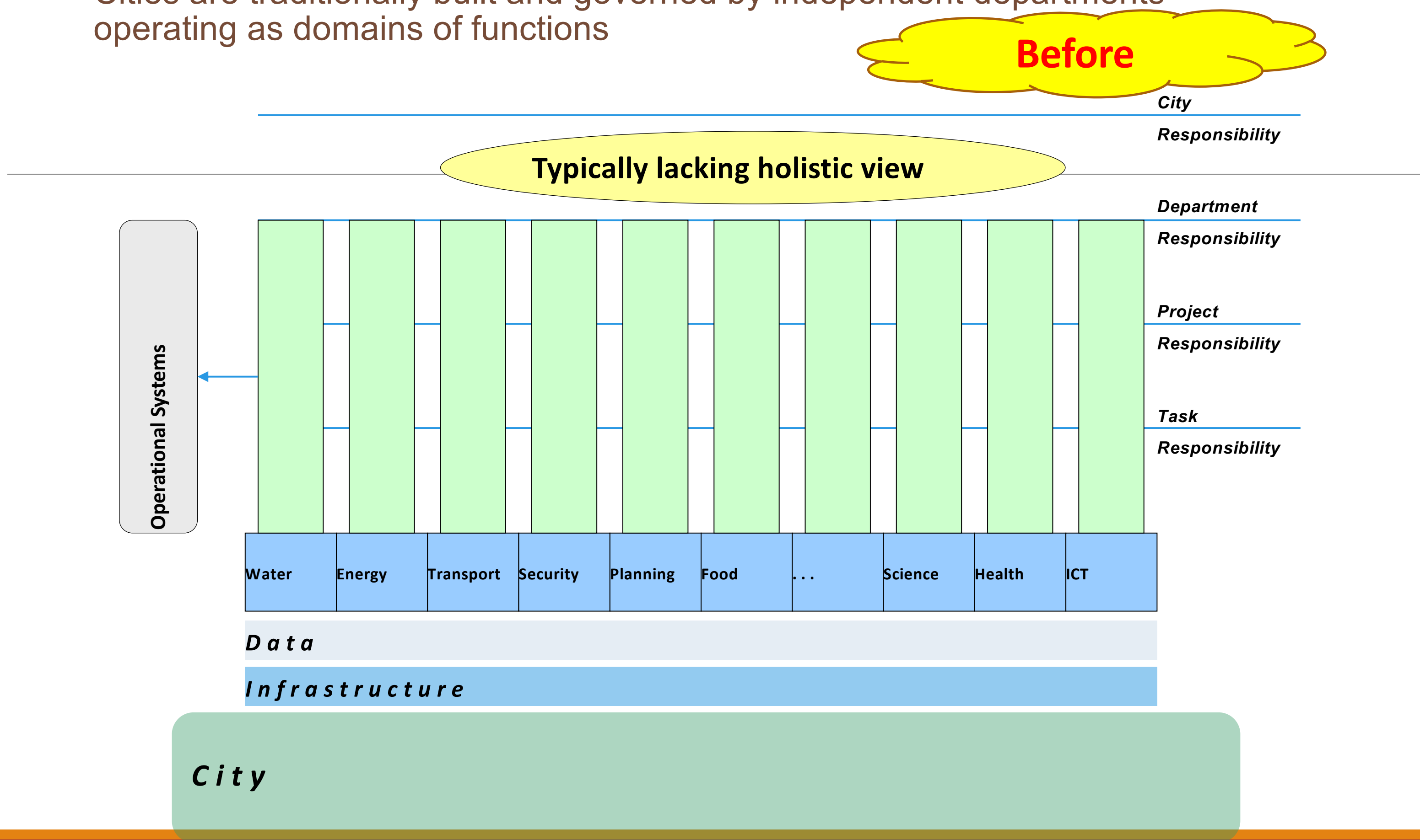
What is a Smart City?

- Smart city can mean one or more of the following:
  - As a **resource optimization objective**, it is to know and manage a city's resources using data.
  - As a **caring objective**, it is about improving standard of life of citizens with health, safety, etc indices and programs.
  - As a **vitality objective**, it is about generating employment and doing sustainable growth.

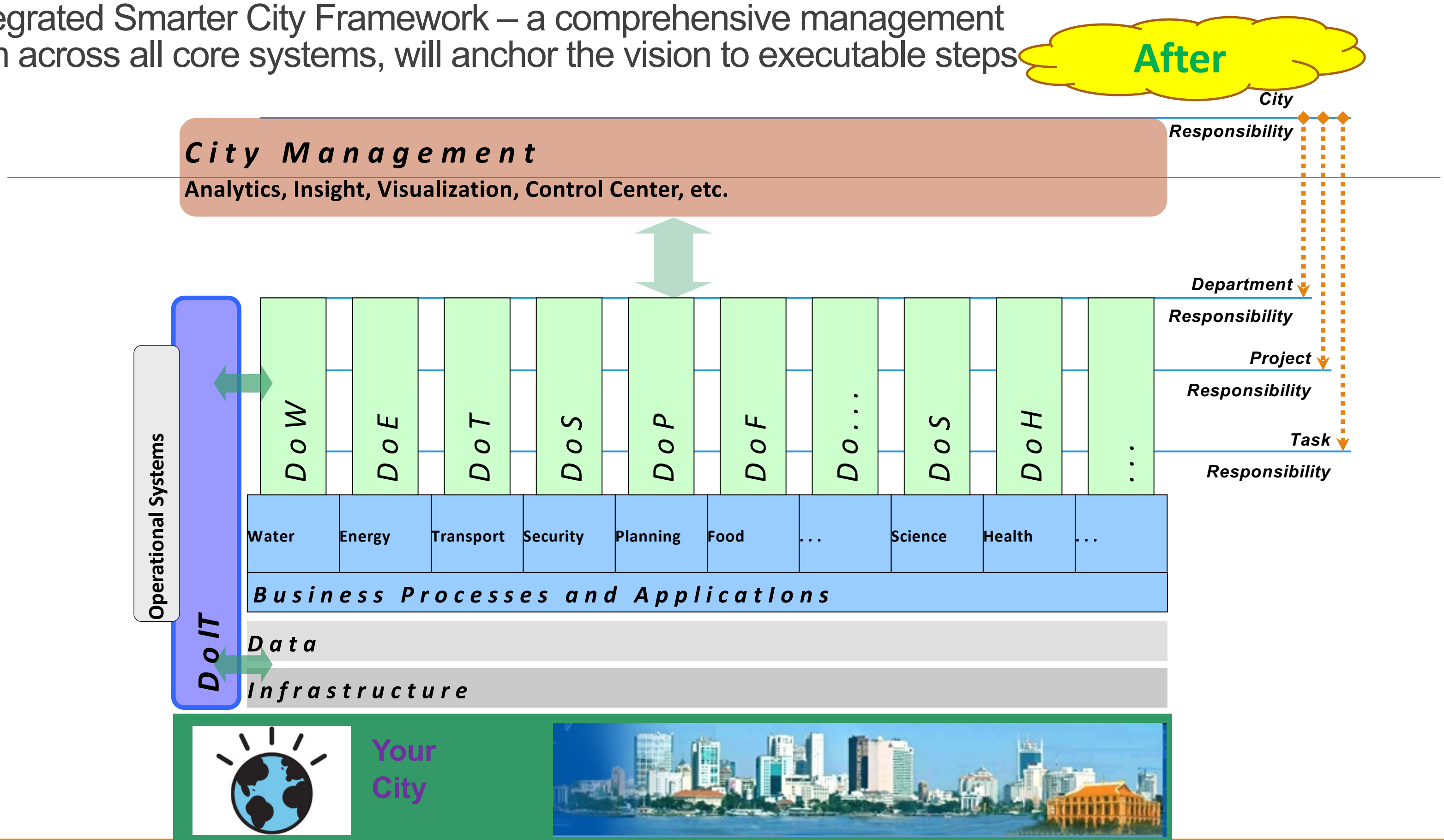
- A city leadership can choose among these or define their own objective(s) and manage with measurements to pro-actively achieve it

See other FAQs at: https://sites.google.com/site/biplavsrivastava/research-1/intelligent-systems/scfaqs

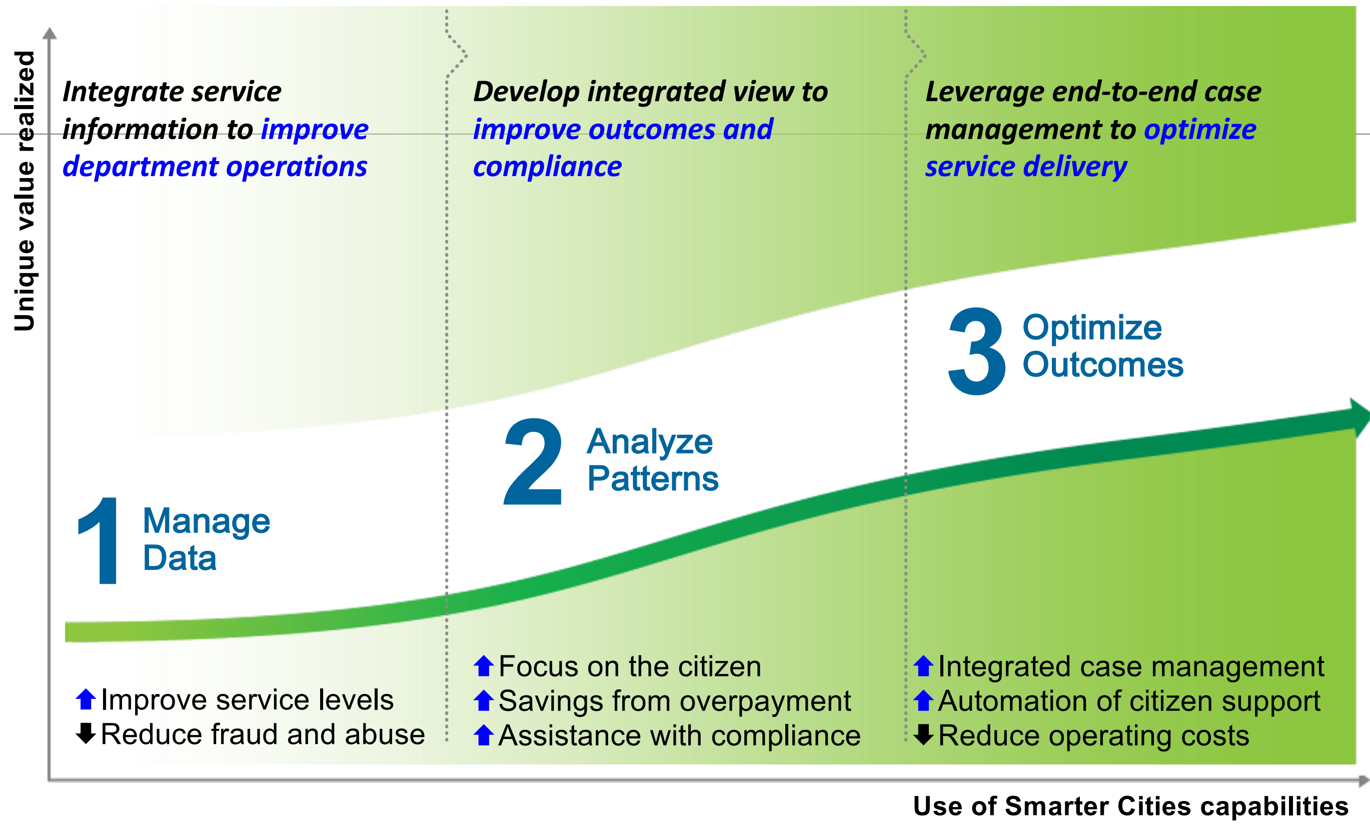
Cities are traditionally built and governed by independent departments operating as domains of functions



An integrated Smarter City Framework – a comprehensive management system across all core systems, will anchor the vision to executable steps



Smarter Cities solution paths leverage a similar approach



A Framework

A Systematic Approach

- **Identify:** Value of decision: before and after
- **Assess:** Data-needed
- **Explore**: Methods
- **Conduct**: Evaluation
- **Integrate**: solution with overall process

IJCAI 2015 Tutorial: https://sites.google.com/site/aismartcitytutorial/

Better Health

Two Tales from (Public) Health

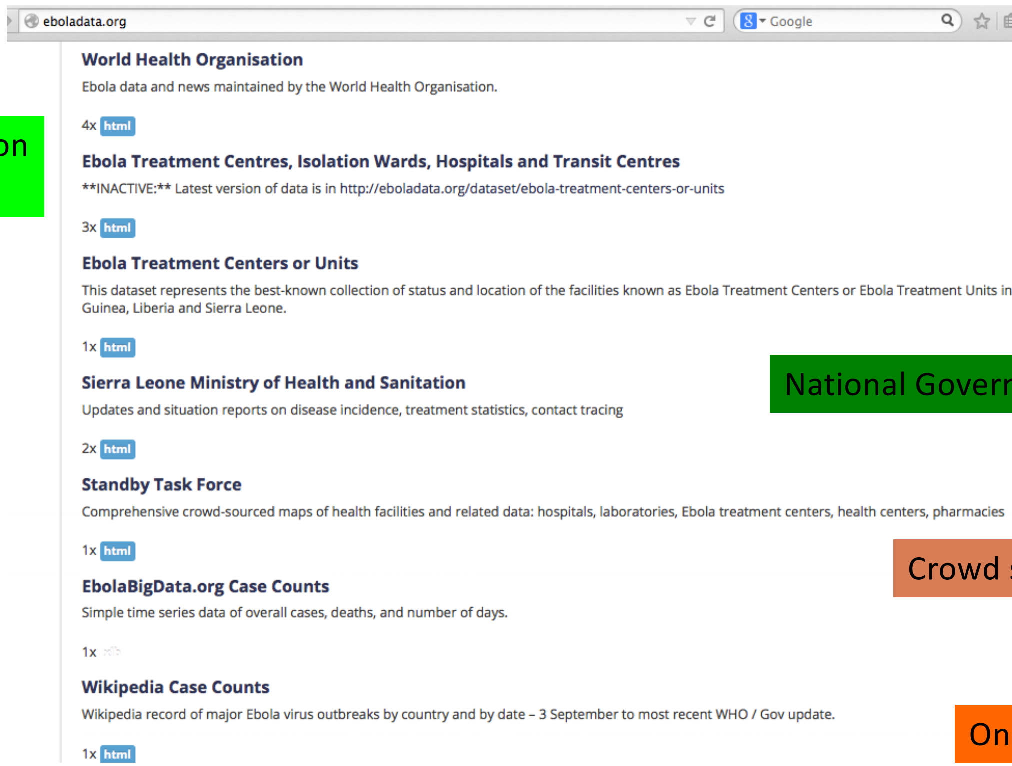
Cutting-edge Technical Progress

- Enormous improvement in our understanding of diseases. E.g., Computational epidemiology
- Enormous advances in treating diseases are being made
  - We are living longer - A baby girl born in 2012 can expect to live an average of 72.7 years, and a baby boy to 68.1 years. This is 6 years longer than the average global life expectancy for a child born in 1990. (**Source**: WHO 2014 Health Statistics)
- Data on disease outbreaks is more available than ever before thanks to open data movement (E.g., data.gov, data.gov.in)

Stone-age Ground Reality

- Half of the top 20 causes of deaths in the world are infectious diseases, and maternal, neonatal and nutritional causes, while the other half are due to noncommunicable diseases (NCDs) or injuries. (**Source**: WHO 2014 Health Statistics)
- Worse – Indifference, mismanagement in response to communicable diseases - late response to known diseases, in known period of the year
  - E.g.: Japanese Encephalitis (JE) has been prevalent for ~3 decades in some parts of India killing 600+ every year
  - District level health experience is not reused over time and in similar regions

Ebola Data

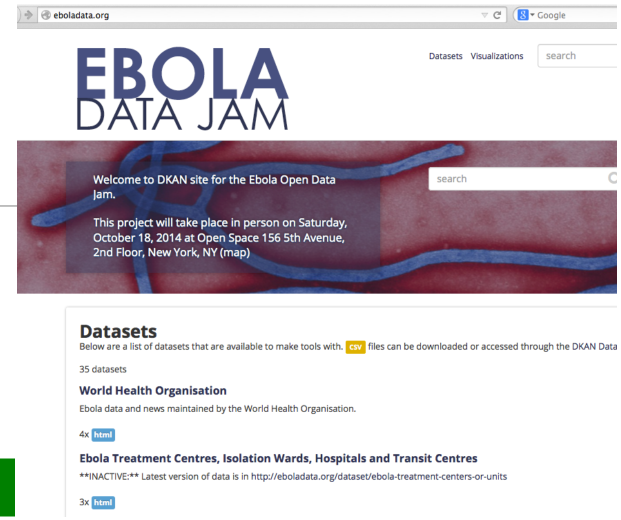


International Bodies

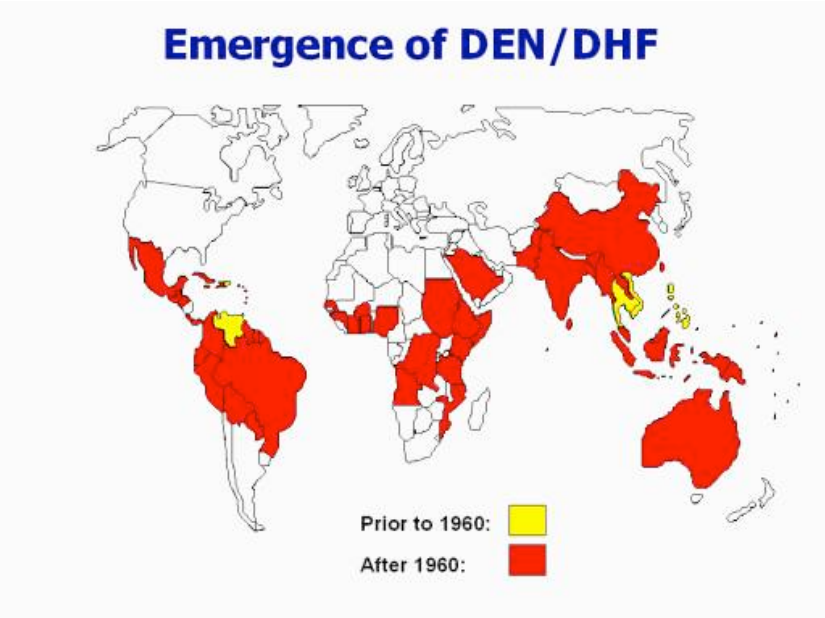
National Government

Crowd sourced

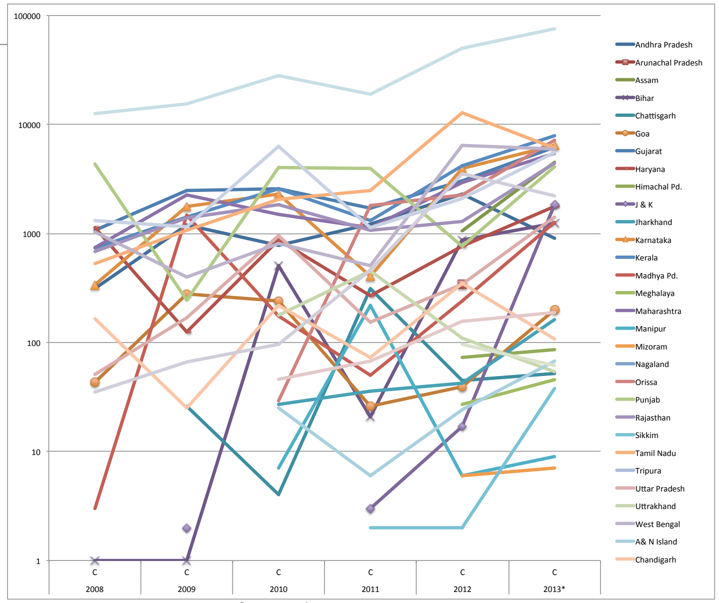
Online



Dengue



So, Do We Control Dengue Effectively? NO



Data for India

- Increasing number of states every year
- No consistent reduction of cases

Source: http://nvbdcp.gov.in/den-cd.html

Decisions in Public Health - Dengue

- **Identify:** Value of decision
  - What specific steps can we take to reduce Dengue and when?
  - How to choose among a set of possible options?
- **Assess:** Data-needed
  - Directly available
  - Available by proxy
- **Explore**: Methods
  - Decision theory (benefit v/s cost), minimize risks (deaths)
- **Conduct**: Evaluation
  - What metrics are relevant?
- **Integrate**: solution with overall process
  - How can the solution be integrated in overall process at a City?

Case Study: Dengue (Mosquito-borne)

Overall cost of a Dengue case is US$ 828 (Sabchareon et al 2012).
From 9 countries in 1960s, it has spread to more than 110 countries now

Prevention methods

COMMUNITY

1. **Mosquito Coils & Candles**: The use of mosquito coils, candles & vapor mats indoors and outdoors of homes to combat mosquitoes.
2. **Window screens & Bed Nets:** The use of window screens in homes and bed nets in bedrooms to keep mosquitos out.
3. **Insecticide Application:** Application of insecticide to kill mosquitos that invade homes and surrounding areas.
4. **Larviciding at Home**: Application of larvicide in homes to kill larvae that live in stagnant water breeding sites like small ponds, gutters, cisterns, barrels, jars, and urns.
5. **Household/Community Cleanup:** Organize cleanups within communities in the surrounding housing areas and individual homes to recycle potential breeding sites like discarded plastic bottles, cans, old tyres, and any trash that can hold water for mosquitoes to breed in.

GOVERNMENT

6. **Surveillance For Mosquitoes**: Conduct periodical surveillance in hotspot areas and other communities to look for signs of mosquitoes.
7. **Medical Reporting**: To collate and compile reports of dengue cases and statistics to prioritize and focus dengue and vector mosquito control efforts and actions for best results.
8. **Effective Publicity & Campaigns**: To foster and champion effective campaigns amongst communities and create adequate public awareness of combating dengue.
9. **Enforcement**: Support and enforce the public and communities to practice effective dengue vector elimination under existing laws and implement new laws as appropriate for public health.
10. **Insecticide Fogging**: Conduct fogging in areas that have mosquitoes and dengue outbreak hotspots to kill adult mosquitoes.
11. **Public Education**: Foster, promote, and participate in public education in schools and all possible public meeting places to inform communities how to eliminate dengue vector mosquitoes, recognize early symptoms of the disease, and proper medical care and reporting.

CORPORATE

12. **Education**: To undertake community service initiatives and campaigns through marketing expertise and the media of TV, radio, and newspapers.
13. **PR/CSR**: To use public relations and customer service relations to reach communities on the fight against dengue.
14. **Adult Mosquito Traps**: To provide adult mosquito traps and other measures within the work areas to protect employees and workers from mosquitoes bites that transmit dengue.
15. **Mosquito Repellants**: Provide mosquito repellants to employees and workers within the work areas for further protection.
16. **Mosquito Control Materials, Methods, and Agents**: To provide the tools to the public and government that are necessary for dengue mosquito vector control like pesticides, biocontrol agents, mosquito traps, repellants, and other means to prevent dengue by eliminating the mosquito vectors.

WHO, 2013, Dengue Control. At http://www.who.int/Denguecontrol/research/en/, Accessed 21 June 2013.

Entogenex, 2013, Integrated Mosquito Management. At http://www.entogenex.com/what-is-integrated-mosquito-management.html,

Accessed 21 June 2013.

(ROI) Metrics - Illustration

Expense for disease control

- $/person spent: How much money (in $) is spent for a given method divided by the population of the region. Lower is better.

Impact of a disease control method

- Reduction: What is the magnitude of reduction in disease cases due to a method, expressed as a percentage, in a time period (e.g., year, disease season)? Higher is better.
- Cases/ person: How many reported cases of a disease occurred in a time period divided by the population of the region when a method was adopted? Lower is better.

Cost-effectiveness:

- Cases / $: how many cases were reported for a disease per dollar spent on controlling it in a given time period? Lower is better.

Major Methods to Tackle Dengue

M1: Public awareness campaigns: to prevent conditions conducive to disease propagation, to improve reporting

M2: Chemical Control: Aerosol space spray

M3: Biological Control: Use of biocides

M4: Distributing equipments: bednets, insecticide- treated curtains

M5: Vaccination against the disease

Dengue Control Case Studies from Literature

Approach	Methods used (Mi)	Nature (Region, Population, area, year)	Expense per person	Reduction in number of cases
A1	M1, M2, M3	Sau Paulo, Brazil; 10,927,985; 2005	US $1.14	34%
A2	M3	Puerto Rico; -; 2003	< US$ 2.50	50% (in Dengue transmission)
A3	M2	Songkhla, Thailand; 162,645; 2009	US$ 1.24	
A4	M5	Bang Phae, Thailand; 207,000; - AND Thailand; 4002; 2009 - 2014		0-70%, 30.2%

- An approach may use 1 or more method(s)
- They incur different costs per person
- Their efficacy is subject to various factors

Still, can we reuse these results in new areas?

Details:

Vandana Srivastava and Biplav Srivastava, Towards Timely Public Health Decisions to Tackle Seasonal Diseases With Open Government Data , International Workshop on the World Wide Web and Public Health Intelligence (W3PHI-2014), AAAI 2014

Challenge: Prescribe Methods to Use for a Hypothetical, Illustrative Area - Sundarpur

- City is Sundarpur
  - Made up of 10 districts
  - 10,000 people in each district.
- Disease control
  - Each district allocates $10,000 per annum to prevent disease.
  - The city has a district-level health administrator per district and then an overall citywide public health administrator.
- What approach/ method should the district health officer use? What should the city health officer recommend?
  - a mix of control methods to produce the maximum reduction feasible.
  - Default option is to do nothing. This is unfortunately followed a lot!

Cost-benefits for Different Approaches

Appro ach Optio n	Population P	Amount available for expenditure (in USD) (a)	Expense per person for each method (in USD) (b)	Number of people exposed to the given method in the given amount c = (a)/(b)	Reduction in number of cases for each method (d)	Reduction in number of cases among exposed persons (e) = (c)*(d)	Effectiveness of the method E = (e) / P
O_def	10,000	10,000	0	0	0%	0	0%
O1_A 1	10,000	10,000	1.14	8772	34%	2982	30%
O2_A 2	10,000	10,000	2.5	4000	50%	2000	20%
O3_A 3	10,000	10,000	1.24	8065	10% *	806	8%
O4_A 4	10,000	10,000	8*	1250	70%	875	9%

* represents assumption made to compensate for missing data.

Prescription for Sundarpur

Best tactical option for administrators at Sundarpur (at district and the whole city level)

- is O1_A1 since it brings the maximum reduction.
- If the administrators are interested to cover the maximum number of people in the given budget, the best method is still O1_A1.
- If the administrators are interested to show maximum reduction in cases for a pocket of the city (sub-district level which may be more prone to the disease), they may choose O4_A4 but it costs maximum and thus can be perceived as taking resources away from the not- directed areas.

Strategic option

- Select top-2 (O1_A1 and O2_A2), and try them in 5 districts each in one year. It hedges risk of variability between Sundarpur and old location of previous studies.
- Based on efficacy, decide the single best option for Sundarpur in subsequent year.
- She may also use the vaccine option only when the disease outbreak is above certain threshold.

Details:

Vandana Srivastava and Biplav Srivastava, Towards Timely Public Health Decisions to Tackle Seasonal Diseases With Open Government Data , International Workshop on the World Wide Web and Public Health Intelligence (W3PHI-2014), AAAI 2014

New Data Practices

- Find correlation among methods (positive or negative)
  - We assumed independence
  - Needs: Historic Data, Experiment Design
- Learn rate of return for approaches and methods (new combinations not tried in health literature)
  - Need: Collect data on efficacy of method individually
- Find similarity among regions
  - Data Need: Spatio-temporal modeling/ STEM
- Multi-objective optimization
  - Examples: Effectiveness of approach, Reduction of case, people coverage
  - Needs: Data about approaches tried historically

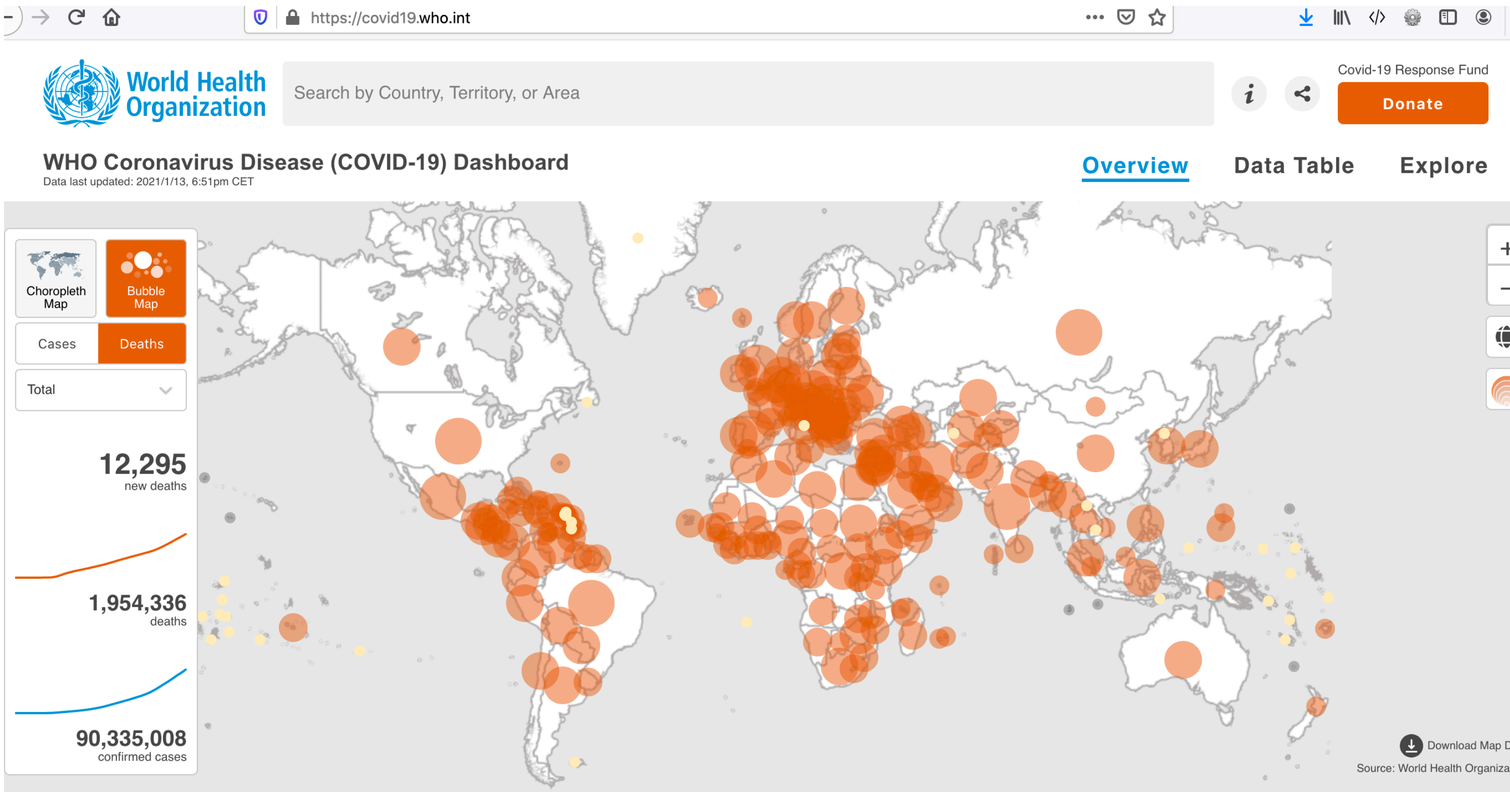
Request to Medical Community on Data

- Report both cost and effectiveness of approaches and methods
  - Overlooking one hampers reuse of results
- Interact with AI community to learn and try mixed approaches that reduce cost and improve overall effectiveness
  - All combinations cannot be tried on the ground due to practical constraints
  - Get more effective approaches rolled out faster targeted to new regions

Decision-Support During COVID-19

Resource: https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19

COVID-19 Pandemic



Globally, as of **6:51pm CET, 13 January 2021**, there have been **90,335,008 confirmed cases** of COVID-19, including **1,954,336 death** reported to WHO.

AI-Based Decision-Support for COVID-19

- Understanding the disease
  - Disease spread and simulation models
  - Insights by visualization
- Tackling the disease
  - Tracking people's movement
  - Fever detection via images
  - Understanding mental depression from social posts
  - Fighting fake news
- Understanding impact
  - Economic – job loss, industrial growth
  - Supply Chain
  - Risks

Resource: https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19

- Individual actions
  - *Screening/ triage tools*
- Group actions
  - *Models for when to open economy*
  - *Contact tracing*
  - *Matching producers and consumers: food, medical supplies*
- Policy actions
  - *Understanding impact of policy choices (e.g. lockdowns, travel restrictions)*
  - *Design of economic interventions*
- AI Community's Learning
  - *Data sources: Structured, Text – Research papers, Image / Video*
  - *Sharing and reuse of models and data is important*
  - *Lots of hackathons*

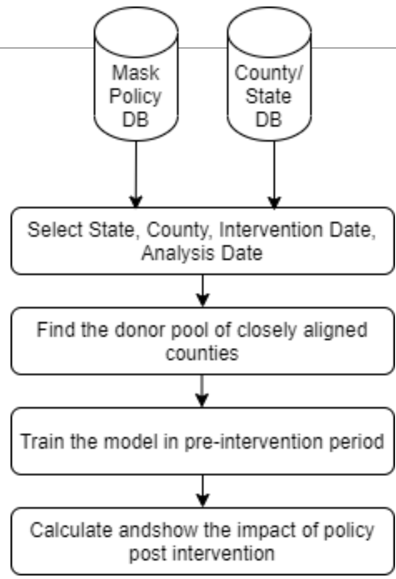
Does Wearing of Mask Help?

CMAP tool: https://sites.google.com/site/biplavsrivastava/research-1/ai-and-covid19

Covid Mask Assessment Program (CMAP) Tool

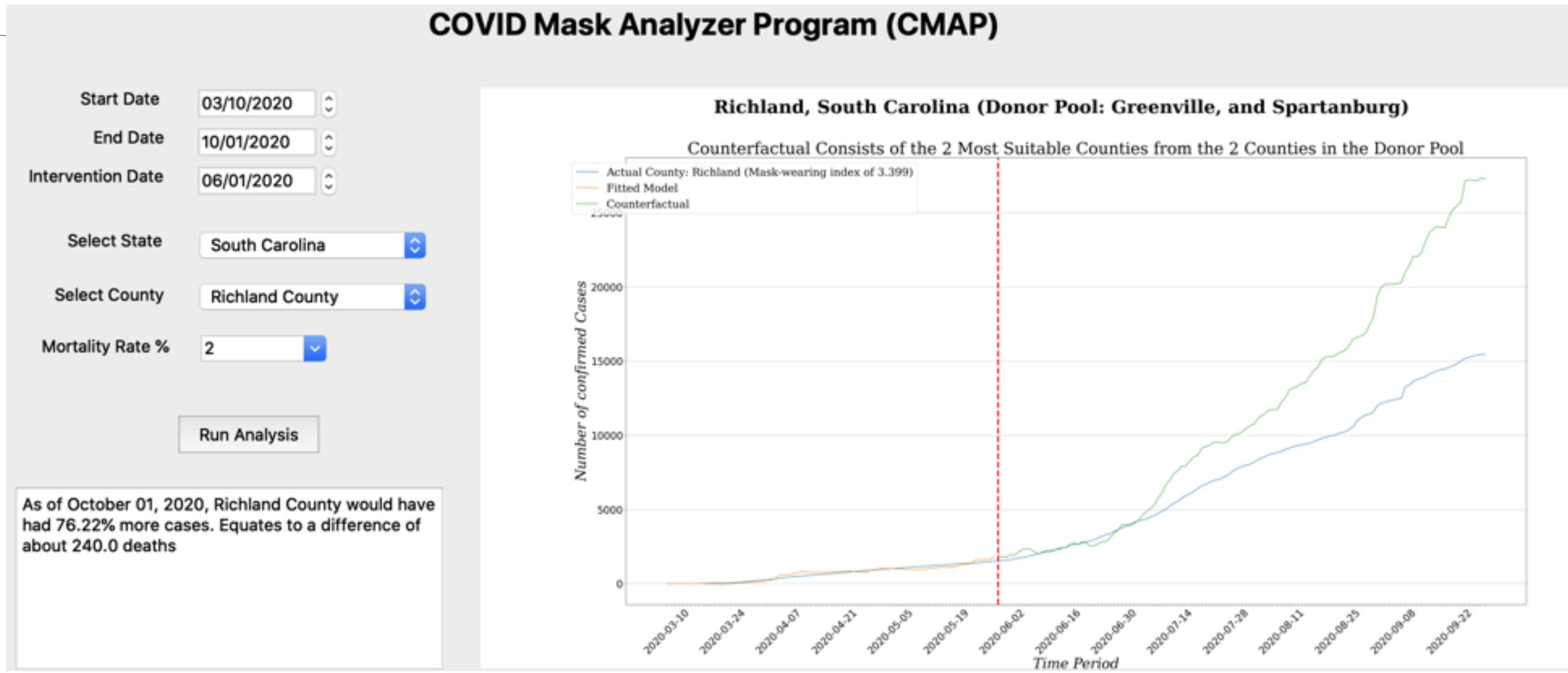
State	Pop. Rank	Av. Mask Score	Cases Rank
California	1	(3.49; 0.22)	1
Kansas	35	(2.66; 0.33)	34
New York	4	(3.61; 0.12)	4
Ohio	7	(2.81; 0.3)	14
Pennsylvania	6	(3.41; 0.23)	12
South Carolina	23	(2.99; 0.25)	16
Texas	2	(3.23; 0.3)	2

Table 1: US States and their statistics related to COVID19: Population ranking as per 2019 estimate[4], Average mask score and standard deviation, Case-based ranking with data as on Sep 17, 2020 from [5]. CMAP is most insightful in states with medium mask adherence and cases.



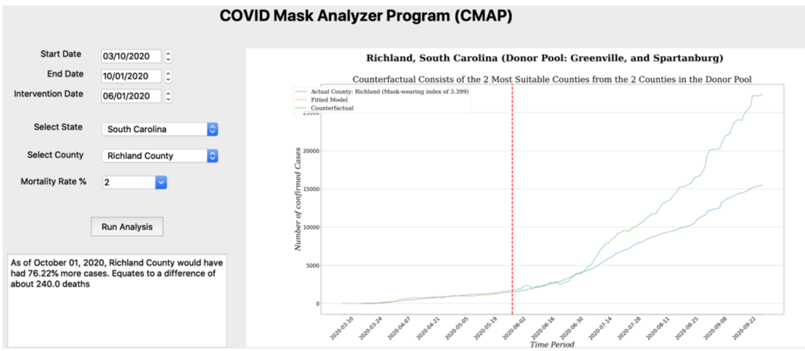
Video: https://drive.google.com/file/d/1Id9bm5400buCNqFg-8AojZGEddM8VMbm/view

The Case of Richland County, SC

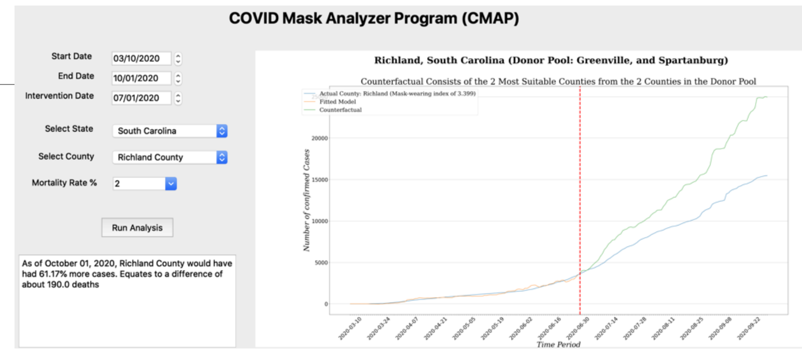


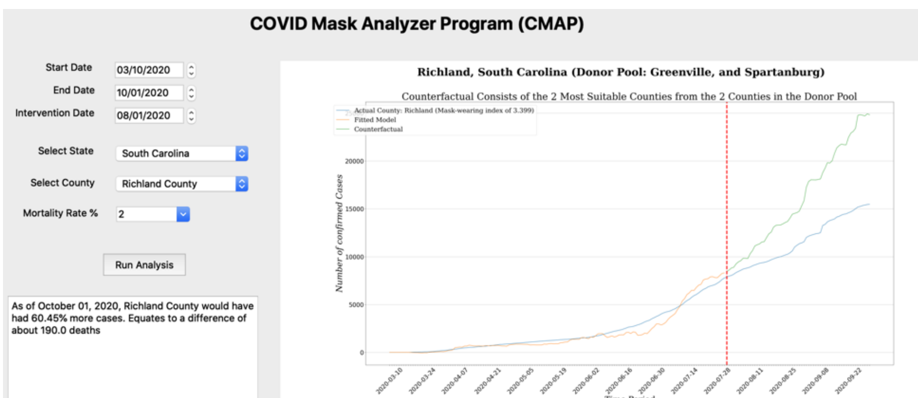
The Case of Richland County, SC

June 2020 Intervention



July 2020 Intervention





Aug 2020 Intervention

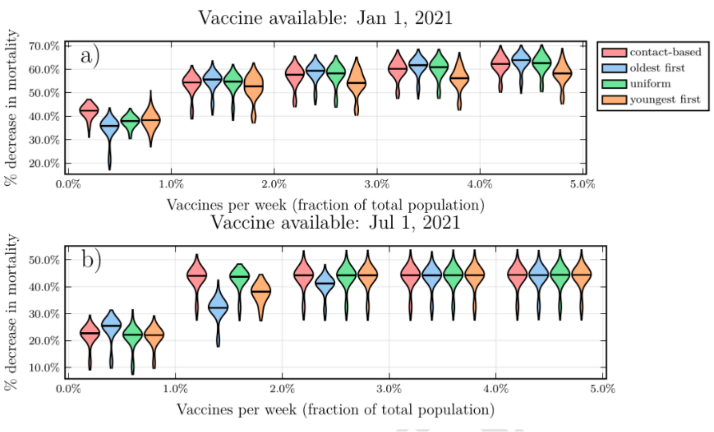
The earlier the intervention, the more lives were saved

Pressing Issue: Distribution of Vaccines

- Problem: Limited supply, larger demand; How do distribute equitably, fairly and efficiently
- Possible (automated) solutions
  - Random: pick receiver based on random choice
    - **Benefit**: Easy to implement
    - **Problems**: Equitable but not fair, receiver may not be at risk or not want it, others wanting it may not get it
    - **Question**: assumes we can give vaccine quickly to the selected person
  - Prioritized random: make a prioritized list of groups, assign randomly in each group
    - **Benefit**: identifies affected groups
    - **Problems**: receiver may not want the vaccine
    - **Question**: who comes up with groups?, is it rewarding groups who have not been taking precautions ? Assumes we can give vaccine quickly to the selected person
  - ...
  - Benefit-cost: based on contribution to economy
    - **Benefit**: efficient

Vaccine Distribution

- Article: [‘The Pandemic Is a Prisoner’s Dilemma Game’]
- [Prioritising COVID-19 vaccination in changing social and epidemiological landscapes], Sep 2020.
- Choices
  - impact of vaccinating 60+ year-olds first;
  - <20 year-olds first;
  - uniformly by age; and
  - a novel contact-based strategy
- Insights
  - Vaccination reduces median deaths by 32%-77% (22%-63%) for January (July) availability, depending on the scenario.
  - Vaccinating 60+ year-olds first prevents more deaths (up to 8% more) than transmission-interrupting strategies



Lecture 3: Concluding Comments

- We looked at problems of the real world
- Understood what is meant by "Smart City". Also referred to as "Cyber-Physical System".
- Introduced a systematic framework to introduce data-driven innovations
- Looked at examples in Health

Concluding Segment

About Next Lecture – Lecture 4

Lecture 4: Systematic Approach - Water

- Real-World Problems – Water
- A Systematic Approach on water problems
  - Value of decision: before and after
  - Data-needed
  - Method
  - Evaluation
  - Integrating with overall process