Leveraging Data and Health Artificial Intelligence Technologies for Chronic Disease

NETHIS
NÚCLEO DE ESTUDOS SOBRE BIOÉTICA E DIPLOMACIA EM SAÚDE

Ministério da Saúde
FIOCRUZ
Fundação Oswaldo Cruz
Brasília

bioeticaediplomacia.org

Opportunities, Challenges, and Implications with AI Technologies for Health Equity

26º Ciclo de Debates do Nethis: Vigilância em Saúde na Era Digital em Tempos de Pandemia
21 de Setembro 2023

Irene Dankwa-Mullan MD MPH
Adjunct Professor, The George Washington University Milken Institute School of Public Health
Chief Health Officer, Marti Health

Presentation Outline

1. Background
2. Artificial Intelligence and Chronic Disease Surveillance and Management
3. Data Science, Artificial Intelligence and Machine Learning Concepts
4. Harnessing data insights and analytics to address complex health system challenges – example of case studies
5. Challenges: Addressing AI Bias and Ethical Issues

Data is transforming every aspect of our world

Public Health Practice, Population Health and Healthcare Operations for Chronic Disease Management & Surveillance are being driven by Advanced Technologies

Operations now need to be shifted to meet the needs of individuals living with chronic disease conditions

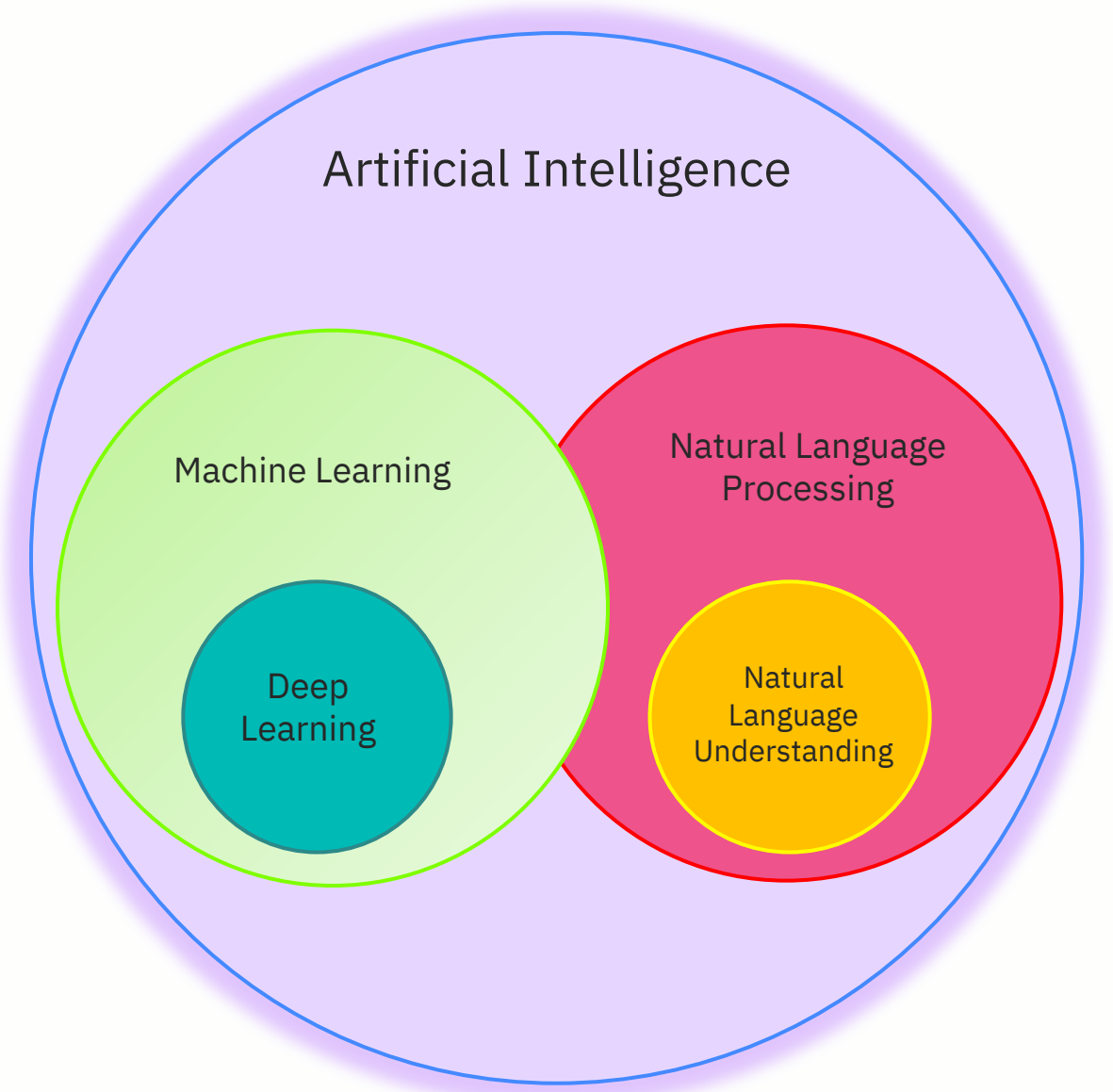
Few examples:

- Transition from clinician's full authority to multi-disciplinary teams managing the patient and emphasizing patient-centered care
- Transition from occasional patient monitoring during clinic visits to continuous monitoring of patient's condition and data outside of the healthcare setting
- Transition from prioritizing volume of care to providing value of care for patients to improve quality and promote health equity
- Transition from inadequate allocation of public health resources to improve and optimize efficiency in health resource allocation through geographic and asset area mapping with real-time population data and optimize care in underserved areas

ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, DEEP LEARNING, NATURAL LANGUAGE PROCESSING:

What are they, and how are they being used in health and healthcare?

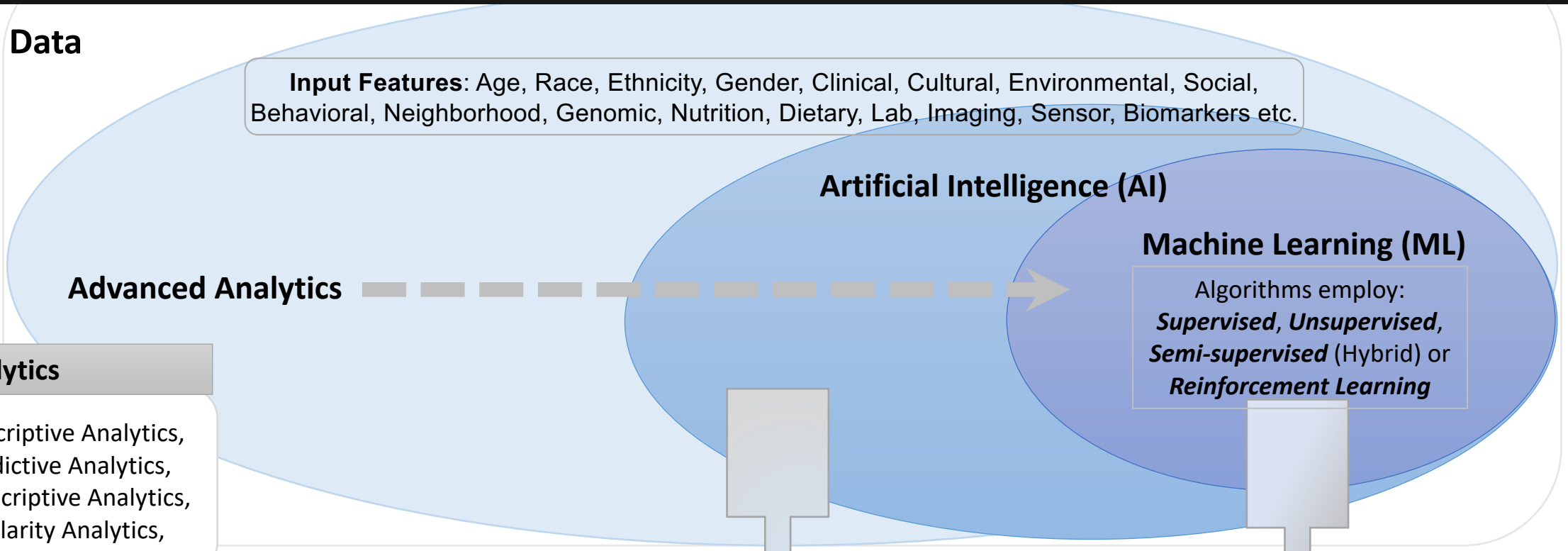
Artificial Intelligence (Machine Learning, Natural Language & Deep Learning)



- **Artificial Intelligence (AI)** – simulating human intelligence and its application for a specific practical purpose[1]
- **Natural Language Processing (NLP)** – a system of algorithms or programs that understands language, syntax, semantics[2]
- **Natural Language Understanding (NLU)** – is a subset of NLP that deals with machine comprehension that includes relationships, inferencing, sentiment[2]
- **Machine Learning (ML)** –a subdomain of AI pertaining to the family of algorithms that share a capacity to iteratively elucidate patterns (i.e., learn) to optimize tasks like prediction or classification[3]
- **Deep Learning (DL)**– a subset of ML that learns using very large data sets using artificial neural networks[3]

1. Benjamins JW et al. Neth Heart J. 2019;27(9):392-402
2. Jurafsky D, Martin JH. Speech and Language Processing. 2nd Ed. Prentice Hill. 2009.
3. Mitchell TM. Machine Learning. McGraw Hill. 1997.

Data Analytics: Spectrum and Complexity



Analytics

Descriptive Analytics,
Predictive Analytics,
Prescriptive Analytics,
Similarity Analytics,

Models

Predictive Models,
Causal Inference Models,
Hypothesis Generation, Testing and Scoring, Personalized Predictive Modeling
Disease Progression Modeling

Algorithm Types: commonly used (*supervised learning*)

Support Vector Machine (SVM)
Decision Trees or Classification
K-Nearest neighbors (KNN)
Naïve Bayes
Fuzzy Logic

ML: Conventional Learning

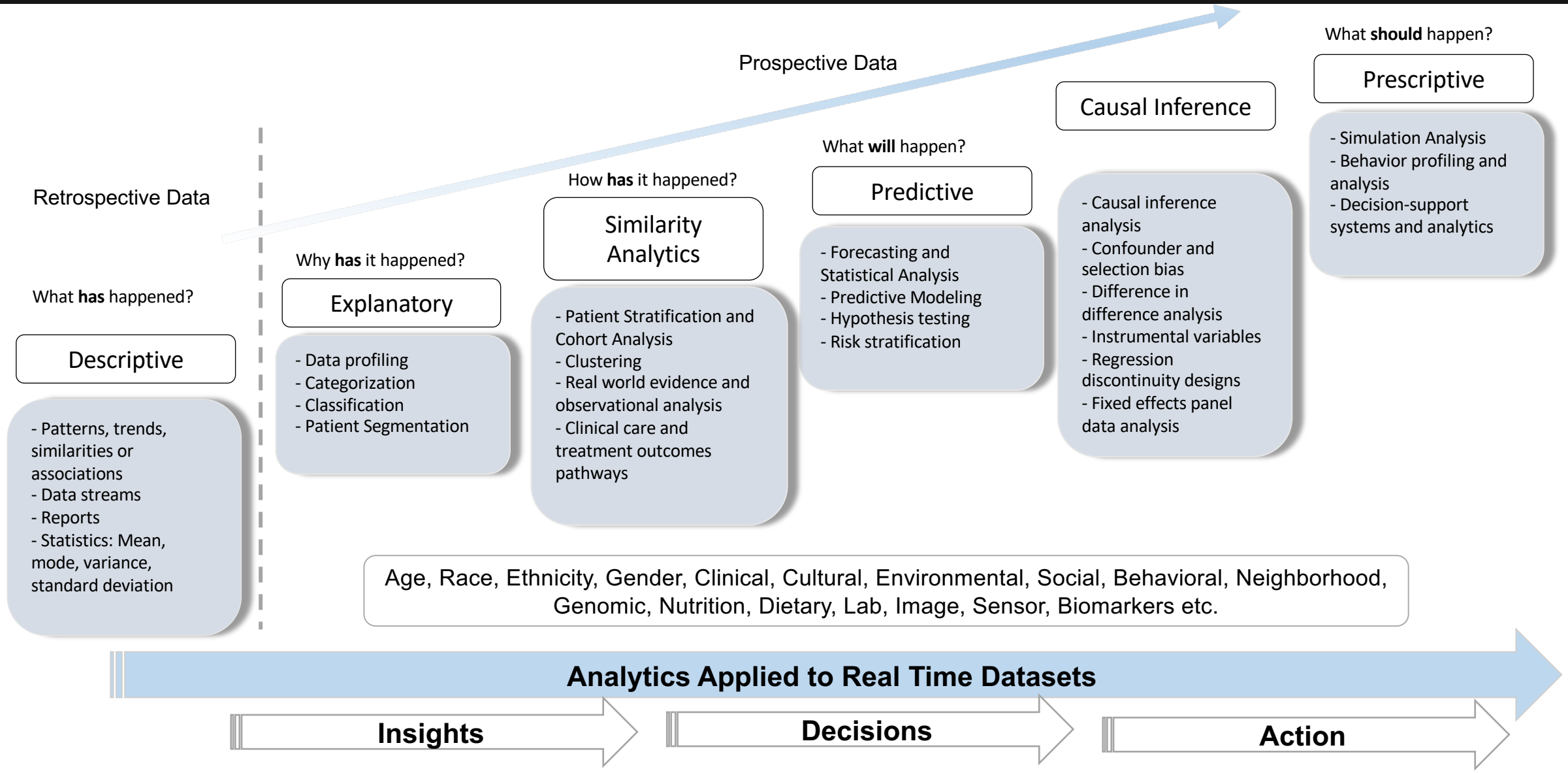
Neural networks with a single hidden layer or support vector machines

ML: Deep Learning

Neural networks with many hierarchical layers of nonlinear information processing

References: Dankwa-Mullan, I. Perez-Stable E et. al eds The Science of Health Disparities Research; 2021 Wiley

Spectrum of Advanced Analytics - Machine Learning and AI Techniques

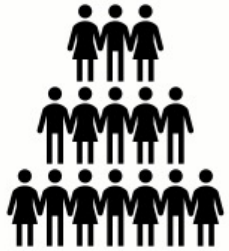


References: Dankwa-Mullan, I. Perez-Stable E et. al eds The Science of Health Disparities Research; 2021 Wiley
Dankwa-Mullan, I -

Some applications of Artificial intelligence and Machine Learning in medicine and health care

1 **Pre-clinical and translational research** e.g., drug discovery and genomic medicine, precision medicine

2 **Population-level applications** e.g., understanding non-communicable chronic diseases, identifying epidemics



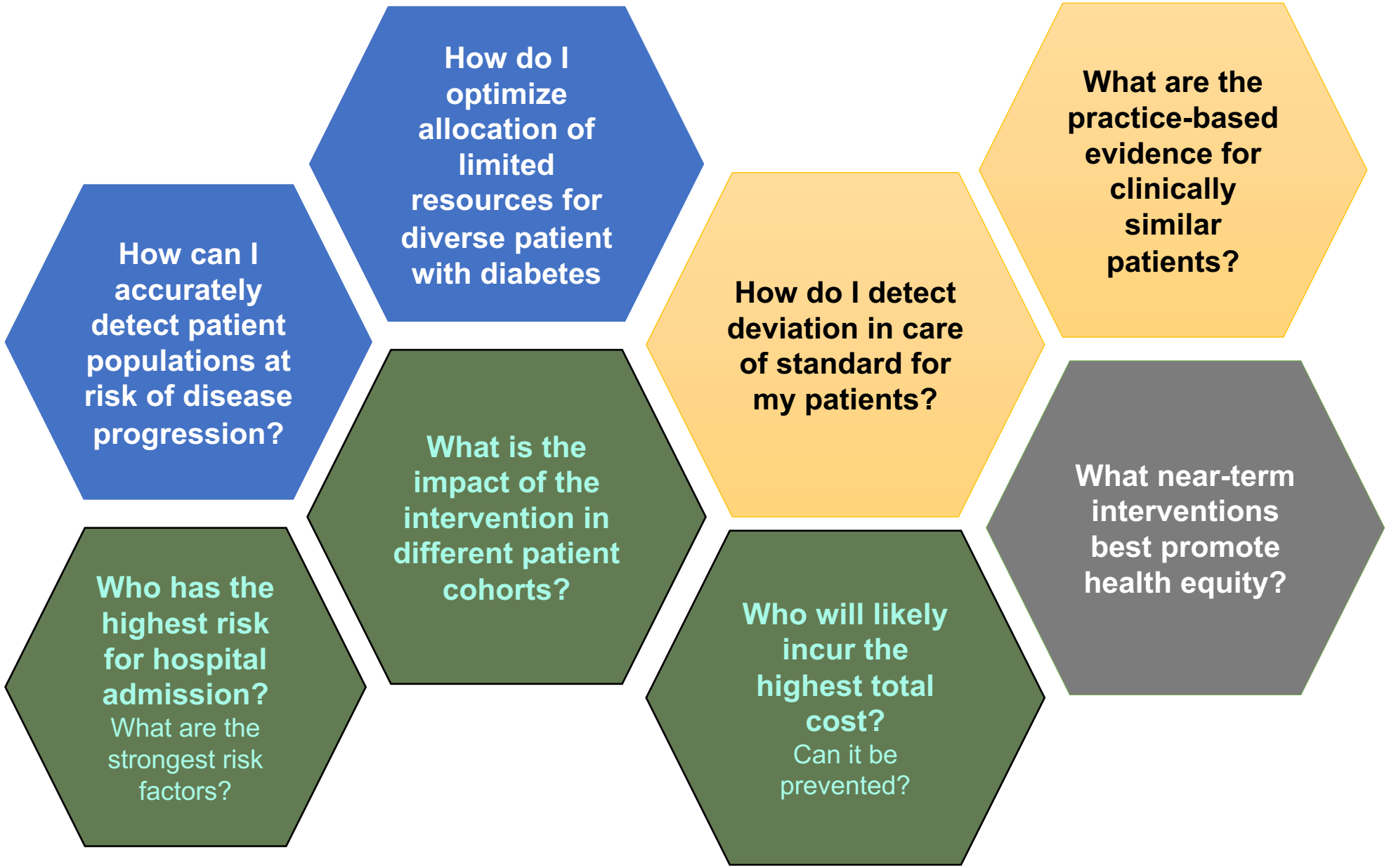
3 **Clinical pathways** e.g., diagnostics, making predictions and identifying risk or patient stratification

4 **Interpretation of Medical Images** Improved accuracy for radiology tasks such as screening, diagnosis, risk prediction and early intervention

5 **Patient-facing applications** e.g., delivery of therapies or the provision of information, clinical decision-making

6 **Optimizing healthcare delivery processes** Process optimization e.g., procurement, logistics, and staff scheduling

AI Insights for Chronic Disease Management & Surveillance



Public Health Surveillance and Management

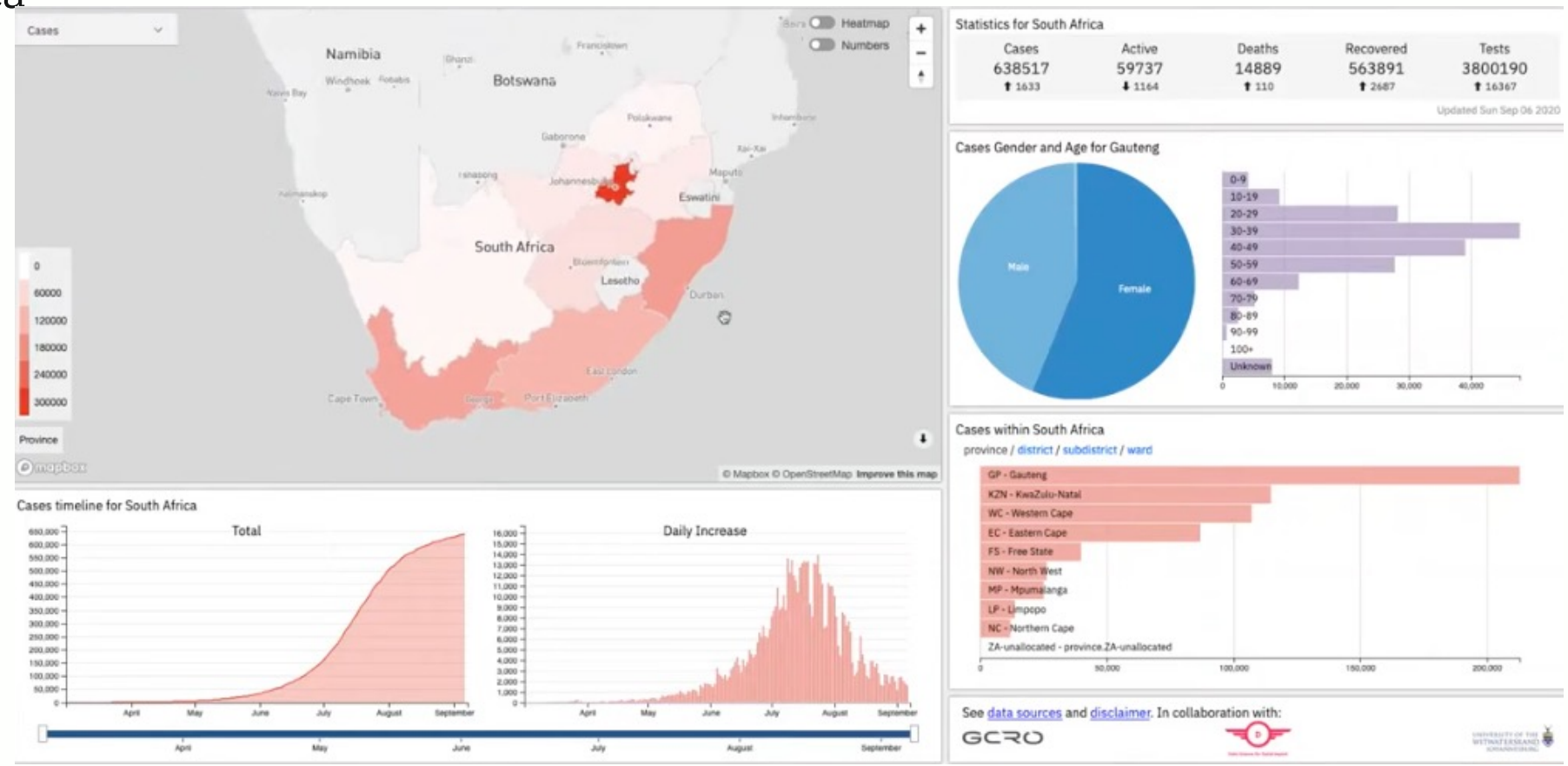
Population-level applications e.g., identifying epidemics and understanding non-communicable chronic disease prevalence, associated social determinants and risk factors

Using Data, AI and Technology to monitor and respond to COVID-19

Gauteng Province, South Africa

The dashboard was designed to address three key questions for policy makers:

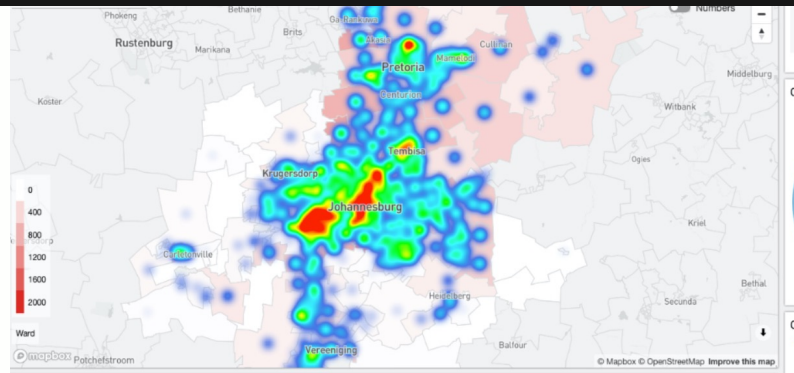
1. Where are the current hot spots of infection?
2. What are the predictions for the spread of the virus?
3. What are the risk factors that make certain communities more susceptible than others?



Using Data, AI and Technology to monitor and respond to COVID-19

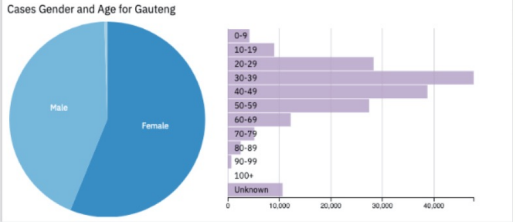
Gauteng Province, South Africa

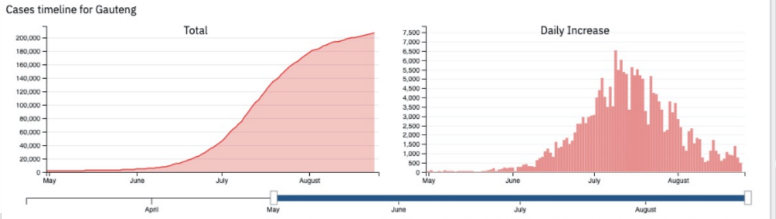
- How many active cases are there in my neighborhood?
- When is the predicted peak for my neighborhood?
- Are the number of cases rising or falling in the area of my local supermarket?

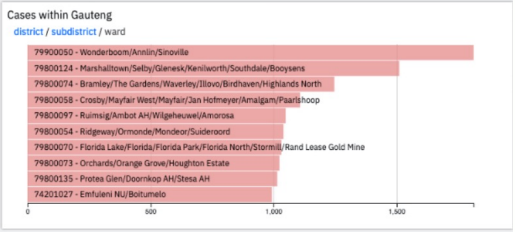


Cases	Active	Deaths	Recovered	Tests
206525	28997	3280	174248	215959
↑ 507	↓ 1141	↑ 12	↑ 1636	→ 0

Updated Mon Aug 24 2020

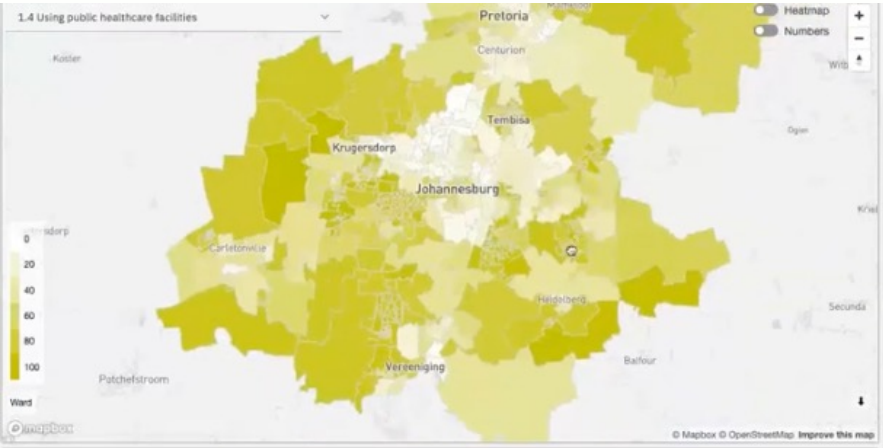






See data sources and disclaimer. In collaboration with:

GCRO



1.4 Using public healthcare facilities

Percentage of respondents per ward who normally use public health services.

Public health facilities are usually far more crowded than private facilities, often requiring long waits in queues in crowded waiting areas. Resource constraints and supply chain challenges mean that protective equipment is more likely to be in short supply than in the private sector, a challenge likely to be exacerbated by the fact that a far larger proportion of the population will attend public health facilities in the event that they do get sick. These factors may mean that the challenges of maintaining appropriate social distancing and preventative hygiene will be relatively greater in public health facilities than private ones.

Data source: GCRO QoL V (2017/18)

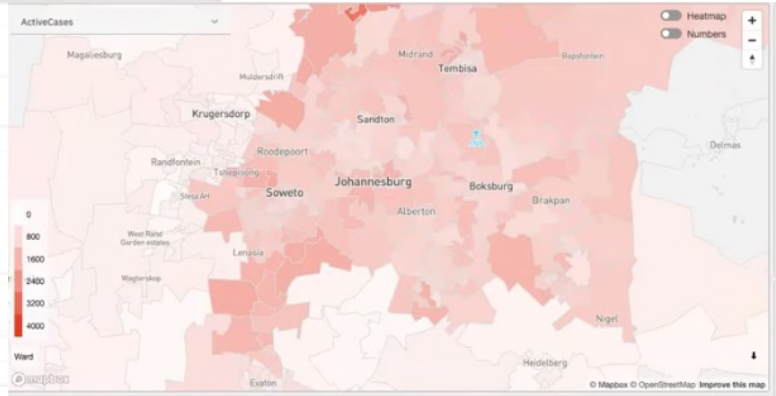
Ward
74205001 - Randfontein NU
74205031 - Mandela Section/Spoke Town
74205034 - Tambo Section
74801030 - Vlakplaas AH/Hillside AH/Bagale/Mogale City NU
79900019 - Slovo/Winterveld Ext 3/Lebanon/Mabopane SP
79800128 - Tshepisong SP
74201026 - Emfuleni NU/Evaton West
74205032 - Panvlak Gold Mine/Skierlik Section
74205029 - Venterspost Gold Mine/Holomisa Section
74801035 - Mayibuye
74804020 - Wedela SP/Elandsrand Mine
74205012 - Randfontein SP1/Pelzvale AH
79700047 - Vosloorus Ext 3/Vosloorus SP1/Nguni Section
74203006 - Ratanda Ext 7/Lesedi NU
79800035 - Jabavu
74203005 - Ratanda Ext 7
79700003 - Winnie Mandela Ext 7/Winnie Mandela Ext 5/Winnie Man

GCRO's Quality of Life V (2017/18) survey data was used to compile two based survey with randomly selected adults (18+ years old) as respondent respondents and is representative of the population of Gauteng.

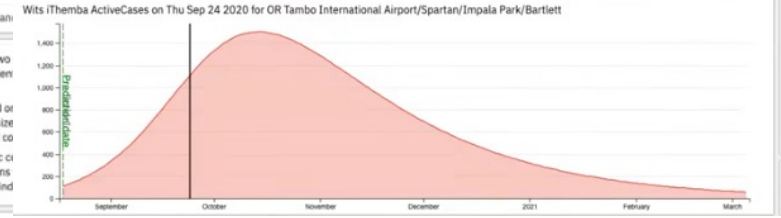
This analysis highlights the average level of risk at the ward level, based o each ward. Ward-level estimates are based on relatively small sample size Ward-level estimates should always be considered along with local and co

Furthermore, while ward-level analysis shades entire wards in a specific c vulnerability are likely. It is important that spatially targeted interventions consideration to ensure that they are directed towards the appropriate ind

See data sources and disclaimer.



	Scenario	Model	Prediction Date
☑	Gauteng Level 1	Wits iThemba	2020-08-18
☐	Gauteng Level 2	Wits iThemba	2020-08-18
☐	Gauteng Level 3	Wits iThemba	2020-08-18



Wards with highest predicted ActiveCases in Ekurhuleni Metropolitan Municipality on Thu Sep 24 2020

Ward	Municipality	ActiveCases
Winnie Mandela Ext 12/Winnie Mandela Ext 4/Winnie Mandela Ext 6/Duduza	Ekurhuleni Metropolitan Municipality	1536
Tokoza Ext 2/Palm Ridge/Kwenele	Ekurhuleni Metropolitan Municipality	1457
Windmill Park/Villa Liza/Rookraal/Mapleton AH	Ekurhuleni Metropolitan Municipality	1385
Rietspruit AH/Zonkizizwe/Palm Ridge	Ekurhuleni Metropolitan Municipality	1369
Hospital View/Ekurhuleni NU/Winnie Mandela/Winnie Mandela Ext 4/Sterkfontein Mines/Elandsfontein SH	Ekurhuleni Metropolitan	1341

See data sources and disclaimer.

Predictions supplied by the University of the Witwatersrand

Race, Social Determinants and COVID-19 Mortality Patterns in the United States

- The objective of the study was to introduce predictive / algorithmic modeling and machine learning approaches to the study of health determinants and COVID-19 mortality across the United States.
- We used a hierarchical machine-learning clustering algorithm approach to identify the population-level demographics, disease risk factors and SDoH characteristics that drove COVID-19-related risks for mortality in different county clusters.

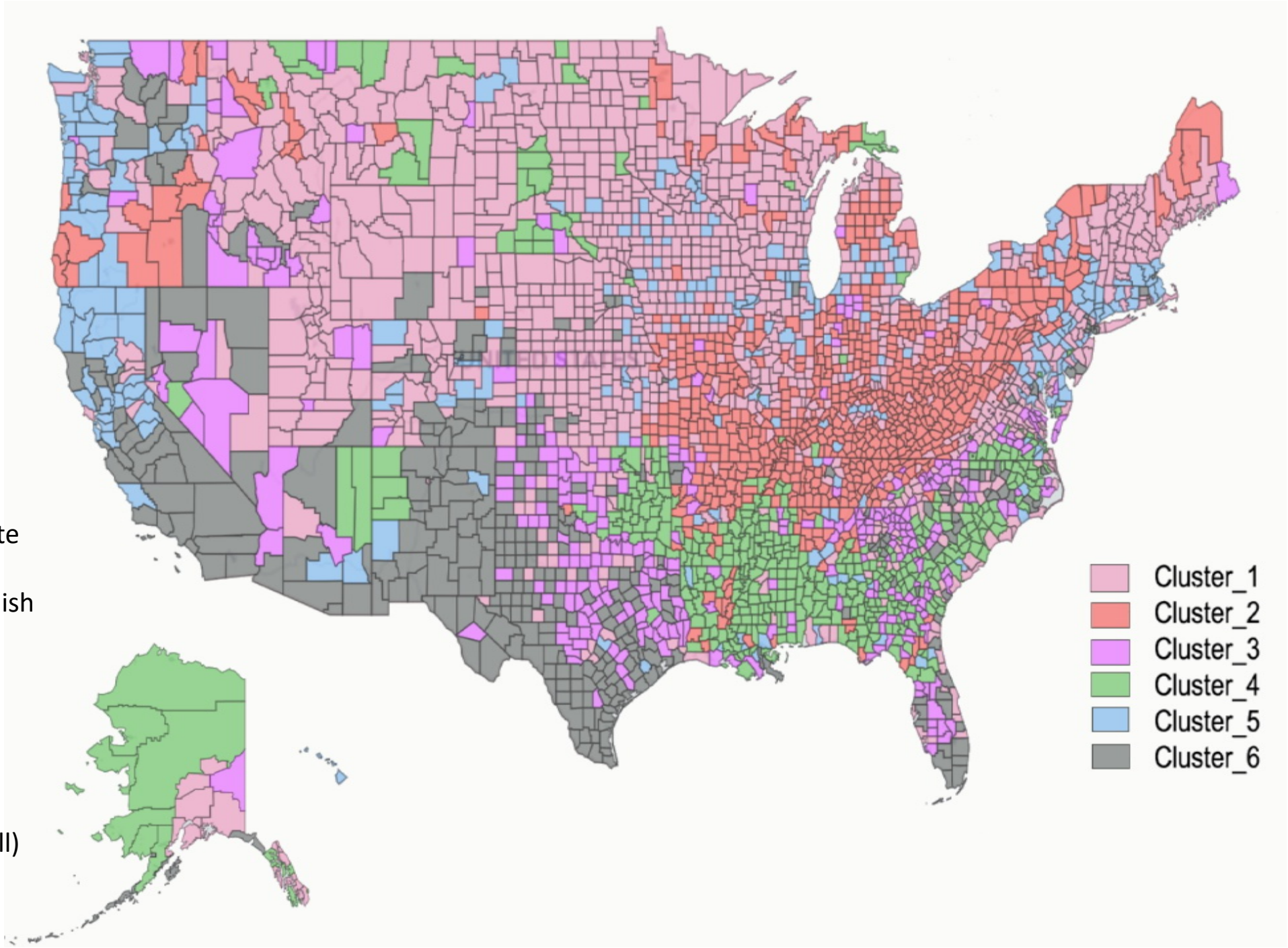
TEAM: Hu T. Huang PhD[1], Sarah Kefayati PhD[1], Cheryl R. Clark MD ScD[2], Anita M. Preininger PhD[1], Tiffani J. Bright PhD[1], Gretchen Jackson MD PhD[1,3], Irene Dankwa-Mullan MD MPH[1]

Results

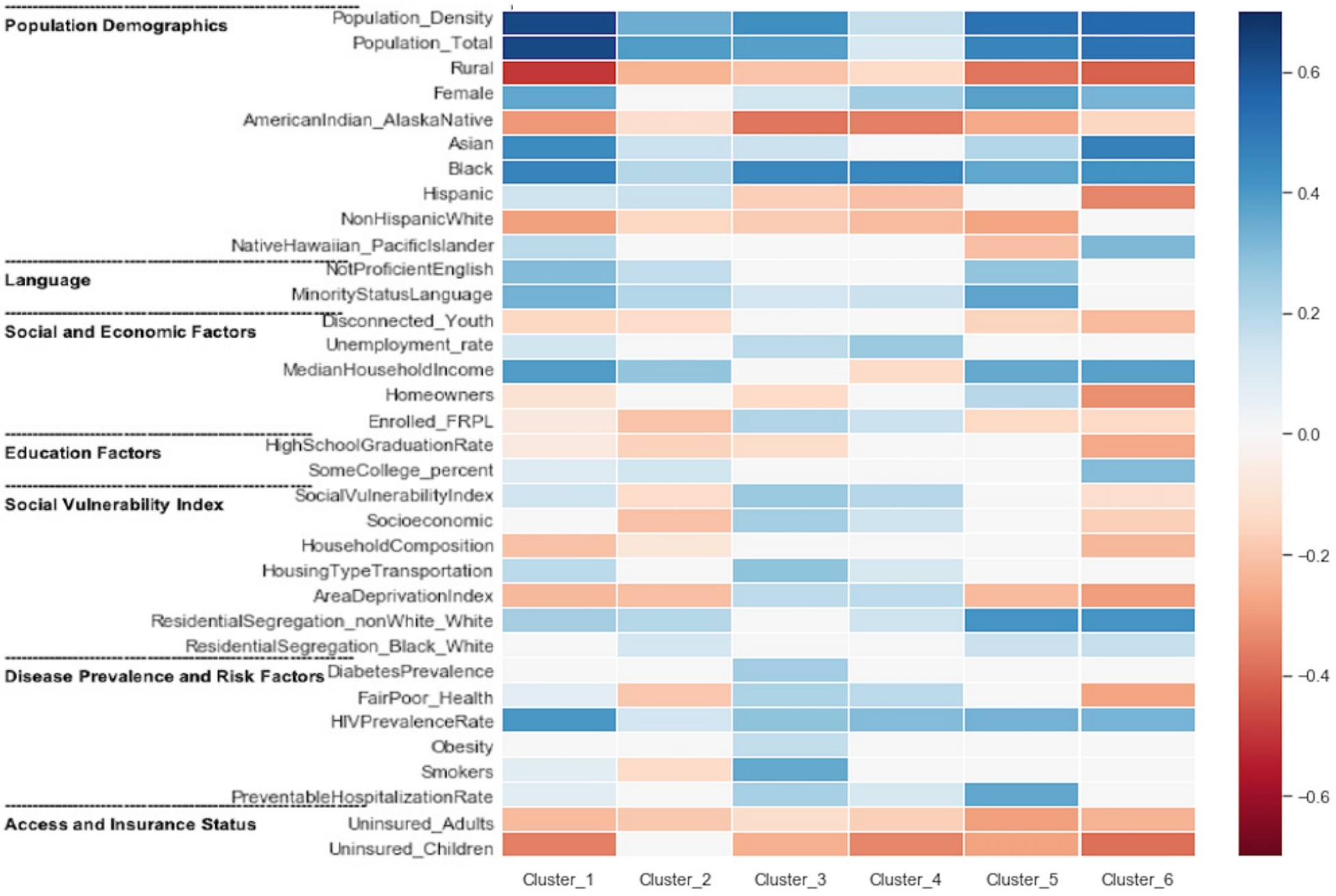
Clusters of U.S. counties based on disease risk, health prevalence and sociodemographic similarities

<u>Top Ten distinctive features were</u>:

- Proportion of the Hispanic population
- Proportion of non-Hispanic White population
- Population not proficient in English
- Population with some college degree
- Uninsured adults
- % Smokers
- % in Fair-poor health,
- Area Deprivation Index
- Social Vulnerability Index (overall)
- Minority Status & Language



Correlation heatmap of the county clusters and COVID-19 mortality



LEVERAGING AI INSIGHTS FROM DATA:

Examples of collaborative opportunities to address chronic disease in community and public health?

Collaboration with Health Systems and Stakeholders

Identifying Collaborative Strategies to Address Social Needs and Health Equity: Developing a Hospital Social Needs Index

Mahil Senathirajah, Irene Dankwa-Mullan, Gary Pickens, Richele Benevent, Bruce Spurlock

CalHospitalCompare: Multistakeholder collaborative of CA Hospitals, Health Plans, Employers, CA Dept. of Healthcare Access and Information (HCAI), Patients, Quality Innovation Networks (QIN/QIO), Covered California, Hospital Quality Institute etc.

Acknowledgements: Public Health Alliance of Southern California

Addressing Social Needs of Patients and communities

California Healthy Places Index

- Developed by Public Health Alliance of Southern California
- 25 component measures, 8 domains, multiple data sources
- Domain weighting based on prediction of Life Expectancy at Birth

Economic	Education	Healthcare	Housing	Neighborhood	Environment	Social	Transportation
Poverty Employment Income	Pre-school enrollment High school enrollment Bachelor's degree attainment	Insured adults	Severe cost burden low income; renters, owners Homeownership Kitchen and plumbing Crowding	Retail jobs Supermarket access Parks Tree canopy Alcohol establishments	Diesel PM Ozone PM 2.5 Drinking Water	Two parent household Voting	Healthy commuting Automobile access
0.32	0.19	0.05	0.05	0.08	0.05	0.10	0.16

Healthy Places Index Policy Action Areas (Domains), Weights and Individual Indicators

Martin Luther King, Jr. Community Hospital

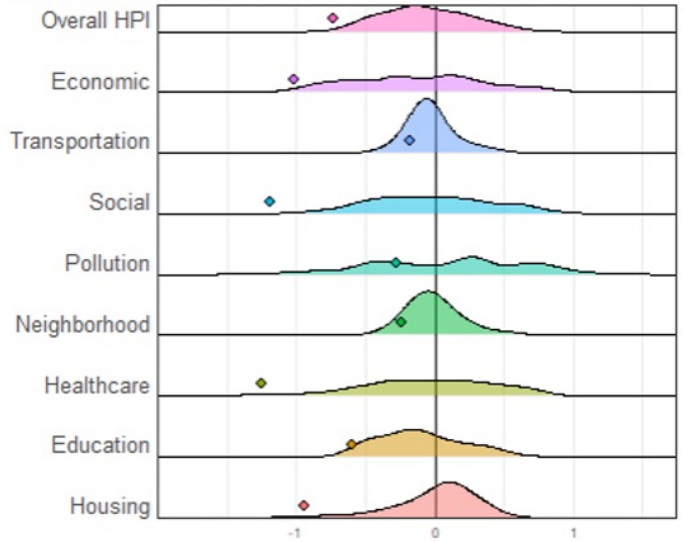
Hospital HPI Score: **-0.74**

Total Admissions: **8,221**

Metric Shown

Proportion of Cases from ZIP

Hospital Healthy Places Index Scores

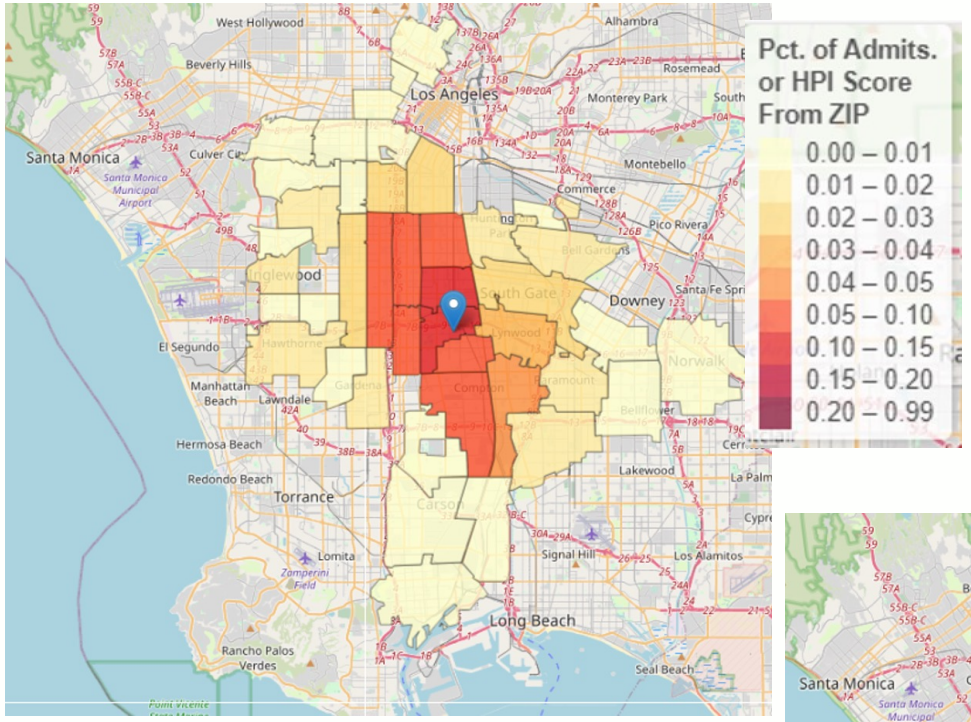


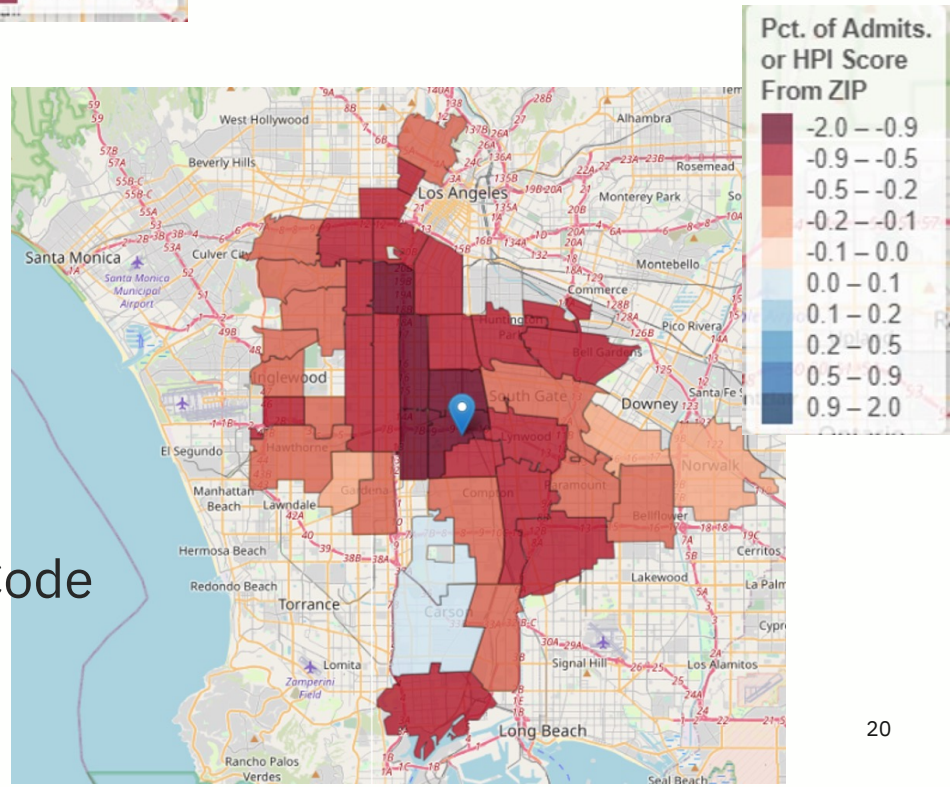
PO Name	ZIP	Prop. Total Admits	Admits	HPI Score
Los Angeles	90059	13%	1057	-0.96
Los Angeles	90002	11%	883	-0.91
Los Angeles	90003	9%	718	-0.98
Compton	90222	8%	689	-0.76
Compton	90220	7%	575	-0.49

1–5 of 44 rows Previous 1 2 3 4 5 ... 9 Next

Level of penetration : Proportion of patients being seen at a hospital by zip code

Urban Hospital- High Social Needs





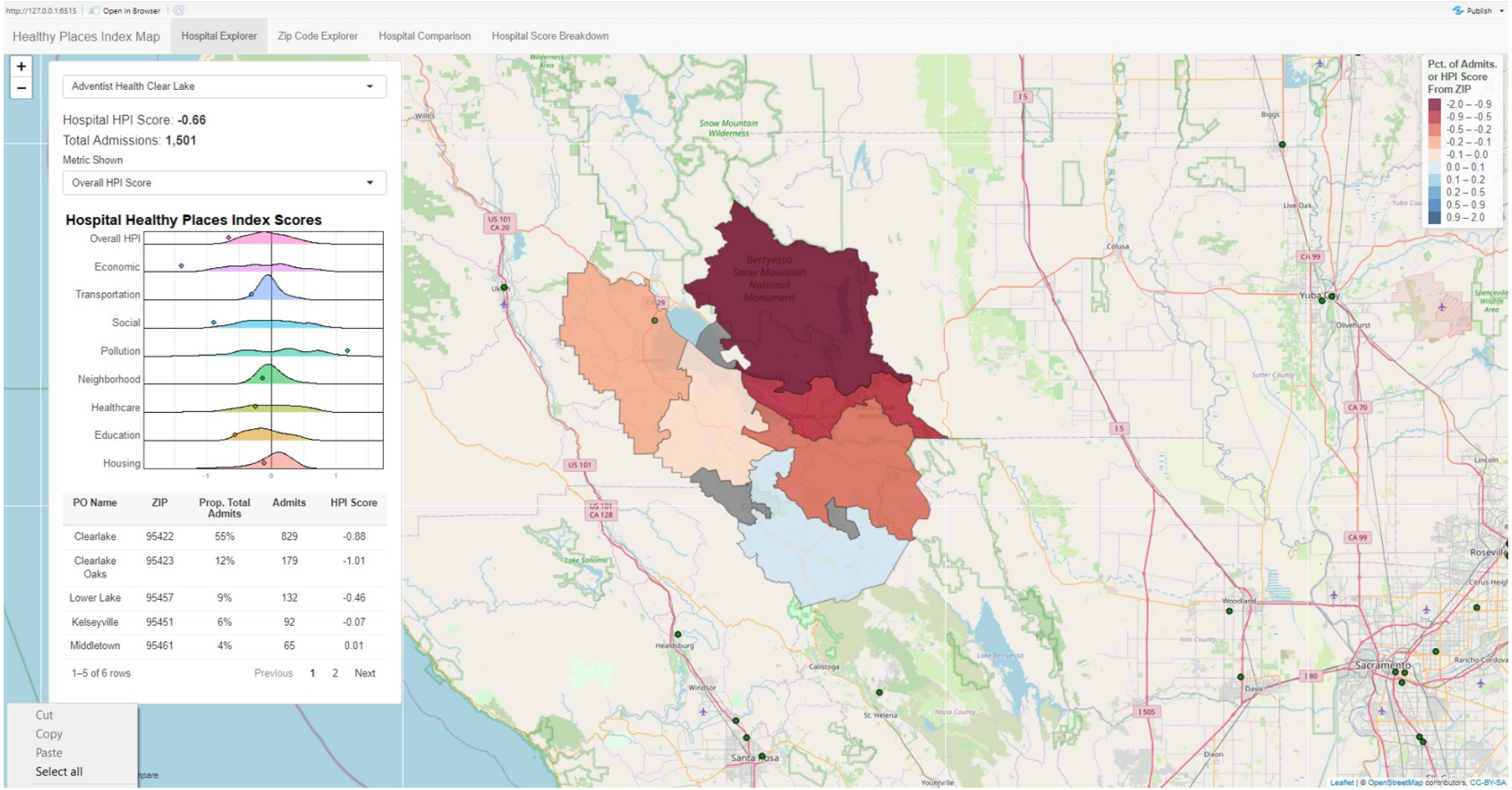
Level of penetration : Zip Code Healthy Places Index by hospital service area

Addressing Social Needs of Patients and communities

Figure: Geographic mapping of zip code-level social needs index (SNI) based on Hospital-specific Patient Origin

Rural Hospital

The map illustrates the social needs, based on the Healthy Places Index score of zip codes served by the hospital. Darker color indicates greater social need



Diabetes Equity

Edit

California

Alameda Pop. 1,682,115

View county details

Colusa
Chronic Condition Diabetes 14.70

Alameda
Chronic Condition Diabetes 6.70

Colusa Pop. 21,843

View county details

1668627177
Chronic Condition Diabetes 170

Chronic Condition Diabetes
50 170

Chronic Condition Diabetes
0 195

Chronic Condition Diabetes
3.50 15.80

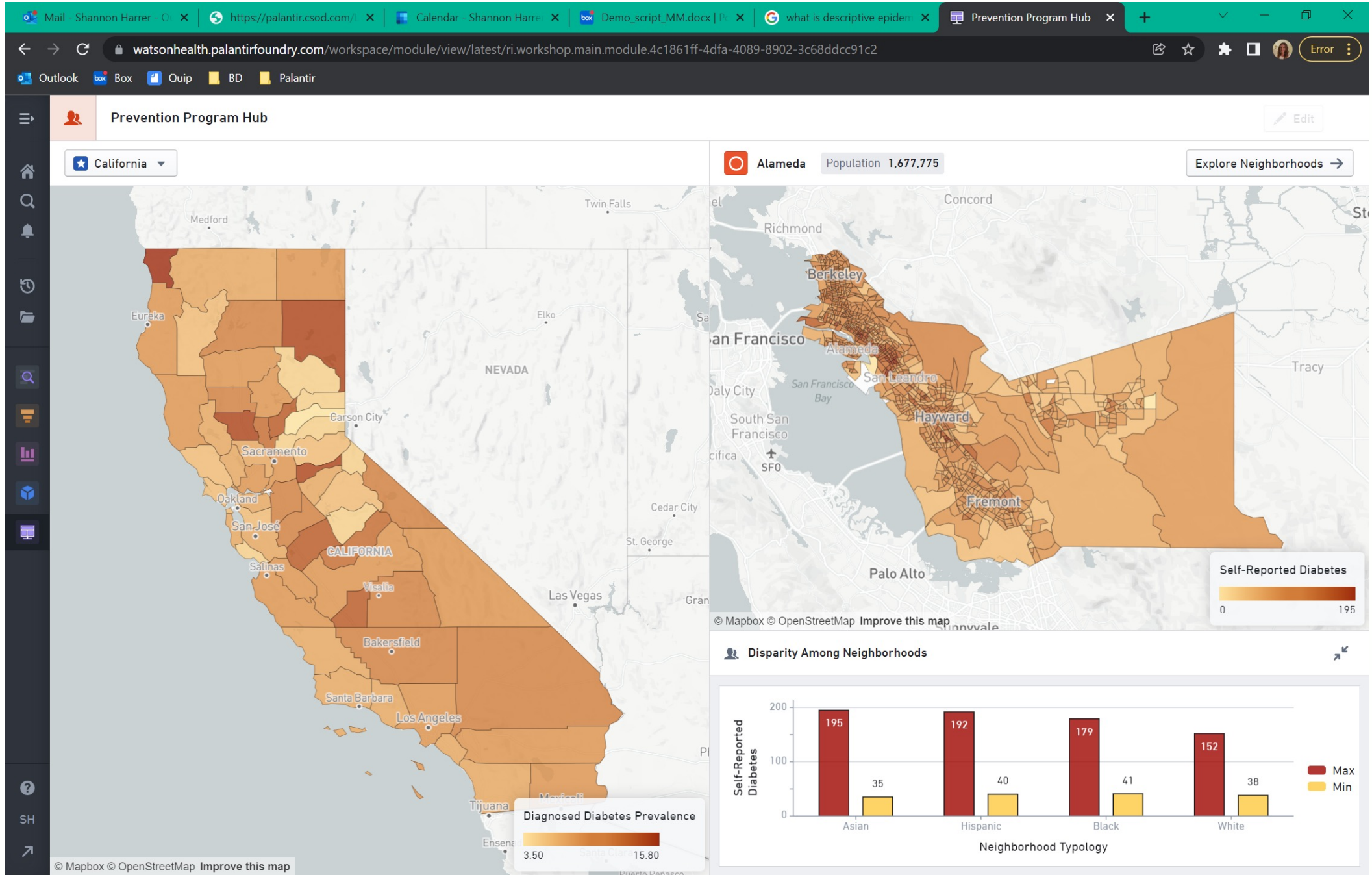
© Mapbox © OpenStreetMap Improve this map

% County SVI Percentiles

Below Poverty	36%	Income	29%
Limited English	76%	Minority	78%
Unemployed	43%		

Other Mapping – Chronic Disease - E.g. Diabetes

Collaboration with Technology partners



Mail - Shannon Harrer - O | Demo_script_MM.docx | P | Calendar - Shannon Harre | Demo_script_MM.docx | P | what is descriptive epidem | Prevention Program Hub

watsonhealth.palantirfoundry.com/workspace/module/view/latest/ri.workshop.main.module.4c1861ff-4dfa-4089-8902-3c68ddcc91c2

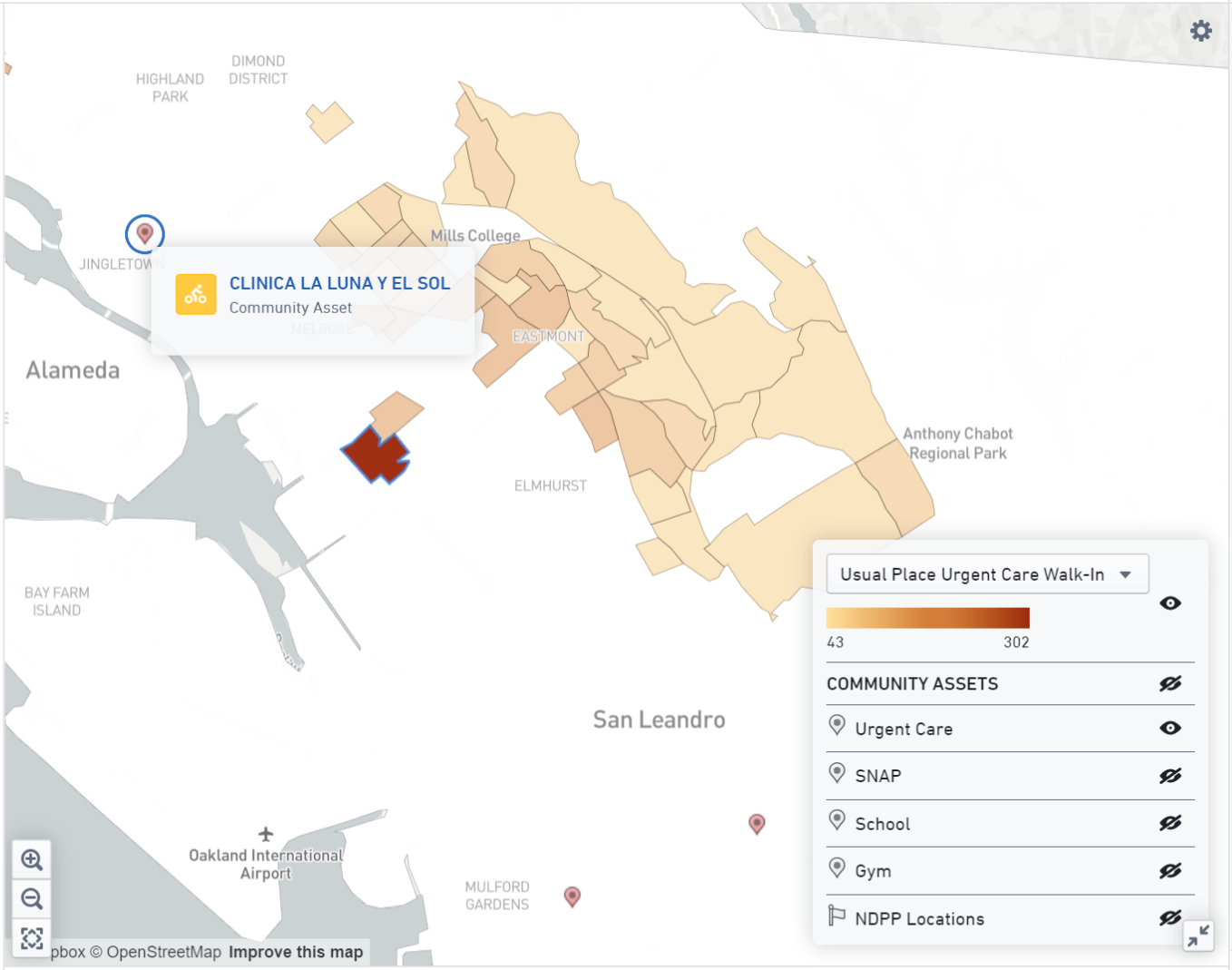
Outlook Box Quip BD Palantir

Prevention Program Hub

Edit

Return to State View | Alameda

Filters



Usual Place Urgent Care Walk-In

43 302

COMMUNITY ASSETS

- Urgent Care
- SNAP
- School
- Gym
- NDPP Locations

Data Drill Down

Actions

Block Group Details 060014088002

Internet Use Regularly	103
My Job Take Care My Health	102
Attended Wellness Program	168
Worried Food Will Run Out	142
Hard to Pay For Basic Needs	313

Comorbid Conditions, Behaviors & Environment

Depressed All/Most of Time	252
Poor Emotional Well Being	147
Personally Afflicted With Anxie...	111
Extremely Stressed	143
Tobacco: Uses Cigarettes	156
Alcohol Use Past 30 Days	145
Availability Of Fresh Food	95
Exercise Is Top Priority	116
Healthy Eating Habits	84
Lack Of Transportation Affecte...	208

Community Assessments and Interventions

Barriers to Health Screening

View

Create Community Assessment

Population Health Management – Decision-support tools, Health Equity Dashboards

Poverty Dashboard Series, San Diego County

Areas of Persistent Poverty (APP) and Historically Disadvantaged Communities (HDC)

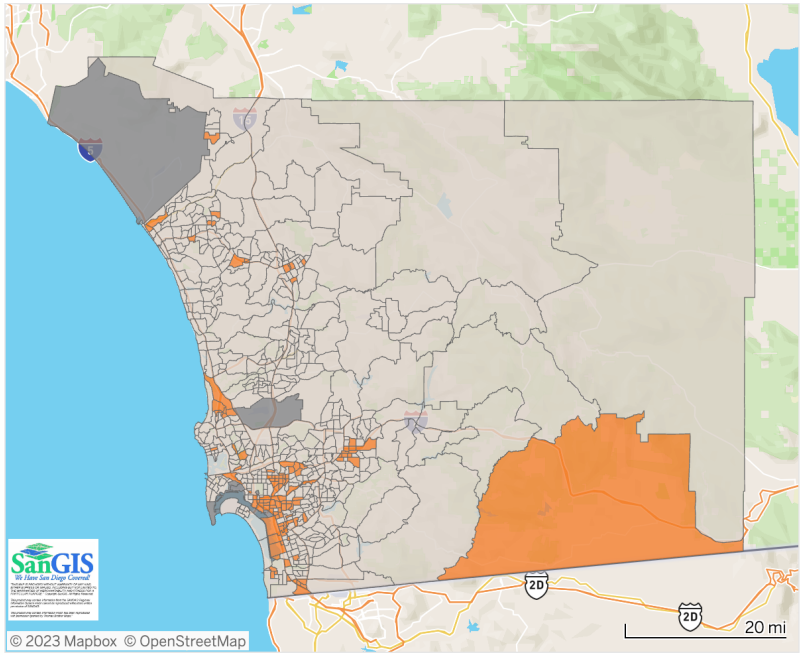
Hover over a census tract to view census tract name, Health and Human Services Agency (HHSA) Region, Subregional Area (SRA), Supervisorial District to which it belongs, Gini Index, proportion of population below 100% Federal Poverty Level (FPL), proportion of population below 200% FPL, and proportion of households receiving food stamps. Alternatively, use the search bar below the map to search for a census tract.

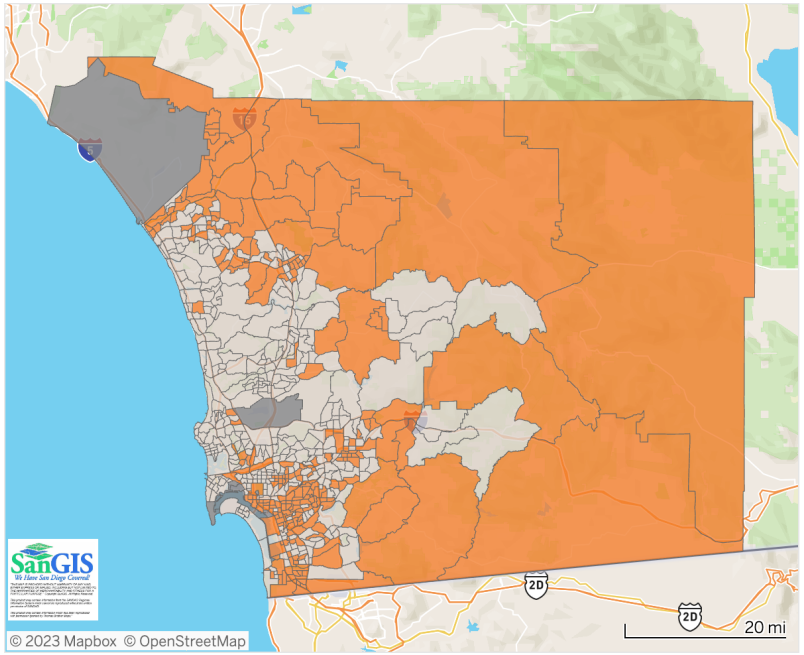
Areas of Persistent Poverty

An "**Area of Persistent Poverty**" is defined for the RAISE grant program by the 2021 Consolidated Appropriations Act. A project is located in an Area of Persistent Poverty for the RAISE grant program if the census tract in which the project is located has a poverty rate of at least 20 percent as measured by the 2014-2018 5-year data series available from the American Community Survey of the Bureau of the Census.^

Historically Disadvantaged Communities

Historically Disadvantaged Community: For the purpose of the 2022 Notices of Funding Opportunity, and consistent with OMB's Interim Guidance for the Justice40 Initiative, Historically Disadvantaged Communities include (a) certain qualifying census tracts, (b) any Tribal land, or (c) any territory or possession of the United States.^





Area of Persistent Poverty
- Yes
- No
- Military Census Tract*

Select census tract of interest to highlight on map

Highlight Census Tract

Historically Disadvantaged Community
- Yes
- No
- Military Census Tract*

Gini Index of Income Inequality: A higher Gini index indicates greater inequality, with high-income individuals receiving much larger percentages of the total income of the population.

****San Diego County Health Equity Zip Codes:** Based on how much area in each zip code is also a Healthy Places Index Health Equity Quartile (HEQ) census tract, 34 zip codes were identified to have at least 25% of the area in a HEQ census tract. Additionally, 5 zip codes were identified to have high burden of COVID-19 (defined as a cumulative case rate of at least 10,000 COVID-19 cases per 100,000 population).

^An "Area of Persistent Poverty" is defined by the RAISE (Rebuillding American Infrastructure with Sustainability and Equity) grant program by the 2021 Consolidated Appropriations Act. A "Historically Disadvantaged Community" is defined for the RAISE program in the 2022 NOFO, consistent with OMB (Office of Management and Budget)'s Interim Guidance for the Justice40 Initiative.

*Data not shown for census tracts with military barracks or majority population living in military group quarters.

Sources: U.S. Department of Transportation, Areas of Persistent Poverty Project (APP) and Historically Disadvantaged Community (HDC) Status Tool, https://datahub.transportation.gov/stories/s/tsyd-k6ij Accessed 5/16/2022. San Diego County Health Equity Zip Codes, https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/Epidemiology/COVID-19%20Health%20Equity%20Zip%20Codes%20Summary%20and%20Vaccinations%20Report.pdf. Accessed 5/27/2022. 2015-2019 American Community Survey 5-Year Estimates, Tables B17024, B19083, DP03.
Prepared by: County of San Diego, Health and Human Services Agency, Public Health Services, Community Health Statistics Unit, June 2022.

HHSA

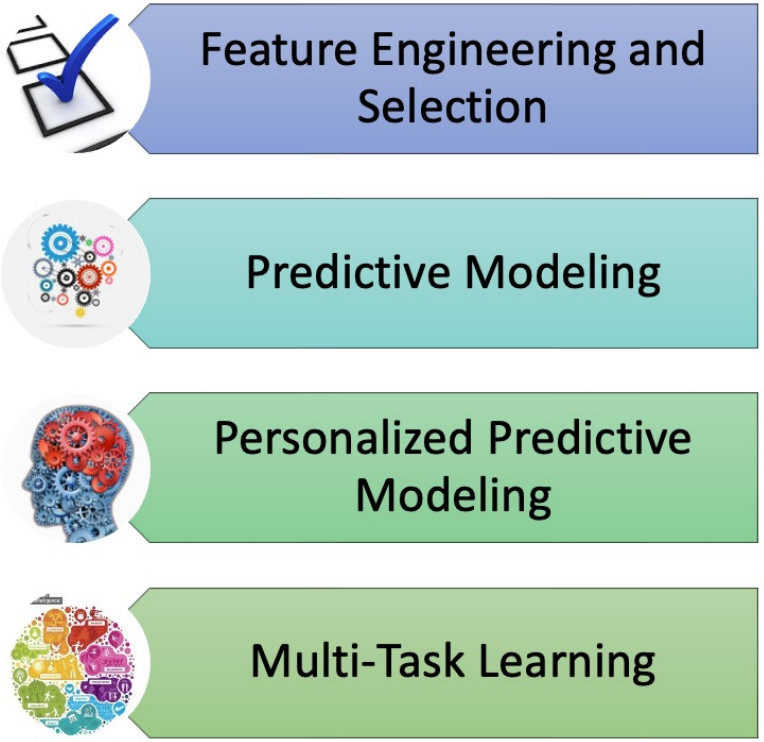
LIVE WELL SAN DIEGO

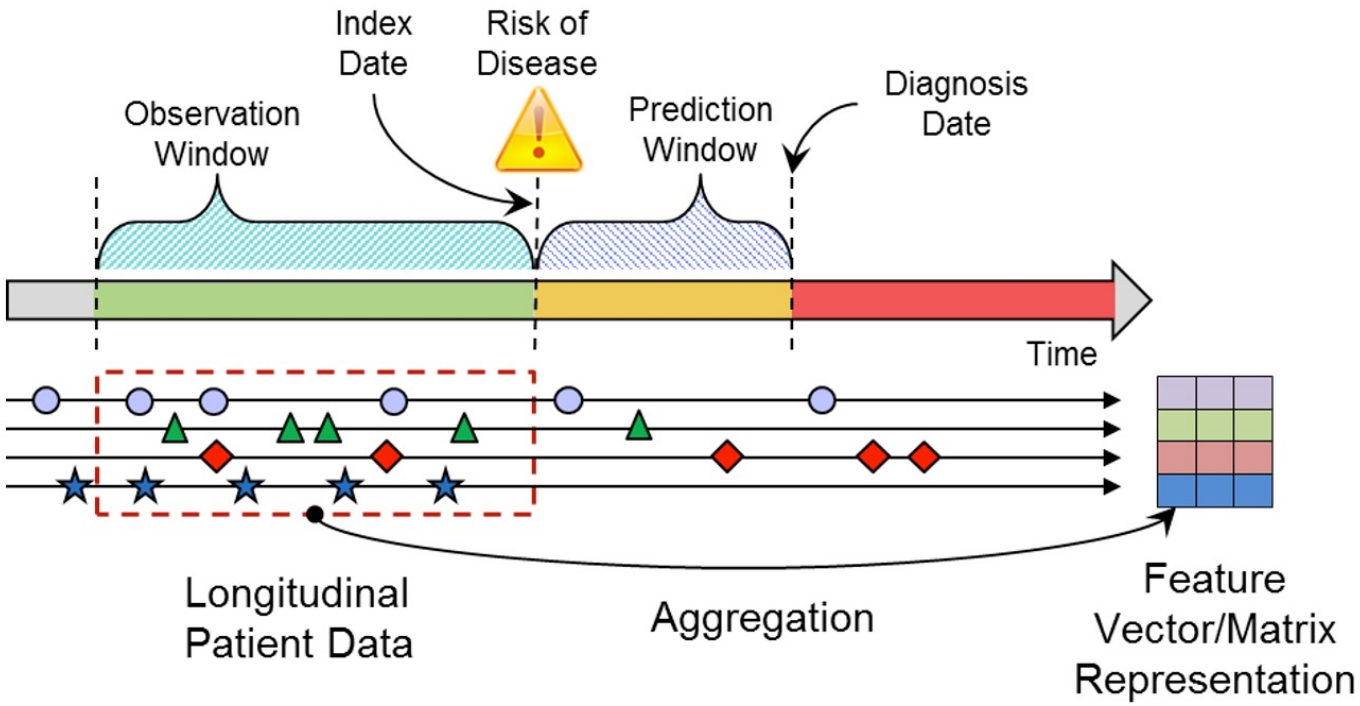
tableau

Population health management and patient-centered care

Developing Predictive Models, Risk and Clinical pathways e.g., diagnostics, making predictions and identifying patients at risk for worse outcomes

What is my patient's risk of developing condition X?



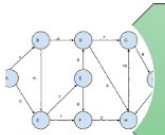


Courtesy: Kenny Ng

Question: What are my patient's top risk factors?

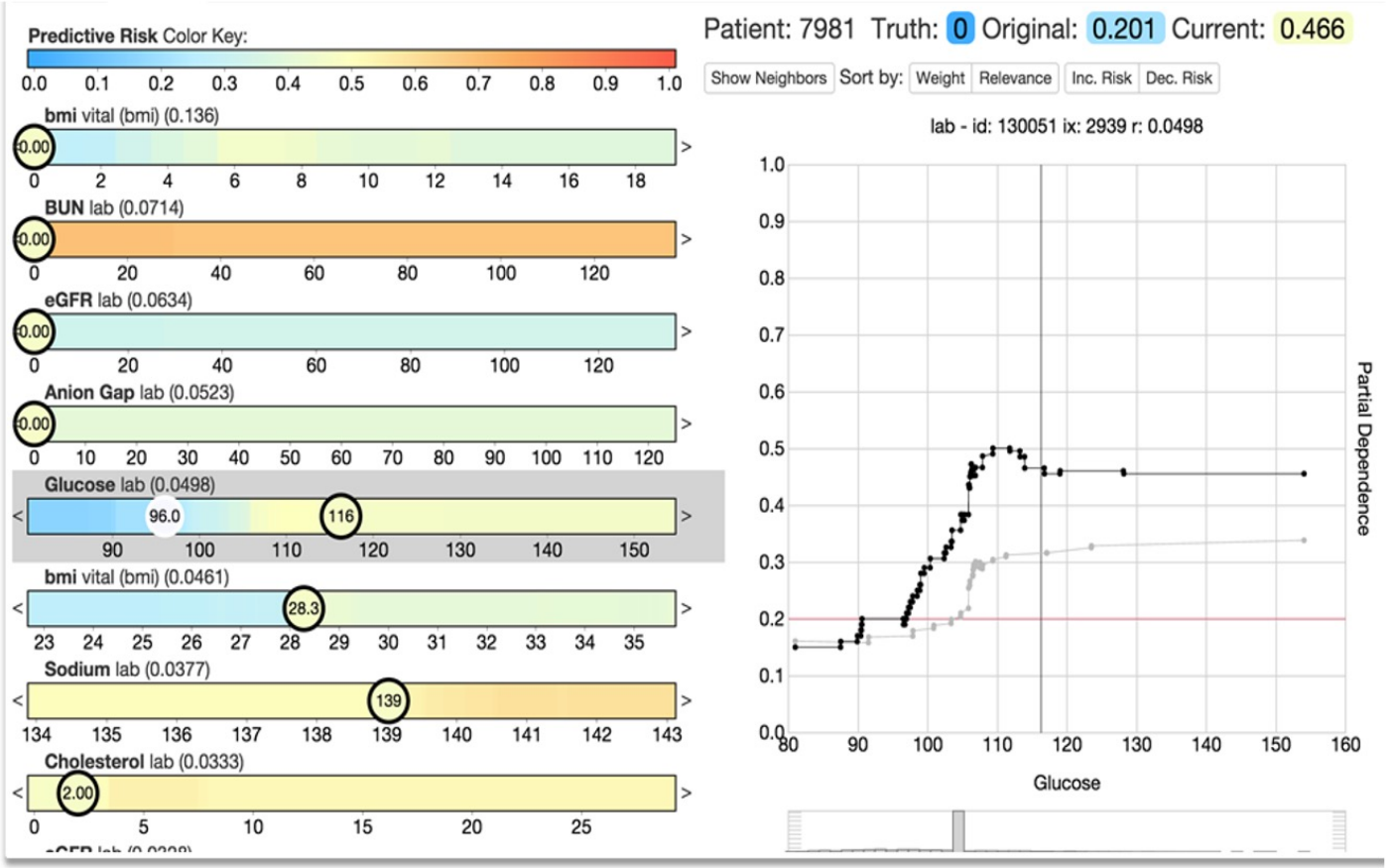
Feature Engineering

Feature and Model Selection



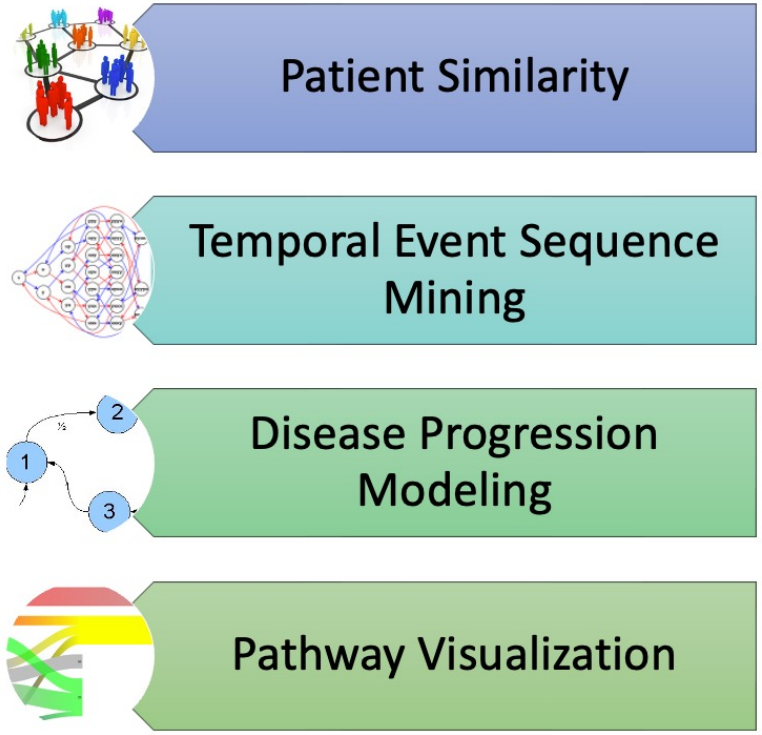
Causal Inference

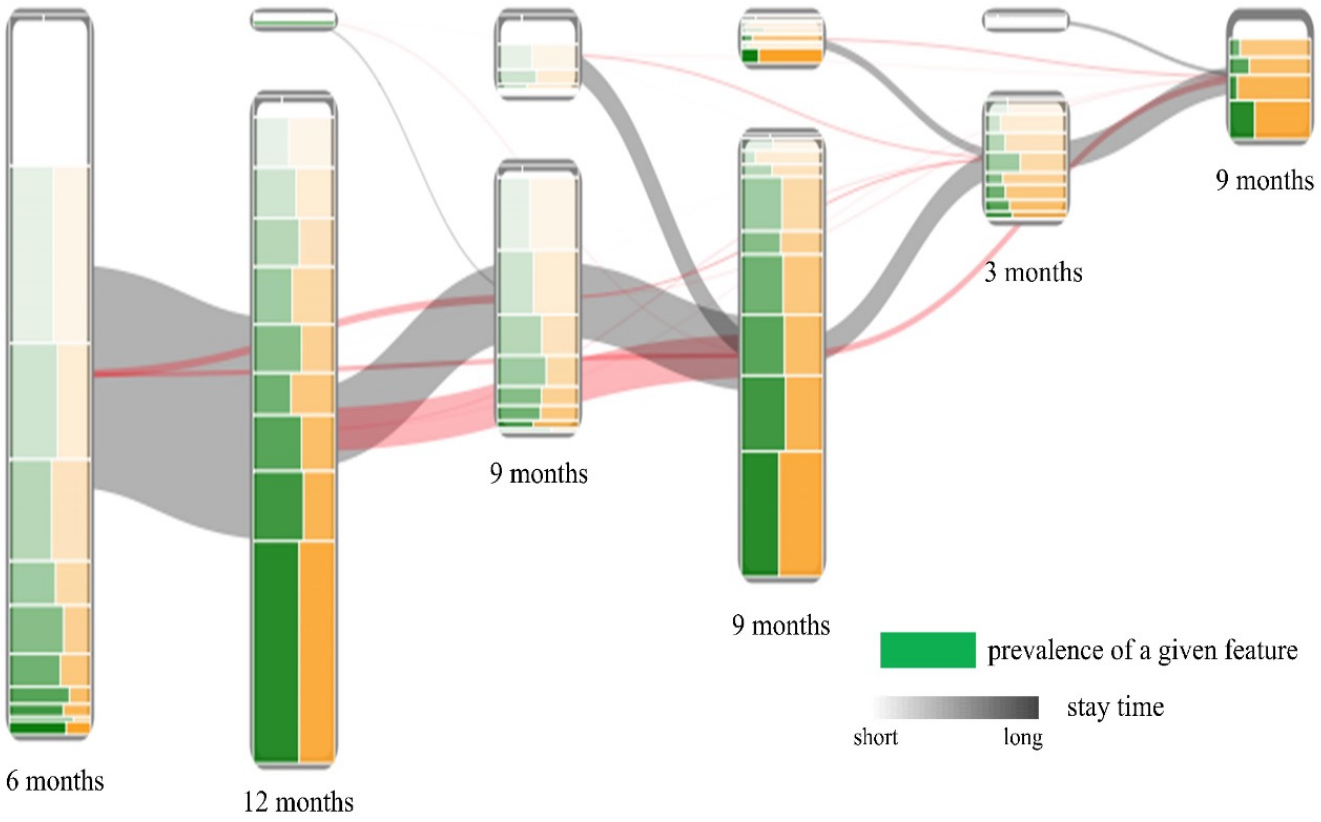
Model Introspection and Interpretation



Courtesy: Kenny Ng

Predict Clinical Pathways with a treatment plan or intervention
Question: What happened to patients with similar features as my patient with a chronic condition?





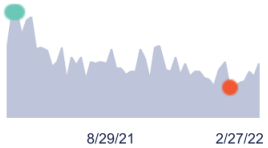
Courtesy: Kenny Ng

Population Health Management – Decision-support tools, Health Equity Dashboards

Patient 360
Overview

- Overview
- Patient Portal
- Download
- Show Filters

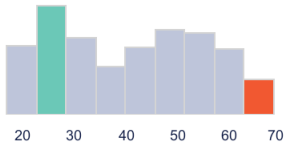
615
Total Patients



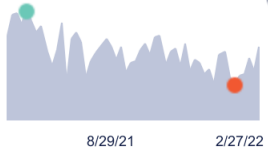
333
Female Patients

282
Male Patients

41 years old
Average Age

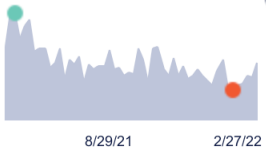


31,134
Total Procedures



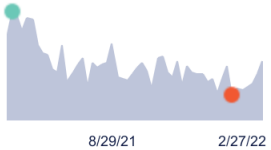
Assessment of health and social care needs (procedure)	2,531
Depression screening (procedure)	2,122
Depression screening using Patient Health Questionnaire..	1,905
Assessment of substance use (procedure)	1,779
Auscultation of the fetal heart	1,473

2,761
Total Encounters



Stress (finding)	559
Full-time employment (finding)	521
Limited social contact (finding)	377
Social isolation (finding)	365
Received higher education (finding)	357

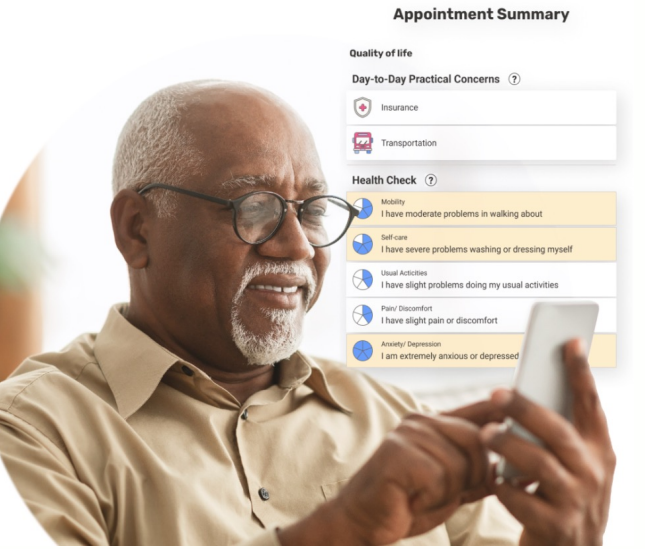
14,922
Total Visits



Hallmark Health System	776
Lowell General Hospital	547
Cambridge Health Alliance	532
Baystate Medical Center	513
Newton-Wellesley Hospital	448

Service Provider	Gender	Procedures	Encounters	Visits
Hallmark Health System	14 / 16	1,981	131	776
Signature Healthcare Brockton Hospital	6 / 9	1,586	42	359
Lowell General Hospital	11 / 12	1,466	68	547
Cambridge Health Alliance	7 / 18	1,421	84	532
Mount Auburn Hospital	10 / 12	1,165	74	413

Overall Average



Patient Discovery Connecting Life to Care Platform-©
Patient Discovery Solutions

Health Data Interoperability Brickbuilder by Lovelytics

Population Health Management and Decision-support tools

Patient Similarity

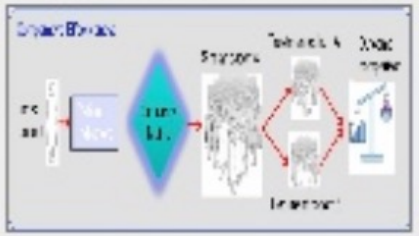
Objective
Find clinically similar patients for decision support and Comparative Effectiveness

Visualize Disease Pathways

Objective
Predict and visualize patient disease progression

Personalized Comparative Effectiveness

Objective
Identify most effective treatment option for a given patient

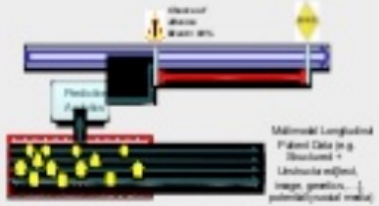


Visualize Population Cohorts

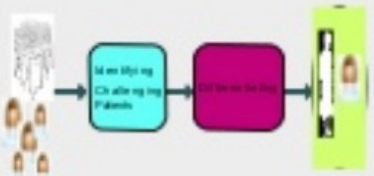
Objective
Visualize populations through interactive multi-dimensional exploration of inter-cluster and intra-cluster relationships

Predict Patient Clinical Pathway

Objective
Analyze patient's longitudinal records to model the future risk of developing adverse conditions



Patient / Provider Matching



Objective
Match patients with providers based on similarity analytics and optimal performance characteristics

And many more Clinical Decision-support tools in healthcare settings

- Medication management, drug interactions and toxicology – e.g., Micromedex®
- Alerts and Reminders
- Checklists and Recommendations
- Clinical workflow- e.g., cardiovascular care via workflow
- Bedside clinical assessment

AI Bias

AI bias is a general concept that refers to the fact that an AI system has been designed, intentionally or not, in a way that may make the system's decisions or use unfair.

Algorithm Model Assessments:

- Label Bias
- Modeling Bias
- Population Bias
- Measurement Bias
- Missing validation bias
- Human Use bias

Bias can be present both in the **algorithm** of the AI system and in the **data** used to train and test it. AI bias can emerge in an AI system as a result of **cultural, social, or institutional expectations**;
because of **technical limitations** of its **design**;
by being **used in unanticipated contexts** or by **making decisions about communities that are not considered in the initial design**.

Five Broad Aspects of Bias Across the Data Generation and Technology Development Continuum

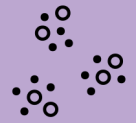
Evidence

Research bias:
Lack of equitable standards around how our science is funded, conducted, reviewed, published and disseminated; lack of inclusion in clinical trials and researcher diversity, evidence-base & real-world data

Experience/ Expertise

Provider bias:
Provider expertise and experience; cognitive biases and in-group biases;
Lack of health data insights and evidence; unconscious biases, pre-existing stereotypes or discriminatory practices from providers or health professionals

Exclusion



Embedded data bias:
incomplete health data, e.g., missing data or incomplete data in EHR's Favoring those groups who have robust health data profiles; Data bias in sample selection, modeling structure and selection of metrics for predictions Lack of cohort diversity; training data not representative

Environment

Data invisibility:
Lack of data on those important factors – such as the social determinants of health or other environmental factors that can trigger discriminatory outcomes

Empathy

Data empathy:
Lack of knowledge, understanding and/or experience about the people, places, factors that make up the data – unable to recognize the bias and optimize analysis; lack of knowledge of data source and real-world evidence or social implications

Ref: Dankwa-Mullan I, Weeraratne D., Artificial Intelligence and Machine Learning Technologies in Cancer Care: Addressing Disparities, Bias and Data Diversity, Cancer Discov. 2022 12(6)

Integrate Equity and Racial Justice Principles into Ethical AI Framework for tool development in healthcare

AI ethics

Goal is to understand how to **optimize AI's beneficial impact** while **reducing risks and adverse outcomes** for **all stakeholders** in a way that prioritizes human agency and well-being, as well as environmental flourishing.

Ethical AI Dimensions

- Accountability
- Impact of Algorithms
- Data Responsibility
- Design equity
- Discrimination and Bias
- Empathy
- Explainability
- Fairness
- Human Oversight

- Human Autonomy
- Inclusion
- Social Cohesion
- Inclusive Technology
- Moral Agency
- Privacy Protection
- Robustness, Safety
- Transparency and Trust
- Value Alignment

Dankwa-Mullan, I. Scheufele E., et al. "A Proposed Framework on Integrating Health Equity and Racial Justice into the Artificial Intelligence Development Lifecycle." *Journal of Health Care for the Poor and Underserved*, vol. 32 no. 2, 2021, p. 300-317..

The Human Part of Artificial Intelligence

Humanity and Empathy in Al and ML Technologies

Our patients and populations are the recipients of care and as such should be at the center of health care. The “high-tech’ is only part of the solution. Our AI and machine learning technologies should function at the service of humanity.

Thank you!

Irene Dankwa-Mullan MD MPH
Email: Irene.Dankwamullan@gwu.edu
LinkedIn - irene-dankwa-mullan-md-mph-98324027/

37