

Appsilon

ET13

Health Equity Explorer

Technical Deep Dive into an R Shiny Application for Revealing
Structural Inequity and Systematic Racism in Healthcare

Pavel Demin

Senior Full Stack Developer at Appsilon

Agenda

1. About authors
2. Project overview
3. Application demonstration
4. System design and architecture
5. Technical highlights
6. The future
7. Q&A

Authors

Pavel Demin

Full Stack Developer at Appsilon

5+ years of experience

R, Python, Javascript, Typescript... Markdown!



William Adams, MD

Pediatrician, epidemiologist, and medical informatician at Boston Medical Center (BMC).

Professor at Boston University

Director of Clinical Research Informatics for Boston University
Director of Child Health Informatics for the Boston University



Disclaimer

All numbers presented on the screenshots from the application are fictional. The demonstration is made using the development build connected to a synthetic database, where data is often *realistic* but sometimes completely random.

None of the demonstrated figures represent real world data.

Project Overview

Project Overview

The purpose

“Advancing health equity is a national priority, and the fundamental causes of health inequity, such as racism, are increasingly being recognized as public health crises”

Adams WG, Gasman S, Beccia AL, and Fuentes L. The Health Equity Explorer: An open-source resource for distributed health equity visualization and research across common data models



Project Overview

The purpose

Install and use Health Equity Explorer at various medical and research facilities to identify structural and individual racism - as well as other factors contributing to health inequity.



Project Overview

The purpose

Install and use Health Equity Explorer **at various** medical and research **facilities** to identify structural and individual racism - as well as other factors contributing to health equity disbalance.



Project Overview

The purpose

Install and use Health Equity Explorer **at various** medical and research **facilities** to identify structural and individual racism - as well as other factors contributing to health equity disbalance.

Direct effect on architectural and technical decisions discussed in the system design section later.



Application Demonstration

Application Demonstration

 BMC H2E

 pavel

App details

Restart app

Stop app

Sign Out

 **H2E**
Health Equity Explorer

Home

Health Outcomes

Advanced Analytics

Data Exploration

Statistical Inference

Machine Learning

Neighborhood Data

About the Health Equity Explorer

The health care we receive often has only a small impact on our overall health. Combining electronic health record (EHR) data with place-based data about the neighborhoods where we live, learn, and work can help us better understand the many factors that drive human health and well-being.

- [Health Outcomes](#): explore trends in health outcomes over time by different patient populations
- [Advanced Analytics](#): build a multivariate logistic regression model to explore relationships between variables and outcomes
- [Neighborhood Data](#): visualize how health outcomes and drivers of health differ by neighborhood

Limitations

Results from this tool are not to be used for patient care or for research purposes. To request access to this data for research purposes, please contact Bill Adams.

Documentation

You can find more information in [official documentation](#).

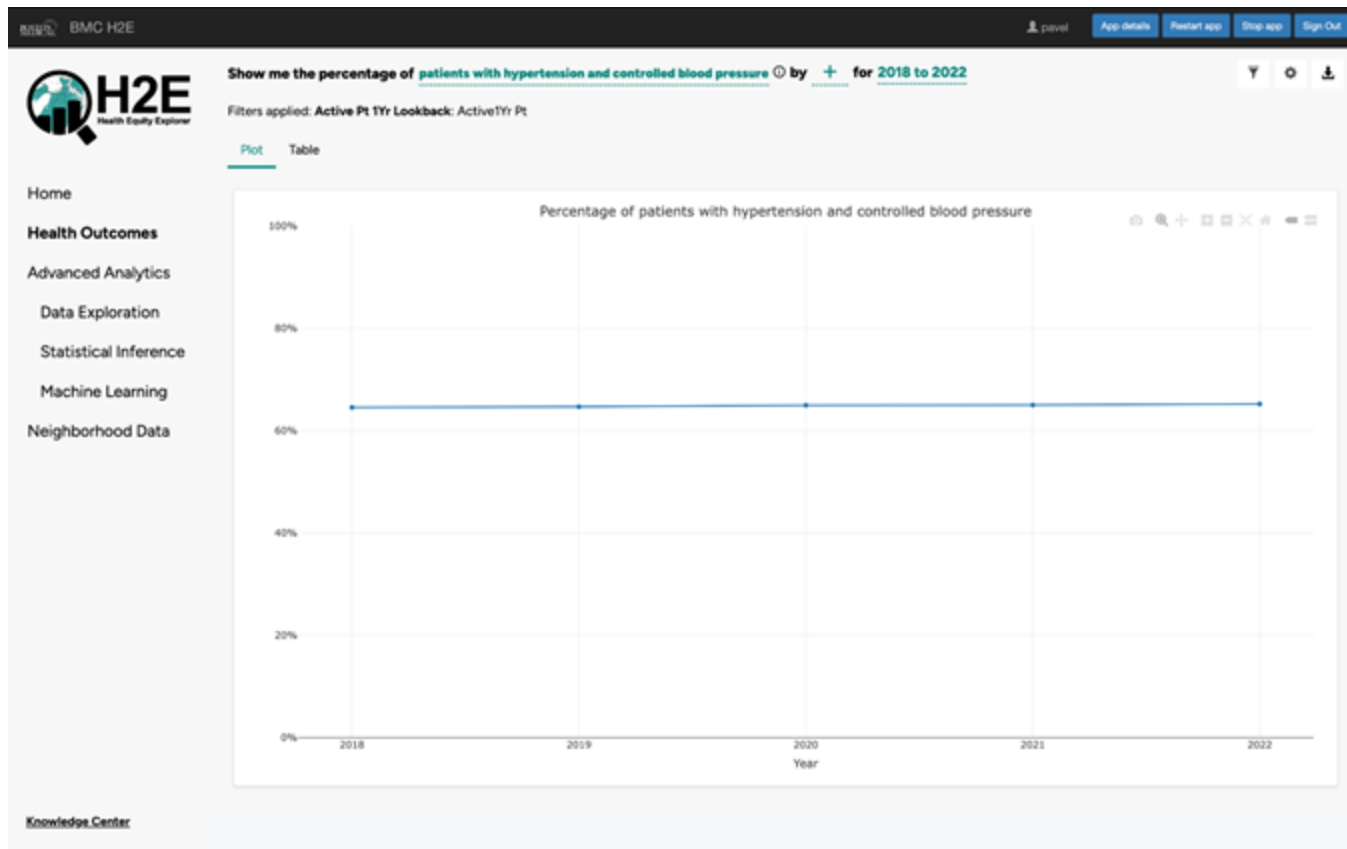
Licensing and Acknowledgements

Please review licensing details in the [licensing and acknowledgement document](#).

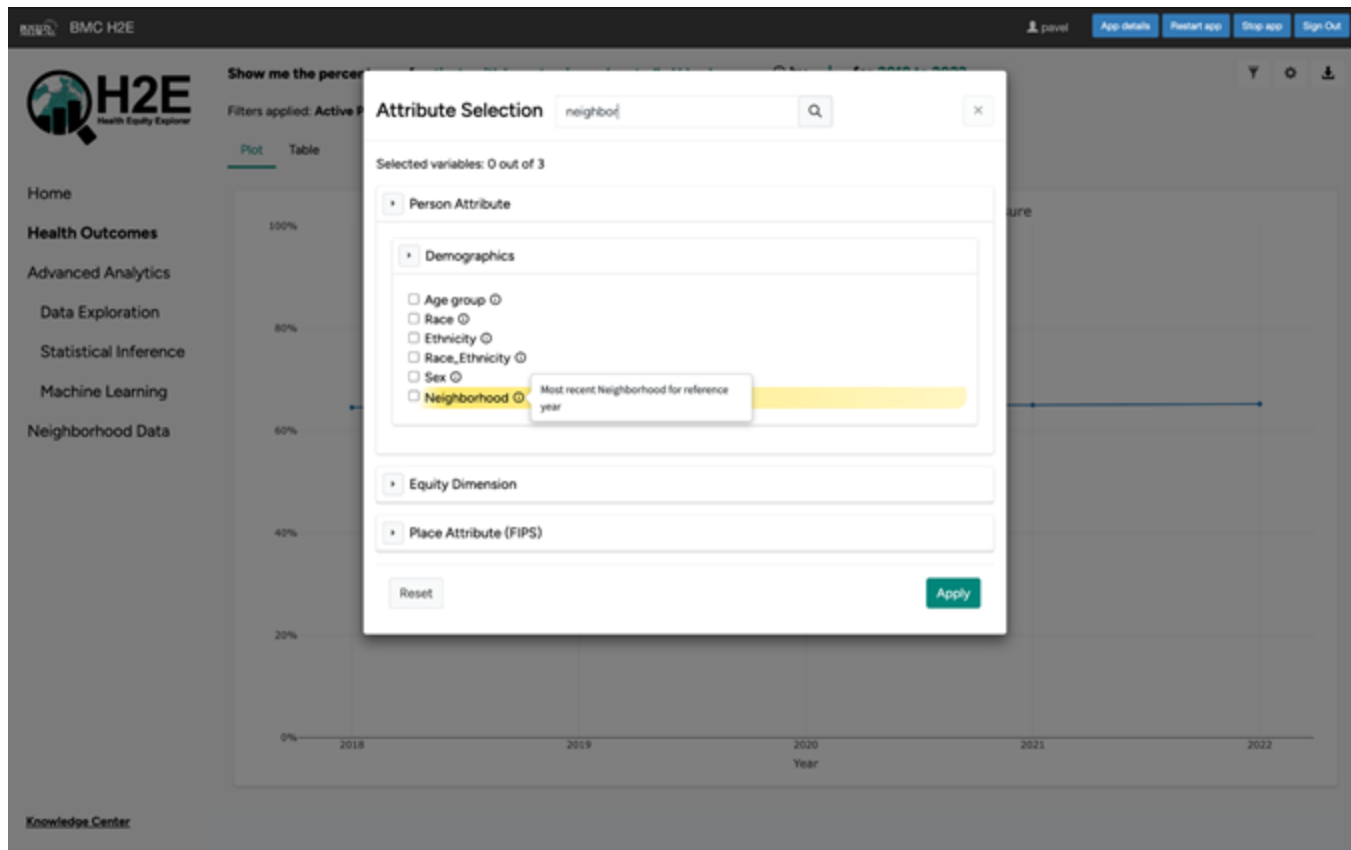
Knowledge Center



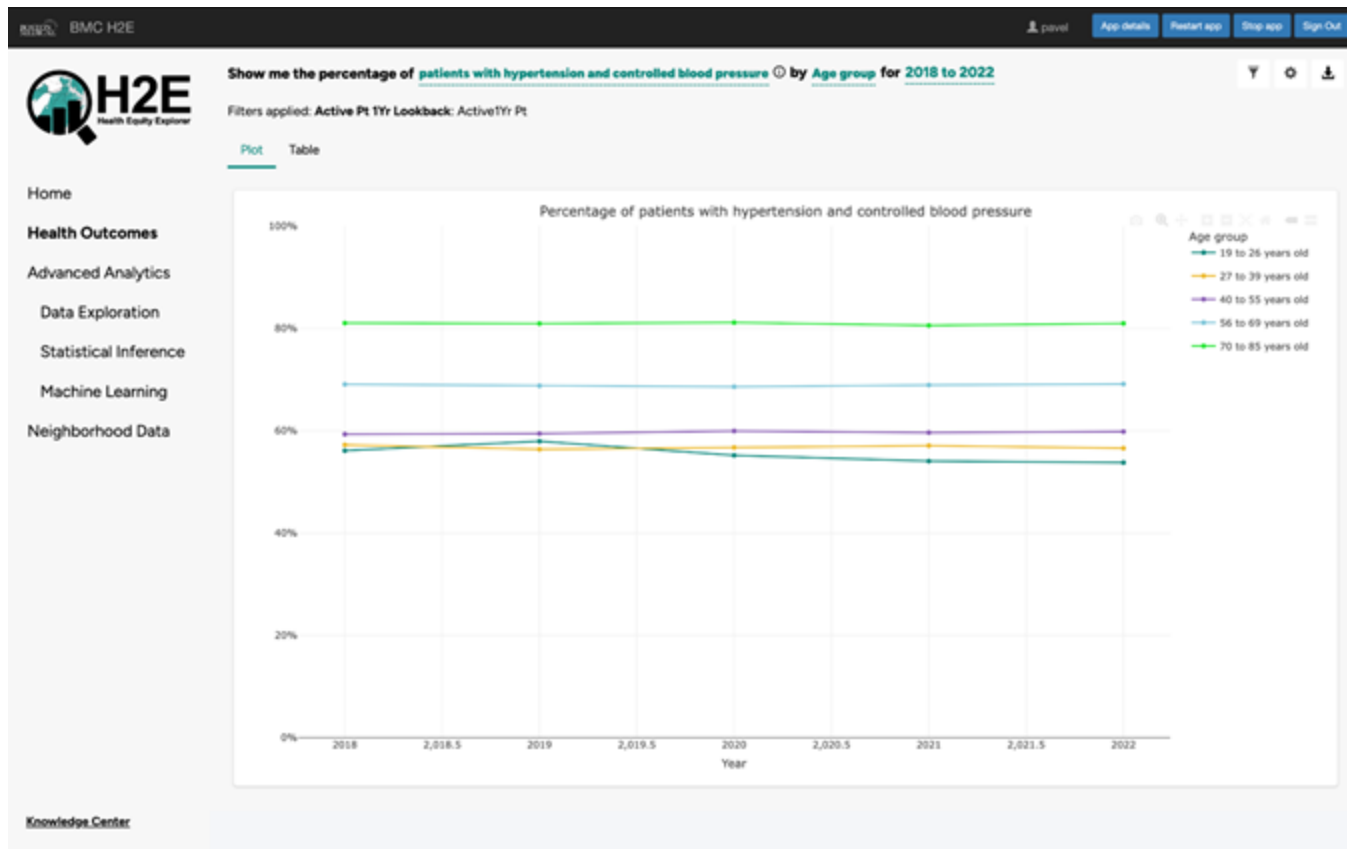
Application Demonstration



Application Demonstration



Application Demonstration



Application Demonstration

 BMC H2E

[App details](#)
[Restart app](#)
[Stop app](#)
[Sign Out](#)





Show me the percentage of patients with hypertension and controlled blood pressure by Age group for 2018 to 2022

[Plot](#)
[Table](#)

Filters applied: Active Pt 1Yr Lookback: Active1Yr Pt

Age group	2018	2019	2020	2021	2022
19 to 26 years old	56.1% 957	57.9% 963	55.2% 942	54.1% 910	53.8% 885
27 to 39 years old	57.2% 3,272	56.4% 3,345	56.7% 3,453	57.1% 3,568	56.6% 3,591
40 to 55 years old	59.3% 7,761	59.4% 7,635	60.0% 7,583	59.6% 7,286	59.8% 7,136
56 to 69 years old	69.1% 6,040	68.8% 6,193	68.6% 6,419	68.9% 6,497	69.1% 6,659
70 to 85 years old	81.1% 2,777	81.0% 3,046	81.2% 3,274	80.6% 3,343	81.0% 3,445

☐ Show Missing Data
 % meeting metric / Screened patients

Home

Health Outcomes

 Advanced Analytics

 Data Exploration

 Statistical Inference

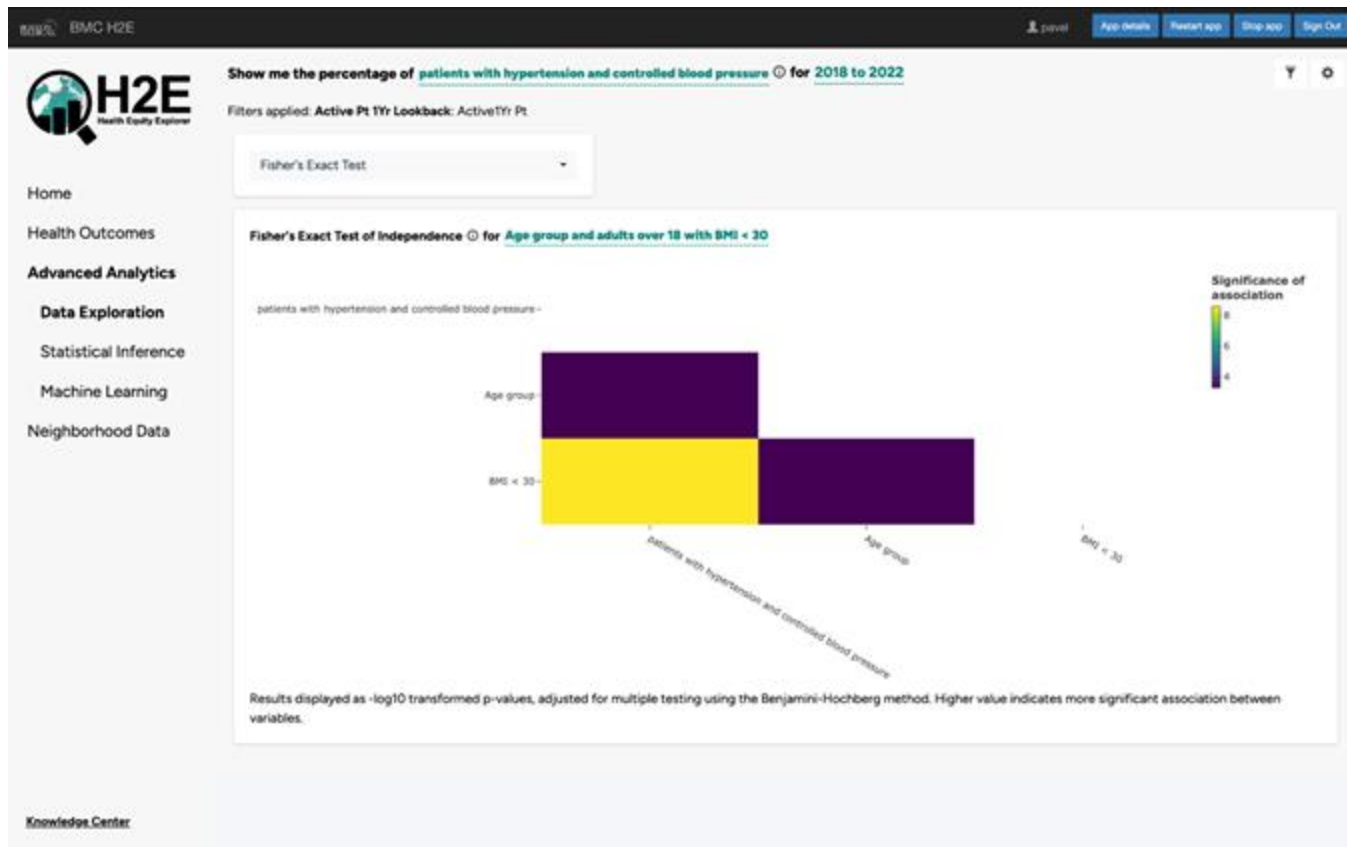
 Machine Learning

 Neighborhood Data

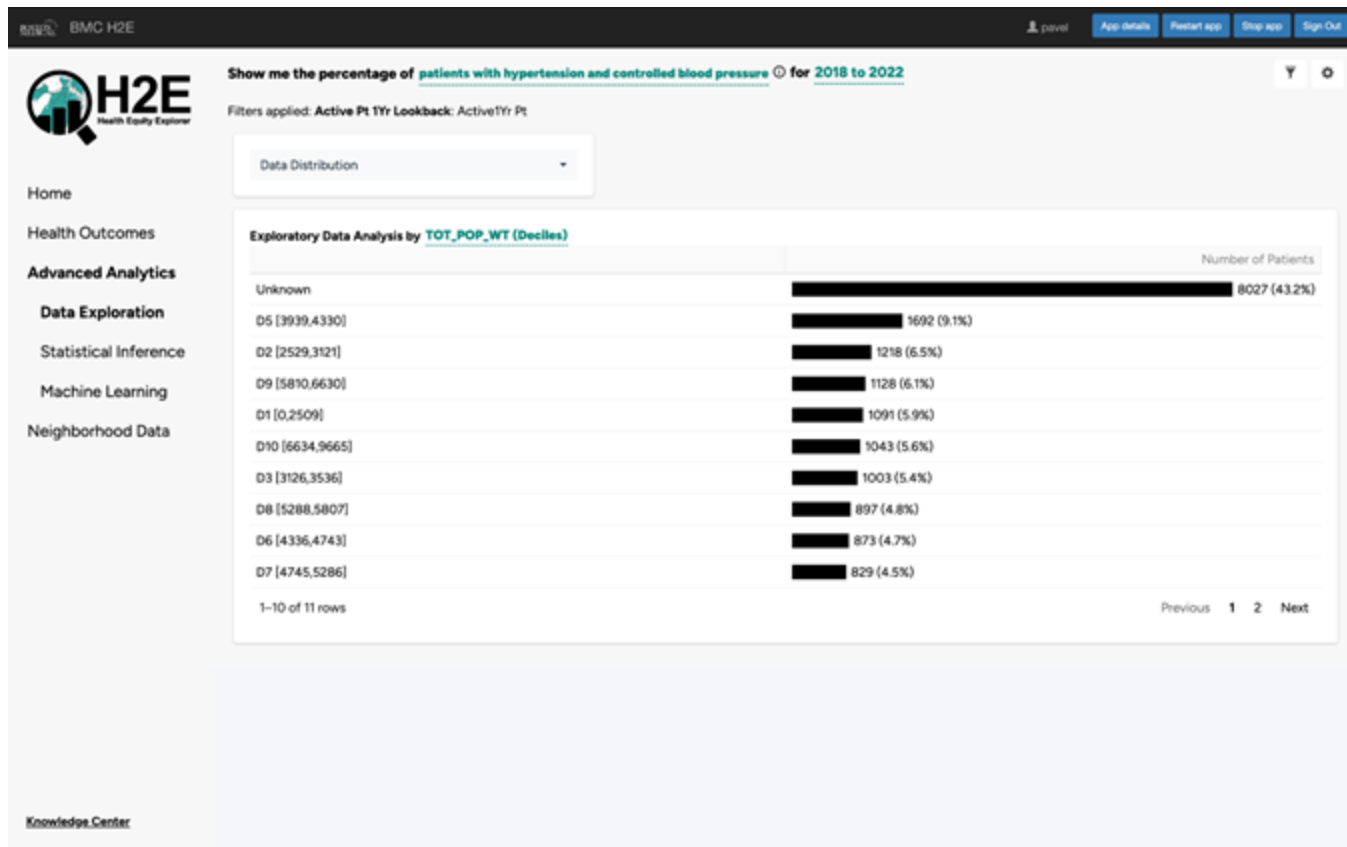
Knowledge Center



Application Demonstration



Application Demonstration



Application Demonstration



[Home](#)
[Health Outcomes](#)
[Advanced Analytics](#)
[Data Exploration](#)
[Statistical Inference](#)
[Machine Learning](#)
[Neighborhood Data](#)

[Show me the percentage of patients with hypertension and controlled blood pressure](#) for 2018 to 2022

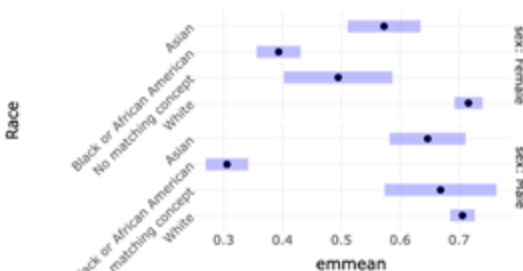
Filters applied: Active Pt Yr Lookback: ActiveYr Pt

Multivariate Regression

- Model [patients with hypertension and controlled blood pressure](#) using [Race and Sex](#) where the primary variable of interest is [Race](#)
- Group the output by Optional [Sex](#)
- Create an interaction variable between Optional [Race and Sex](#)

Go

$$\log \left[\frac{P(\text{status} = 1)}{1 - P(\text{status} = 1)} \right] = \alpha + \beta_1(\text{race}_{\text{Black or African American}}) + \beta_2(\text{race}_{\text{No matching concept}}) + \beta_3(\text{race}_{\text{White}}) + \beta_4(\text{sex}_{\text{Male}}) + \beta_5(\text{race}_{\text{Black or African American}} \times \text{sex}_{\text{Male}}) + \beta_6(\text{race}_{\text{No matching concept}} \times \text{sex}_{\text{Male}}) + \beta_7(\text{race}_{\text{White}} \times \text{sex}_{\text{Male}})$$



race	sex	emmean	SE	df	Lower CI	Upper CI
Asian	Female	0.57	0.03	Inf	0.51	0.63
Black or African American	Female	0.39	0.02	Inf	0.36	0.43
No matching concept	Female	0.49	0.05	Inf	0.40	0.59
White	Female	0.72	0.01	Inf	0.69	0.74
Asian	Male	0.65	0.03	Inf	0.58	0.71
Black or African American	Male	0.31	0.02	Inf	0.27	0.34
No matching concept	Male	0.67	0.05	Inf	0.57	0.76
White	Male	0.71	0.01	Inf	0.69	0.73

Show EMM as [logit](#)

- Results are given on the logit scale.
- Confidence level used: 0.95

Knowledge Center

Application Demonstration



[Home](#)
[Health Outcomes](#)
[Advanced Analytics](#)

[Data Exploration](#)
[Statistical Inference](#)
[Machine Learning](#)

[Neighborhood Data](#)



[App details](#)
[Restart app](#)
[Stop app](#)
[Sign Out](#)

Show me the percentage of patients with hypertension and controlled blood pressure for 2018 to 2022

Filters applied: Active Pt 1Yr Lookback: Active 1Yr Pt

About

Machine Learning

To get started, choose a module from the dropdown above.

[Learn more about available modules by expanding sections below.](#)

▼ About Feature Importance Module

The feature importance module helps you identify the valuable features in predicting the target outcome. The **Variable Importance Plot (VIP)** highlights these features in order of their **Importance** values. This outcome is achieved by employing the Random Forest machine learning algorithm.

In technical terms, the **Importance** value of a **Variable** is the total reduction in the **Gini impurity** brought by that **Variable** across all Decision Trees in the Random Forest. Hence, a higher mean decrease in impurity indicates that the feature is more important. Values presented in the module do not have a unit of measure.

Module inputs

- **Number of trees:** Indicates the number of Decision Trees you want to build in the Random Forest model. A higher number can provide a more accurate outcome, but can also significantly increase computational time depending upon the sample size. By default we use 501 decision trees. Notice that the number is odd - it helps to avoid a decision parity.
- **Sample size:** Allows the user to choose the data sample size. A larger sample size will result in longer computational time for two reasons: more data to download, and more data to train the model on.
- **Variable selection options:** Provides the user with the ability to control which variables are used in the model. By default, we will try to compute feature importance across all variables available to us. However, the user can either exclude certain variables from the calculation, or even provide a limited subset of variables.

Knowledge Center

Application Demonstration



Home

Health Outcomes

Advanced Analytics

Data Exploration

Statistical Inference

Machine Learning

Neighborhood Data

Knowledge Center

BMC H2E

person | [App details](#) | [Restart app](#) | [Stop app](#) | [Sign Out](#)

Show me the percentage of patients with hypertension and controlled blood pressure for 2018 to 2022

Filters applied: Active Pt 1Yr Lookback: Active1Yr Pt

Feature Importance

Number of trees: 501 Sample size: 2500

▼ Show Variable Selection Options

Use specific variables for model training?

☐ No

Exclude variables: patients with hypertension and controlled blood pressure and Neighborhood

[Re-run model](#)

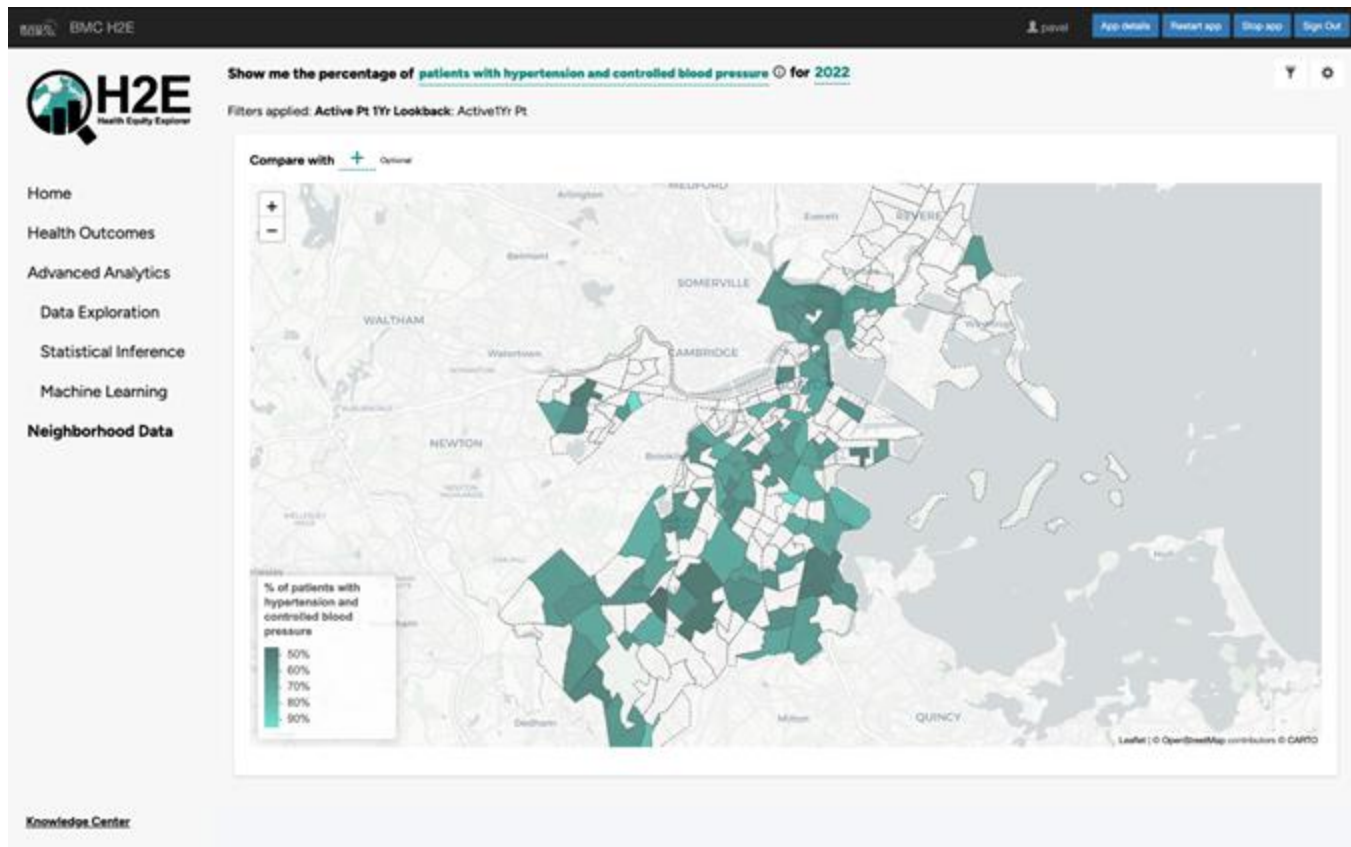
Variable Importance

Results based on running the model 1 time(s)

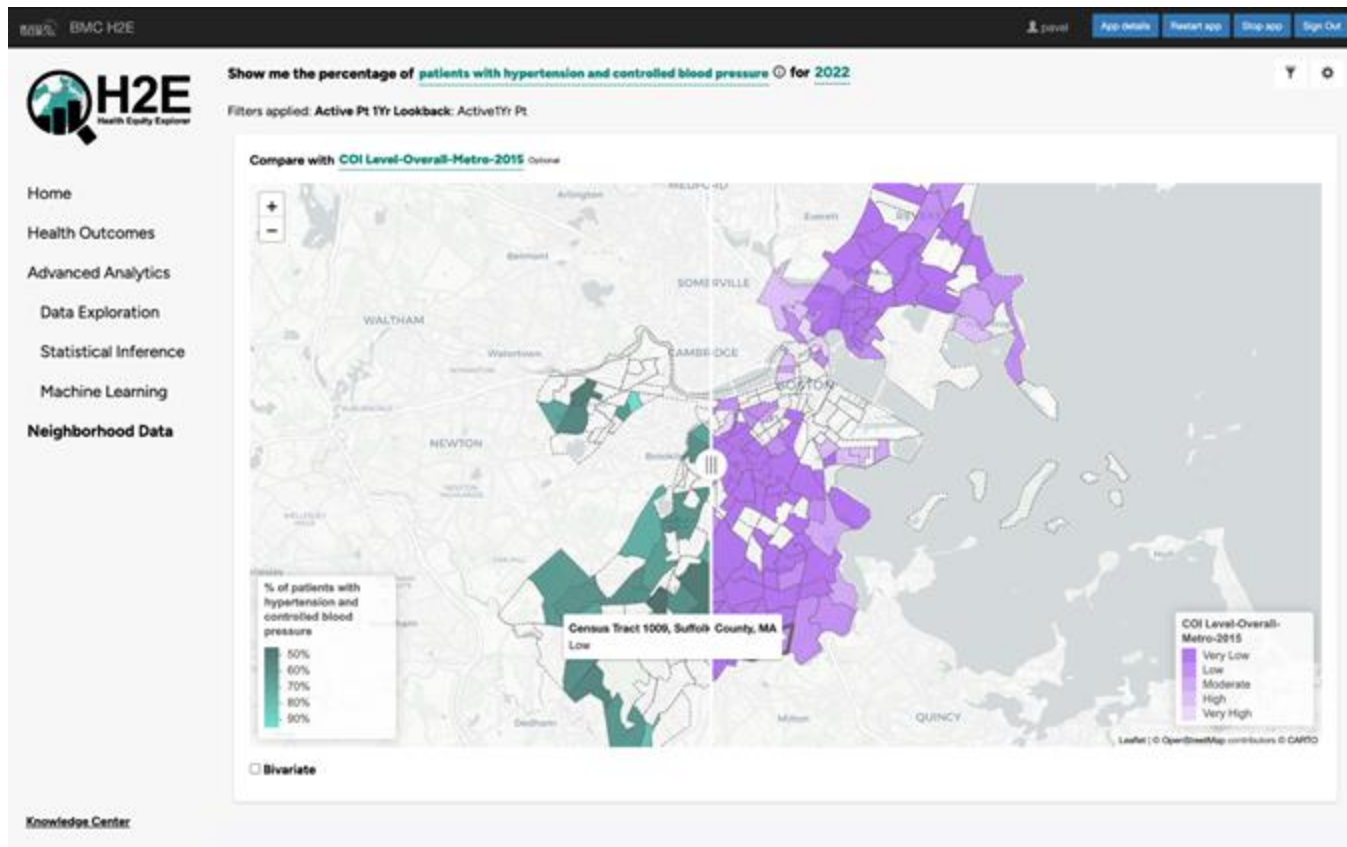
Search

Variable	Importance
age	4.61
Race_Ethnicity	2.81
Race	1.93
Age group	1.7
PCT_PUB_COMMT_59MIN	1.49
MEDIAN_HH_INC	1.09
PCT_HH_FOOD_STMP_BLW_POV	1.07

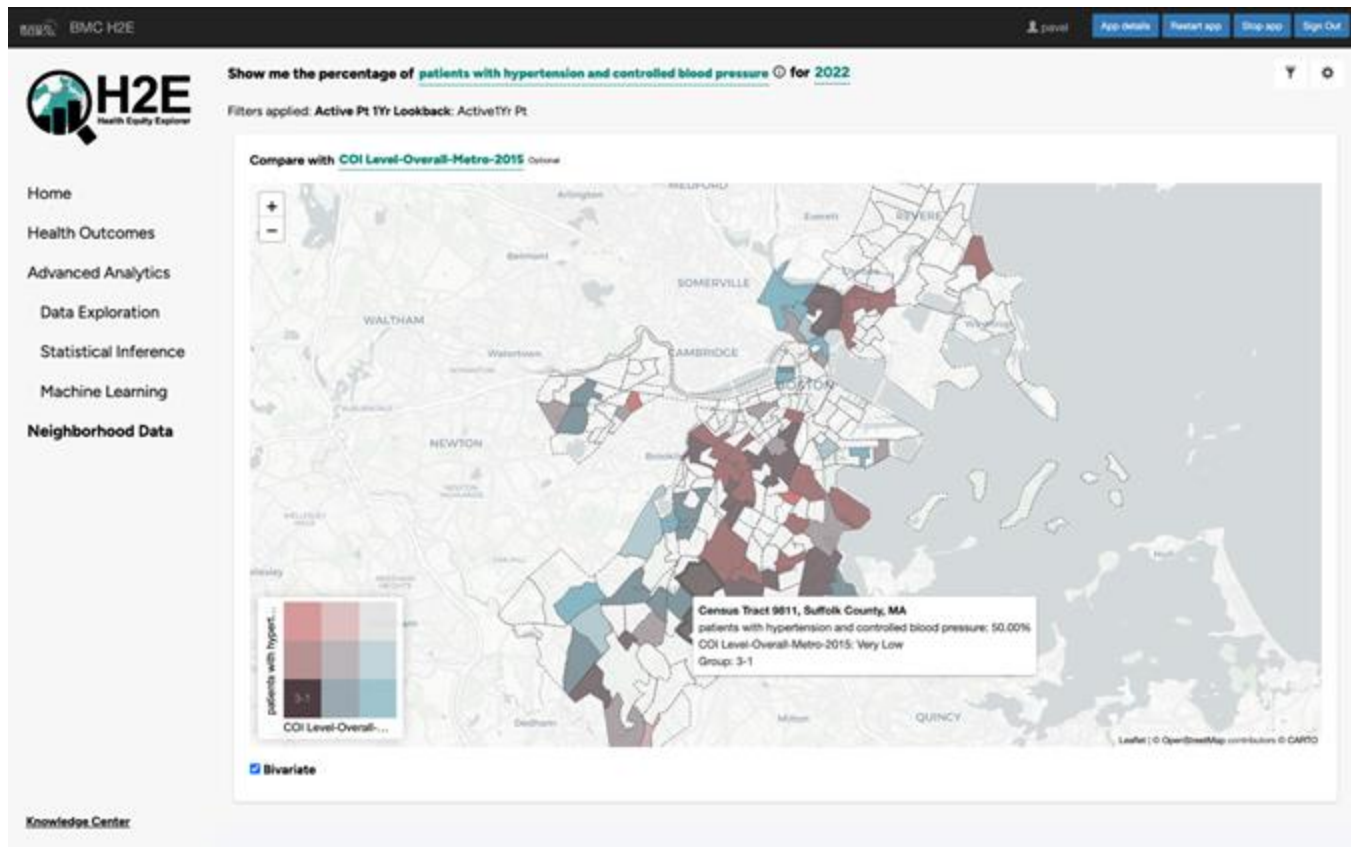
Application Demonstration



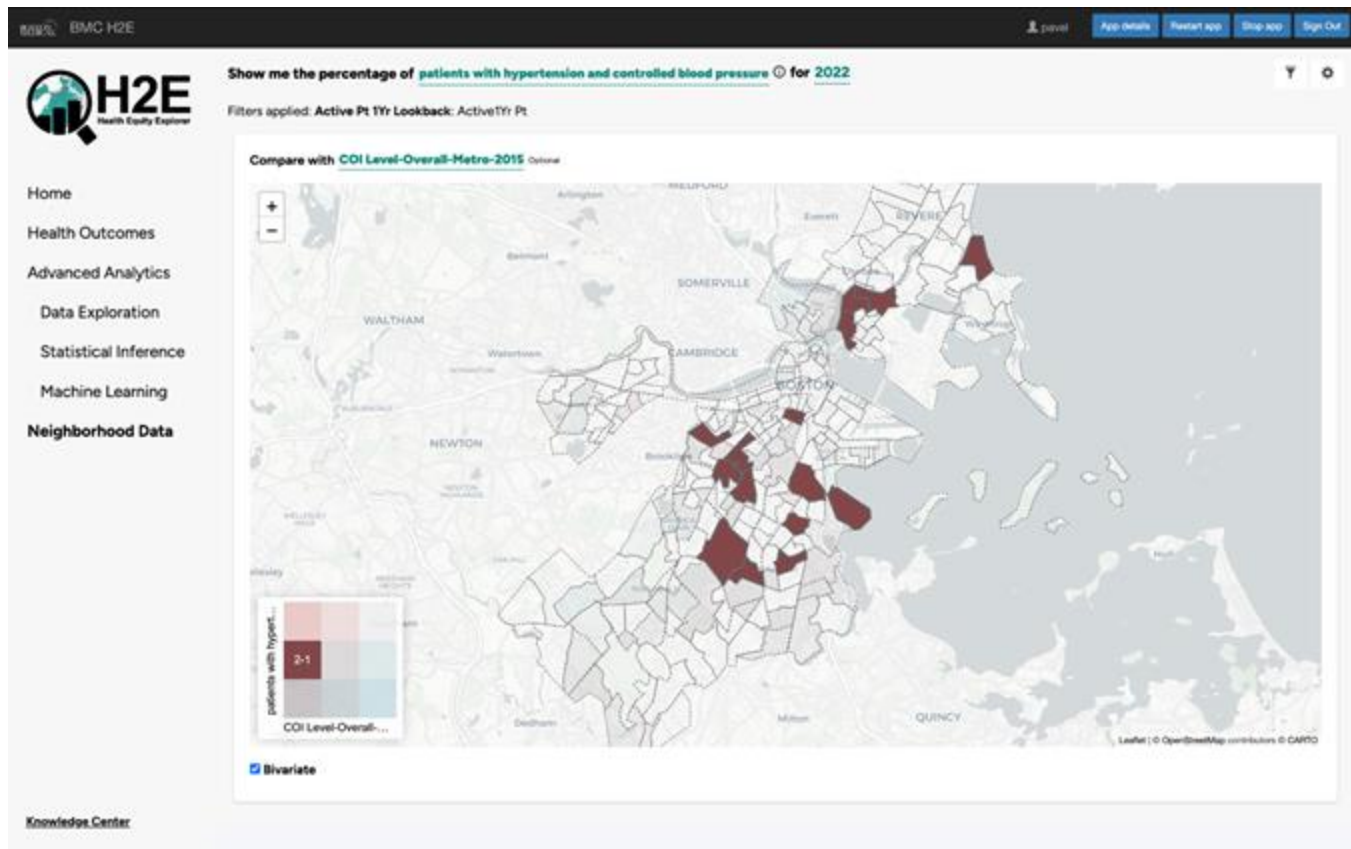
Application Demonstration



Application Demonstration

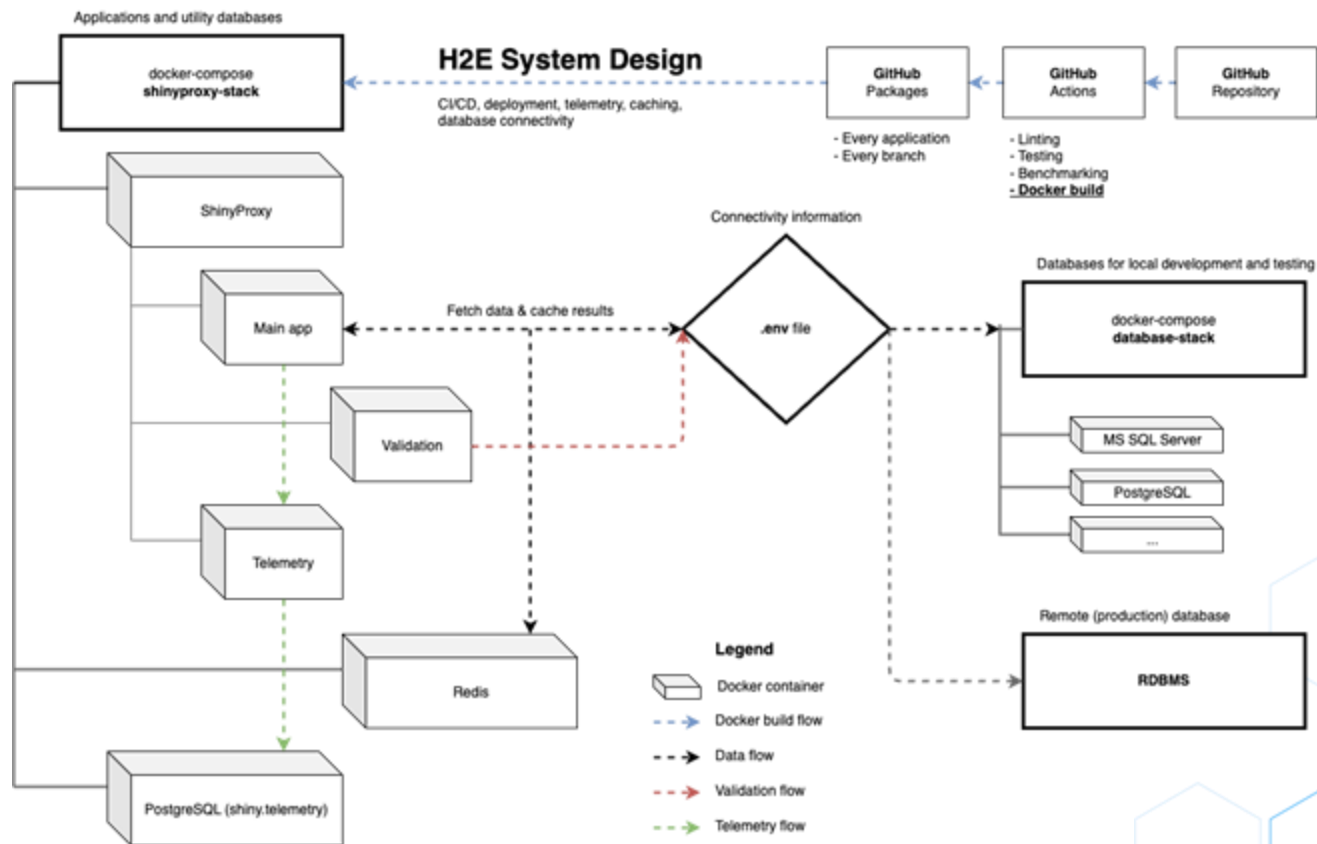


Application Demonstration



System Design and Architecture

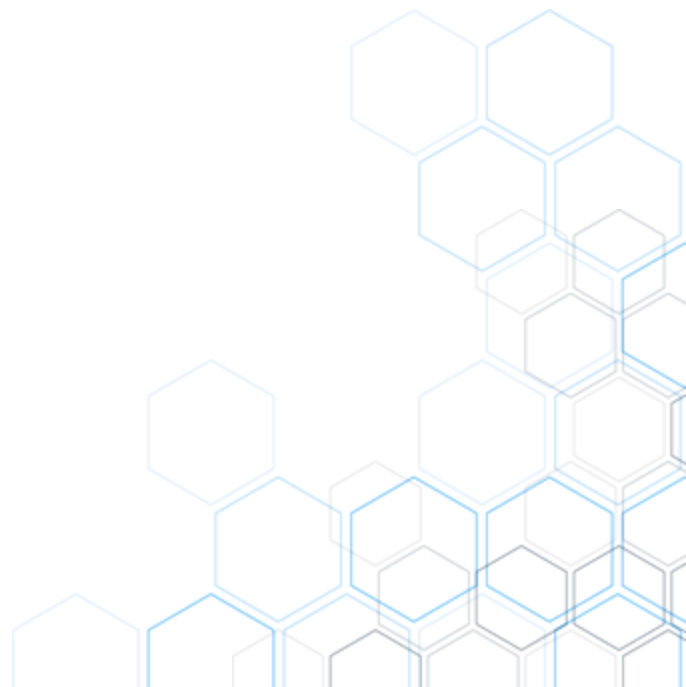
System Design and Architecture



Technical Highlights

Technical Highlights

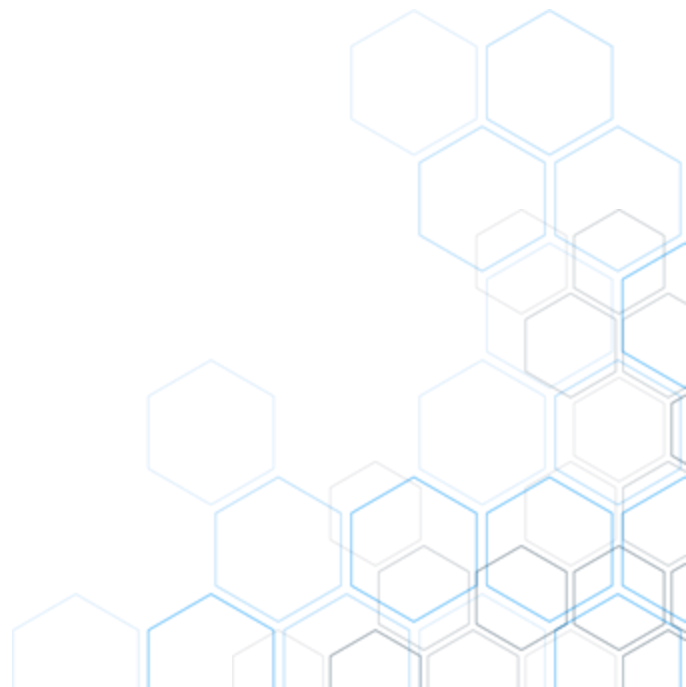
Profiling, Optimization, and Benchmarking



Technical Highlights

Profiling, Optimization, and Benchmarking

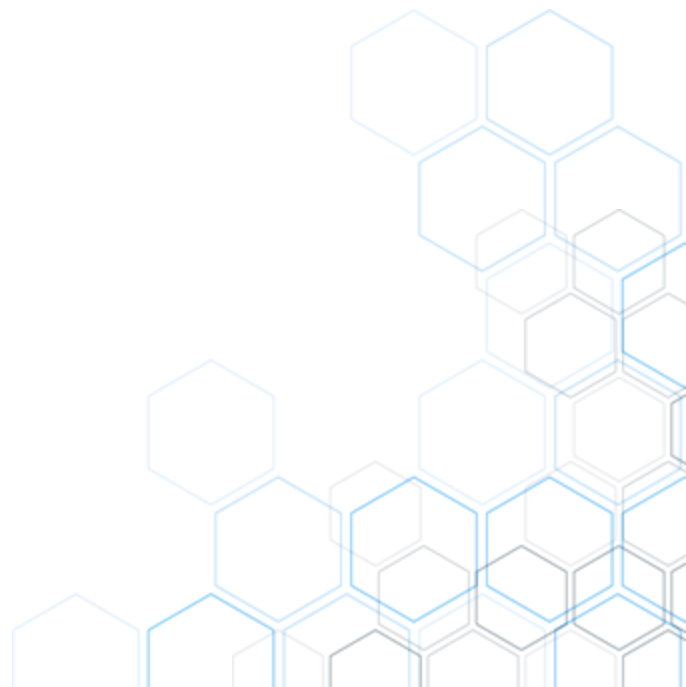
Intra-session measurements using Cypress



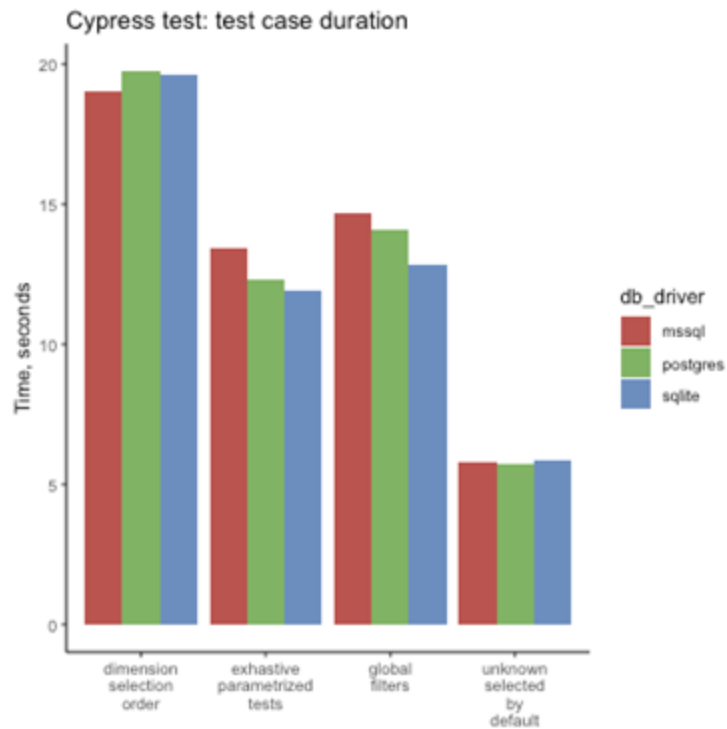
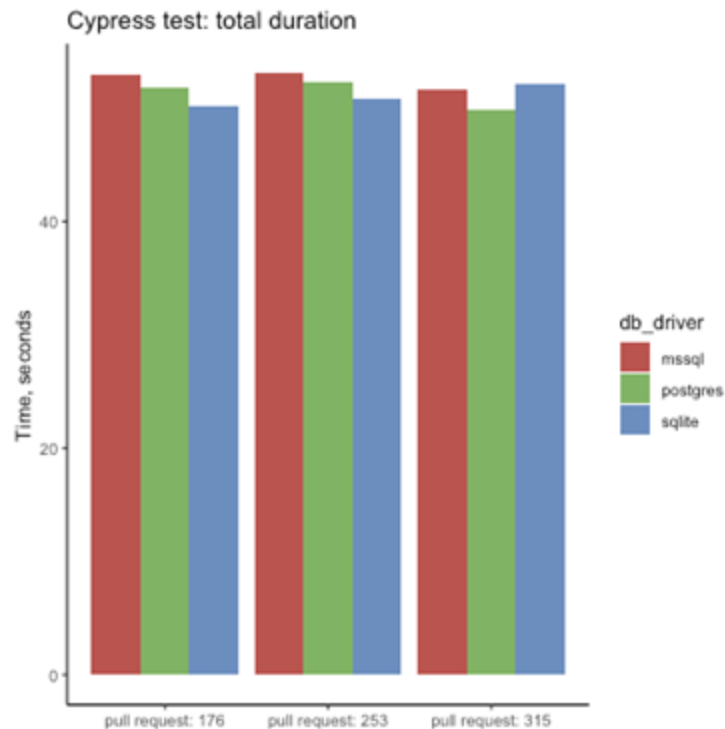
Technical Highlights Profiling, Optimization, and Benchmarking

Intra-session measurements using Cypress

- Teamcity reporter
- Logs saved in a file
- Parsed by an R script
- Timestamps extracted
- Test duration calculated and visualized
- Automatically in CI

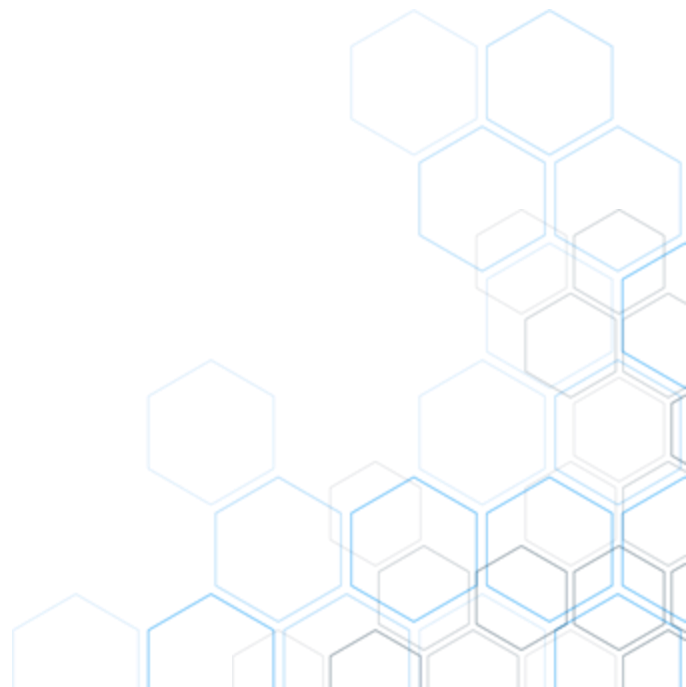


Technical Highlights Profiling, Optimization, and Benchmarking



Technical Highlights

Custom React Component

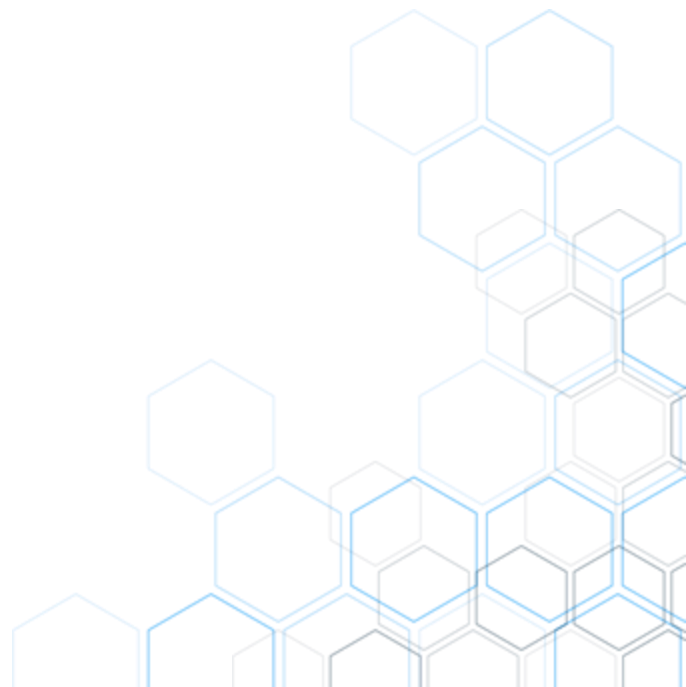


Technical Highlights

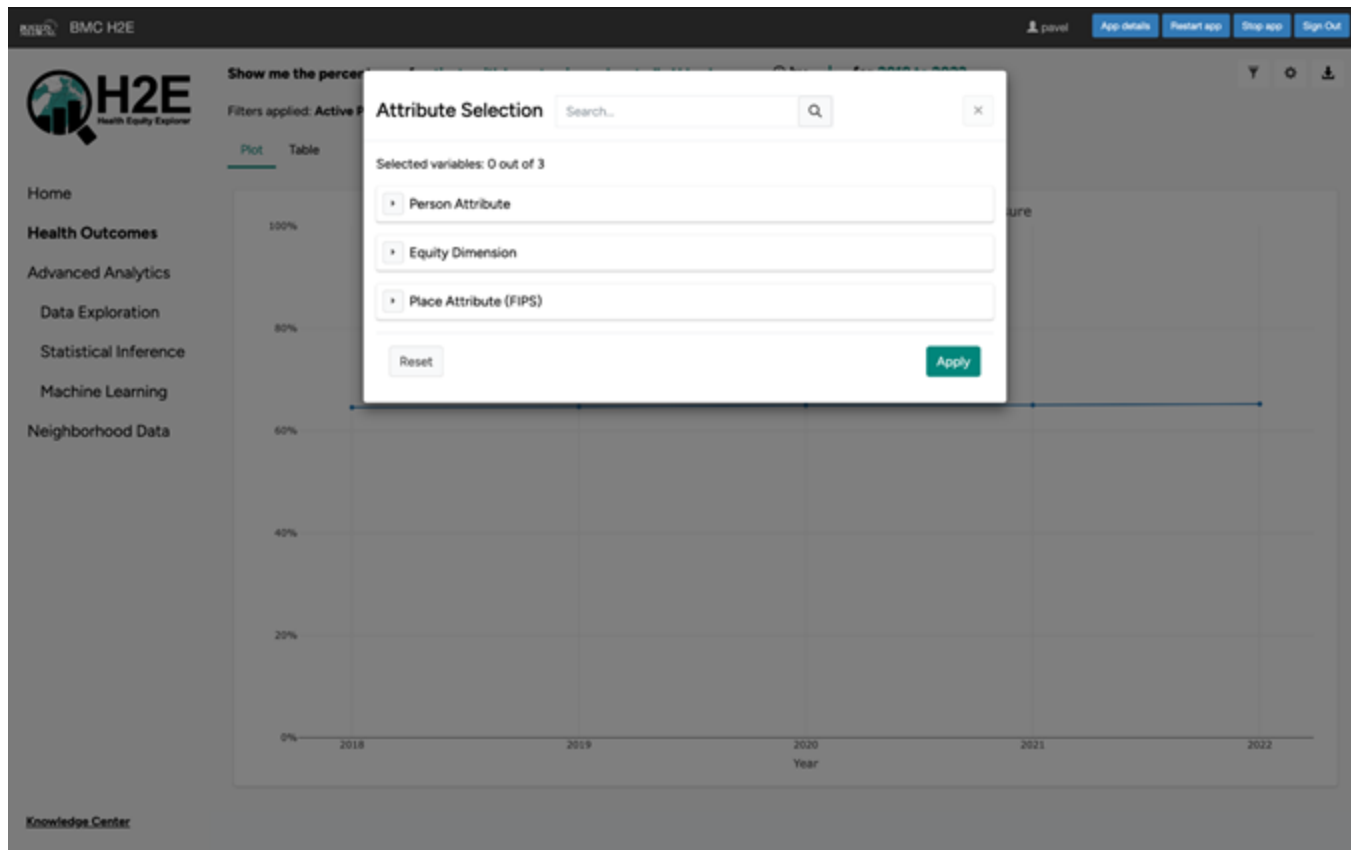
Custom React Component

Why React?

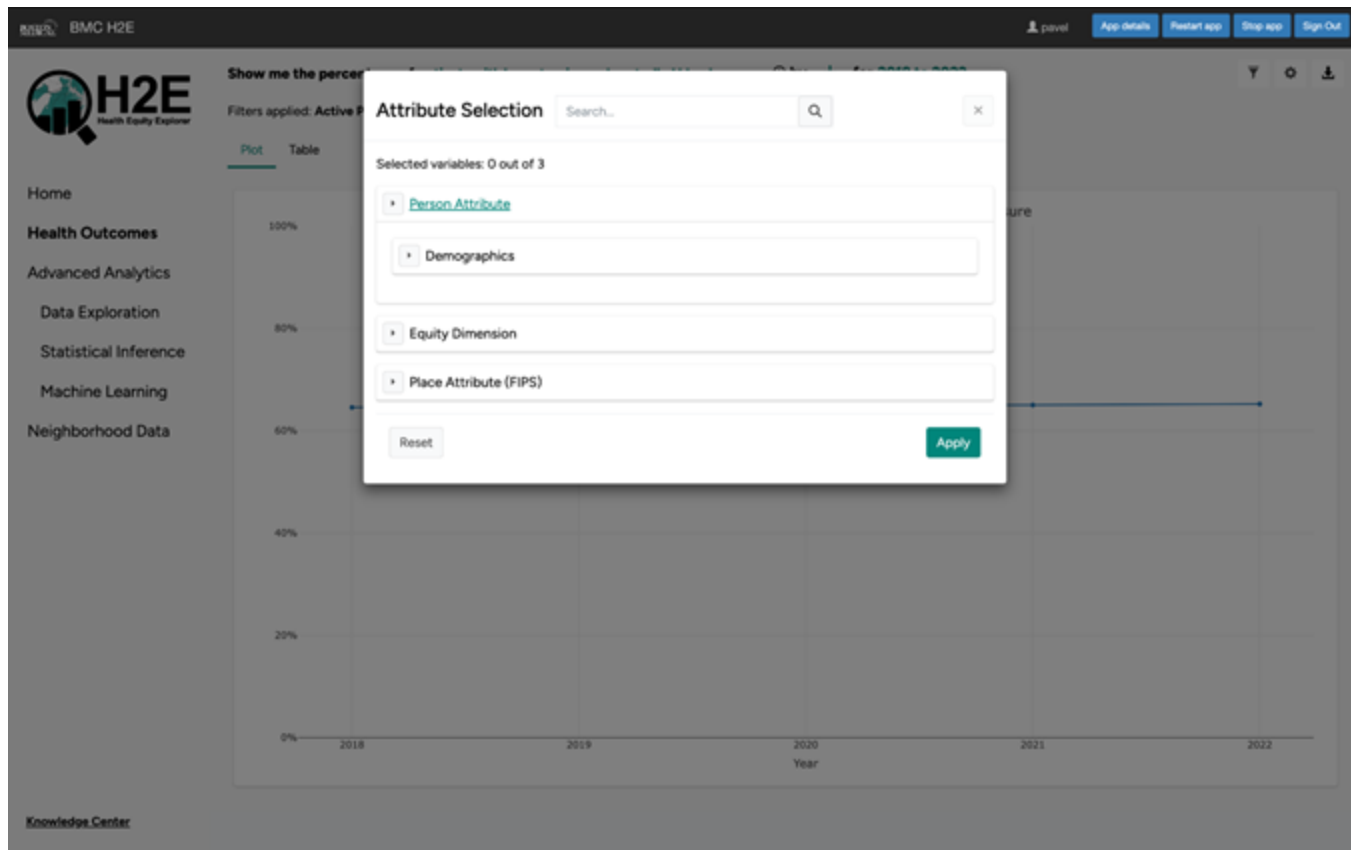
- Lots of inputs at once
- Complicated UI
- Statefulness
- Extra capabilities
- Conditional rendering



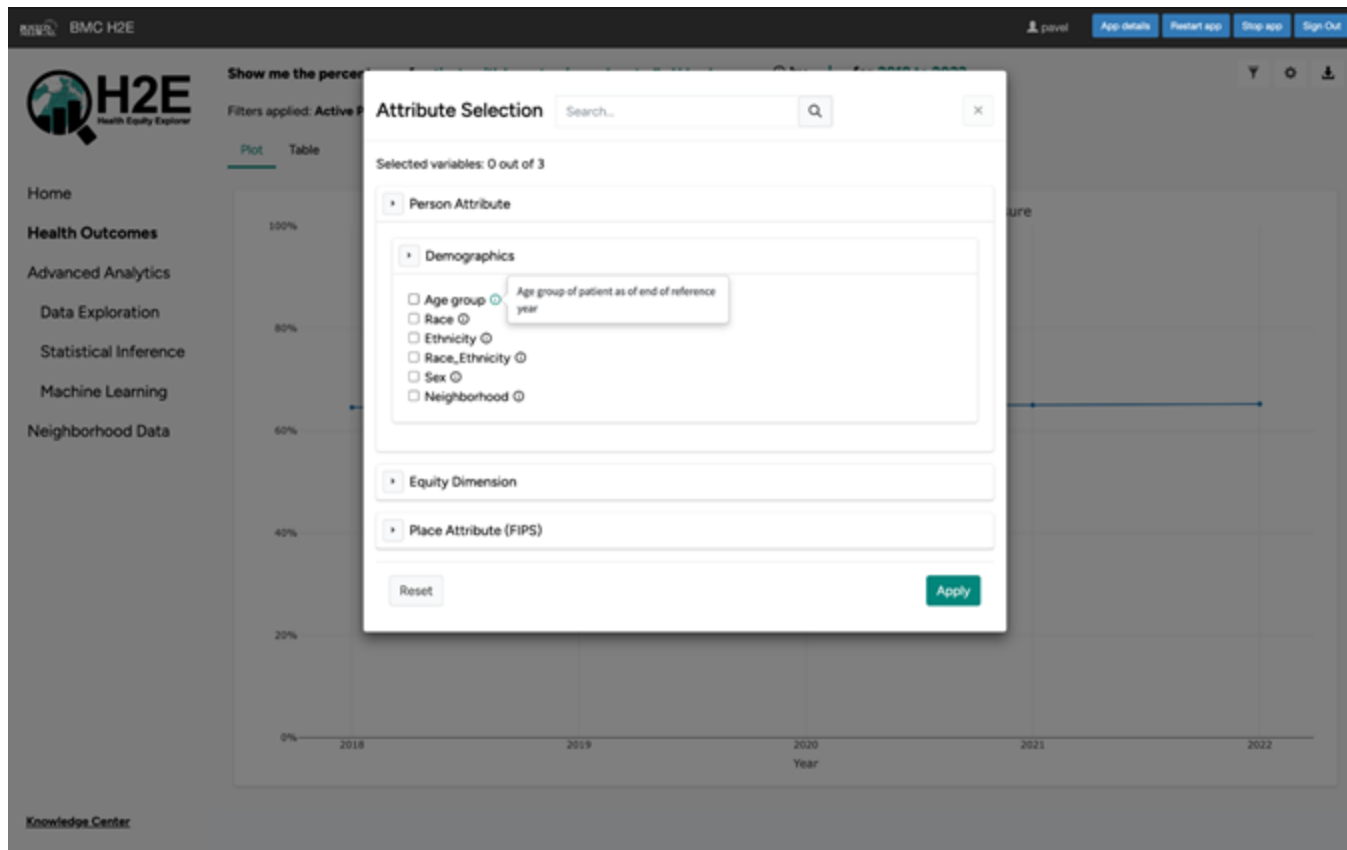
Technical Highlights Custom React Component



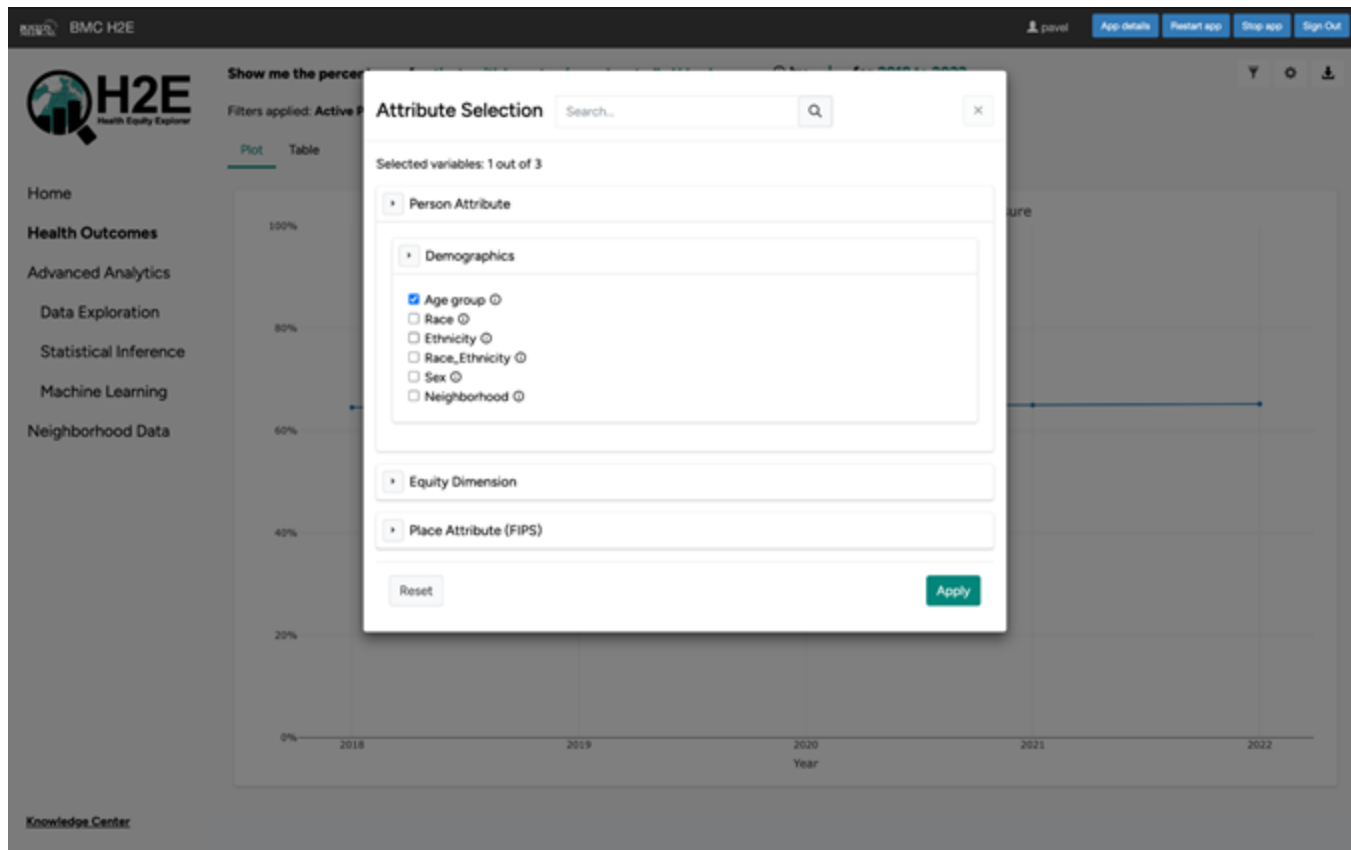
Technical Highlights Custom React Component



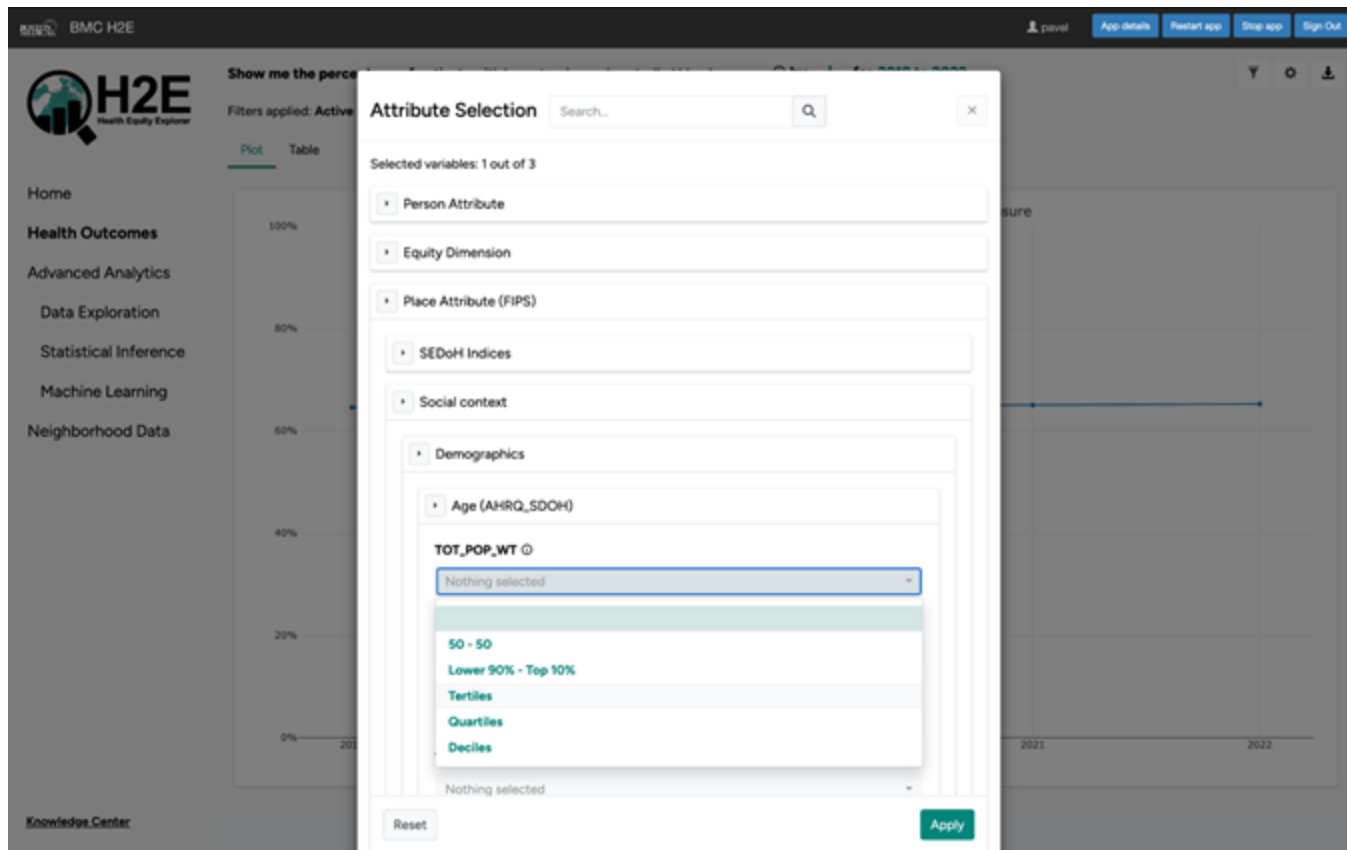
Technical Highlights Custom React Component



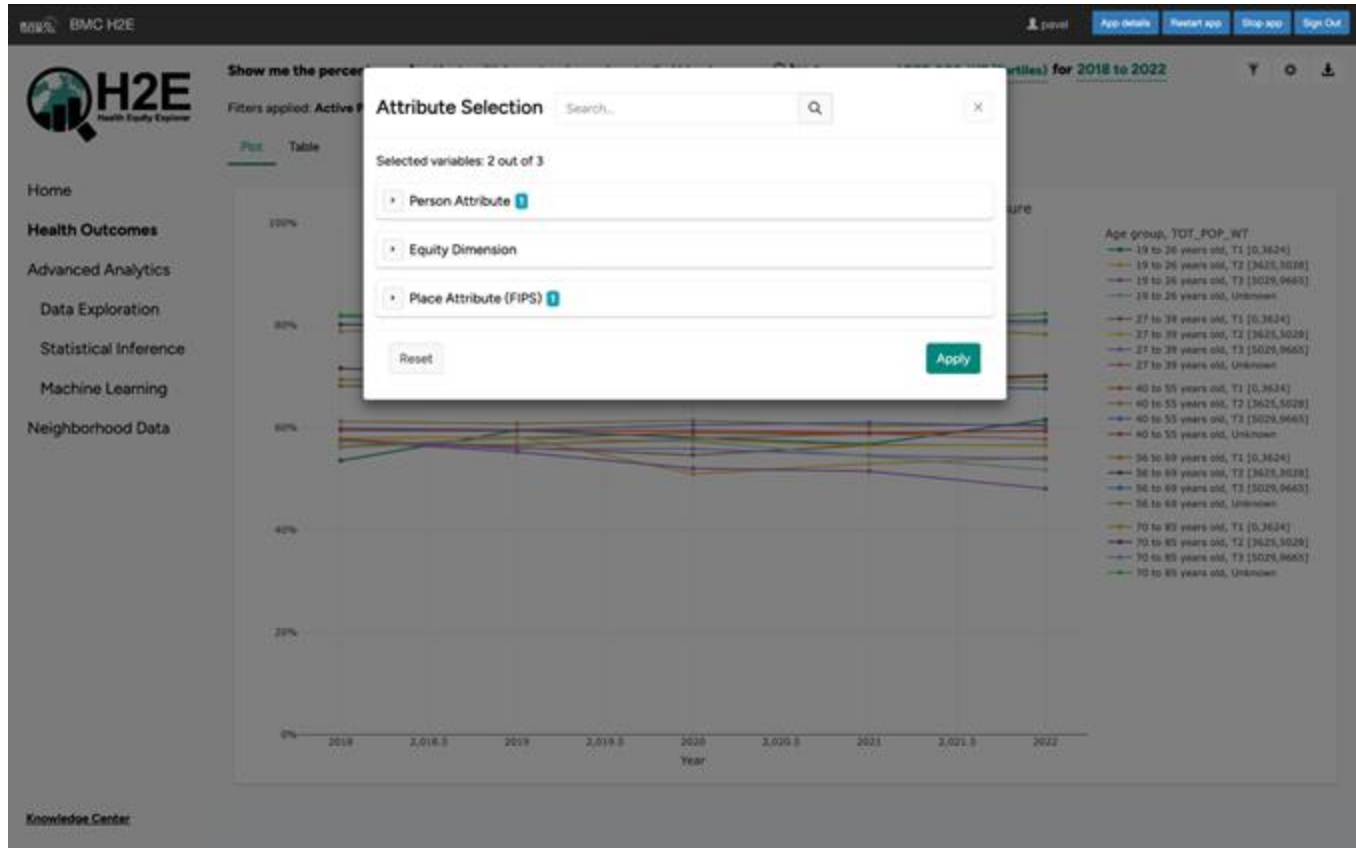
Technical Highlights Custom React Component



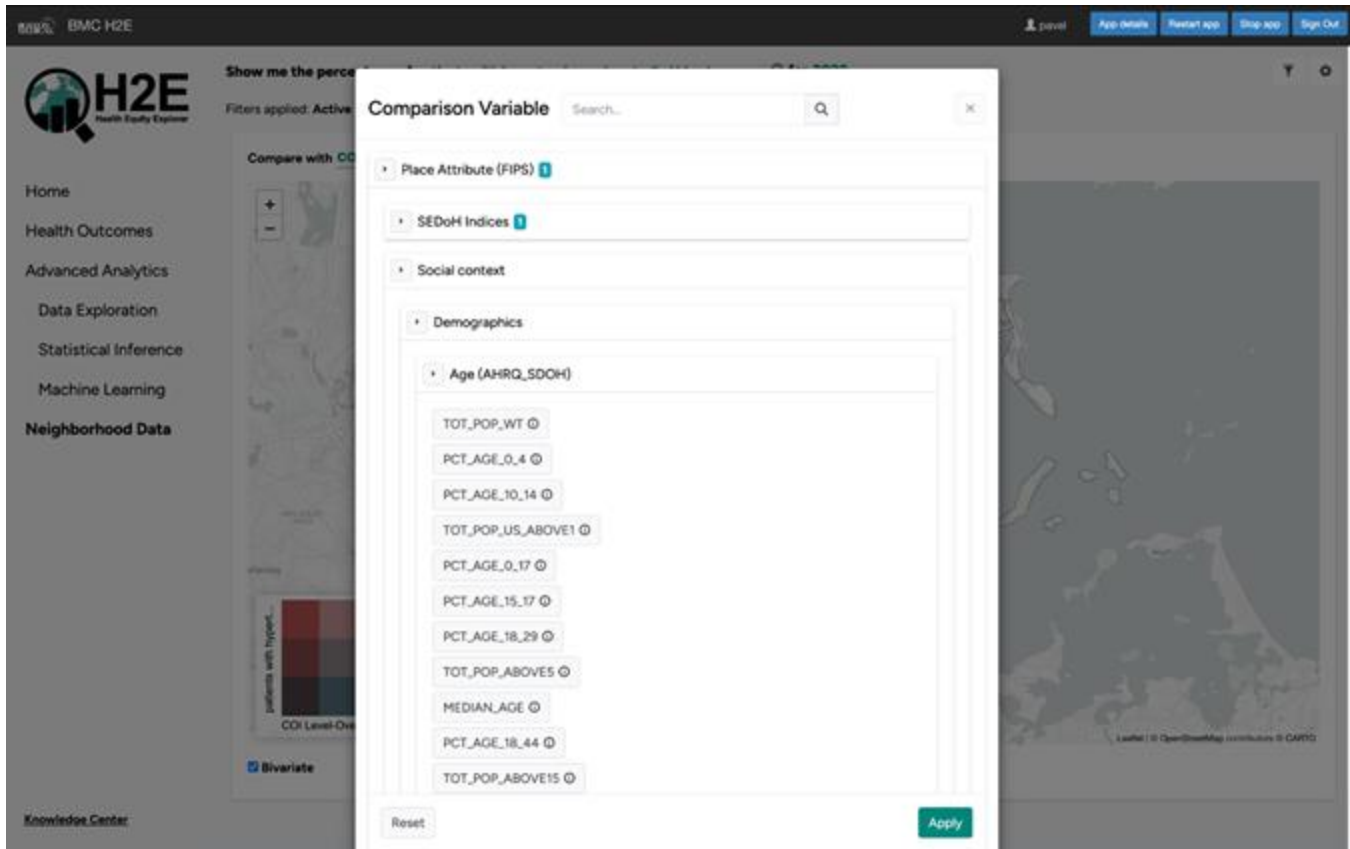
Technical Highlights Custom React Component



Technical Highlights Custom React Component



Technical Highlights Custom React Component



The screenshot displays the BMC H2E application interface. A modal titled "Comparison Variable" is open, allowing users to select variables for comparison. The modal includes a search bar and a list of variables under the "Age (AHRQ_SDOH)" category. The background shows a map of the United States and a sidebar with navigation options.

Comparison Variable

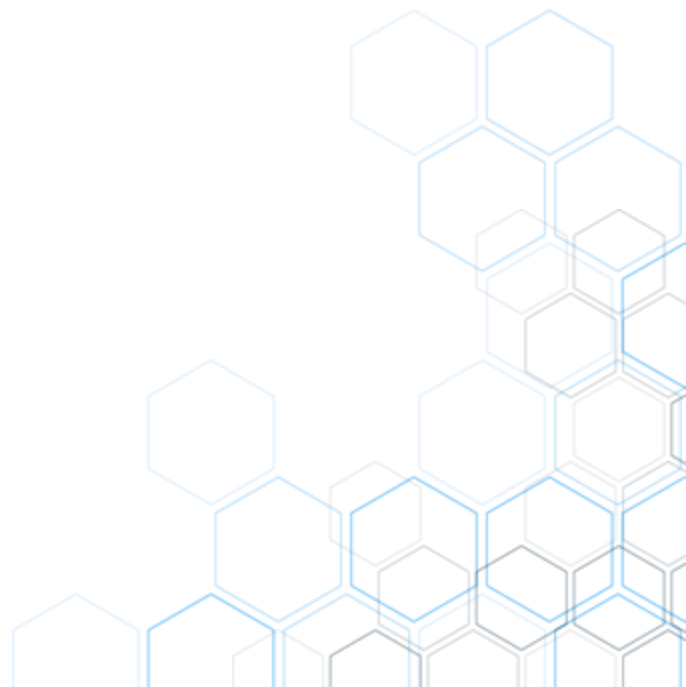
Search...

- Place Attribute (FIPS)
- SEDoH Indices
- Social context
- Demographics
 - Age (AHRQ_SDOH)
 - TOT_POP_WT
 - PCT_AGE_0_4
 - PCT_AGE_10_14
 - TOT_POP_US_ABOVE1
 - PCT_AGE_0_17
 - PCT_AGE_15_17
 - PCT_AGE_18_29
 - TOT_POP_ABOVE5
 - MEDIAN_AGE
 - PCT_AGE_18_44
 - TOT_POP_ABOVE15

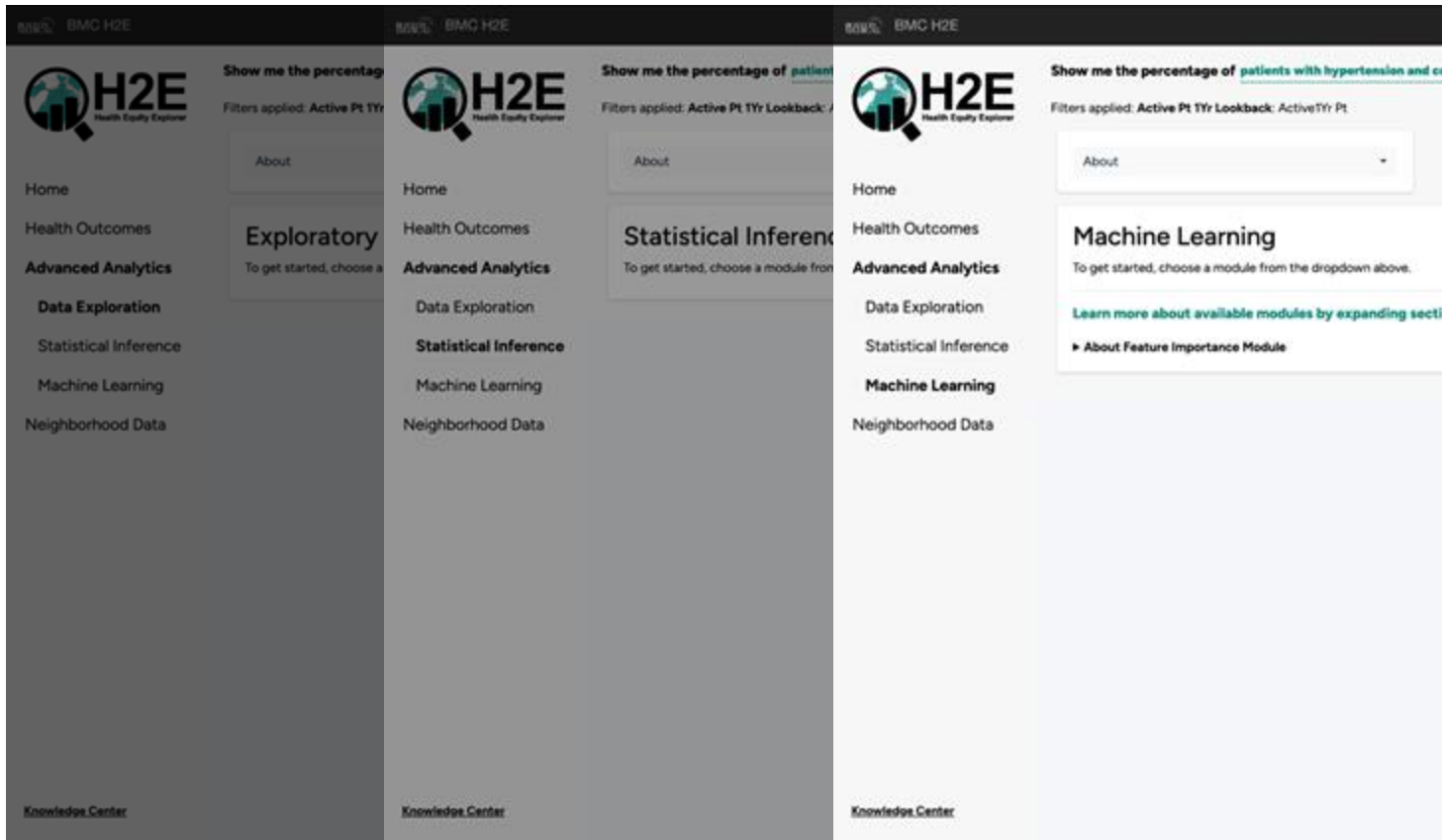
Reset Apply

Technical Highlights

Machine Learning Framework



Technical Highlights Machine Learning Framework



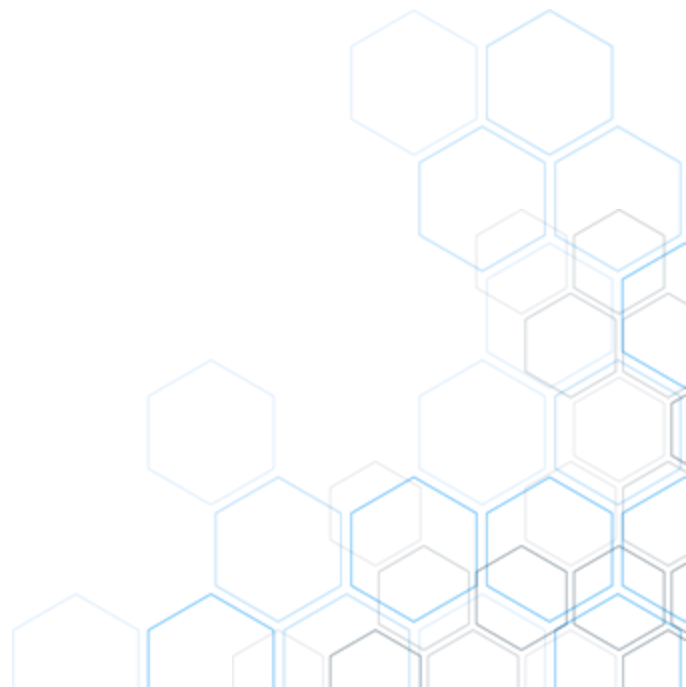
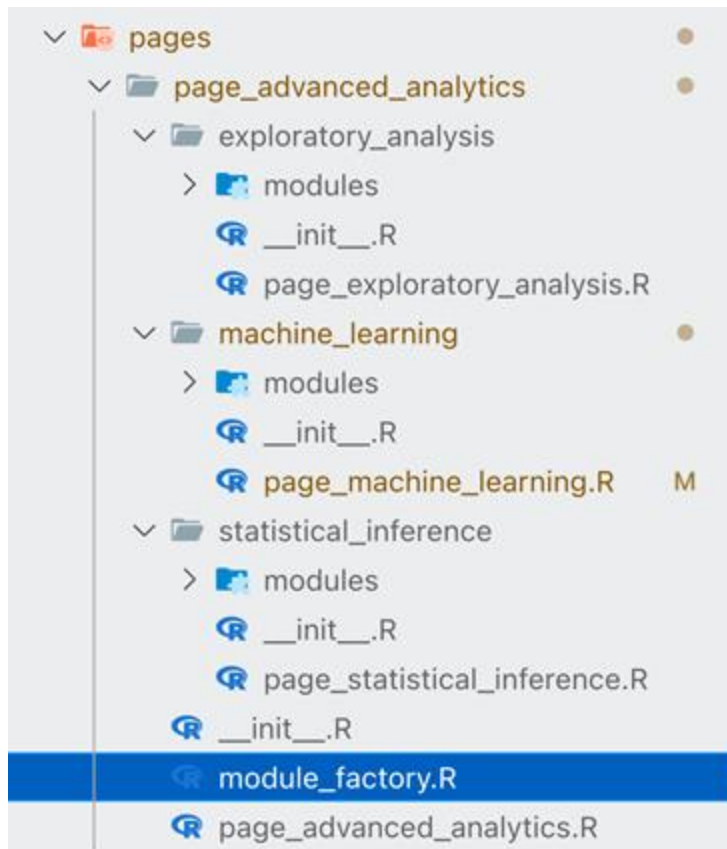
The image displays three sequential screenshots of the H2E Health Equity Explorer web application interface, illustrating its technical capabilities in a machine learning framework.

Screenshot 1 (Left): Exploratory View
 The interface shows the H2E logo and a sidebar with navigation options: Home, Health Outcomes, Advanced Analytics, Data Exploration, Statistical Inference, Machine Learning, and Neighborhood Data. The main content area displays the title "Show me the percentage of" followed by a dropdown menu. Below the dropdown, it indicates "Filters applied: Active Pt Tfr" and provides an "About" button. The "Exploratory" section is highlighted, with a prompt "To get started, choose a" and a list of modules: Data Exploration, Statistical Inference, Machine Learning, and Neighborhood Data.

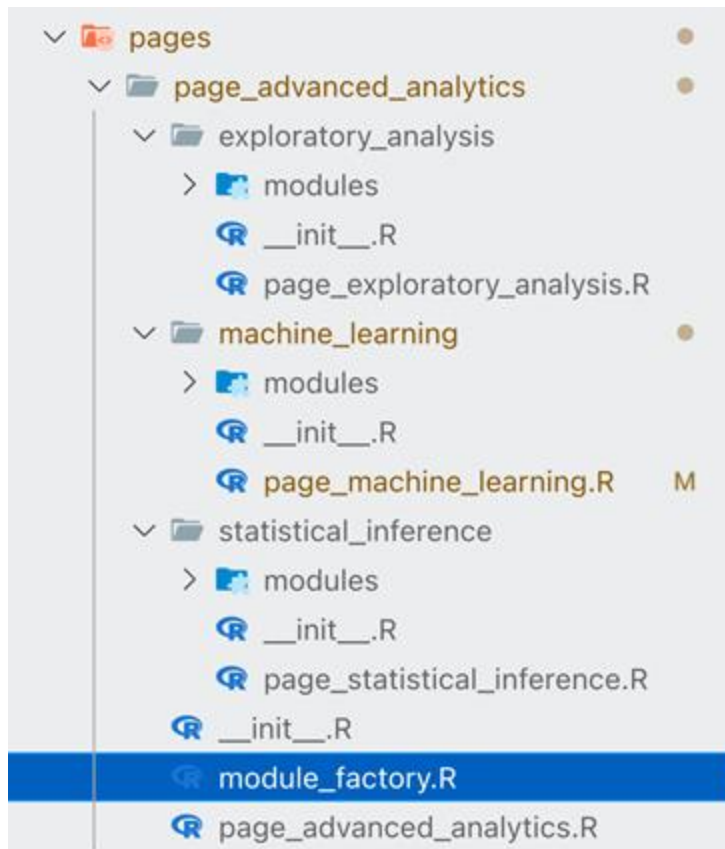
Screenshot 2 (Middle): Statistical Inference View
 This view shows the same interface but with the "Statistical Inference" module selected. The main content area displays the title "Show me the percentage of patients" followed by a dropdown menu. Below the dropdown, it indicates "Filters applied: Active Pt Tfr Lookback: /" and provides an "About" button. The "Statistical Inference" section is highlighted, with a prompt "To get started, choose a module from" and a list of modules: Data Exploration, Statistical Inference, Machine Learning, and Neighborhood Data.

Screenshot 3 (Right): Machine Learning View
 This view shows the same interface but with the "Machine Learning" module selected. The main content area displays the title "Show me the percentage of patients with hypertension and ex" followed by a dropdown menu. Below the dropdown, it indicates "Filters applied: Active Pt Tfr Lookback: ActiveTfr Pt" and provides an "About" button. The "Machine Learning" section is highlighted, with a prompt "To get started, choose a module from the dropdown above." and a link "Learn more about available modules by expanding secti". Below this, there is a link "About Feature Importance Module".

Technical Highlights Machine Learning Framework



Technical Highlights Machine Learning Framework



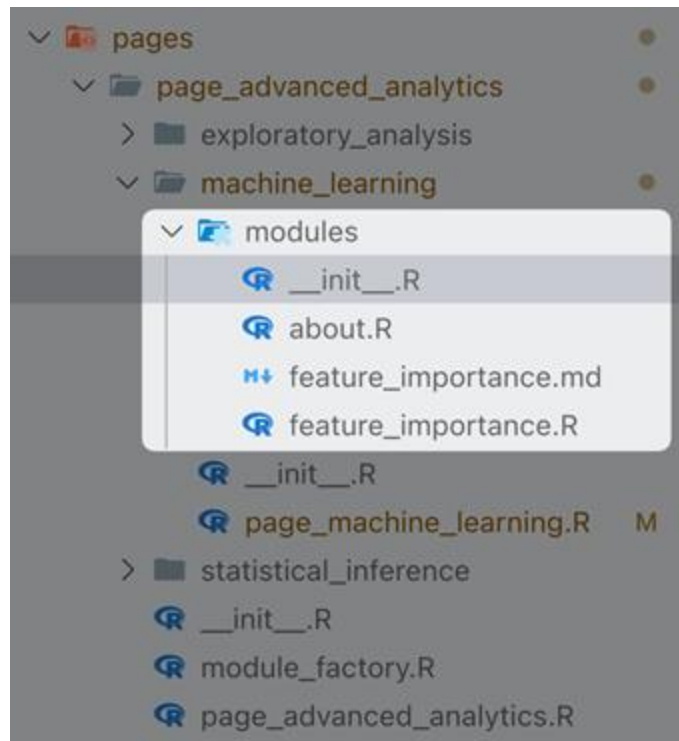
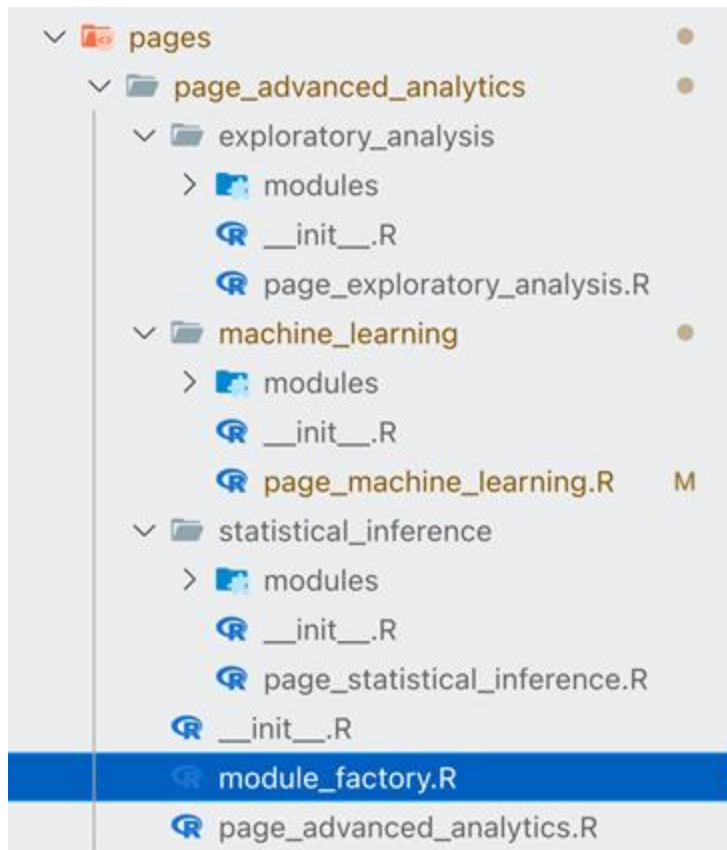
```
#' @param box_mod Box module
#' @export
factory <- function(box_mod) {
  exported_modules <- get_module_names(
    box_mod
  )

  ui <- function(id) {
    pickerInput(
      choices = exported_modules
    )
  }

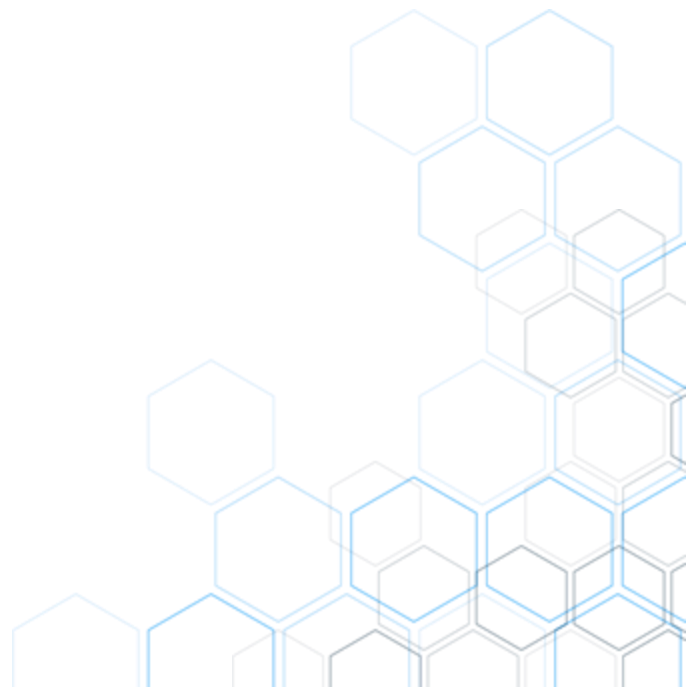
  server <- function(id, ...) {...}

  list(ui = ui, server = server)
}
```

Technical Highlights Machine Learning Framework



The Future



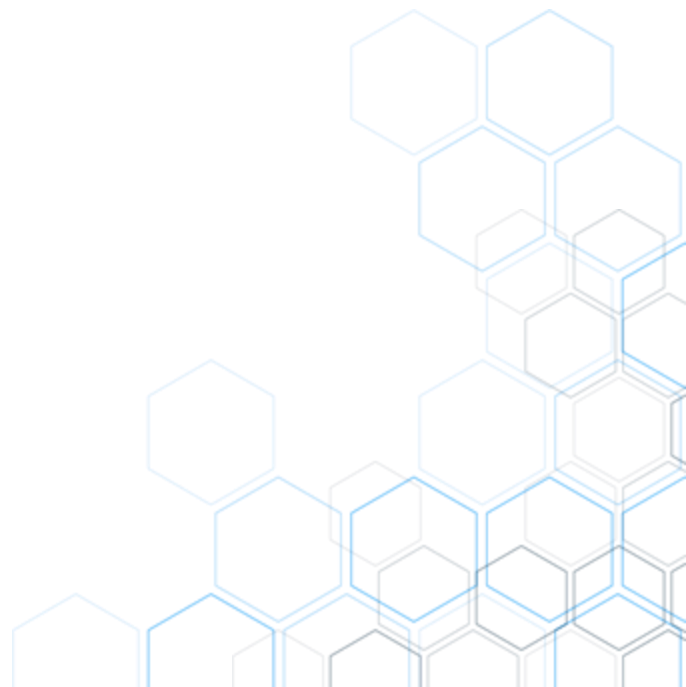
The Future

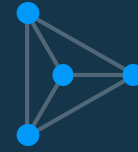
Rebranding

- **Population Health Explorer** - PHX
- Wider use and application

Open Source

- Source code available on GitHub
- **Summer 2025**





Appsilon

Thank you!