Prototype Data Dashboard for Multi-Source Transportation and Community Health Data Analytics

Chengyue Wang[1a], Talha Azfar[2b], Ruimin Ke[2c],

[1]Department of Civil Engineering, The University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968
[2]Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute, 110 8th St, Troy, NY 12180
e-mails: [a]cwang5@miners.utep.edu, [b]azfart@rpi.edu, [c]ker@rpi.edu

ABSTRACT

This study had a primary objective of creating a versatile and scalable data dashboard prototype that highlights transportation and community health data, using El Paso as a case study. The dashboard serves as a centralized hub for various data sources within the city, facilitating visual analysis to uncover the intricate connections between these two crucial domains. It's important to note that the primary aim was not to develop a final product for the city, but rather to establish a proof-of-concept, engage potential users, and conduct scenario analysis. The true impact of this study is expected to extend beyond the creation of the dashboard itself, as it has the potential to influence significant city-wide or regional policy changes based on the valuable insights derived from the comprehensive data analysis. This initiative aims to provide a robust tool for decision-makers and stakeholders to make informed choices that could ultimately benefit the community's transportation and health outcomes.

INTRODUCTION

City and county-wide data on transportation and community health are currently dispersed across various agency repositories, each with its own data format and resolution. This fragmentation hinders the analysis of cross-departmental or interdisciplinary indicator relationships. To address this, independent dashboards for transportation (WSDOT, 2023; Zuo et al., 2020), environment (Oberlin, 2023), and community health (EPTX, 2023) have been developed. This study aims to amalgamate, house, and synchronize spatial and temporal data from these three sectors into a unified dashboard.

A dashboard serves as a centralized data hub, streamlining the analysis of data spanning transportation, environment, and community health. It is more than just a repository; it combine open data sources with tools for data management, mapping, visualization, and modeling. Typically, a dashboard operates on three layers of data processing. The first layer focuses on data collection, preprocessing incoming data, and storing it in databases. The second layer is tasked with computing intermediate performance metrics or indicators, linking diverse data sets using

SQL (Structured Query Language), and executing spatial-temporal modeling. The third layer is dedicated to application-specific tasks, such as visualization and analysis.

The study's goal is to create a customizable and scalable prototype data dashboard, facilitating the visualization of data related to transportation and community health. Using El Paso, Texas as a case study, the research team demonstrated how the prototype dashboard could be utilized to intersect different data sets, enabling the analysis of the interplay among transportation and community health sectors.

LITERATURE REVIEW

A city dashboard is a sophisticated tool that aggregates, examines, and illustrates key performance indicators of a city's or region's infrastructure and services. It serves to enlighten both the public and policymakers about the current state and efficiency of vital resources such as transportation, healthcare, energy, and water, among others. Presented through an interactive and user-friendly graphical interface, these dashboards often utilize maps, classifying them under geospatial dashboards. They are instrumental in delivering real-time data, with some applications requiring updates as frequently as every 24 hours to maintain their relevance and accuracy.

According to Farmanbar and Rong (2020), city dashboards can be categorized into two distinct types: city-level and analytical dashboards. City-level dashboards provide a comprehensive view of a city's overarching performance, intended for a general audience. In contrast, analytical dashboards are designed for in-depth, specialized analysis, enabling users to delve into data across various layers, catering to analysts and decision-makers.

Globally, major cities have developed their dashboards to address diverse applications. However, not all platforms labeled as dashboards meet the criteria of showcasing live, operational data. The criteria for inclusion in an overview of such dashboards, as summarized in the referenced Table 1, are stringent: they must display live data, and should not be mistaken for mere landing pages of broader smart city initiatives. An often-cited exemplar of a city dashboard is Rio de Janeiro, Brazil's Operation Center, although, as of Mattern's 2015 publication, its online presence was not accessible.

Table 1. Operational Dashboards in Selected Cities

City	Dashboard name	Reference	Application
Baltimore, U.S.	CitiStat	City of Baltimore (2022), Gullino (2009)	Community development, environment, public safety
Boston, U.S.	Boston COVID-19	City of Boston (2023a)	COVID-19 cases and vaccinations
Boston, U.S.	Imagine 2030 Boston Metrics Dashboard	City of Boston (2023b)	A variety of city's KPIs
Charlotte, U.S.	Capital Project Dashboard	City of Charlotte (2022)	Building projects
Dublin, Ireland	Building City Dashboard	Maynood University (2023)	City infrastructures' KPIs

Las Vegas, U.S.	Community Dashboard	City of Las Vegas (2023)	A variety of city's KPIs
London, U.K.	CityDashboard	CASA (2023)	A variety of city's KPIs
Melbourne, Australia	Melbourne Dashboard	Crisis Dashboard (2021)	Traffic, weather
New York, NY	NYC Open Data	City of New York (2022)	Various city services
New York City and Seattle, U.S.	none	Zuo et al. (2020)	Transportation and sociability
Singapore	LIVE Singapore	Ratti et al. (n.d.)	Communication, environment, transportation
Washington D.C., U.S.	DCStat	ADT (2006)	Property maintenance, public safety

Zuo et al. (2020) crafted dashboards for New York City and Seattle to explore the impact of COVID-19 social distancing policies on various traffic metrics including volume, travel time, average speed, crash occurrences, and bikeshare usage, using data provided by local agencies. They also incorporated additional datasets pertaining to pedestrian density and social distancing adherence. The dashboards were utilized to analyze changes in social distancing compliance, traffic speed variations, crash severity, and bicycle usage during the pandemic. The dashboards developed by Zuo et al. feature a tripartite structure comprising:

1. **Data Integration and Access Layer**: This foundational layer consolidates and facilitates access to the collected data.
2. **Data Mining and Cloud Computing Layer**: At this intermediate level, the data undergoes processing and analysis, leveraging cloud computing resources.
3. **All-in-One Data Dashboard and Analytics Layer**: The topmost layer presents a comprehensive view of the analytics, offering users a unified interface for data interpretation.

Farmanbar and Rong (2020) introduced the Triangulum City Dashboard, designed to assist residents and policymakers in discerning the interconnections among mobility, energy usage, and environmental stewardship to inform policy decisions aimed at enhancing quality of life. This cloud-based platform also operates on a three-tier architecture:

- **Data Layer**: Engages in the acquisition, storage, and preliminary handling of data.
- **Application Layer**: Processes the data from the lower layer to extract actionable insights.
- **Presentation Layer**: Displays the processed data in a user-friendly format, facilitating an intuitive understanding for end-users.

The data layer integrates five distinct datasets: traffic, electricity, parking management, energy, and cargo e-bike data. Farmanbar and Rong also navigate through challenges such as varying data resolutions and formats, incomplete datasets due to sensor limitations, and the need to uphold privacy laws.

DATA COLLECTION

The foundational layer of the dashboard comprises demographic data, predominantly sourced from the U.S. Census website (Census, 2023). The most up-to-date data available was from the 2020 American Community Survey. As the official results of this survey were not yet public, estimates spanning six years, from 2016 to 2021, were utilized. The U.S. Census provides data at the census tract level, and for the City of El Paso, Texas, this equates to 128 distinct tracts. The acquired data included estimates of population size and median household income for each tract over the years 2016 through 2020. This data was subsequently consolidated by zip code. In the City of El Paso, there are 23 zip codes, which are identified by numbers ranging from 79901 to 79936, albeit not sequentially.

The crash data set encompasses 35,422 recorded traffic collisions that took place within the City of El Paso, Texas, from January 1, 2016, to October 18, 2021. These records originated from CR-3 crash report forms, completed by police officers at the scene of each incident, and were subsequently uploaded to the Texas Department of Transportation (TxDOT) Crash Report Information System (CRIS). This data compilation is a crucial element of the City of El Paso's Intersection Safety Improvement Plan (Cheu et al., 2021). It specifically includes crashes that occurred at city-managed street intersections and excludes incidents on highways under TxDOT jurisdiction, non-intersection-related accidents, and those on private properties. The dataset was furnished in a Microsoft Excel CSV format, with each row detailing the attributes of an individual crash.

The research team acquired a comprehensive set of vehicle trajectory data from Wejo (2023), capturing vehicular movements across El Paso County, Texas. This dataset spans from October 25, 2021, to November 21, 2021, and records round-the-clock vehicle activity, amassing a total of 968,715,648 individual records. Each record documents various attributes of a vehicle, with data points collected at three-second intervals. Wejo provided this extensive dataset via the Amazon Web Service S3 database in an efficiently compressed JSON format. To verify the accuracy of the data's geolocation, the points were plotted on a street map.

In addition to traffic data, the dashboard integrates community health metrics, particularly those pertaining to COVID-19 within the City of El Paso, Texas. The local health data, available to the public on the El Paso Strong website (EPTX, 2023), align with the standard statistics requested by the CDC. This includes the community risk level, new and cumulative counts of positive test results, new and cumulative deaths attributed to COVID-19, along with the number of hospitalizations, ICU admissions, and ventilator utilizations.

PROTOTYPE DATA DASHBOARD FRAMEWORK

This paper presents a web-based dashboard prototype designed to perform cross-layer analyses of extensive datasets, facilitating the tracking and identification of trends within transportation, environment, and community health in the City of El Paso, Texas. The proposed dashboard's architecture, depicted in Figure 1, is structured into three principal layers: data collection, modeling, and application, aligned in a bottom-up configuration.

The foundational data collection layer is tasked with the aggregation, preprocessing, and storage of data from diverse infrastructural systems. This includes data quality assurance measures such as outlier removal, deduplication, and gap filling to ensure the data's consistency, accuracy, completeness, and timeliness. This refined data is then stored across various layers, utilizing cloud-based solutions or local drives for the prototype.

The intermediary modeling layer serves as the nexus between the raw data and the user interface. It processes data from the collection layer, applying algorithms and models to derive key metrics like incident frequencies, safety indices, congestion levels, vehicle trajectories, and health indices. Integration within this layer is pivotal, merging multi-source data into formats suitable for modeling. This layer can employ pre-programmed or user-interactive models, leveraging predefined models from Python libraries like Scikit-Learn or custom models tailored to specific requirements, ensuring efficient resource utilization and minimal response latency.

At the forefront is the application layer, which acts as the interactive visual interface of the dashboard. It offers users the capability to navigate and manipulate displayed information through a one-page or drill-down design, constructed on a web-based application framework. Depending on the programming language, this framework may vary, such as Vaadin for Java or Streamlit for Python. It incorporates common interface elements like buttons, dropdowns, tabs, checkboxes, date selectors, sliders, and text boxes, with a unique layout for each function driven by the modeling layer algorithms. Interactivity between modules, along with the integration of charts and maps for data representation, is crucial. Tools like Google Map, Open Street Map, and the Folium library in Python are instrumental in visualizing geospatial data. The design of this layer is focused on the optimal visualization of model outputs, including map overlays, heatmaps, and a variety of charts, with the flexibility for data to be displayed on the map, in pop-up windows, or as downloadable files.

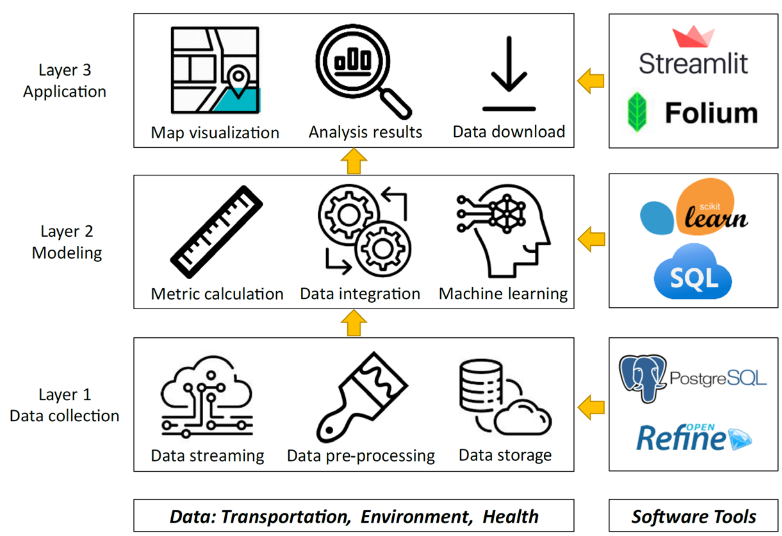

Figure 1 Concept Diagram of the Dashboard

PROTOTYPE DATA DASHBOARD DEVELOPMENT

The prototype data dashboard is now live and can be visited at https://chengyuew-el-paso-dashboard-case-study-dashboard-5ukgqq.streamlit.app/. This section will guide you through the dashboard's essential functionalities. Upon arrival, the landing page greets visitors with a welcome message, illustrated in Figure 2, and offers three informative tabs: About, People, and Disclaimer. The About tab provides insight into the dashboard's background and aims. The People tab credits the research team behind the study, while the Disclaimer tab mentions the funding source and clarifies any data usage and interpretation constraints.

To dive into the dashboard, users are prompted to double-click the [Double Click Enter El Paso Dashboard] button at the landing page's foot. This action transitions the user to the main interface of the dashboard, which maintains a consistent layout across all data layers. The interface is strategically partitioned into three sections:

- The left panel, known as the Main Menu, outlines the four data layers available: Demographic, Crash Data, Traffic Data, and Health Data. Users have the option to focus on one data layer at a time. The lower segment of this panel presents dropdown menus for selecting or inputting attributes relevant to the chosen data layer.
- The right-top panel is dedicated to the dashboard's title block, "El Paso Data Dashboard," accompanied by the subtitle "Transportation, Environment and Community Health," establishing the dashboard's thematic focus.
- The right-bottom panel, the primary display area, showcases the maps or visualizations corresponding to the selections made in the left panel, enabling users to visualize the data objects or attributes in detail.

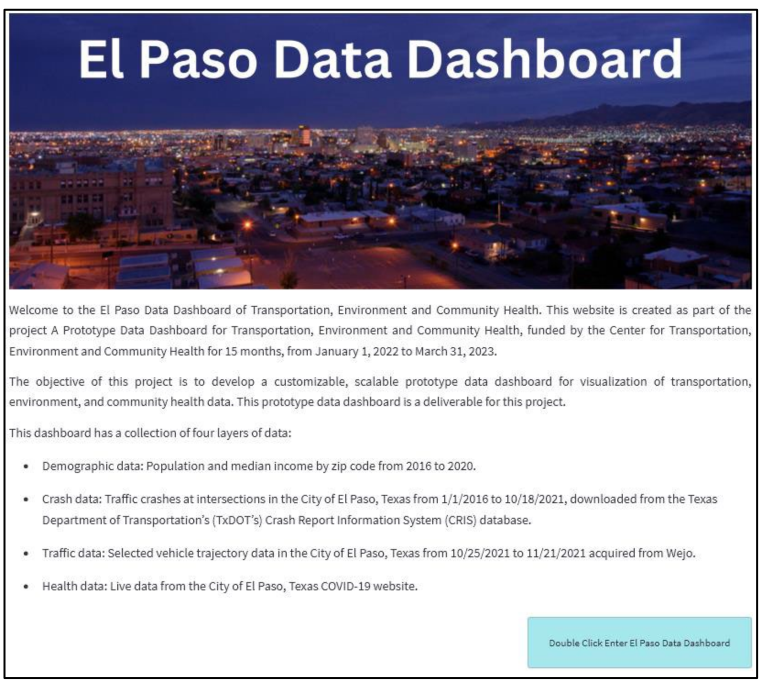

Figure 2 The landing page of the dashboard

Figures 3 contains screenshots as examples of the Demographic layer. The base map divides the City of El Paso by zip code. Users can select to display either Population or Median Household Income by a calendar year from 2016 to 2020. The data of the selected Attribute and Year are plotted by zip code on a map. Figure 4 (left) displays the Population in 2016. Figure 3 (right) displays the Median Household Income in 2016.

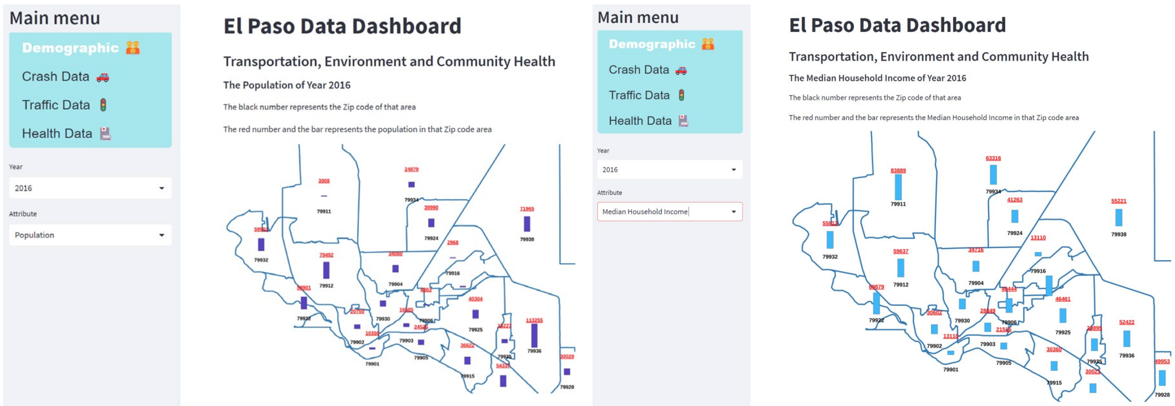

Figure 3 Demographic – population and median household income by zip code

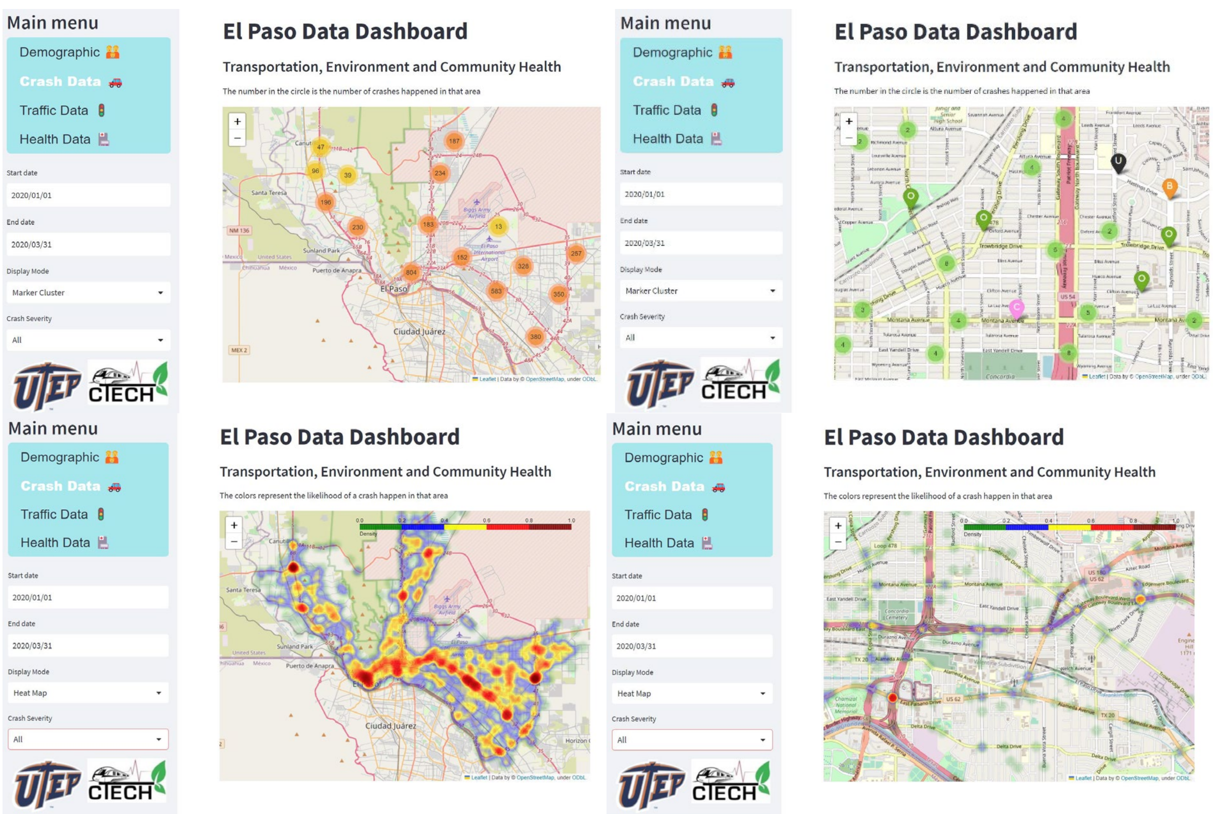

Figure 4 Crash data display on the dashboard with interactive functionality

Figure 4 demonstrates the diverse methodologies for visualizing crash data within the dashboard. Users are presented with three distinct display options for crash data visualization: Cluster Marker, Heat Map, and Marker. These can be selected via the dropdown menu labeled 'Display Mode' found in the left panel. For each mode, there are additional customization options available, such as specifying the Start Date, End Date, and Crash Severity.

When it comes to Traffic Data, there are three modes of display: Table, Density Heat Map, and Trajectory Visualization. The dashboard allows users to analyze traffic data by mode and by zip code, offering a granular view of traffic patterns.

The default screen for Traffic Data, shown in Figure 5, starts in 'Table' mode. This table format displays individual data points for each vehicle, including details like journey ID, latitude, longitude, speed, and heading. The data for each vehicle is sampled at three-second intervals. As an example, Figure 5 illustrates vehicle trajectory data from zip code 79912, captured on November 10, 2021, from 10:00:00 a.m. to 10:02:59 a.m. The table header indicates the total number of data points, while key statistics such as average, maximum, minimum speeds, and speed standard deviation are calculated for all data points within the selected zip code and displayed beneath the table.

Additionally, Figure 5 showcases a density heat map representing all zip codes. This visualization can be accessed by selecting 'Density Heat Map' in the Display Mode and opting for 'All' under Zip Code in the left panel. The density metric is derived from the number of vehicle trajectory data points in a given zip code, adjusted for the area of the zip code. It's important to note that this density represents the vehicles tracked by Wejo and does not directly equate to the actual number of vehicles per mile or kilometer. The density values are normalized to a range between 0 and 1, in increments of 0.2, with varying colors depicting different density levels.

Figure 5 (bottom) displays the sampled vehicles' trajectories in a selected area. This is obtained when the Display Mode is changed to Trajectory Visualization. The trip origins and destinations are represented by the "O" markers and the "D" markers, respectively. The origin and destination of a trip, and the travel path are indicated by one "O" maker, one "D" marker and a color-coded line. Only trips that have at least an "O" or a "D" marker in the selected display area are shown on the map. The data points for the same vehicles are collected at 3 second intervals.

The Health Data section of the dashboard presents the City of El Paso's COVID-19 data in a variety of visual formats: Case Table, Bar Chart, Cumulative Positive Curve, Cumulative Recovery Curve, and Cumulative Death Curve. Figure 6 showcases the default view of the Health Data, which is a table format detailing COVID-19 statistics across three columns: cumulative positive cases, cumulative recovered cases, and cumulative deaths. Each row corresponds to a different zip code in El Paso, and the numbers represent the total new cases accumulated between selected start and end dates.

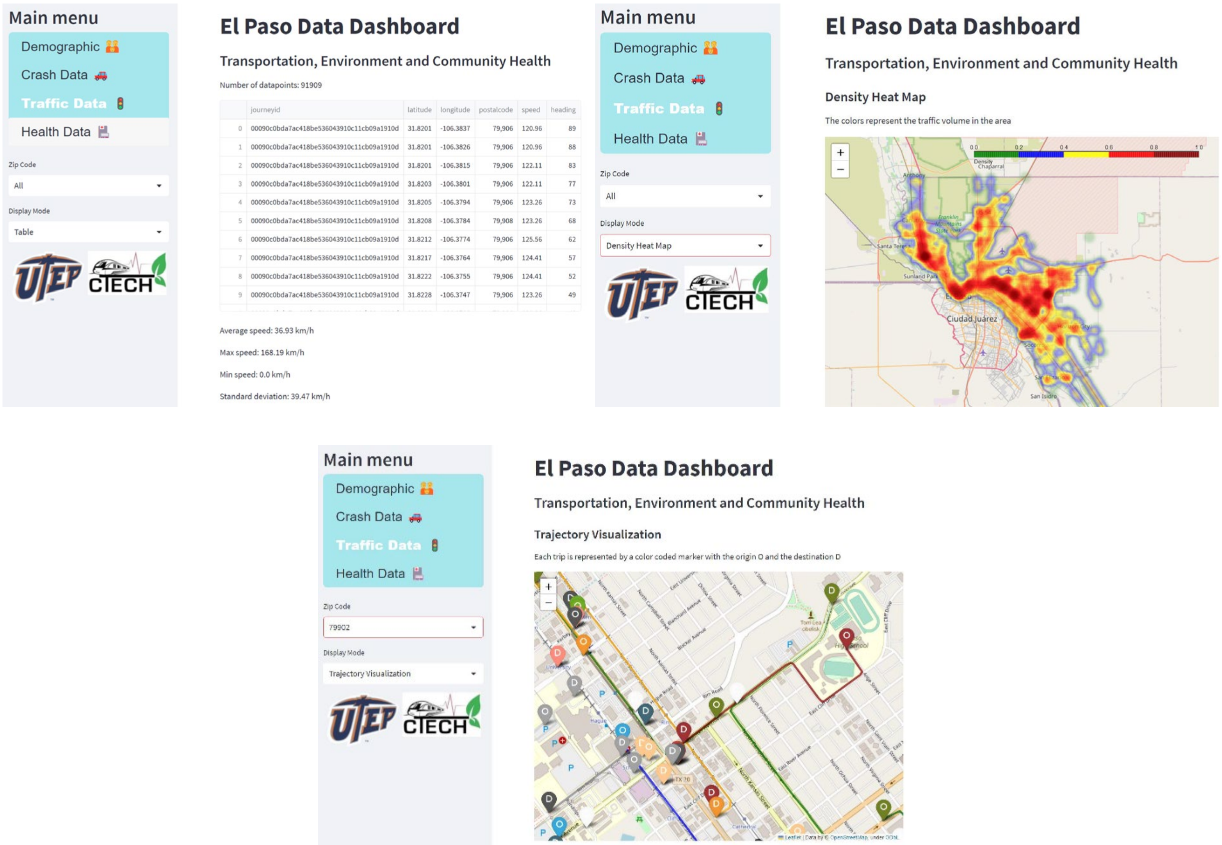

Figure 5 Traffic data displayed on the dashboard

In the example illustrated in the top left of Figure 6, the dashboard is set to display the data in Table mode, encompassing all zip codes, spanning from September 1, 2021, to March 31, 2022. This view highlights nine zip codes, with the option to scroll down for more. The top right section of Figure 6 displays a bar chart visualizing cumulative positive cases by zip code, offering a clear perspective on which areas experienced the highest number of cases during the specified timeframe.

The Health Data feature enables users to delve into detailed statistics about cumulative positive cases, recoveries, and deaths within specific timeframes. Users can tailor the display by selecting the desired zip code, choosing the type of curve under Display Mode, and setting the start and end dates. For instance, the bottom left of Figure 6 presents a Cumulative Positive Cases curve for zip code 79902, charting the period from March 1, 2020, to March 31, 2021. Similarly, the bottom right part of the figure illustrates the Cumulative Recovery Curve for the same timeframe, but for zip code 79911. These visualizations provide an insightful overview of the pandemic's impact across different areas of El Paso.

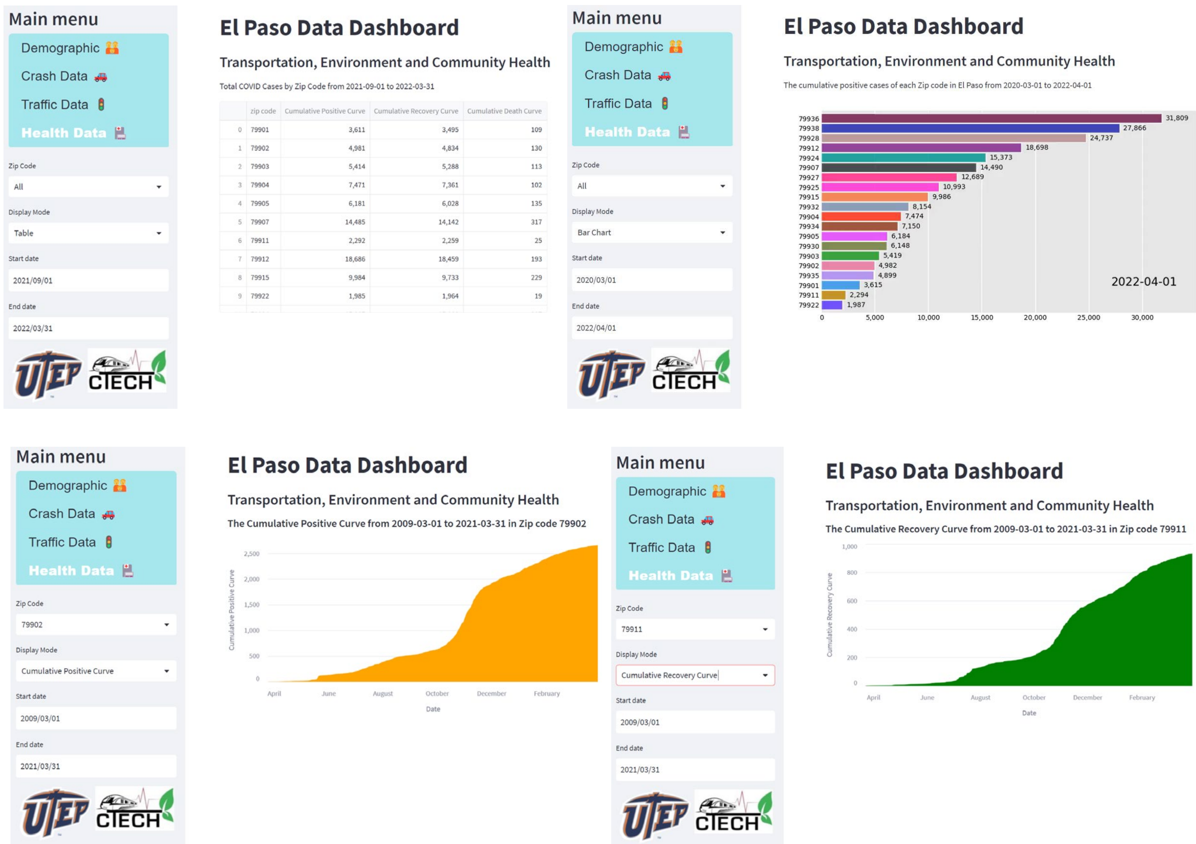

Figure 6 Health Covid-19 data displayed on the dashboard

CONCLUSION

In this study, a prototype dashboard has been developed that integrates demographic, crash, traffic, and health data within the City of El Paso. This tool was employed to engage stakeholders from El Paso and solicit their feedback, demonstrating its customizability and scalability through two case studies. The dashboard showcases the possibility of hosting, visualizing, and analyzing diverse datasets, illustrating the integration and cross-layer analysis of data as its principal output. This capacity makes it an indispensable tool for urban decision-makers, allowing them to discern patterns and trends through a convenient and intuitive interface. Beyond the dashboard itself, this research has yielded valuable processed data. Despite originating from disparate sources, the dashboard unifies these data streams, offering filtering and visualization options to clarify emerging patterns. The integration and analytic capabilities offered by the dashboard could revolutionize urban decision-making, providing a holistic tool that benefits decision-makers and, by extension, all city residents. This research holds profound implications for the use of data in informed urban planning and policy. Future enhancements could include:

- Direct connections between the dashboard and data sources for real-time updates.
- An expanded dataset timeframe and scope, such as including post-October 13, 2021, crash data and incidents on TxDOT-managed properties.

- Enhanced data layers with additional attributes, like demographic distributions by gender and age, and traffic volume for a richer analysis.
- Integration of additional layers covering transportation and infrastructure data, such as bus stops, parking, and traffic signal timings.
- Implementation of new applications within the dashboard, like service level and equity analyses.
- Application of machine learning techniques to analyze and integrate data layers, revealing novel insights and patterns.

REFERENCES

ADT (2006). In depth DCStat. *Application Development Trend Magazine*. April 5, 2—6. https://adtmag.com/Articles/2006/04/05/DCStat.aspx?Page=2&p=1. Accessed January 26, 2023.

CASA (2022). CityDashboard: London. Centre for Advanced Spatial Analysis, University College London. https://citydashboard.org/london.

Census (2023). Data and tables. U.S. Census Bureau. https://www.census.gov/data.html. Accessed January 26, 2023.

City of Baltimore (2022). About CitiStat. Office of Performance and Innovation. City of Baltimore, Maryland. https://www.baltopi.com/about-citistat. Accessed January 26, 2023.

City of Boston (2023a). COVID-19 in Boston. City of Boston. https://www.boston.gov/government/cabinets/boston-public-health-commission/covid-19-boston. Accessed January 26, 2023.

City of Boston (2023b). Imagine Boston 2030 metrics dashboard. City of Boston. https://data.boston.gov/showcase/imagine-boston-2030-metrics-dashboard. Accessed January 26, 2023.

City of Charlotte (2022). Capital projects. City of Charlotte, North Carolina. https://charlottenc.gov/charlottefuture/Pages/Capital_Project.aspx. Accessed January 26, 2023.

City of Las Vegas (2023). City of Las Vegas community dashboard. https://communitydashboard.vegas. Accessed January 26, 2023.

City of New York (2022) NYC open data dashboard. City of New York. https://opendata.cityofnewyork.us/dashboard. Accessed January 26, 2023.

Crisis Dashboards (2021) https://www.citydashboard.com.au/melbourne.php. Accessed January 26, 2022.

Delpeuch, A. (2022). OpenRefine User Manual. https://openrefine.org/docs. Accessed February 23, 2023. 70

EPTX (2023). El Paso Strong. City of El Paso, Texas. https://epstrong.org. Accessed February 15, 2023.

Farmanbar, M. and Rong, C. (2020). Triangulum city dashboard: an interactive data analytic platform for visualizing smart city performance. *Processes*, 8(2), 250. doi:10.3390/pr 8020250.

Jing, C., Du, M. . (2019). Geospatial dashboards for monitoring smart city performance. *Sustainability,* 11, 5648. doi:10.3390/su1120564.

Fischer, F. (2021). UTEP, City of El Paso opening new COVID-19 mega-testing site. KFOX14 News, August 19, 2021. https://kfoxtv.com/news/local/utep-city-of-el-paso-opening-new-covid-19-mega-testing-site.

Microsoft (2023). Azure Database for PostgreSQL. Microsoft Corp., https://azure.microsoft.com/en-us/products/postgresql/#overview. Accessed February 23, 2023.

Mattern, S. (2015). Mission control: a history of the urban dashboard. *Places Journal*, March 2015. Accessed 24 Jan 2023. https://doi.org/10.22269/150309.

Maynood University (2023). Building city dashboard. https://dashboards.maynoothuniversity.ie. Accessed January 26, 2023.

Oberlin (2023). Oberlin environment dashboard - bringing environmental dashboard to your community. Oberlin College, Ohio. https://environmentaldashboard.org/bring-dashboard-to-your-community#_Toc446430146. Accessed February 15, 2023.

Ratti, C., Kloeckl, K., Anwar, A., Biderman, A., Britter, R., Chen, X., Cruz, P., Dunnam, J., Lee, D., Nagel, T., Santani, D., Senn, O., Siegel, A., Sommer, C., Vanky, A., York, D. (n.d.). LIVE Singapore! Senseable City Lab, Massachusetts Institute of Technology. https://senseable.mit.edu/livesingapore/visualizations.html. Accessed 26 January 2023.

Sanromã, L. (2021). The principles of visual design for dashboards. Adnia Solutions. Retrieved November 18, 2022, from https://adniasolutions.com/dashboard-design-principles/the-principles-of-visual-design-for-dashboards.

Scikit-Learn (2023) Scikit-learn, machine learning in Python. Scikit-Learn Developers. https://scikit-learn.org/stable/ Accessed February 23, 2023. 71

Streamlit (2023). Streamlit documentation. Streamlit Inc. https://docs.streamlit.io. Accessed February 23, 2023.

Story, R. (2013). Folium 0.14.0 documentation. https://python-visualization.github.io/folium. Accessed February 23, 2023.

Wejo (2023). Trafic Movements. Wejo Limites. https://www.wejo.com/products/vehicle-movements. Accessed February 15, 2023.

WSDOT (2023). COVID-19 Multimodal transportation system performance dashboard. Washington State Department of Transportation. https://www.wsdot.wa.gov/about/covid-19-transportation-report. Accessed 2/15/2023.

Zuo, F., Wang, J., Gao, J., Ozbay, K., Ban, X. J., Shen, Y., Yang, H. and Iyer, S. (2020). An interactive data visualization and analytics tool to evaluate mobility and sociability trends during COVID-19. UrbComp 2020: *The 9th SIGKDD International Workshop on Urban Computing*, San Diego, California, USA.