INTERACTIVE DASHBOARD FOR CHRONIC DISEASE DATA

CSE6242 | Data and Visual Analytics | Polo Chau
Team003: Jinglin Xu, Haijiao Tao, Mingwei Zhu, Maggie Xia, Yang Lu

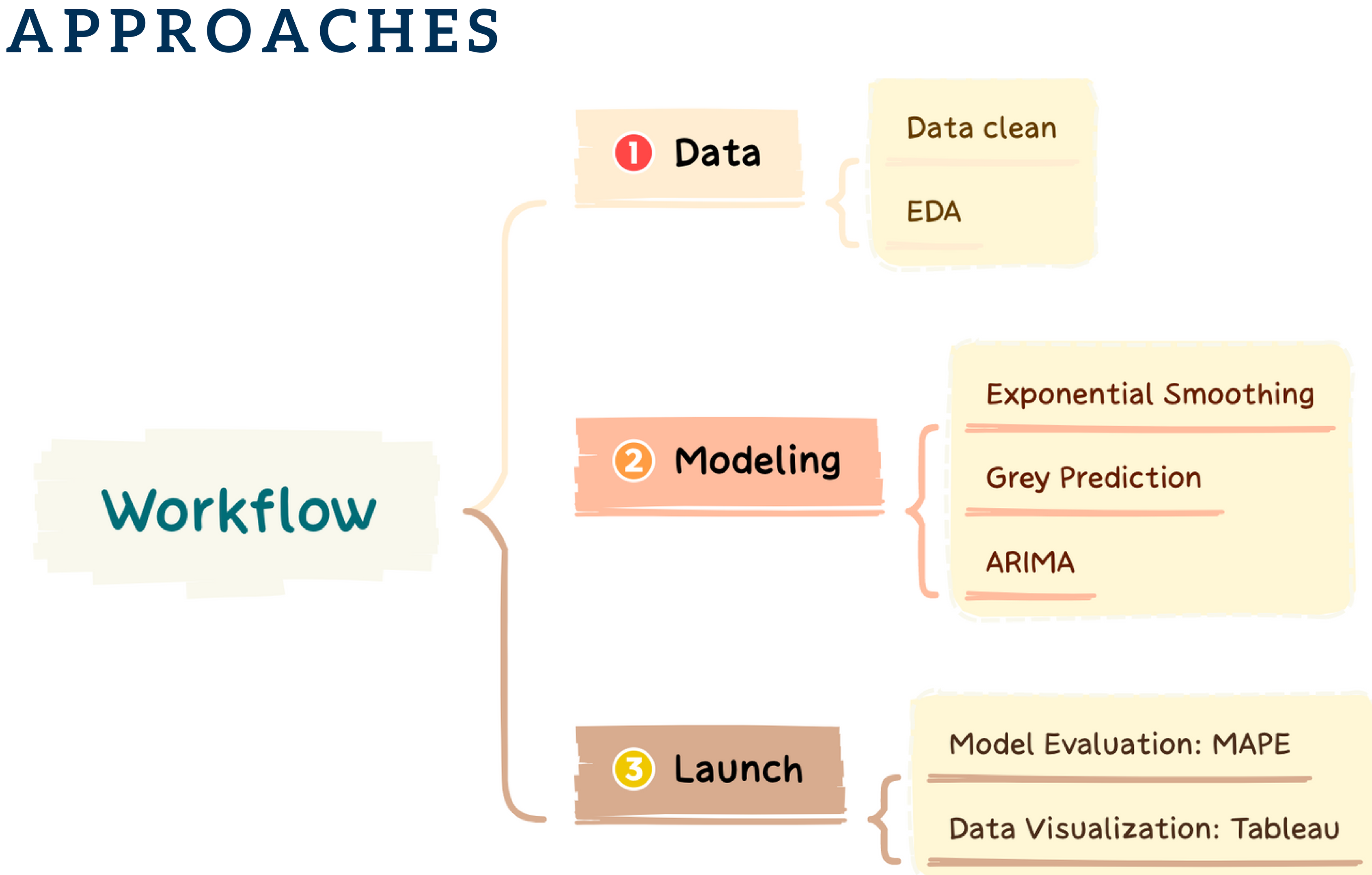
INTRODUCTION

The project features an interactive map visualizing chronic disease spread across the United States, highlighting regions with significant disease prevalence and mortality. Designed for public health enthusiasts, government agencies, and policymakers, the tool provides state-level data and forecasts to aid in resource allocation, planning, and prevention. Success will be evaluated based on user feedback and the impact of resources on high-risk areas.

DATA

The dataset was obtained from the CDC (Centers for Disease Control and Prevention) website. This dataset is sizable, taking up 359.3MB of disk space, and containing 1,232,801 distinct records. It is a temporal dataset, covering years 2007 to 2021, and includes detailed information on various chronic diseases across all 50 states.

APPROACHES



Our project employs Exponential Smoothing, Grey Prediction, and ARIMA models to forecast state-level chronic disease trends using historical data. Exponential Smoothing handles non-seasonal data, Grey Prediction works well with few data points to reveal trends, and ARIMA is a time-series specific model that addresses non-stationarity in time series data.

These models are chosen for their robustness in handling different data characteristics, such as non-stationarity and minimal historical data points. The combination allows for tailored predictions that reflect the unique attributes of chronic disease spread and outcomes. The interactive dashboard enables stakeholders, from policymakers to public health professionals, to access and interpret complex data easily, facilitating informed decision-making and resource allocation.

The innovative aspect of our approach is the melding of time-series predictions with real-time, interactive visualization, tailored to regional nuances, enhancing data utility for effective public health strategies.

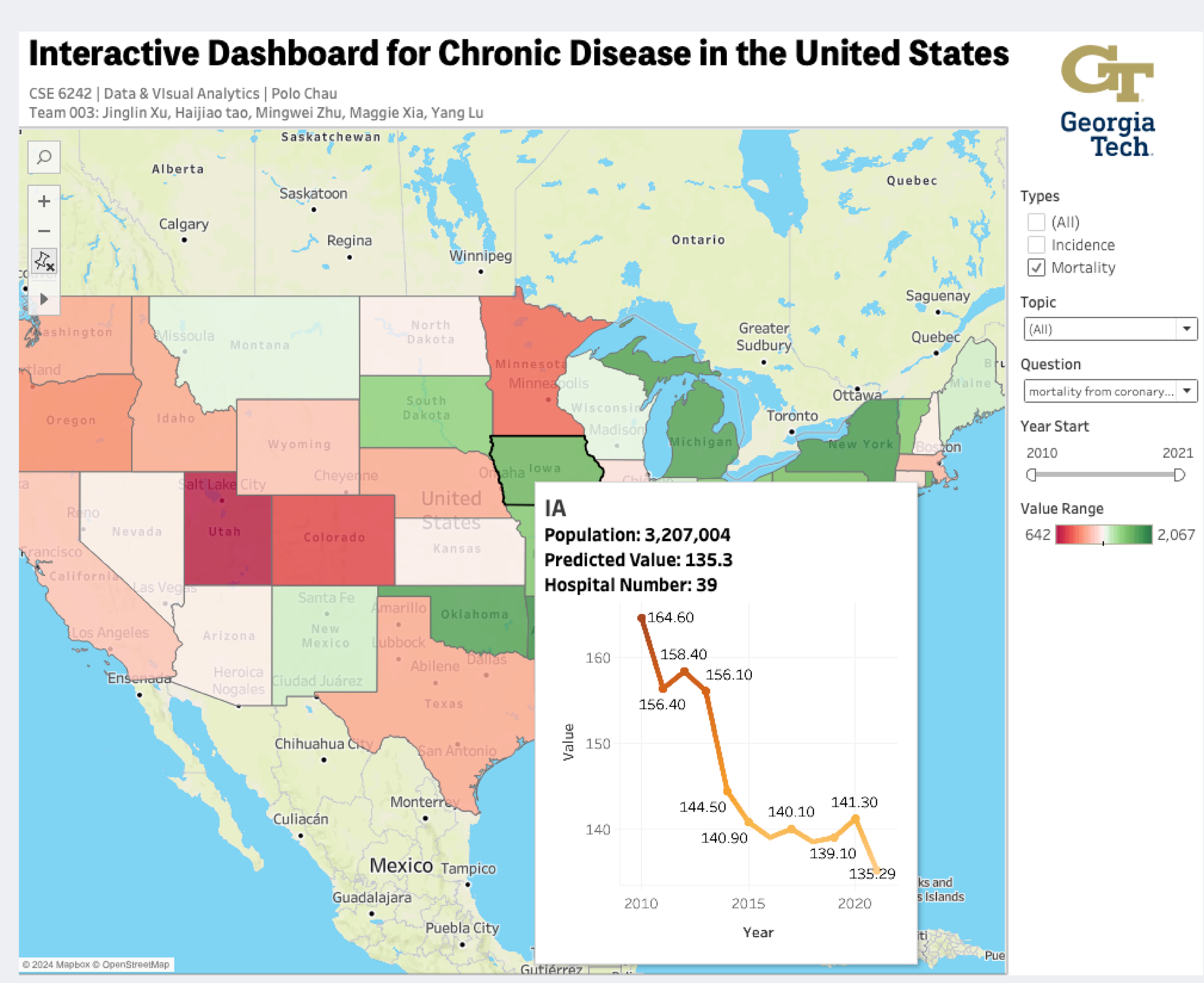
EXPERIMENTS & RESULTS

We evaluate the models using Mean-Absolute-Percentage-Errors (MAPE), chosen over MSE or R^2 due to its scale adaptability and interpretability as percentage differences. We average the MAPEs for mortality and incidence predictions across diseases and states to identify the most accurate model.

	Exponential Smoothing	Grey Prediction	ARIMA
MAPE for Mortality	5.076%	3.673%	NA
MAPE for Incidence	5.498%	5.285%	NA

Grey Prediction gives the best MAPE for both mortality and incidence rates. Exponential Smoothing performs similarly for incidence and slightly worse for mortality. ARIMA does not perform well because it assumes a positive trend over time, while neither mortality nor incidence conform to that trend. We also have very limited yearly data for ARIMA since ARIMA only uses the Y variable during modeling. Thus, Grey Prediction and Exponential Smoothing are better fits given the purpose of our project. Our methods are better suited for time-series data and more robust compared to traditional methods such as linear regression.

DATA VISUALIZATIONS



We create an interactive data dashboard to present the incidence number grouped by disease topic and states, also adding "year" as a filter. By clicking on the states, we can see the trend and our prediction results as well as the state hospital numbers and population so that we can get an idea of how medical resources can be allocated properly.