Election Data Analysis using Tableau
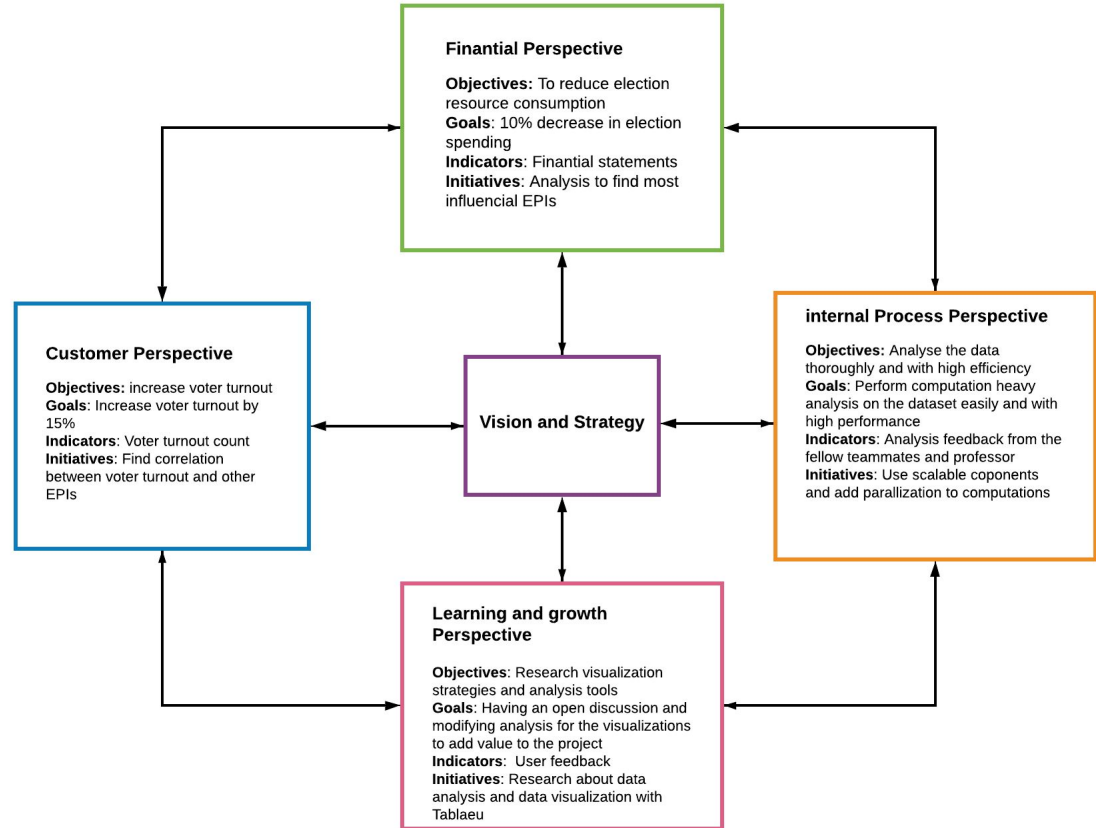
CMPE 274 Project under Prof. Weider Yu

Ajinkya Thakare (012436489)
Akshay Jaiswal (013774137)
Akshay Pagar (013717379)
Janhavi Dahihande (013732030)
Koushik Kumar Kamala (013766571)
Sai Harshith Reddy (013533026)
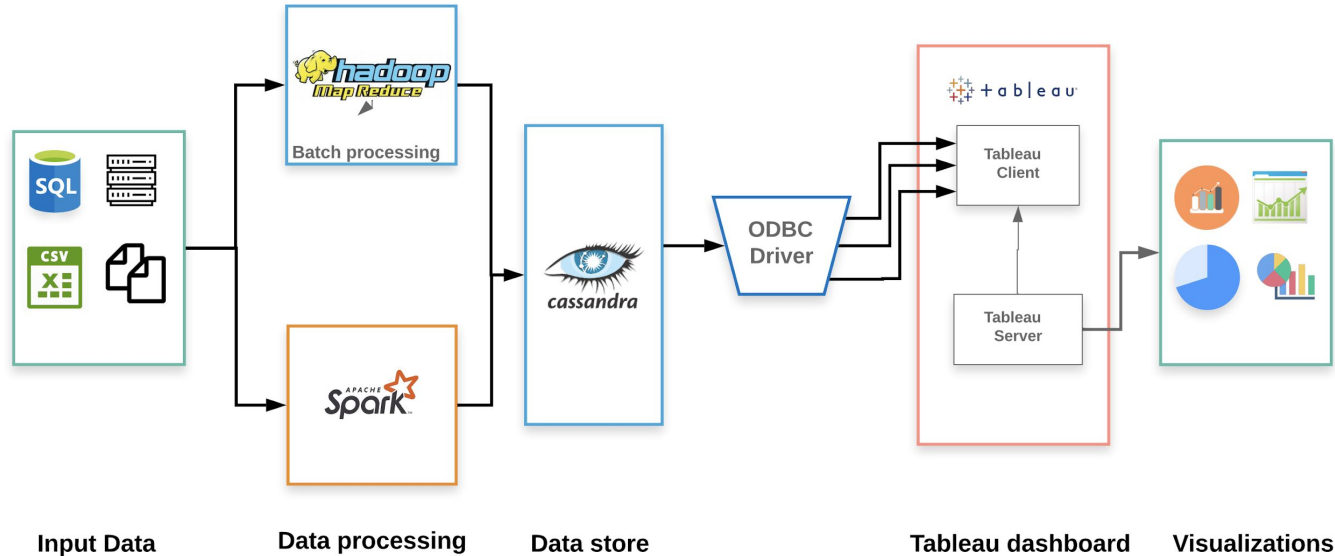Daniel Sampreeth Eadara(013702988)
Madhusudhan Shagam (013707187)

Problem Statement

- Elections are imperative to democracy
- Low voter turnouts impair the objectives of election
  - Let citizens make a contribution in the democratic governance
  - Let citizens voice out opinions
- Waste government resources on elections
- Important to understand the reasons for low voter turnouts

- **Problem Statement:**
  - Use Business Intelligence and Analytical tools to visualize and understand data
  - Derive possible voter turnout patterns that help understand low turnout reasons
  - Suggest plausible solutions that may help fix the problem areas
 - Reduce unnecessary government spending

BPM Scorecard

Finantial Perspective

Objectives: To reduce election resource consumption
Goals: 10% decrease in election spending
Indicators: Finantial statements
Initiatives: Analysis to find most influencial EPIs

Customer Perspective

Objectives: increase voter turnout
Goals: Increase voter turnout by 15%
Indicators: Voter turnout count
Initiatives: Find correlation between voter turnout and other EPIs

Vision and Strategy

internal Process Perspective

Objectives: Analyse the data thoroughly and with high efficiency
Goals: Perform computation heavy analysis on the dataset easily and with high performance
Indicators: Analysis feedback from the fellow teammates and professor
Initiatives: Use scalable coponents and add paralllization to computations

Learning and growth Perspective

Objectives: Research visualization strategies and analysis tools
Goals: Having an open discussion and modifying analysis for the visualizations to add value to the project
Indicators: User feedback
Initiatives: Research about data analysis and data visualization with Tablaeu

System Architecture

Input Data   **Data processing**   **Data store**   **Tableau dashboard**   **Visualizations**

Input Data

- Input data includes simple flat files
- MIT Election Data + Science Lab
  - https://github.com/MEDSL/election-performance-index
  - Contains various key point indices
  - Number of records - 25000
  - Number of columns - 25
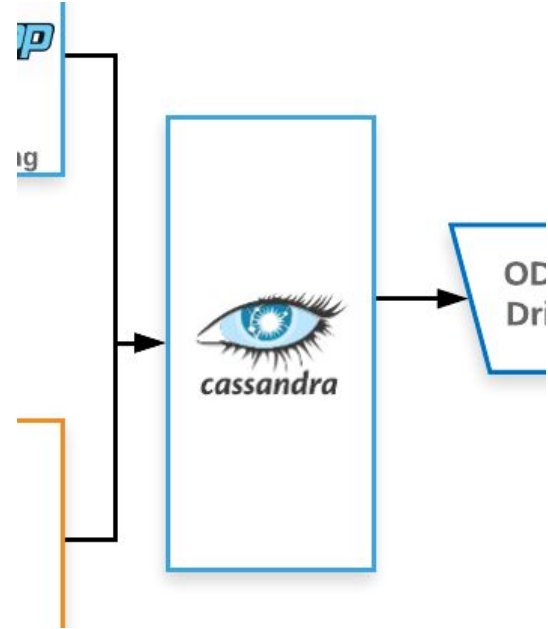- Kaggle 2016 US election dataset (consists of democratic data on counties from the US census)
  - https://www.kaggle.com/benhamner/2016-us-election
  - Number of columns - 54
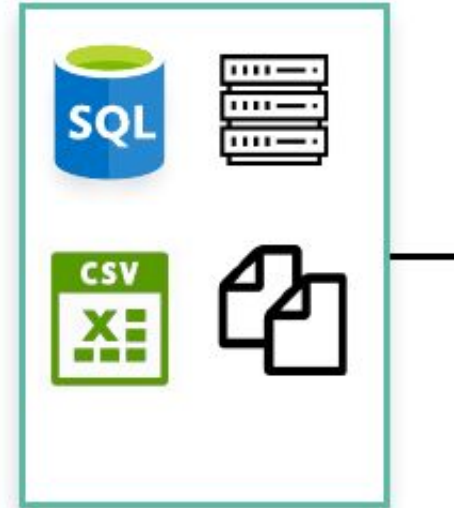
Data Processing

- Anticipating queries is an important part of Cassandra's storage model.
- Spark performs various operations using Hadoop's map-reduce to generate data to be stored in the Cassandra database.
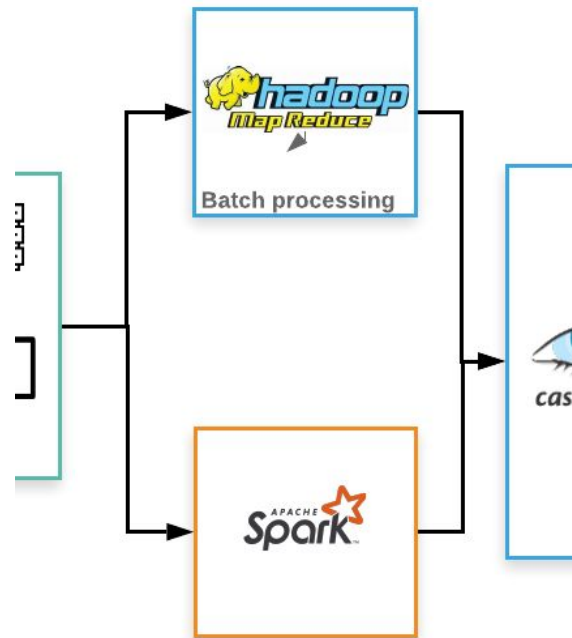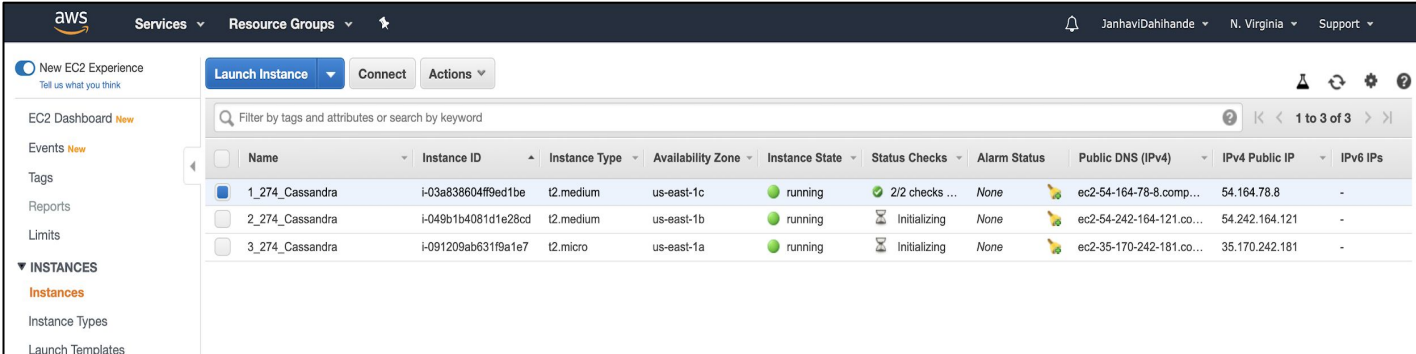
Data Store

- We have used Cassandra as our data store.
- Cassandra is a NoSQL column-oriented database.
- It doesn't have any single point of failure and every node in the cluster is identical.
- We chose Cassandra as our data store because it outperforms the other NoSQL alternatives without compromising performance

Cassandra Cluster on EC2

Database Design

Column families

- Wait time by state
- Voter turnout by state
- Data completion by state
- Ballot rejection by state
- Disability related problems by state
- Registration rejection by state

Cassandra Keyspace

- Focuses on data replication on nodes

- Every node in our cluster contains a keyspace

- Elections keyspace consists of different tables satisfying a particular query

- Query-focussed approach

- Every key performance indicator will have a different column-family as we can see in the database design
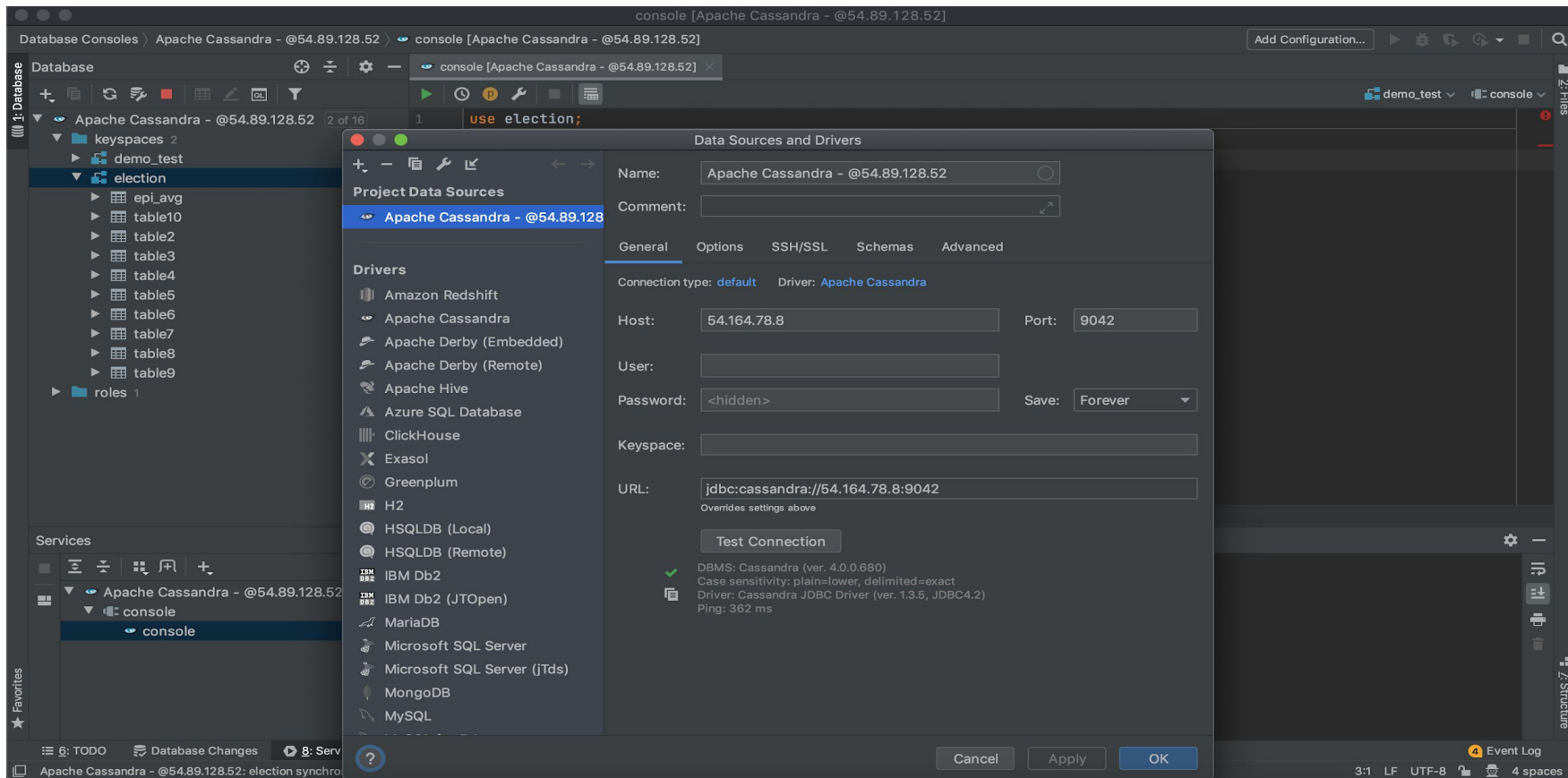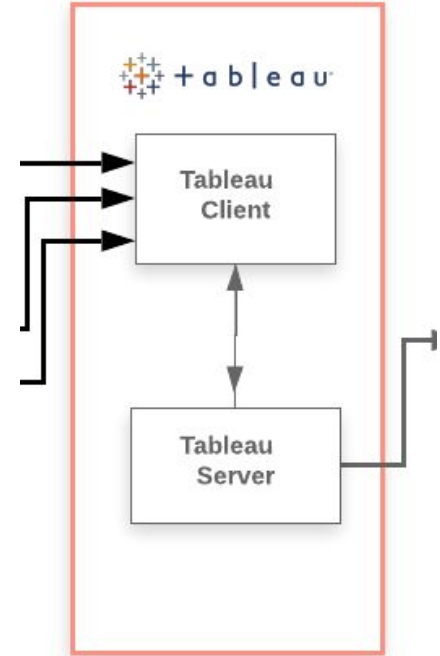
DataGrip

Tableau Dashboard

- Consolidated display of many worksheets
- For comparing and monitoring various data simultaneously
- Data stored in Cassandra can be accessed using ODBC drivers
- Provide various Visualizations using this data

Visualizations

- Visualizations should tell a story
- Use graphs
- Various graphs in our project:
  - Voter turnout trends
  - Ballot rejection, etc

Data Analysis Approach

- Relationship between disability vs turnout group by state

- Registration Rejection vs Turnout

- Wait time vs Turnout

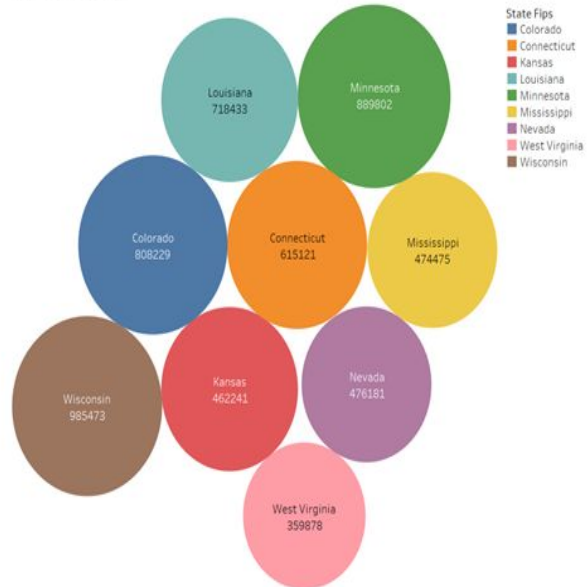- Non-Voter Illness Problems vs Turnout
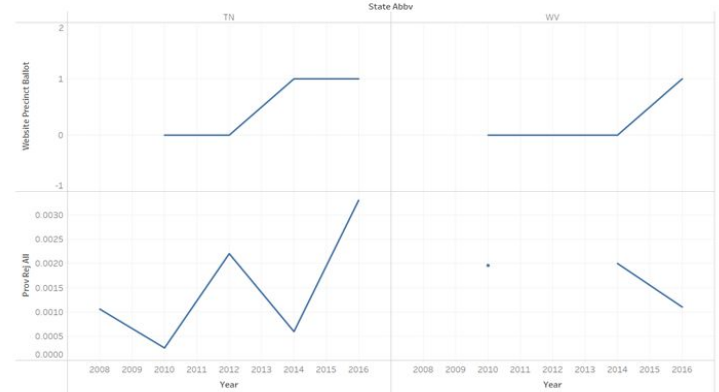
Visualization Design Indicators

- Disability-or Illness-related Voting Problems

- Online Registration Available

- Residual Vote Rate

- Turnout

- Voter Registration Rate and more

Sample Visualizations

age vs turnout

State Fips
- Colorado
- Connecticut
- Kansas
- Louisiana
- Minnesota
- Mississippi
- Nevada
- West Virginia
- Wisconsin

Louisiana 718433
Minnesota 889802
Colorado 808229
Connecticut 615121
Mississippi 474475
Wisconsin 985473
Kansas 462241
Nevada 476181
West Virginia 359878

State Fips and maximum of Data (Sheet1 (state_vs_age)). Color shows details about State Fips. Size shows sum of Vep Turnout. The marks are labeled by State Fips and maximum of Data (Sheet1 (state_vs_age)). The view is filtered on State Fips, which keeps 9 of 51 members.

ballot vs prov_rej

State Abbv

TN   WV

The trends of sum of Website Precinct Ballot and sum of Prov Rej All for Year broken down by State Abbv. The view is filtered on State Abbv, which keeps TN and WV.

Perspectives	Strategic Objectives		Measures / Key Performance Indicators	Targets	Initiatives	Status
	Category	Objective				
Financial	Database	Find free to use datasets and database	No payment for the use of dataset	Using datasets without having to pay the source or creator	Search for datasets and databases	100%
	Data Analysis	Find free to use data analysis tools	No payment for the use of tools	Using tools for analysis without having to pay the source or creator	Search for analysis tools	100%
	Visualization	Find free to use data visualization tools	No payment for the use of tools	Using tools for visualization without having to pay the source or creator	Search for visualization tools	100%
Customer	Database	Use datasets from legitimate, reputed and frequently updating sources	Rating for the references for this data and number of times it is updated every election cycle	Use a legitimate and highly reliable source of data which updates the data frequently	Verify source references and reputation	100%
		Use a database which is reliable, scalable and high performing	Database read/write times	Perform frequent and heavy read/write tasks and get good performance	Perform tests on the database before finalizing	100%
	Data Analysis	Perform thorough analysis of the problem of low voter turnout	Correlation between the EPI indicators	Get meaningful insights from the dataset	Exploratory data analysis	100%
	Visualization	Dashboard that provide all the results to the team and the customers at one glance	Readability and correlation of the EPIs	Showcase all the results of the analysis	Deciding right visualization components	100%
Internal Business Process	Database	To be able to use the dataset and database easily and effectively	Ease of access to the dataset and databases	Use datasets to perform analysis	Make dataset and database easy to access by giving proper access to both	100%
	Data analysis	Analyzing any correlation between indicators	Correlation visualization	Analyzing any correlation between indicators	Analyzed statewise indicators distribution grouped by age	100%
		To be able to perform thorough analysis on the dataset with high efficiency	Analysis feedback from the fellow teammates and professor	Perform computation heavy analysis on the dataset easily and with high performance	Use scalable components and add parallelization to computations	100%
	Visualization	Make visualization results readily accessible in a shared location	Ease of access to the visualization	Easily navigable visualization dashboard	Prioritize the visualization results based on key performance indicators.	100%
		To be able to use visualization tools and features to effectively visualize results	Visually interpretable results	Easy to understand visualization	Strictly follow deadlines to not affect tasks that are dependent on visualization and related processes.	100%
Learning and growth	Database	Finding reliable sources of data and research about various databases	Comparing databases for the same scenario	Find the best possible datasets and databases for the use case	Search on data.org and kaggle for datasets and read technical articles for right databases	100%
	Data Analysis	Sharing techniques and methodolgies related to project progress.	Analyzing any correlation between indicators	Analyze the voter's distribution based on voter registration for each state	Explore and compare various data analysis tools	100%
	Visualization	Research about data analysis and data visualization with tablaeu	Having an open discussion and modifying analysis for the visualizations to add value to the project	Having an open discussion and modifying analysis for the visualizations to add value to the project	Research about data analysis and data visualization with Tablaeu	100%
		Learning about the tablaeu environment setup	Proper install and usage	Use tablaeu to its full potential	100%	100%

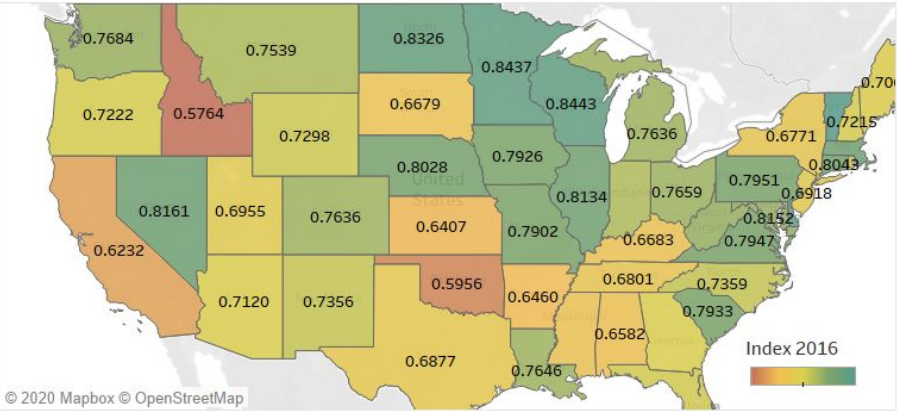# Challenges and new learnings

- Finding key performance indicators
- Finding correlation between indicators
- Performing computation heavy tasks on the dataset
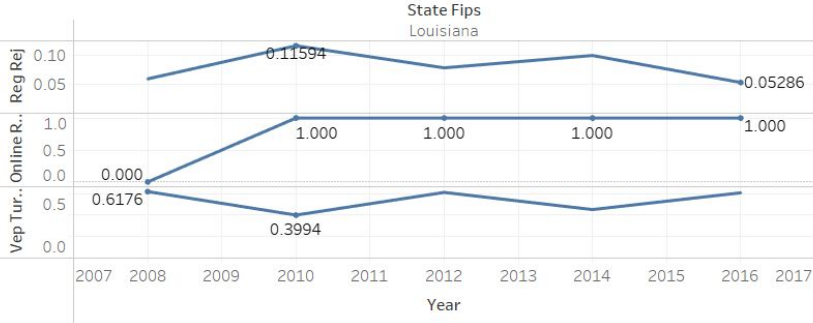- Using cassandra with Apache PySpark and Hadoop

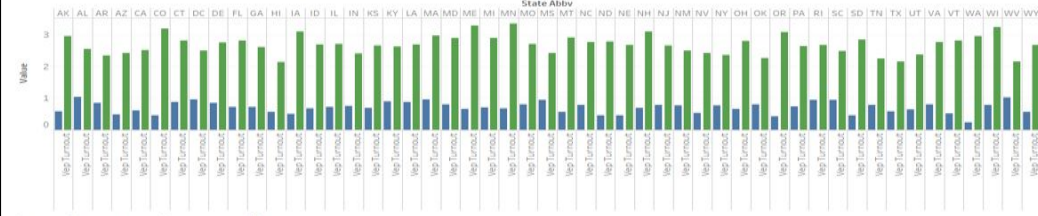# Demo

State Wise - Indicator Average

Registration_Rej vs Turnout

State Fips
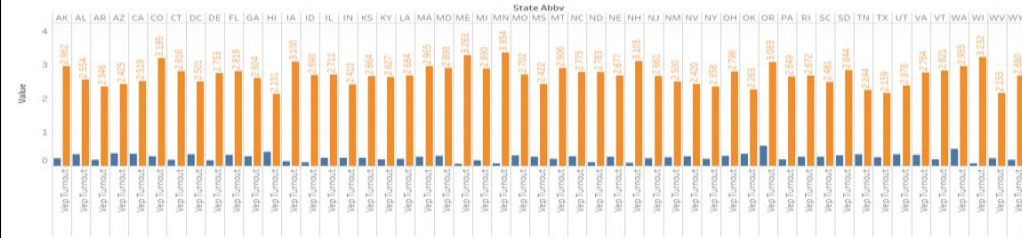Louisiana

Index 2016
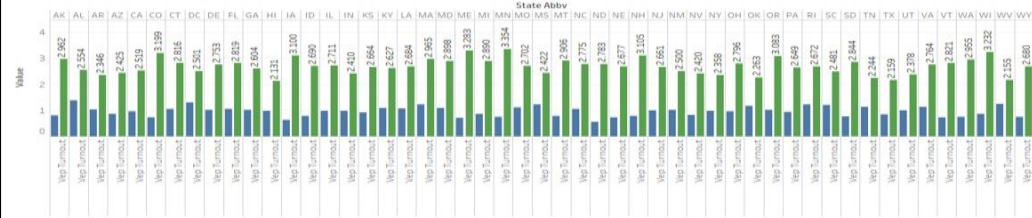
© 2020 Mapbox © OpenStreetMap
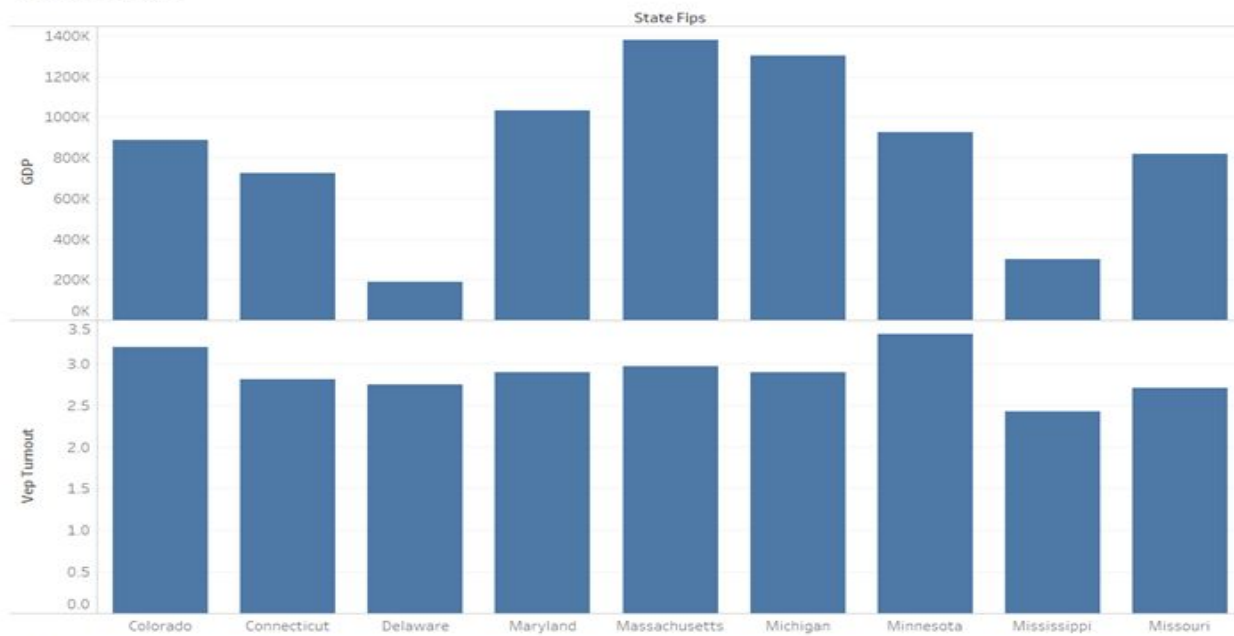
State - Turnout vs Nonvoters-Reg

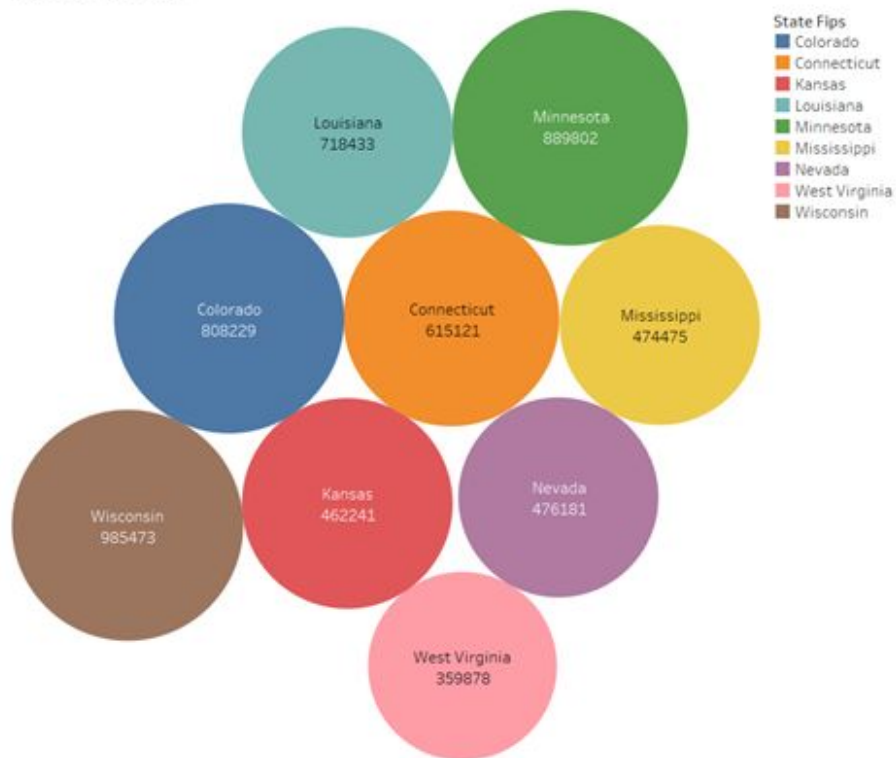State - Turnout vs Nonvoters-Illness
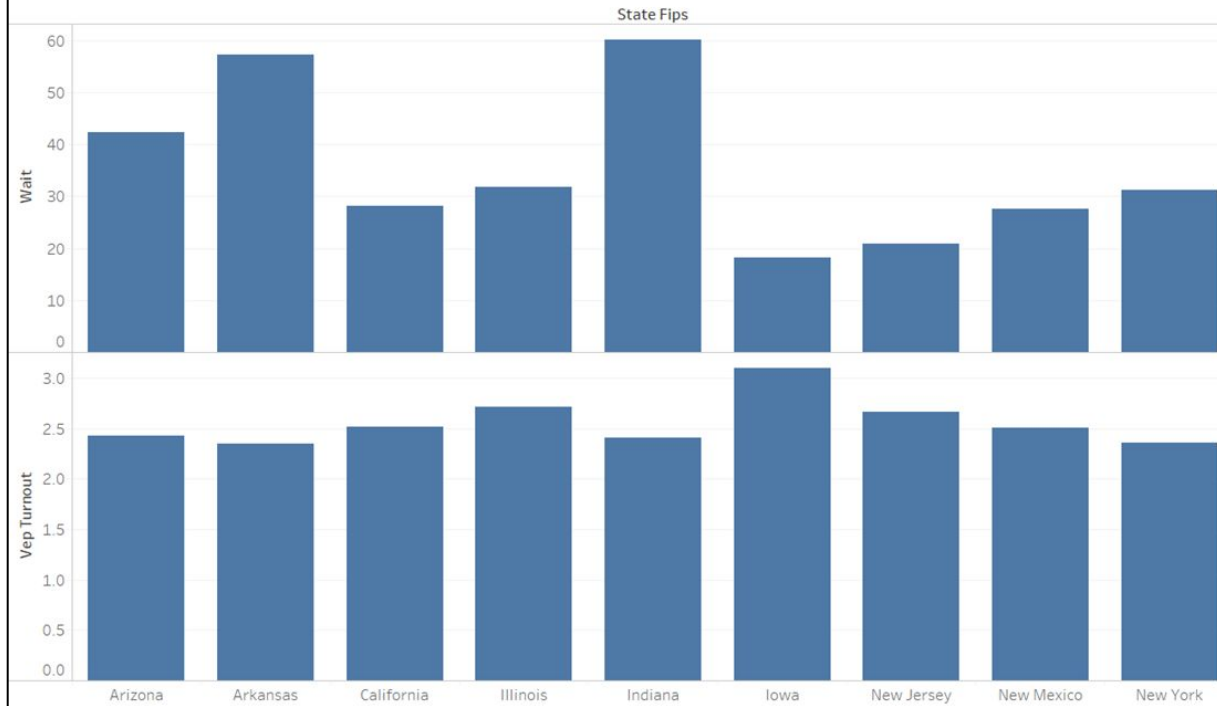
State - Turnout vs Nonvoters-Both

gdp vs turnout

Sum of GDP and sum of Vep Turnout (epi_indicators-all_years) for each State Fips. The view is filtered on State Fips, which keeps 9 of 59 members.

age vs turnout

State Fips
- Colorado
- Connecticut
- Kansas
- Louisiana
- Minnesota
- Mississippi
- Nevada
- West Virginia
- Wisconsin

Louisiana
718433

Minnesota
889802

Colorado
808229

Connecticut
615121

Mississippi
474475

Wisconsin
985473

Kansas
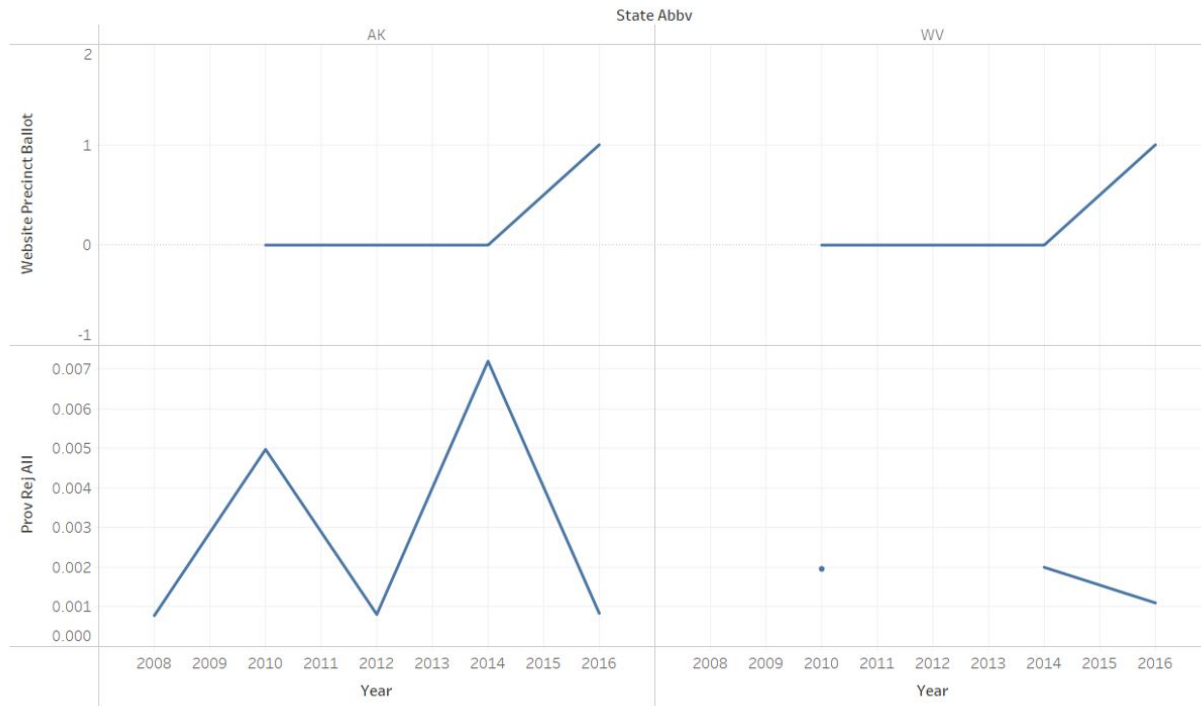462241

Nevada
476181

West Virginia
359878

State Fips and maximum of Data (Sheet1 (state_vs_age)). Color shows details about State Fips. Size shows sum of Vep Turnout. The marks are labeled by State Fips and maximum of Data (Sheet1 (state_vs_age)). The view is filtered on State Fips, which keeps 9 of 51 members.

waittime vs turnout

ballot vs prov_rej

The trends of sum of Website Precinct Ballot and sum of Prov Rej All for Year broken down by State Abbv. The view is filtered on State Abbv, which keeps AK and WV.

References

- *The Measure of American Elections. Eds. Barry C. Burden and Charles Stewart III (New York: Cambridge University Press, 2014)*
- *Environmental Performance Index, http://epi.yale.edu.*
- *The World Justice Project Rule of Law Index, http://worldjusticeproject.org/rule-of-law-index.*
- *Government Accountability Office. Voters With Disabilities: Additional Monitoring of Polling Places Could Further Improve Accessibility. GAO-09-941 (September 2009), http://www.gao.gov/assets/300/296294.pdf.*
- *U.S. Election Assistance Commission, Uniformed and Overseas Citizens Absentee Voting Act (2008 report), 10, http://www.eac.gov/assets/1/Page/2008%20Uniformed%20and%20Overseas%20Citizens%20Absentee%20Voting%20Act%20Survey.pdf. U.S. Elections Assistance Commission, Uniformed and Overseas Citizens Absentee Voting Act (2010 report), 8, http://www.eac.gov/assets/1/Documents/EAC%202010%20UOCAVA%20Report_FINAL.pdf U.S. Election Assistance Commission, Uniformed and Overseas Citizens Absentee Voting Act (2012 report), 9, http://www.eac.gov/assets/1/Documents/508compliant_Main_91_p.pdf.*

Thank You !