

README

Project Title: Minnesota's Congressional District 3 Election Data Analysis (2012–2020)

Overview

This project analyzes election data for Minnesota's Congressional District 3, combining shapefiles from 2012 to 2020. It involves data preparation, visualization, and spatial analysis to uncover voting trends, turnout rates, and party performance across precincts. The analysis is split into two parts, each documented in separate R Notebooks.

Part 1: Data Preparation and Consolidation

Notebook: Data Preparation and Shapefile Consolidation

This notebook focuses on combining shapefiles, filtering for District 3, and creating a dataset for further analysis.

Key Steps:

1. **Data Loading:** Load shapefiles for each election year (2012, 2014, 2016, 2018, 2020) from the base directory.
 2. **Column Name Consistency:** Identify common columns across shapefiles to ensure uniformity.
 3. **Shapefile Merging:** Combine all shapefiles into a single dataset.
 4. **Filtering District 3:** Filter data to include only Minnesota's Congressional District 3.
 5. **Export Dataset:** Save the combined dataset as a new shapefile for subsequent analysis.
-

Part 2: Data Analysis and Visualization

Notebook: Election Data Analysis and Visualization

This notebook continues the analysis by exploring trends, calculating metrics, and visualizing results.

Key Features:

1. **Exploratory Data Analysis:**
 - Explore dataset structure and variables.
 - Summarize key metrics, such as total votes and turnout rates.
 2. **Data Preparation:**
 - Calculate voter turnout rate for each precinct.
 - Handle missing and non-finite values in critical columns.
 3. **Visualizations:**
 - **Interactive Maps:** Visualize precinct-level voting using the `tmap` package.
 - **Voting Trends:** Plot trends in total votes and turnout rates over time.
 - **Party Performance:** Compare votes for Republican, Democratic, and Write-In candidates.
 - **Percentage Breakdown:** Display vote percentage by party across precincts.
 4. **Spatial Analysis:**
 - Perform clustering analysis to identify spatial patterns in voter turnout.
 - Visualize clusters using Moran's I statistics.
 5. **Export Summarized Data:**
 - Save a summarized dataset (e.g., total voters, average turnout rate, party performance) as a CSV file for further analysis.
-

Prerequisites

Software:

- R and RStudio

R Packages:

- **Data Manipulation:** `sf`, `dplyr`
- **Visualization:** `ggplot2`, `tmap`
- **Spatial Analysis:** `spdep`

Input Data:

- Shapefiles for Minnesota's election results (2012–2020) and boundary data for District 3.
-

Running the Analysis

Steps:

1. Clone or download the project files, including shapefiles and R Notebooks.
 2. Install required R packages if not already installed.
 3. Open **Data Preparation and Shapefile Consolidation** notebook:
 - Define the base path for your shapefiles.
 - Execute the notebook to create the combined dataset.
 4. Open **Election Data Analysis and Visualization** notebook:
 - Load the combined shapefile.
 - Execute the notebook to explore data, visualize trends, and perform spatial analysis.
 5. Export results (e.g., summarized CSV, visualizations) as needed.
-

Outputs

1. **Combined Shapefile:** `district_3_combined_2012_2020.shp`
 2. **Summarized Dataset:** `district3_summary.csv`
 3. **Visualizations:**
 - Total voting trends (line plots)
 - Party performance over time
 - Turnout rates by year
 - Precinct-level maps of voter turnout and party percentages
-

Contact

For questions or contributions, contact **Norah Khalaf**.