STATS 290 Project Proposal

Mack Smith, Kate Esbenshade, Luke McFaul

Project Overview

Project Type

☒ New package proposal
☐ Implementation of suggested topic: [Topic name if applicable]

Description

This package will be used to analyze electricity prices in New England by location (locational marginal prices, also known as LMPs). The package will pull from the ISO New England API to get the most recent data on electricity prices and provide functions to analyze and visualize this data. We will use spacial data to visualize the prices on a map of New England. We will also combine this with shapefiles of transmission lines in order to visualize transmission congestion, which is a major driver of electricity prices.

Dependencies

1. httr2

   - What it does: Tools for creating and modifying HTTP requests, then performing them and processing the results.
   - What functionality will you use: Perform HTTP requests for API data retrieval.

2. jsonlite

   - What it does: Parser and generator for JSON.
   - What functionality will you use: Parsing JSON data from API responses.

3. dplyr

   - What it does: Tool for working with data frame like objects for data manipulation.

- What functionality will you use: Cleaning and pre-processing data for analysis and visualization.

4. tidyr

   - What it does: Tools for creating tidy data.
   - What functionality will you use: Transforming data for analysis and visualization.

5. ggplot2

   - What it does: ggplot2 is used for data visualization and mapping
   - What functionality will you use: We will use ggplot2 to create maps of New England with electricity prices and transmission congestion. We will also use it to create other visualizations of the data.

6. sf

   - What it does: Standardized way to encode spatial vector data
   - What functionality will you use: Reading shapefiles for creating graphical visualizations

7. leaflet

   - What it does: Create and customize interactive maps
   - What functionality will you use: Creating interactive maps for visualizing geographical data

8. ggspacial

   - What it does: ggspacial is used for mapping spatial data
   - What functionality will you use: We may use ggspacial to aid in the creation of our maps and visualizations.

9. maps

   - What it does: maps is used for providing basic map data
   - What functionality will you use: We may use maps to aid in the creation of our maps and visualizations for the mapping data inputs.

10. tmap

- What it does: tmap is used for creating thematic maps
- What functionality will you use: We will use tmap to help make our maps easier to visualize and read

11. rgdal

- What it does: rgdal is used for reading and writing spatial data
- What functionality will you use: We will use rgdal to help read in transmission shapefiles for our maps

Similar Packages

Describe any similar existing packages and how your package differs or extends their functionality:

1. USgrid

   - Current capabilities: Provides a set of regular time-series datasets, describing the US electricity grid, including the total demand and supply as well as the demand by energy source (coal, solar, wind, etc.). Additionally, the package only provides data for U.S. mainland states (excluding Alaska and Hawaii).
   - How your package differs: Our package will focus on a specific region within the United States (New England) and will provide data on regarding the pricing of electricity rather than its demand. Additionally, we will be primarily visualizing data using maps rather than graphs.

2. eia

   - Current capabilities: provides API access to data from the US Energy Information Administration (EIA). The datasets returned by these functions are provided in a tidy format. It also offers helper functions for working with EIA API date strings and time formats and for inspecting different summaries of data metadata.
   - How your package differs: Both our package and eia will provide access to an API, but we will be focusing on a different API and focus more on geographical data and visualizations. However, the eia provides a baseline for how we should use R to wrap the data from the API.

Project Planning

Team Member Responsibilities

Luke McFaul (lmcfaul)

Responsibilities:

1. Create ways to visualize LMP data and transmission congestion data on maps.
2. Build functions that allow for easier presentation of different electricity data cartographically.
3. Combining and visualizing data with optionality to allow users to explore and analyze New England grid data.
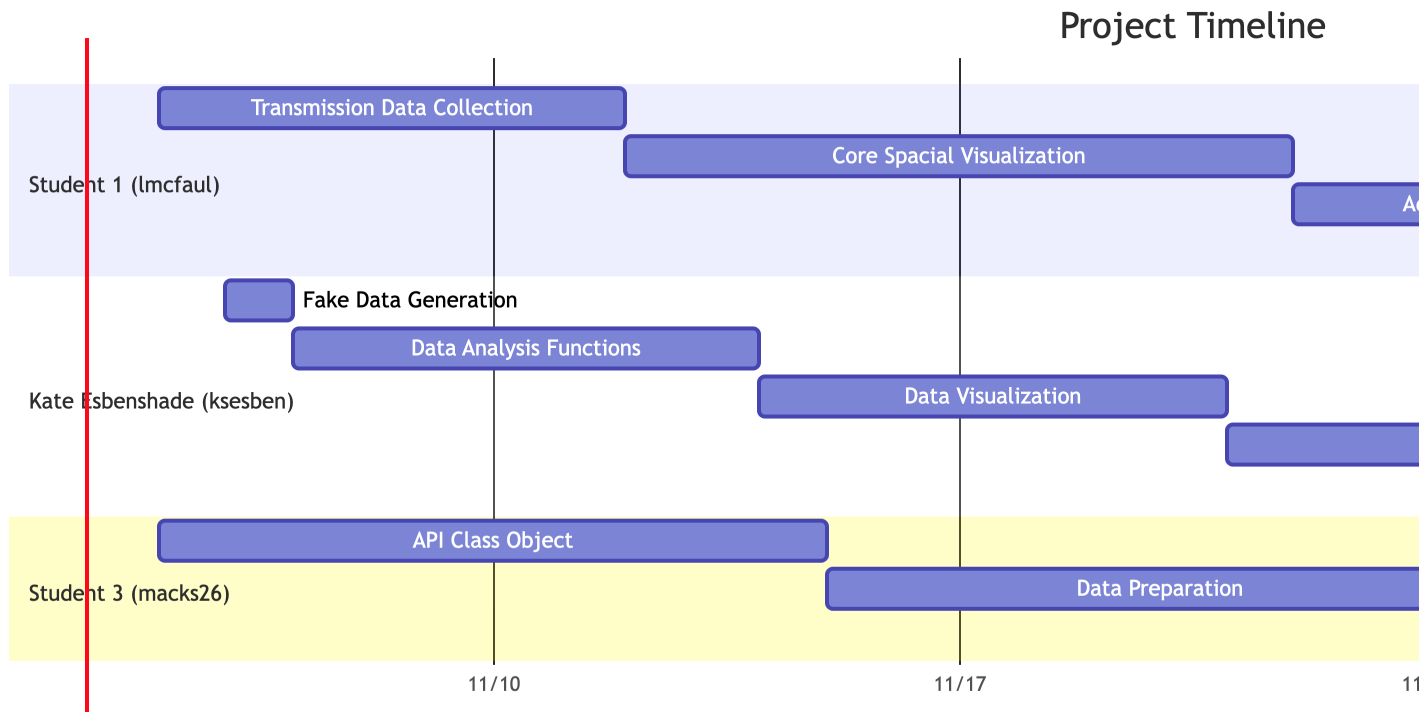
Mack Smith (macks26)

Responsibilities:

1. Developing class for collecting and parsing data from API.
2. Cleaning, pre-processing, and transforming of data (once parsed) for analysis and visualizations.
3. Writing documentation and demo for using the developed API class.

Kate Esbenshade (ksesben)

Responsibilities:

1. Data analysis: Creating functions that allow for easy data analysis, such as price comparisons between locations and time periods.
2. Data visualization: Creating functions that allow for easy visualization of the data, outside of the maps.

Timeline

Project Timeline

Student 1 (lmcfaul)
- Transmission Data Collection
- Core Spacial Visualization
- A...

Kate Esbenshade (ksesben)
- Fake Data Generation
- Data Analysis Functions
- Data Visualization

Student 3 (macks26)
- API Class Object
- Data Preparation

11/10   11/17   11

Additional Considerations

Scope and Limitations

- This project will only be focusing on the states in the New England Independent System Operator (ISO), which is Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. The data is separate for each ISO, so to do a full analysis of the United States, we would need to expand the package to include data from other ISOs.
- There will not be spacial data on generating stations, so we will only be estimating congestion through prices and transmission lines.

Future Extensions

- Future expansions to this package could include more geographic areas of the U.S., or more ways to visualize the data. We are kind of imitating the website https://www.gridstatus.io/ but with more ways to visualize the data, so in theory this could use all of the data that grid status does.
- Similarly, an expansion could add generating station data (if available) to the package to get a more complete picture of the electricity grid.
- Lastly, an expansion could make modeling changes such as what the impacts of adding a new transmission line on prices would be, or methods of price prediction such as weather data.