



DPP2 Project Presentation

A closer look at the US household energy consumption using RECS data

Ellie Chen, Xiaoqi Zhou, Lee Ko

A large orange circle is positioned on the left side of the slide, partially cut off by the edge.

Research Question

How do **climate** and **house age** affect household energy consumption across the US states?

A series of four yellow dashed line segments are arranged in a curved, upward-sloping pattern in the bottom right corner of the slide.

Data Source

- Datasets
 - 2015 Residential Energy Consumption Survey (RECS)
 - 2021 American Housing Survey (AHS)
- NLP, Web scraping
 - US Energy Information Administration Annual Energy Outlook reports
 - State Energy Data Report energy consumption table



Project Layout

: how to satisfy key
requirements

1. Data wrangling

- 2015 RECS data, 2021 AHS data
- state-level energy consumption data (Web-scraping)
- Common key column: state id

2. Model analysis

- Regression: Energy consumption – Household characteristics, climate categories
- (E.g.) Household energy consumption level \sim house age + age of major A/C facilities + size of house (# of rooms, area) + climate type (categorized values)

3. Text processing

- Multiple years of energy outlook reports (same source)
- Keyword analysis
- Sentiment analysis

4. Plotting

- Shiny plot #1: US map with average energy consumption level by each state
 - Options to choose type of energy, sub-categories of household characteristics
- Shiny plot #2: Stacked bar chart for the composition of the overall energy consumption
- Static ggplots: text processing results
- (TBD) Shiny or stats plots for AHS analysis

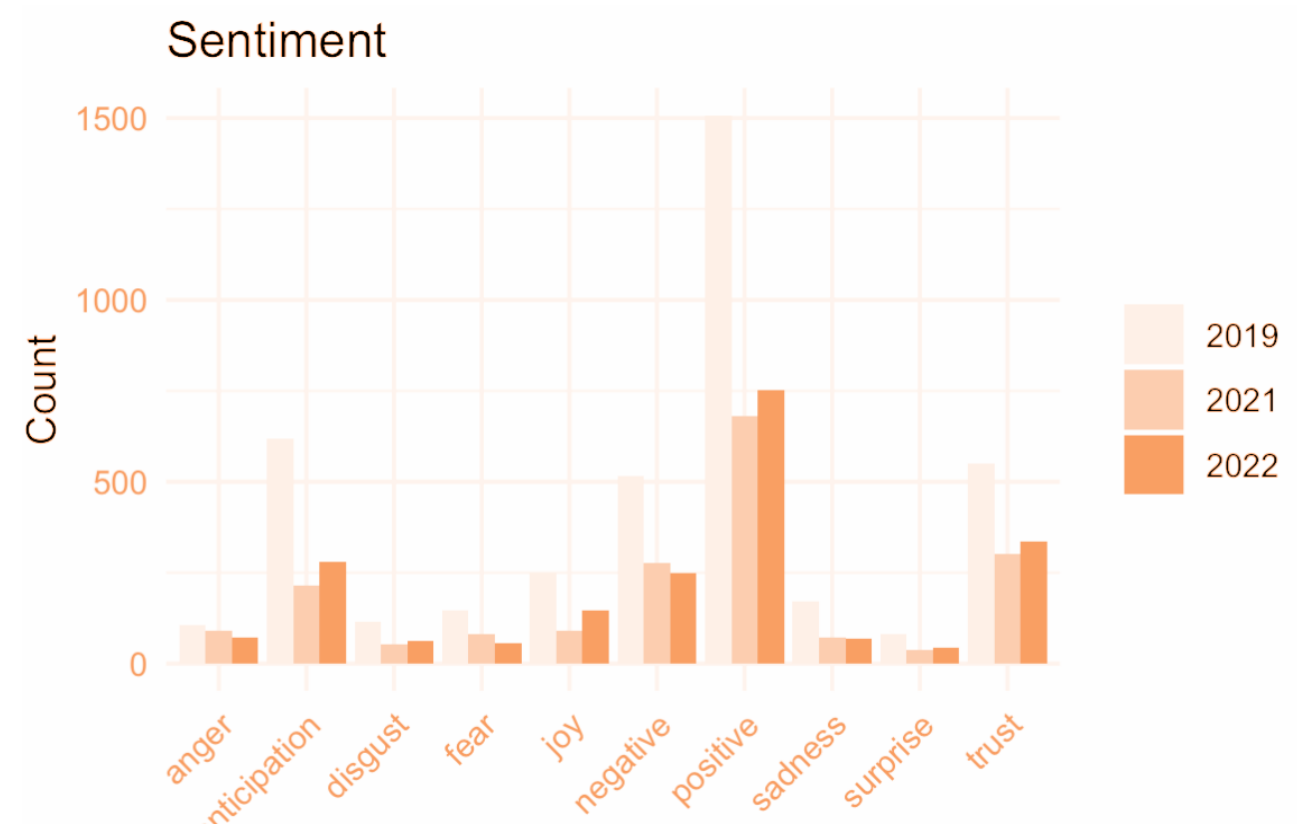


Project Layout

: how to satisfy key requirements

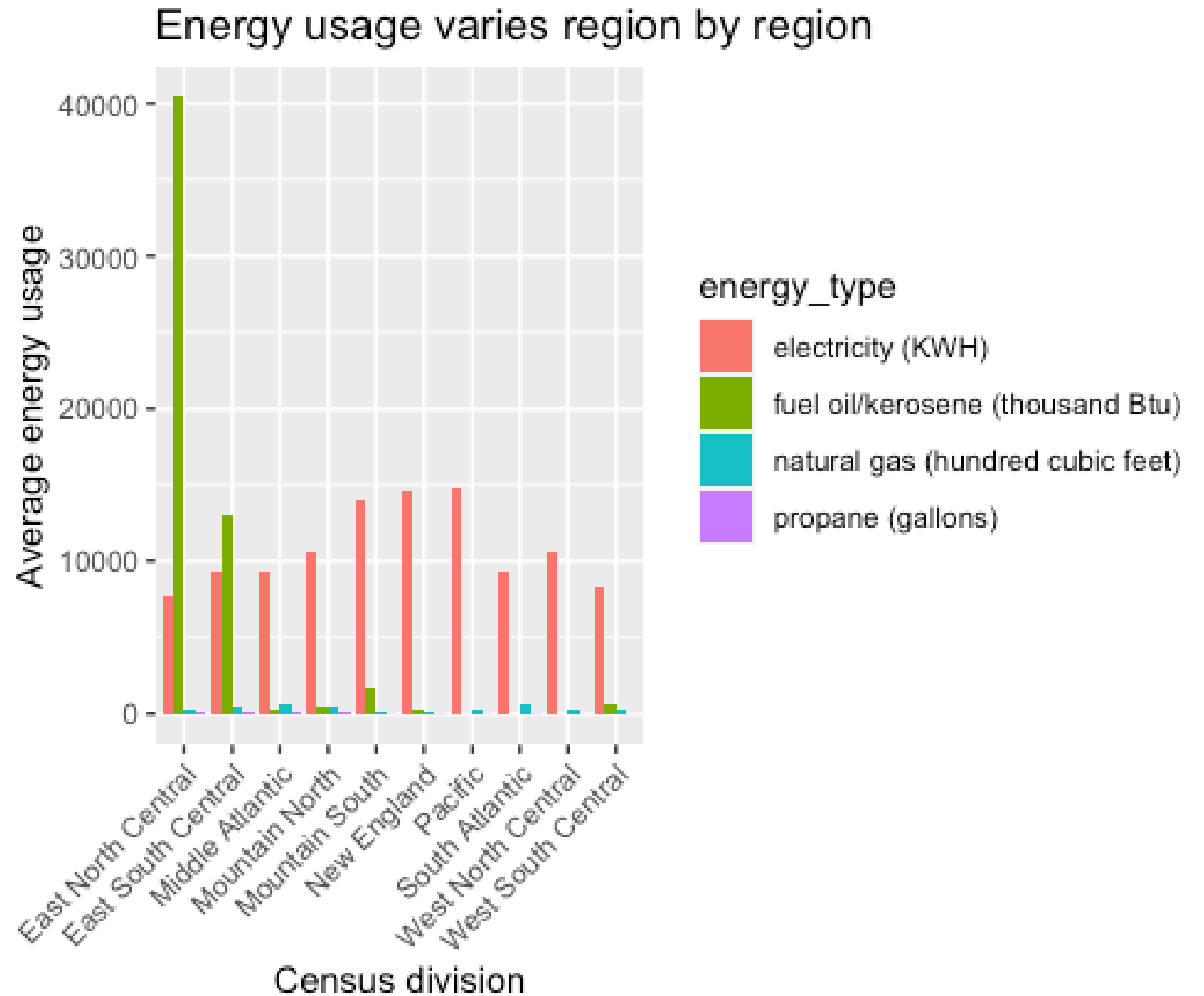
Sentiment Analysis Sample Chart

Source: Annual Energy Outlook by EIA (2011-2022)



Project Layout

: how to satisfy key requirements




Project Status

What we have done

- 10/16 Project topic discussion
- 10/28 Specified data set, research question
- 11/3 RECS data preliminary analysis, Regressions using major variables
- 11/9 Interpreted individual work results (RECS regression, NLP on a single report, web-scraping a table), Decided on the project layout

Tasks for the following meetings

- AHS data statistical analysis
 - Design and build Shiny apps structure, code effective plots
 - Further sentiment & keyword analysis on multiple reports
 - Write-up for the analysis
- 

Challenges

- Relatively outdated RECS data
 - Unable to use the most recent one – missing expenditure/consumption statistics
 - Compare the accessible subset with the previous one (2015 RECS data)
- Extracting meaningful implications from NLP on multiple years
 - How to check the validity of report
 - Test with a critical energy-relevant event





Questions & Feedback