**The Roux Institute at Northeastern University**

**User Guide for The State of Maine Governor’s Energy Office**

6/30/2022

Alaina Cellini, Emma Christakis, Henry Gent, Ryan Goebel, Molly Goulet, Sarah Hogan, Amy Macdonald, Ian Mohler,

*Analytics XN Team - ALY6980, Spring 2022*

**Data Sources:**

*Note: Please refer to the data hub spreadsheet for instructions on how to navigate to the original source of a given dataset as well as how to clean the data via R-scripts.*

1. **Download R**

Note: *If you have already completed this section, please move section 2.0*

* 1. Download R at <https://cloud.r-project.org/>

**Note**: Choose the correct processing system (Windows, Mac, Linux)

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

* 1. Download R Studio at <https://www.rstudio.com/products/rstudio/download/>. Scroll all the way to the bottom and install RStudio Desktop for your correct operating system

Graphical user interface, application

Description automatically generated

1. **Download Power BI**
   1. Download Power BI Desktop at <https://powerbi.microsoft.com/en-us/downloads/>
   2. Select Advanced Download Options under Microsoft Poweer BI Desktop

Graphical user interface, text, application, email

Description automatically generated

* 1. Select the correct language and press Download

Graphical user interface, text, application

Description automatically generated

1. **Steps to create the Excel File titled “Historical\_dollars\_per\_million\_btu.xlsx”**
   1. Find the following files
      1. historical propane - k1 - #2 price tables\_charts1.xlsx (file provided to Roux Institute by GEO)
      2. Electricity Data of Sales and Revenue from <https://www.eia.gov/electricity/data.php#summary> from 1900 through current in an XLS file
      3. Natural Gas Data from <https://www.eia.gov/dnav/ng/hist/n3010me3m.htm> in the form of XLS.
   2. Navigate and open the R file titled “pricing - historical\_dollars\_per\_million\_btu”
   3. Run the R file following the steps in it utilizing the updated Excel files gathered in step 3.1
   4. Convert the file to an excel spreadsheet

*Note: we chose to convert all our files to excel files before loading our data into PowerBI. If you do not wish to do this, you can skip step 3.4 and complete step 3.5 in PowerBI.*

* 1. In the excel file select “Find and Replace”
     1. Replace all NA’s with null values (“”)
     2. Save as “historical\_dollars\_per\_million\_btu.xlsx” and close

1. **Gather the Excel File titled “all\_SEDS\_Consumption.xlsx” and locate the Historical Events Excel File**
   1. Navigate to the CSV file titled “all\_SEDS\_Consumption.csv”
   2. Convert the file to an excel spreadsheet to now be titled “all\_SEDS\_Consumption.xlsx”
   3. Locate CSV file called “Historical Events.xlsx”
2. **Steps to Create Power BI Visualizations for Visual Communication Section:** 
   1. Updating the VisualCommunications.pbix file
      1. Open the “VisualCommunications.pbix” file
      2. Select “Transform Data” in the top ribbon

Graphical user interface, application

Description automatically generated

* + 1. On the right side, select the table that you would like to update the source for.
    2. Once you have selected the correct table, go to the left menu called “Applied Steps” and select “Source”

Graphical user interface, text, application

Description automatically generated

* + 1. Expand the top bar so you see the file path of the current file. Copy and paste your new file path over the current one. Press the check mark.

Graphical user interface, application, Word

Description automatically generated

*Note: As long as your file stays in the same place, this should only need to be updated once.*

* + 1. The files to be replaced are as follows
       1. all\_SEDS\_Consumption.xlsx
       2. Historical Events.xlsx
       3. historical\_dollars\_per\_million\_btu.xlsx
  1. Updating Energy Cost Calculator
     1. Open the file titled “EnergyCostCalculator.pbix”
     2. Publish to where you would like and update the numbers accordingly

*Note: No updated files are needed at any given time to update this Power BI report as this is an interactive report for the end user to update the amount of usage as necessary for their household.*

1. **Steps to Create Power BI Visualizations for Historical Comparisons Section:** 
   1. Updating the HistoricalComparisons.pbix file
      1. Open the “HistoricalComparisons.pbix” file
      2. Select “Transform Data” in the top ribbon

Graphical user interface, application

Description automatically generated

* + 1. On the right side, select the table that you would like to update the source for.
    2. Once you have selected the correct table, go to the left menu called “Applied Steps” and select “Source”

Graphical user interface, text, application

Description automatically generated

* + 1. Expand the top bar so you see the file path of the current file. Copy and paste your new file path over the current one. Press the check mark.

Graphical user interface, application, Word

Description automatically generated

*Note: As long as your file stays in the same place, this should only need to be updated once.*

* + 1. The files to be replaced are as follows
       1. SEDS\_CPEP\_AllStates
       2. SEDSConsumption\_AllStates\_wPop

1. **Heating Oil and Electricity Price Forecasts and Comparisons** 
   1. **Model Selection and Analysis**

*This section provides instructions on how to run the .R code that was used to analyze various forecast models for heating oil and electricity prices, including how to obtain the required data files.*

* + 1. To analyze and choose an appropriate forecast model for heating oil prices:
       1. Data Collection:

*Place the following files in the “price forecasts\data” folder:*

* + - * Monthly Maine heating oil #2 prices:
        + Use the internal Maine GEO spreadsheet for tracking propane, kerosene, and heating oil #2 prices.
        + File should be named "historical propane - k1 - #2 price tables\_charts1.xlsx"
      * Monthly New England on-highway diesel prices:
        + Navigate to <https://www.eia.gov/petroleum/gasdiesel/>
        + Scroll down to "U.S. On-Highway Diesel Fuel Prices (dollars per gallon)" and click on "full history XLS"
        + File should be named "psw18vwall.xls"
      * Monthly WTI spot prices in Cushing, Oklahoma:
        + Navigate to <https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=M>
        + Click "Download Data (XLS File)" to obtain historic monthly
        + File should be named "RWTCm.xls"
      * Monthly average temperatures for Maine:
        + Navigate to <https://www.ncdc.noaa.gov/cag/statewide/time-series/17/tavg/all/12/2004-2022>
        + In drop-down menu, verify "End Year" is current year
        + Click "Plot”
        + Below the plot, click on the Excel icon to download data in CSV format
        + File should be named something similar to "17-tavg-all-12-2004-2022.csv".
        + Note: There should only be one file in the data directory with "tavg-all" in the filename.
      * Global Economic Uncertainty Index (GEPU):
        + Navigate to <https://www.policyuncertainty.com/global_monthly.html>
        + Click "Download Global EPU Data"
        + File should be named "Global\_Policy\_Uncertainty\_Data.xlsx"
      * Geopolitical Risk Index (GPR):
        + Navigate to <https://www.policyuncertainty.com/gpr.html>
        + Click “Download Data (New Methodology)”
        + File should be named “data\_gpr\_export.xls”
      * Monthly average closing price for the S&P 500:
        + Navigate to <https://www.marketwatch.com/investing/index/spx/download-data>?
        + Data can only be downloaded in one-year chunks
        + Scroll down to "Historical Quotes"
        + Select Start Date of 01/01/2004 and End Date of 12/31/2004
        + Click "Update Results"
        + "Daily" should be selected for Result Frequency
        + Click "DOWNLOAD DATA (.CSV)"
        + Update the Start and End Dates years to 2005 and do the same
        + Continue for all years up to and including the current year

Note: If your data folder already has these files, you only need to download an updated version of the current year

* + - * + There should now be several files in the data folder named "Download Data - INDEX\_US\_S&P US\_SPX.csv" or similar.
      1. From the “price forecasts” folder, run “import\_heatingoil\_data.R” in RStudio.
      2. From the “price forecasts” folder, run “import\_extfactors\_data.R” in RStudio.
      3. From the “price forecasts” folder, run “model\_selection\_heatingoil.R” in RStudio.
    1. To analyze and choose an appropriate forecast model for electricity prices:

*Place the following files in the “price forecasts\data” folder:*

* + - Monthly average electricity prices:
      * Navigate to <https://www.eia.gov/electricity/data.php#summary>
      * Expand the menu for "Sales (consumption), revenue, prices & customers"
      * Click "XLS" under "Monthly Form EIA-861M (formerly EIA-826) detailed data (1990 - present)"
      * File should be named "sales\_revenue.xlsx"
      1. From the “price forecasts” folder, run “import\_electricity\_data.R” in RStudio.
      2. From the “price forecasts” folder, run “model\_selection\_electricity.R” in RStudio.
  1. Export forecasts in CSV format:
     1. Heating Oil Price Forecast:
        1. Data Collection:

*Place the following files in the “price forecasts\data” folder:*

* + - * Monthly Maine heating oil #2 prices:
        + Use the internal Maine GEO spreadsheet for tracking propane, kerosene, and heating oil #2 prices.
        + File should be named "historical propane - k1 - #2 price tables\_charts1.xlsx"
      * Monthly New England on-highway diesel prices:
        + Navigate to <https://www.eia.gov/petroleum/gasdiesel/>
        + Scroll down to "U.S. On-Highway Diesel Fuel Prices (dollars per gallon)" and click on "full history XLS"
        + File should be named "psw18vwall.xls"
      1. From the “price forecasts” folder, run “import\_heatingoil\_data.R” in RStudio.
      2. From the “price forecasts” folder, run “export\_heatingoil\_price\_forecast.R” in RStudio.
    1. Electricity Price Forecast:
       1. Data Collection:

*Place the following files in the “price forecasts\data” folder:*

* + - * Monthly average electricity prices:
        + Navigate to <https://www.eia.gov/electricity/data.php#summary>
        + Expand the menu for "Sales (consumption), revenue, prices & customers"
        + Click "XLS" under "Monthly Form EIA-861M (formerly EIA-826) detailed data (1990 - present)"
        + File should be named "sales\_revenue.xlsx"
      1. From the “price forecasts” folder, run “import\_electricity\_data.R” in RStudio.
      2. From the “price forecasts” folder, run “export\_electricity\_price\_forecast.R” in RStudio.
  1. Heating Oil and Electricity Price Forecast Document

*This section provides instructions on how to update the R Markdown (.Rmd) file that displays the heating oil and electricity price forecasts in a user-friendly html format.*

* 1. If you have not already done so, collect the following files and place them in the “price forecasts\data” folder:
     1. Monthly Maine heating oil #2 prices:
     2. Use the internal Maine GEO spreadsheet for tracking propane, kerosene, and heating oil #2 prices.
     3. File should be named **"historical propane - k1 - #2 price tables\_charts1.xlsx"**
  2. Monthly New England on-highway diesel prices:
     1. Navigate to <https://www.eia.gov/petroleum/gasdiesel/>
     2. Scroll down to "U.S. On-Highway Diesel Fuel Prices (dollars per gallon)" and click on "full history XLS"
     3. File should be named "psw18vwall.xls"
  3. Monthly average electricity prices:
     1. Navigate to <https://www.eia.gov/electricity/data.php#summary>
     2. Expand the menu for "Sales (consumption), revenue, prices & customers"
     3. Click "XLS" under "Monthly Form EIA-861M (formerly EIA-826) detailed data (1990 - present)"
     4. File should be named "sales\_revenue.xlsx"
  4. Open “forecast.Rmd” in RStudio
  5. Click “Knit” on the RStudio menu

1. **Consumption Forecasting Code and Data**

*Note that both scripts in this folder are heavily commented with steps to follow to execute them and locations to get data. The .Rmd file is an RMarkdown that formats the forecast as a report, and the .R file produces the results in the R console and allows for generating csv’s of the prepared data.*

To execute, three files will be needed.

* 1. Obtain SEDS Consumption Data from https://www.eia.gov/state/seds/seds-data-complete.php?sid=ME#Consumption by clicking 'all consumption estimates'
  2. Obtain SEDS Codes from the same page by clicking on 'codes and descriptions' at the bottom of the page. Select just the data in the table 'MSN Descriptions' and paste into a csv (I have included this file in the folder).
  3. Obtain Heat Pump Data from Dan at EMT (Data through 2021 is included in the folder).
  4. When Running the R script

*If you do not have the required libraries, R Studio will ask you to install them when you open the R file. Install all of them (this will only need to be done once*

*You can run all the code in the script all at once. As it runs, it will on three occasions open a file explorer window. For the first prompt, select the SEDS file. For the second, select the codes file. For the third, select the Heat Pumps file. The code will take care of everything else, and plots will be available in the plots pane*

* 1. To export prepared datasets to csv

*To generate a csv that is simply the SEDS data in “vertical format” rather than with each year being a column (shown below) uncomment line 74 and run it. It will ask where you want the file located and the name and then put a copy there.*

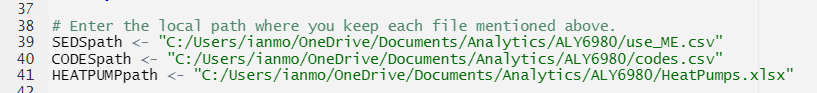
Text

Description automatically generated with medium confidence

*To generate a file with actual and forecast data in it, uncomment line 271 and run it, much like above. This file includes all of history like the above file, with 10 years of forecast for each type. The forecast contains the fitted points, and the upper and lower 95% confidence interval results.*

* 1. When Running the Rmd file

*This will be very similar, except that you will need to edit the filepaths within quotes on lines 39-41 to have the filepath to the three files needed where they are stored on your computer (see below, this is where I kept them on my computer).*



*After you do this, you will not need to worry about selecting the files from the file explorer when the code runs, it will know where to find them. Then you can click run > run all. After that you can select Knit and it will create a report.*

1. **Historical Comparisons Code, Data, and PowerBI**

*Note that the single R script in this folder is heavily commented with steps to follow to execute and locations to get data. The .R file produces the results in the R console and allows for generating csv’s of the prepared data that are input into the Power BI to compare Maine to other states*

To execute, four files will be needed.

* 1. Obtain SEDS Consumption Data for all States from <https://www.eia.gov/state/seds/seds-data-complete.php?sid=US#CompleteDataFile> listed under Data Files, Consumption, in Btu, and click on CSV. CSV file should be titled *use\_all\_btu.csv*
  2. Obtain SEDS Codes from the same page by clicking on 'codes and descriptions' at the bottom of the page. Select just the data in the table 'MSN Descriptions' and paste into a csv (I have included this file in the folder titled as *codes.csv*)
  3. Obtain SEDS Population data from the same page by going up to the Key Statistics and Indicators section, down to the Other section listing Population and GDP, and clicking on the XLSX to download.. Select just the data and headers on the Total Population tab i paste into a csv (I have included this file in the folder titled as *total\_population\_thousand.csv*)
  4. Obtain SEDS Consumption, Pricing, Expenditure, and Production data from the same page by going back down to Data Files section, down to the final Consolidated Data File section listing Consumption, price, expenditure, and production estimates and clicking on the CSV to download.. CSV file should be titled *Complete\_SEDS.csv*
  5. When Running the R script

*If you do not have the required libraries, R Studio will ask you to install them when you open the R file. Install all of them (this will only need to be done once*

*You can run all the code in the script all at once. As it runs, it will on four occasions open a file explorer window. For the first prompt, select the use\_all\_btu.csv file. For the second, select the codes file. For the third, select the Population file. For the fourth, select the Complete\_SEDS.csv file. The code will take care of everything else and write the two output CSVs required for the PowerBI into the same folder the script is housed in.*