Nationwide Visualization of Home Energy Data

Jonathan Cohen, *REU Participant Student, MassBay Community College*
Michael Kane, *Civil & Environmental Engineering, Northeastern University*
Krissy Govertsen, *PhD Student, Northeastern University*

Abstract

<u>Motivation</u>: The format in which data is presented visually to decision makers is incredibly important. The visualization must present the information in an accurate and easy to understand format that can be published by several media outlets.

<u>Challenge</u>: The dataset ABLE Lab utilizes is geographically and time sensitive, meaning it is crucial to develop a method to easily and accurately display both the location and time of characteristics of the dataset to publish our findings.

<u>Vision</u>: Create static and dynamic data visualizations able to convey results from smart thermostat data analysis in a meaningful and easily interpretable way that can be replicated for other analysis and published when necessary.

<u>Goal</u>: Write two MATLAB functions whose inputs are the home identifier and corresponding information to be graphed. The output of these MATLAB functions is either a static plot or a dynamic plot with respect to time.
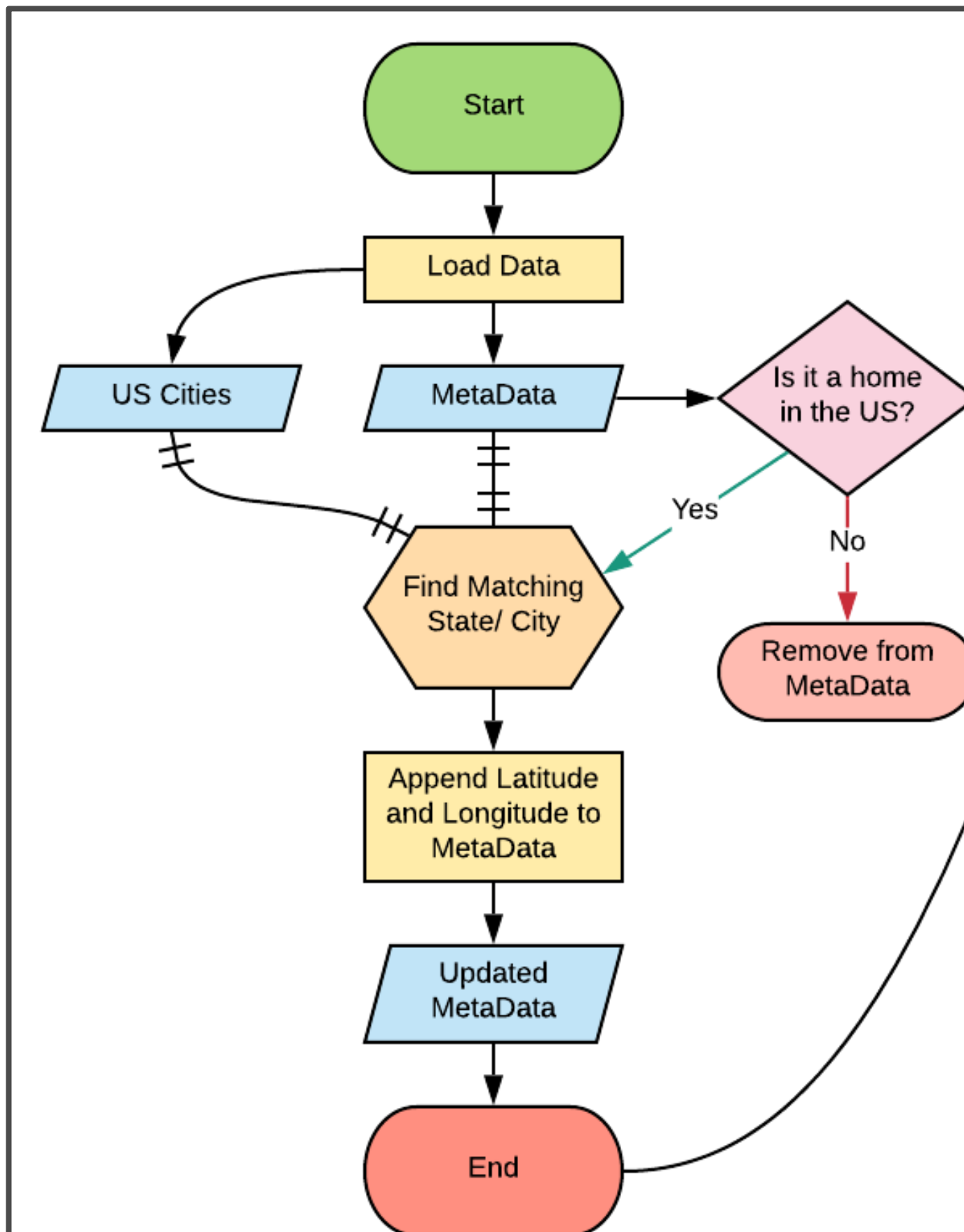
Methods

Figure (1)
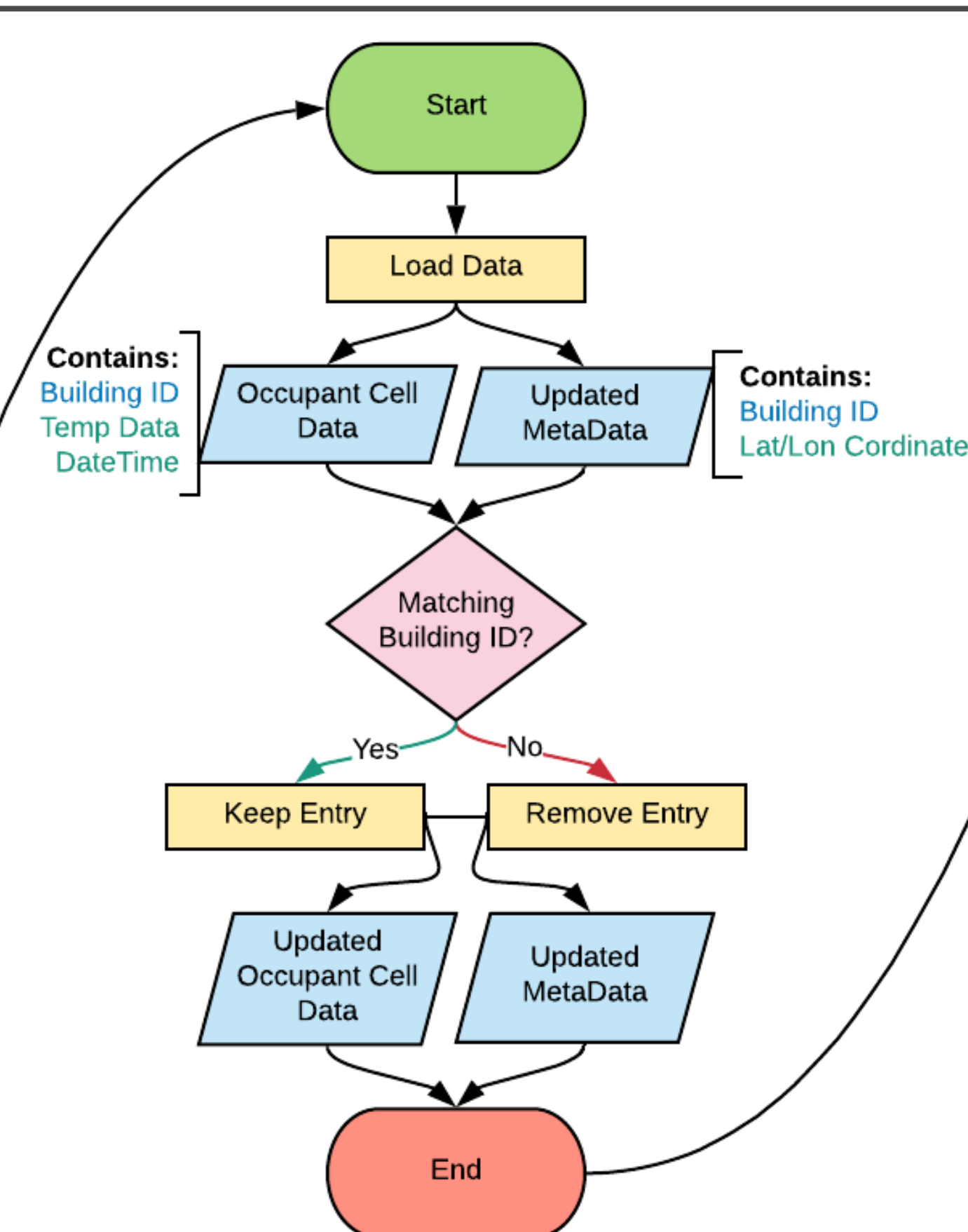Latitude Longitude
Identification FlowChart

Figure (2)
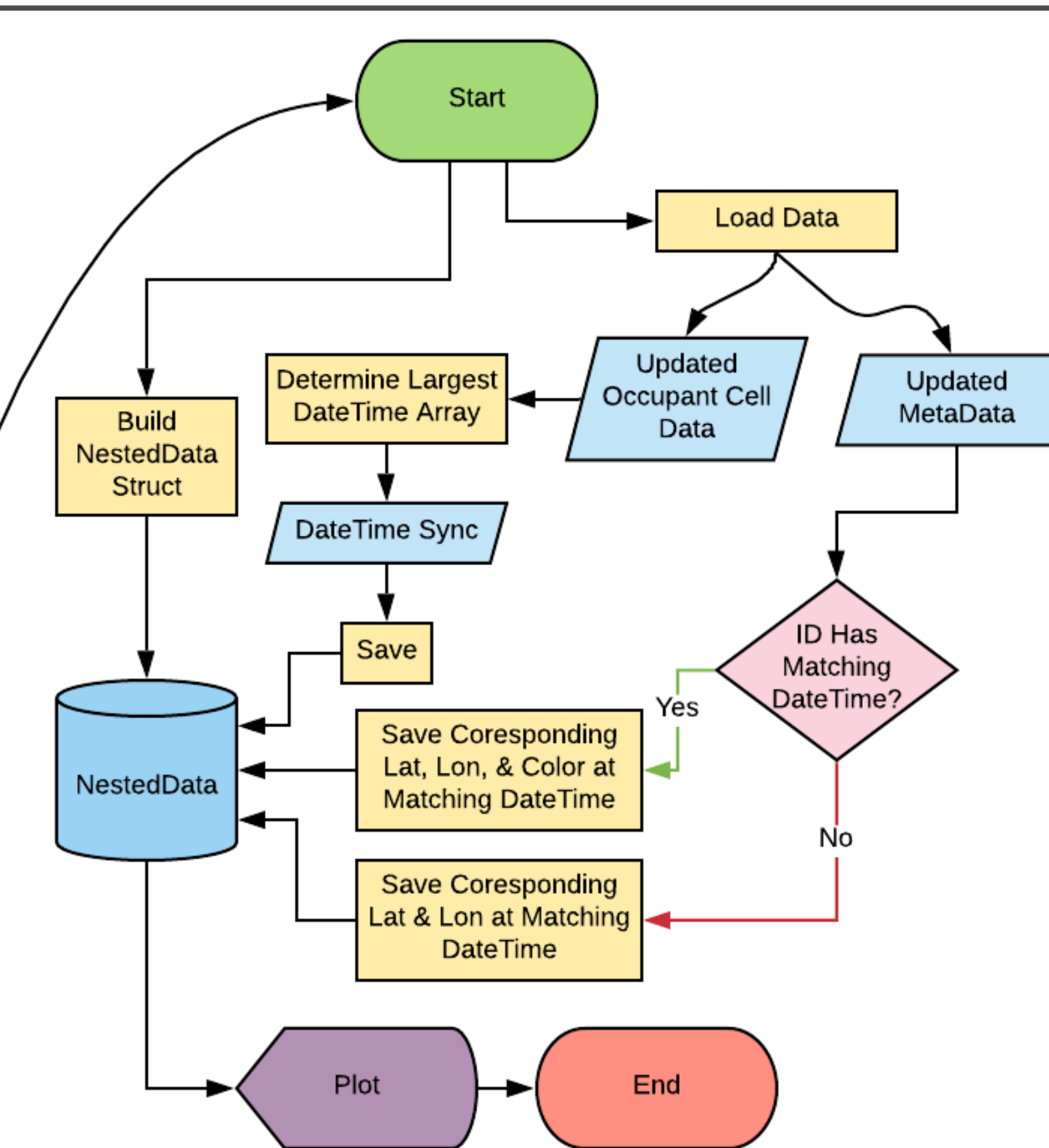MetaData & Occupant Cell
Data filtering FlowChart

Figure (3)
Struct Assembler
FlowChart

Best Practices for Data Visualization:

Keep It Simple

<u>Data ink ratio:</u> Erase everything you can until you do not lose critical information.

<u>Utilize cultural conventions:</u>
- **Red** means stop and **Green** means go.
- **Orange**/ **Red** colors indicate heat.
- **Blues** and **Purples** indicate cold.

Visualization Principles

<u>Gestalt Theory:</u> Group by similarity, by proximity, and by enclosure.

<u>Colors:</u>
- Same kind of data: ▲ = ▲ :Same color
- Contrasting data: ▮ = ● :Contrasting color
- Groups of data: 🏠 = 🏠 :Similar colors

<u>Titles, Labels, Legends and Tables:</u>
- Use bold type only to emphasize something.
- Avoid rotating text & lengthy labels.
- Use a legend only when space is tight.
- Align numbers on the decimal point.
- Never order entities randomly.

Data
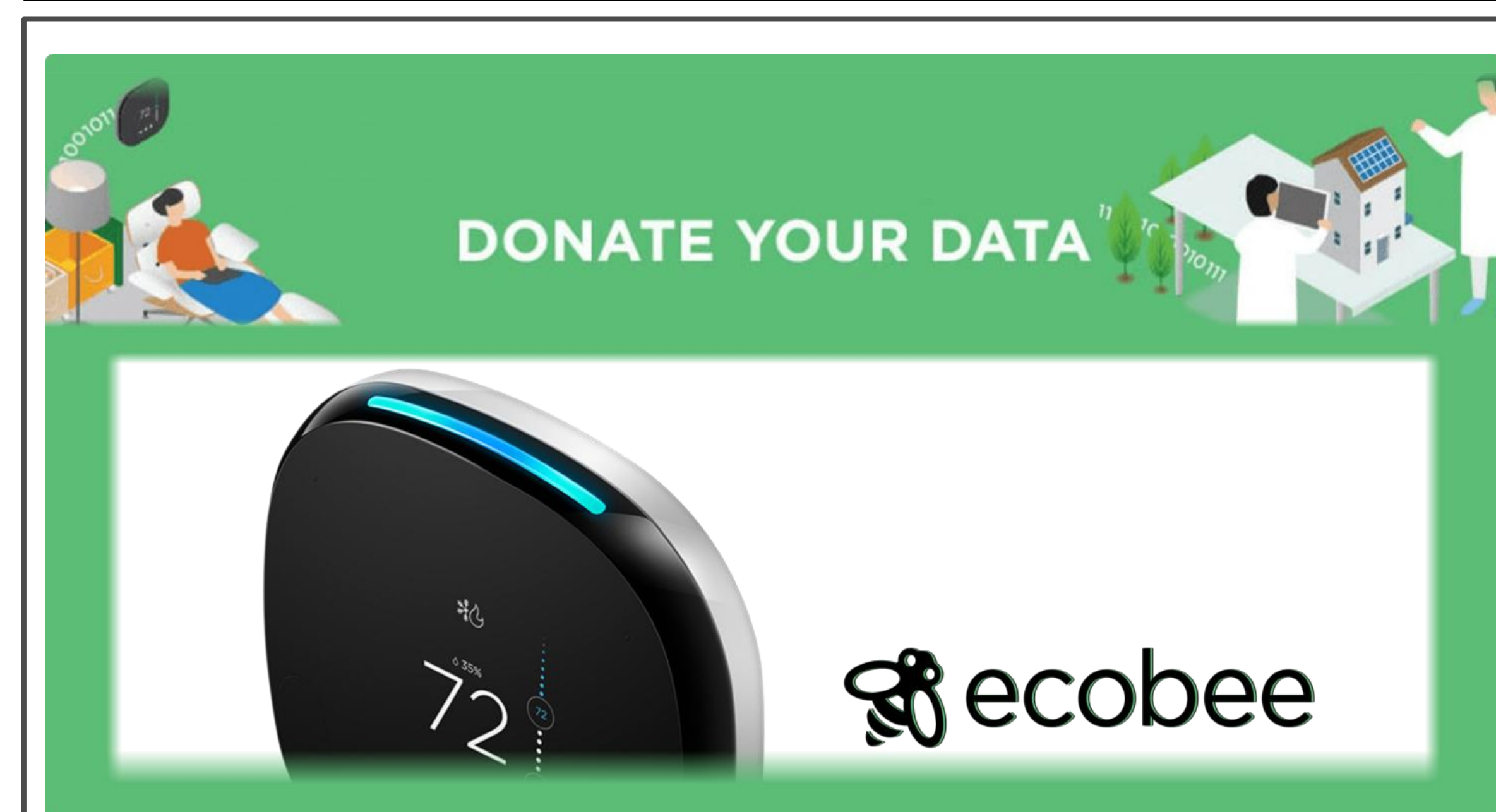
More than **70,000** unique buildings each containing:

Multiple Room Temps.	Room Occupancy
Control Temp.	Events
Setpoints (Heat & Cool)	Equipment Runtime
Outdoor Temp.	Country, State, City
Indoor Temp.	Date Time

Data is collected every 5 minutes at all locations.
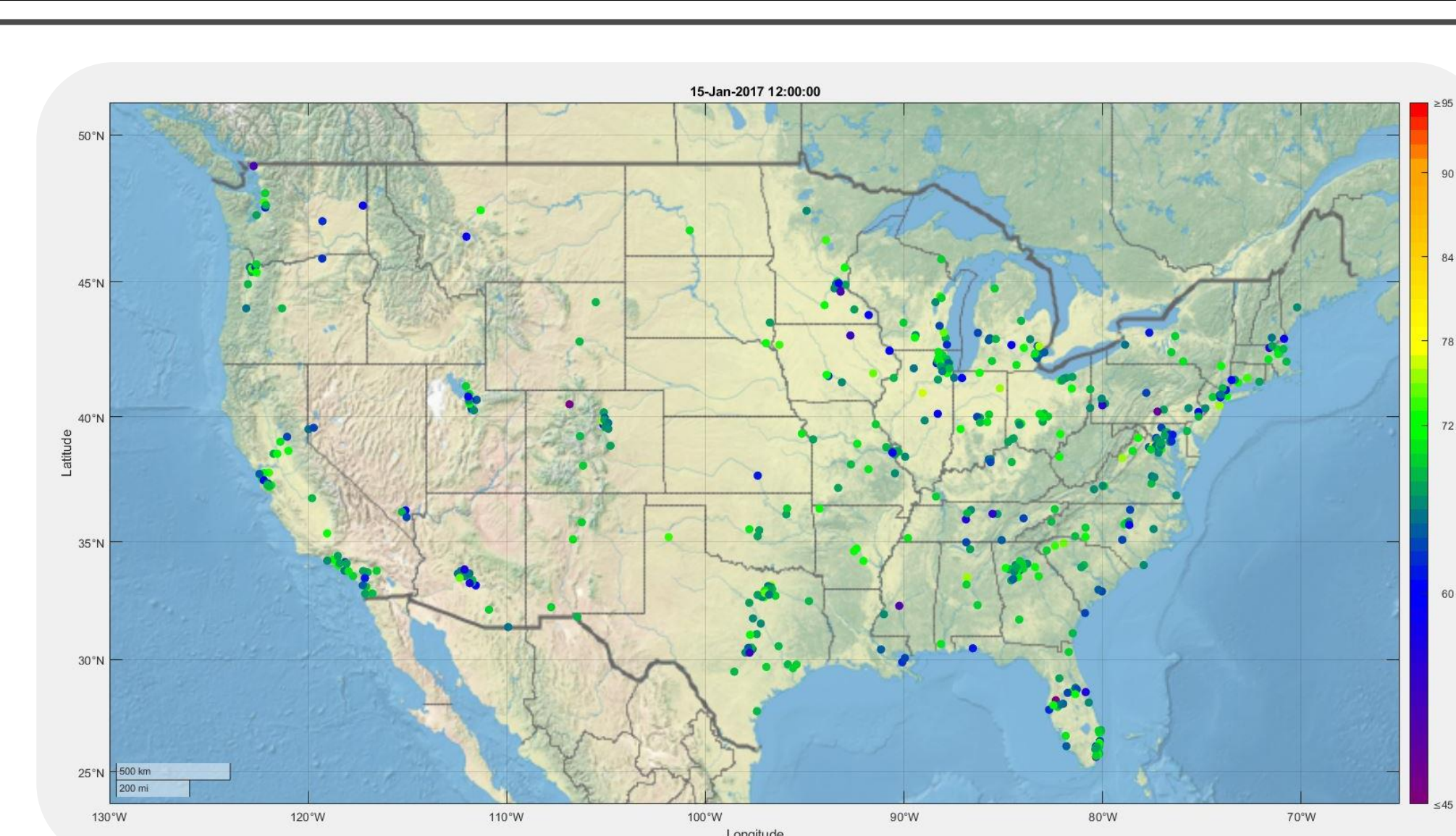
Results

Figure (4 & 5)
Top: Thermostat setpoints at noon January 2017
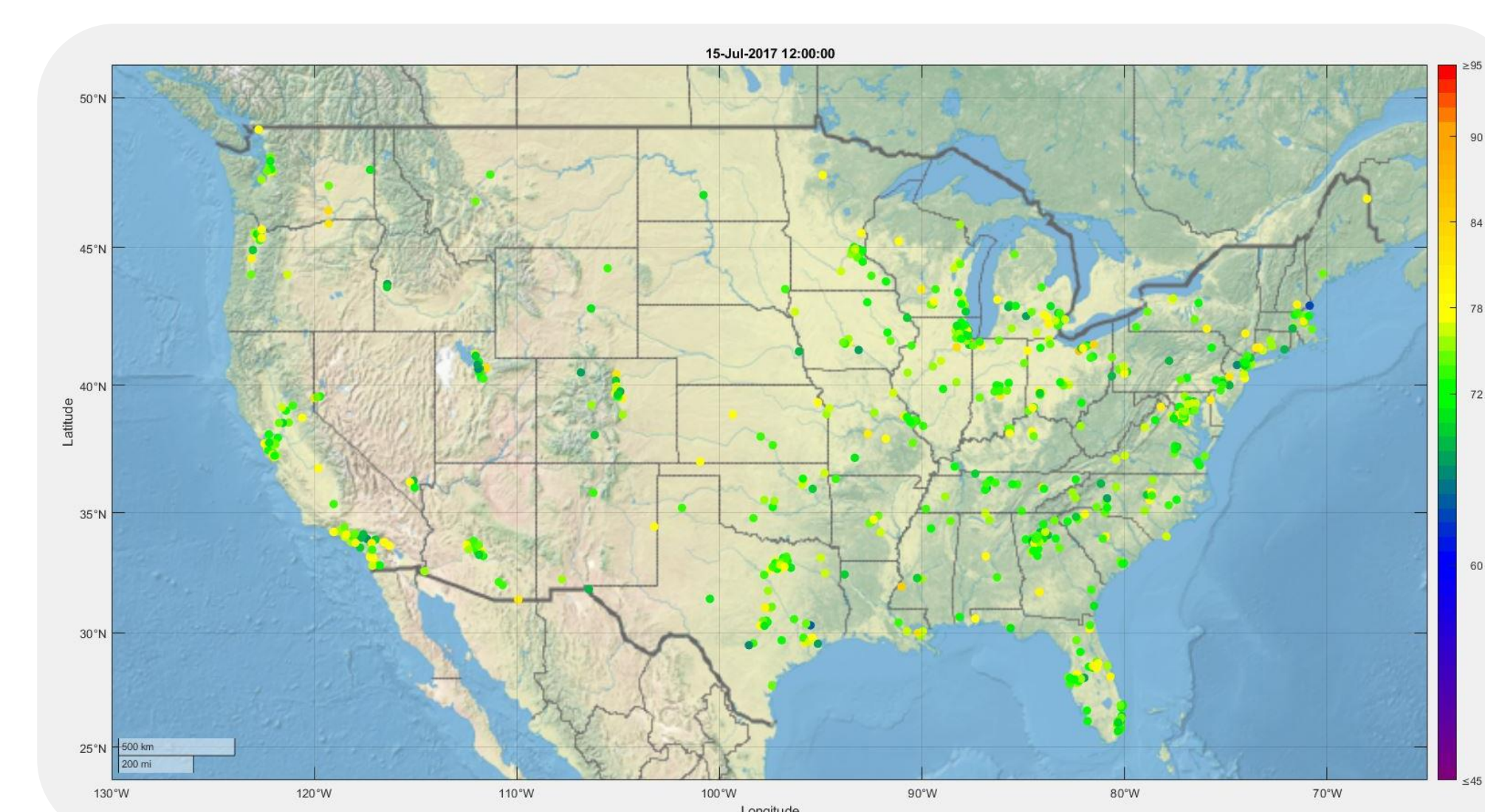Bottom: Thermostat setpoints at noon July 2017

Figure (6)
Capacity of insulating material to
resist heat flow in IL, IN, MI, WI

Conclusion

Accurate data representation of the Ecobee data set used in ABLE Lab required calculated visualization techniques as time and spatial reference can be particularly hard concepts to convey, resulting in the best practices guidelines. These guidelines enabled the development of high quality figures in MATLAB that are both animated and geographically based. The scripts used to create the figures are also scalable, ensuring that as more data becomes available the figures will be easily updated.

Future Research

Geo-Scatter Plotting Upgrades
- Increase the size of cities with multiple thermostats and display average temperature.
- Create custom base maps to reduce to specific geographic locations

Surface Plotting
- displaying the concentration and density of Ecobee thermostats.
- Challenging to bound within the United States.