**Design of Intuitive Visualizations for Residential Heating and Cooling Demands**

I 379C Informatics Capstone

Michael Chen

Host: University of Texas School of Information

Supervisor: Dr. James Howison

Capstone Instructor: Dr. John Neumann

29 April 2024

**I. Executive Summary**

A lack of intuitive visualizations of the loads that a Heating, Ventilation, and Air Conditioning (HVAC) system faces throughout a year may be contributing to the problem of oversized systems. This project aims to develop intuitive time-series level visualizations of the comfort implications of differently sized systems throughout a year. Data was collected through hourly simulation of mock suburban houses in Austin, Texas under a variety of HVAC system sizes. Several types of visualizations were generated, including calendar-shaped heatmaps and daily hour-by-hour snapshots of indoor and outdoor temperatures. Both types of visualizations were used in tandem in order to intuitively display the comfort implications of the indoor conditions of a residential building. This system is currently limited to simulations based on historical weather data and does not consider the implications of latent heat and humidity in comfort. Future development of the project includes generating visualizations that include humidity comfort considerations as well as requests for collaboration and user testing with other previously developed visualization systems and contractors.

**II. Project Overview**

A prevalent issue in residential America is the presence of oversized Heating, Ventilation, and Air Conditioning (HVAC) systems. Indeed, the majority of homes with internet-connected thermostats that donated their data operate at only 50% of maximum capacity during the coldest hours, a potential consequence of oversized systems as well as homeowners adding insulation over time (Meier et al., 2019). Many homeowners and contractors often opt for larger systems in order to avoid discomfort during the hottest and coldest days of the year, but at the consequence of higher upfront cost and risks to humidity and long-term home sustainability (Djunaedy et al., 2011).

A potential source of oversized systems is a lack of intuitive visualizations, and so the primary objective of this project is to create a visualization system of the heating and cooling loads that a home will face throughout the year. In particular, developing a visualization of HVAC loads on a time-series level throughout a year can help inform better understanding and thus decisions regarding equipment sizing from both a homeowner and contractor perspective. This project aims to create a generalizable time-series visualization system based on inputs such as home schematics, local weather and climate conditions, and simulated system loads.

The main body of work for this project are as follows:

* **Background research**, such as into current visualization systems and open-source data availability. Current visualizations have indeed continued to be unintuitive to the average consumer, such as those developed by the Northeast Energy Efficiency Partnership (NEEP, 2021). While providing helpful information for contractors and HVAC experts, there is still an opportunity in providing intuitive visualizations to homeowners lacking prior knowledge.
* **Design of prototype visualizations and sketches**. Following design principles, this project worked within an iterative manner, moving from paper sketches to rough generated visualizations before continuing to enhance visualization detail and quality.
* **Simulation of HVAC loads**. While peak loads for many homes in the US are available through open-source storage, such as from the Open Energy Data Initiative (2022), hourly loads are often sparse or unavailable. As such, this project utilizes hourly energy loads generated by the Building Energy Optimization Tool (BEopt), developed by the National Renewable Energy Laboratory (2024). The data generated through simulation was refined, transformed, and visualized according to the design process, experimenting with alternative ways of representing the time-series data, such as through a calendar-shaped heatmap or daily hour-by-hour visualizations.
* **Reporting and Collaboration**. Having developed a working system, this project is reaching out to relevant stakeholders in the HVAC industry, such as the Northeast Energy Efficiency Partnership, to gauge interest in conducting user testing and/or implementation of the new visualization option.

The work on this project took place within the Spring 2024 semester, ranging from January to May 2024. The work was conducted between the student, Michael Chen, and supervisor, Dr. James Howison. For a more granular timeline, most background research was conducted in February, with prototype visualization sketches and preliminary simulations as well as requests for collaboration being sent out in March. Further refinement of output data and visualization details took place throughout April and May, as well as continuing to reach out and work with interested collaborators.

**III. Relevance of UT iSchool Informatics Courses**

There were four iSchool Informatics courses that were particularly applicable to the work in this capstone project.

**I 305: Research Methods for Informatics.** While this capstone is not a direct research project, the research process described in I 305 has been particularly applicable in the early stages of this project, especially in conducting a review of prior resources and literature as well as developing a value proposition for this project. The course also touched on research ethics and principles when designing research, which acted as a guiding element when looking for open-source data availability and deciding where to source project data from. In the instance of looking for open-source data and working with open-source Python packages, **I 320D Open-Source Software** was also particularly applicable, and the source files for this project have been uploaded to be open source as well.

**I 320D: Data Engineering.** Working with several sources of data, many of which are very large, has required many foundational elements taught in Data Engineering. Data Engineering skills were utilized in differentiating between the different datasets as well as working with data manipulation and transformation. The manipulation of simulation and open-source data in tandem will require several skills that were applicable to Data Engineering, such as general data wrangling as well as linking large datasets. Finally, the final project of the Data Engineering course was a dashboard of several visualizations of a cleaned dataset, which is highly relevant to this project itself as it directly relates towards data visualization and design.

**I 310D: Introduction into Human-Centered Data Science.** Although this project does involve direct work with predictive analysis, it also carries a heavy human element of it through its main goal of making more intuitive visualizations. Creating final products that mesh well with human intuition is a core tenant of the concepts in Human-Centered Data Science, where the interpretability of the products created out of data, whether that be a visualization or a machine learning model, is granted heavy emphasis. Introduction into Human-Centered Data Science also carried with it an introduction into data exploration and choosing data visualization methods, which has proven helpful in the design of the time-series product of this project.

**I 310U: Introduction into User Experience Design.** The principles of design, including using an iterative process that moves from paper to rough wireframes to a more refined process, was a core component of Introduction into User Experience Design that has been extremely helpful in the process of designing the visualizations of this project itself. Creating visualizations that are both usable and explainable to stakeholders regardless of experience is a core part of this project drawn from I 310U. This was particularly relevant in the design of the heatmap/calendar visualization that served as a core component of the final product: a key challenge when working with the time series data was how to represent it clearly on the large time frame of a year, and the design principles taught in I 310U proved to be helpful. Furthermore, if continued collaboration with external actors takes place, continued user testing and interviews will likely be conducted to fully assess the usability of the new visualization system.

**IV. Deliverables**

The main deliverables of this project reside in a [GitHub repository](https://github.com/alchemicHen/hvac-viz) that documents the iterative process. There were four types of deliverables completed:

1. **Initial Research and Design Documents**
   * The beginning of the project included research into prior visualization methods and data source availability. For instance, the visualization system provided by the Northeast Energy Efficiency Partnership served as a good example of existing visualization systems that utilize weather and load data. This served as a stepping stone for moving into time-series scale visualizations that utilized similar data sources. Although there were some open-source historical records of HVAC load data available, large-scale availability was sparse. As such, the project opted towards simulating loads and temperatures based on weather data instead. This research process was documented as a deliverable.
2. **Collected Data and Simulations**
   * Having opted towards utilizing simulated data to create the time-series level scale visualizations, the project opted towards the BEopt systems to generate hourly data. Data was generated based on 2023 weather data for a variety of locations, including mock suburban houses in both Austin and New York. Data was generated for several types of systems, including those with both oversized and undersized HVAC capacities. This enabled visualizations with progressively decreasing system sizes, as displayed through the calendar heatmaps in the final poster. The data generated as well as the simulation files used were stored and uploaded as another deliverable.
3. **Data Cleaning and Visualization Jupyter Notebooks**
   * Having generated simulated data, the Jupyter notebooks were utilized in an iterative process of cleaning the data and generating candidate visualizations. Data cleaning involved managing units and selecting out the critical columns to be then visualized. The data was then transformed into a more usable time series form through manipulation in the Python Pandas package. The visualization process involved iterating on noisy hourly-by-hourly data visualizations, transforming data into day-by-day summaries, and working with a variety of visualization types, including calendar-like heatmaps as well as snapshots of temperatures within a single day. In doing so, the visualization process aimed to create several visualizations that would work in tandem within the poster to create an intuitive understanding of what was happening within the simulated homes throughout the year. The documented Jupyter notebooks that include the full iterative process were uploaded as a deliverable.
   * During the data visualization process, this project utilized the open-source Python package calplot, which was helpful in generating heatmap-like visualizations of a calendar (Kwok, 2024). In line with open-source collaboration, the package was forked with the edited version also being linked as a sub-directory within the deliverable repository.
4. **Reports and Documentation**
   * The generated visualizations, textual explanations of the problem statement and simulations, conclusions, and limitations, were then selected and documented within reports such as this one as well as within a final poster. Although requests for collaboration were sent out, further communication is still pending. Future work may include further collaboration as well as continued development of the current visualizations. All of these prior deliverables are made available at the public GitHub repository.

**V. Contribution to Career Goals**

My career goals reside in looking for an industry experience that fits within my passion for working within the intersection of technology and businesses. As such, this capstone has provided a valuable experience that fits exactly within that intersection.

From a technology standpoint, this project fully utilized my interests in data manipulation and analysis, including the design process. Thinking of candidates for designing visualizations and ways to represent data and then implementing those visualizations through code was a particularly applicable part of the project towards my career goals. This fits inherently within my interest of working within the intersection of people and technology, because the premise of the visualizations is that they should be intuitive to humans. Implementing this through code within both the data transformation and visualization process was especially rewarding. Furthermore, working with open-source packages and collaboration also fits within my professional interests, as I will be continuing to make my non-confidential work open-source and contributing to other open source projects. The utilization of the technologies within this project serve my future career outlooks well by enabling me to work with data in a way that best serves human intuition. I am passionate about communicating conclusions from technical processes in an intuitive manner, which this project has trained well.

Finally, from a business standpoint, this project has many implications towards client and business needs. My professional background and future experiences revolve around technology consulting, and so the element of explaining conclusions in an intuitive manner fits well towards this career path. This is furthered by the fact that the HVAC industry itself is highly client-focused, where being able to sell to clients and explain things intuitively is key to success. Explaining technical conclusions to non-technical clients is key to success when working with clients as well as before starting any engagements, and so this project serves as a good introduction into a specific industry niche where such skills would be necessary. In the consulting industry, client types can vary heavily, and so getting exposure into having to research and quickly become familiar with and industry before moving into developing a solution has proved a helpful contribution into my likely future of working with clients.

**References**

Chen, M. (2024). hvac-viz*.* *GitHub* [Computer Software] <https://github.com/alchemicHen/hvac-viz>

Djunaedy, E., van den Wymelenberg, K., Acker, B., & Thimmana, H. (2011). Oversizing of HVAC system: Signatures and penalties. *Energy and Buildings*, *43*(2–3), 468–475. <https://doi.org/10.1016/j.enbuild.2010.10.011>

Kwok, T. (2024). Calplot (Version 0.1.7.5) [Computer software] [DOI: 10.5281/zenodo.11074893](https://doi.org/10.5281/zenodo.11074893)

Meier, A., Rainer, L., Daken, A., Ueno, T., Pritoni, M., & Baldewicz, D. (2019). *What can connected thermostats tell us about American heating and cooling habits?*

National Renewable Energy Laboratory. (2024). *BEopt: Building Energy Optimization Tool.* <https://www.nrel.gov/buildings/beopt.html>

NEEP’s Cold Climate Air Source Heat Pump List. (2021). *Cold Climate Heat Pump Sizing Support Tools*. <https://ashp-production.s3.amazonaws.com/NEEP_ccASHP+Heating+Visualization+User+Guide_v2.2_TRC_04.01.22.pdf>

Open Energy Data Initiative. (2022). *NREL PDS Building Stock*. <https://data.openei.org/s3_viewer?bucket=oedi-data-lake&prefix=nrel-pds-building-stock%2Fend-use-load-profiles-for-us-building-stock%2F2022%2Fresstock_amy2018_release_1.1%2Fmetadata_and_annual_results%2F>