footprints: Visualizing Energy Use in Commercial Buildings in the U.S.

JinYoung Baik, Prateek Kakirwar, Janani Vasudev, Gopal Vaswani

Need

Energy is undoubtedly one of the most pressing issues of our time. Research shows that buildings account for nearly half of the energy (41%) consumed in the United States, far more than industry (31%) or even transportation (28%). Recognizing the dark side of the growing ecological footprint of this sector, the federal government has unleashed aggressive energy policies to curtail resource use and improve the energy performance of buildings. Nevertheless, awareness of energy consumption and associated costs is still limited partly because the data is hidden in large and complex datasets that are unfathomable to the general public. In the recent years, we are seeing new information visualization methods and tools being applied in business, science and academia for understanding trends and relationships in these data sets. These innovations provide users with interactive capability for filtering, sorting, and visualizing information, taking advantage of the powerful “bandwidth of human vision."

Goals

- Support exploration of ecological *footprints* of commercial buildings, in terms of energy intensity and expenditures in dollars, along various dimensions (per sq.ft, per occupant, per building, etc.)
- Allow comparisons of the historical energy profiles of different building types across U.S.

- Use the visualization as a basis to identify patterns in energy use and tell interesting stories of relationships (if any) between energy consumption, expenditures and locations of buildings

Users

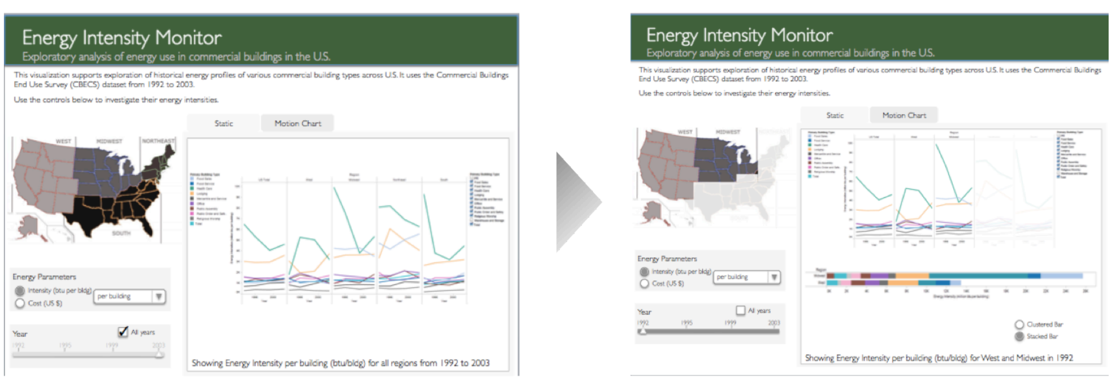
The broad audience for our visualization include anyone that has an interest in energy monitoring and analysis. More specifically, this includes energy researchers, environmentalists, architects, building managers, and potentially, even policy makers.

Data source

footprints is an exploratory visualization to track the energy performance in commercial buildings across the United States. It is built on data from the Commercial Building End-Use Survey or the CBECS database that is a survey of existing building stock conducted once in every four years by the U.S. Department of Energy (DOE). The CBECS defines commercial buildings to include those that have at least one half of their floor space used for activities that are not residential, industrial, and agricultural. Consequently, they include building types that might not traditionally be considered "commercial," such as schools, correctional institutions, and buildings used for religious worship. Even while the CBECS contains useful information on the building energy sources and consumption, energy-using equipment and expenditures in over 5000 commercial buildings in the U.S. sorted by region and space type, their availability as purely raw energy information can be overwhelming to digest and use effectively. Since the human visual system is a pattern seeker of enormous power and subtlety, graphical visualizations of this dataset provide greater opportunities to spot trends and reveal relationships that are otherwise obscured in tabular representations.

Conceptual Design

Using CBECS data between 1992 and 2003, this visualization supports visual discovery of energy performance, in terms of quantity and dollar costs, of various types of commercial buildings in the four main census divisions in the U.S. Available filters for display include unit energy intensity and costs for a specific building, square foot area and occupant. The core design paradigm of this visualization is to encourage 'learning by exploration.' The interactions supported align with Schneiderman's design philosophy for visualizations providing an “overview first, zoom and filter, then details-on-demand.” At any time, the user drives the interaction and retains complete control of what is being displayed. Each visualization offers different perspectives on the same dataset and by doing so, captures subtleties that make the discovery process more meaningful.

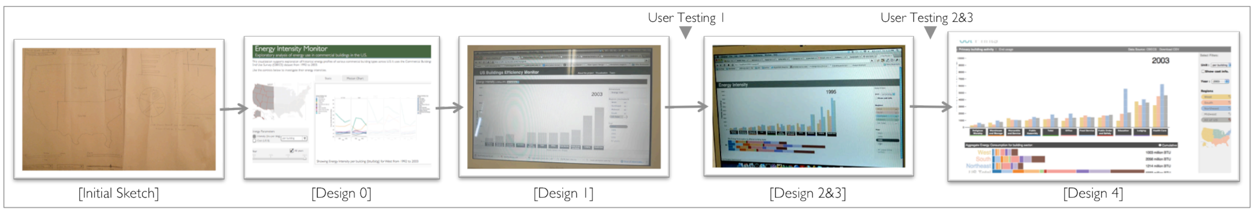


[Overview of Energy Intensity per building for all regions from 1992 to 2003]   [Details-on-demand: Energy Intensity per building for West and Midwest in 1992]

[Figure 1] Conceptual design

Design Timeline

[Figure 2] is an overview of our iterative design process. As this trail shows, we started out with rough paper sketches of our design. We iterated this to a low-fidelity digital prototype (Design 0), that was further iterated to a higher fidelity (Design 1) based on the feedback received during the midterm presentation. This higher-fidelity prototype went through a series of iterations (Designs 2,3) based on results from three user tests resulting in additional functionality and GUI tweaks. Our current design is at version 4 (Design 4).



[Figure 2] Design Timeline

Implementation

- Data organization : MS Excel
- Data visualization : Protovis is a visualization toolkit for JavaScript using SVG. It takes a graphical approach to data visualization, composing custom views of data with simple graphical primitives like bars and dots. These primitives are called marks, and each mark encodes data visually through dynamic properties such as color and position. (http://vis.stanford.edu/protovis/)
- Web application : Ruby on Rails

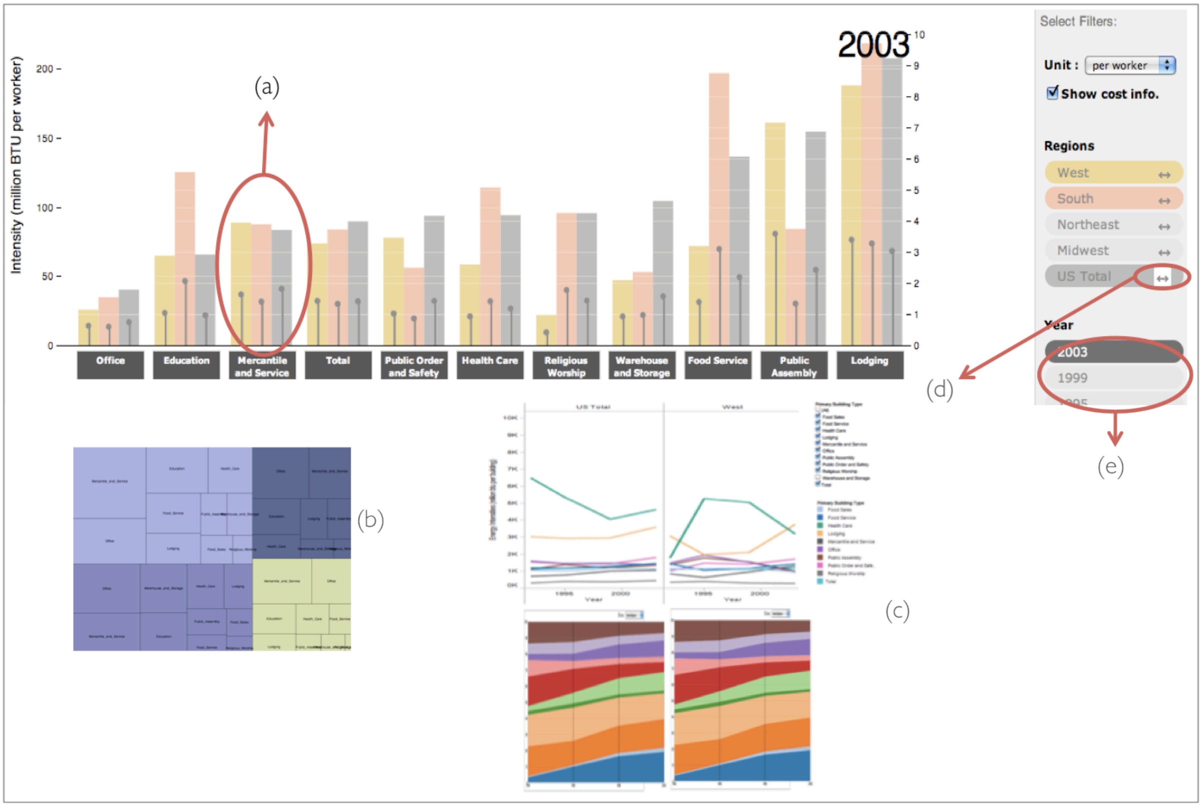
Design Evaluation

Usability tests

We conducted three user tests, each 45 minutes in length, with researchers in energy labs in campus, including the Energy Resources Group and the Center for the Built Environment. The goal of our user tests was to understand how well *footprints* afforded exploratory interactions and to evaluate both its usability and usefulness. Instead of giving our subjects specific tasks, we let them play around with the interface and explore the visualizations. As they used the interface, we asked them to think-aloud to get a more contextual sense of their thought process and learn what they liked or disliked in the interface. In most cases, we let the subjects figure out ways to recover from problems and other points of confusion, helping them out only when they came to a complete roadblock. Following the test, we also ran a quick exit survey that asked about *footprints*' effectiveness in providing the information they were looking for, points of confusion, their likes and dislikes and additional information that might be useful to include in a tool like *footprints*. In this section, we summarize out findings from these evaluations and our design iterations.

User Test #1

Design [1]



[Figure 3] Design [1]

(a)

- Default selections are energy intensity per building in all of U.S. for 2003
- Separated annual energy intensity visualizations from historical energy profile
- Distinguished regions by colors
- Cost overlays on energy intensity data

(b) Treemap to show the contribution of energy use in each building type

(c) (When a user selects 'all years' in the year section) display trend lines of energy intensity data and stacked graph

(d) Allow users to sort by value

(e) Support toggling between years as opposed to sliding

User Profile

[Figure 4] User 1 exploring *footprints*

Graduate Student, Energy Resources Group, UC Berkeley: Our first subject was first year Master's student at the Energy Resources Group in UC Berkeley. She works with energy data on a regular basis and is knowledgeable about the energy use patterns and trends.

Findings

Energy costs overlays work well

One aspect of our visualization that worked well for this user were the overlays of energy cost information on the bars representing the energy intensities. Having separate visualizations for annual and trend energy intensity data and the ability to toggle between them were useful to her.

Treemap not effective for comparison

This user pointed out that the absence of order in the individual blocks of data in treemaps made them confusing to those who didn't know how they worked. In addition, while treemaps are useful in displaying hierarchical data, they are not so effective in supporting comparisons of these datapoints. She recommended using stacked bars instead to represent the same information.

Map would add more contextual information

While toggling between the regions with multi-select buttons was intuitive to this user, she pointed out that having a map would help a reader immediately associate a region with its geographical location. Also, it would make it easier to immediately identify the states that compose a particular census division.

No meaning in the order of data points in Stacked graph

The order of data points in the stacked graph appeared somewhat random to this user. Why this specific order, if the data points are not sorted by their values?

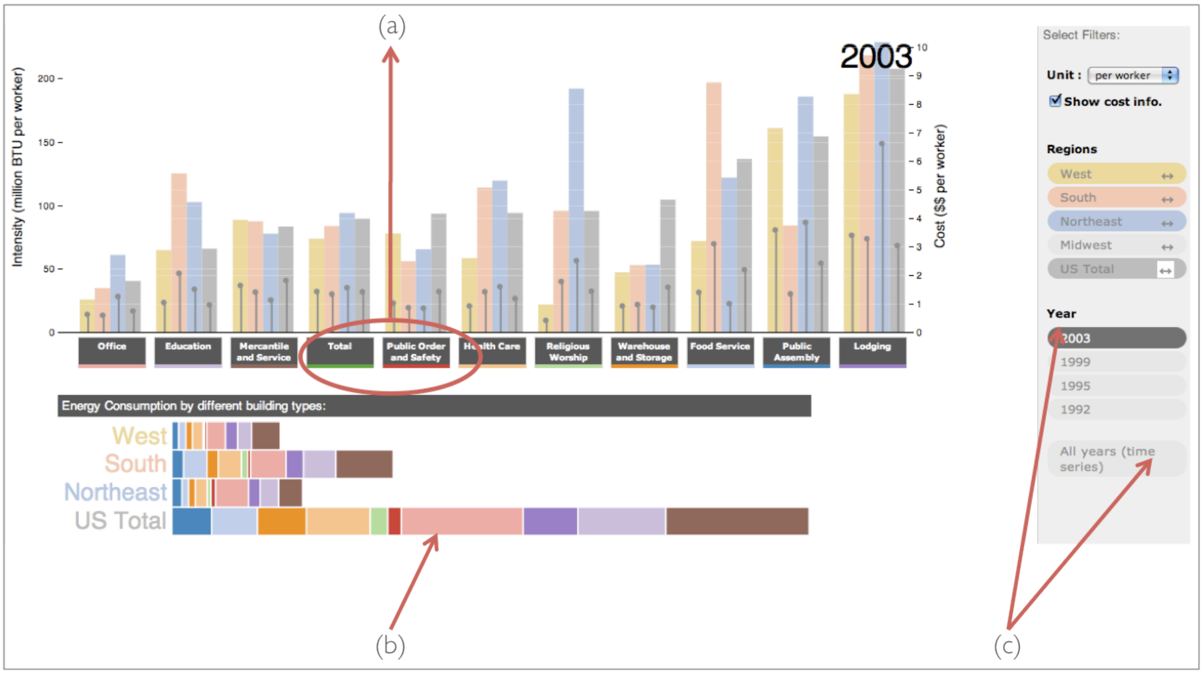
UI ambiguities

The ‘Sort by’ icon was unclear to this user. Labels such as ‘All years’ (for trend data) and ‘U.S. Total’ (a selection of U.S Total data point, NOT all regions)
were misleading.

This user also mentioned that since the trend and annual visualizations were completely different types of visualizations, it was important to visually separate the controls that operate them. Since the buttons to toggle between the visualizations were not visually distinct and were in close proximity, she had expected to see them produce similar 'looking' results. However, she was surprised to see how they different these visualizations actually were.

User Test #2

Design [2]



[Figure 5] Design [2]

(a) Added color strips to labels in x-axis to make relating building types in the bar chart easier with the stacked bar

(b) Replaced Treemap to Stacked bar to improve the effectiveness of comparisons

(c) Visual separation of controls for historical trends and yearly energy intensities

User Profile

Ph.D Student, Energy Resource Group, UC Berkeley: Our second participant was a first year PHD student at the Energy Resource Group at Berkeley. She does extensive energy research and is interested in exploring how energy consumption and expenditures have changed over the years.

[Figure 6] User 2 exploring *footprints*

Findings

Support ability to download data as a CSV/ .xls format

Disclaimer to note erroneous data points

Since some of the original data points were missing in the original CBECS data set, we had to extrapolate them to ensure that our visualization still worked for these cases. Because of this, the data for some building types appeared erroneous. Including a disclaimer not only informs the reader that we are aware of these errors but also warns him or her to not be completely trusting of the data.

Support more interactions

Allow interaction with individual data points to get actual values and additional information such as number of buildings, etc. This user also suggested that it might be useful to support a 'click and lock' functionality that selects a particular and track how it has changed across dimensions over the years.

Maps for more context

Once again, which the multi-select buttons for selecting regions was intuitive, this user pointed out that maps can still be useful in adding more geographical context to the data.

UI ambiguities

This user did not notice the 'sort' function until we told her about it. She wanted the functionality to be made more 'visible' by the use of clearer and less ambiguous icon.

User Test #3

Design [3]

[Figure 7] Design [3]

(a)

- Allow users to download CSV data file
- Added data source
- Added disclaimer
- Added units and actual value of energy consumption and # of buildings

(b) (When a user checks 'Cumulative' checkbox) display stacked bar to show annual cumulative energy consumption of buildings

(c) (When a user selects 'all years' in the year section and chooses 'Cumulative consumption over years') display stacked graph to show historical cumulative energy consumption

User Profile

[Figure 8] User 3 exploring *footprints*

Director of Communications, Center for the Built Environment, UC Berkeley: Our third participant was the Director of Communications at the Center for the Built Environment, UC Berkeley.

Findings

Support more interactions

- Support ability to drill down to get detailed information on specific data points

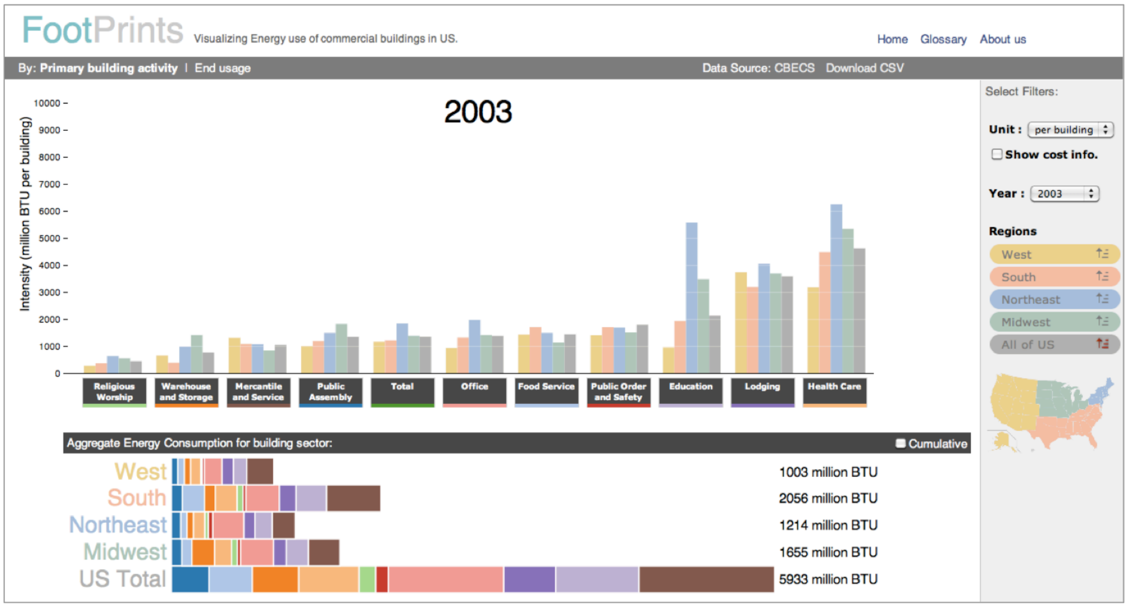
Include more filtering capabilities for selection of primary building type

- This user pointed out that including an “All” check box to turn the selection of all buildings on or off is useful particularly when the user wants to compare the performance of only a few building types. This is easier to use than a default selection of all buildings, in which case the user has to manually deselect from the long list of building types.

Maps for more context

Yet again, while multi-select buttons for selecting regions was intuitive, this user found maps to be useful in adding more geographical context to the data.

Design [4] : Final iteration



[Figure 9] Design [4]

- Added Glossary I About us section
- Added a map with colors to give more contextual information to users
- Changed Sorting option icon to be more intuitive

Challenges

- One significant limitation of this dataset is the unavailability of data on specific buildings of each state. Because we only have aggregate building data only based on the U.S. census divisions (West, Midwest, South, and Northeast), multivariate data analysis is nearly impossible.
- Some data entries and tables were missing in the original CBECS database. To make the visualization appear consistent and complete, we had to extrapolate the data for the missing entries. In some cases, our extrapolations created misleading data points that struck our users as odd.

- Another challenge was limited functionalities of Tableau Public. Initially, we considered creating a mashup that was a combination of visualizations generated using Tableau Public and Protovis. However, as our design progressed, we found ourselves being constricted in terms of what we could show because of Tableau's limited features. Some of the However, since Tableau doesn't support some of dynamic graphs (for example, a dynamic map without having latitude and longitude information, treemap and plotting two y-axis dimensions data in one graph), and linking visualizations created from Tableau Public onto website has limited interaction with users, we have decided to use Protovis only for visualization.
- Ambiguity in the use of colors. Our first prototype included 27 different colors for each dimension: regions, primary building type, and end use. Using a lot of colors, we reckoned would confuse users. To deal with color better, we used light pastel colors for regions in the annual energy intensity graphs and brighter colors for the primary building types in the stacked bars.
- End-use data was available for specific regions and primary building types, but not together. Because it was not along the same dimensions as the data for energy intensity per.worker, per building or per sq.ft. data, we had to create a different style of visualization for end-use data.
- Implementation issues with using a map to drive interactions. We also ran into a risk of ambiguous interactions in using a map - for example, differentiating between selection of all regions (northeast, south, midwest and west) and U.S Total which is a separate datapoint in itself. If the user selected all the regions,

would that mean he or she selected all regions or U.S. Total (which logically is a selection of all regions)?

- Supporting "details on demand" was problematic because of limited data in varying dimensions. In other words, we did not have multivariate data, we were stumped on what to display on interactions.
- Data spread over 200 tables, and to get comparable and homogenous data we had to scrape through a large database
- Although protovis provides a low level control of the visual elements in the visualization, it has a significant learning curve to understand and imbibe how it renders the graphics.

Future Work

While we would have liked to incorporate all the feedback we received from our users, there were some we could not because of time limitations and unavailability of data. These are potential areas of future work. These include:

- *Providing comparisons of energy consumed from thermal and hydro sources, carbon multipliers, etc.* One user pointed out that it might be useful to include visualizations that track the energy performance of different sources, such as thermal or hydro particularly because the sources of energy is diverse for different regions.
- *Investigate the possibility of supporting more data exploration/interaction with current data*. We want to be able to provide more interaction with existing set of data. Features like displaying the actual values for energy intensity, energy cost,

and energy consumption for a particular region and year and the ability to drill down for more detailed information on the datapoint were not implemented partly because of poor availability of data and limited time.

- *Explore multivariate analysis by merging data from different tables.* Merging data from different tables will involve extensive data mining and in some cases even extrapolation to account for missing data. Nevertheless, such visualizations can be useful. Some additional dimensions can be year of building construction, number of operating hours, climate zone, etc.

Conclusion

We were successfully able to design an interactive visualization tool which supports exploration of ecological *footprints* of commercial buildings along different dimensions. Our users were able to use *footprints* to detect trends in energy intensity and expenditures and spot outliers. Some visualizations confirmed hypotheses such as higher energy consumption in places of religious worship in the South as opposed to the Northeast reflect the more 'religious' and 'conservative' character of these states. On the other hand, our visualization was also able to show the decreasing energy intensities of Healthcare buildings raising the possibility of aggressive energy policies and more energy efficient equipment. Outliers were also spotted in the extremely high energy intensities per worker in the Northeast for the Lodging sector. Because large amounts of raw data can be overwhelming and difficult to digest, we are confident that *footprints* has presented them in a more accessible fashion, while still making the process of discovery of interesting trends and relationships more meaningful.