Cybersensitive Electricity Consumption Patterns

Produced for the California Energy Commission by Indicia Consulting in fulfillment of Task 5 Deliverable 1 of the project Cybernetic Research across California: Documenting Technological Adoption and Behavior Change across Diverse Geographies and Populations to Inform Energy Efficiency Program Design Funded by EPIC PON 14-306

January 31, 2018

Contents

Abstract

This report on Cybersensitive Electricity Consumption Patterns describes the difference in behaviors and outcomes experienced by households participating in our research in terms of their electricity consumption. This report uses electricity consumption data from Pacific Gas and Electric (PG&E) and Southern California Edison (SCE) to identify patterns of consumption among populations of California electricity consumers. The data sets include electricity consumption by specific households in this study, as well as aggregated electricity consumption data for zip codes in Marin County and the City of Long Beach, the same areas where the specific households in our study are located. We also compared our data with data collected by the California Public Utility Commission for California Building Climate Zones 2 and 3.

This report analyzes the average monthly and annual electricity use for PG&E and SCE customers who participated in our research. It analyzes whether there is a statistically significant monthly electricity usage difference between participants identified previously as cybersensitive and non-cybersensitive. We found that cybersensitives use less electricity than other consumers. There was a statistically significant difference between cybersensitives and non-cybersensitives in their average monthly and annual electricity usage.

This report also compares electricity usage between participants and non-participants in the same communities to understand how the participants energy consumption compares to the general population. Cybersensitives in the PG&E have a statistically significant difference in consumption from the general population.

1.Introduction

1.1. Goals

Our goals for Task 5 were to:

- Organize and analyze a quantitative dataset consisting of electricity usage data.
- Place cybersensitive participation and performance against a representative sample of local electricity consumers' data to establish the delta of electricity consumption and responsiveness.
- Measure cybersensitive participation and performance in past energy efficiency programs using partner utility electricity data.
- Set ethnographic data alongside larger numerical data sets.
- Identify a consistently repeating set of characteristics, including behavioral, demographic, and energy usage, which can be attributed to the cybersensitive profile.

1.2. Report structure

We begin by revisiting our definition of cybersensitivity, cybersensitives potential for saving energy, and the segment classifications of our participants We then discuss our methods for achieving the goals for this project, the results we found, and what the implications of these results are in the remainder of this report.

Establishing the distinction between the populations of the two utility territories, for instance, supports our thesis that cybersensitivity, since it shows up in both territories, is independent of geography. We discuss the impact of a single outlier on small sample sizes and what happened when we performed the analysis with the outlier included versus removing it. We discuss the role of size in our samples, and why we were able to draw statistically valid conclusions from the larger (and more complete) PG&E sample size. Comparing our dataset to that of the CPUC helped us to analyze in more detail how to climate and geography influence patterns of electricity consumption, as well as set our results against those of other studies for a potential narrative of reliability.

We demonstrate that we achieved the goals set out above for Task 5. We acquired, organized, and analyzed four quantitative data sets consisting of electricity usage data. We placed cybersensitive (and non-cybersensitive) participation and performance against a representative sample of local electricity consumers' data to establish the delta of electricity consumption and responsiveness. We demonstrated a consistent pattern of a delta existing between cybersensitive electricity usage when set against the general population for their local utility (at zip code level), or the sample from the CPUC. We acquired utility electricity data for our participating households and used a 36-month sample to measure cybersensitives against other members of their cohort. We established that non-cybersensitives are a reasonable match to the general population for the respective utilities, and that they are also a reasonable match to the sample used by the CPUC.

We set ethnographic data alongside larger numerical sets by first establishing the validity of our segmentation statistically, and then reviewing the interview transcripts, setting cybersensitive responses alongside non-cybersensitive responses to establish patterns. Finally, we identified a consistently repeating set of characteristics, including behavioral, demographic, and energy usage, attributable to the cybersensitive profile.

2.Background

In this section we revisit our definition of cybersensitivity, cybersensitives' potential for saving energy, and review the segment classifications of our participating households. We then discuss our methods for achieving the goals for this project, the results we found, and what the implications of these results are in the remainder of this report.

2.1. Identifying a new personality trait (cybersensitivity)

The original PON for this project asked, "What are the principal factors among the wide range of cultures and behaviors in California that explain diversity of energy efficiency measure uptake, as compared to purely economic formulations?"

Our hypothesis, based upon an extensive review of energy efficiency and behavior literature (including several pilot projects conducted by utilities) is that cybersensitivity is a personality trait which distinguishes its possessor from other members of demographic cohorts such as age, gender, and income strata, as discussed in our Task 2 report, "Preliminary Ethnographic Report on Cybersensitives and Technology Detailing the Fieldwork and Early Findings" (Indicia Consulting, 2016).

2.1.1. Review of cybersensitive definition and characteristics

We previously defined cybersensitivity in our Task 4 report, "Cybersensitive Response to Technology" (Indicia Consulting, 2017b). Cybersensitivity is a propensity to have a greater emotional engagement with personal (mostly handheld) technology, such as a smartphone or tablet. Using the combined survey/interview data set, we statistically validated our hypothesis that cybersensitivity does not correlate with the demographic variables of age, gender, or income. Nor does cybersensitivity correlate with a specific geography in our sample.

2.1.2. Cybersensitives and energy saving potential

Cybersensitives, though a relatively small proportion of the population—10 to 15% of the population was our assumption at the outset of this research—have larger than average energy savings rates in response to a variety of interventions tested across several piloted energy efficiency programs (8-12%) (ACEEE, 2013). This response rate is several times higher than the average aggregated energy savings (1-2%) delivered through established feedback-based behavior change programs such as that offered by Opower (Alcott, 2011). Looking beyond cybersensitives, we also found a related segment we call 'cyberawares' which includes 10-15% of the population. They have double or triple the energy savings (4-8%)[1] when compared to the results listed by Opower. Our fieldwork observations have led us to conclude that cybersensitives are interested in more ambitious and innovative energy efficiency measures, while cyberawares appear to be interested in tools for tracking energy consumption and savings.

2.1.3. Segments of cybersensitivity

In prior reports[2] we showed how we used the data from ethnographer observations and the Atlas.ti coding of transcribed interviews to identify ways in which interviewees interact with technology in their everyday lives and to determine the psychological drivers behind that usage. Using Atlas.ti to code the data allowed us to quantify behavioral observations and run frequency counts independently of the subjective perceptions of the ethnographers. The resultant clustering of cybersensitives, cyberawares, nulls, and the group in between, who we think of as 'mainstream' showed up strongly in these code frequency counts. In In the report, "Cybersensitive Response to Technology," (Indicia Consulting, 2017b) we showed that the relationships between code frequency and distribution and cybersensitive status held up across multiple operations and was also demonstrably statistically valid. Therefore, we consider our identification of a set of segments to be well grounded. Those segments are:

- Cybersensitives
- Cyberawares
- Mainstream
- Low Mainstream
- Null

[2] In both the Task 3 Report, "Psychosocial Drivers of Behavior," (Indicia Consulting, 2017a) and the Task 4 Report, "Cybersensitive Response to Technology" (Indicia Consulting, 2017b).

We also reduced the five segments into two contrasting segments: cybersensitives and 'non-cybersensitives'. Cybersensitives being the combined set of cybersensitives and cyberawares, and non-cybers comprised of the Mainstream, Low Mainstream, and Null segments.

3.Methods

In this section of this report, we discuss our sample, their distribution across the two utility territories, and their membership in one of two segments, cybersensitives and non-cybersensitives. We then discuss the four sets of electricity data we analyzed for this report:

- 12-36 months of electricity data for 23 consenting participating households
- Anonymized residential customer data by zip code for Marin County (PG&E Public Data Sets: Electric Usage by Zip)
- Anonymized residential customer data by zip code for City of Long Beach (SCE Quarterly Customer Data Reports)
- California Public Utilities Commission (CPUC) 2016 Residential Electric Usage and Bill Statistics by Climate Zone

In the Methods section, we also outline the statistical tests used in our analysis and share tables with output for comparing usage monthly, annually, by cybersensitive status, by utility, by climate zone, and season. We share several histograms showing the 'normal' distribution of energy usage, which was a necessary pre-condition for being able to conduct t-tests[3]. We also compare our participants to their respective general populations. We conclude the Methods section with our methodology for analyzing qualitative data drawn from the in-depth interviews and survey answers given by our participants. We define 'quantifying behavioral observation,' revisit the coding of ethnographic data, and discuss the statistical operations previously carried out on the coded data. We also reviewed the original transcripts for this Task, mining them for examples of cybersensitive behavior in the arena of energy efficiency programs.

3.1. Sample

Our study interviewed 45 households. While we originally had signatures on forms consenting to the acquisition and use of energy data for all households, we ran into administrative complications with the utilities who required use of their consent forms. We therefore had to follow-up and request the households to complete the new utility consent forms several months later. This led to a significant reduction in the number of consenting households—we received only 23 new consent forms. We have *some* electricity data for fourteen households in Pacific Gas & Electric (PG&E) territory, and nine households in Southern California Edison (SCE) territory. Data received was inconsistent across the sample: For PG&E only 12 households had 36 months of data for the analysis. For SCE only 4 households had 36 months of data for the analysis. We were unable to get an answer from the utilities for

[3] "A t-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistics (under certain conditions) follow a Student's t distribution. The t-test can be used, for example, to determine if two sets of data are significantly different from each other." (Wikipedia, 2018a).

the different datasets delivered (12 months vs. 36 months for some customers). We only ended up using the customers with 36 months of usage in our sample.

Table 1. Distribution of *the cyber and non-cyber households that were analyzed* across *the PG&E and SCE service* territories.

	Cybersensitive	**Non-Cybersensitive**
PG&E	4	8
SCE	2	2

3.2. Electricity usage data

Recognizing that electricity data acquisition was a critical component for our project, we initially approached the three California investor owned utilities (IOUs) in November 2014 to be partners on the project. All three ultimately passed on collaborating with us. Without preexisting utility agreements in place, we collected consents from our participants and reached out directly to PG&E and SCE to request the household electricity consumption data.

As of this writing, we have acquired electricity consumption data for 23 of our participating households in both PG&E and SCE territories. For households in PG&E territory, we were able to collect monthly and hourly data for 2016-2017. For general population electricity usage data, we obtained two large data sets containing the anonymized data for our main geographic regions of interest: Marin County (PG&E Public Data Sets: Electric Usage by Zip) and the City of Long Beach (SCE Quarterly Customer Data Reports). We also accessed the California Public Utilities Commission (CPUC) "2016 Residential Electric Usage and Bill Statistics by Climate Zone" data to find average and median household electricity consumption by climate zone.

3.2.1. General population residential electricity usage data

We began by acquiring the anonymized residential customer data by zip code for Marin County (PG&E) and Long Beach (SCE)—the areas where our participants were most concentrated. These are pre-produced, zip-code sortable anonymous customer datasets that served as a control group source.

We collected general population residential electricity usage data to compare with our participating households' electricity consumption. These datasets gave us more certainty of average and median electricity consumption rates in our study areas for our analysis. We acquired both PG&E (PG&E Public Data Sets: Electric Usage by Zip, 2017) and SCE residential aggregated electricity usage data for the zip codes (SCE Quarterly Customer Data Reports) to contrast with our participating households in Marin County and City of Long Beach. This data is easily downloadable by quarter from their respective websites. Our original fieldwork consent forms, beginning in 2015, incorporated language around energy data access, did not request utility account numbers (or service numbers) at that time. In February of 2017 a utility representative informed us that we needed a signature and account number, whereupon we sent a revised consent form to all participants.[4] Submitting this final round of consent forms to PG&E and SCE in late summer of 2017, we received our participants' utility data late in the third quarter of 2017.

[4] Personal communication between Lena Lopez at PG&E Data Governance and Products and Chris Granda in February 2017.

3.3. Statistical Tests on Electricity Data

3.3.1. Electricity Usage Distribution

To help with analyzing the electricity usage data, we wanted to confirm that the electricity usage data for the general population was normally distributed. This would allow us to perform t-tests on the data.

Figure 1 - Figure 2 display histograms of household electricity consumption.[5] As displayed, the histograms demonstrate that household electricity consumption in fact is normally distributed for the general population of the areas of interest in SCE and PG&E territories combined (Figure 1), general population for PG&E customers in Marin County (Figure 2), and general population for SCE customers in City of Long Beach (Figure 3). For each histogram, the x-axis is the monthly energy usage (in kWh), and the y-axis is the frequency (a.k.a. number of instances that month a month had that average kWh, a.k.a. the number of months in which someone used that much energy).

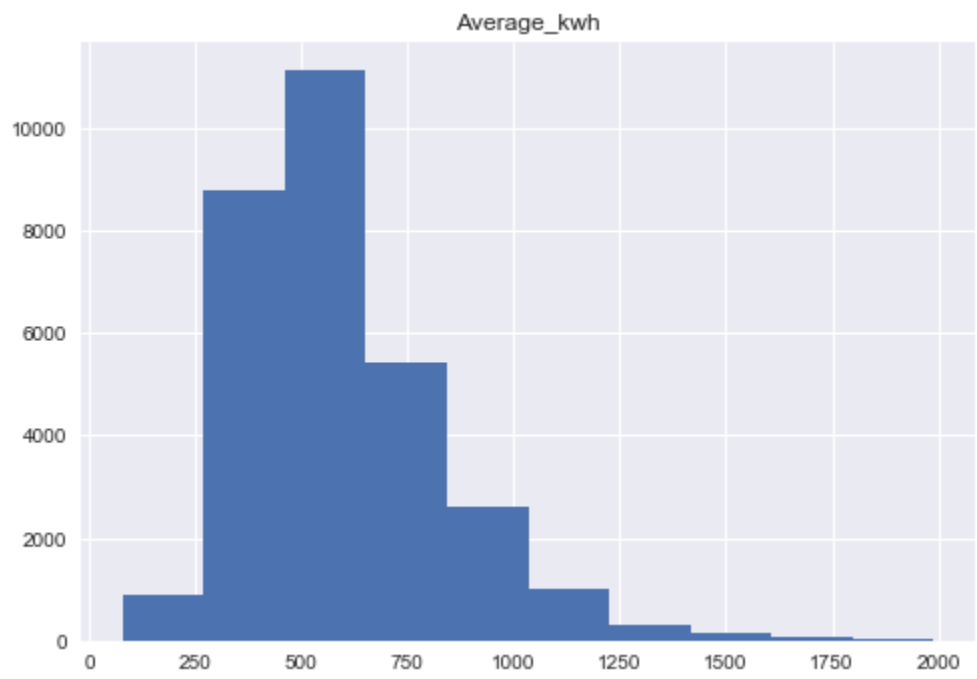

Figure 1. Histogram of the aggregated, general population electricity usage data.

[5] Some monthly electricity usage data exceed 2000. In the graphs below, we excluded them as outliers for visualization purposes.

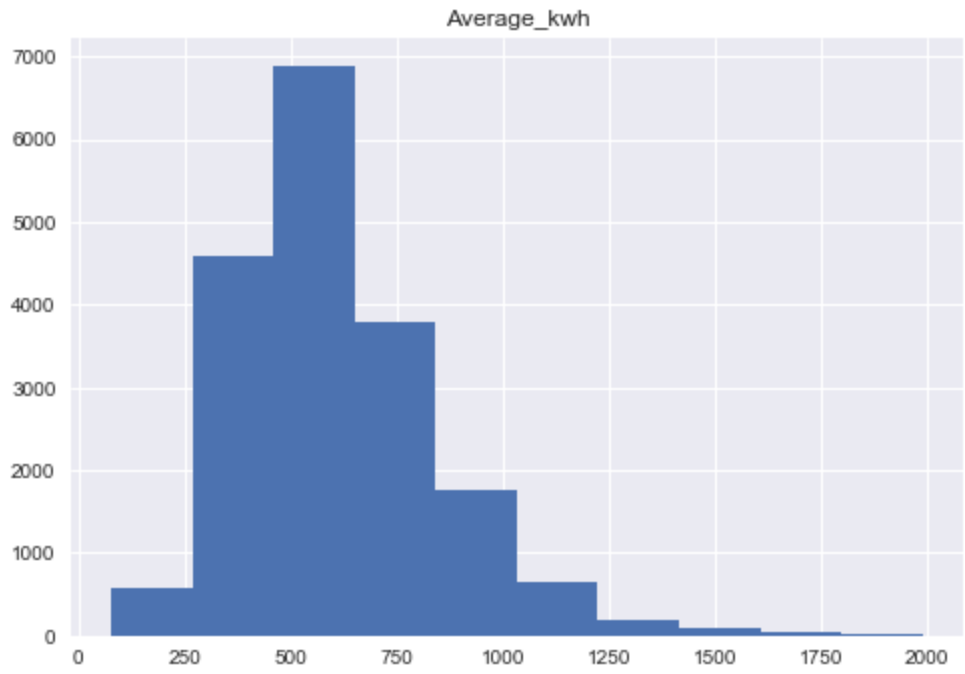

Figure 2. Histogram of PG&E general Electricity Consumption

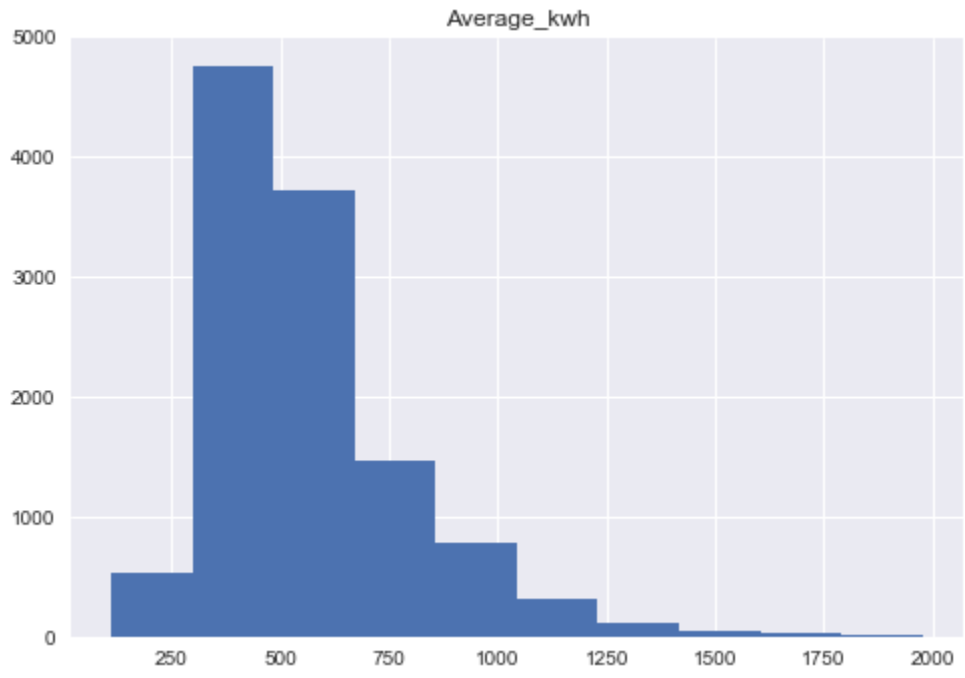

Figure 3. Histogram of SCE general population data usage

3.3.2. PG&E vs. SCE data usage

The two utility company customers demonstrate two distinct levels of electricity use, therefore we compared PG&E electricity usage data from participating households with PG&E electricity usage data from the general population. We similarly compared SCE electricity usage data from participating

households with SCE electricity usage data from the general population. We applied a two-sided t-test to determine that the average household energy usage for the general population of City of Long Beach in SCE is statistically different than for the population of Marin County in PG&E territory (Table 2). This agrees with our experience since the areas of interest, Marin County, and City of Long Beach, are in very different climates with different building constructions and occupancy. We used the PG&E and SCE residential aggregated electricity usage data for the zip codes to arrive at an average use of kWh per household by territory as seen in Table 2.

Table 2. Monthly User Average kWh by Company

	Monthly Average kWh (household)
PG&E	617
SCE	571

Because they are statistically different, we assessed electricity usage by PG&E and SCE households separately. We only compared participating household data to general population data within the same utility territory.

3.4. Monthly Usage

Below is the monthly average usage for the PG&E Marin County population, SCE City of Long Beach population, and the average of the two populations combined. The F-test[6] performed on this data confirms that the variation by month between the two utility averages is statistically significant.

Table 3. Monthly average household kWh usage for PG&E Marin County population, SCE City of Long Beach population, and the average of the two populations combined.

Months	PG&E General Population Mean kWh	SCE General Population Mean kWh	Combined Mean kWh
January	662	595	636
February	525	491	512
March	543	461	511
April	500	458	484
May	534	450	501
June	659	538	612
July	725	706	718
August	693	751	715
September	597	728	648

[6] "An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled." (Wikipedia, 2018c)

Months	PG&E General Population Mean kWh	SCE General Population Mean kWh	Combined Mean kWh
October	541	627	575
November	611	511	572
December	813	533	705

We set out to analyze the data by comparing the:

- PG&E household cybersensitive average monthly energy use to the general population PG&E data in Marin County,
- PG&E household non-cybersensitive average monthly energy use to the general population PG&E data in Marin County,
- SCE household cybersensitive average monthly energy use to the general population SCE data in the City of Long Beach, and
- SCE household non-cybersensitive average monthly energy use to the general population SCE data in the City of Long Beach.

3.5. PG&E Monthly and Annual Usage

We first examined the data collected from our participating households residing in PG&E territory. We analyzed the monthly and annual averages of energy consumption for these households. We grouped and averaged monthly electricity use for the segments cybersensitives and non-cybersensitives We were interested in learning whether there was a statistically significant difference in monthly energy usage between the two segments. Table 4 summarizes the average monthly and annual electricity usage among participating households who were PG&E customers along with their cybersensitivity status.

Table 4. Participating PG&E territory households' average monthly and annual electricity and cybersensitivity status.

Household	Monthly Average kWh (household)	Annual Average kWh (household)	Cybersensitivity Status
1	1180	14,167	Cybersensitive
2	1012	12,146	Non-Cybersensitive
3	811	9732	Non-Cybersensitive
4	683	8198	Non-Cybersensitive
5	651	7823	Non-Cybersensitive
6	579	6958	Non-Cybersensitive
7	247	2969	Non-Cybersensitive
8	225	2705	Non-Cybersensitive
9	198	2386	Non-Cybersensitive
10	169	2028	Cybersensitive
11	137	1655	Cybersensitive
12	121	1458	Cybersensitive

Table 5 shows the aggregated average monthly and annual electricity usage for PG&E cybersensitives and non-cybersensitives.

Table 5. PG&E participating households' average monthly and annual electricity by cybersensitivity status.

	Monthly Average kWh (household)	**Annual Average kWh (household)**
Cybersensitive	402	4827
Non-Cybersensitive	551	6615

We used a two-sample t-test to determine whether there is a statistical difference between the cybersensitives and non-cybersensitives. As shown by the Table 6, they have very different variances, so we used Welch's version of the two-sample t-test because it explicitly adjusts for unequal variances. Table 6 presents the standard deviation for the cybersensitives and non- cybersensitives. The standard deviation for cybersensitives in PG&E territory is 519 which is close to its mean of 402. This indicated a large range of household electricity usage and merited further investigation of outliers.

Table 6. Standard deviations for electricity usage by participating PG&E customers, broken out by cyber status

	Monthly Average kWh (household)	**Annual Average kWh (household)**
Cybersensitive	519	6231
Non-Cybersensitive	300	3602

We found one major outlier among the PG&E cybersensitives data. Household "1" had the highest average electricity consumption of all participants—and was almost ten times that of the next highest cybersensitive. The other cybersensitives rank among the three lowest average electricity users per month (Table 4). If we remove Household 1 as an outlier, our sample size is 10 and the number of cybersensitives to 3. This is small but produces more valid results.

Table 7 and Table 8 show the new mean cybersensitives and non-cybersensitives when the outlier is removed. The mean and standard deviation for Cybersensitives' average monthly electricity consumption becomes 142 kWh with a standard deviation of 24 kWh. After excluding the outlier, the standard deviation becomes closer to what one would expect for a dataset of this size and magnitude.

Table 7. PG&E customer participating household means after excluding outlier

	Monthly Mean kWh (household)	**Annual Mean kWh (household)**
Cybersensitive	142	1714
Non-Cybersensitive	551	6615

Table 8. PG&E customer participating household standard deviations in electricity usage after removing outlier

	Monthly Average kWh (household)	**Annual Average kWh (household)**
Cybersensitive	24	289

	Monthly Average kWh (household)	Annual Average kWh (household)
Non-Cybersensitive	300	3602

3.5.1. Statistical Tests

We used a Mann-Whitney U-test[7] to determine whether there is a statistical difference between the cybersensitives and non-cybersensitives. The small sample size makes it difficult to determine whether both are normally distributed. The U-test is the best test to use when you cannot confirm that the data is normally distributed. Being nonparametric, the test also does not assume similar standard deviations. We used the Mann-Whitney U-test on the revised dataset (after excluding the outlier).

3.6. Study Participants vs General Local Population Sample

3.6.1. PG&E

We compared the PG&E participants with PG&E household customers in Marin County. The first two-sided t-test analyzed whether PG&E cybersensitives had a statistically significant difference in electricity usage compared with the local population, and the second two-sided t-test analyzed whether PG&E non-cybersensitives had a statistically significant difference in electricity usage compared with the local population.[8]

Table 9. Mean monthly electricity usage for PG&E participants by cyber status

	Monthly Average kWh (household)
Cybersensitive	402
Non-Cybersensitive	549

3.6.2. SCE

We compared the SCE participants with SCE household customers in City of Long Beach. The first two-sided t-test analyzed whether SCE cybersensitives had a statistically significant difference in electricity usage with the local population. The second two-sided t-test analyzed whether SCE non-cybersensitives had a statistically significant difference in electricity usage with the general population. The table below shows the mean monthly electricity usage for SCE cybersensitives and non-cybersensitives.

Table 10. Mean monthly electricity usage for SCE participants by cyber status

	Monthly Average kWh (household)

[7] "The Mann-Whitney U-test is a test of the null hypothesis that it is equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample. Unlike the t-test it does not require the assumption of normal distributions. It is nearly as efficient as the t-test on normal distribution. This test can be used to determine whether two independent samples were selected from populations having the same distribution." (Wikipedia, 2018b).

[8] Because different individuals have different lengths of their monthly bills (in addition to months being of different lengths, certain billing cycles are more than 31 days), the statistical tests below analyze the average daily electricity usage, calculated by dividing monthly electricity usage by the number of days in that monthly period. This provides a standardized unit to compare different month intervals.

	Monthly Average kWh (household)
Cybersensitive	670
Non-Cybersensitive	946

3.7. CPUC

The California Public Utilities Commission (CPUC) has provided extended data on electricity usage for California's Building Climate Zones (CPUC, 2016). We analyzed their findings and compared them with the data from our study and with the general population PG&E customer electricity usage for Marin County. This dataset normalizes energy consumption by climate. It is helpful to corroborate or negate our findings from the comparison of our participants with the local population data in PG&E.

We used addresses and zip codes from participants to find their respective California Building Climate Zone and then accessed the corresponding CPUC average residential electric usage data. The data from CPUC does not include income-qualified customers participating in the California Alternate Rates for Energy (CARE) program. We then compared the cybersensitives' average monthly electricity usage to the CPUC building climate zone data. We also compared non-cybersensitives' average monthly electricity usage to the CPUC building climate zone data.

Our PG&E participants are located primarily in Climate Zone 2 and 3. Thus, the table below includes the averages for Climate Zone 2 and 3. One PG&E participant was in Climate Zone 12, but one participant is not enough to generate statistically representative comparisons. Thus, we will focus on Climate Zone 2 and 3 in this report.

One notices, in Table 3, that for most groups, the means are greater than the medians. Any large sampling naturally includes a few months with extremes of electricity consumption, and these outliers disproportionately increase the mean (but not the median). The results from our study have a less present distinction between mean and median. This is because our study has a leaner sample size and studies with extremely large samples, like the CPUC and PG&E general population datasets, have more diversity in household electricity consumption.

We compare data from the CPUC study with data from the PG&E participants in our study. [9,]The CPUC dataset reports the mean and median household energy consumption for winter and summer seasons. Therefore, we are group our PG&E data seasonally as follows: Winter is December, January, and February; Spring is March, April, and May; Summer is June, July, and August; and Fall is September, October, and November. We are interested in whether the means found for winter and summer in Climate Zone 2 and 3 are comparable. Table 11 and Table 12 present the Winter and Summer comparisons of the CPUC data with the average PG&E participant household energy consumption. We bolded the cybersensitive averages in Table 11 for easier comparison.

Table 11. Winter comparison of the CPUC data with the average PG&E participant household energy consumption broken out by cyber status.

	Climate Zone	CPUC Study Average (kWh)	PG&E Study Participants Average (kWh)	Cybersensitives Average (kWh)	Non-Cybersensitives Average (kWh)

[9] CPUC only reports summer and winter averages but not the spring and fall.

	Climate Zone	CPUC Study Average (kWh)	PG&E Study Participants Average (kWh)	Cybersensitives Average (kWh)	Non-Cybersensitives Average (kWh)
Winter	2	686	551	**192**	731
Winter	3	469	539	**147**	696

Table 12. Summer comparison of the CPUC data with the average PG&E participant household energy consumption broken out by cyber status.

	Climate Zone	CPUC Study Average(kWh)	PG&E Study Participants Average(kWh)	Cybersensitives Average(kWh)	Non-Cybersensitives Average(kWh)
Summer	2	597	421	**147**	552
Summer	3	371	423	**122**	539

Table 13. Seasonal median and average monthly electricity consumption for all PG&E populations under study

Season	**Climate Zone**	**CPUC Study Average (kWh)**	**CPUC Study Median (kWh)**	**PG&E General Population Average (kWh)**	**PG&E General Population Median (kWh)**	**PG&E Study Participant Average (kWh)**	**PG&E Study Participant Median (kWh)**	**Cybersensitive Average (kWh)**	**Cybersensitive Median (kWh)**	**Non-Cybersensitive Average (kWh)**	**Non-Cybersensitive Median (kWh)**
Fall	2	-	-	750	614	479	543	**196**	180	620	618
Fall	3	-	-	423	399	454	276	**133**	135	582	687
Spring	2	-	-	646	594	423	496	**142**	133	563	559
Spring	3	-	-	412	388	439	218	**116**	114	569	598
Summer	2	597	424	669	629	421	506	**147**	142	552	552
Summer	3	371	287	396	376	423	243	**122**	113	539	603
Winter	2	686	491	1070	716	551	609	**192**	204	731	670
Winter	3	469	372	500	475	539	306	**147**	146		

3.8. Ethnographic Data

3.8.1. Quantifying Behavioral Observation

One of the goals of this project was to set ethnographic data alongside larger numerical data sets. Quantifying behavioral observation is the process of converting qualitative data into codes. Generically speaking, qualitative data amenable to such treatment may include themes from in-depth interviews, self-reports from time diaries, or observed number of instances of a behavior (e.g. grooming) that are calculated as tallies, frequencies, distributions, percentages, etc. They can also be combined into cross tabs with other forms of quantitative data, such as size of household or income, and analyzed statistically using methods such as Pearson's Chi Square or Goodness of Fit.

As we presented in the Task 3 report "Psychosocial Drivers of Behavior," our ethnographers used the software program Atlas.ti to tag interview transcripts with codes. We used the data from ethnographer observations and the Atlas.ti coding of transcribed interviews to identify ways in which interviewees interact with technology in their everyday lives and to determine the psychological drivers behind that usage. Using Atlas.ti to code the data allowed us to quantify behavioral observations and run frequency counts independently of the subjective perceptions of the ethnographers. The resultant clustering of cybersensitives, cyberawares, nulls, and the group in between, who we think of as 'mainstream' showed up strongly in these code frequency counts. Then, in Task 4, "Cybersensitive Response to Technology," (Indicia Consulting, 2017b) we integrated the datasets from the recruitment survey and the in-depth interviews. We performed statistical analysis on the combined dataset. In this report, we are using these statistically validated segments (cybersensitive, cyberaware, etc.) to organize and group the electricity data provided for participating households.

3.8.2. Reviewing Transcripts

Returning to the original in-home interview transcripts for this report, we reviewed the answers made by cybersensitives and non-cybersensitives to ethnographers' specific questions about appliances, energy bills, energy programs they may have heard about, and their tracking of household energy usage. This is where the value of qualitative data becomes apparent, because it was readily noticeable that cybersensitive interviewees responded to these questions with better recall, more detail, and higher levels of engagement and enthusiasm. As an example, let us compare an answer made by a cybersensitive to a non-cybersensitive about refrigerators:

Cybersensitive response:

> Respondent: For the refrigerator, I wanted something that was eligible for the CEC's cash for clunkers program. And so, meeting the specs for that and… which weren't that onerous. But so, I wanted an energy star fridge and one that was not that big because as you've now seen we have a small house. And the kitchen's small so I did not want any extra features that are energy consuming features like water or an ice maker and it comes with an ice maker, but we've disabled it completely. I didn't want those things because I know from my work that those features use a fair amount of energy and I just don't value those services at all. So, a basic fridge. I also wanted the freezer on bottom style of fridge just for kind of convenience of where I'm opening which doors. So, there was like the aesthetic aspect of it but also small and no extra features that would consume energy.

Non-cybersensitive response:

Respondent: That refrigerator is a pretty new refrigerator.

Interviewer: [Crosstalk] When was this replaced?

R: 5 years maybe.

Interviewer: So, at that time did you get any rebates or anything?

Respondent: [Crosstalk] Yeah, we got PG&E I think they even paid for that refrigerator at any rate.

In another example, for illustrative purposes, we also compared answers to the question, "How did you hear about Program X?"

Cybersensitive response:

Interviewer: Let's talk about the changes you're making in your lawn right now. How did you find out about the rebates on that?

Respondent: So, the water district had a couple different rebates so like they'll all give you rebates for getting new appliances, for getting rain barrels. I had already gotten rain barrel. I bought one from my neighbor for $15. They'll just… so they have the… and then they mentioned that there was turf rebate. And then my neighbor up the street got it. **So that's when I went, and I went online, and I investigated it and applied for it**.[10]

Non-cybersensitive response:

Interviewer: How do you find information about taking measures like that?

Respondent: Well… we get stuff in the mail that does that. There may be an article in a magazine. If I'm interested I can google it and get a lot of information about that. But I think a lot of it comes from public awareness campaigns that come in a variety of different media.

4.Results

In this section we share our findings from the two methods of analysis undertaken for this report, which are quantitative and qualitative.

4.1.1. Quantitative analysis

We establish that the participants from the two utility companies demonstrate distinctly different levels of electricity usage, which we attribute to climate zone and geography.

The means of the average monthly electricity usage for all five cyber status categories: Cybersensitives, Cyberawares, Mainstream, Low Mainstream, and Null appear to be different in that, the greater the degree of cybersensitivity, the lower the average monthly usage.

Cybersensitives consistently use a statistically significantly lower amount of electricity than either the CPUC study participants, or the PG&E general population, or the PG&E non-cybersensitives.

[10] Emphasis ours

SCE cybersensitives are quite a bit lower on average in absolute terms of electricity usage than are the non-cybersensitives.

PG&E cybersensitives use a statistically significant lower amount of electricity usage per month than PG&E non-cybersensitives or PG&E general population. PG&E non-cybersensitives possess a statistically insignificant difference with the PG&E general population. PG&E customers who were our study participants, and who lived in climate zone three were similar in usage to CPUC study participants. In Climate Zone 2, PG&E territory participants in our study use electricity at a different rate than those studied by the CPUC. In our test, we concluded that the non-cybersensitives in our study may have a similar mean to the CPUC results.

4.1.2. Qualitative Analysis

While both cybersensitives and non-cybersensitives may participate in energy efficiency programs, the pathways to uptake are very different, as is their receptivity to offers, their engagement with data from their utility, and other dimensions. Summing up, cybersensitives are more active in acquiring information about energy efficiency measures and programs, than are non-cybersensitives. They use more detail and description in their answers, and more often report taking direct action in response to information received from their utility.

4.2. PG&E customer usage vs. SCE customer usage

We evaluated PG&E and SCE participant and local population household energy consumption separately. We found with statistical tests a small p-value, around 0.000004, which indicated with high probability that PG&E and SCE customers have statistically significantly different electricity usage. We compared electricity usage data for PG&E participating households with local population PG&E data for Marin County. Likewise, we compared electricity usage data for SCE participating households with local population SCE data for the City of Long Beach.

4.3. Monthly and Annual Averages

4.3.1. PG&E

The average monthly electricity usage for PG&E per month is reported in Table 14. The second column lists average household electricity usage for the local population of PG&E customers located in Marin County. The third and fourth column list the average monthly electricity usages for the PG&E cybersensitives and non-cybersensitives in our study. One can see simply looking at these, that in general, PG&E cybersensitives average significantly lower electricity usage per month throughout the year.

Table 14. Mean PG&E household monthly electricity usage throughout the year (36 months of data).

Months	PG&E Local Population Household Mean (kWh)	PG&E Cybersensitives Household Mean (kWh)	PG&E Non-Cybersensitives Household Mean (kWh)
January	662	165	622
February	525	145	583
March	543	126	521
April	500	121	521
May	534	127	511
June	659	130	487
July	725	126	487
August	693	135	530
September	597	144	505
October	541	151	518
November	611	168	604
December	813	176	732

We tested this observation for each month throughout the year. We found the following results:

- PG&E cybersensitives have a mean monthly electricity use of 402 kWh. This is lower than the mean general PG&E monthly electricity usage of 618 kWh, which is statistically significant. PG&E cybersensitives exhibit a statistically significant difference in electricity consumption when contrasted with the PG&E general population, except during November, December, and April. The former two may represent a tendency during the winter for PG&E cybersensitives to exhibit electricity usage that is less different (or simply be anomalies), and the April result is likely an anomaly.
- PG&E non-cybersensitives possess a statistically insignificant difference with the PG&E general population, except during June, July, and August. These could represent a statistically significant difference in electricity usage over the summer months, potentially indicating a significant change in electricity usage behavior by non-cybersensitives over the summer (or be anomalies).

4.3.2. Monthly Electricity Usage across all Segments

Table 15 lists monthly average electricity usage for all five cyber status categories: Cybersensitives, Cyberawares, Mainstream, Low Mainstream, and Null. They appear to be different in that, the greater the degree of cybersensitivity, the lower the average monthly usage. The group sizes are too small to reliably test this difference with a statistical test, but it confirms initial impressions about the influence of cybersensitivity on electricity usage. Thus, among the PG&E research participants, the cybersensitives have a statistically significant difference in electricity usage but non-cybersensitives do not.

Table 15. Average household monthly electricity usage by cyber status segments

	Monthly Average household electricity consumption (kWh)
Cybersensitive	121
Cyberaware	153
Mainstream	544
Low Mainstream	225
Null	731

4.4. Participants and General Population Electricity Usage

4.4.1. PG&E

A small p-value of .013 demonstrates that there is a statistically significant difference in electricity usage between PG&E cybersensitives and local population PG&E household electricity usage. However, the non-cybersensitives cannot be shown to have a statistically significant difference in electricity usage from the general population. Even though PG&E non-cybersensitives have a lower average monthly electricity usage than the general PG&E population, this does not seem to be statistically significant.

4.4.2. SCE

The average household monthly electricity usage for the SCE local population was 571 kWh, which is higher than both the cybersensitives and non-cybersensitives averages, 670 kWh and 947 kWh, respectively. However, cybersensitives in SCE territory consumed quite a bit less electricity on average than the non-cybersensitives.

4.5. CPUC

Analysis showed that PG&E participating households who lived in Climate Zone 3 consumed similar rates of electricity to the CPUC sampled households. As shown in Table 3, there is variation in electricity usage by month and season. This should not be surprising considering the difference in household electricity demands during different seasons and months throughout the year.

In our analysis, we concluded that non-cybersensitives may have similar average energy consumption rates when compared to the CPUC results, thus not statistically different. The results of our analysis seem to indicate that in Climate Zone 2, our participants consume electricity at a different rate than those households sampled by the CPUC. The means between the two studies are different.

The PG&E cybersensitive participants in our study, consume less electricity than the CPUC sampled households for Climate Zones 2 and 3, the PG&E local customers, and the PG&E non-cybersensitives (Table 11).

5. Discussion

5.1. PG&E vs. SCE

Geographic and climatic differences between the two parts of California served by these two companies likely account for the different patterns of consumption. It may seem to be unnecessary to establish this with our own data set, but by showing the difference (and then comparing data only within territories) we can show that cybersensitivity is visible as a pattern in the data irrespective of geographic location. This adds support to our hypothesis that cybersensitivity is a previously undiscovered personality trait, and not a gloss on a data artifact.

5.2. Monthly and Annual Averages Across Cybersensitive Status

5.2.1. PG&E

When including outliers, there was no statistically significant difference between cybersensitives and non-cybersensitives. This would seem to imply that the two groups have the same average monthly and annual electricity usage. Yet looking at the actual kWh data patterns of difference were apparent. The team discussed what we felt were anomalous statistical findings and thus looked closely at some of the context for clues as to why the apparent differences in average usage between cybersensitives and non-cybersensitives disappeared when tested statistically. One of the issues was the variances. The eight non-cybersensitives had a much smaller variance, meaning that they clustered together, mostly at higher levels of usage than cybersensitives – which is good news for our hypothesis. The cybersensitives, however, had a variance almost three times as large, but when we looked at the individual data we discovered that one household was an extreme outlier, using more energy than any of the other Northern California households.

Removing Household 1 instantly reduced the variance to something much closer to the non-cybers' variance. An outlier in a small sample has an outsized effect on the variance, and thus on the test. Removing Household 1 meant that we do have statistical significance for the difference between cybersensitives and non-cybersensitives (which matches what one would intuitively expect from looking at the monthly/yearly usage averages, but also comparing households directly. For example, cybersensitives clustering in the lower range of usage vs. non-cybersensitives. Looking more closely at House 0 we see that this was one of the larger ones in the sample with four members. Further, upon examination, the participant's age range was 45-54. We hypothesized that two of those members are teenagers, with many energy-consuming devices and behaviors. This hypothesis would go a long way towards explaining how she could be a 'deep green' cybersensitive, but still have larger than average consumption.[11]

After excluding the outlier, there is a statistically significant difference between cybersensitives and non-cybersensitives in their average monthly and annual electricity usage. When including the outlier,

[11] We also looked at household sizes as a possible explanation for lower consumption, and two of the cybers were singletons, but the third was our 'deepest green' cybersensitive, HG, with a household size of three persons.

however, the results are insignificant. This is most likely due to the disproportionate influence of the outlier and a low sample size. The p-value is small, 0.9%, which means we can be confident that there is a statistical difference between the average monthly and annual electricity usage of cybersensitives and non-cybersensitives (and by implication, the general population). Excluding the outlier, the statistical test suggests a distinct difference between the Cybersensitives and non-Cybersensitives. Simply looking at their monthly and annual averages makes them appear different, and the small sample size may be influence the statistical tests. We believe that since excluding one outlier changed the result to being statistically significant, when placed in conjunction with our ethnographic observations of behavior – supports our hypothesis that cybersensitivity influences patterns of electricity consumption.

Excluding Outliers

When excluding the outlier, there seems to be a statistically significant difference between Cybersensitives and non-Cybersensitives. The outlier would explain why the initial tests did not indicate a difference between the groups, despite a seemingly apparent distinction looking at the numbers. An outlier like this influences the results far more with smaller group sizes, so more statistical testing with the full sample amount could help further confirm (or negate) this difference.

5.2.2. SCE

Both the SCE cybersensitives and non-cybersensitives potentially have a statistically significant difference in electricity usage when compared to the local general population, but the sample size is too small to be able to verify this.

5.3. Participants and General Population Electricity Usage

5.3.1. PG&E

We would need to conduct further research to understand why PG&E participating households and the local population did not yield a significant difference in November, December, and April. Although this could potentially occur because of a difference in electricity use behavior during the winter among cybersensitives.

5.3.2. SCE

Most tests, except for tests on the PG&E summer months and the SCE usage, indicate a lack of a significant difference in electricity usage. Whether (or if the case, why) non-cybersensitives exhibit a more distinct type of electricity usage behavior than the general population during the summer months would also require further research but based on these findings, could be the case as well. Again, this sample size was quite small, and SCE participant households were the most affected by the discrepancy in terms of the data delivered to us (12 months vs. 36 months). A larger sample would be preferable, but even with this small sample, we see the same trend that cybersensitives use less energy than their peers.

5.4. CPUC Comparison by Climate Zone and Season

We analyzed the extent to which the CPUC sampled household data corroborates our study findings in terms of electricity usage. We asked whether the CPUC's mean electricity usages for summer and winter in each Climate Zone are compatible with the electricity usages we found in our PG&E study participants for their respective Climate Zones. Comparing our dataset to that of the CPUC helped us to analyze in

more detail how climate, region and building construction influence patterns of electricity consumption. We also were able to set our results against those of other studies for a potential narrative of reliability

First, understanding the role climate plays in household electricity consumption is insightful. It helps to explain the different patterns of electricity consumption found with our cybersensitives and non-cybersensitives in Marin County and City of Long Beach.

Second, we can confirm that other studies support our findings. By comparing means between the two studies, we are comparing whether to what extent our findings are similar. For example, analyzing whether our study has similar mean electricity usages, we are determining whether the two studies represent similar or different random samples. This type of comparison is commonly done to determine whether the sampling fairly represented the population or whether it was biased in some way: if it compares similarly with our (hopefully) fair and roughly accurate samples from the CPUC study, there is a greater chance that there is no bias in who we chose as well (the opposite -concluding potential bias because of different means - is harder to show). This starts a kind of statistical cross-comparison between other statistical studies common in many studies.

5.5 Ethnographic Data

We had previously validated our hypothesized segments by 'quantifying behavioral observation' through a set of coded ethnographic data and survey response data. We established that a certain set of our respondents demonstrated statistically significant different energy consumption patterns. We have also validated the hypothesis that cybersensitivity does not correlate with the demographic variables of age, gender, or income. Nor does cybersensitivity correlate with a specific geography in our sample (Indicia Consulting, 2017b).

The initial Program Opportunity Notice requested research that evaluated energy efficiency measures people have installed, such as LEDs and thermostats, or energy efficiency programs they participated in. Therefore, we looked closely at how the two segments, cybersensitives and non-cybers, responded with respect to the topics such as rebates, LEDs, thermostats, and home energy audits (particularly those with product installation such as lights or low flow). In our review of the transcripts, it soon became clear, that while both cybersensitives and non-cybersensitives may participate in energy efficiency programs, the pathways to uptake are very different, as is their receptivity to offers, their engagement with data from their utility, and other dimensions. Summing up, cybersensitives are more active in acquiring information about energy efficiency measures and programs, than are non-cybersensitives. They use more detail and description in their answers, and more often report taking direct action in response to information received from their utility.

5.6 Future Research

In terms of interpreting these findings, they mainly point us in a new direction to dig further. While we were able to statistically validate many of our hypotheses, our sample size was small. These findings would benefit from a follow-up study with a larger sample size to confirm the patterns we are seeing and establish reliability. In our final report we will discuss Ethnographic Decision Tree Modeling, and how it could be used to test these findings. Further, testing this with utility data from other companies/regions of the countries would expand the sample size and inform reliability.

6.Conclusion

Our goals for Task 5 were to:

- Organize and analyze **a** quantitative data set consisting of electricity usage data
- Place cybersensitive participation and performance against a representative sample of local electricity consumers' data to establish the delta of electricity consumption and responsiveness.
- Measure cybersensitive participation and performance in past energy efficiency programs using partner utility electricity data.
- Set ethnographic data alongside larger numerical data sets.
- Identify a consistently repeating set of characteristics, including behavioral, demographic, and energy usage, which can be attributed to the cybersensitive profile.

We acquired, organized, and analyzed **four** electricity usage datasets:

- 12-36 months of electricity data for 23 consenting participating households
- Anonymized residential customer data by zip code for Marin County (PG&E Public Data Sets: Electric Usage by Zip)
- Anonymized residential customer data by zip code for City of Long Beach (SCE Quarterly Customer Data Reports)
- California Public Utilities Commission (CPUC) 2016 Residential Electric Usage and Bill Statistics by Climate Zone

We placed cybersensitive (and non-cybersensitive) participation and performance against a representative sample of local electricity consumers' data to establish the delta of electricity consumption and responsiveness. We demonstrated a consistent pattern of a delta existing between cybersensitive electricity usage when set against the general population for their local utility (at zip code level), or the sample from the CPUC. This distinction persisted regardless of season (Winter or Summer) or Climate Zone (2 or 3). While we do not have specific energy efficiency program participation data for our participating households, we did acquire utility electricity data for those households and used a 36-month sample to measure cybersensitives against other members of their cohort. We have established that non-cybersensitives are a reasonable match to the general population for their respective utilities, and that they are also a reasonable match to the sample used by the CPUC.

We set ethnographic data alongside larger numerical sets by first establishing the validity of our segmentation statistically. We then reviewed the interview transcripts and set the cybersensitive responses alongside non-cybersensitive responses to establish patterns. Repeatedly, the transcripts show that cybersensitives are much more likely to be aware of the availability of energy efficiency programs, to be interested in participating, to conduct cost-benefit analyses on their own behalf, and to be willing to actively seek out additional efficiency measures that they can undertake.

Finally, it was one of our goals for Task 5 to identify a consistently repeating set of characteristics, including behavioral, demographic, and energy usage, which can be attributed to the cybersensitive profile. Combining the statistical findings in this report with our extensive qualitative data allows us to make the case that distinct behavioral patterns exist between cybersensitives and non-cybersensitives when it comes to energy efficiency measure installation, and program awareness and participation.

References

Allcott, H. 2011. "Social Norms and Energy Conservation." Journal of Public Economics, Special Issue: The Role of Firms in Tax Systems, 95 (9–10): 1082–95. https://doi.org/10.1016/j.jpubeco.2011.03.003

California Public Utilities Commission, 2016, "2016 Residential Electric Usage and Bill Data Statistics by Climate Zone" http://www.cpuc.ca.gov/res2016/ (accessed January 30, 2018).

Foster B, Mazur-Stommen S. 2012. Results from Recent Real-Time Feedback Studies. Research Report B122. American Council for an Energy-Efficient Economy Washington, DC. Retrieved 2016-08-20 from http://aceee.org/research-report/b122

Indicia Consulting, 2016 "Preliminary Ethnographic Report on Cybersensitives and Technology Detailing the Fieldwork and Early Findings" Produced for the California Energy Commission by Indicia Consulting in fulfillment of Task 2 Deliverable 1 of the project Cybernetic Research across California: Documenting Technological Adoption and Behavior Change across Diverse Geographies and Populations to Inform Energy Efficiency Program Design Funded by EPIC PON 14-306, August 29, 2016 http://indiciaconsulting.com/downloads/Indicia-EPIC-Task-Deliverable-2-Report-on-Cybersensitives-and-Technology.pdf (accessed January 30, 2018).

Indicia Consulting, 2017a, "Psychosocial Drivers of Technology Engagement among Cybersensitives" Produced for the California Energy Commission by Indicia Consulting in fulfillment of Task 3 Deliverable 1 of the project Cybernetic Research across California: Documenting Technological Adoption and Behavior Change across Diverse Geographies and Populations to Inform Energy Efficiency Program Design Funded by EPIC PON 14-306, February 28, 2017 http://indiciaconsulting.com/downloads/Indicia-EPIC-Task-Deliverable-3-Psychosocial-Drivers-of-Technology-Engagement-Among-Cybersensitives.pdf (accessed January 30, 2018).

Indicia Consulting, 2017b "Cybersensitive Response to Technology" Produced for the California Energy Commission by Indicia Consulting in fulfillment of Task 4 Deliverable 1 of the project Cybernetic Research across California: Documenting Technological Adoption and Behavior Change across Diverse Geographies and Populations to Inform Energy Efficiency Program Design Funded by EPIC PON 14-306, June 30, 2017 http://indiciaconsulting.com/downloads/Indicia-EPIC-Task-Deliverable-4-Cybersensitive-Response-to-Technology.pdf (accessed January 30, 2018).

Wikipedia contributors, 2018a "Student's t-test," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Student%27s_t-test&oldid=823182456 (accessed January 30, 2018).

Wikipedia contributors, 2018b "Mann–Whitney U test," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Mann%E2%80%93Whitney_U_test&oldid=820599537 (accessed January 30, 2018).

Wikipedia contributors, 2018c "F-test," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=F-test&oldid=820799379 (accessed January 30, 2018).