Predicting Solar Panel Adoption

Donnie Meyer

November 1, 2018

Introduction

There are many reasons why consumers adopt, and choose not to adopt rooftop solar. For example, some adopt for financial gain while others adopt for environmental reasons. Understanding roof top solar adoption characteristics of consumers seems to be an extremely worth while endeavor. Volatile energy prices, environmental concerns, and lack of jobs in traditional energy sectors are all problems that could be alleviated with an increase in roof top solar adoption. This project seeks to determine which factors play a significant role in the rooftop solar adoption process. The cost of manufacturing panels has decreased substantially over time, and therefore has become more affordable for average households. Although the cost of manufacturing solar panels has decreased, the cost of obtaining customers for solar panels has remained relatively high. It is likely that consumers lack knowledge of the technology, have skepticism about making the switch, are unsure of its investment return, etc. If this is the case, then a better understanding of the consumer traits that influence roof top solar adoption could help solar companies push this technology forward. Solar installation firms can use this information as it would help them pinpoint marketing strategies and in turn increase sales. Knowing which groups of consumers to target with advertising could ultimately encourage consumers to adopt rooftop solar at a faster rate.

Data

Data Description

The name of the data set is "Understanding the Evolution of Customer Motivations and Adoption Barriers in Residential Solar Markets: Survey Data". It has been made public by the National Renewable Energy Laboratory (NREL) and can be found here https://data.nrel.gov/submissions/68.

For this project a survey data set will be used where all features are categorical, i.e. there are no continuous variables. The data is a compilation of three surveys titled an adopter survey, a considerer survey, and a general population survey. The adopter survey is limited to households who have adopted solar, the considerer survey is limited to households who have seriously considered adopting solar but have yet to do so, and the general population survey is limited to those who have never seriously considered solar. The surveys where conducted in California, Arizona, New York, and New Jersey. The general population survey is limited to households that are single family homes that are owned by the residents who occupy them. These households are more likely than other households to adopt rooftop solar making it possible to compare across groups(adopters and non-adopters).

The main features that will be used are composed of nominal and ordinal variables. Most variables are ordinal and many are represented in likert form questions such as "Do you believe installing solar panels will save you money?" where respondents have the choices strongly agree, agree, neutral, disagree, and strongly disagree. The nominal variables are mainly demographic characteristics such as gender, age, education, etc. The variables have been broken up into the six categories which are Demographics, Economics, Ethics, Consumer traits, Personality traits , and Environmental variables. Two categories did not make it into the final analysis because they had no statistical relationship with the dependent variable. These two categories were the Personality and Ethics categories. *HAVESOLAR* is a survey question asking "Do you currently have roof top solar installed at your current home?" and is the dependent variable in this project.

Data Wrangling

The first step was to append the three data sets. The data sets were combined keeping only those columns (survey questions) that where common across all three surveys. This allows for comparisons across groups because each group is answering the exact same questions. This eliminated about 200 variables leaving 49 in my final data set. Though many features where lost, the remaining 49 where the most relevant as these will allow for comparison across adopters and non-adopters. A small amount of recoding was performed on the dependent variable *HAVESOLAR*. One important note is that I decided to put people who had rooftop solar in the past but had moved to a new home without rooftop solar, into the group who had it. The thinking here is that this group is probably more similar to rooftop solar adopters than those who had not adopted, as they had adopted in the past.

The data is fairly clean, but there are quite a few NA's and some values that are large relative to the rest of the data. For example many questions contain responses such as "I don't know" and "prefer not to answer" that are equal to 98 and 95 respectively. There are many values equal to 99 in the data that was later discovered to be NA as well. The value of 99 was spread over multiple columns and rows. The r-package *naniar* has a function called replace_with_na_all(). This function will go through each column and find the value you assign, in my case 99, and replace each of those values with NA. The NA's are left in the data set as opposed to using complete.cases() because most columns only have between 0-2 percent NA's. Removing the NA's would have eliminated 25% of the rows in my data. I also recoded the value of 98 ("I don't know) to NA's because this is similar to"No_Answer" and made little difference in the final analysis. Removing 98 from a specific set of variables allowed for the creation of ordered factors.

The next step was to rename the variables and convert each column into factor variables. I took on the burdensome task of renaming all my variables so that they were more descriptive. Though burdensome, a value of "climate_change"" as opposed to "PN1"" is much more informative and saves me time in the long run as opposed to looking up a variable every time in the code book. There is now at least a good sense of what this variable represents. A factor data set was created by converting all integers to factors and ordered factors. The final data sets structure is below. This is the full data set with all features. The final data set will contain only 18 features that are correlated with the dependent variable *HAVESOLAR*.

```
Observations: 3,533
Variables: 49
$ X1                      <ord> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,...
$ CASE_ID                 <ord> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,...
$ GPS_NAC_ADOPTER         <ord> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
$ STATE                   <ord> 1, 4, 2, 4, 4, 1, 1, 4, 1, 2, 3, 3, 1,...
$ HAVESOLAR               <ord> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
$ winter_bill             <ord> 6, 7, 5, 3, 4, 7, 3, 8, 6, NA, 6, 5, 4...
$ summer_bill             <ord> 4, 8, 5, 2, 3, 8, 4, 7, 6, NA, 10, 9, ...
$ renewable_energy        <ord> 3, 5, 3, 3, 4, 1, 5, 4, 4, NA, 3, 3, 3...
$ climate_change          <ord> 3, 5, 3, 3, 4, 1, 4, 4, 3, 3, 3, 2, 3,...
$ waste_energy            <ord> 3, 3, 4, 3, 4, 2, 5, 4, 5, 3, 4, 4, 4,...
$ climate_change_serious  <ord> 4, 5, 4, 4, 4, 2, 5, 3, NA, 5, 5, 3, 4...
$ evironment_improve      <ord> 4, 5, 3, 3, 3, 1, 3, 5, NA, NA, 5, 4, ...
$ slow_climate_change     <ord> 4, 5, NA, 3, 3, 1, 3, 4, NA, NA, 5, NA...
$ reduce_footprint        <ord> 3, 5, 3, 3, 4, 2, 4, 3, 4, 4, 4, 4, 4,...
$ slow_energy_price       <ord> 2, 4, 3, 3, 4, 1, 4, 4, 4, 5, 4, 3, 4,...
$ return_investment       <ord> 3, 4, 3, 2, 3, 1, NA, 3, 4, 3, 4, 4, 3...
$ save_money              <ord> 3, 5, NA, 3, 4, 2, 3, 3, 4, 4, 4, 4, 3...
$ Co3                     <ord> 3, 5, 3, 2, 4, 1, 4, 3, 4, NA, 4, 4, 3...
$ protect_environment     <ord> 3, 4, 4, 3, 4, 3, 4, 4, 4, 5, 4, 3, 5,...
$ respect_earth           <ord> 3, 4, 4, 3, 4, 4, 4, 3, 4, 3, 4, 3, 3,...
$ unity_nature            <ord> 3, 4, 3, 3, 4, 3, 3, 3, 4, 2, 4, 3, 4,...
$ world_peace             <ord> 3, 4, 5, 3, 4, 3, 4, 3, 5, 5, 5, 4, 2,...
```

```
$ social_justice          <ord> 2, 4, 4, 3, 4, 3, 4, 4, 4, 5, 5, 5, 1,...
$ equality                <ord> 3, 5, 5, 3, 4, 3, 5, 5, 4, 5, 5, 4, 2,...
$ respect_elders          <ord> 2, 4, 5, 4, 4, 4, 4, 4, 5, 5, 5, 5, 3,...
$ family_security         <ord> 3, 4, 4, 4, 4, 5, 4, 3, 5, 5, 5, 5, 3,...
$ self_discipline         <ord> 3, 4, 3, 3, 4, 4, 3, 4, 4, 2, 4, 4, 2,...
$ right_to_lead           <ord> 3, 3, 4, 3, 4, 3, 2, 3, 5, 2, 4, 3, 2,...
$ influential             <ord> 3, 3, 4, 3, 3, 3, 4, 4, 4, 3, 4, 3, 2,...
$ wealth                  <ord> 4, 1, 4, 4, 3, 2, 4, 3, 0, 4, 4, 1, 2,...
$ varied_life             <ord> 4, 4, 3, 4, 4, 3, 4, 4, 3, 3, 5, 3, 4,...
$ exciting_life           <ord> 3, 4, 3, 4, 3, 3, 4, 4, 4, 2, 5, 3, 5,...
$ curious                 <ord> 3, 5, 4, 3, 3, 3, 5, 5, 4, 3, 5, 3, 5,...
$ ask_someone_brand       <ord> 4, 4, 1, 2, 4, 3, 4, 4, 4, 2, 4, 5, 4,...
$ ask_someone_service     <ord> 4, 4, 2, 2, 3, 3, 4, 4, 4, 3, 4, 5, 3,...
$ trust_opinions          <ord> 3, 4, 2, 2, 3, 3, 4, 3, 4, 2, 4, 5, 3,...
$ look_new_products       <ord> 3, 4, 2, 2, 2, 3, 3, 4, 3, 2, 4, 3, 3,...
$ new_experience_products <ord> 3, 3, 3, 2, 3, 3, 3, 4, 3, 2, 3, 3, 3,...
$ visit_places_products   <ord> 3, 4, 3, 2, 3, 3, 4, 4, 3, 4, 4, 3, 2,...
$ sqft_house              <ord> 3, 3, 4, 2, 3, 3, 1, 2, 4, 4, 4, 2, 3,...
$ political_party         <ord> 2, 2, 4, 4, 97, 4, 2, 1, 4, 2, 2, 3, 4...
$ three_people_house      <ord> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, NA...
$ child_under_18          <ord> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, NA...
$ GENDER                  <ord> 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1,...
$ age                     <ord> 4, 3, 4, 4, 4, 3, 2, 1, 4, 4, 3, 1, 2,...
$ education               <ord> 4, 4, 1, 2, 1, 2, 3, 4, 2, 4, 3, 1, 2,...
$ financial_situation     <ord> 2, 2, 1, 1, 2, 1, 2, 3, 1, 1, 1, 3, 1,...
$ INCOME_BINNED           <ord> 3, 4, 95, 4, 3, 95, 3, 1, 4, 95, 95, 1...
$ retired                 <ord> 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0,...
```

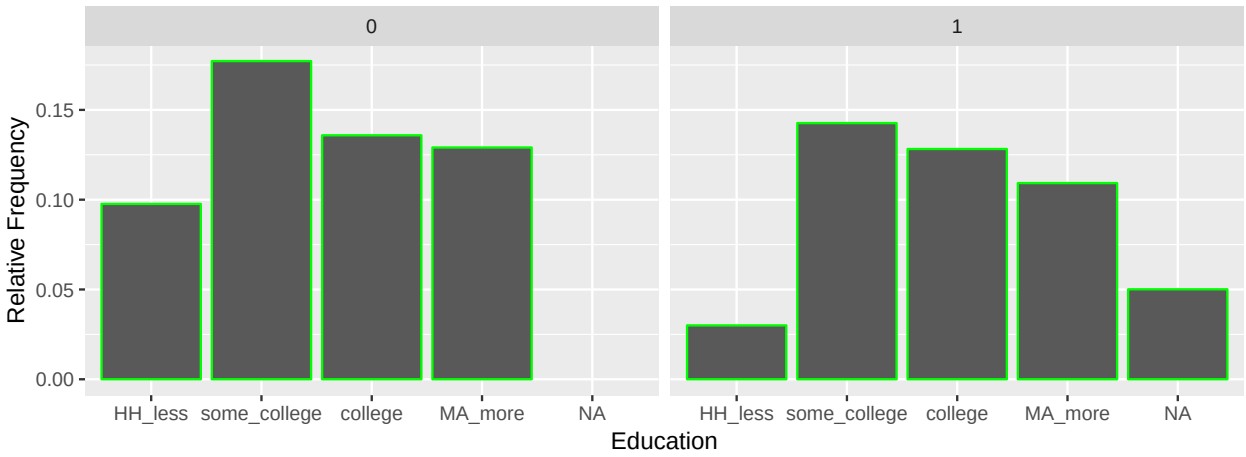
Explaratory Data Analysis (EDA)

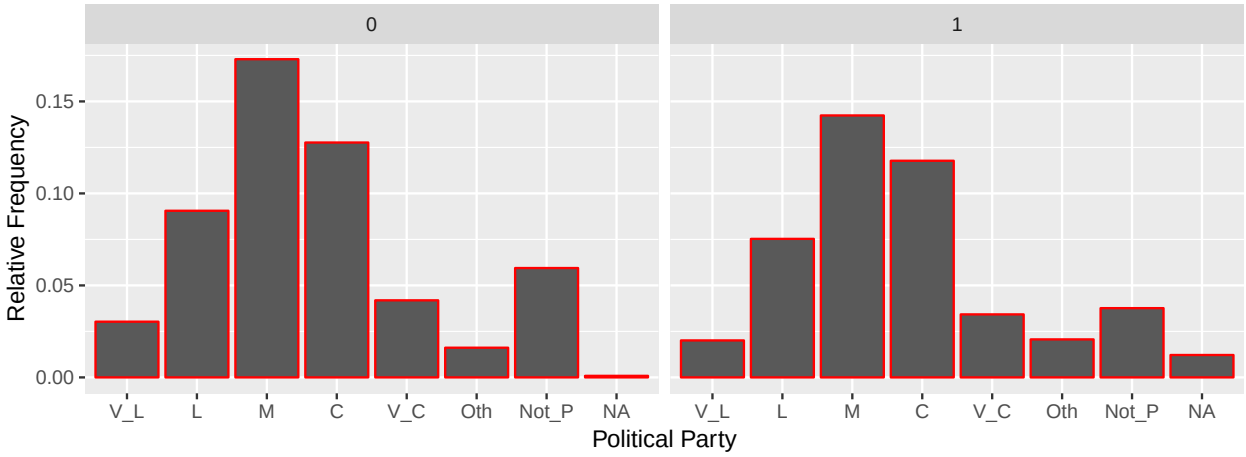
To select features for this project I used visualization techniques, independence tests, and measures of association. Due to the nature of the data being categorical, different tests of independence and association were required depending on whether the variable was nominal or ordinal. Using these strategies I identified 18 variables that were statistically dependent with *HAVESOLAR* and at a minimum weakly associated with it. The variable *HAVESOLAR* is the dependent variable in this research and is equal to 1 if solar was was adopted and equal to 0 otherwise.

Adopeters vs. Non_Adopters by Demographics

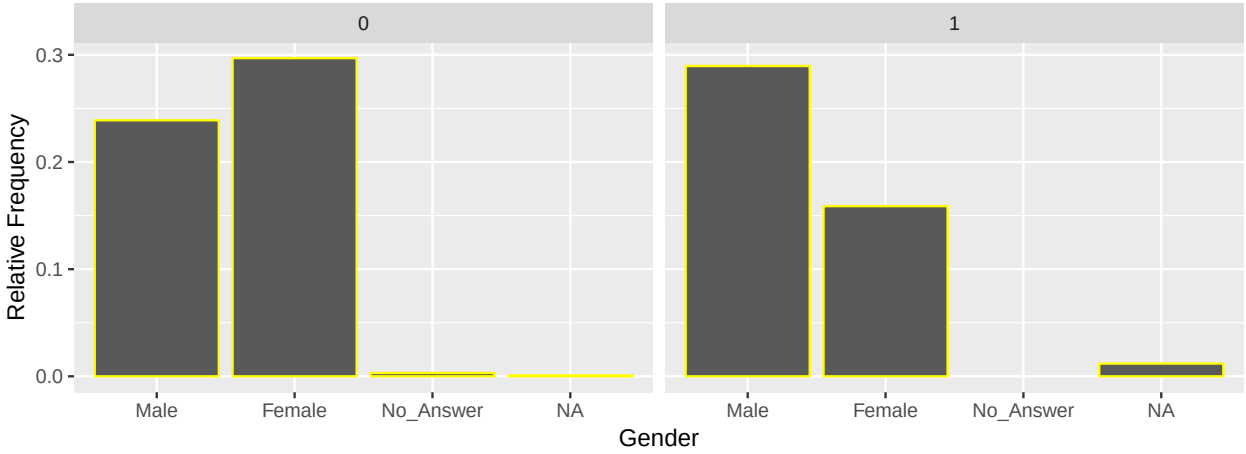
Solar adopters and non-adopters are compared below by demographic factors. The adopters are on the right with a value of 1 and the non-adopters are on the left with a value of 0. Visualizing the two groups by there demographics can reveal insightful information about how the groups are similar and how they differ.

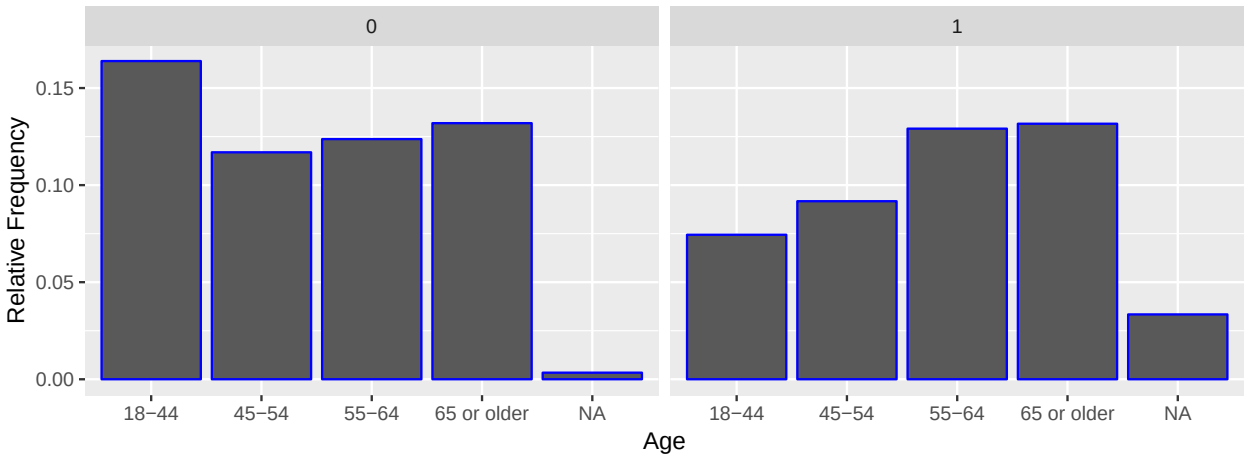
The two groups have very similar distributions with respect to education and political party. This suggests that these factors most likely are not playing a role in whether a consumer adopts solar or not. Though these bar plots are suggestive of not playing a role, we will confirm this with tests of independence and measures of association later in the paper.

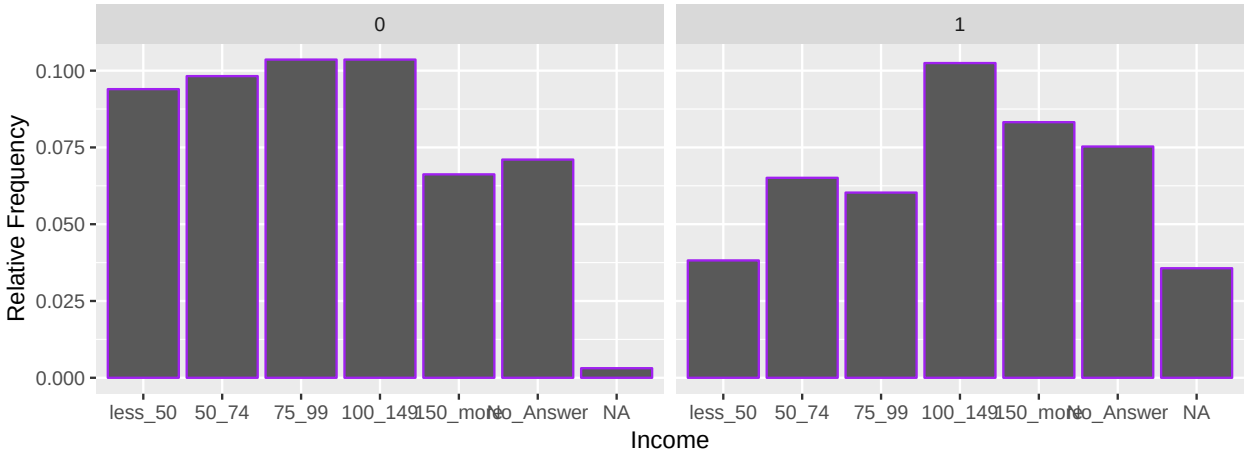


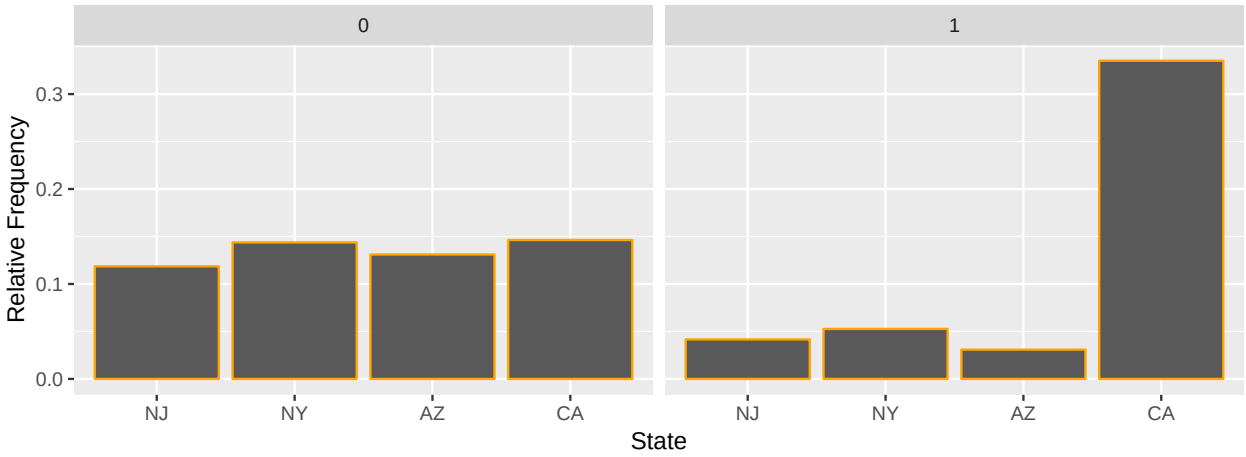


The two groups also display very different distributions with respect to age, gender, income, and the state in which they live. These variables will most likely be more useful as predictors of rooftop solar adoption. We can make some quick observations such as there are more male solar adopters, more younger people who have not adopted solar, more poor people who have not adopted solar, and more people from California who have adopted solar. These demographic differences between these distributions do a good job jumping out you, implying further investigation.







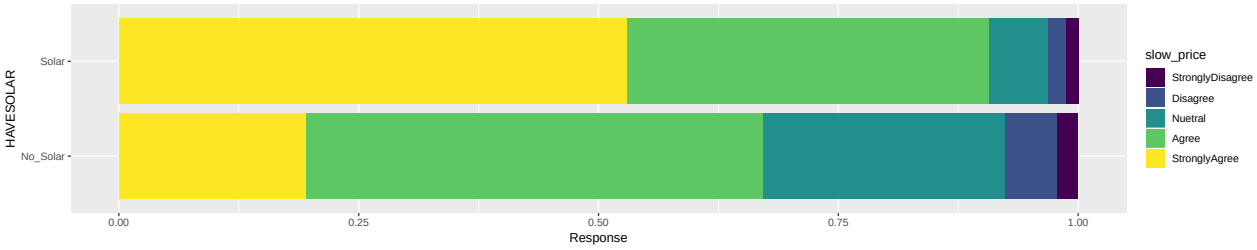


Likert Questions

Another way to get a feel for the data is to visualize survey responses in the form of stacked bar charts. I have subjectively selected three questions to visualize as an example. One question is from the economics category, one from the consumer category, and one from the environmental category. The three survey questions are listed below.

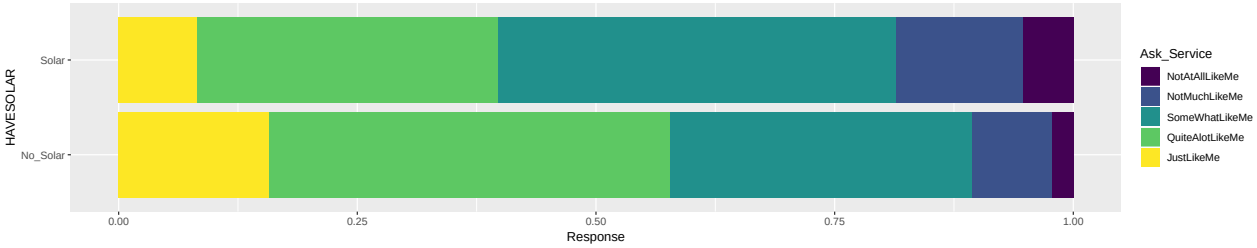
Using solar will help protect my family from rising electricity prices in the future

Here we see that about 55% of people who strongly agree adopted solar while only 20% have not adopted. It appears that there is large differences in the way people feel about solar panel adoption protecting rising energy prices in the future. Characteristics that display large differences should be investigated further.



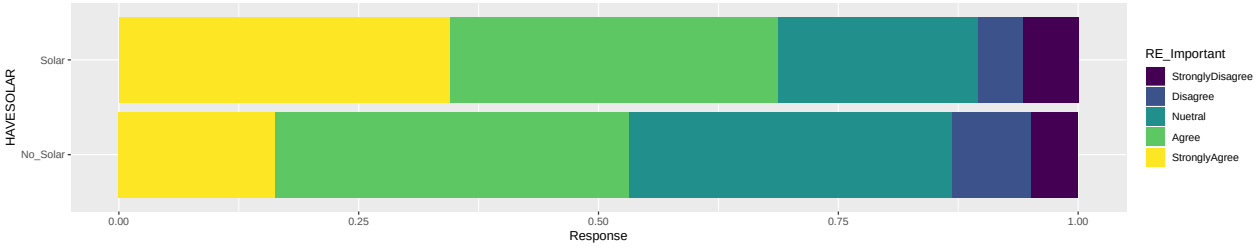
Before I buy a new product or service, I often ask acquaintances about their experiences with that product

This survey question has results that are interesting about consumer behavior. Those who have adopted solar inquire less with others about their experience with that product or service, as compared to non-adopters. We see that 55% of non-adopters responded that this is "Quite Alot Like ME" as compared to 35% of adopters. It could be that those who adopt solar are more self reliant than those who do not.



I feel a personal obligation to do my part to move the country to a renewable energy future.

This question is not surprising and suggests that consumers who are concerned about the environment adopt solar at a higher rate.



Examining survey responses in this fashion help identify which variables have explanatory power over the dependent variable. The three variables above definitely show differences in opinions between adopters and non-adopters.

Visualizing the demographic distributions and survey responses in the form of stacked bar charts is a good starting point to understanding the data, but it can only take us so far. More advanced statistical techniques will be conducted that help us determine what factors are most important for roof top solar adoption.

Contingency Tables

Contingency tables are a great way to display the relationship between two variables. As an example, the variable that had the highest correlation with the response variable, *save_money* to demonstrate the process that took place to select features. The *save_money* variable is a survey response question that asked "Using

solar would save me money". The respondent had the choices Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree. A two way contingency table of *save_money* and *HAVESOLAR* is displayed below.

Table 1: Using solar would save me money

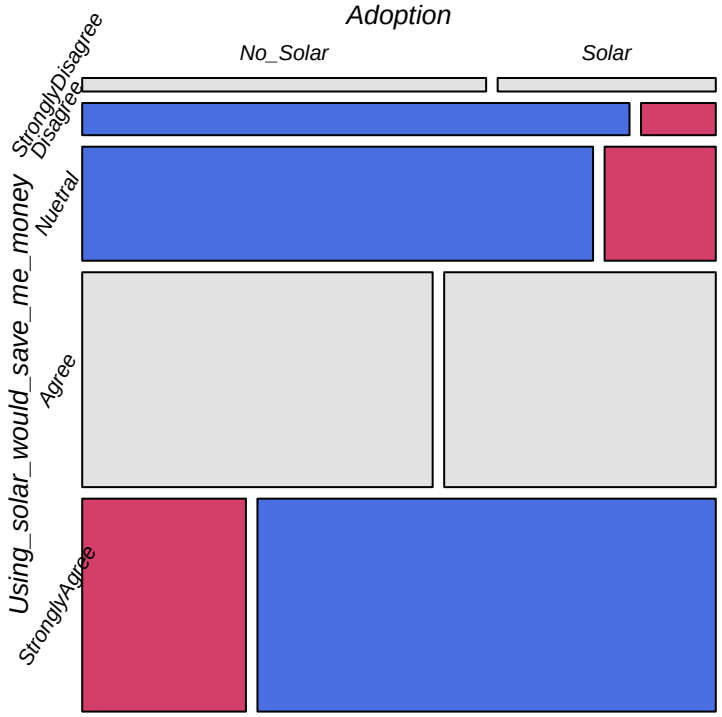
	StronglyDisagree	Disagree	Nuetral	Agree	StronglyAgree
No_Solar	50	161	533	689	319
Solar	27	22	116	534	893

It appears that people who increasingly agree that solar saves money adopt solar at a higher rate. Contingency tables allow the researcher to test the independence between two variables. This is done with independence test such as Pearson's Chi Square for nominal variables and Cochran-Mantel-Haenszel for ordinal variables. The null hypothesis is rejected if the p-value is less than 0.05, implying that there is indeed a statistical relationship between the variables.

Mosaic Plots

Visualizing mosaic plots is a very informative way to look for relationships between categorical variables and determining the direction of the relationship. Mosaic plots represent a contingency table, where the association between two variables can be inspected. Each box represents a conditional relative frequency from a contingency table. A box with color (blue/red) represents positive and negative Pearson's residuals which is the difference between the observed observation and the expected observation. If the box is dull or gray, its telling us that the two groups are probably independent.

Here we can see that individuals who strongly agree own more solar panels than those groups who agree less. There appears to be a positive relationship between beliefs that solar saves money and solar panel ownership. This variable appears to have explanatory power over solar panel adoption, and will be confirmed with independence tests and measures of association.

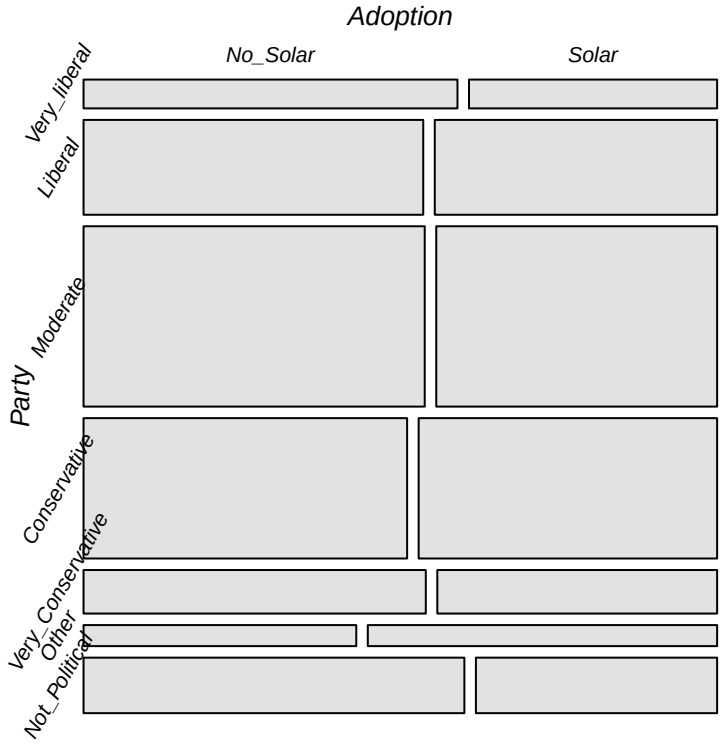


Below is a two way contingency table and mosaic plot of a variable that appears to have no impact over solar panel adoption below. The variable is political party, which was deemed unimportant when we compared

the distributions. We can do this more formally here. It appears that political affiliation has no predictive power over solar panel adoption. Each box in the plot is grey indicating that there is independence between political party and solar panel adoption. For this reason, it most likely will not be included. Again we will perform statistical independence tests to confirm our suspicions.

Table 2: Politcal Party

	Very_liberal	Liberal	Moderate	Conservative	Very_Conservative	Other	Not_Political
No_Solar	107	320	611	451	148	57	210
Solar	71	266	503	416	121	73	133



Tests of Independence and Association

For nominal data the Pearson's Chi-Square to test the independence between groups. Pearson's Chi-Square tests the null hypothesis that the two groups are independent where we reject the null if the p-value is less than 0.05. One draw back is the Chi-Square test is that it is sensitive to sample size. For example, if we increase our sample size the Chi-Square test statistic will also increase. We can use a measure of association called Cramers V, which is not sensitive to sample size, that tells us how strong the relationship between our groups are. Cramers V ranges from 0-1 with values closer to 0 indicative of no association and closer to 1 indicative of a strong association.

With ordinal data there may be a linear trend. The Cochran-Mantel-Haenszel test can be used to test the independence between ordinal variables. We reject the null hypothesis of independence if the p-value is less than 0.05. The value **cor** in the CHMtest is the linear by linear test which ranges from -1 to 1. 0 means that the two groups are independent, while a value of 1 is a perfect positive relationship, and values at -1 is a perfect negative relationship. The GKgamma function in R is a measure of association for ordinal data which tests the strength of relationship as well as its direction. The value of gamma ranges from -1 to 1 with a value of zero indicating no relationship.

We have been looking at the relationship between *HAVESOLAR* and *save_money* which is an ordinal variable. Therefore, the CHMtest to test for independence and the GKgamma the measure of association. Here we reject the null hypothesis of independence as the cor p-value is very small and conclude that there is a strong positive relationship with a gamma value of0.645. This will be a good variable to use when building the model.

```
Cochran-Mantel-Haenszel Statistics for Adoption by Using_solar_would_save_me_money

                 AltHypothesis  Chisq Df        Prob
cor        Nonzero correlation 544.63  1 1.8572e-120
rmeans  Row mean scores differ 544.63  1 1.8572e-120
cmeans  Col mean scores differ 665.54  4 1.0066e-142
general    General association 665.54  4 1.0066e-142

gamma        : 0.645
std. error   : 0.019
CI           : 0.609 0.682
```

The other variable we have been looking at is the political_party variable. This is a nominal variable so I use the Chi-Square test to test independence and Cramers for the measure of association. Here we have a p-value of 0.01046 which is not less than 0.05. We cannot reject the null hypothesis of independence. Cramers V has a value of 0.069 which is very close to 0, meaning there appears to be no association between political party and solar panel adoption.

```
    Pearson's Chi-squared test

data:  political_party_tab
X-squared = 16.697, df = 6, p-value = 0.01046

                    X^2 df P(> X^2)
Likelihood Ratio 16.746  6 0.010265
Pearson          16.697  6 0.010465

Phi-Coefficient   : NA
Contingency Coeff.: 0.069
Cramer's V        : 0.069
```

Feature Selection

The process just described of creating a contingency table, evaluating a mosaic plot, testing for independence, and tests for association was the process used to select features for the model. This process was conducted for all features in the data set. In the end, 18 variables were hand selected that will be used to build the model. In the Appendix, I have presented all contingency tables, mosaic plots, and the statistical tests for the variables that will be used in the modeling process. Below is a list of the four categories of variables included in the analysis. Each variable lists the original name from the data set, the new name used in the analysis, and what type of variable it is i.e. ordinal or nominal.

Demographic Variables

- GENDER - GENDER, nominal
- AGE_BINNED - age, ordinal
- STATE - STATE, nominal
- INCOME_BINNED, ordinal

Economic Variables

- WINTER_NOPV_BINNED - winter_bill, ordinal
- SUMMER_NOPV_BINNED - summer_bill, ordinal
- BTE8 - slow_energy_price, ordinal
- BE13 - return_investment, ordinal
- BE10 - save_money, ordinal, ordinal

Consumer Variables

- CIJM1 - ask_someone_brand, ordinal
- CIJM2 - ask_someone_service, ordinal
- CIJM3 - trust_opinions, ordinal
- CNS1 - look_new_products, ordinal
- CNS2 - new_experience_products, ordinal
- CNS3 - visit_places_products, ordinal

Environmental Variables

- PN1 - renewable_energy, ordinal
- BB1 - environment_improve, ordinal
- BB2 - slow_climate_change, ordinal
- BB3 - reduce_footprint, ordinal

Model

Baseline Method

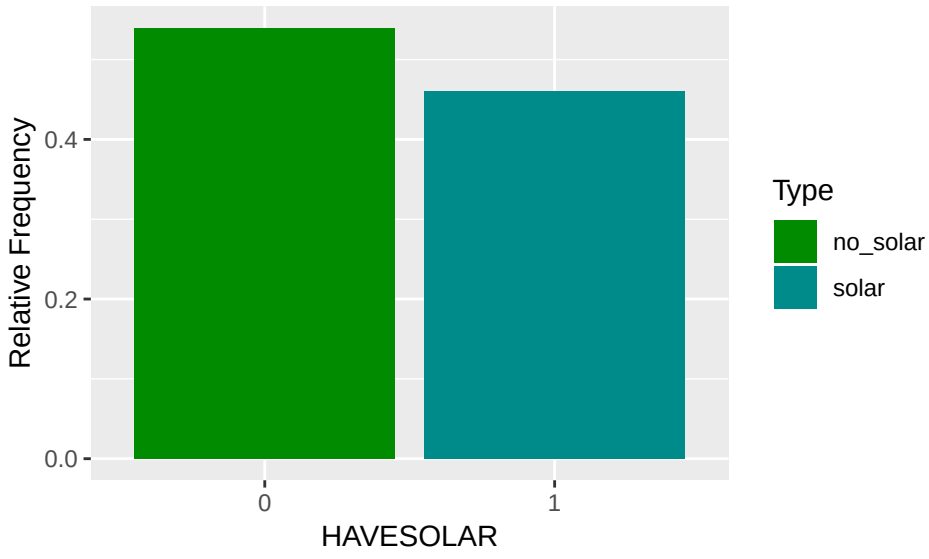
First lets find the standard base line method. For a classification problem, we predict the most frequent outcome for all observations. We see here that the base line model has an accuracy of 54 percent. A logistic regression model will be built in attempt to beat the base line prediction.

```
Adoption
No_Solar    Solar
    1907     1626
```

```
[1] 0.5397679
```

Dependent Variable

The graph below displays the distribution of the dependent variable *HAVESOLAR*. The sample is slightly unbalanced and gives us a visual of the baseline method. When we split the data into the training and test data sets, each data set will maintain this distribution of the dependent variable. In other words, both the training set and the test set will have 54% non-adopters and 46% adopters.



Training and Test Data

Next we create training and test data sets. The sample.split() function will be used which is part of the catools package in R. The first argument is the dependent variable, and second argument is the percentage of data we want in the training set. This also makes sure that the outcome variable is well balanced, as mentioned previously. The training data will consist of 75% of the original data with the remaining 25% saved for the test data.

Here we view the two data sets and notice that they both have the same number of columns/variables but differ in row counts. We will use the training data to build the model, and then see how our model predicts solar adoption on the test data.

Training data

```
Observations: 1,886
Variables: 20
$ X1                      <int> 1, 2, 6, 8, 14, 15, 18, 19, 21, 22, 23...
$ HAVESOLAR               <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
$ INCOME_BINNED           <int> 3, 4, 95, 1, 4, 3, 95, 3, 5, 5, 2, 5, ...
$ GENDER                  <int> 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0,...
$ age                     <int> 4, 3, 3, 1, 4, 3, 2, 1, 1, 1, 1, 4, 4,...
$ STATE                   <int> 1, 4, 1, 4, 2, 4, 3, 1, 4, 3, 1, 3, 4,...
$ winter_bill             <int> 6, 7, 7, 8, 6, 3, 5, 3, 7, 7, 3, 5, 8,...
$ summer_bill             <int> 4, 8, 8, 7, 6, 4, 5, 4, 5, 7, 6, 7, 9,...
$ slow_energy_price       <int> 2, 4, 1, 4, 4, 3, 3, 4, 4, 5, 5, 2, 3,...
$ return_investment       <int> 3, 4, 1, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3,...
$ save_money              <int> 3, 5, 2, 3, 4, 3, 3, 4, 4, 5, 4, 4, 4,...
$ ask_someone_brand       <int> 4, 4, 3, 4, 4, 5, 3, 4, 4, 5, 5, 1, 4,...
$ ask_someone_service     <int> 4, 4, 3, 4, 5, 5, 3, 4, 4, 5, 4, 1, 4,...
$ trust_opinions          <int> 3, 4, 3, 3, 3, 3, 3, 4, 4, 4, 4, 1, 4,...
$ look_new_products       <int> 3, 4, 3, 4, 4, 4, 3, 3, 4, 4, 4, 5, 4,...
$ new_experience_products <int> 3, 3, 3, 4, 4, 4, 3, 4, 4, 4, 2, 5, 3,...
$ visit_places_products   <int> 3, 4, 3, 4, 4, 4, 3, 4, 4, 4, 3, 3, 4,...
$ renewable_energy        <int> 3, 5, 1, 4, 4, 4, 3, 3, 4, 5, 2, 4, 4,...
$ slow_climate_change     <int> 4, 5, 1, 4, 3, 3, 4, 3, 4, 3, 3, 3, 4,...
$ reduce_footprint        <int> 3, 5, 2, 3, 4, 3, 4, 4, 4, 4, 4, 4, 4,...
```

Test Data

```
Observations: 629
Variables: 20
$ X1                      <int> 4, 5, 11, 32, 35, 37, 40, 54, 56, 57, ...
$ HAVESOLAR               <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
$ INCOME_BINNED           <int> 4, 3, 95, 4, 3, 95, 4, 4, 2, 95, 1, 2,...
$ GENDER                  <int> 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1,...
$ age                     <int> 4, 4, 3, 3, 3, 4, 4, 2, 1, 4, 4, 2, 2,...
$ STATE                   <int> 4, 4, 3, 4, 3, 2, 3, 3, 4, 4, 3, 4, 3,...
$ winter_bill             <int> 3, 4, 6, 4, 2, 11, 4, 2, 6, 5, 3, 4, 4...
$ summer_bill             <int> 2, 3, 10, 3, 2, 11, 7, 3, 7, 5, 6, 5, ...
$ slow_energy_price       <int> 3, 4, 4, 4, 3, 3, 2, 3, 3, 2, 4, 4, 3,...
$ return_investment       <int> 2, 3, 4, 3, 3, 4, 1, 3, 2, 3, 3, 3, 3,...
$ save_money              <int> 3, 4, 4, 4, 3, 3, 2, 4, 3, 3, 3, 4, 2,...
$ ask_someone_brand       <int> 2, 4, 4, 3, 5, 3, 4, 3, 3, 4, 4, 5, 2,...
$ ask_someone_service     <int> 2, 3, 4, 3, 5, 3, 3, 3, 3, 4, 4, 5, 2,...
$ trust_opinions          <int> 2, 3, 4, 4, 3, 4, 4, 4, 4, 3, 3, 5, 3,...
$ look_new_products       <int> 2, 2, 4, 3, 3, 2, 4, 2, 4, 1, 2, 5, 1,...
$ new_experience_products <int> 2, 3, 3, 3, 2, 2, 3, 2, 3, 1, 2, 4, 1,...
$ visit_places_products   <int> 2, 3, 4, 4, 4, 3, 3, 3, 4, 1, 2, 4, 2,...
$ renewable_energy        <int> 3, 4, 3, 3, 4, 3, 3, 3, 3, 4, 4, 3, 3,...
$ slow_climate_change     <int> 3, 3, 5, 4, 3, 3, 2, 2, 1, 3, 4, 2, 3,...
$ reduce_footprint        <int> 3, 4, 4, 4, 4, 4, 2, 3, 2, 3, 4, 4, 4,...
```

Logistic Regression

Logistic regression predicts the probability of an outcome variable being true. The logistic regression model will predict the probability that the consumer has adopted solar i.e. P(y=1). P(y=0) = 1 - P(y=1) where the P(y=0) is the probability the consumer has not adopted solar. Positive parameter estimates are predictive of class 1, while negative values are predictive of class 0. For example, a positive value increases the probability that HAVESOLAR = 1 or that a consumer has adopted solar.

The first model will include the 18 selected features from the data set. I include all selected features in the first model since we rejected the null hypothesis of independence for each one and the dependent variable *HAVESOLAR*. The models are located in the appendix due to their length and you can view them here **Appendix**. After running the first model, the goal will be to check the model for multicollinearity and then perform statistical significance test to determine which variables should be removed from the model. In the process of doing this, the AIC value will be evaluated.

Detecting Multicollinarity

A VIF of 1 means that there is no correlation among the predictor and the remaining predictor variables, and hence the variance is not inflated at all. The general rule of thumb is that VIFs exceeding 4 warrant further investigation, while VIFs exceeding 10 are signs of serious multicollinearity requiring a correction. As you can see below, there are many variables that have a VIF greater than 4 and some greater than 10. This is not surprising as many of the variables where survey questions that were very similar. Its not hard to image that some of these variables are strongly correlated. Coefficients with VIF values over 10 will be removed. Ideally I would like to get the VIF values below 4 without significantly increasing the AIC value.

In table 3 we can see there are three VIF values well over 10. The three variables are from the consumer category. It makes sense to keep the one that has the strongest measure of association with the dependent variable, so two will be removed. It appears that the *ask_someone_service* variable is more correlated than the *ask_someone_brand* and *new_experience_product* variables, therefore I will remove the latter two. The AIC value in model 1 is 1543, and in model 2 where the *ask_someone_brand* and *new_experience_product* variable are removed, the AIC value remained unchanged. Therefore, this is a significant improvement to the model. You can compare model 1 and model 2 in the **Appendix**.

Table 3: VIF Values

	GVIF	Df	GVIF^(1/(2*Df))
ordered(INCOME_BINNED)	1.584449	5	1.047099
factor(GENDER)	1.164484	2	1.038803
ordered(age)	1.438599	3	1.062486
factor(STATE)	1.508333	3	1.070901
ordered(winter_bill)	4.266467	10	1.075235
ordered(summer_bill)	4.541089	10	1.078594
ordered(slow_energy_price)	7.273596	4	1.281500
ordered(return_investment)	5.663080	4	1.242029
ordered(save_money)	8.066370	4	1.298180
ordered(ask_someone_brand)	8.757583	4	1.311590
ordered(ask_someone_service)	10.777963	4	1.346068
ordered(trust_opinions)	5.499642	4	1.237490
ordered(look_new_products)	11.008798	4	1.349639
ordered(new_experience_products)	12.326236	4	1.368844
ordered(visit_places_products)	5.817539	4	1.246213
ordered(renewable_energy)	5.603086	4	1.240376
ordered(slow_climate_change)	4.488738	4	1.206467
ordered(reduce_footprint)	6.991003	4	1.275168

We now see all the VIF values are well below 10 so we have definitely removed multicollinearity from the model. We can attempt to further remove variables that look suspicious and compare the models. We go through two more series where we remove variables until the VIF values are below 4. During this process we generated model 3 in the **Appendix**. Model 3 had the largest jump in AIC value increasing to 1583. The final VIF values are in table 4, where all variables have a VIF below 4.

Table 4: VIF Values

	GVIF	Df	GVIF^(1/(2*Df))
ordered(INCOME_BINNED)	1.291246	5	1.025890
factor(GENDER)	1.123317	2	1.029498
ordered(age)	1.274501	3	1.041254
factor(STATE)	1.198638	3	1.030658
ordered(winter_bill)	1.478723	10	1.019751
ordered(save_money)	1.602754	4	1.060739
ordered(ask_someone_service)	1.531109	4	1.054692
ordered(look_new_products)	1.634632	4	1.063353
ordered(renewable_energy)	2.837969	4	1.139268
ordered(slow_climate_change)	2.849556	4	1.139849

Wald Test

We can test for the overall significance for each categorical variable using the Wald Test. In the table below we see that the only variable that is insignificant is the *slow_climate_change* variable.

```
Analysis of Deviance Table (Type II tests)

Response: factor(HAVESOLAR)
                          Df    Chisq Pr(>Chisq)
ordered(INCOME_BINNED)     5  31.5435  7.315e-06 ***
```

```
factor(GENDER)                2  29.0354  4.955e-07 ***
ordered(age)                  3  21.7349  7.406e-05 ***
factor(STATE)                 3 201.3586  < 2.2e-16 ***
ordered(winter_bill)         10  40.1665  1.584e-05 ***
ordered(save_money)           4 172.6559  < 2.2e-16 ***
ordered(ask_someone_service)  4  45.2165  3.584e-09 ***
ordered(look_new_products)    4  48.0508  9.210e-10 ***
ordered(renewable_energy)     4  23.5378  9.885e-05 ***
ordered(slow_climate_change)  4   7.0474     0.1334
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After removing the variable *slow_climate_change* from the model we again use the Wald Test and find that all of our categorical predictors are statistically significant. This new set of variables will generate model 4 in the **Appendix**.

```
Analysis of Deviance Table (Type II tests)

Response: factor(HAVESOLAR)
                            Df   Chisq Pr(>Chisq)
ordered(INCOME_BINNED)       5  28.280  3.209e-05 ***
factor(GENDER)               2  28.588  6.197e-07 ***
ordered(age)                 3  22.915  4.206e-05 ***
factor(STATE)                3 182.219  < 2.2e-16 ***
ordered(winter_bill)        10  38.896  2.648e-05 ***
ordered(slow_energy_price)   4  21.280  0.0002786 ***
ordered(save_money)          4  97.865  < 2.2e-16 ***
ordered(ask_someone_service) 4  32.691  1.382e-06 ***
ordered(look_new_products)   4  11.345  0.0229523 *
ordered(visit_places_products) 4  20.523  0.0003936 ***
ordered(renewable_energy)    4  32.001  1.913e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For simplicity, the rest of the report will use model 4 as it is the simplest model, has all statistically significant predictors, has no multicollinearity, and its AIC value is only slightly higher than models 1 and 2.

Model Evaluation and Prediction Performance

Summary of Predicted Values

Lets see if we are predicting higher probability for HAVESOLAR = 1, and lower probability for HAVESOLAR = 0. We are predicting an average probability of 0.72 for HAVESOLAR = 1 and 0.22 for HAVESOLAR = 0. This a good sign because we are predicting a higher probability for the actual HAVESOLAR = 1 cases.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00000 0.09378 0.38159 0.44168 0.80174 0.99321
```

```
        0         1
0.2195201 0.7225034
```

Confussion Matrix

Next we will evaluate a confusion matrix. A confusion matrix compares actual outcomes to predicted outcomes. The classification matrix is composed of true negative (TN) cases, true positive (TP) cases, false negative(FN)

cases, and false positive (FP) cases. TN cases predict $HAVESOLAR = 1$ and the case is $HAVESOLAR = 0$, TP cases predict $HAVESOLAR = 1$ and the case is $HAVESOLAR = 1$, FN cases predict $HAVESOLAR = 0$ and the case is $HAVESOLAR = 0$, and FP predict $HAVESOLAR = 0$ and the case is $HAVESOLAR = 1$.

In order to build a confusion matrix we have to choose a threshold value **t**. Since our predictions are between 0 and 1 the threshold value **t** should be between 0 and 1. Values great than **t** will predict **HAVESOLAR** = 1 and values below **t** will predict $HAVESOLAR = 0$. Below is the confusion matrix. The rows are the actual values of *HAVESOLAR* and the columns are the predicted values of *HAVESOLAR*. As you can see the first row and column of the matrix is the TP rate, the number of cases we predicted $HAVESOLAR = 1$ where the case is $HAVESOLAR = 1$.

Table 5: Predict HAVESOLAR, t = 0.5

	Predicted = 0	Predicted = 1
Actual = 0	908	145
Actual = 1	165	668

From the confusion matrix we can calculate various measures. I will focus on sensitivity, specificity, and over all accuracy.

sensitivity - the percentage of TRUE cases classified correctly

```
[1] 0.8019208
```

specificity - the percentage of FALSE cases classified correctly

```
[1] 0.8622982
```

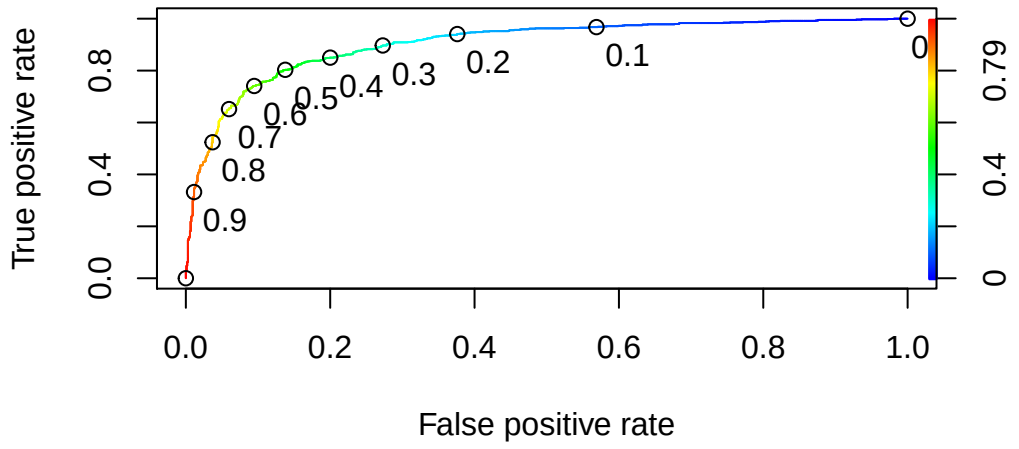
overall accuracy - the percentage of overall cases classified correctly

```
[1] 0.835631
```

The model with a threshold of $\mathbf{t} = 0.5$ has an overall all accuracy of 84%, with a sensitivity of 80%, and a specificity of 86%. Which threshold should we choose for our predictions? In the next section we cover the ROC curve which can help a researcher determine the value of **t**.

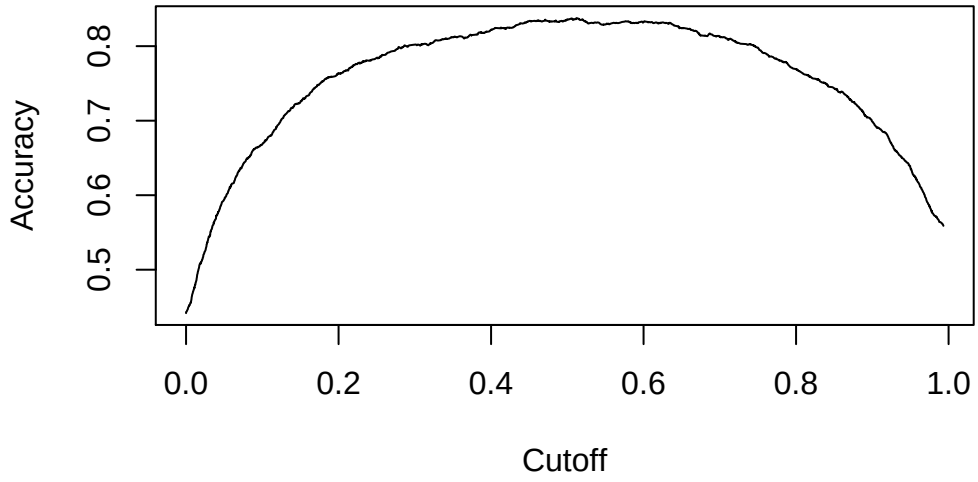
ROC Curve

The ROC curve has a true-positive rate on the y-axis and a false positive rate on the x-axis. $\mathbf{t} = 1$ is at the point (0,0) while $\mathbf{t} = 0$ is at the point (1,1). The ROC curve is a performance measure of a classification model at all thresholds **t**. This provides information to the researcher in selecting a threshold **t** depending on the preferences for specificity and sensitivity measures. In our case we are not concerned with having a specific sensitivity or specificity rate so we will use another method to choose our **t** value. We will however use the ROC to calculate the AUC value.



Calculating the Threshold

We will use a measure that maximizes the overall accuracy. R allows us to calculate the accuracy, the cutoff, and the ability to visualize this. Below is a plot to help us determine our threshold value **t**.




```
##     accuracy cutoff.1228
##    0.8377519   0.5121446
```

Table 6: Predict HAVESOLAR, t = 0.51

	Predicted = 0	Predicted = 1
Actual = 0	917	136
Actual = 1	170	663

AUC

The AUC provides an aggregate measure of performance across all possible classification thresholds. It measures the probability that a random TRUE value, in our case $HAVESOLAR = 1$, is to the right of a random NEGATIVE value $HAVESOLAR = 0$. We have an AUC of 0.9011, so 90% of the time our random true value is to the right of our random negative value. A simpler interpretation is that when the AUC is 0 we have predicted none of our values, and when the AUC is 1 we predict 100 % of our values correctly.

```
Area under the curve: 0.9043
```

Using Model on Test Set

Summary of Predicted Values

We first compare the statistical summary of the predicted results from the training data and the test data, using model 4. We see that the measures are very close indicating our model is doing a good job of predicting *HAVESOLAR* in the test data.

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.0000079 0.0830335 0.3483680 0.4355143 0.8089502 0.9957319
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00000 0.09378 0.38159 0.44168 0.80174 0.99321
```

Confusion Matrix

Lets next examine the confusion matrix as we did with the training data. We see that our model is performing well on the test data with an overall accuracy of 82%.

Table 7: Predict HAVESOLAR, t = 0.5

	Predicted = 0	Predicted = 1
Actual = 0	297	54
Actual = 1	60	218

sensitivity - the percentage of TRUE cases classified correctly

```
[1] 0.8461538
```

specificity - the percentage of FALSE cases classified correctly

```
[1] 0.7841727
```

overall accuracy - the percentage of overall cases classified correctly

```
[1] 0.8187599
```

Test Set Precition Accuracy

We have an AUC of 89% so our model seems to be performing well. The model is predicting 89% of the *HAVESOLAR* = 1 cases correctly.

```
Area under the curve: 0.8938
```

Reccomendations

It is apparent that economic factors play the strongest role in rooftop solar adoption. The final model had the two predictors **save_money** and **slow_energy_price** which had the strongest correlation with the dependent variable *HAVESOLAR*. Each had a positive relationship with rooftop solar adoption. Adopters believe that rooftop solar saves them money and protects them from rising energy prices in the future. Solar companies should continue to educate potential customers about the economic benefits of adopting rooftop solar. This can be done through advertising, social media, information on websites etc.

Predictors from the consumer category **included ask_some_service**, **look_new_products**, and **visit_places_products**. Surprisingly these all shared a negative relationship with *HAVESOLAR*. For example, those who described themselves as "not like me at all" for the question "I continually look for new

products" adopted solar more than those who described themselves as "Just like me". The same goes for **ask_someone_service** which asked "Do you ask others about their experiences with products before you purchase the product" and **visit_places_product** which asked "do you visit places to find new products". This implies that solar adopters are not actively looking or seeking new products. Potential solar adopters may be harder to reach through traditional advertising strategies. Efforts to reach these customers may be essential to increase solar adoption and more research should be done by solar companies to understand this behavior.

The two variables *summer_bill* and *winter_bill* where positively associated with solar adoption. Solar companies should target areas where electricity bills are higher than average and solar generation potential is high. This can be done with increased advertising and increased development to create more exposure to potential solar panel consumers. Electricity bills are a important topic because consumers find it easy to understand savings when using this information. Information should be made available about how consumers will lower monthly utility bills when adopting solar.

Environmental reasons also play a factor. The most highly correlated factor from the environmental category was **renewable_energy**, though factors where correlated with *HAVESOLAR**. Solar companies should maintain an image of being green. This may encourage those who already know of the financial benefits to ultimately make the switch. It may be that consumers need more than one reason that is important to them when making such a large investment.

Some demographics are important while others are not. Surprisingly education levels and political party played no role in solar adoption. Other demographic factors did however play a role. For example, men adopted solar more than females, California adopted solar more than the other states, older aged consumers adopted solar more than younger consumers, and consumers with higher incomes adopted more solar than lower income consumers. When developing marketing strategies keep these demographic factors in mind.

Appendix

Model Summary

Model

```
Calls:
Model 1: glm(formula = factor(HAVESOLAR) ~ ordered(INCOME_BINNED) + factor(GENDER) +
    ordered(age) + factor(STATE) + ordered(winter_bill) + ordered(summer_bill) +
    ordered(slow_energy_price) + ordered(return_investment) +
    ordered(save_money) + ordered(ask_someone_brand) + ordered(ask_someone_service) +
    ordered(trust_opinions) + ordered(look_new_products) + ordered(new_experience_products) +
    ordered(visit_places_products) + ordered(renewable_energy) +
    ordered(slow_climate_change) + ordered(reduce_footprint),
    family = "binomial", data = dfTrain)
Model 2: glm(formula = factor(HAVESOLAR) ~ ordered(INCOME_BINNED) + factor(GENDER) +
    ordered(age) + factor(STATE) + ordered(winter_bill) + ordered(summer_bill) +
    ordered(slow_energy_price) + ordered(return_investment) +
    ordered(save_money) + ordered(ask_someone_service) + ordered(trust_opinions) +
    ordered(look_new_products) + ordered(visit_places_products) +
    ordered(renewable_energy) + ordered(slow_climate_change) +
    ordered(reduce_footprint), family = "binomial", data = dfTrain)
Model 3: glm(formula = factor(HAVESOLAR) ~ ordered(INCOME_BINNED) + factor(GENDER) +
    ordered(age) + factor(STATE) + ordered(winter_bill) + ordered(save_money) +
    ordered(ask_someone_service) + ordered(look_new_products) +
    ordered(renewable_energy) + ordered(slow_climate_change),
```

```
    family = "binomial", data = dfTrain)
Model 4: glm(formula = factor(HAVESOLAR) ~ ordered(INCOME_BINNED) + factor(GENDER) +
    ordered(age) + factor(STATE) + ordered(winter_bill) + ordered(slow_energy_price) +
    ordered(save_money) + ordered(ask_someone_service) + ordered(look_new_products) +
    ordered(visit_places_products) + ordered(renewable_energy),
    family = "binomial", data = dfTrain)
```

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-1.739***	-1.721***	-1.672***	-1.558***
	(0.282)	(0.279)	(0.244)	(0.251)
ordered(INCOME_BINNED): .L	0.645**	0.643***	0.821***	0.758***
	(0.197)	(0.193)	(0.183)	(0.186)
ordered(INCOME_BINNED): .Q	-0.269	-0.199	-0.194	-0.206
	(0.191)	(0.187)	(0.177)	(0.180)
ordered(INCOME_BINNED): .C	-0.189	-0.135	-0.108	-0.144
	(0.178)	(0.175)	(0.166)	(0.168)
ordered(INCOME_BINNED): ^4	-0.300	-0.277	-0.317*	-0.306
	(0.166)	(0.165)	(0.156)	(0.159)
ordered(INCOME_BINNED): ^5	0.364*	0.373*	0.345*	0.348*
	(0.163)	(0.162)	(0.152)	(0.154)
factor(GENDER): 1/0	-0.729***	-0.712***	-0.752***	-0.755***
	(0.149)	(0.147)	(0.139)	(0.141)
factor(GENDER): 95/0	-11.656	-11.919	-12.228	-12.497
	(403.392)	(399.798)	(380.547)	(382.805)
ordered(age): .L	0.677***	0.690***	0.576***	0.623***
	(0.149)	(0.148)	(0.138)	(0.140)
ordered(age): .Q	-0.136	-0.128	-0.209	-0.183
	(0.143)	(0.141)	(0.133)	(0.135)
ordered(age): .C	0.063	0.084	0.117	0.057
	(0.143)	(0.141)	(0.134)	(0.136)
factor(STATE): 2/1	0.201	0.215	-0.007	-0.033
	(0.250)	(0.247)	(0.231)	(0.235)
factor(STATE): 3/1	-0.295	-0.322	-0.005	-0.069
	(0.264)	(0.261)	(0.244)	(0.249)
factor(STATE): 4/1	1.941***	1.939***	1.904***	1.824***
	(0.212)	(0.210)	(0.196)	(0.200)
ordered(winter_bill): .L	0.053	0.053	1.562***	1.527***
	(0.551)	(0.548)	(0.455)	(0.462)
ordered(winter_bill): .Q	0.351	0.401	0.345	0.562
	(0.494)	(0.493)	(0.431)	(0.437)
ordered(winter_bill): .C	-0.542	-0.482	-0.449	-0.556
	(0.438)	(0.438)	(0.397)	(0.400)
ordered(winter_bill): ^4	0.258	0.274	0.390	0.355
	(0.402)	(0.402)	(0.377)	(0.383)
ordered(winter_bill): ^5	-0.837*	-0.846*	-0.663	-0.767*
	(0.383)	(0.385)	(0.363)	(0.373)
ordered(winter_bill): ^6	-0.154	-0.192	-0.071	-0.103
	(0.348)	(0.348)	(0.330)	(0.340)
ordered(winter_bill): ^7	-0.306	-0.341	-0.279	-0.312
	(0.297)	(0.296)	(0.279)	(0.287)
ordered(winter_bill): ^8	0.260	0.247	0.283	0.242
	(0.260)	(0.257)	(0.242)	(0.247)

ordered(winter_bill): ^9	0.139	0.103	0.008	-0.047
	(0.227)	(0.225)	(0.211)	(0.215)
ordered(winter_bill): ^10	-0.114	-0.112	-0.231	-0.232
	(0.188)	(0.186)	(0.174)	(0.177)
ordered(summer_bill): .L	2.120***	2.202***		
	(0.464)	(0.458)		
ordered(summer_bill): .Q	0.593	0.606		
	(0.402)	(0.397)		
ordered(summer_bill): .C	-0.173	-0.167		
	(0.357)	(0.353)		
ordered(summer_bill): ^4	0.264	0.260		
	(0.334)	(0.332)		
ordered(summer_bill): ^5	-0.348	-0.337		
	(0.332)	(0.329)		
ordered(summer_bill): ^6	0.119	0.143		
	(0.306)	(0.302)		
ordered(summer_bill): ^7	0.024	0.026		
	(0.268)	(0.266)		
ordered(summer_bill): ^8	0.152	0.167		
	(0.241)	(0.238)		
ordered(summer_bill): ^9	-0.434	-0.438*		
	(0.223)	(0.221)		
ordered(summer_bill): ^10	-0.235	-0.278		
	(0.193)	(0.191)		
ordered(slow_energy_price): .L	-0.299	-0.283		-0.116
	(0.600)	(0.591)		(0.531)
ordered(slow_energy_price): .Q	0.913	0.915*		1.008*
	(0.467)	(0.461)		(0.413)
ordered(slow_energy_price): .C	-0.202	-0.188		-0.241
	(0.335)	(0.330)		(0.308)
ordered(slow_energy_price): ^4	-0.369	-0.381		-0.390
	(0.258)	(0.255)		(0.244)
ordered(return_investment): .L	1.011*	0.775		
	(0.446)	(0.435)		
ordered(return_investment): .Q	-0.083	0.001		
	(0.356)	(0.349)		
ordered(return_investment): .C	0.110	0.098		
	(0.273)	(0.268)		
ordered(return_investment): ^4	0.067	0.045		
	(0.186)	(0.183)		
ordered(save_money): .L	1.459**	1.528**	2.098***	2.059***
	(0.535)	(0.529)	(0.327)	(0.478)
ordered(save_money): .Q	1.025*	0.929*	1.259***	0.926*
	(0.414)	(0.408)	(0.291)	(0.377)
ordered(save_money): .C	-0.478	-0.462	-0.504	-0.384
	(0.347)	(0.343)	(0.296)	(0.320)
ordered(save_money): ^4	0.007	0.029	-0.046	0.011
	(0.263)	(0.261)	(0.240)	(0.248)
ordered(ask_someone_brand): .L	-0.049			
	(0.456)			
ordered(ask_someone_brand): .Q	-0.036			
	(0.370)			
ordered(ask_someone_brand): .C	0.185			
	(0.272)			

ordered(ask_someone_brand): ^4	0.429*			
	(0.175)			
ordered(ask_someone_service): .L	-0.501	-0.612	-1.206***	-0.939**
	(0.446)	(0.388)	(0.289)	(0.302)
ordered(ask_someone_service): .Q	-0.532	-0.526	-0.330	-0.344
	(0.360)	(0.315)	(0.253)	(0.266)
ordered(ask_someone_service): .C	0.394	0.463*	0.343	0.425*
	(0.265)	(0.234)	(0.199)	(0.206)
ordered(ask_someone_service): ^4	-0.177	-0.070	-0.044	-0.075
	(0.172)	(0.158)	(0.144)	(0.147)
ordered(trust_opinions): .L	-0.690	-0.756		
	(0.480)	(0.453)		
ordered(trust_opinions): .Q	0.540	0.542		
	(0.380)	(0.358)		
ordered(trust_opinions): .C	-0.228	-0.226		
	(0.279)	(0.270)		
ordered(trust_opinions): ^4	0.132	0.226		
	(0.179)	(0.175)		
ordered(look_new_products): .L	-0.537	-1.003***	-1.518***	-0.840**
	(0.362)	(0.294)	(0.233)	(0.281)
ordered(look_new_products): .Q	-0.377	-0.213	-0.210	-0.247
	(0.261)	(0.229)	(0.196)	(0.220)
ordered(look_new_products): .C	-0.021	-0.045	-0.063	-0.018
	(0.199)	(0.181)	(0.160)	(0.175)
ordered(look_new_products): ^4	-0.073	-0.122	-0.010	-0.066
	(0.147)	(0.137)	(0.127)	(0.131)
ordered(new_experience_products): .L	-0.795*			
	(0.397)			
ordered(new_experience_products): .Q	0.220			
	(0.279)			
ordered(new_experience_products): .C	-0.083			
	(0.204)			
ordered(new_experience_products): ^4	-0.153			
	(0.145)			
ordered(visit_places_products): .L	-0.967**	-1.098**		-1.290***
	(0.348)	(0.334)		(0.316)
ordered(visit_places_products): .Q	-0.116	-0.136		0.044
	(0.272)	(0.263)		(0.251)
ordered(visit_places_products): .C	-0.080	-0.060		-0.105
	(0.201)	(0.196)		(0.188)
ordered(visit_places_products): ^4	0.111	0.112		0.146
	(0.141)	(0.139)		(0.134)
ordered(renewable_energy): .L	0.885**	0.859*	0.839**	0.818***
	(0.342)	(0.339)	(0.282)	(0.237)
ordered(renewable_energy): .Q	0.565*	0.571*	0.440	0.492*
	(0.260)	(0.257)	(0.225)	(0.213)
ordered(renewable_energy): .C	-0.102	-0.079	-0.048	-0.071
	(0.237)	(0.236)	(0.219)	(0.216)
ordered(renewable_energy): ^4	-0.025	-0.015	0.009	-0.032
	(0.194)	(0.192)	(0.179)	(0.180)
ordered(slow_climate_change): .L	-0.003	-0.003	-0.090	
	(0.288)	(0.282)	(0.249)	
ordered(slow_climate_change): .Q	0.253	0.228	0.498*	
	(0.236)	(0.231)	(0.206)	

ordered(slow_climate_change): .C	-0.131	-0.104	-0.353	
	(0.224)	(0.222)	(0.205)	
ordered(slow_climate_change): ^4	0.010	0.026	0.090	
	(0.178)	(0.176)	(0.163)	
ordered(reduce_footprint): .L	0.064	0.103		
	(0.442)	(0.434)		
ordered(reduce_footprint): .Q	-0.016	0.075		
	(0.316)	(0.311)		
ordered(reduce_footprint): .C	-0.155	-0.240		
	(0.309)	(0.306)		
ordered(reduce_footprint): ^4	0.005	0.069		
	(0.254)	(0.251)		
AIC	1543.701	1547.747	1597.793	1571.204
N	1886	1886	1886	1886

Contingency Tables

Table 8: State

	NJ	NY	AZ	CA
No_Solar	419	508	463	517
Solar	147	186	109	1184

Table 9: Age

	18-44	45-54	55-64	65 or older
No_Solar	579	413	437	466
Solar	263	324	456	465

Table 10: Gender

	Male	Female	No_Answer
No_Solar	844	1049	11
Solar	1023	561	0

Table 11: Income

	less_50	50_74	75_99	100_149	150_more	No_Answer
No_Solar	332	347	366	366	234	251
Solar	135	230	213	362	294	266

Table 12: Using solar would save me money

	StronglyDisagree	Disagree	Nuetral	Agree	StronglyAgree
No_Solar	50	161	533	689	319
Solar	27	22	116	534	893

Table 13: Using solar will protect my family from rising energy prices

	StronglyDisagree	Disagree	Nuetral	Agree	StronglyAgree
No_Solar	38	98	452	851	350
Solar	20	29	98	597	839

Table 14: How much do you spend on your electricity bill during winter months

	less_25	25-49	50-74	75-99	100-149	150-199	200-249	250-299	300-349	350-399	400_more
No_Solar	15	138	264	314	435	234	161	89	58	21	52
Solar	8	31	119	142	319	298	208	126	100	48	94

Table 15: You ask others about their experiences before buying a new product

	NotAtAllLikeMe	NotMuchLikeMe	SomeWhatLikeMe	QuiteAlotLikeMe	JustLikeMe
No_Solar	43	159	602	799	300
Solar	84	212	663	504	131

Table 16: I continually look for new products

	NotAtAllLikeMe	NotMuchLikeMe	SomeWhatLikeMe	QuiteAlotLikeMe	JustLikeMe
No_Solar	123	452	687	448	193
Solar	208	478	572	264	69

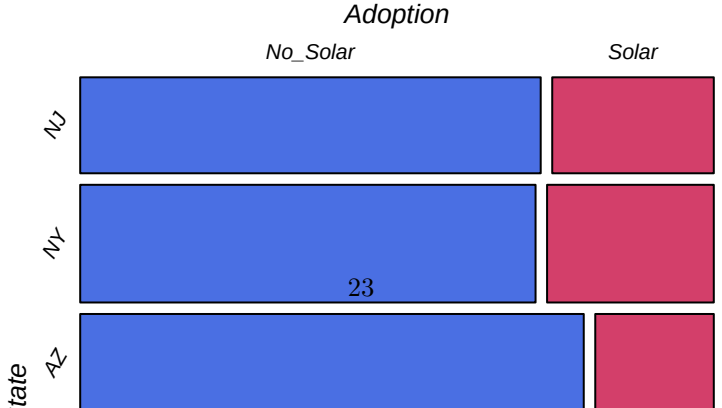
Table 17: I like to visit new places to learn about new products

	NotAtAllLikeMe	NotMuchLikeMe	SomeWhatLikeMe	QuiteAlotLikeMe	JustLikeMe
No_Solar	71	293	708	608	223
Solar	168	374	612	342	98

Table 18: I feel obligated to move the country towards renewable energy

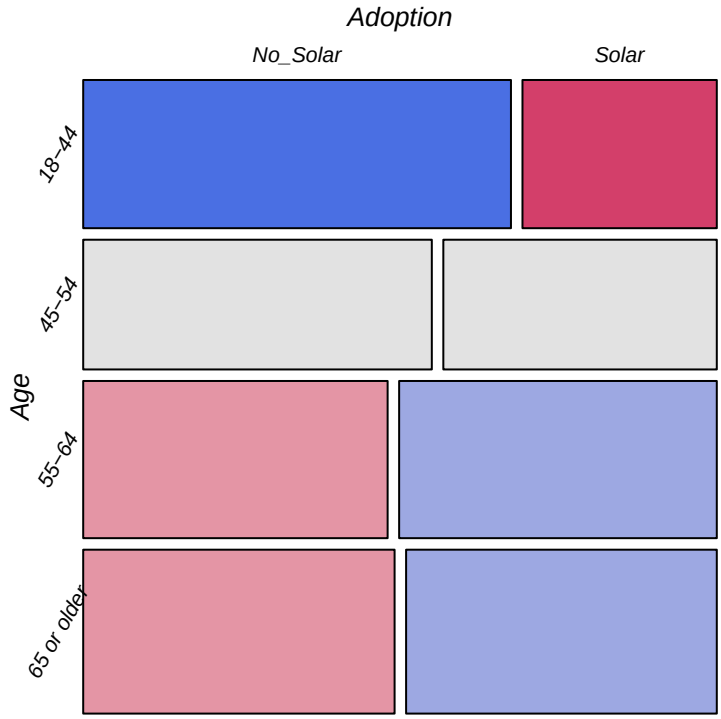
	NotAtAllLikeMe	NotMuchLikeMe	SomeWhatLikeMe	QuiteAlotLikeMe	JustLikeMe
No_Solar	71	293	708	608	223
Solar	168	374	612	342	98

Mosaic Plots and Independence Tests/ Associataion



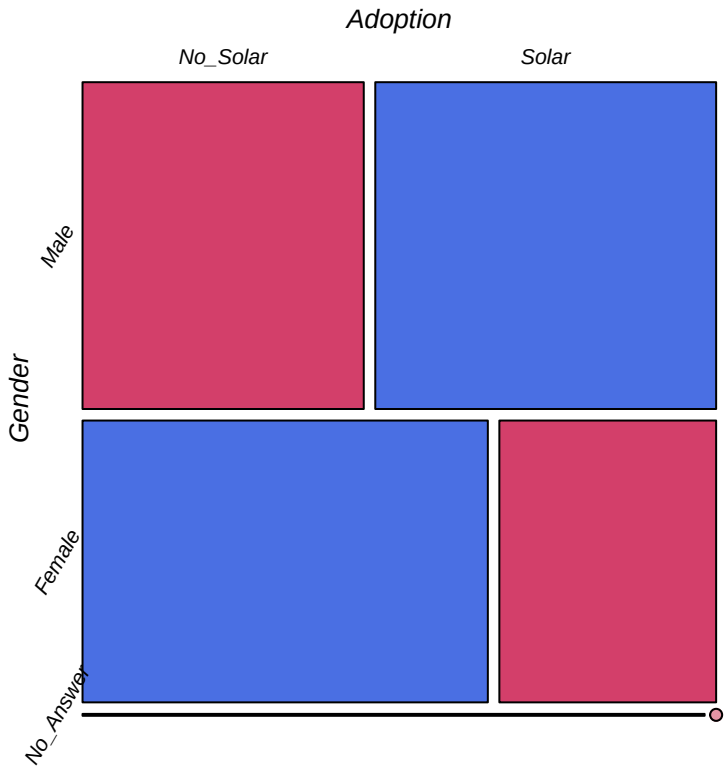
```
##
##  Pearson's Chi-squared test
##
## data:  state_tab
## X-squared = 743.1, df = 3, p-value < 2.2e-16
```

```
##                      X^2 df P(> X^2)
## Likelihood Ratio 773.68  3        0
## Pearson          743.10  3        0
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.417
## Cramer's V        : 0.459
```



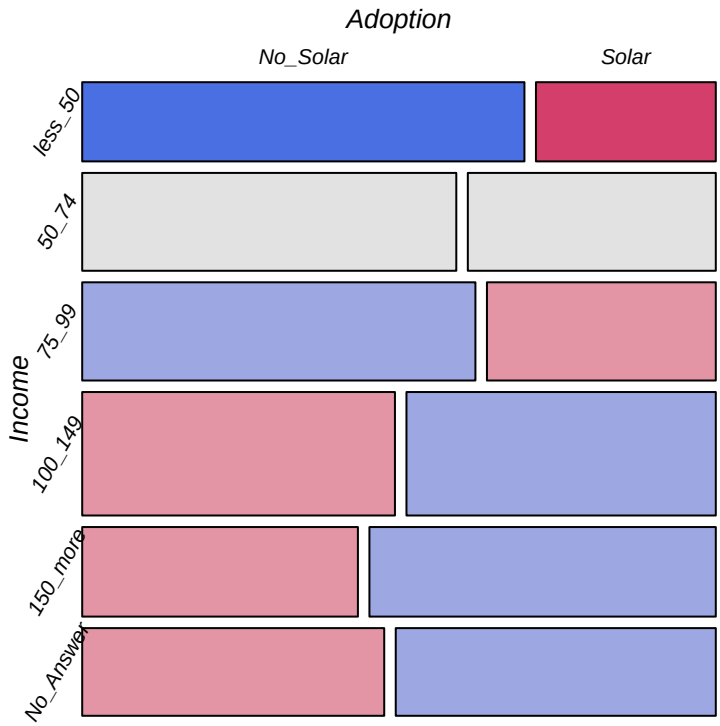

```
## Cochran-Mantel-Haenszel Statistics for Adoption by Age
##
##                  AltHypothesis  Chisq Df       Prob
## cor        Nonzero correlation 70.031  1 5.8373e-17
## rmeans  Row mean scores differ 70.031  1 5.8373e-17
## cmeans  Col mean scores differ 86.834  3 1.0483e-18
## general    General association 86.834  3 1.0483e-18
```

```
## gamma        : 0.21
## std. error   : 0.025
## CI           : 0.162 0.259
```



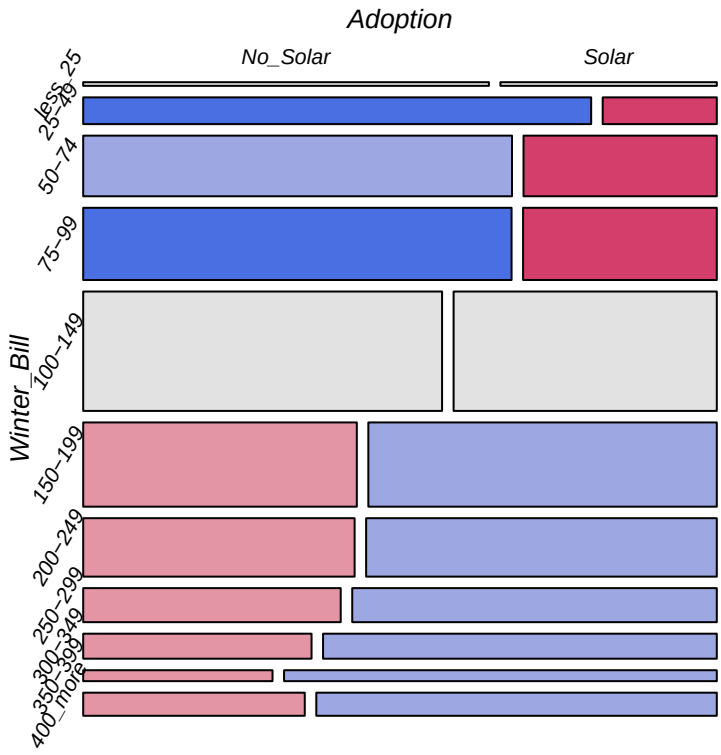

```
##
##  Fisher's Exact Test for Count Data
##
## data:  gender_tab
## p-value < 2.2e-16
## alternative hypothesis: two.sided
```

```
##                      X^2 df P(> X^2)
## Likelihood Ratio 153.31  2        0
## Pearson          147.96  2        0
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.202
## Cramer's V        : 0.206
```



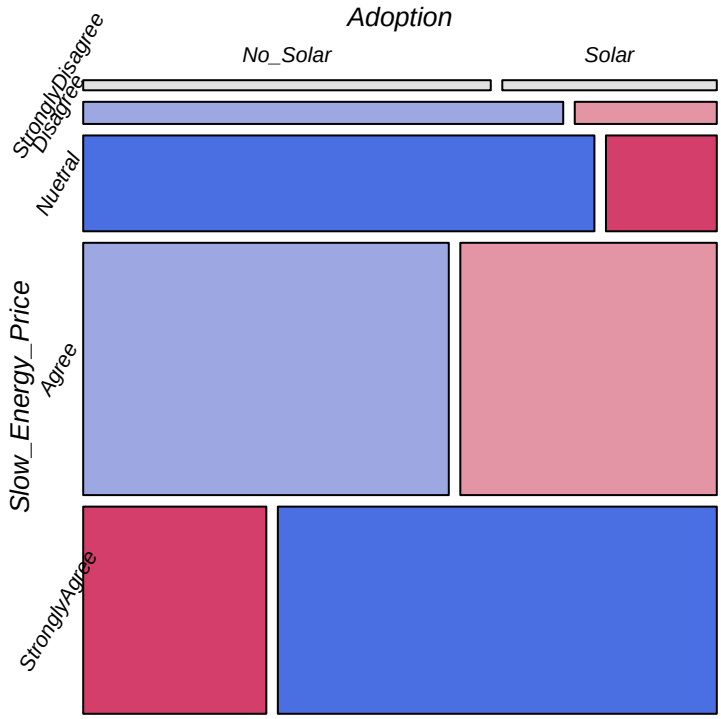

```
## Cochran-Mantel-Haenszel Statistics for Adoption by Income
##
##                   AltHypothesis   Chisq Df       Prob
## cor         Nonzero correlation  88.547  1 4.9631e-21
## rmeans   Row mean scores differ  88.547  1 4.9631e-21
## cmeans   Col mean scores differ 109.817  5 4.4783e-22
## general     General association 109.817  5 4.4783e-22
```

```
## gamma        : 0.223
## std. error   : 0.023
## CI           : 0.179 0.268
```



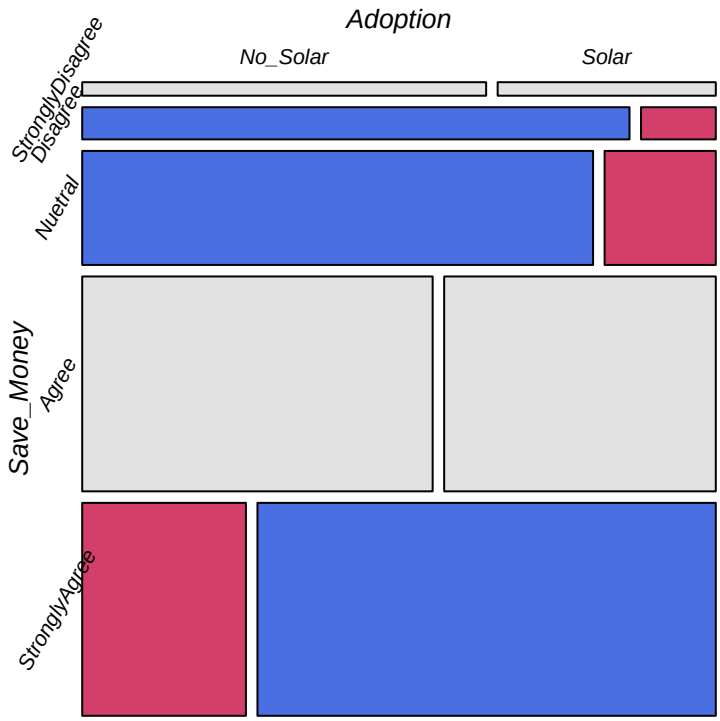

```
## Cochran-Mantel-Haenszel Statistics for Adoption by Winter_Bill
##
##                  AltHypothesis  Chisq Df       Prob
## cor        Nonzero correlation 203.27  1 4.0467e-46
## rmeans  Row mean scores differ 203.27  1 4.0467e-46
## cmeans  Col mean scores differ 237.79 10 1.9904e-45
## general    General association 237.79 10 1.9904e-45

## gamma        : 0.345
## std. error   : 0.021
## CI           : 0.303 0.387
```



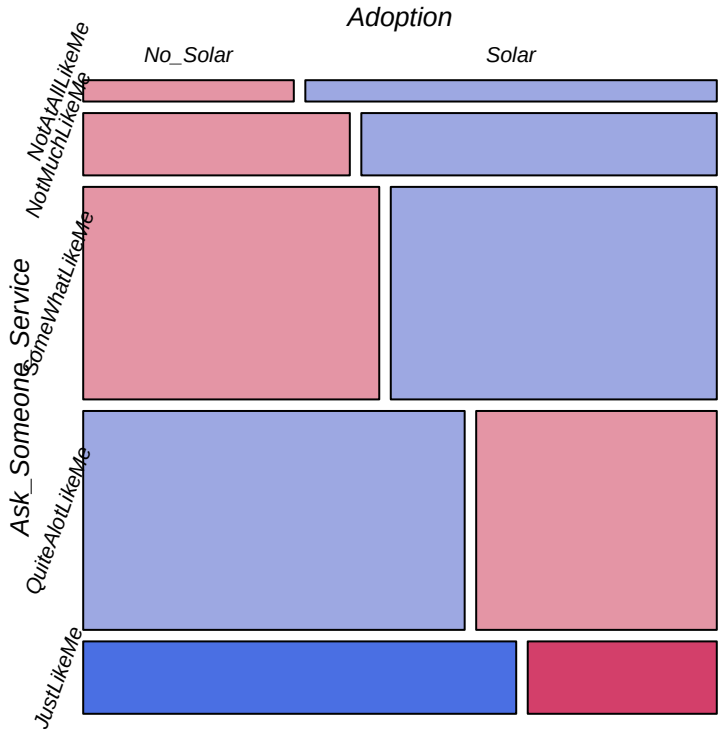

```
## Cochran-Mantel-Haenszel Statistics for Adoption by Slow_Energy_Price
##
##                  AltHypothesis  Chisq Df        Prob
## cor         Nonzero correlation 399.83  1  5.9857e-89
## rmeans   Row mean scores differ 399.83  1  5.9857e-89
## cmeans   Col mean scores differ 505.74  4 3.8399e-108
## general     General association 505.74  4 3.8399e-108
```

```
## gamma        : 0.587
## std. error   : 0.021
## CI           : 0.546 0.628
```



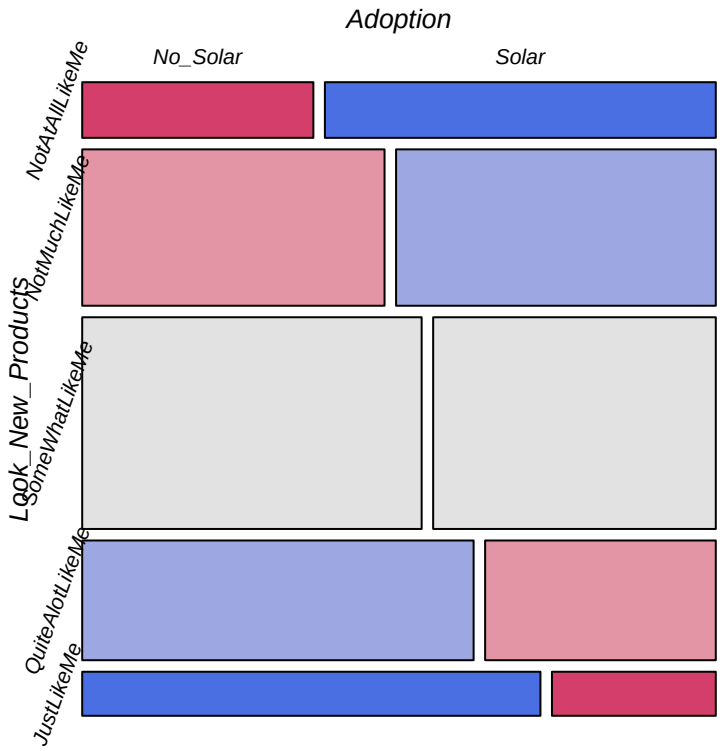

```
## Cochran-Mantel-Haenszel Statistics for Adoption by Save_Money
##
##                  AltHypothesis  Chisq Df        Prob
## cor         Nonzero correlation 544.63  1 1.8572e-120
## rmeans   Row mean scores differ 544.63  1 1.8572e-120
## cmeans   Col mean scores differ 665.54  4 1.0066e-142
## general     General association 665.54  4 1.0066e-142
```

```
## gamma        : 0.645
## std. error   : 0.019
## CI           : 0.609 0.682
```



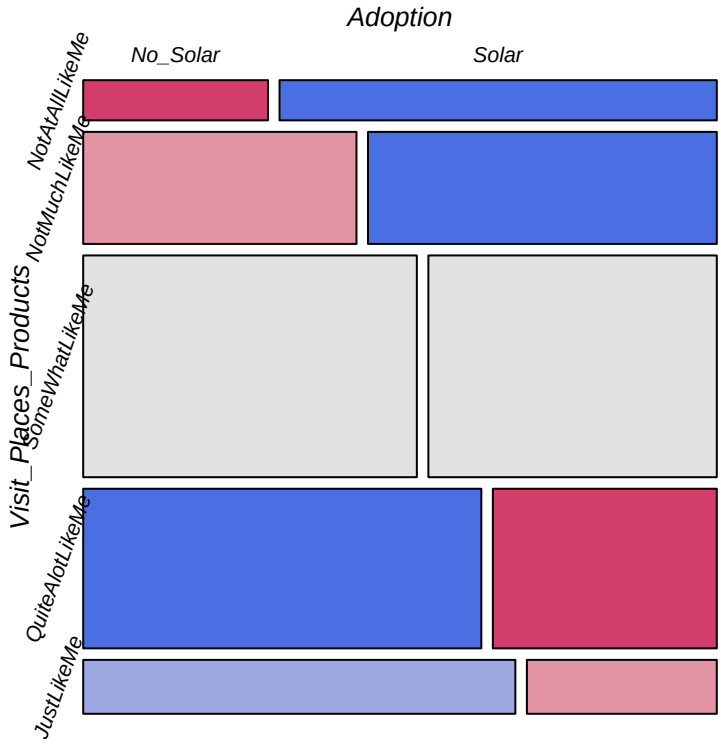

```
## Cochran-Mantel-Haenszel Statistics for Adoption by Ask_Someone_Service
##
##                  AltHypothesis  Chisq Df       Prob
## cor         Nonzero correlation 124.65  1 6.0647e-29
## rmeans   Row mean scores differ 124.65  1 6.0647e-29
## cmeans   Col mean scores differ 130.48  4 3.0710e-27
## general     General association 130.48  4 3.0710e-27
```

```
## gamma        : -0.296
## std. error   : 0.025
## CI           : -0.345 -0.248
```



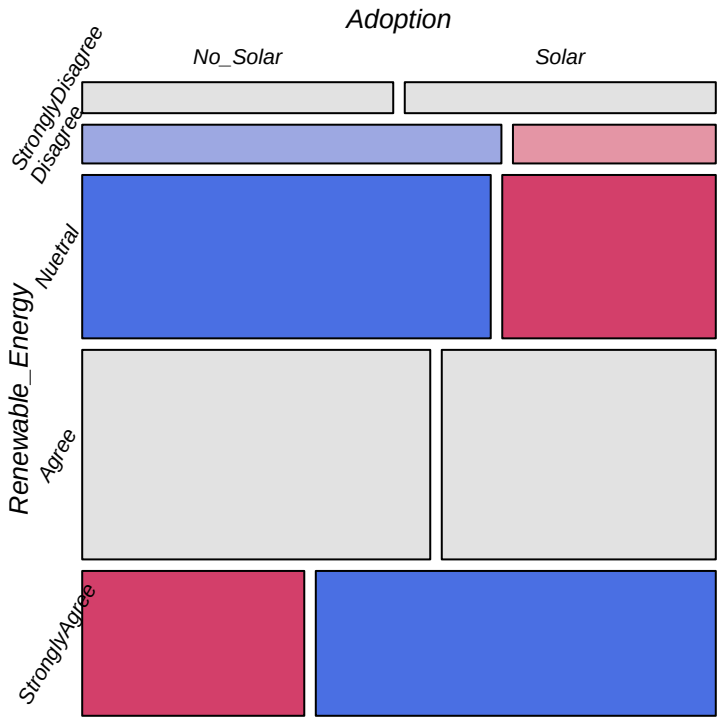

```
## Cochran-Mantel-Haenszel Statistics for Adoption by Look_New_Products
##
##                  AltHypothesis  Chisq Df       Prob
## cor         Nonzero correlation 110.29  1 8.4742e-26
## rmeans   Row mean scores differ 110.29  1 8.4742e-26
## cmeans   Col mean scores differ 112.30  4 2.3519e-23
## general     General association 112.30  4 2.3519e-23
```

```
## gamma        : -0.258
## std. error   : 0.024
## CI           : -0.305 -0.211
```




```
## Cochran-Mantel-Haenszel Statistics for Adoption by Visit_Places_Products
##
##                  AltHypothesis  Chisq Df       Prob
## cor        Nonzero correlation 149.75  1 1.9640e-34
## rmeans  Row mean scores differ 149.75  1 1.9640e-34
## cmeans  Col mean scores differ 153.19  4 4.2143e-32
## general    General association 153.19  4 4.2143e-32
```

```
## gamma        : -0.306
## std. error   : 0.024
## CI           : -0.353 -0.259
```




```
## Cochran-Mantel-Haenszel Statistics for Adoption by Renewable_Energy
##
##                  AltHypothesis   Chisq Df       Prob
## cor         Nonzero correlation  95.029  1 1.8761e-22
## rmeans   Row mean scores differ  95.029  1 1.8761e-22
## cmeans   Col mean scores differ 188.841  4 9.4064e-40
## general     General association 188.841  4 9.4064e-40
```

```
## gamma        : 0.294
## std. error   : 0.024
## CI           : 0.246 0.341
```