Research Proposal

As the cost of producing and installing solar panels continues to decrease, more homeowners have turned to solar as a means to effectively lower their total energy bill. This prompted Google to use their extensive resources in data and mapping to come up with a way for homeowners to easily assess the impact of installing solar on their own roofs. The endeavor turned into Project Sunroof, which effectively calculates how much sunlight hits your roof in a year. This takes into account such factors as shadows cast by nearby structures, all possible sun positions over the course of a year and historical data on cloud and temperature patterns in your area. The dataset below incorporates all of this information for every house at a city level, and also shows the potential for solar growth based off available roof space. This allows us to visualize the current state of solar generation in the United States, as well as the potential impact solar could have in the future. As we exhaust finite energy resources, renewable energy may become a neccessity of everyday life in the future.

In [1]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns

%matplotlib inline
```

In [2]:
```
solar = pd.read_csv(r'C:\Users\mike1\Documents\Thinkful\project-sunroof-city-09082017.csv', low_memory = False)

solar.dropna(inplace = True)
solar.head()
```

Out[2]:

	region_name	state_name	lat_max	lat_min	lng_max	lng_min	lat_avg	lng_avg	yearly_sunlight_kwh_kw_threshold_avg	count_qualified
2	Aberdeen	North Carolina	35.183959	35.053605	-79.388482	-79.538347	35.143597	-79.424747	1083.750000	1078
3	Abilene	Texas	32.614325	32.236656	-99.589583	-100.086436	32.435016	-99.750683	1252.411447	42802
4	Abington	Massachusetts	42.147124	42.088107	-70.918676	-71.001744	42.116700	-70.954330	974.091632	3285
5	Abington	Pennsylvania	40.141026	40.105290	-75.098139	-75.154789	40.127027	-75.129703	995.225877	4290
6	Absecon	New Jersey	39.449684	39.392896	-74.477543	-74.528522	39.430349	-74.501735	1045.500000	500

5 rows × 31 columns

In [3]:
```
solar_states = solar.groupby('state_name', as_index = False).sum()

#Number of predicted buildings with solar/number of buildings suitable to solar
solar_states['utilization%'] = (solar_states['existing_installs_count']/solar_states['count_qualified']) * 100
solar_states.head()
```

Out[3]:

	state_name	lat_max	lat_min	lng_max	lng_min	lat_avg	lng_avg	yearly_sunlight_kwh_kw_threshold_avg	count_qualif
0	Alabama	5954.123641	5935.675211	-15597.454886	-15619.569833	5944.770799	-15608.162936	195109.417999	6895
1	Alaska	256.345448	255.207702	-590.503201	-594.097973	255.736510	-593.297461	3002.200000	274
2	Arizona	2345.363253	2334.275026	-7822.175216	-7834.052368	2338.723960	-7827.499890	96969.740304	16236
3	Arkansas	2639.069130	2631.506433	-6956.430151	-6965.785992	2635.212587	-6961.189488	80590.663207	3182
4	California	25302.942246	25241.729426	-84014.545647	-84094.429872	25272.475595	-84055.798886	888458.938328	78106

5 rows × 30 columns

Potential Impact

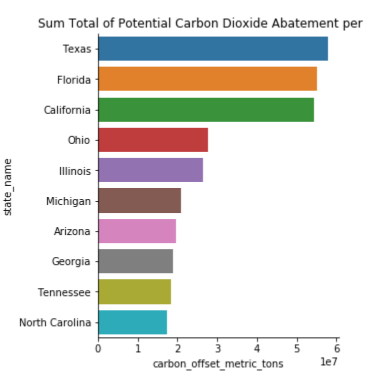
Its no secret that climate change is one of the most debated and potentially most challenging problems facing the world to date. Decreasing our reliance on fossil fuels and becoming more dependent on renewable energy has the potential to greatly impact climate in a positive way. This dataset from Project Sunroof allows us to see just how big of an impact there could be in the U.S. if all roof space was utilized for solar energy generation. The U.S. alone has the potential to eliminate 561.5 billion metric tons of carbon dioxide while generating 1.07 trillion kwh of energy per year. At a state level, Texas, California, and Florida have the highest potential impact in both these measureables.

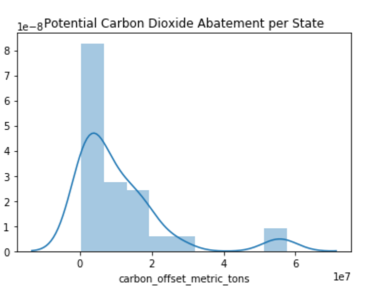
In [4]:
```
print("Sum Total of U.S. Carbon Dioxide Abatement \n", solar['carbon_offset_metric_tons'].sum())
print("Sum Total of the Yearly Solar Energy Generation Potential in the U.S. \n", solar['yearly_sunlight_kwh_total'].sum())
```

```
Sum Total of U.S. Carbon Dioxide Abatement
 561509730.3703218
Sum Total of the Yearly Solar Energy Generation Potential in the U.S.
 1069411604583.3407
```

In [5]:
```
# Sum of the potential carbon dioxide abatement per state of the solar capacity that meets the technical potential criteria.
sum_carbon_offset = solar['carbon_offset_metric_tons'].groupby(solar['state_name']).sum()
print(sum_carbon_offset.nlargest(10))

sns.catplot(x = "carbon_offset_metric_tons", y = "state_name", kind = 'bar', data = solar_states.sort_values(by = ['carbon_of
fset_metric_tons'], ascending = False, axis = 0).head(10))
plt.title('Sum Total of Potential Carbon Dioxide Abatement per State')
plt.show()

sns.distplot(solar_states['carbon_offset_metric_tons'])
plt.title("Potential Carbon Dioxide Abatement per State")
plt.show()
```

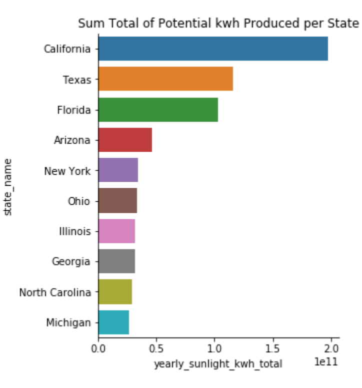
```
state_name
Texas             5.779106e+07
Florida           5.509479e+07
California        5.425624e+07
Ohio              2.775254e+07
Illinois          2.648314e+07
Michigan          2.085491e+07
Arizona           1.973379e+07
Georgia           1.880518e+07
Tennessee         1.835439e+07
North Carolina    1.732582e+07
Name: carbon_offset_metric_tons, dtype: float64
```

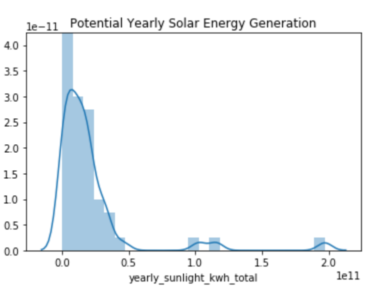




In [6]:
```
# Sum of the yearly total solar energy generation potential for all roof space in per state
sum_kwh = solar['yearly_sunlight_kwh_total'].groupby(solar['state_name']).sum()
print(sum_kwh.nlargest(10))

sns.catplot(x = "yearly_sunlight_kwh_total", y = "state_name", kind = 'bar', data = solar_states.sort_values(by = ['yearly_su
nlight_kwh_total'], ascending = False, axis = 0).head(10))
plt.title('Sum Total of Potential kwh Produced per State')
plt.show()

sns.distplot(solar_states['yearly_sunlight_kwh_total'])
plt.title("Potential Yearly Solar Energy Generation")
plt.show()
```

```
state_name
California        1.971434e+11
Texas             1.156487e+11
Florida           1.026393e+11
Arizona           4.597492e+10
New York          3.431742e+10
Ohio              3.352405e+10
Illinois          3.215390e+10
Georgia           3.147178e+10
North Carolina    2.939512e+10
Michigan          2.658946e+10
Name: yearly_sunlight_kwh_total, dtype: float64
```



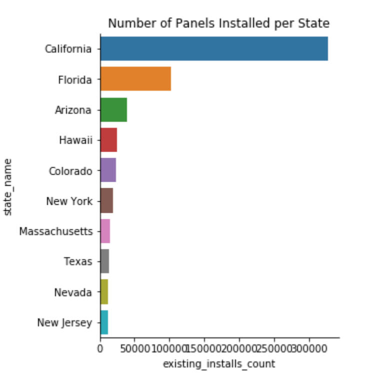


Current Utilization

Despite the potential impact solar panels could have on carbon dioxide abatement and energy generation, there is still surprisingly low utilization across the U.S., both in terms of the actual number of panels installed and the percentages of utilization. California leads the way in solar panels by a large margin with a total of 327,142 panels. The next closest being Florida at 103,019 panels, still 3x lower than California. However, what is even more glaring is when we examine the percentages of roofs utilizing solar. Hawaii is far ahead of all states at 13.15%, with California coming in second at 4.19%. Only six states have a utilization percentage rate above 2%, and the bottom 10 states are all significantly below half of a percent.

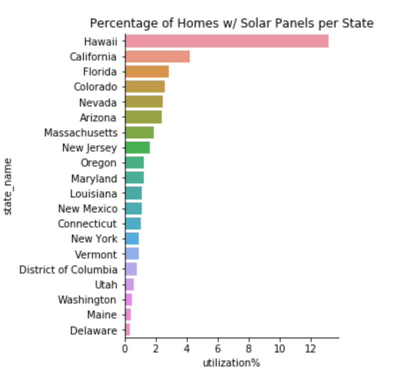
In [7]:
```
# Total number of buildings per state estimated to have a solar installation, at time of data collection
sum_installs = solar['existing_installs_count'].groupby(solar['state_name']).sum()
print("Top ten states with solar panels installed \n", sum_installs.nlargest(10))

sns.catplot(x = "existing_installs_count", y = "state_name", kind = 'bar', data = solar_states.sort_values(by = ['existing_in
stalls_count'], ascending = False, axis = 0).head(10))
plt.title('Number of Panels Installed per State')

plt.show()
```

```
Top ten states with solar panels installed
 state_name
California       327142
Florida          103019
Arizona           39367
Hawaii            25234
Colorado          24182
New York          19675
Massachusetts     15325
Texas             13799
Nevada            12755
New Jersey        12227
Name: existing_installs_count, dtype: int64
```



In [8]:
```
utilization_percentage = (sum_installs/(solar['count_qualified'].groupby(solar['state_name']).sum())) * 100
print("Percentage of buildings with solar installed per state: Top 10 \n", utilization_percentage.nlargest(20))
print("Percentage of buildings with solar installed per state: Bottom 10 \n", utilization_percentage.nsmallest(10))

sns.catplot(x = "utilization%", y = "state_name", kind = 'bar', data = solar_states.sort_values(by = ['utilization%'], ascend
ing = False, axis = 0).head(20))
plt.title('Percentage of Homes w/ Solar Panels per State')

plt.show()
```

```
Percentage of buildings with solar installed per state: Top 10
 state_name
Hawaii                  13.151613
California               4.188414
Florida                  2.872319
Colorado                 2.606726
Nevada                   2.447495
Arizona                  2.424532
Massachusetts            1.867600
New Jersey               1.664767
Oregon                   1.270981
Maryland                 1.217057
Louisiana                1.127573
New Mexico               1.094716
Connecticut              1.061184
New York                 0.953558
Vermont                  0.917495
District of Columbia     0.828921
Utah                     0.631105
Washington               0.447881
Maine                    0.423819
Delaware                 0.338914
dtype: float64
Percentage of buildings with solar installed per state: Bottom 10
 state_name
South Dakota     0.032506
North Dakota     0.033732
Wyoming          0.059740
Arkansas         0.074477
Nebraska         0.077125
Alabama          0.078462
Iowa             0.084875
West Virginia    0.094627
Indiana          0.114210
Oklahoma         0.121720
dtype: float64
```



All of this begs the question: "How do we increase solar installation rate?" Is it simply a matter of decreasing the initial costs? Or is it essential that we develop solar panels that are even more efficient so the upfront costs are negated more quickly? While the former may be solvable through government incentives and tax write offs, the long term efficacy of solar panels as a reliable energy source will depend on developing new and more efficient panels. Current solar panels are estimated to be about 15% efficient in terms of energy generation, and this is the same calculation used by Project Sunroof. However, new technology such as SunPower's Residential X-Series: X22-370 panel has an estimated efficiency greater than 22%. This experiment will not only test new solar technology against existing technology to determine if there is a statistically significant difference between the two, but also look at the efficiency of solar panels across various regions using Project Sunroof as a guide for where to test.

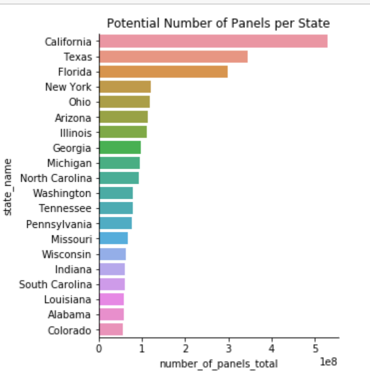
Hypothesis

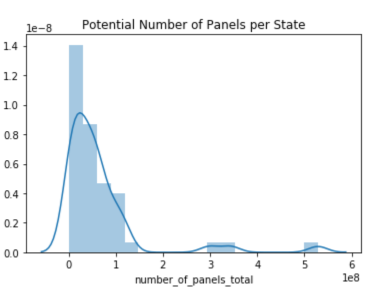
The null hypothesis for this experiment is that energy generation will be the same between solar panels.

Identifying States & Cities for Experiment

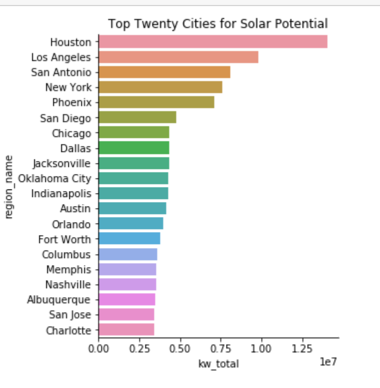
The first essential step for this experiment will be to identify where to conduct it. We have shown previously which states have a high potential impact of using solar, as well as the current utilization percentages amongst states. I believe the best scenario for this experiment is to eventually run it across several regions/cities where the potential for solar energy is high, but there is currently a low percentage of actual solar panel use with the hope that this will encourage these areas begin investment in solar energy. Luckily, these factors can be easily displayed from the data. We have already shown that Texas, California, and Florida have the greatest potential for solar energy generation, however Texas has a percentage of utilization outside of the top 20 states. This makes Texas a prime spot for the initial rollout of the experiment and in fact, if we dive deeper into the data, Houston and San Antonio are number 1 and 3 respectively in terms of total kw potential. Texas is also number two in terms of the available roof space for solar panels. Not far behind it in this category are Ohio and Illinois at 5 and 7. Going back to our previous displays, these states are both top 5 in terms of the potential for carbon dioxide abatement, top 7 in terms of yearly solar potential and surprisingly, like Texas, neither state has a utilization percentage in the top 20. Because of this, Chicago and Columbus, both cities in the top 15 in regards to kw potential, seem to be exciting prospects for the experiment as well.

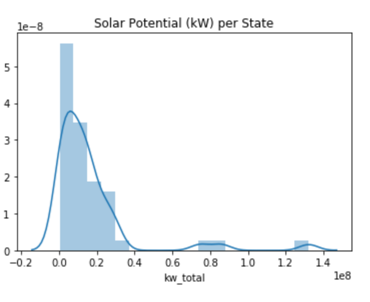
In [9]:
```
sns.catplot(x = "number_of_panels_total", y = "state_name", kind = 'bar', data = solar_states.sort_values(by = ['number_of_pa
nels_total'], ascending = False, axis = 0).head(20))
plt.title('Potential Number of Panels per State')
plt.show()

sns.distplot(solar_states['number_of_panels_total'])
plt.title('Potential Number of Panels per State')
plt.show()
```





In [10]:
```
sns.catplot(x = "kw_total", y = "region_name", kind = 'bar', data = solar.sort_values(by = ['kw_total'], ascending = False, a
xis = 0).head(20))
plt.title('Top Twenty Cities for Solar Potential')
plt.show()


sns.distplot(solar_states['kw_total'])
plt.title('Solar Potential (kW) per State')
plt.show()
```





Method of Testing

To test the hypothesis, existing solar panels will be placed on a roof next to SunPower's new XSeries panels. Data on individual energy generation from both panels will be collected everyday. Because the panels will be in close proximity, conditions should be identical for both panels, however upon installation this assumption should be verified by looking at any potential issues such as areas that may experience more shade, etc. The experiment will begin in Houston with a sample of 500 rooftops and data will be collected and analyzed after one month. If the experiment is progressing well and there are no problems with the data or its collection, the experiment will then be extended to San Antonio and another sample of 500 rooftops. After another month if the experiment has not experienced any issues, the final phase of the rollout will commence in Chicago and Columbus. This will give us a sample size of 2,000 rooftops across four regions with high potential in terms of solar generation. To test the significance between energy generation of the panels, the average kw produced per panel will be found at varying time periods (3, 6 and 12 months). The t-values will be calculated from these averages and if they lead to a p-value < 0.05, it will be concluded that there is a statistical difference between new and existing panels and the null hypothesis will be rejected.

Additional Thoughts

Besides examining how new and existing solar panels compare to one another, this experiment will also show how solar energy generation differs between states and regions. The data has shown us cities where there is a high potential in terms of energy production, but it is not exactly clear if this is more a product of ideal conditions for solar energy production or due to the sheer volume of available roof space. Weather patterns and conditions will always play a major role in any type of renewable energy production which means some areas will naturally be more efficient. Another important thing to consider is the energy requirements of a city. More densely populated regions will also require greater amounts of energy, so while potential may be high in some of these cities, it may not exceed current or future requirements. This makes the efficiency by which we gather and store solar energy a constant area where we should strive for improvement.

In [ ]: