Sample_Project_2

March 20, 2023

1 Sample Project 2

[1]:
```
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Read in data from csv's along with a few adjustments.

[2]:
```
df_gross = pd.read_csv('UNdata_TotalElectricity.csv')
df_solar = pd.read_csv('UNdata_Solar.csv')
df_wind = pd.read_csv('UNdata_Wind.csv')
df_geothermal = pd.read_csv('UNdata_Geothermal.csv')
df_wave = pd.read_csv('UNdata_Wave.csv')
df_nuclear = pd.read_csv('UNdata_Nuclear.csv')

df_gross = df_gross.filter(['Country or Area', 'Year', 'Quantity'])
df_solar = df_solar.filter(['Country or Area', 'Year', 'Quantity'])
df_wind = df_wind.filter(['Country or Area', 'Year', 'Quantity'])
df_geothermal = df_geothermal.filter(['Country or Area', 'Year', 'Quantity'])
df_wave = df_wave.filter(['Country or Area', 'Year', 'Quantity'])
df_nuclear = df_nuclear.filter(['Country or Area', 'Year', 'Quantity'])

df_gross = df_gross.rename(columns={'Quantity':'Gross Electricity'})
df_solar = df_solar.rename(columns={'Quantity':'Solar'})
df_wind = df_wind.rename(columns={'Quantity':'Wind'})
df_geothermal = df_geothermal.rename(columns={'Quantity':'Geothermal'})
df_wave = df_wave.rename(columns={'Quantity':'Wave/Tidal'})
df_nuclear = df_nuclear.rename(columns={'Quantity':'Nuclear'})
```

[3]:
```
#many nonrenewable stats are actually in the UNdata_Thermal.csv
df_thermal = pd.read_csv('UNdata_Thermal.csv')
print(df_thermal['Commodity - Transaction'].unique())
```

```
['Electricity - total thermal production'
 'Electricity production from hard coal'
 'Electricity production from gas oil/ diesel oil'
 'Electricity production from fuel oil'
```

```
'Electricity production from oil shale'
'Electricity production from other oil products'
'Electricity production from natural gas'
'Electricity production from non-renewable waste'
'Electricity production from renewable municipal waste'
'Electricity production from bagasse'
'Electricity production from solid biofuels'
'Electricity production from liquid biofuels'
'Electricity production from manufactured gases'
'Electricity production from biogases'
'Electricity production from solid coal products'
'Electricity production from brown coal'
'Electricity production from peat'
'Electricity production from crude oil, NGLs, other hydrocarbons'
'Footnote' 'Estimate']
```

[4]:
```
df_totalfossil = df_thermal.loc[df_thermal['Commodity - Transaction'] ==␣
 ↪'Electricity - total thermal production']
df_naturalgas = df_thermal.loc[df_thermal['Commodity - Transaction'] ==␣
 ↪'Electricity production from natural gas']

df_totalfossil = df_totalfossil.filter(['Country or Area', 'Year', 'Quantity'])
df_naturalgas = df_naturalgas.filter(['Country or Area', 'Year', 'Quantity'])

df_totalfossil = df_totalfossil.rename(columns={'Quantity':'Total Fossil␣
 ↪Fuels'})
df_naturalgas = df_naturalgas.rename(columns={'Quantity':'Natural Gas'})
```

[5]:
```
df_coal = df_thermal.loc[(df_thermal['Commodity - Transaction'] == 'Electricity␣
 ↪production from hard coal')]
df_coal2 = df_thermal.loc[(df_thermal['Commodity - Transaction'] ==␣
 ↪'Electricity production from brown coal')]

df_coal = df_coal.filter(['Country or Area', 'Year', 'Quantity'])
df_coal2 = df_coal2.filter(['Country or Area', 'Year', 'Quantity'])
df_coal = df_coal.rename(columns={'Quantity':'Hard Coal'})
df_coal2 = df_coal2.rename(columns={'Quantity':'Brown Coal'})

df_coal = pd.merge(df_coal, df_coal2, on=['Country or Area', 'Year'])
df_coal.insert(2, 'Total Coal', df_coal['Hard Coal'] + df_coal['Brown Coal'])
```

[6]:
```
df = pd.merge(df_gross, df_solar, on=['Country or Area', 'Year'], how='outer')

#need a left merge for subsequent merges, otherwise data doesn't line up well
df = df.merge(df_wind, on=['Country or Area', 'Year'], how='left')
df = df.merge(df_geothermal, on=['Country or Area', 'Year'], how='left')
df = df.merge(df_wave, on=['Country or Area', 'Year'], how='left')
```

```
df = df.merge(df_nuclear, on=['Country or Area', 'Year'], how='left')
df = df.merge(df_totalfossil, on=['Country or Area', 'Year'], how='left')
df = df.merge(df_naturalgas, on=['Country or Area', 'Year'], how='left')
df = df.merge(df_coal, on=['Country or Area', 'Year'], how='left')

#majority of countries have this recorded span of years
df = df.loc[(df['Year'] >= 1990) & (df['Year'] <= 2018)]

display(df)
```

```
     Country or Area    Year  Gross Electricity  Solar  Wind  Geothermal  \
1        Afghanistan  2018.0            1265.72    NaN   NaN         NaN
2        Afghanistan  2017.0            1420.04    NaN   NaN         NaN
3        Afghanistan  2016.0            1393.90    NaN   NaN         NaN
4        Afghanistan  2015.0            1338.70    NaN   NaN         NaN
5        Afghanistan  2014.0            1356.00    NaN   NaN         NaN
…                …     …                  …     …   …          …
7232        Zimbabwe  1994.0            7815.00    NaN   NaN         NaN
7233        Zimbabwe  1993.0            7468.00    NaN   NaN         NaN
7234        Zimbabwe  1992.0            8237.00    NaN   NaN         NaN
7235        Zimbabwe  1991.0            8886.00    NaN   NaN         NaN
7236        Zimbabwe  1990.0            9362.00    NaN   NaN         NaN

      Wave/Tidal  Nuclear  Total Fossil Fuels  Natural Gas  Total Coal  \
1            NaN      NaN              199.93          NaN         NaN
2            NaN      NaN              176.26          NaN         NaN
3            NaN      NaN              154.70          NaN         NaN
4            NaN      NaN              149.20          NaN         NaN
5            NaN      NaN              159.00          NaN         NaN
…          …      …                  …          …          …
7232         NaN      NaN             5440.00          NaN         NaN
7233         NaN      NaN             5406.00          NaN         NaN
7234         NaN      NaN             5076.00          NaN         NaN
7235         NaN      NaN             5772.00          NaN         NaN
7236         NaN      NaN             4993.00          NaN         NaN

      Hard Coal  Brown Coal
1           NaN         NaN
2           NaN         NaN
3           NaN         NaN
4           NaN         NaN
5           NaN         NaN
…         …          …
7232        NaN         NaN
7233        NaN         NaN
7234        NaN         NaN
7235        NaN         NaN
```

```
7236         NaN          NaN

[7158 rows x 13 columns]
```

The above dataframe will remain the same and will be used to create new dataframes from this point on.

I want to first look at data solely in the United States.

[7]:
```
df_US = df.loc[df['Country or Area'] == 'United States']
#there is no data on Wave and Tidal energy in the US
df_US = df_US.drop('Wave/Tidal', 1)

df_US = df_US.astype({"Year":'int', "Gross Electricity":'int', "Solar":'int',␣
 ↪"Wind":'int', "Geothermal":'int',
                "Nuclear":'int', "Total Fossil Fuels":'int', "Natural Gas":
 ↪'int',
                "Total Coal":'int', "Hard Coal":'int', "Brown Coal":'int'})

display(df_US)
```

	Country or Area	Year	Gross Electricity	Solar	Wind	Geothermal	\
6921	United States	2018	4455439	85184	275834	18773	
6922	United States	2017	4286423	70980	257249	18726	
6923	United States	2016	4322038	50334	229471	18584	
6924	United States	2015	4317159	35635	192992	18727	
6925	United States	2014	4340371	25764	183892	18710	
6926	United States	2013	4306371	15872	169713	18422	
6927	United States	2012	4290660	10145	141922	18135	
6928	United States	2011	4349463	6215	120854	17840	
6929	United States	2010	4378430	3942	95148	17577	
6930	United States	2009	4188215	2515	74226	17046	
6931	United States	2008	4368261	2092	55696	16873	
6932	United States	2007	4349841	1674	34603	16798	
6933	United States	2006	4300831	1287	26676	16581	
6934	United States	2005	4294368	1120	17881	16778	
6935	United States	2004	4174856	965	14291	15487	
6936	United States	2003	4081764	848	11300	14870	
6937	United States	2002	4051120	830	10459	14939	
6938	United States	2001	3865307	785	6806	14246	
6939	United States	2000	4052667	709	5650	14621	
6940	United States	1999	3897518	689	4802	15717	
6941	United States	1998	3830489	890	3018	15369	
6942	United States	1997	3697728	897	3254	14907	
6943	United States	1996	3677022	906	3410	15746	
6944	United States	1995	3582114	828	3196	14941	
6945	United States	1994	3473435	828	3483	17479	
6946	United States	1993	3411280	901	3053	17774	
6947	United States	1992	3291109	749	2917	17168	

6948	United States	1991	3275840	782	3051	16267
6949	United States	1990	3218621	666	3066	16012

	Nuclear	Total Fossil Fuels	Natural Gas	Total Coal	Hard Coal	\
6921	841329	2912042	1519218	1267748	526593	
6922	838861	2770050	1337703	1317448	573491	
6923	839918	2886040	1418100	1349707	625715	
6924	830288	2962870	1372570	1466931	689901	
6925	830584	2995556	1161333	1708397	845169	
6926	822004	2985891	1158454	1708088	806840	
6927	801129	3016908	1264552	1639371	780993	
6928	821405	3034500	1045254	1871581	920700	
6929	838931	3132755	1017869	1990647	1019425	
6930	830210	2965173	949776	1890058	939466	
6931	837804	3173013	910176	2128335	1137972	
6932	836634	3183440	915196	1743972	802113	
6933	816195	3121205	842774	1851336	910229	
6934	810726	3149290	782829	1808701	860163	
6935	813339	3032176	731552	1789729	867778	
6936	787818	2960065	670192	1798286	906615	
6937	804519	2928244	712432	1825779	957829	
6938	792604	2836138	659914	1972709	1160834	
6939	797718	2953983	634290	2122298	1252950	
6940	771811	2802706	581933	2011336	1227073	
6941	714124	2775008	558449	2004148	1285200	
6942	666363	2653622	505647	1973060	1281002	
6943	715212	2565046	478765	1920678	1252317	
6944	713806	2511487	528844	1829080	1198890	
6945	678920	2488352	492096	1809179	1198104	
6946	646987	2439505	441000	1807356	1218285	
6947	655970	2339422	426908	1728055	1191742	
6948	649399	2297186	402230	1700335	1181603	
6949	611589	2298328	381669	1699648	1182944	

	Brown Coal
6921	741155
6922	743957
6923	723992
6924	777030
6925	863228
6926	901248
6927	858378
6928	950881
6929	971222
6930	950592
6931	990363
6932	941859
6933	941107

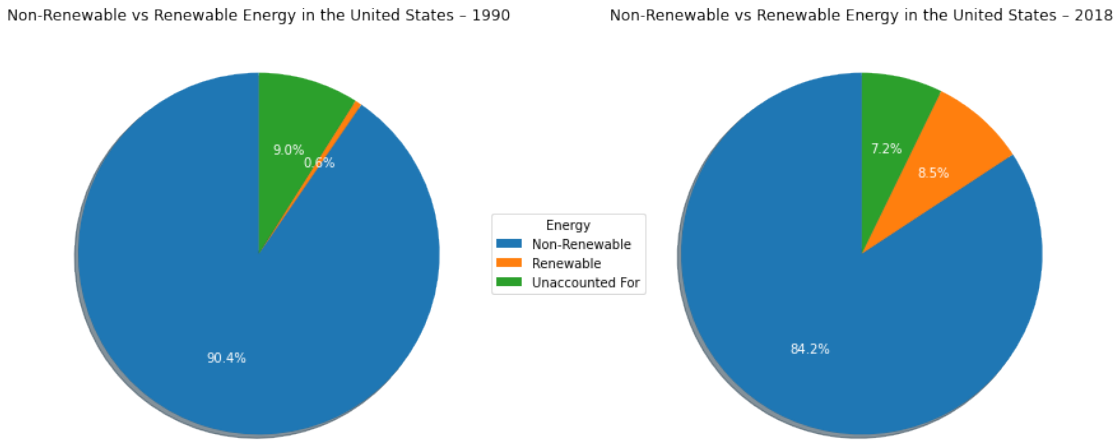
```
6934      948538
6935      921951
6936      891671
6937      867950
6938      811875
6939      869348
6940      784263
6941      718948
6942      692058
6943      668361
6944      630190
6945      611075
6946      589071
6947      536313
6948      518732
6949      516704
```

This dataframe will be used for a few experiments.

Looking at non-renewable energy vs renewable energy in the US

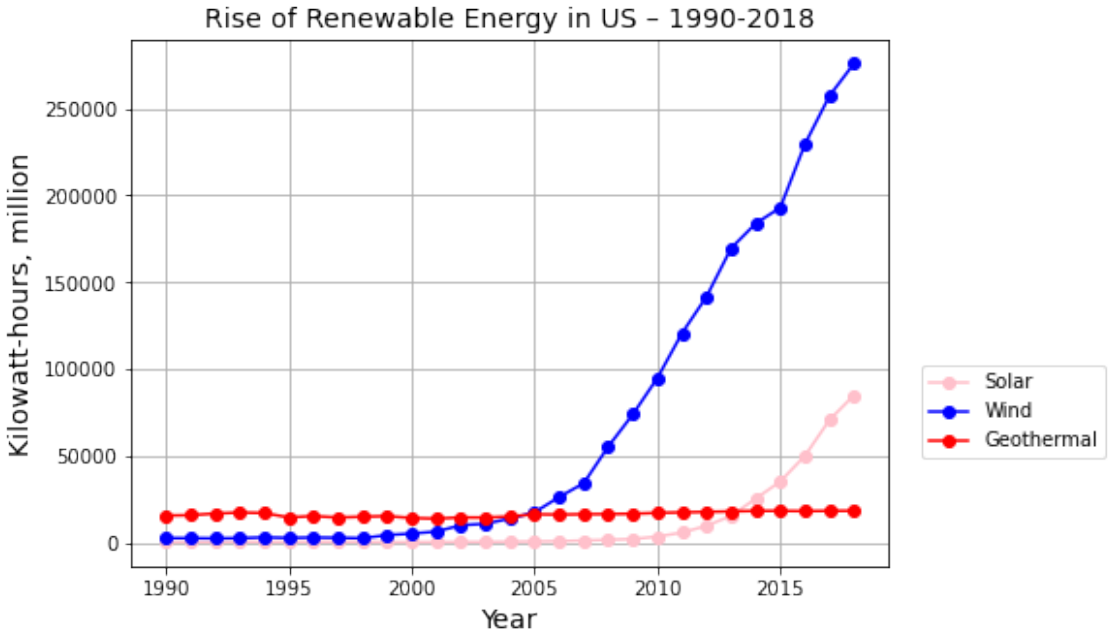
[8]:
```
fig = plt.figure(figsize=(15, 15))
grid = plt.GridSpec(2, 2, wspace=0.3, hspace=0.3)

df_US_1990 = df_US.loc[df_US['Year'] == 1990]

non_renewable = (df_US_1990['Nuclear'] + df_US_1990['Total Fossil Fuels']) /␣
 ↪df_US_1990['Gross Electricity']
renewable = (df_US_1990['Solar'] + df_US_1990['Wind'] +␣
 ↪df_US_1990['Geothermal']) / df_US_1990['Gross Electricity']
unaccounted = 1 - non_renewable - renewable

pie_inputs = [non_renewable.values[0], renewable.values[0], unaccounted.
 ↪values[0]]

ax1 = plt.subplot(grid[0,0:1])
labels = ['Non-Renewable', 'Renewable', 'Unaccounted For']
wedges, texts, autotexts = ax1.pie(pie_inputs,
                                   autopct = '%1.1f%%',
                                   shadow = True,
                                   startangle = 90,
                                   textprops = dict(color ="white"))
ax1.set_title("Non-Renewable vs Renewable Energy in the United States - 1990")

ax1.legend(wedges, labels, title="Energy", loc="center left",␣
 ↪bbox_to_anchor=(1, 0, 0.5, 1))
```

```
df_US_2018 = df_US.loc[df_US['Year'] == 2018]

non_renewable = (df_US_2018['Nuclear'] + df_US_2018['Total Fossil Fuels']) /␣
 ↪df_US_2018['Gross Electricity']
renewable = (df_US_2018['Solar'] + df_US_2018['Wind'] +␣
 ↪df_US_2018['Geothermal']) / df_US_2018['Gross Electricity']
unaccounted = 1 - non_renewable - renewable

pie_inputs = [non_renewable.values[0], renewable.values[0], unaccounted.
 ↪values[0]]

ax2 = plt.subplot(grid[0,1:2])
labels = ['Non-Renewable', 'Renewable', 'Unaccounted For']
wedges, texts, autotexts = ax2.pie(pie_inputs,
                                   autopct = '%1.1f%%',
                                   shadow = True,
                                   startangle = 90,
                                   textprops = dict(color ="white"))
ax2.set_title("Non-Renewable vs Renewable Energy in the United States - 2018")

plt.show()
```

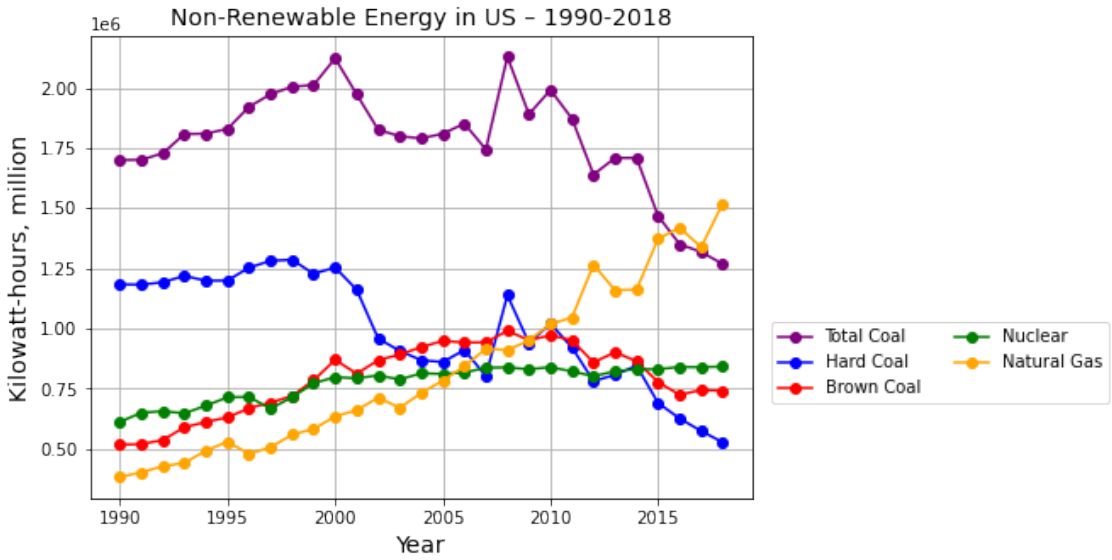


[9]:
```
plt.figure(figsize=(7, 5))
plt.plot(df_US['Year'], df_US['Solar'], color='Pink', marker='o')
plt.plot(df_US['Year'], df_US['Wind'], color='Blue', marker='o')
plt.plot(df_US['Year'], df_US['Geothermal'], color='Red', marker='o')
plt.title('Rise of Renewable Energy in US - 1990-2018', fontsize=14)
plt.xlabel('Year', fontsize=14)
plt.ylabel('Kilowatt-hours, million', fontsize=14)
```

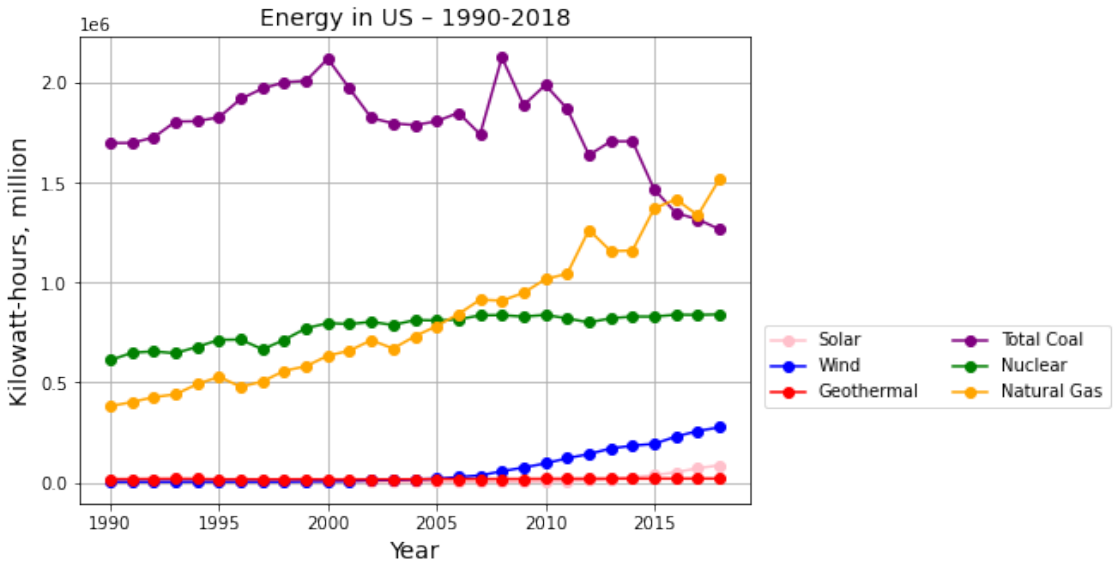
```
plt.grid(True)

plt.legend(['Solar', 'Wind', 'Geothermal'], bbox_to_anchor=(1.3, 0.4))
plt.show()
```



[10]:
```
plt.figure(figsize=(7, 5))
plt.plot(df_US['Year'], df_US['Total Coal'], color='Purple', marker='o')
plt.plot(df_US['Year'], df_US['Hard Coal'], color='Blue', marker='o')
plt.plot(df_US['Year'], df_US['Brown Coal'], color='Red', marker='o')
plt.plot(df_US['Year'], df_US['Nuclear'], color='Green', marker='o')
plt.plot(df_US['Year'], df_US['Natural Gas'], color='Orange', marker='o')
plt.title('Non-Renewable Energy in US - 1990-2018', fontsize=14)
plt.xlabel('Year', fontsize=14)
plt.ylabel('Kilowatt-hours, million', fontsize=14)
plt.grid(True)

plt.legend(['Total Coal', 'Hard Coal', 'Brown Coal', 'Nuclear', 'Natural Gas'],␣
 ↪bbox_to_anchor=(1.55, 0.4), ncol=2)
plt.show()
```

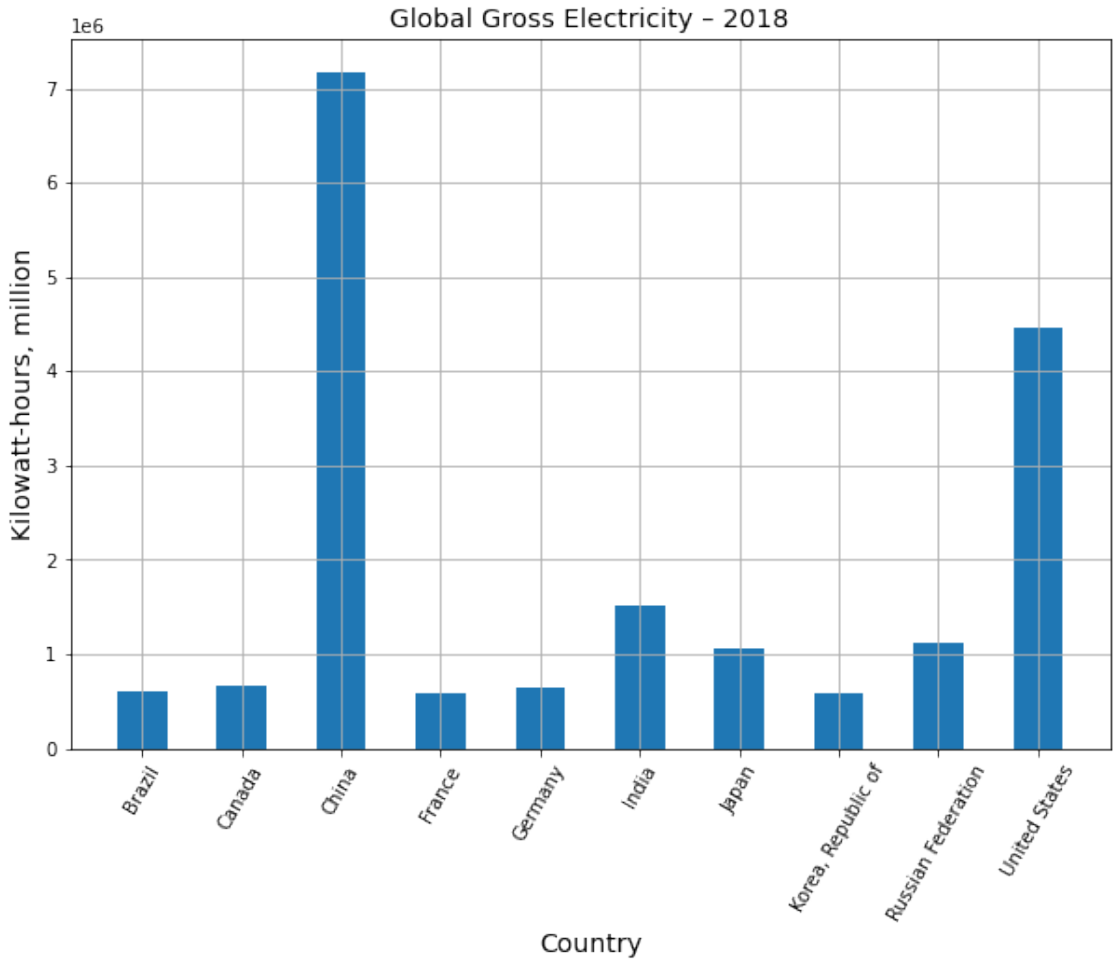


[11]:
```
plt.figure(figsize=(7, 5))

plt.plot(df_US['Year'], df_US['Solar'], color='Pink', marker='o')
plt.plot(df_US['Year'], df_US['Wind'], color='Blue', marker='o')
plt.plot(df_US['Year'], df_US['Geothermal'], color='Red', marker='o')
plt.plot(df_US['Year'], df_US['Total Coal'], color='Purple', marker='o')
plt.plot(df_US['Year'], df_US['Nuclear'], color='Green', marker='o')
plt.plot(df_US['Year'], df_US['Natural Gas'], color='Orange', marker='o')
plt.title('Energy in US - 1990-2018', fontsize=14)
plt.xlabel('Year', fontsize=14)
plt.ylabel('Kilowatt-hours, million', fontsize=14)
plt.grid(True)

plt.legend(['Solar', 'Wind', 'Geothermal', 'Total Coal', 'Nuclear', 'Natural␣
 ↪Gas'], bbox_to_anchor=(1.55, 0.4), ncol=2)
plt.show()
```



Comparison of top ten countries in electricity production.

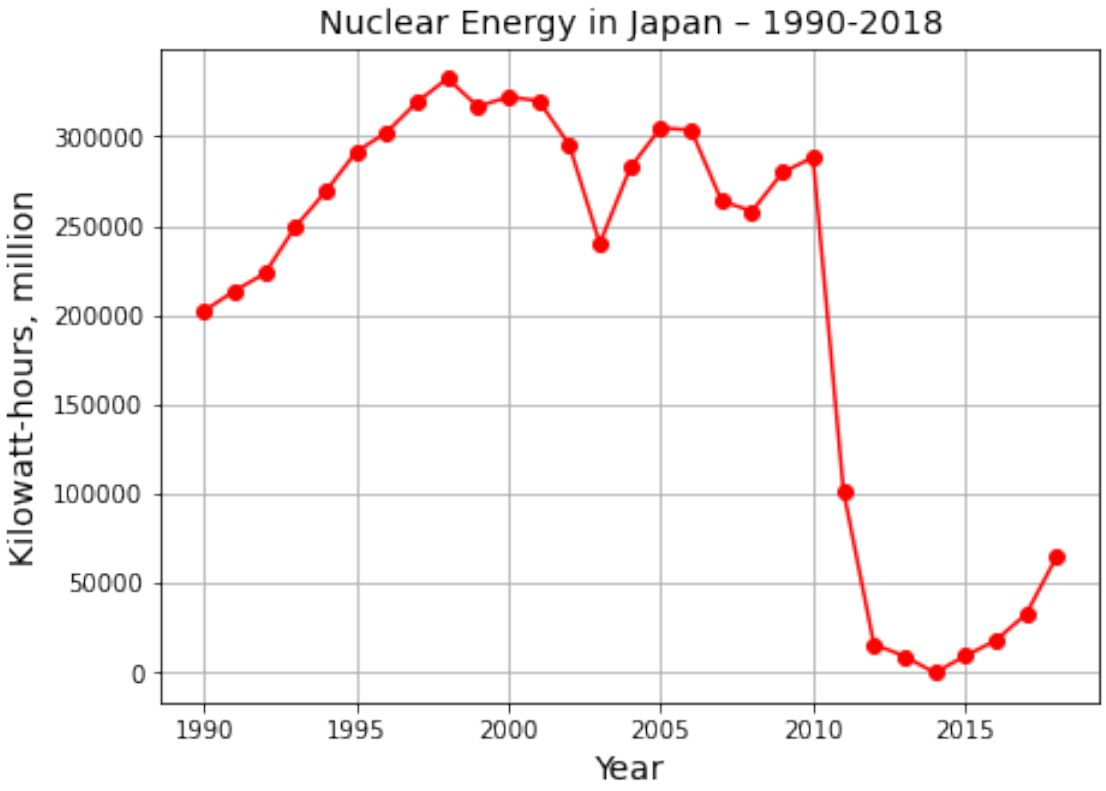
[12]:
```
plt.figure(figsize=(10, 7))

#only show data from 2018 and provide the top 10 producers of gross␣
 ↪electricity, which are all over 500,000 (in millions)
df_2018 = df.loc[(df['Year'] == 2018) & (df['Gross Electricity'] >= 500000)]
plt.bar(df_2018['Country or Area'],df_2018['Gross Electricity'],␣
 ↪align='center', width=0.5)
plt.xlabel('Country', fontsize=14)
plt.xticks(rotation=60)
plt.ylabel('Kilowatt-hours, million', fontsize=14)
plt.title('Global Gross Electricity - 2018', fontsize=14)
plt.grid(True)
plt.show()
```



An interesting side note that I enjoyed investigating was the sudden, complete drop in nuclear energy in Japan in the early 2010's.

This was due to the Tōhoku earthquake and tsunami that occured in 2011, which caused the subsequent Fukushima Daiichi nuclear disaster.

[13]:
```
japan = df.loc[(df['Country or Area'] == 'Japan')]
plt.figure(figsize=(7, 5))

plt.title('Nuclear Energy in Japan - 1990-2018', fontsize=14)
plt.plot(japan['Year'], japan['Nuclear'], color='Red', marker='o')
plt.xlabel('Year', fontsize=14)
plt.ylabel('Kilowatt-hours, million', fontsize=14)
plt.grid(True)
plt.show()
```


Nuclear Energy in Japan – 1990-2018
Kilowatt-hours, million
300000
250000
200000
150000
100000
50000
0
1990
1995
2000
2005
2010
2015
Year

[ ]: