US Energy Production Visualization

- How is our energy produced across the US?

- Kaggle inspired; web scraped

    - US Energy Information Administration

- 6 main types of production plus "Other"

Web Scraping

```python
if ".zip" in link:
    if "b" in link: pass
    else:
        wholeLink = "https://www.eia.gov/electricity/data/" + link
        zipLinks.append(wholeLink)
```

Code bypasses 'b' files

Create new directories according to the year being scrapped

Extract scrapped zip file into created directory

```python
path = os.path.join(parentDir, yearFile)
os.mkdir(path)
```

```python
with zipfile.ZipFile(path, 'r') as zippedFile:
    zippedFile.extractall(path)
```

ETL

- Jupyter
  - Used Pandas for import and transformation

- Transform
  - Renamed columns
  - Added a Year column
  - Dropped Duplicates
  - Concatenation
  - Separation into energy types

- Exported into CSVs

```python
y2007 = "Resource/y2007.csv"
y2007_df = pd.read_csv(y2007)
y2007_df
```

```python
y2007_df.rename(columns={'PLNTCODE': 'Plant_ID'}, inplace=True)

for row in y2007_df:
    y2007_df["Year"] = 2007

y2007_df.drop_duplicates(subset=['Plant_ID', 'Technology'], inplace=True)

y2007_df
```

```python
frames = [y2001_df, y2002_df, y2003_df, y2004_df, y2005_df, y2006_df, y2007_df, y2008_df, y2009_df,
          y2010_df, y2011_df, y2012_df, y2013_df, y2014_df, y2015_df, y2016_df, y2017_df, y2018_df]
combined_df = pd.concat(frames, ignore_index=True)
combined_df
```

```python
coal_df = combined_df.copy()
coal_df = coal_df.loc[coal_df['Technology'] == 'Coal']
coal_df
```

```python
coal_df.set_index("Plant_ID", inplace=True)
coal_df
```

```python
coal_df.to_csv('Resource/Cleaned/coal.csv', index=True, header=True)
```

ETL cont.

```
engine = create_engine(f'postgresql://postgres:{password}@localhost:5432/energy_db')

engine.table_names()
```

- ● SQL
  - ○ Used sqlalchemy
  - ○ Created new Database in pgAdmin
  - ○ .to_sql took care of the schema

- ● Json
  - ○ .to_json
  - ○ Orient type

∨ ⊞ Tables (7)
  › ⊞ coal
  › ⊞ hydroelectric
  › ⊞ naturalgas
  › ⊞ nuclear
  › ⊞ other
  › ⊞ solar
  › ⊞ wind

```
coal_df.to_sql(
    name = "coal",
    schema = 'public',
    con = engine,
    if_exists = 'replace',
    index = True
    )
```

```
coal_df.to_csv('Resource/Cleaned/coal.csv', index=True, header=True)
coal_df.to_json('Resource/Json/coal.json', orient='records', indent=4 )
```

```
{
    "Nameplate":0.2,
    "Technology":"Solar",
    "Latitude":33.5561,
    "Longitude":-112.2153,
    "State":"AZ",
    "Year":2001
},
```
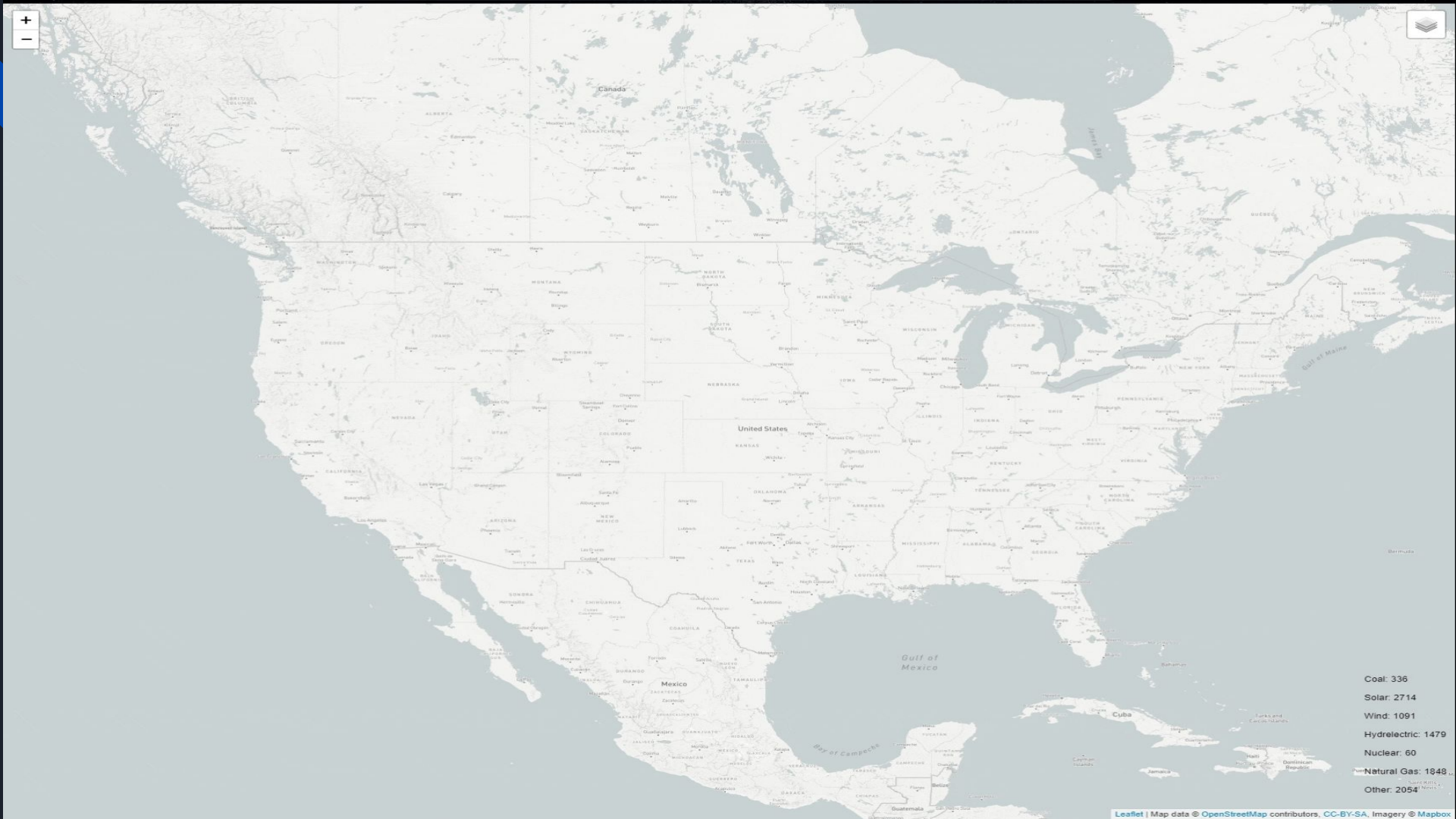
Javascript

- Used Leaflet and CitiBike as reference

- Finalized on what to visualize
  - Over time
  - Select energy types comparison

- Bloopers/Troubleshooting
  - JSONs couldn't be found
  - Try to use multiple JSONs

Javascript cont.

- For Looping and JSON format

- Extra Markers.min

- Extra Markers.CSS

- Icons

Coal: 336

Solar: 2714

Wind: 1091

Hydrelectric: 1479

Nuclear: 60

Natural Gas: 1848 ..

Other: 2054'

Possible Expansions

- Compare and contrast change in different energy plants over time

- Focus on single state vs. all of US

- Does proximity to plant type affect rates/prices?

- Track the flow of renewable energy sales to major industries?

End