

DATA TASK

Transforming the data-adding new rows

Q. "Which state has the highest number of homeless individuals per 10,000 people in the state?" Combine your new `pandas` skills to find out.

```
# Create indiv_per_10k col as homeless individuals per 10k state pop
homelessness["indiv_per_10k"] = 10000 * homelessness["individuals"] / homelessness["state_pop"]

# Subset rows for indiv_per_10k greater than 20
high_homelessness = homelessness[homelessness["indiv_per_10k"] > 20]

# Sort high_homelessness by descending indiv_per_10k
high_homelessness_srt = high_homelessness.sort_values("indiv_per_10k", ascending=False)

# From high_homelessness_srt, select the state and indiv_per_10k cols
result = high_homelessness_srt[["state", "indiv_per_10k"]]

# See the result
print(result)
```

	state	indiv_per_10k
8	District of Columbia	53.738
11	Hawaii	29.079
4	California	27.624
37	Oregon	26.636
28	Nevada	23.314
47	Washington	21.829
32	New York	20.392

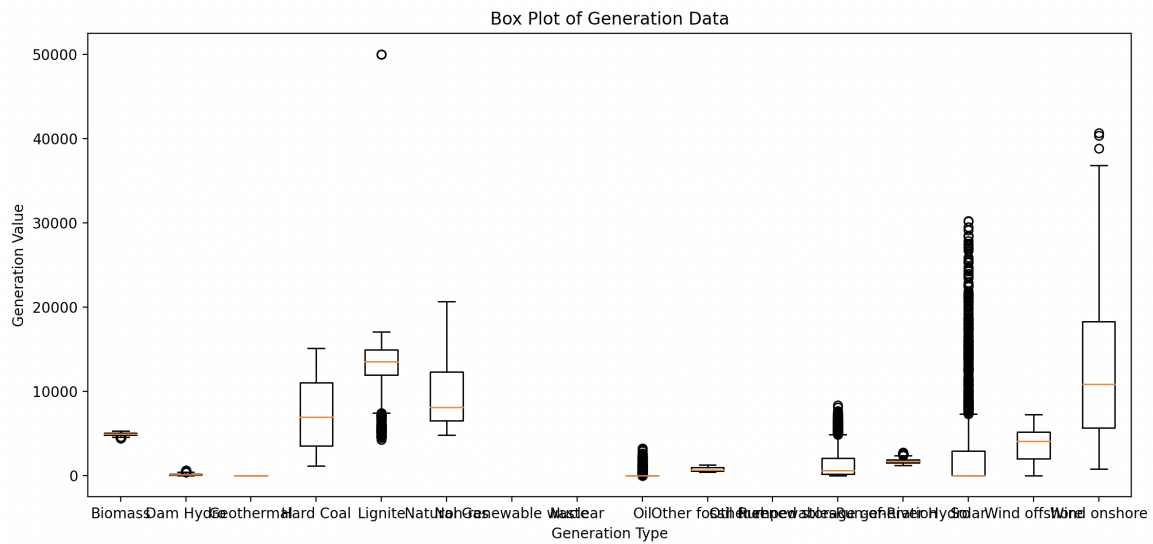
<script.py> output:

	state	indiv_per_10k
8	District of Columbia	53.738
11	Hawaii	29.079
4	California	27.624
37	Oregon	26.636
28	Nevada	23.314
47	Washington	21.829
32		

```
output to delete
['2022-12-30T03:00:00', 'France', 'Wind offshore', None], [
'2022-12-30T03:00:00', 'Denmark', 'Wind offshore', 1863.0],
['2022-12-30T03:00:00', 'Germany', 'Wind offshore', 5293.36], ['2022-12-30T03:00:00', 'France', 'Wind onshore', 8184.0], ['2022-12-30T03:00:00', 'Denmark', 'Wind onshore', 1863.0], ['2022-12-30T03:00:00', 'Germany', 'Wind onshore', 5293.36]]
```

Working with Pivot tables

plot of pivoted data before removing outliers



summary statistics

generation	Biomass	Dam Hydro	Geothermal	Hard Coal	Lignite \
count	2185.000000	2185.000000	2185.00	2185.000000	2185.000000
mean	4919.173844	151.097799	20.81	7412.907666	12915.738851
std	175.302858	118.158766	0.00	4141.446811	3273.990552
min	4441.340000	11.640000	20.81	1185.960000	4251.600000
25%	4845.480000	60.570000	20.81	3559.300000	11939.010000
50%	4949.410000	116.000000	20.81	6936.450000	13547.480000
75%	5038.390000	201.550000	20.81	11063.120000	14939.990000
max	5282.110000	620.750000	20.81	15090.510000	50000.000000

generation	Natural Gas	Non-renewable waste	Nuclear	Oil \
count	2185.000000	2176.000000	2174.000000	2185.000000
mean	9674.306380	1217.610455	3717.808155	176.502357
std	3945.188097	1818.679643	397.485338	401.290371
min	4844.560000	633.520000	2595.380000	30.780000
25%	6509.620000	1103.467500	3612.275000	30.780000
50%	8129.800000	1180.930000	3861.595000	30.780000
75%	12331.950000	1246.757500	3994.870000	30.780000
max	20676.910000	50000.000000	4069.570000	3244.820000

generation	Other fossil fuel	Other renewables	Pumped storage	generation \
count	2185.000000	2161.000000		2185.000000
mean	755.641057	383.509218		1420.811478
std	243.372696	81.807630		1732.063576
min	445.880000	105.100000		0.000000
25%	527.100000	308.930000		198.750000
50%	694.300000	356.710000		596.500000
75%	964.220000	414.030000		2072.500000
max	1263.600000	544.610000		8327.000000

generation	Run-of-River Hydro	Solar	Wind offshore	Wind onshore
count	2185.000000	2185.000000	2185.000000	2185.000000
mean	1729.998238	2891.293808	3635.345611	12449.090815
std	278.859663	5679.034120	1801.492634	8146.480876
min	1242.170000	0.000000	16.170000	774.530000
25%	1539.540000	0.000000	2022.790000	5655.860000
50%	1679.840000	0.000000	4081.200000	10870.980000
75%	1881.050000	2936.070000	5162.970000	18300.520000
max	2769.400000	30253.450000	7239.580000	40670.020000

gillian@Gillians-MacBook-Air prepwork % /usr/local/bin/python3 /Users/gillian/code/AGORA/prepwork/backen

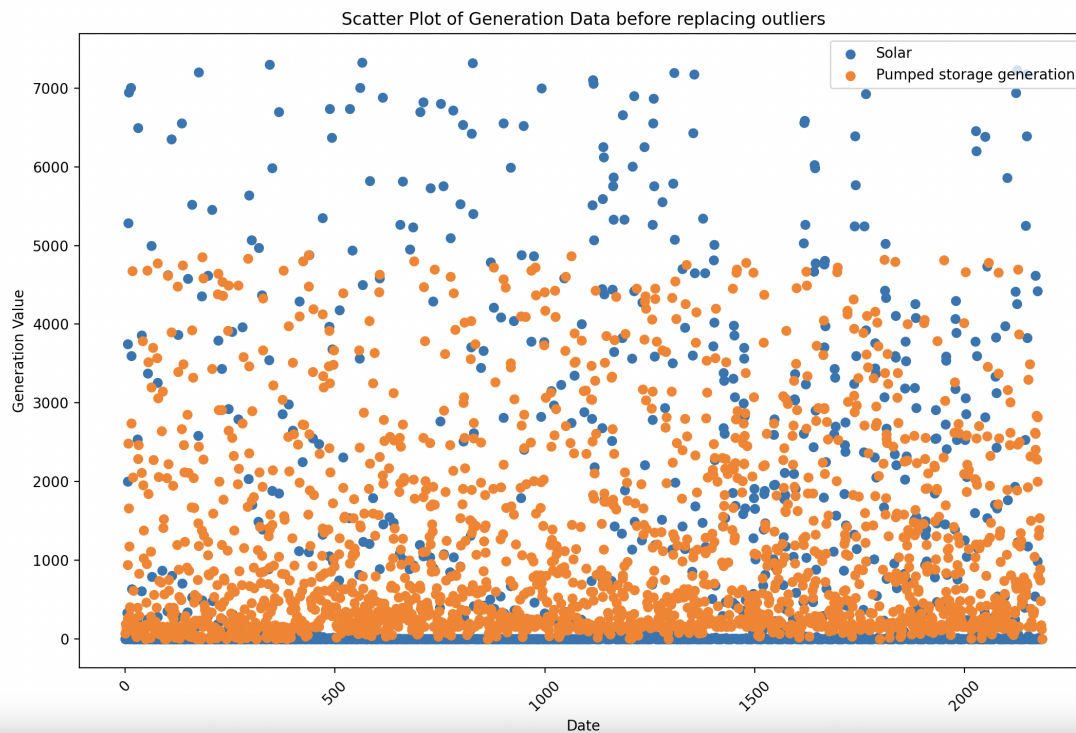
generation	Biomass	Dam Hydro	Geothermal	Hard Coal	Lignite \
count	2185.000000	2185.000000	2185.00	2185.000000	2185.000000

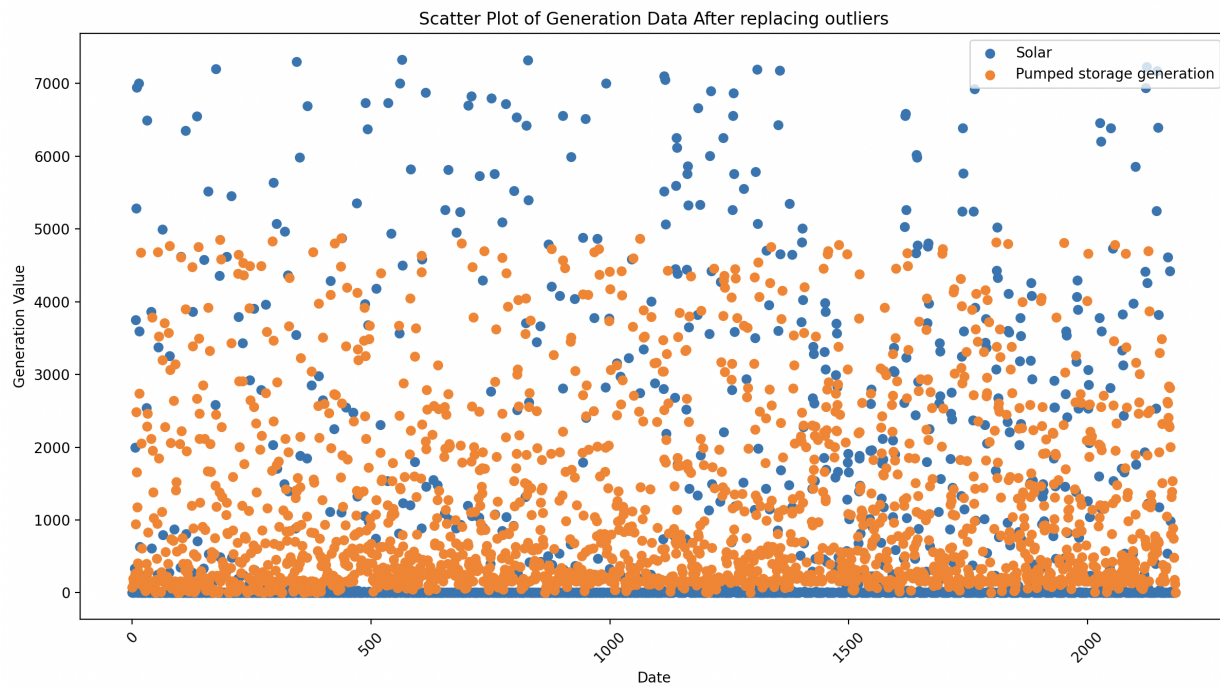
mean	4919.173844	151.097799	20.81	7412.907666	12915.738851
std	175.302858	118.158766	0.00	4141.446811	3273.990552
min	4441.340000	11.640000	20.81	1185.960000	4251.600000
25%	4845.480000	60.570000	20.81	3559.300000	11939.010000
50%	4949.410000	116.000000	20.81	6936.450000	13547.480000
75%	5038.390000	201.550000	20.81	11063.120000	14939.990000
max	5282.110000	620.750000	20.81	15090.510000	50000.000000

generation	Natural Gas	Non-renewable waste	Nuclear	Oil \
count	2185.000000	2176.000000	2174.000000	2185.000000
mean	9674.306380	1217.610455	3717.808155	176.502357
std	3945.188097	1818.679643	397.485338	401.290371
min	4844.560000	633.520000	2595.380000	30.780000
25%	6509.620000	1103.467500	3612.275000	30.780000
50%	8129.800000	1180.930000	3861.595000	30.780000
75%	12331.950000	1246.757500	3994.870000	30.780000
max	20676.910000	50000.000000	4069.570000	3244.820000

generation	Other fossil fuel	Other renewables	Pumped storage generation \
count	2185.000000	2161.000000	2185.000000
mean	755.641057	383.509218	1420.811478
std	243.372696	81.807630	1732.063576
min	445.880000	105.100000	0.000000
25%	527.100000	308.930000	198.750000
50%	694.300000	356.710000	596.500000
75%	964.220000	414.030000	2072.500000
max	1263.600000	544.610000	8327.000000

generation	Run-of-River Hydro	Solar	Wind offshore	Wind onshore
count	2185.000000	2185.000000	2185.000000	2185.000000
mean	1729.998238	2891.293808	3635.345611	12449.090815
std	278.859663	5679.034120	1801.492634	8146.480876
min	1242.170000	0.000000	16.170000	774.530000
25%	1539.540000	0.000000	2022.790000	5655.860000
50%	1679.840000	0.000000	4081.200000	10870.980000
75%	1881.050000	2936.070000	5162.970000	18300.520000
max	2769.400000	30253.450000	7239.580000	40670.020000





analyzing missing data

nullity.sum

generation

Biomass 0

Dam Hydro 0

Geothermal 0

Hard Coal 0

Lignite 0

Natural Gas 0

Non-renewable waste 9

Nuclear 11

Oil 0

Other fossil fuel 0

Other renewables 24

Pumped storage generation 0

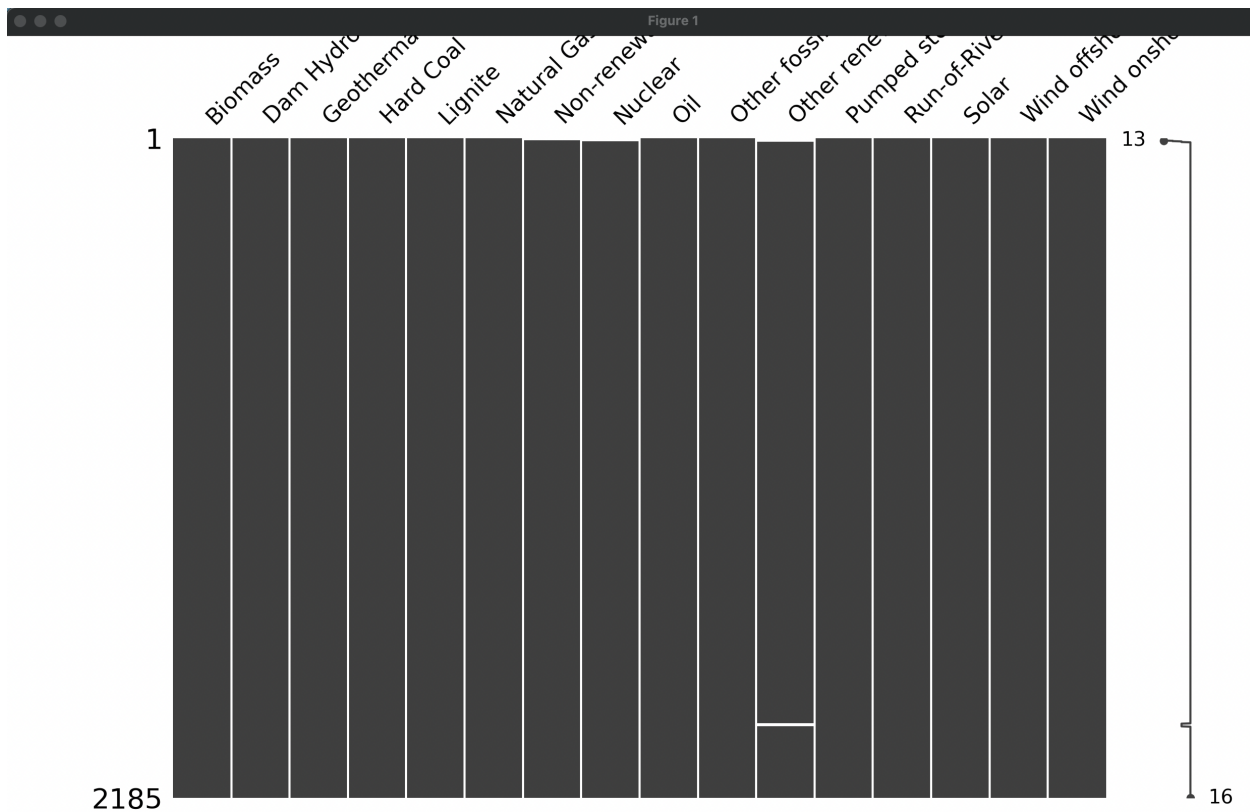
Run-of-River Hydro 0

Solar 0

Wind offshore 0

Wind onshore 0

dtype: int64



there is a correlation between non-renewable and nuclear

for nuclear and non-renewable waste missing data - missing data for nuclear infers missing data for waste due to **operational interdependencies**.

for other renewables and non-renewable waste: **operational interdependencies**-Biofuels, such as biodiesel or bioethanol, are derived from organic materials, typically crops or agricultural residues. The production process of biofuels can generate waste materials, such as spent biomass, wastewater, or solid residues, which may be classified as non-renewable waste.

other renewable and nuclear

