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

In [3]:
```
data=pd.read_csv('dataset.csv')
data
```

Out[3]:

| | VIN (1-10) | County | City | State | Postal Code | Model Year | Make | Model | E V |
|---|---|---|---|---|---|---|---|---|---|
| **0** | JTMEB3FV6N | Monroe | Key West | FL | 33040 | 2022 | TOYOTA | RAV4 PRIME | P E V (I |
| **1** | 1G1RD6E45D | Clark | Laughlin | NV | 89029 | 2013 | CHEVROLET | VOLT | P E V (I |
| **2** | JN1AZ0CP8B | Yakima | Yakima | WA | 98901 | 2011 | NISSAN | LEAF | B E V |
| **3** | 1G1FW6S08H | Skagit | Concrete | WA | 98237 | 2017 | CHEVROLET | BOLT EV | B E V |
| **4** | 3FA6P0SU1K | Snohomish | Everett | WA | 98201 | 2019 | FORD | FUSION | P E V (I |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **112629** | 7SAYGDEF2N | King | Duvall | WA | 98019 | 2022 | TESLA | MODEL Y | B E V |
| **112630** | 1N4BZ1CP7K | San Juan | Friday Harbor | WA | 98250 | 2019 | NISSAN | LEAF | B E V |
| **112631** | 1FMCU0KZ4N | King | Vashon | WA | 98070 | 2022 | FORD | ESCAPE | P E V (I |
| **112632** | KNDCD3LD4J | King | Covington | WA | 98042 | 2018 | KIA | NIRO | P E V (I |
| **112633** | YV4BR0CL8N | King | Covington | WA | 98042 | 2022 | VOLVO | XC90 | P E V (I |

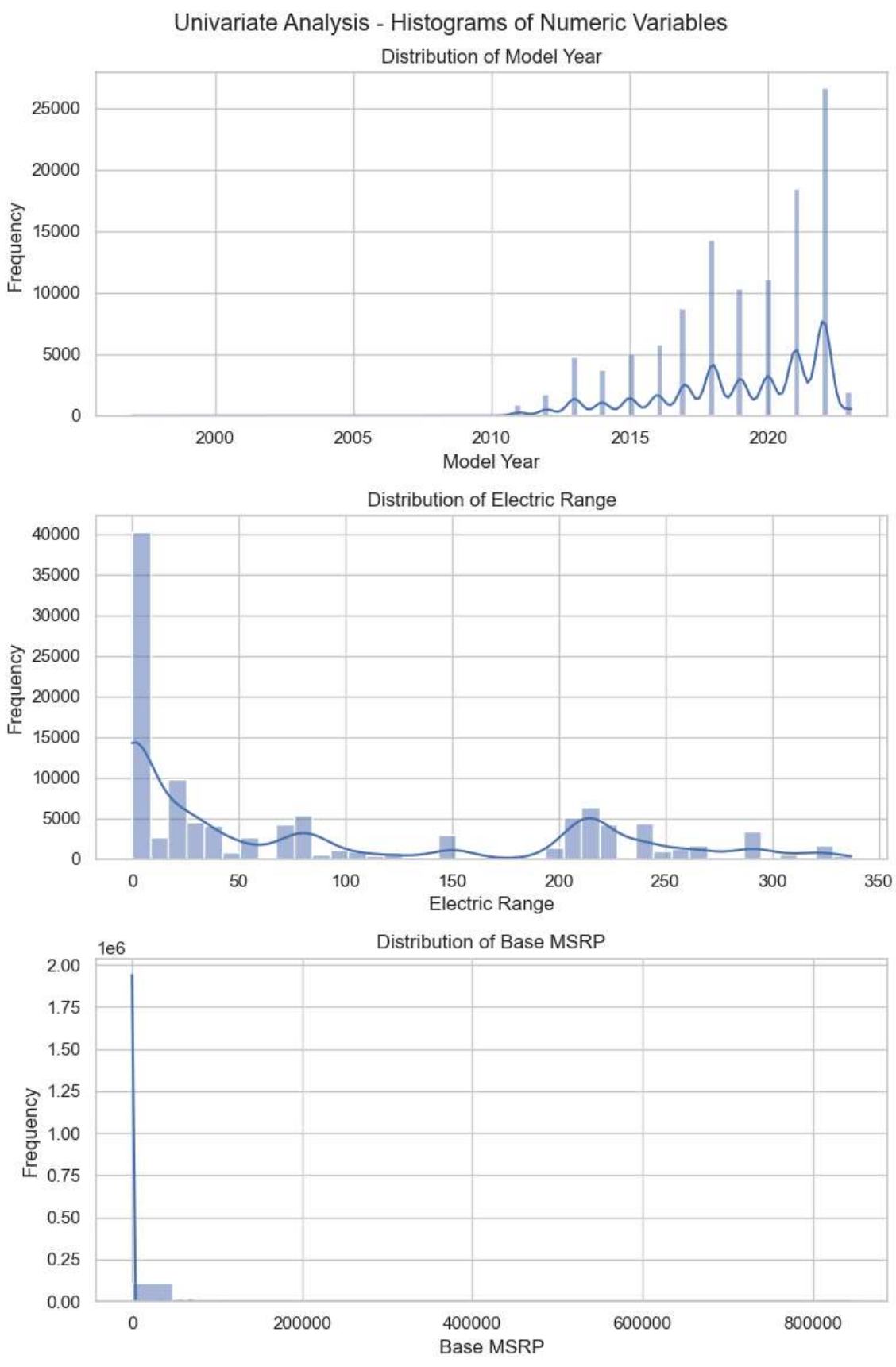
112634 rows × 17 columns

In [4]:
```
data.columns = data.columns.str.strip()
```

# # Univariate Analysis

In [5]:
```
sns.set(style="whitegrid")

# Descriptive statistics for numeric columns
numeric_columns = ['Model Year', 'Electric Range', 'Base MSRP']
print(data[numeric_columns].describe())

# Plot histograms for numeric columns
fig, axes = plt.subplots(len(numeric_columns), 1, figsize=(8, 12))
fig.suptitle('Univariate Analysis - Histograms of Numeric Variables')

for i, column in enumerate(numeric_columns):
    sns.histplot(data[column], kde=True, ax=axes[i])
    axes[i].set_title(f'Distribution of {column}')
    axes[i].set_xlabel(column)
    axes[i].set_ylabel('Frequency')

plt.tight_layout()
plt.show()
```

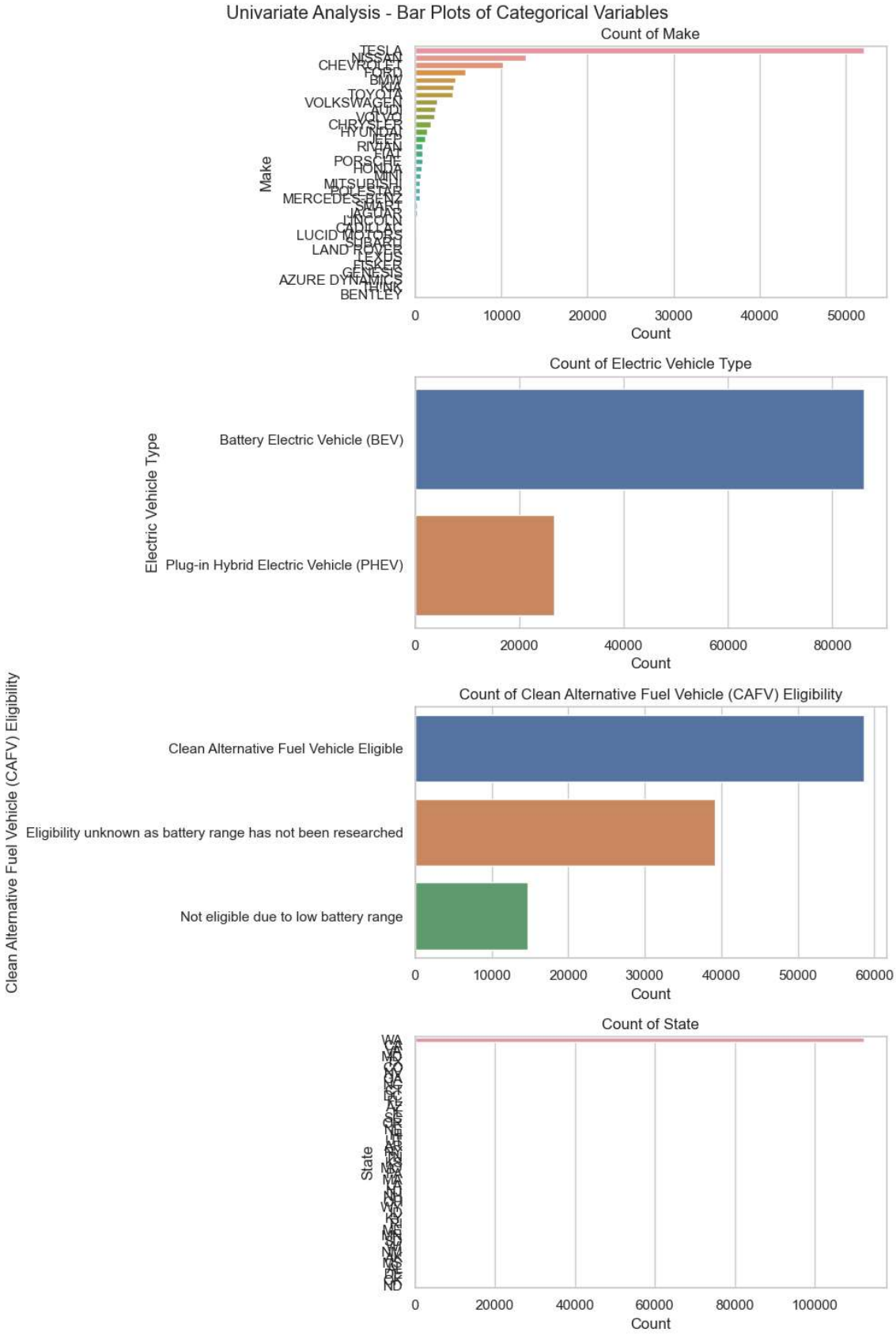
```
          Model Year  Electric Range      Base MSRP
count  112634.000000   112634.000000  112634.000000
mean     2019.003365       87.812987    1793.439681
std         2.892364      102.334216   10783.753486
min      1997.000000        0.000000       0.000000
25%      2017.000000        0.000000       0.000000
50%      2020.000000       32.000000       0.000000
75%      2022.000000      208.000000       0.000000
max      2023.000000      337.000000  845000.000000
```


Univariate Analysis - Histograms of Numeric Variables
Distribution of Model Year
Frequency
Model Year
Distribution of Electric Range
Frequency
Electric Range
Distribution of Base MSRP
1e6
Frequency
Base MSRP

# Frequency distribution for categorical columns

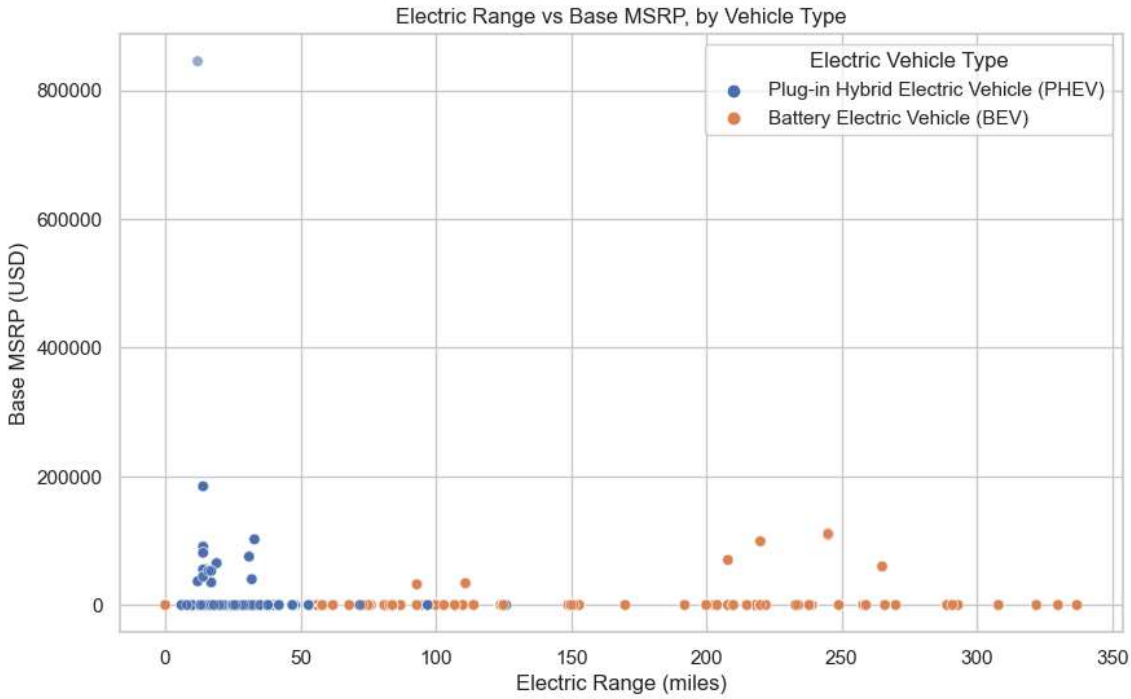
In [6]:
```
categorical_columns = ['Make', 'Electric Vehicle Type', 'Clean Alternative
print(data[categorical_columns].describe())

# Plot bar plots for categorical columns
fig, axes = plt.subplots(len(categorical_columns), 1, figsize=(10, 15))
fig.suptitle('Univariate Analysis - Bar Plots of Categorical Variables')

for i, column in enumerate(categorical_columns):
    sns.countplot(y=data[column], order=data[column].value_counts().index,
    axes[i].set_title(f'Count of {column}')
    axes[i].set_xlabel('Count')
    axes[i].set_ylabel(column)

plt.tight_layout()
plt.show()
```

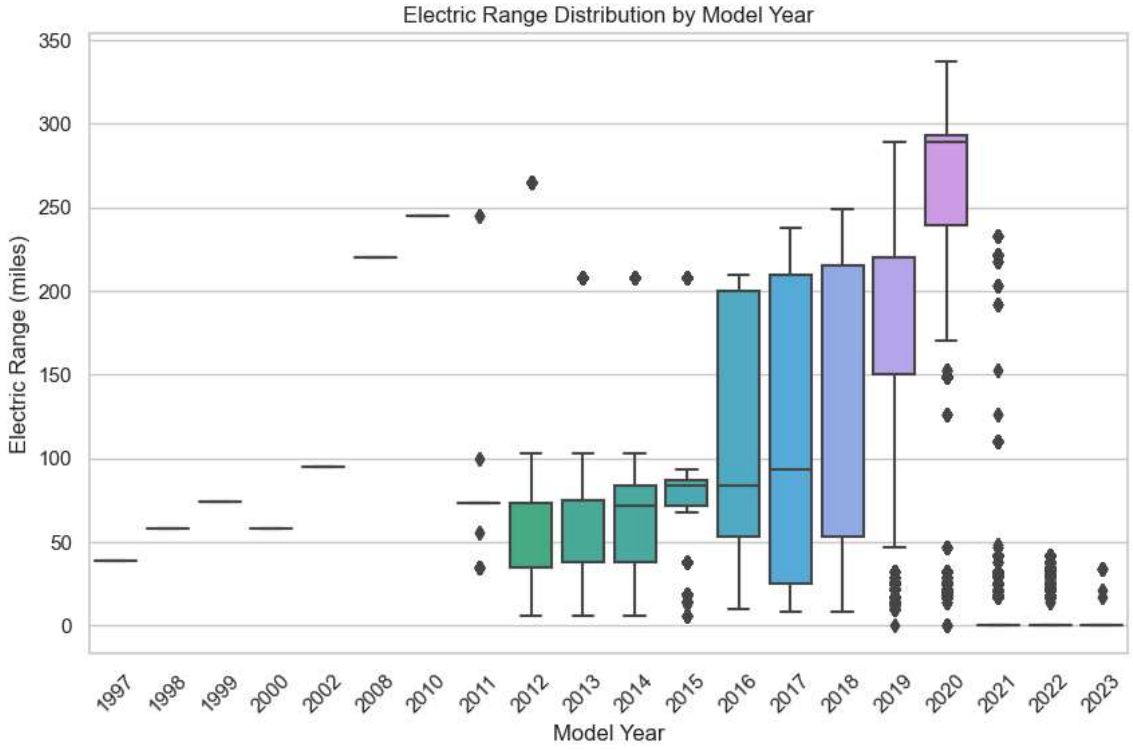
```
          Make           Electric Vehicle Type  \
count   112634                          112634
unique      34                               2
top       TESLA  Battery Electric Vehicle (BEV)
freq      52078                           86044

       Clean Alternative Fuel Vehicle (CAFV) Eligibility   State
count                                             112634  112634
unique                                                 3      45
top              Clean Alternative Fuel Vehicle Eligible      WA
freq                                               58639  112348
```


Univariate Analysis - Bar Plots of Categorical Variables
Count of Make
Make
TESLA
NISSAN
CHEVROLET
FORD
BMW
KIA
TOYOTA
VOLKSWAGEN
AUDI
VOLVO
CHRYSLER
HYUNDAI
JEEP
RIVIAN
FIAT
PORSCHE
HONDA
MINI
MITSUBISHI
POLESTAR
MERCEDES-BENZ
SMART
JAGUAR
LINCOLN
CADILLAC
LUCID MOTORS
SUBARU
LAND ROVER
LEXUS
FISKER
GENESIS
AZURE DYNAMICS
KIA
BENTLEY
0
10000
20000
30000
40000
50000
Count
Count of Electric Vehicle Type
Electric Vehicle Type
Battery Electric Vehicle (BEV)
Plug-in Hybrid Electric Vehicle (PHEV)
0
20000
40000
60000
80000
Count
Count of Clean Alternative Fuel Vehicle (CAFV) Eligibility
Clean Alternative Fuel Vehicle (CAFV) Eligibility
Clean Alternative Fuel Vehicle Eligible
Eligibility unknown as battery range has not been researched
Not eligible due to low battery range
0
10000
20000
30000
40000
50000
60000
Count
Count of State
State
WA
0
20000
40000
60000
80000
100000
Count

# Bivariate Analysis

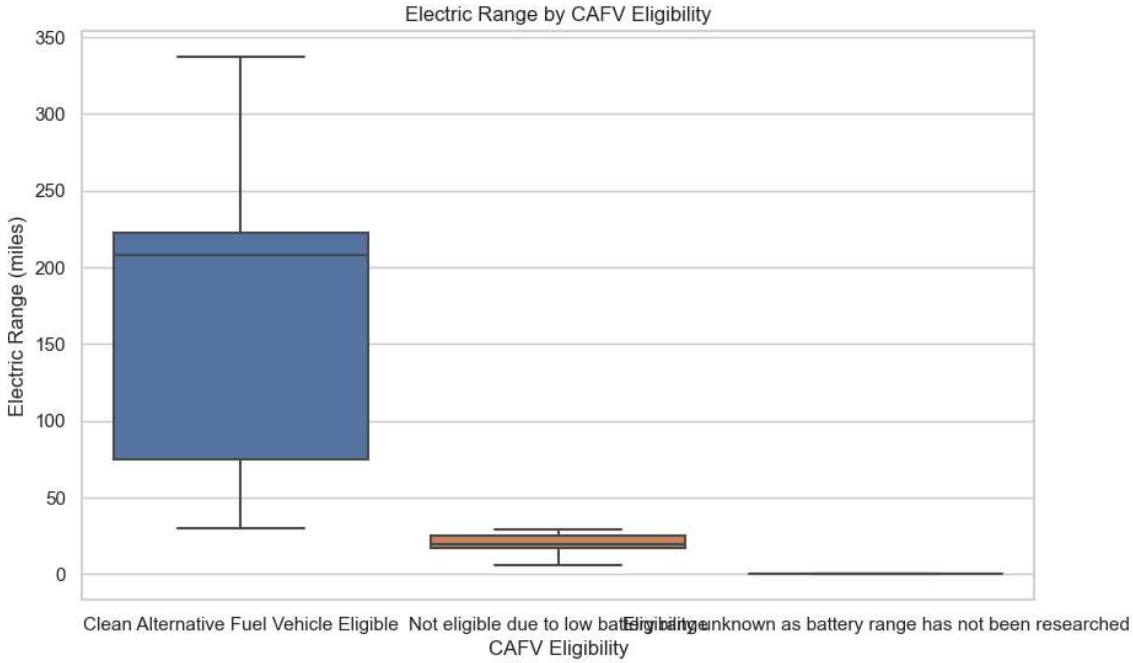
In [7]:
```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Electric Range', y='Base MSRP', hue='Electric Vehicle Ty
plt.title('Electric Range vs Base MSRP, by Vehicle Type')
plt.xlabel('Electric Range (miles)')
plt.ylabel('Base MSRP (USD)')
plt.legend(title='Electric Vehicle Type')
plt.show()

# Box plot of Electric Range grouped by Model Year
plt.figure(figsize=(10, 6))
sns.boxplot(x='Model Year', y='Electric Range', data=data)
plt.title('Electric Range Distribution by Model Year')
plt.xlabel('Model Year')
plt.ylabel('Electric Range (miles)')
plt.xticks(rotation=45)
plt.show()
```






# Box plot of Electric Range grouped by CAFV Eligibility

In [9]:
```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Clean Alternative Fuel Vehicle (CAFV) Eligibility', y='Elect
plt.title('Electric Range by CAFV Eligibility')
plt.xlabel('CAFV Eligibility')
plt.ylabel('Electric Range (miles)')
plt.show()
```



In [ ]: