1. Generate summary statistics (including skewness and kurtosis) of the electric range for the whole dataset. Comment on the results.

- Summary Statistics:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6.0	32.0	84.0	127.1	215.0	337.0

- Skewness: 0.359528
- Kurtosis: -1.392596
- From the observation, we can see that the mean value of the electric range is a wide distance from the median value of the electric range, so there is a high chance this is a right-skewed distribution. This is proved when we calculate and get the result of 0.359528 for the skewness of the electric range. This number is greater than 0 which means it is a positive or right-skewed distribution. In addition, we know that this is a platykurtic distribution when we get the kurtosis value of -1.392596 for the electric range.

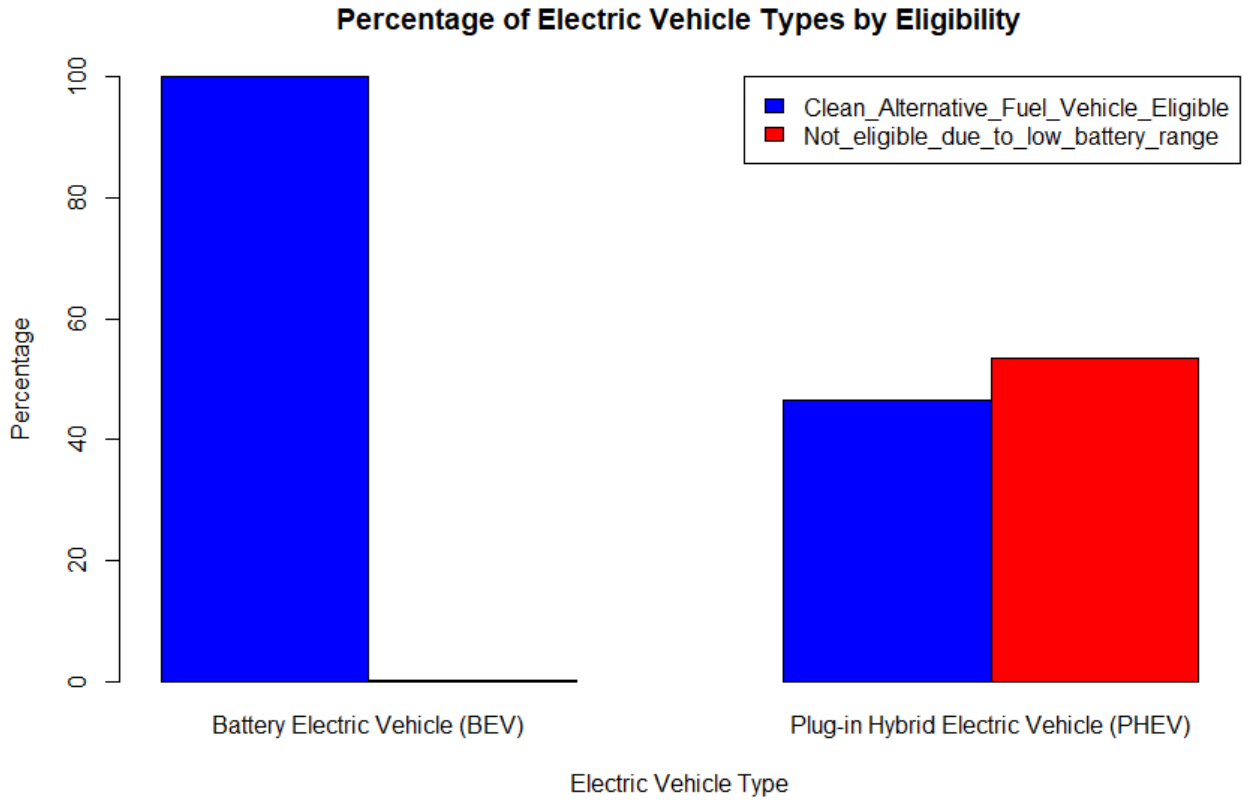
2. Calculate the frequency distribution of the electric range of vehicles (hint: first divide the variables into certain categories and then find out the distribution). Comment on the results.

0-100	101-200	>200
42610	6596	30436

From the observation, we can see that 42610 cars have an electric range less than or equal to 100, 6596 cars have an electric range between 101 and 200, and 30436 cars have an electric range greater than 200.

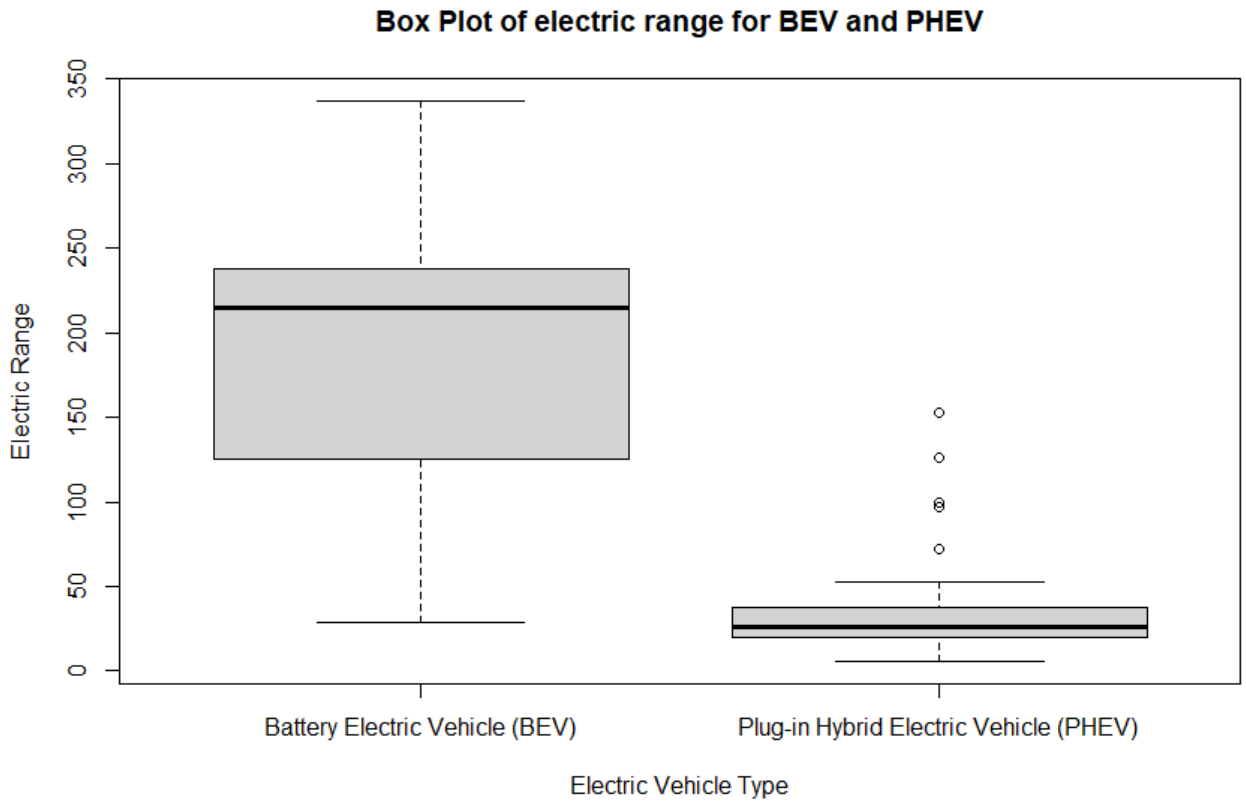
3. Generate a cluster bar plot by placing the percentage of vehicles on the y-axis and electric vehicle types on the x-axis. Each electric vehicle type should be divided into two bars based on the variable "clean alternative fuel vehicle eligibility". Comment on the results.

	BEV	PHEV
Clean_Alternative_Fuel_Vehicle_Eligible	99.98	46.56
Not_eligible_due_to_low_battery_range	0.02	53.44



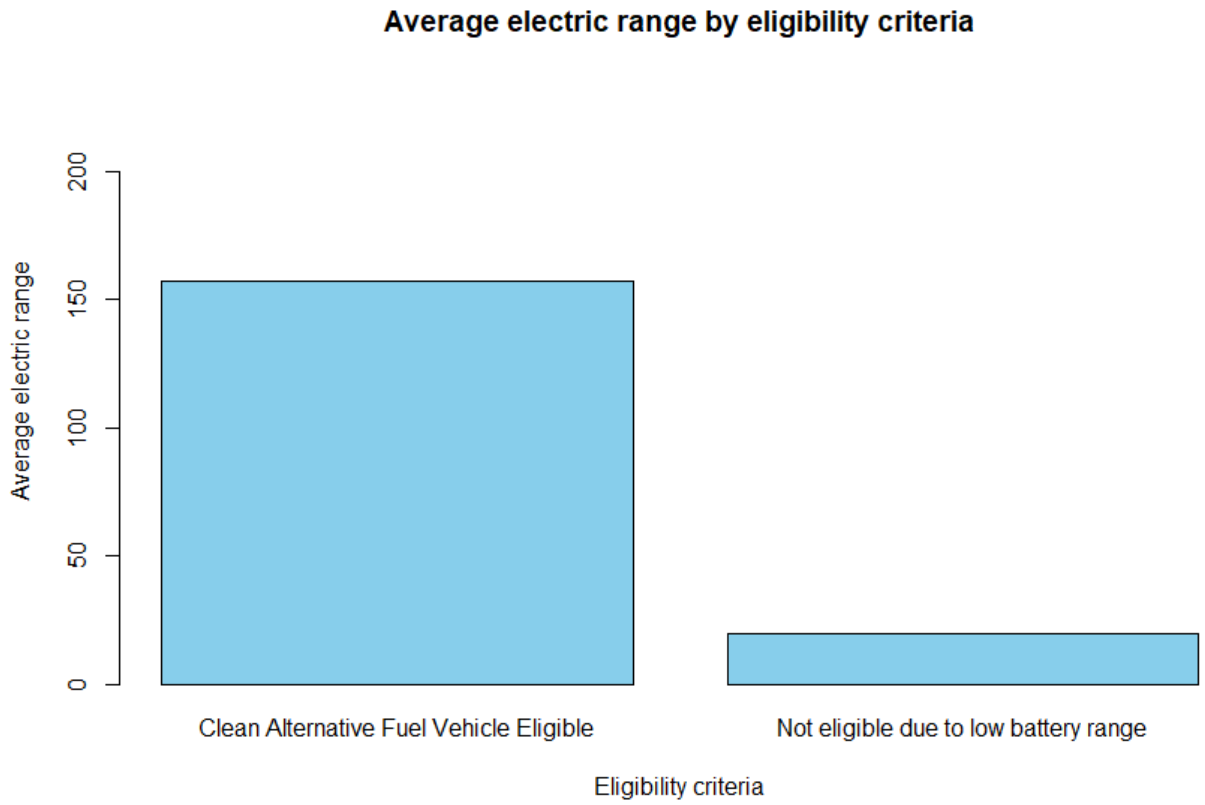
From the observation, we can see that there are two vehicle types: Battery Electric Vehicle (BEV), Plug-in Hybrid Electric Vehicle (PHEV). For BEV, 99.8% of the vehicles are eligible for clean alternative fuel. On the other hand, 53.44% of PHEV are not eligible for clean alternative fuel because of low battery range.

4. Generate box plots of the electric range of vehicles for Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs). Compare the plots in terms of median, spread, skewness, and outliers (if any).



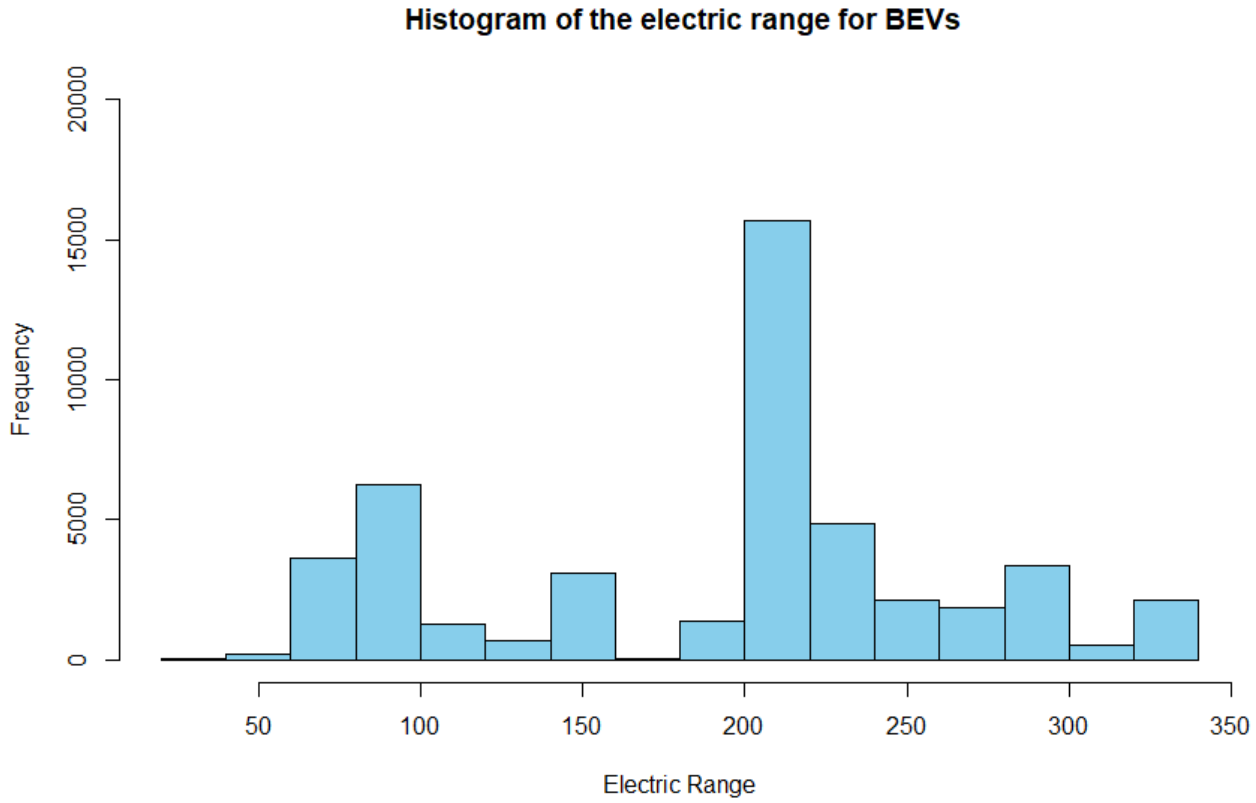
From the observation, we can see the position of the median line of BEV is higher than that of PHEV, indicating a potentially higher electric range for BEV compared to PHEV. In addition, the length of the box for BEV is longer than PHEV which implies a larger spread of the data. It also means the corresponding vehicle type has more variability in the electric range. In addition, from the graph, we can see the median lines of BEV and PHEV are not centered in the box which suggests skewness in the data distribution. Finally, there are no outliers for the box plot of BEV. On the other hand, there are 5 outliers for the ones of PHEV.

5. Find the average values of the electric range of the vehicles that meet the eligibility criteria of "clean alternative fuel vehicle" and that do not meet so, respectively. Generate a bar plot by placing average values of the electric range of vehicles on the y-axis and the two categories of clean alternative fuel vehicle eligibility on the x-axis. Comment on the plot.



From the observation, we can see the average electric ranges of vehicles in the “clean alternative fuel vehicle” criteria are higher than those in the "not eligible due to low battery range." This means the vehicles in the “clean alternative fuel vehicle” criteria have a wider and higher electric range than those in the "not eligible due to low battery range" criteria.

6. Generate a histogram of the electric range for Battery Electric Vehicles (BEVs). Interpret the plot. Also, calculate the percentage of vehicles that fall under a 3-standard deviation of this distribution.



The percentage of vehicles that fall under a 3-standard deviation of this distribution is 100%. From the observation, we can see this histogram has a normal distribution that is symmetric about the mean, with data near the mean being more frequent in occurrence than data far from the mean. This is also proved when we know 100% of vehicles fall under a 3-standard deviation of this distribution which implies that all the data points fall within three standard deviations on either side of the mean.