3. a) 46.8599

b) 50.6625

c) 118.1177

d) 10.8682

e) range $= y_{(n)} - y_{(1)} = 60.174 - 20.030 = 40.144$

f) 12.5905

g) ii) the number of kilometers driven

iii) the frequency that the number of kilometers driven appears

iv) The heights of each bar represent the frequencies of the number of kilometers driven ~~that~~ in the interval from the left number to the right number on the x axis. For example, the height of first bar represent the frequency of the number of kilometers driven in the interval from 200 to 300. They sum to 20, which is the number of S cars.

v) Yes. Because the histograms can show the distribution of the data by PIS in course notes and the

numbers of Car S and Car H in the sample are the same, which means that both of them are 20, and the intervals on x axis and y axis remain the same for S and H cars, and the data in the plot is distance driven for Car S and Car H, this is a good plot to use.

h) The distribution of gas consumption of S is negatively skewed because the part below the median is longer than the part above the median. Also, by observation, we get $q(0.25)$ approximates 40, and mean approximates 50, and $q(0.75)$ approximates 54, and so $IQR = q(0.75) - q(0.25) = 14$. Moreover, we get the minimum value is 20, and the maximum value is 60, and so the range $= 60 - 20 = 40$. Also, there exists outliers, which equal to 20.

i) ii) The explanatory variate is the number of kilometers driven, and the response variate is the gas consumed.

iii) By using cor(gas S $km, gas S $gas) in Rstudio, we get Sample correlation $= r = 0.9720875 > 0$. Use this,

and observation on the scatterplot, we get that there exists positive linear relationship between the two variates since as distance driven ↑ increases, gas consumed ↑ increases as shown on the scatterplot and $r = 0.9720875 > 0$.