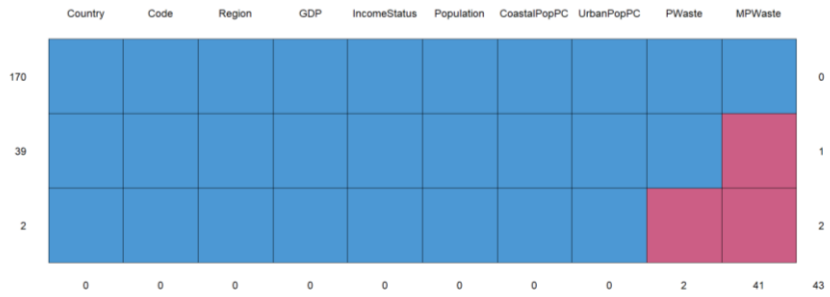
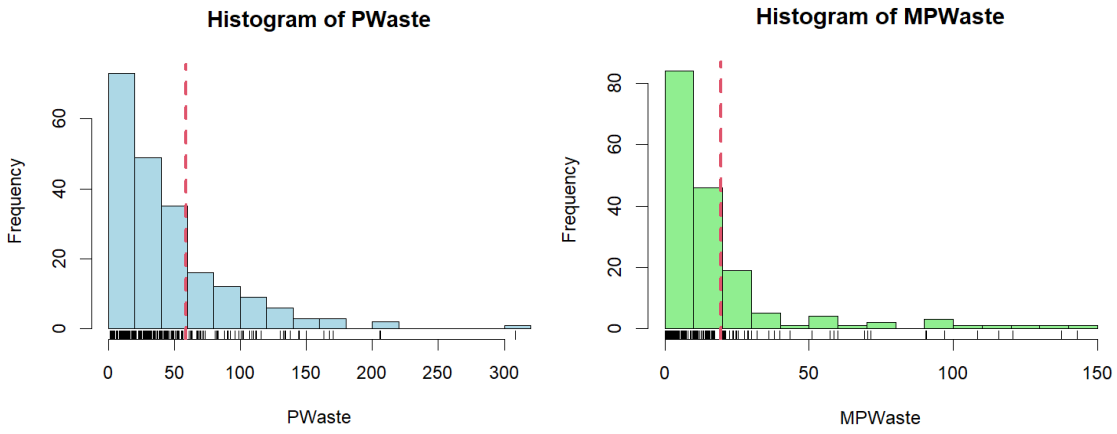
Q1:

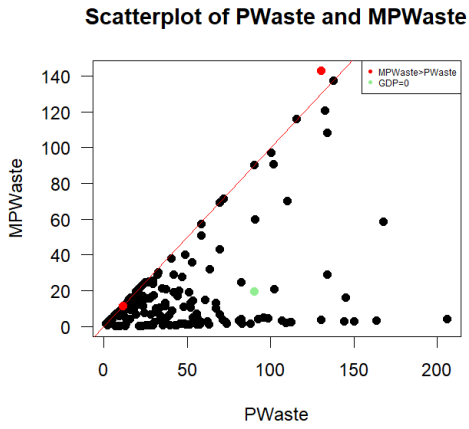


Plot 1.1

Plot 1.2 & Plot 1.3



Plot 1.4



Relevant features:

(1) 2 patterns of missingness. There are 170 values that are completed, and 39 values that are only missing 'MPWaste' and 2 values that are both missing 'MPWaste' and 'PWaste'. Perhaps the value wasn't collect or it was hard to determine whether Plastic Waste is well-managed, which leads to these missing values in 'MPWaste'.

(2) Both distributions of 'PWaste' and 'MPWaste' are right-skewed, which possibly means that most of countries don't produce so much plastic waste and most of the plastic waste are well-managed among them.

(3) There's a huge gap between 200 and 300, and a small gap between 175 and 200. The red

line shows in the plot means the 3rd quantile, which is around 55. Consequently, the observations which plastic waste are more than 200 can be seen as potential outliers. In Plot 1.3, there's also a small gap between 70 and 90, and the 3rd quantile is around 20. Moreover, from values beyond 100, there are typically only 1 to 2 observations, and they are far from the third quartile. Therefore, they can be considered as potential outliers.

(4) Two red points in Plot 1.4 which indicates that the values of their observations' 'MPWaste' is greater than 'PWaste'. The 'MPWaste' is the mismanaged plastic waste, which should be less than the amount of plastic waste. Moreover, there's a green dot shows above the red line, which indicates the value of GDP of that observation is 0. These three observations may be wrong recorded ,which are considered strange values here.

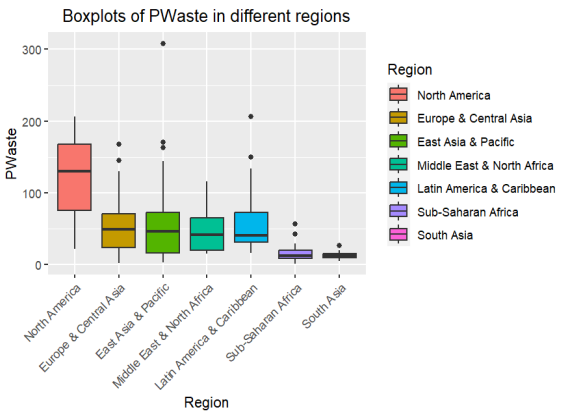
Implications:

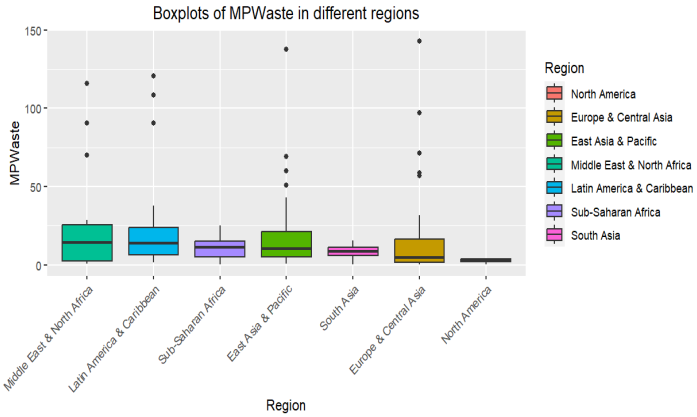
(1) The observations with 0 GDP should be rechecked that this situation shouldn't happen in the real world. This may warrant further investigation for the values of two types of waste are not considered potential outliers.

(2) Most of the missing values are missing the value of 'MPWaste', which may impact the further analysis on the relationship of 'MPWaste' with other variables.

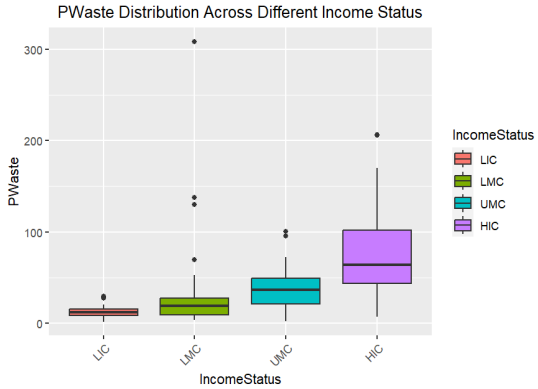
Q2

Plot 2.1 & 2.2





Plot 2.3 & 2.4



Relevant features:

(1) North America generated very large amount of plastic waste, while only few of them are mismanaged. South Asia generated limited amount of plastic waste and most of plastic

waste in Middle East & North Africa are not well-managed.

(2) The difference in plastic waste of North America is greater than other regions, while the ranges in mismanaged plastic waste of all regions except North America are nearly the same. According to this point, we can assume that North America may develop a great system and highly legal cognition to well manage the plastic waste.

(3) The higher the income, the greater the value of plastic waste. Places with low-income levels, MPWaste tends to be higher, while regions with high-income levels have the lowest MPWaste levels. However, more outliers are shown in high income regions, which can attribute to the great difference among different regions.

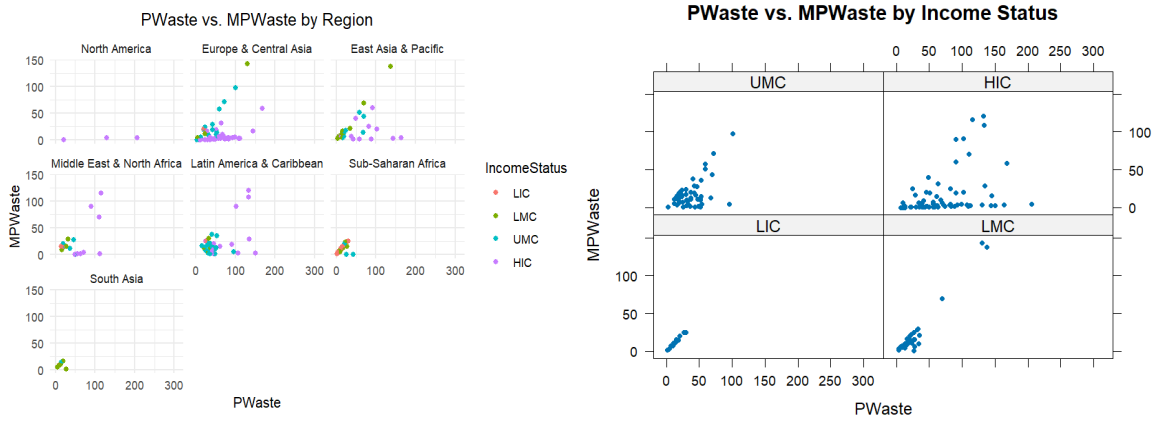
Implications:

(1) The income disparity varies across regions, which can explain why regions like Europe exhibit low MPWaste values but still have a considerable number of very high outliers. This may affect the variance and bias of the model in the future.

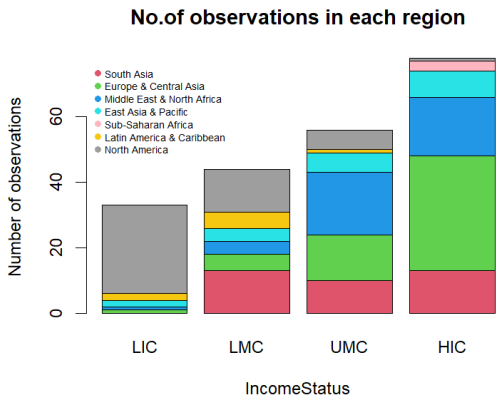
(2) The huge gaps in plot 2.1 between specific regions may helpful to the classification modelling.

Q3

Plot 3.1 & 3.2



Plot 3.3



Relevant features:

(1) 'PWaste' has a slightly positive relationship with 'MPWaste'. With the income growing, the range of two variables also getting larger.

(2) The lower the income, the tighter and more linear the relationship between PWaste and MPWaste, whereas, conversely, the higher the income level, the more dispersed their

relationship becomes.

(3) In contrast to high-income areas, low-income regions exhibit a nearly one-to-one positive relationship between these two variables, whereas in high-income areas, this positive relationship has been gentler, except 'Middle east & North Africa' and 'Latin America & Caribbean'.

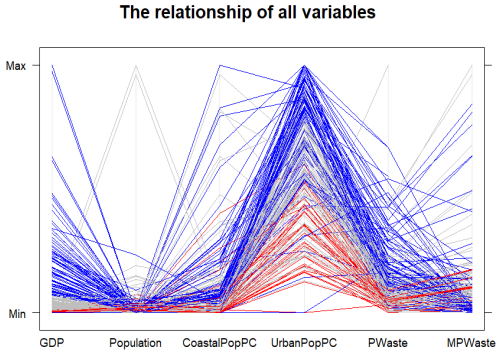
Implications:

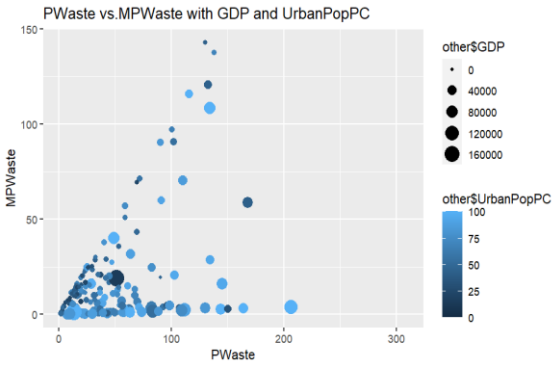
(1) The data is not even in each region. There are only 3 observations in North America and they're all belong to high income category, which may affect the universality of the model.

(2) The relationship between the two variables in high-income areas of 'Middle East & North Africa' and 'Latin America & Caribbean' differs from that of other regions. In these areas, there are relatively few observations with extremely high and irregular MPWaste values among high-income observations. The lack of a reasonable explanation for this phenomenon may contribute to increased variance in the model's data fitting.

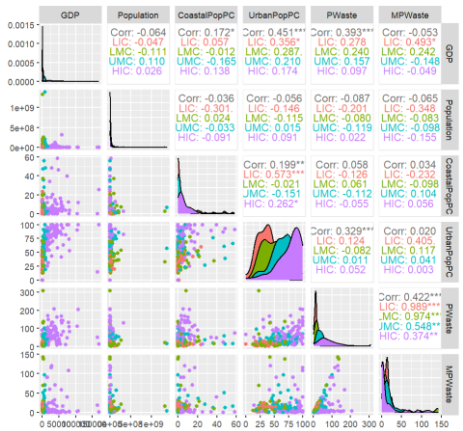
Q4

Plot 4.1 & 4.2





Plot 4.3



Relevant features:

(1) Higher GDP results in higher amount of plastic waste and less proportion of mismanaged plastic waste. Same pattern shows in CoastalPopPC and UrbanPopPC, while population shows the opposite.

(2) GDP has a slightly positive relationship with 'PWaste' but a clearly negative one with 'MPWaste'. UrbanPopPC shows both positive relationship with two variables. More people live in urban, more plastic waste are created and higher chances of mismanagement.

(3) CoastalPopPC has slightly negative relationship with two variables, while with population grows, both two variables decline sharply.

(4) Cities with smaller GDP and fewer UrbanPopPC tend to be closer to a 1:1 positive correlation line between the two variables; conversely, they tend to be closer to the x-axis as these variables decrease. There may be not enough money to support optimal management for those cities with smaller GDP.

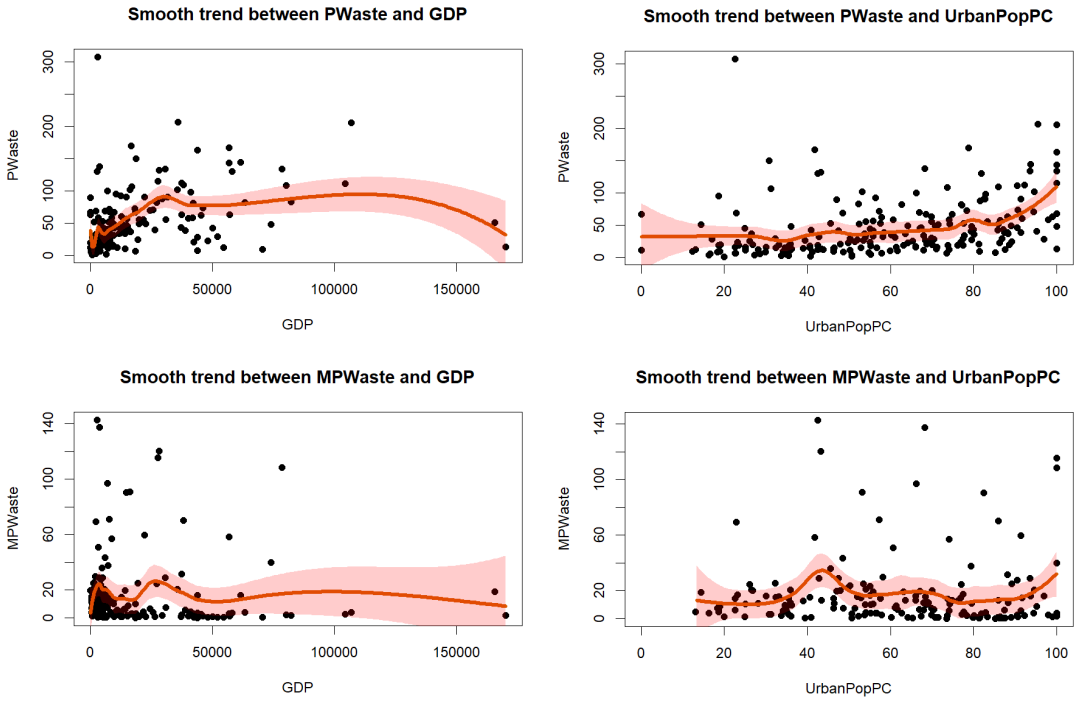
Implications:

(1) The span of population is so limited that can hardly discover any effect it brings to the relationship between 'PWaste' and 'MPWaste', which may overlook some patterns in it.

(2) The clear pattern shows in GDP and UrbanPopPC can be helpful to the regression model and classification model, such as decision trees and random forests.

Q5

Plot 5.1~5.4



Relevant features:

(1) As GDP grows, PWaste initially increases until reaches $50000, and then sharply decreases after stabilizing from $50000 to $120000, while MPWaste fluctuates initially until $50000, stabilizes, and exhibits a slight downward trend.

(2) As UrbanPopPC increases, PWaste remains stable initially before increasing, while MPWaste shows some fluctuation in the middle and a slight increase towards the end.

Implications:

(1) The trend between PWaste & UrbanPopPC shows is helpful to generating a model for regression to discover the potential relationship between them, such as linear regression.

(2) There's a slight trend between MPWaste and GDP, which may be suitable for splines.

(3) There's no clear trend between MPWaste & UrbanPopPC and PWaste & GDP; the variations are too small, requiring more data to support the analysis.