Data Analytics and Mathematical Statistics – Tutorial 2

1. Let's try plotting the bar plot for "Region" to see the counts in each region. Plot as well the histogram for SO_2, and vary the number of bins. What do we observe?

2. Compare the histogram with those for PSI, PM10 and rainfall. What can we say about the shapes?

3. Let's plot the dotplots for all five regions. Can we tell roughly where the means are for each region, and the relative spread between them? How about unusual data features like outliers and skewness?

4. Let us plot the boxplots now for the PM10 data across the different regions. Notice how much more info the boxplot provides. What does the dotplot provide that the boxplot doesn't?

5. Let us plot the scatterplots for PSI vs PM10 and SO_2. What do we notice about the scatterplots?