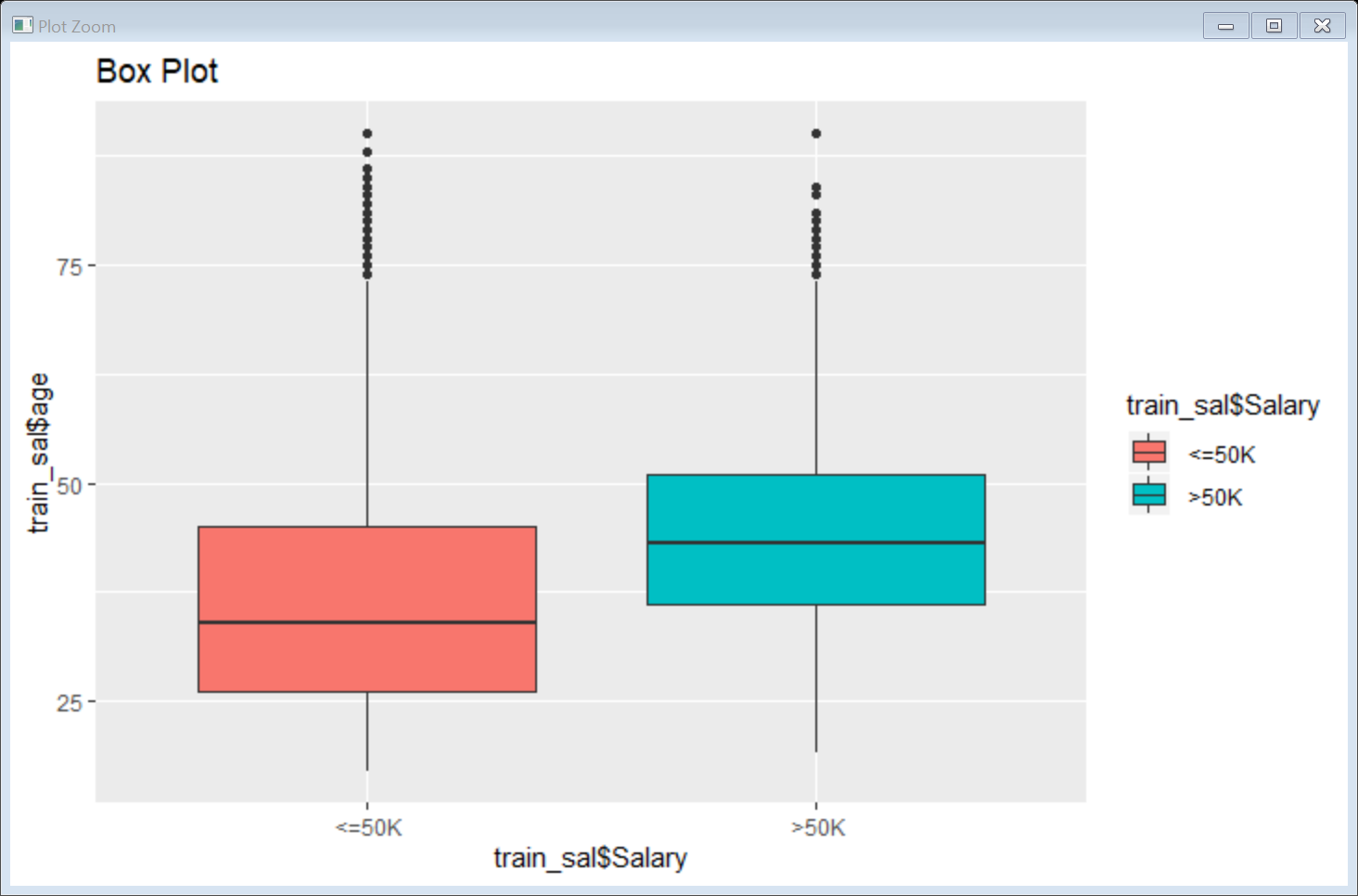
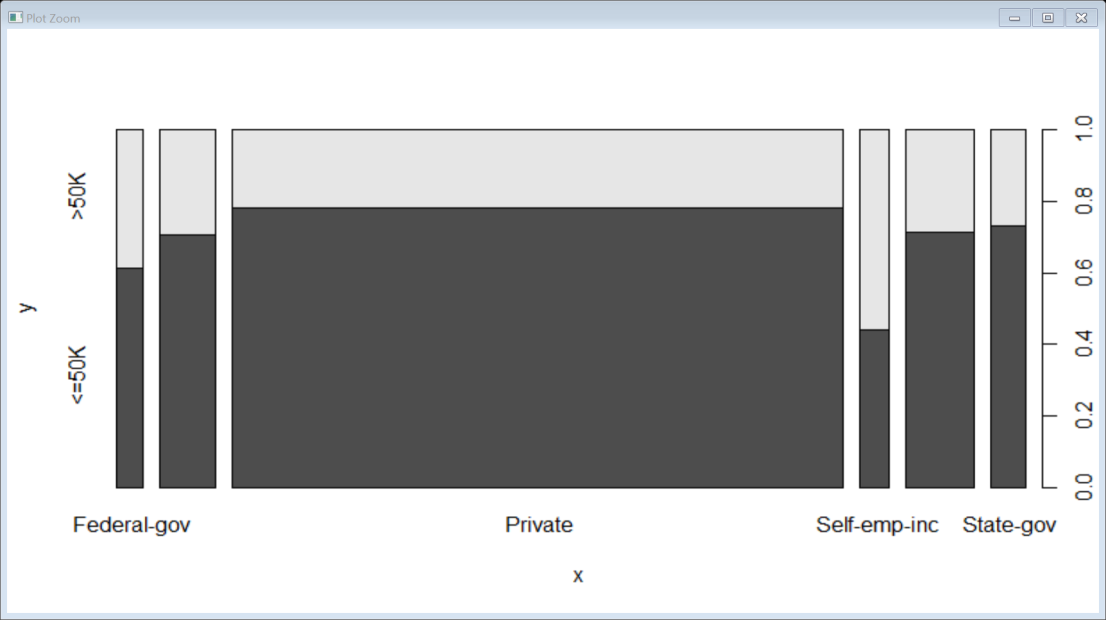
# Naïve Bayes

**Business Objective:** Prepare a classification model using Naive Bayes for salary data

Step 1: Install all the packages and then read the train data first and use factor for education no for converting it to categorical form.

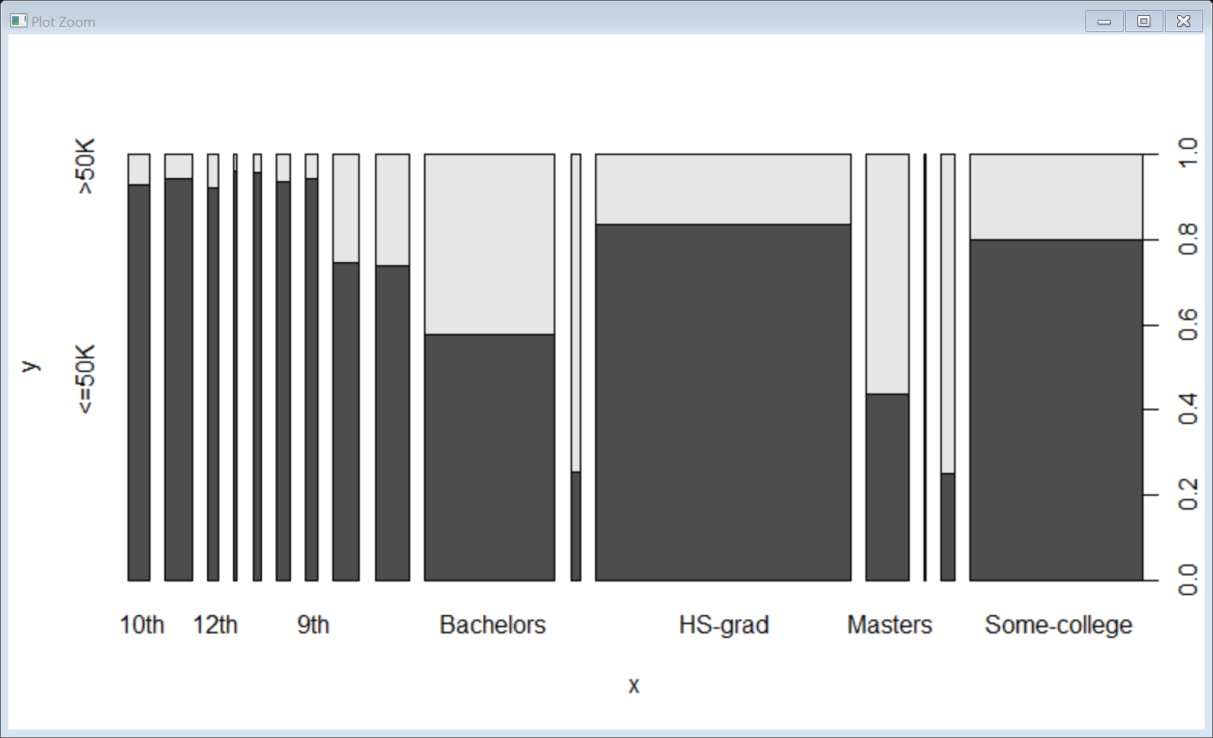
Step 2:Boxplot for determining the outliers with <=50 and >50 salary varaiable.

****

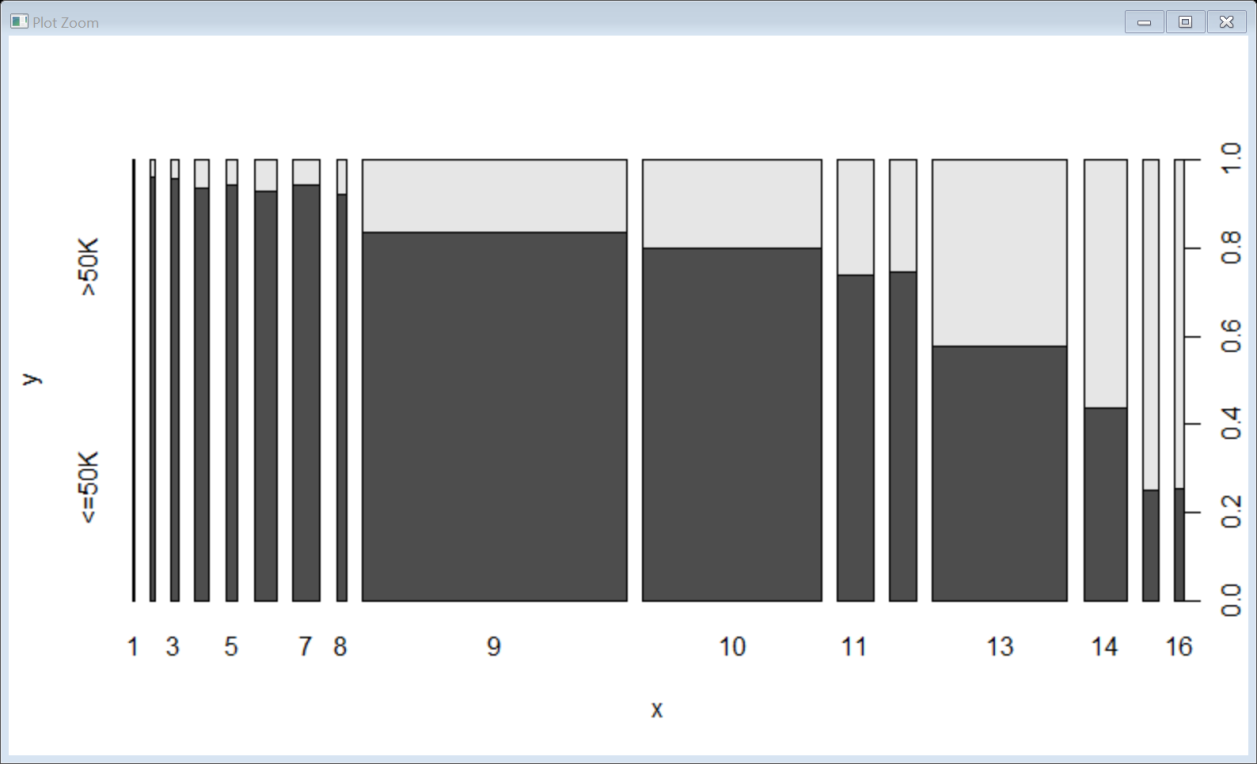


Step 3: The plot of salary with workclass is shown above.

Step 4: The plot of salary with education is shown below.

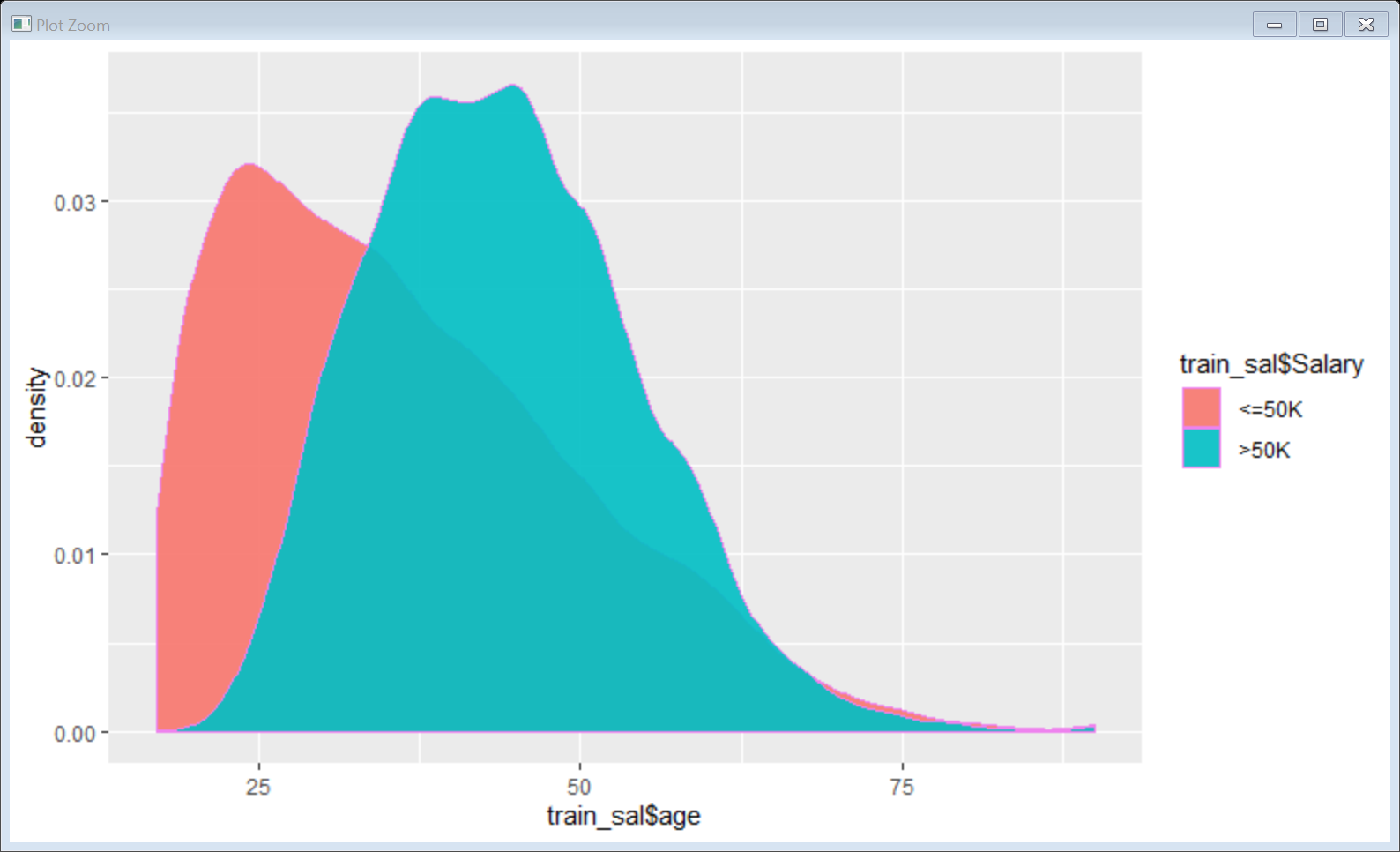
:

Step 5: The plot of salary with education no is shown below.

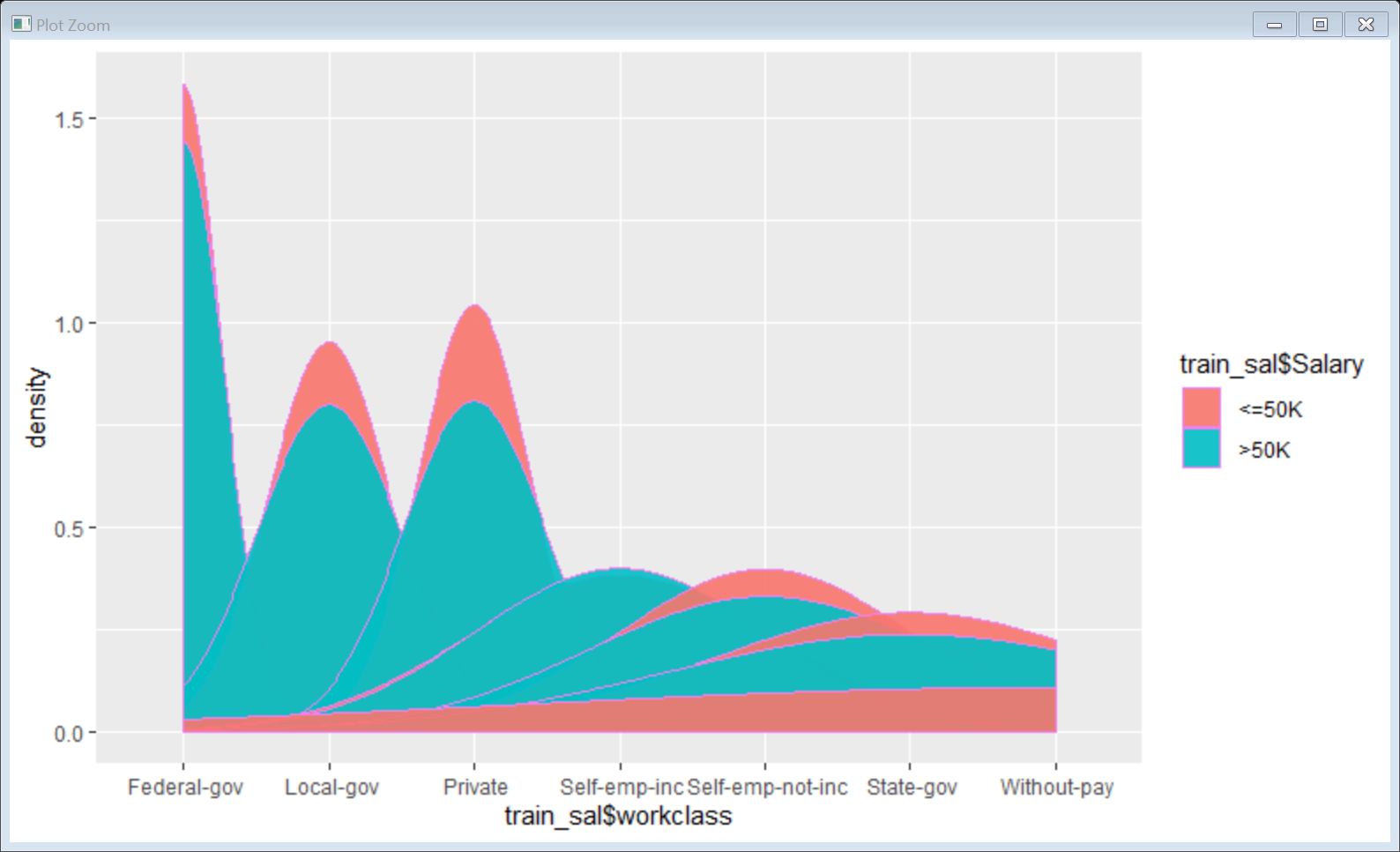


Step 6: Likewise do the plot for every other input variables.

Step 7:Now do the Age density plot as below



Step 8: Workclass Density Plot



Step 9: Likewise do it for rest of varaibles.

Step 10: Naive Bayes Model

Step 11:

Conditional probabilities:

age

Y [,1] [,2]

<=50K 36.60826 13.46489

>50K 43.95911 10.26963

Step 12: Predict the model with test data

Accuracy = 0.8187251

Step 13: Confusion Matrix and Statistics

Reference

Prediction <=50K >50K

<=50K 10549 1919

>50K 811 1781

Conclusion: The model is 81.87% accurate.

**Business Objective:** Build a naive Bayes model on the data set for classifying the ham and spam

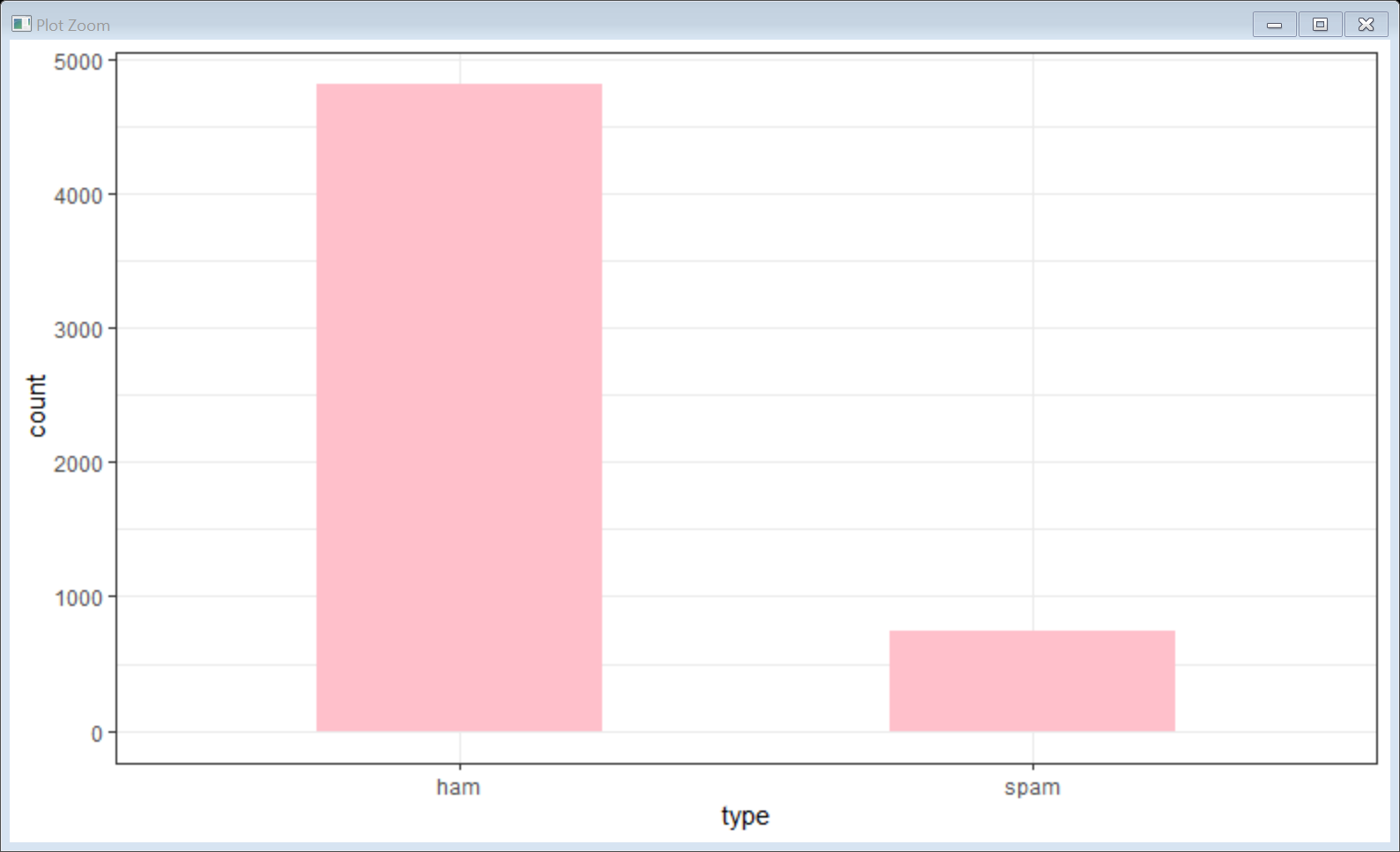
Step 1:Install packages and read the data

Step 2: table of type

ham spam

4812 747

Step 3: Create a ggplot to show the ham and spam sms



Step 4: clean up the corpus using tm\_map()

Step 5:Train data

ham spam

0.8647158 0.1352842

Step 6: Test Data

ham spam

0.8683453 0.1316547

Step 7:

sms\_test\_pred

ham spam

1235 155

Step 8:

| actual

predicted | ham | spam | Row Total |

-------------|-----------|-----------|-----------|

ham | 1203 | 32 | 1235 |

| 0.997 | 0.175 | |

-------------|-----------|-----------|-----------|

spam | 4 | 151 | 155 |

| 0.003 | 0.825 | |

-------------|-----------|-----------|-----------|

Column Total | 1207 | 183 | 1390 |

| 0.868 | 0.132 | |

-------------|-----------|-----------|-----------|