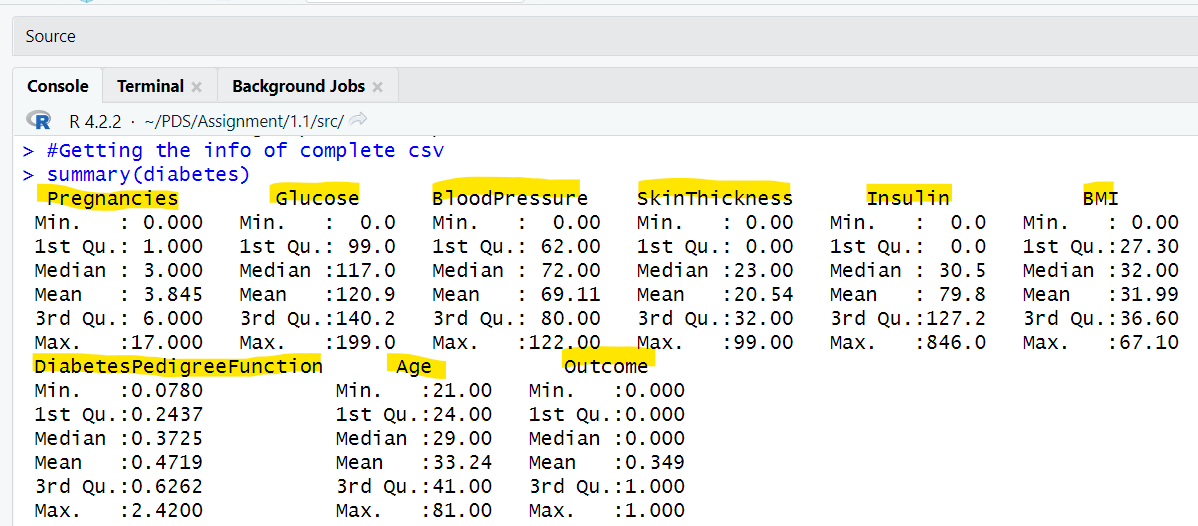
**ASSIGNMENT-2 (5530-0001)**

**KONETI SRI GANESH ID: 16341531**

**DATA SET – DIABETES**

The diabetes data set consists of 8 features with 768 rows of data which includes-

****

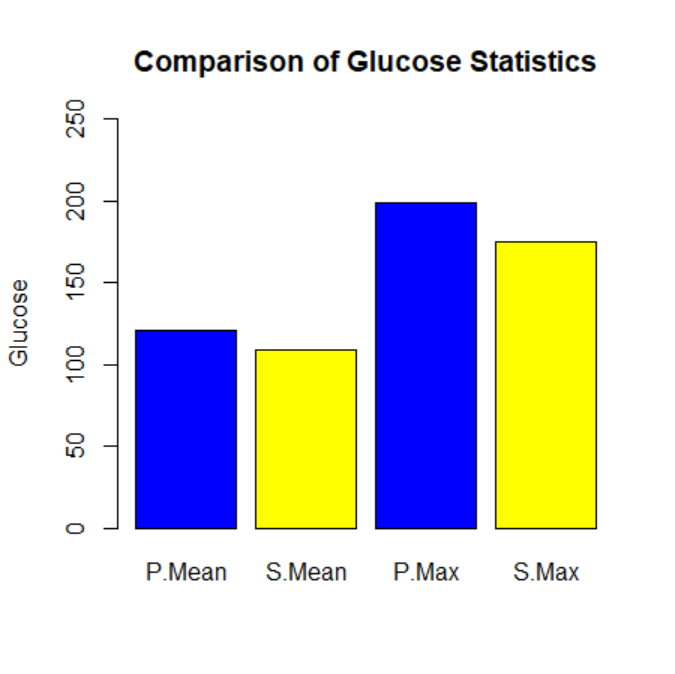
1. find the mean Glucose and highest Glucose values of this sample and compare these statistics with the population statistics of the same variable. You should use charts for this comparison.

Table

Description automatically generated with medium confidence

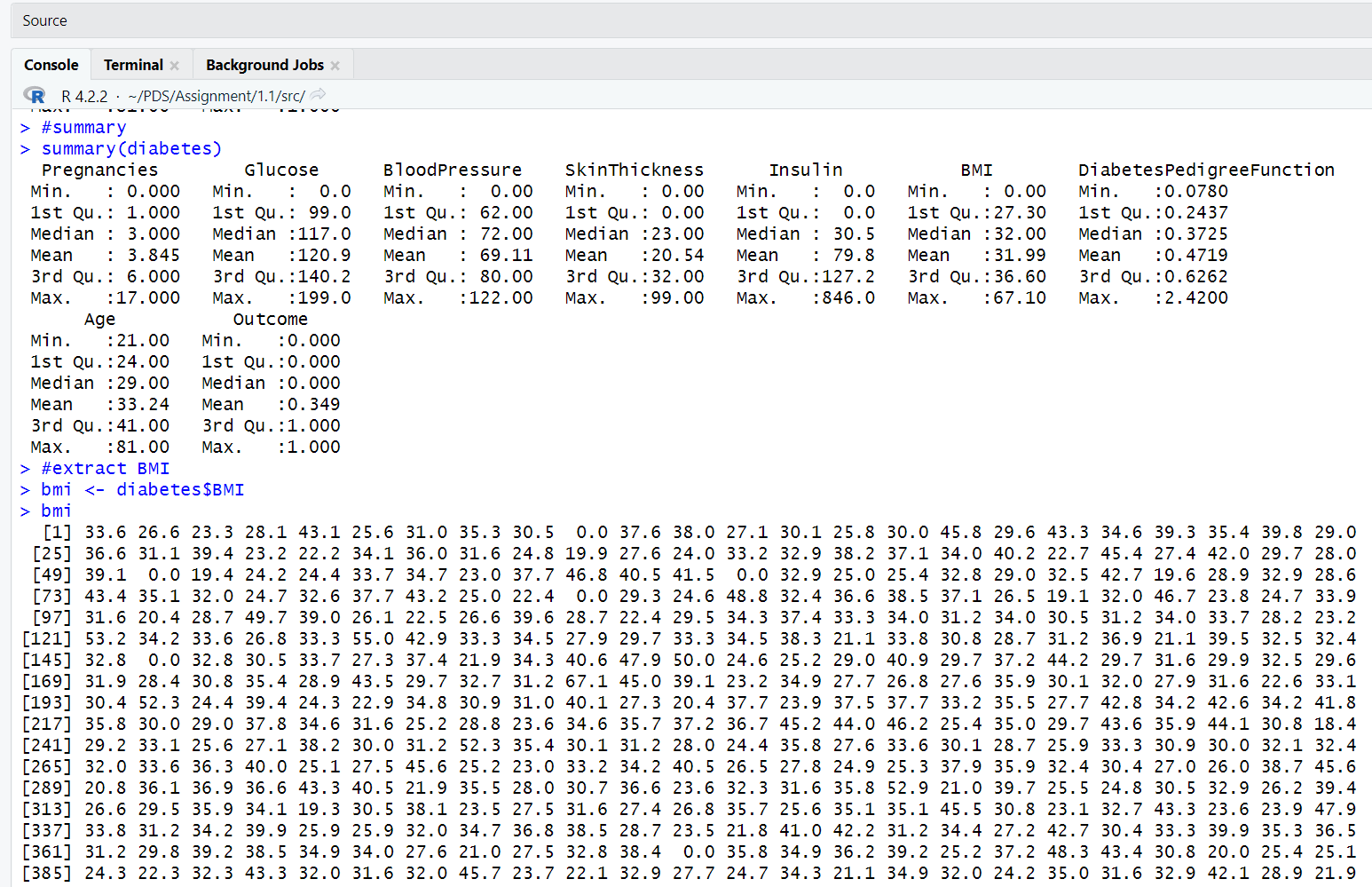
Graphical user interface, text

Description automatically generated



**Outcome: When we compare Glucose Statistics , we conclude that population mean and sample mean are almost similar.**

1. Find the 98th percentile of BMI of your sample and the population and compare the results using charts.



Text

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

C)Using bootstrap (replace= True), create 500 samples (of 150 observation each) from the population and find the average mean, standard deviation and percentile for Blood Pressure and compare this with these statistics from the population for the same variable.

Text

Description automatically generated

Chart

Description automatically generated with medium confidence

**Outcome: We can observe that the bootstrap means, standard deviations, and percentiles are all fairly near to the population values by comparing the histograms and line plots. This implies that the bootstrap samples are representative of the population and that we can reasonably estimate the statistics for Blood Pressure using them.**