ST495/590 Assignment 1 - Solutions

2a *Create a table with the overall mean, standard deviation, and percent missing.*

The following table provides the summary statistics for the ozone data.

Mean[1]	SD	Percent Missing
51.27	17.26	4.32%

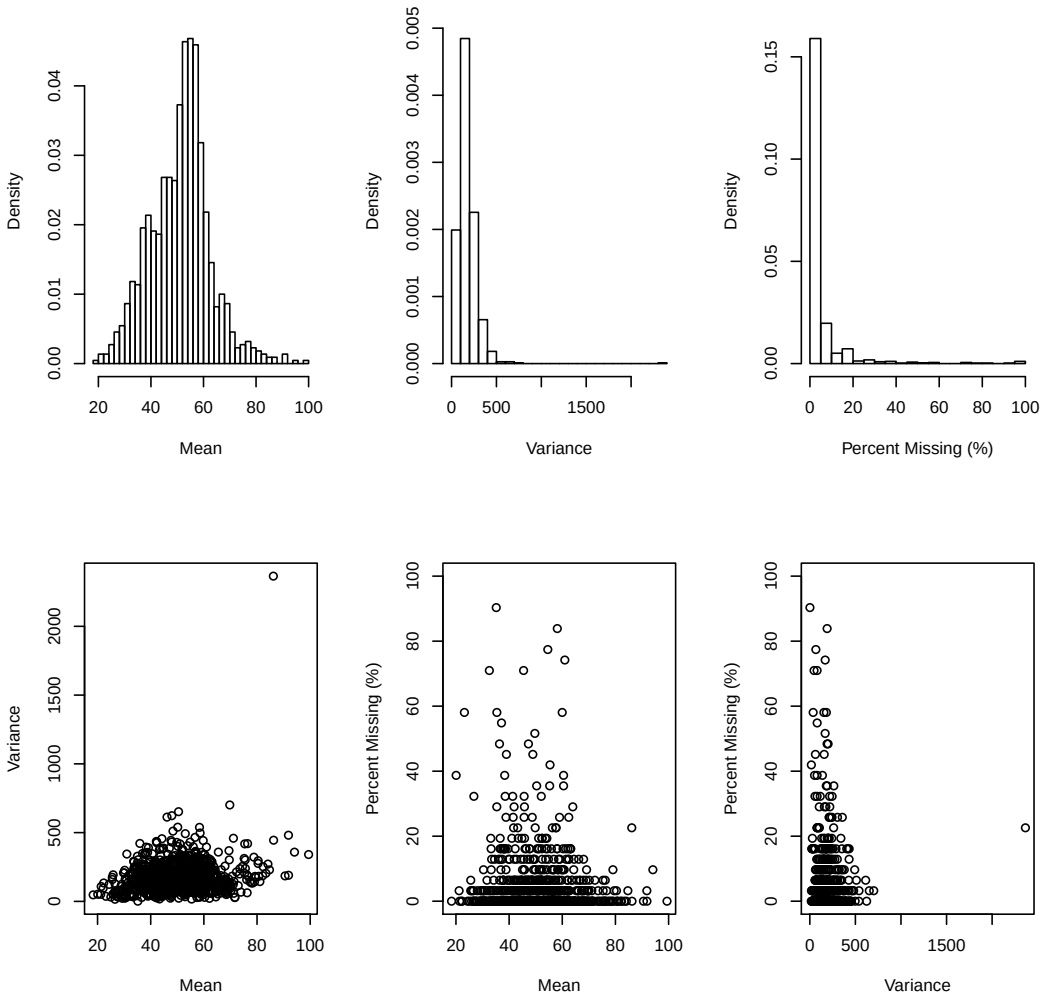
2b *Create scatter plots of each pair of these variables*

For this problem I computed the mean, standard deviation, and percent missing for ozone at each of the $n = 1,106$ locations, and constructed histograms of the n values for each of the three summary measures (top row) and scatter plots for each pair of summary measures (bottom row). The results show that the means and standard deviations are right-skewed, the proportion of missing values is near zero for most sites, and that site with large mean also tend to have large variance.

2c *Conduct a linear regression with response equal to the sites mean and the sites variance and percent missing as covariates*

The results of the linear regression are in the table below. The regression suggests a positive relationship between X and Y and a negative relationship between X and Z.

	Estimate	Standard error	p-value
Intercept	48.59	0.633	< 0.0001
Variance	0.017	0.0030	< 0.0001
Percent Missing (%)	-0.105	0.036	0.036


Density
Mean
Density
Variance
Density
Percent Missing (%)
Variance
Mean
Percent Missing (%)
Mean
Percent Missing (%)
Variance