Exercise 1

a) For resting blood pressure, the mean is 131.3, the median is 130, and the standard deviation is 17.9. The skewness of .723 indicates a tail to the right in the distribution.

Since there is a noticeable skew, the median and IQR should be used as the measure of location and spread. Roughly half of the resting blood pressures are expected to be greater than 130 and roughly half less than 130. The difference between the 75th percentile and 25th percentile is expected to be about 20.

Variable: restingbp

Moments			
N	270	**Sum Weights**	270
Mean	131.344444	**Sum Observations**	35463
Std Deviation	17.8616083	**Variance**	319.037051
Skewness	0.72261801	**Kurtosis**	0.92309674
Uncorrected SS	4743689	**Corrected SS**	85820.9667
Coeff Variation	13.5990588	**Std Error Mean**	1.08702286

Basic Statistical Measures			
Location		**Variability**	
Mean	131.3444	**Std Deviation**	17.86161
Median	130.0000	**Variance**	319.03705
Mode	120.0000	**Range**	106.00000
		Interquartile Range	20.00000

b) Resting blood pressure for those without heart disease has a mean of 128.9, median of 130, and standard deviation of 16.5. The skewness of .414 indicates a small tail to the right in the distribution. The skew is large enough that it would likely be noticed, so again median and IQR should be used. For individuals without heart disease, we expect roughly half to have a resting blood pressure above 130 and roughly half below 130. The middle 50% of blood pressures is expected to span a range of about 20.

Resting blood pressure for those with heart disease has a mean of 134.4, median of 130, and standard deviation of 19.1. The skewness of .889 indicates a tail to the right in the distribution. The skew indicates that median and IQR should again be used. Half of those with heart disease are expected to have resting blood pressures above 130 and half below. The spread for those with heart disease is a bit wider with an IQR of 25.

Variable:
restingbp

heartdisease=absence

Moments			
N	150	**Sum Weights**	150
Mean	128.866667	**Sum Observations**	19330
Std Deviation	16.4576604	**Variance**	270.854586
Skewness	0.41352706	**Kurtosis**	0.26136044
Uncorrected SS	2531350	**Corrected SS**	40357.3333
Coeff Variation	12.7710764	**Std Error Mean**	1.34376235

Basic Statistical Measures			
Location		**Variability**	
Mean	128.8667	**Std Deviation**	16.45766
Median	130.0000	**Variance**	270.85459
Mode	120.0000	**Range**	86.00000
		Interquartile Range	20.00000

Variable:
restingbp

heartdisease=presence

Moments			
N	120	**Sum Weights**	120
Mean	134.441667	**Sum Observations**	16133
Std Deviation	19.0954242	**Variance**	364.635224
Skewness	0.8886235	**Kurtosis**	0.96280272
Uncorrected SS	2212339	**Corrected SS**	43391.5917
Coeff Variation	14.2035015	**Std Error Mean**	1.74316576

Basic Statistical Measures			
Location		**Variability**	
Mean	134.4417	**Std Deviation**	19.09542
Median	130.0000	**Variance**	364.63522
Mode	120.0000	**Range**	100.00000
		Interquartile Range	25.00000

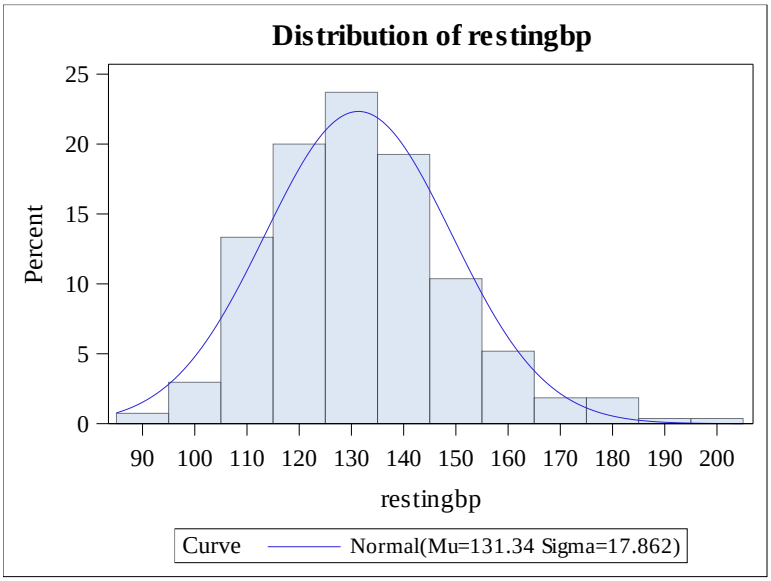
Note: The mode displayed is the smallest of 2 modes with a count of 13.

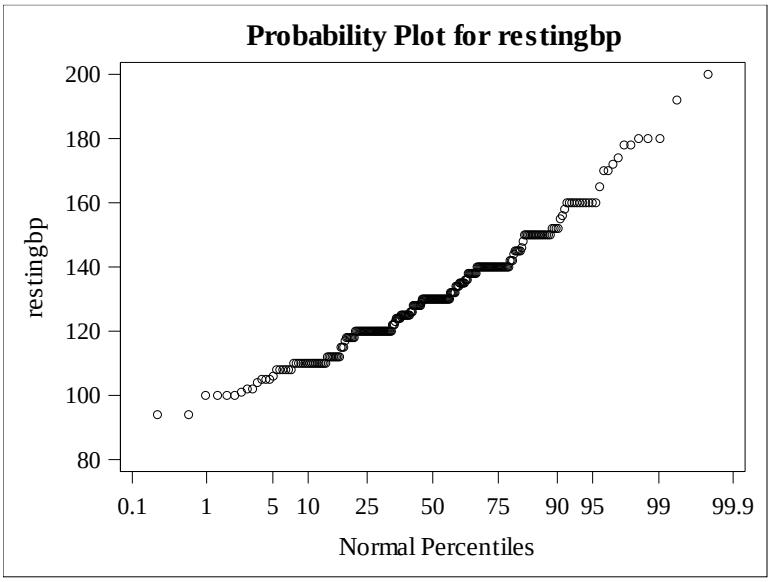
Exercise 2

a) Normality tests, a histogram and a probability plot for the resting blood pressure levels for the entire sample follow. The p-values for the normality tests are much less than .05, so we reject the null hypothesis that the data came from a normal distribution. The histogram and probability plot do not look as bad, but there is some deviation from a bell shape and a straight line, so we see some visual indications of deviation from normality, too. We conclude there is enough evidence of deviation from normality and assumptions of normality for this data will not be fine.

Variable: restingbp

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.964922	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.10037	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.364616	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	2.286146	**Pr > A-Sq**	<0.0050





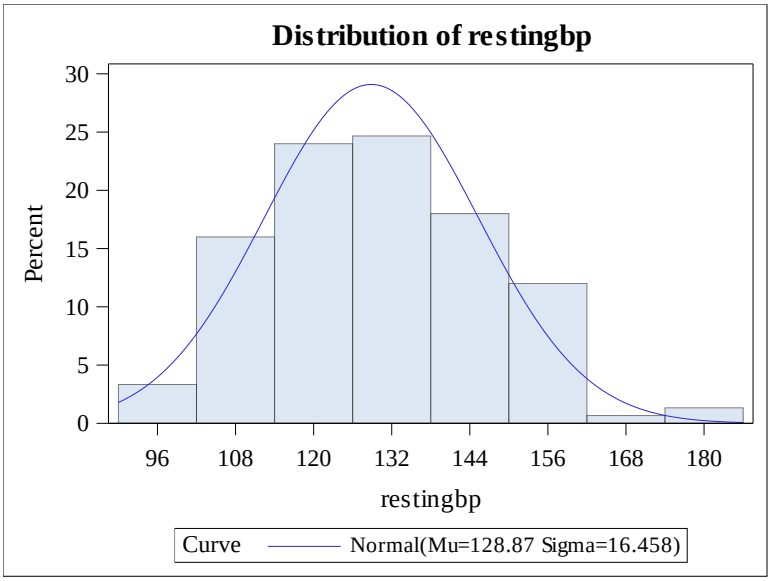
b) Considering the two subpopulations separately, we see that for those with or without heart disease, all four tests would reject normality at a .05 level. The plots for those without heart disease show a little deviation from a bell shape and a straight line. A case could be made for the blood pressures of those without heart disease to not be too far from normal based on the plots, but the tests clearly reject normality. The plots for those with heart disease show more pronounced deviations from normality. Both the plots and tests reject normality for blood pressures of individuals with heart disease. Based on these results, we reject normality for both subpopulations.

Variable:
restingbp

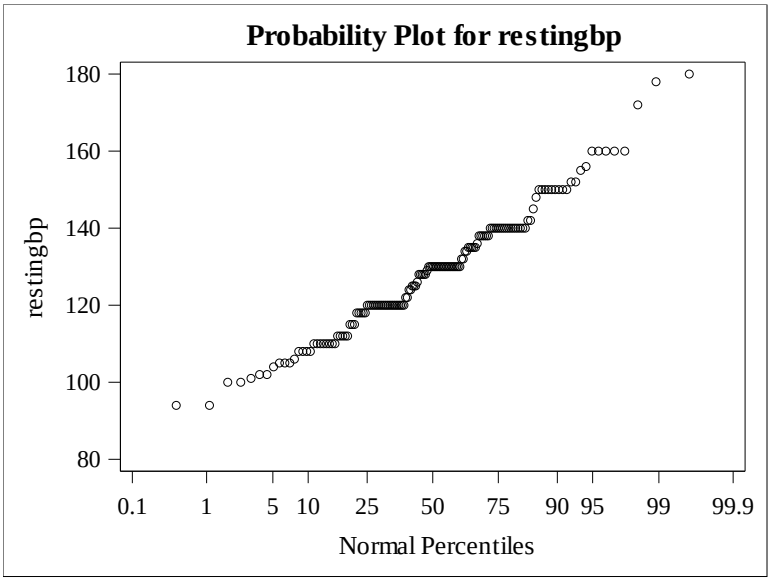
heartdisease=absence

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.978975	**Pr < W**	0.0213
Kolmogorov-Smirnov	**D**	0.091639	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.152709	**Pr > W-Sq**	0.0224
Anderson-Darling	**A-Sq**	0.899307	**Pr > A-Sq**	0.0221

heartdisease=absence



heartdisease=absence

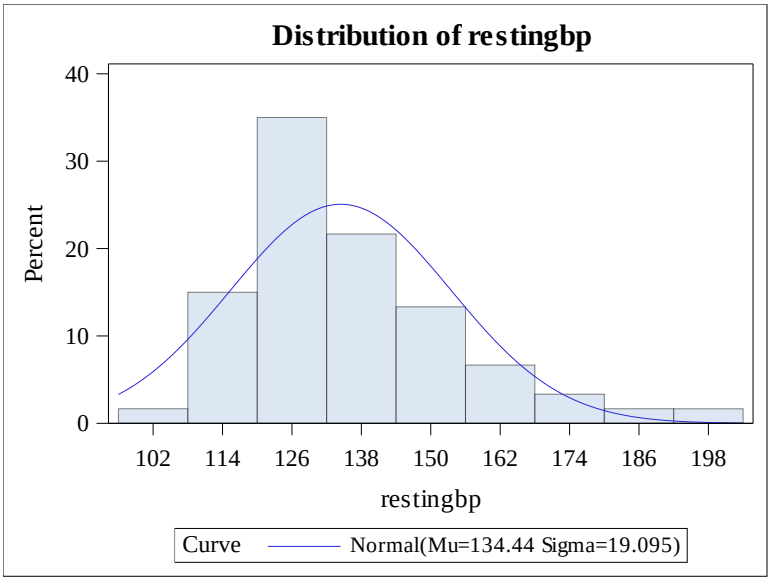


Variable:
restingbp

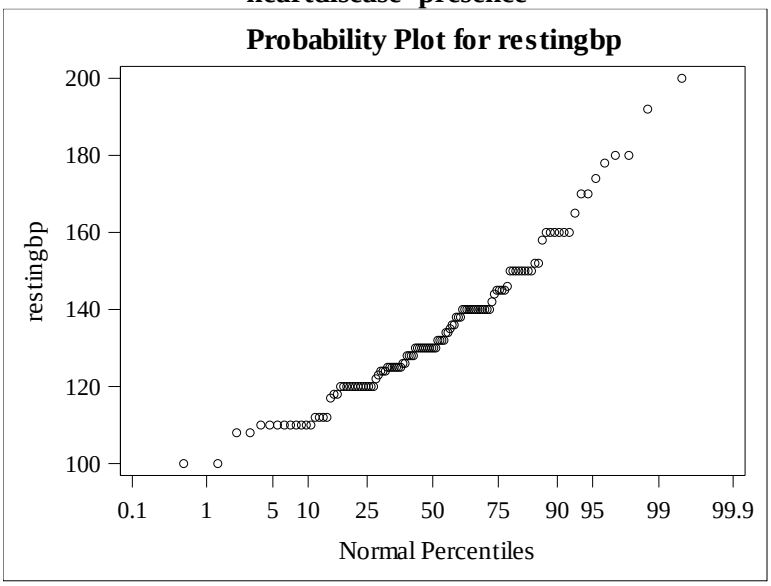
heartdisease=presence

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.948046	**Pr < W**	0.0002
Kolmogorov-Smirnov	**D**	0.110494	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.253846	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	1.610518	**Pr > A-Sq**	<0.0050

heartdisease=presence



heartdisease=presence



Exercise 3

a) Given that the population is right skewed, we should look at one-sided Sign tests when comparing to hypothesized null medians of 120 and 129. The p-values provided are for two-sided tests and the one-sided p-values will need to be computed from those results.

Variable: restingbp

Tests for Location: Mu0=120				
Test	**Statistic**		**p Value**	
Student's t	**t**	8.284735	**Pr > |t|**	<.0001
Sign	**M**	33.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	2244	**Pr >= |S|**	<.0001

For the alternative that the true median is greater than 120, the statistic is positive. This means the difference is in the same direction as the greater alternative, and the p-value is half the two-sided p-value shown. This means the p-value is much less than 0.0001, and we conclude there is strong evidence that the median resting blood pressure for those with heart disease is greater than 120.

Variable: restingbp

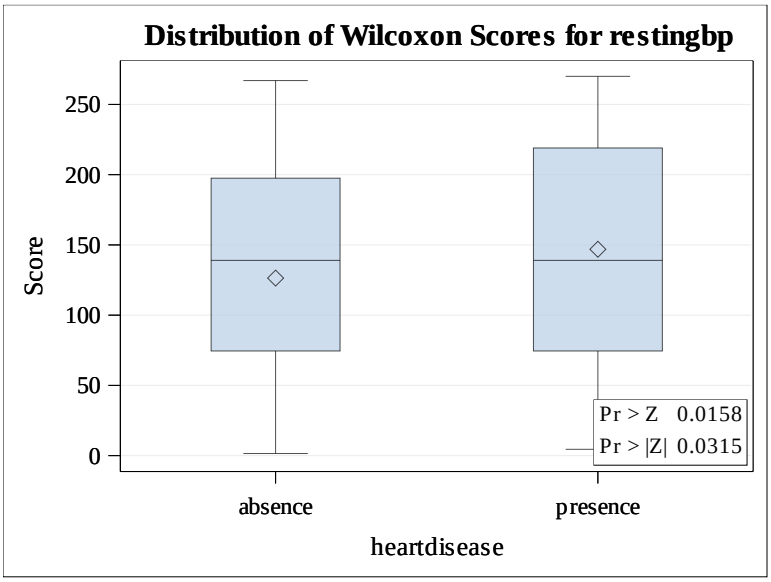
Tests for Location: Mu0=129				
Test	**Statistic**		**p Value**	
Student's t	**t**	3.121715	**Pr > |t|**	0.0023
Sign	**M**	9	**Pr >= |M|**	0.1203
Signed Rank	**S**	934	**Pr >= |S|**	0.0137

When comparing to a hypothesized median of 129, the two-sided p-value is 0.1203. The test statistic is again in the same direction as the alternative, so the one-sided p-value is about 0.06. At a 0.05 level, the evidence is not strong enough to reject the null assumption that the true median is 129. The data do not support a statistically significant difference between the true median and 129. Therefore, the median resting blood pressure is not statistically significantly greater than 129.

b) Given the normality conclusions, the rank sum test will need to be used to perform a hypothesis test of whether those with heart disease have significantly higher resting blood pressure levels than those without. The alternative for the one-sided test is that those with heart disease have significantly higher resting blood pressure levels than those without. The p-value for that test is 0.0158, so we concluded that those with heart disease tend to have higher resting blood pressure levels than those without.

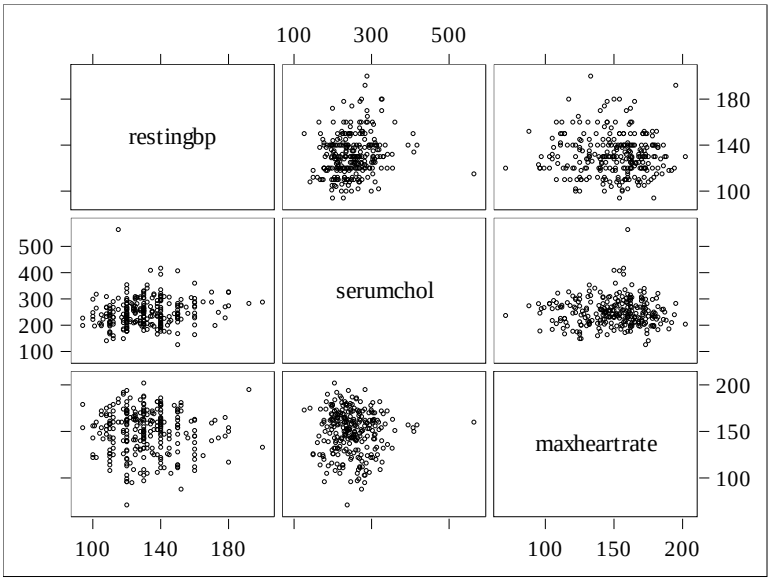
Wilcoxon Scores (Rank Sums) for Variable restingbp Classified by Variable heartdisease					
heartdisease	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
absence	150	18957.50	20325.0	635.759554	126.383333
presence	120	17627.50	16260.0	635.759554	146.895833
Average scores were used for ties.					

Wilcoxon Two-Sample Test					
				t Approximation	
Statistic	**Z**	**Pr > Z**	**Pr > |Z|**	**Pr > Z**	**Pr > |Z|**
17627.50	2.1502	0.0158	0.0315	0.0162	0.0324
Z includes a continuity correction of 0.5.					



Exercise 4

a) From the scatter plot, there are no extreme values and no apparent nonlinear relations, so we use Pearson correlation.



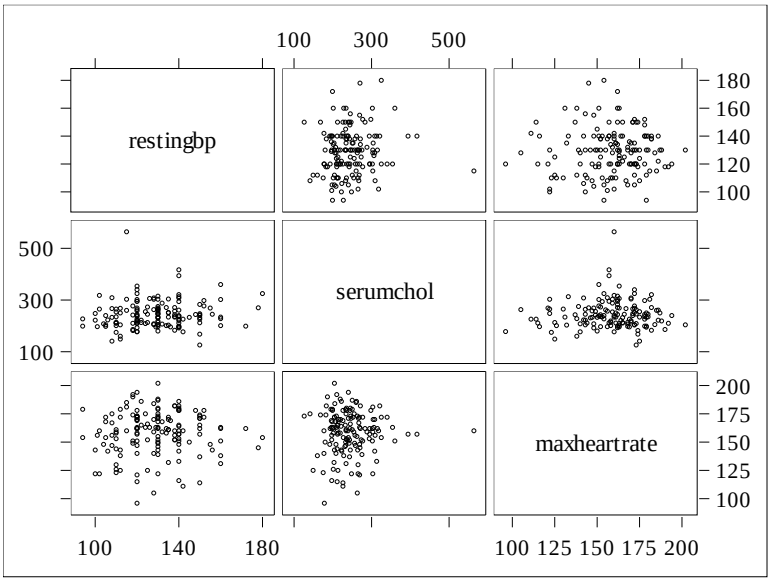
Looking at the correlation matrix for the entire sample, we see there is a small but statistically significant positive correlation between resting blood pressure and serum cholesterol. The other correlations are insignificant.

Based on these results we conclude there is a weak but statistically significant tendency in the broader population for serum cholesterols to be higher when resting blood pressures are higher.

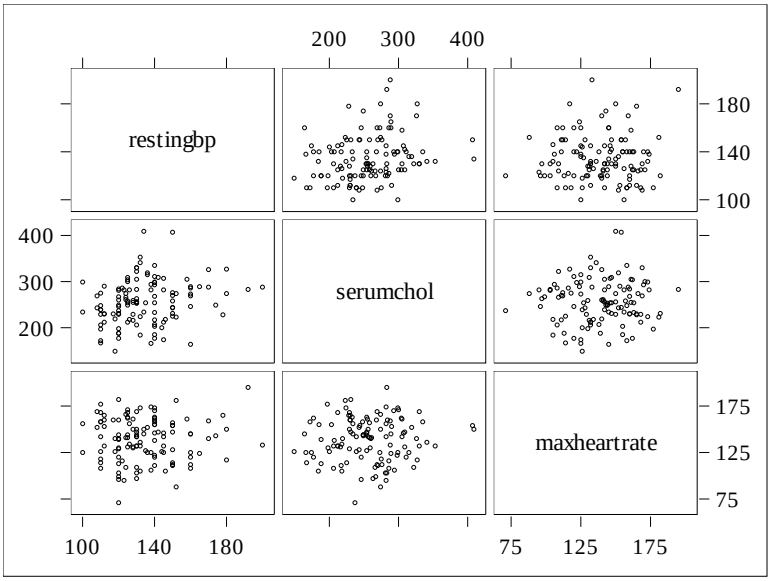
Pearson Correlation Coefficients, N = 270 Prob > |r| under H0: Rho=0			
	restingbp	**serumchol**	**maxheartrate**
restingbp	1.00000	0.17302 0.0044	-0.03914 0.5220
serumchol	0.17302 0.0044	1.00000	-0.01874 0.7592
maxheartrate	-0.03914 0.5220	-0.01874 0.7592	1.00000

b) From the scatter plots for each subpopulation, there are again no extreme values or apparent nonlinear relations, so we use Pearson correlation.

heartdisease=absence



heartdisease=presence



Analyzing by patient group, we see that none of the correlations are statistically significantly different from 0 for those without heart disease, so there is no significant relationship between the three measures for those without heart disease.

For those with heart disease, the correlation between resting blood pressure and serum cholesterol is statistically significant and slightly greater than it was in the full sample. For patients with heart disease, there is a weak but significant tendency for those with higher serum cholesterol to have higher resting blood pressure.

The difference between the two groups is the significance of this one correlation. Those without heart disease showed no relationship between serum cholesterol and resting blood pressure, while those with heart disease showed a weak positive relationship. Likewise, the broader population showed a weak positive relationship between these two measures.

heartdisease=absence

Pearson Correlation Coefficients, N = 150 Prob > |r| under H0: Rho=0			
	restingbp	**serumchol**	**maxheartrate**
restingbp	1.00000	0.10109 0.2184	0.03533 0.6678
serumchol	0.10109 0.2184	1.00000	0.02509 0.7605
maxheartrate	0.03533 0.6678	0.02509 0.7605	1.00000

heartdisease=presence

Pearson Correlation Coefficients, N = 120 Prob > |r| under H0: Rho=0			
	restingbp	**serumchol**	**maxheartrate**
restingbp	1.00000	0.22956 0.0117	0.02305 0.8027
serumchol	0.22956 0.0117	1.00000	0.04528 0.6234
maxheartrate	0.02305 0.8027	0.04528 0.6234	1.00000