List Of Attributes in The Data

Data Set Name	WORK.HEART	**Observations**	5209
Member Type	DATA	**Variables**	17
Engine	V9	**Indexes**	0
Created	11/17/2023 09:14:56	**Observation Length**	168
Last Modified	11/17/2023 09:14:56	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Alphabetic List of Variables and Attributes

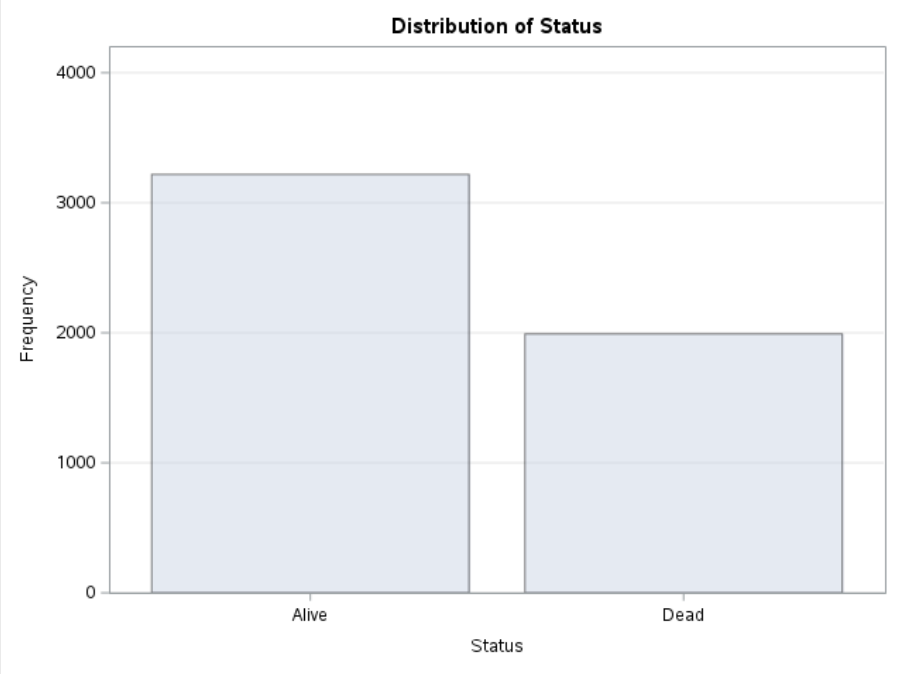
#	Variable	Type	Len	Label
12	AgeAtDeath	Num	8	Age at Death
5	AgeAtStart	Num	8	Age at Start
3	AgeCHDdiag	Num	8	Age CHD Diagnosed
15	BP_Status	Char	7	Blood Pressure Status
14	Chol_Status	Char	10	Cholesterol Status
13	Cholesterol	Num	8	
2	DeathCause	Char	26	Cause of Death
8	Diastolic	Num	8	
6	Height	Num	8	
10	MRW	Num	8	Metropolitan Relative Weight
4	Sex	Char	6	
11	Smoking	Num	8	
17	Smoking_Status	Char	17	Smoking Status
1	Status	Char	5	
9	Systolic	Num	8	
7	Weight	Num	8	
16	Weight_Status	Char	11	Weight Status

Here 17 Variables are Present and observation size 5209

To Analyze Categorical Variables We use proc freq

Frequencies for Categorical Variables

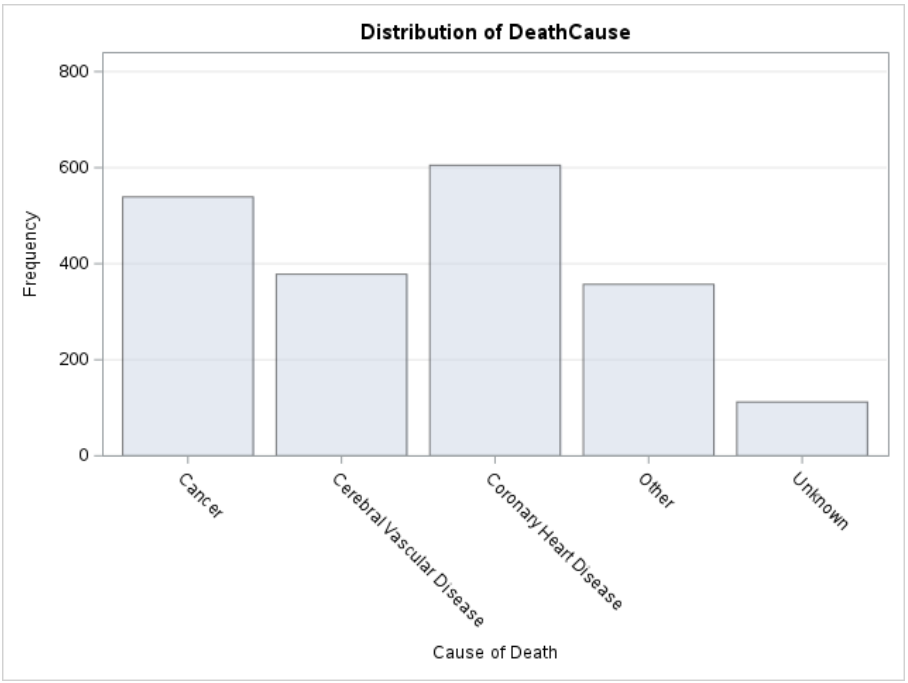
Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Alive	3218	61.78	3218	61.78
Dead	1991	38.22	5209	100.00



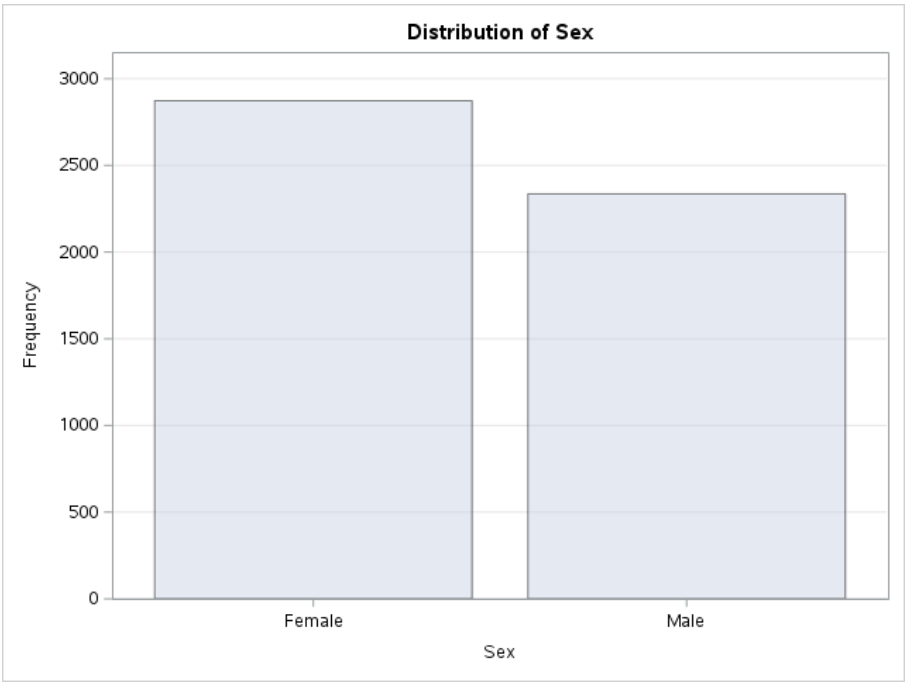
Cause of Death

DeathCause	Frequency	Percent	Cumulative Frequency	Cumulative Percent

Cause of Death				
DeathCause	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Cancer	539	27.07	539	27.07
Cerebral Vascular Disease	378	18.99	917	46.06
Coronary Heart Disease	605	30.39	1522	76.44
Other	357	17.93	1879	94.37
Unknown	112	5.63	1991	100.00
Frequency Missing = 3218				

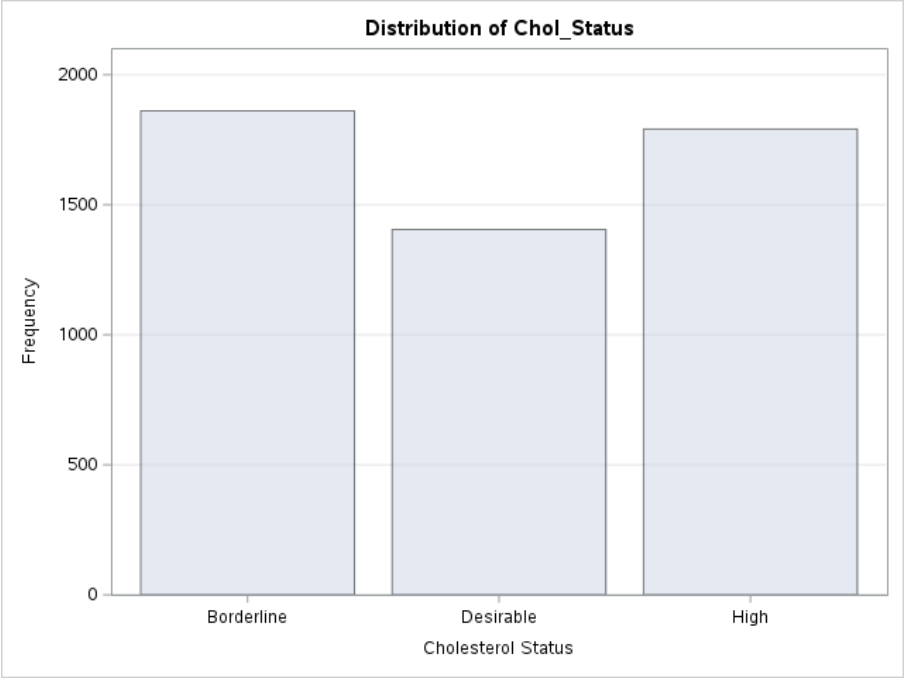


Sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	2873	55.15	2873	55.15
Male	2336	44.85	5209	100.00

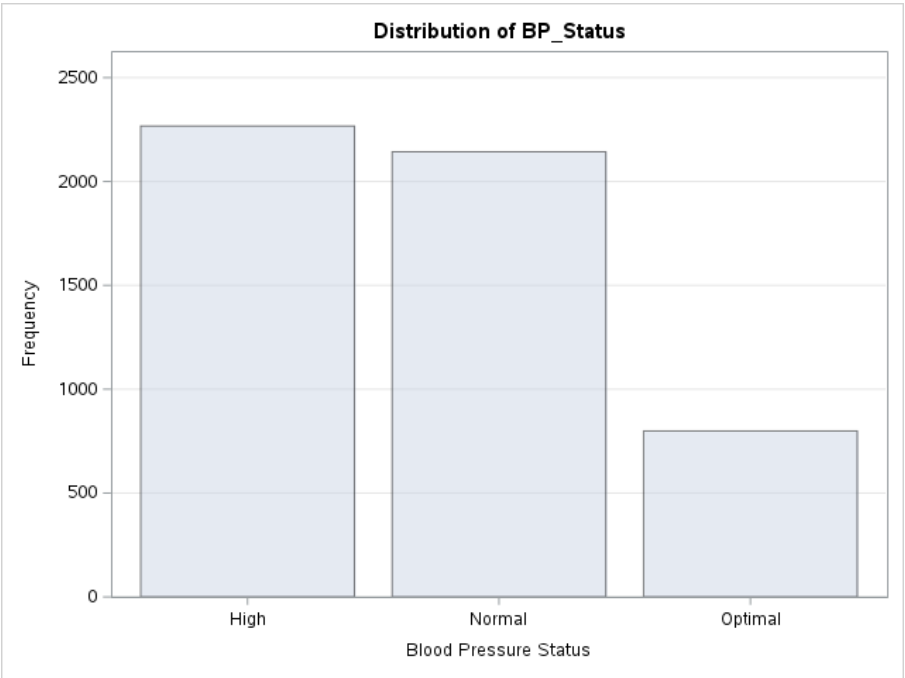


Cholesterol Status				
Chol_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Borderline	1861	36.80	1861	36.80
Frequency Missing = 152				

Cholesterol Status				
Chol_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Desirable	1405	27.78	3266	64.58
High	1791	35.42	5057	100.00
Frequency Missing = 152				

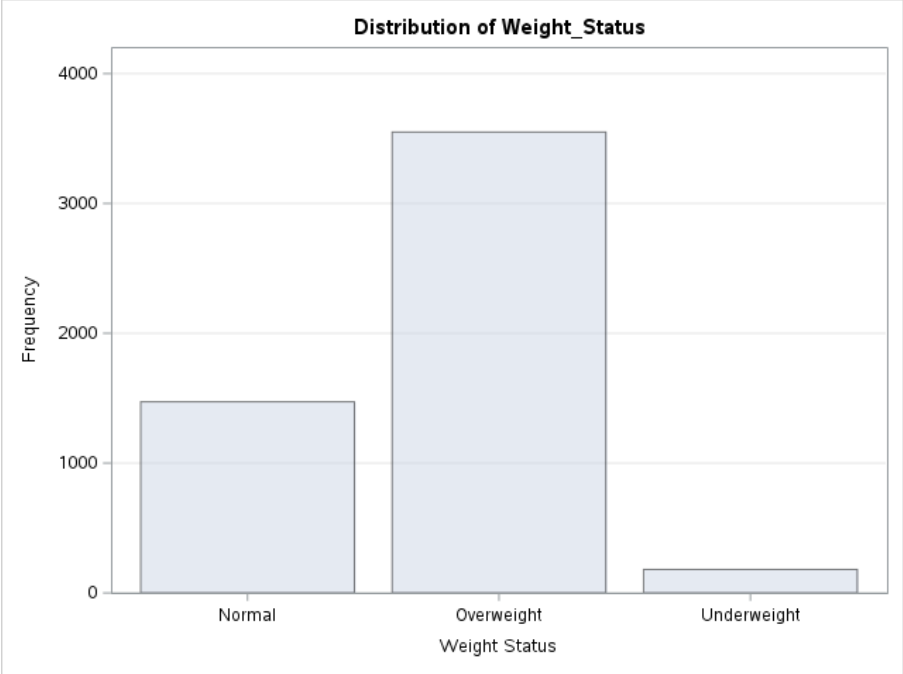


Blood Pressure Status				
BP_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
High	2267	43.52	2267	43.52
Normal	2143	41.14	4410	84.66
Optimal	799	15.34	5209	100.00

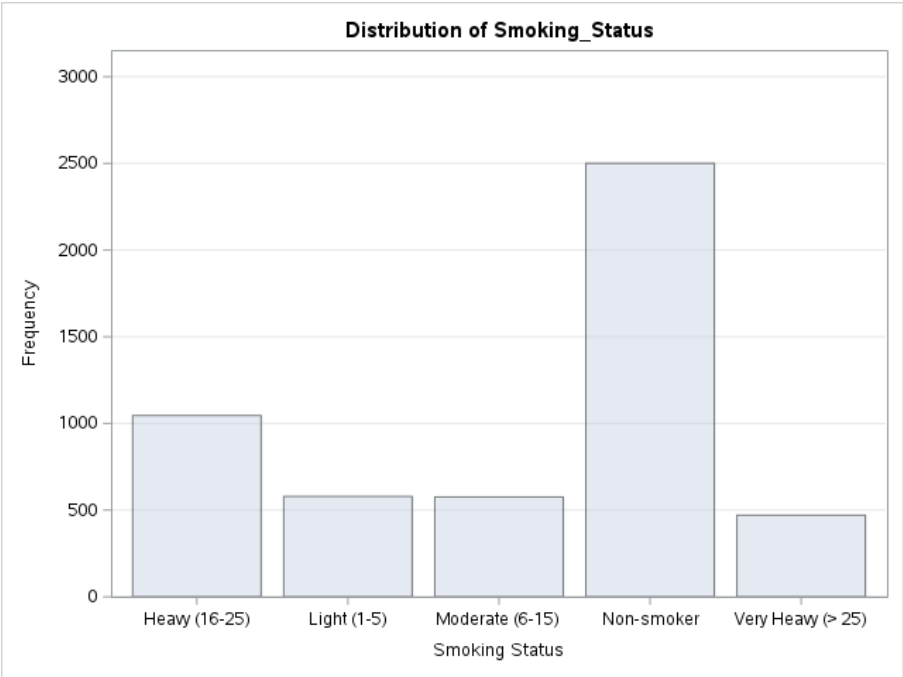


Weight Status				
Weight_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Normal	1472	28.29	1472	28.29
Overweight	3550	68.23	5022	96.52
Frequency Missing = 6				

Weight Status				
Weight_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Underweight	181	3.48	5203	100.00
Frequency Missing = 6				



Smoking Status				
Smoking_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Heavy (16-25)	1046	20.22	1046	20.22
Light (1-5)	579	11.19	1625	31.41
Moderate (6-15)	576	11.13	2201	42.55
Non-smoker	2501	48.35	4702	90.90
Very Heavy (> 25)	471	9.10	5173	100.00
Frequency Missing = 36				



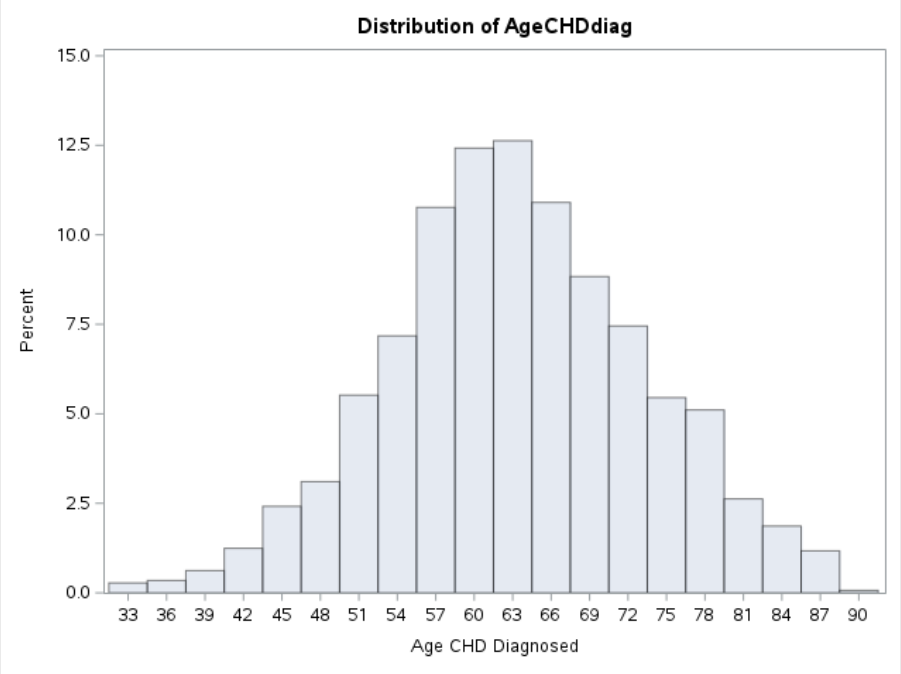
Here we Plotted Simple Bar Graph and its Corresponding Freq..

Descriptive Statistics for Numeric Variables (proc means)

Variable	Label	N	N Miss	Minimum	Mean	Median	Maximum	Std Dev
AgeCHDdiag	Age CHD Diagnosed	1449	3760	32.0000000	63.3029676	63.0000000	90.0000000	10.0182154
AgeAtStart	Age at Start	5209	0	28.0000000	44.0687272	43.0000000	62.0000000	8.5749541
Height		5203	6	51.5000000	64.8131847	64.5000000	76.5000000	3.5827074
Weight		5203	6	67.0000000	153.0866808	150.0000000	300.0000000	28.9154261
Diastolic		5209	0	50.0000000	85.3586101	84.0000000	160.0000000	12.9730913
Systolic		5209	0	82.0000000	136.9095796	132.0000000	300.0000000	23.7395964
MRW	Metropolitan Relative Weight	5203	6	67.0000000	119.9575245	118.0000000	268.0000000	19.9834015
Smoking		5173	36	0	9.3665185	1.0000000	60.0000000	12.0314511
AgeAtDeath	Age at Death	1991	3218	36.0000000	70.5364139	71.0000000	93.0000000	10.5594062
Cholesterol		5057	152	96.0000000	227.4174412	223.0000000	568.0000000	44.9355238

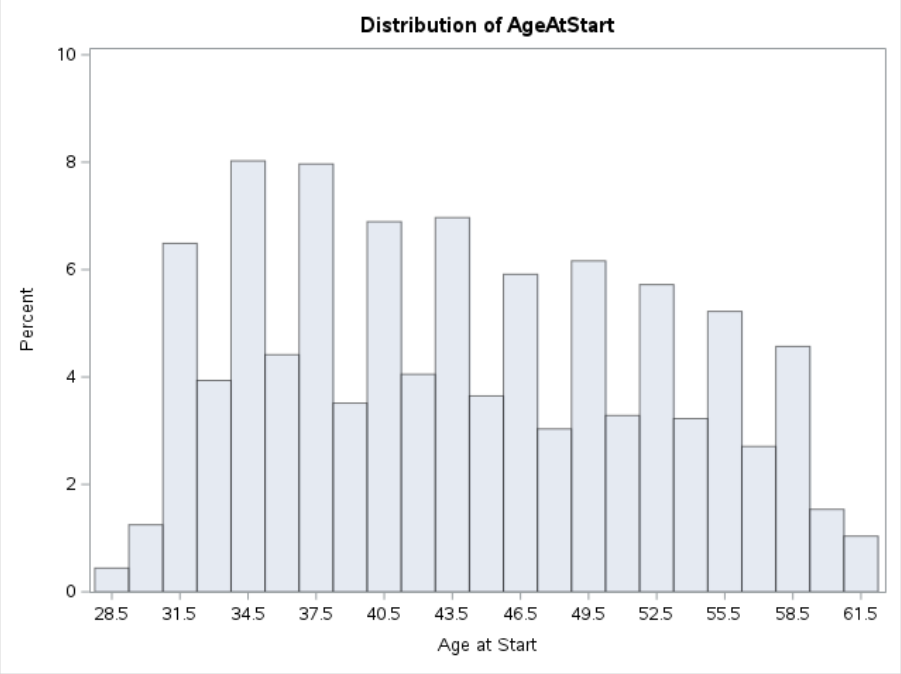
We Analyze The Descriptive Statistics of The Data

To Visualize The Data Using Histogram (proc univariate)



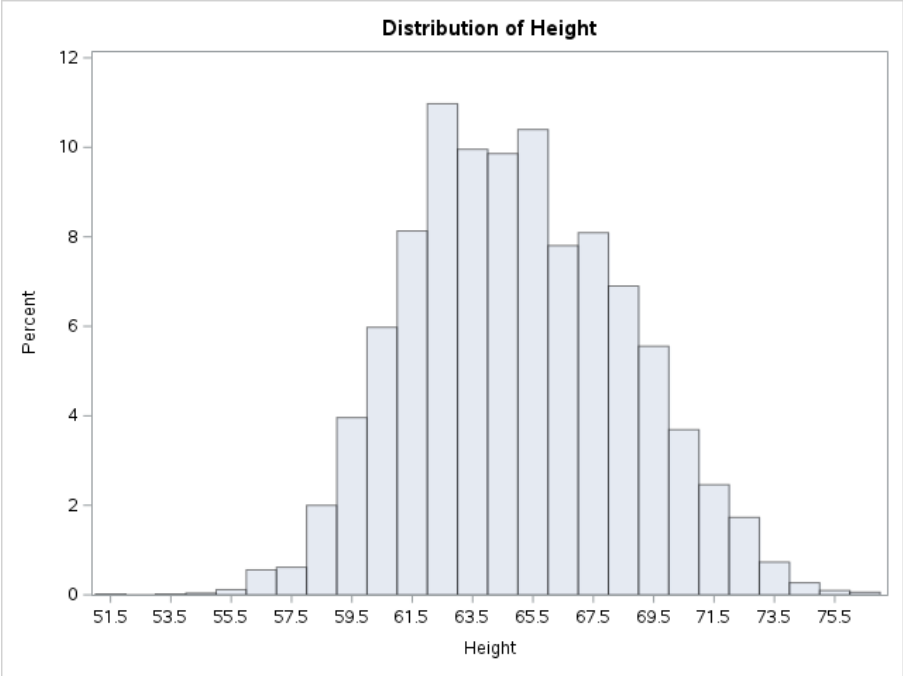
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



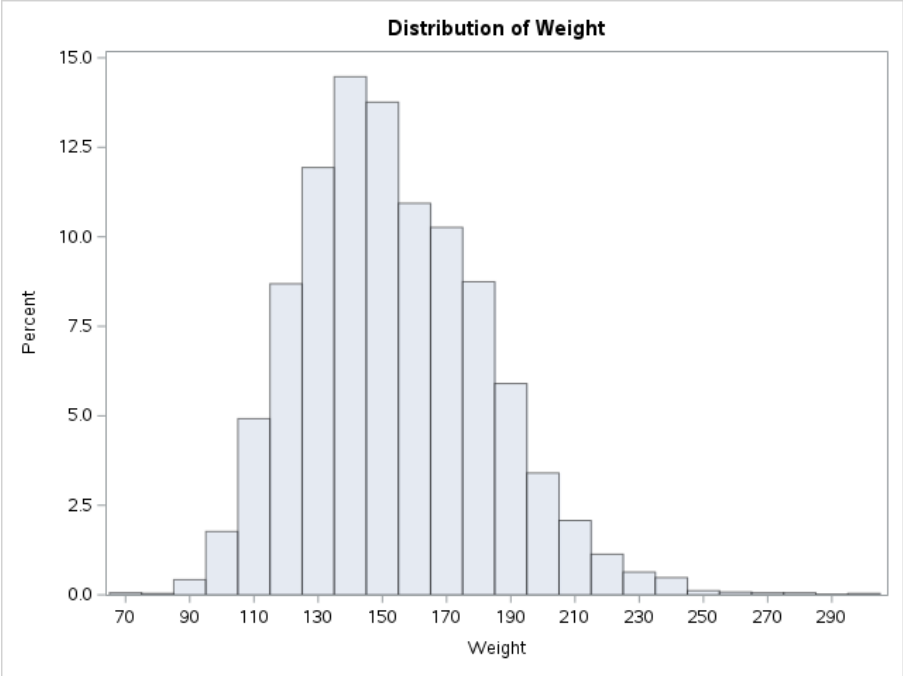
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



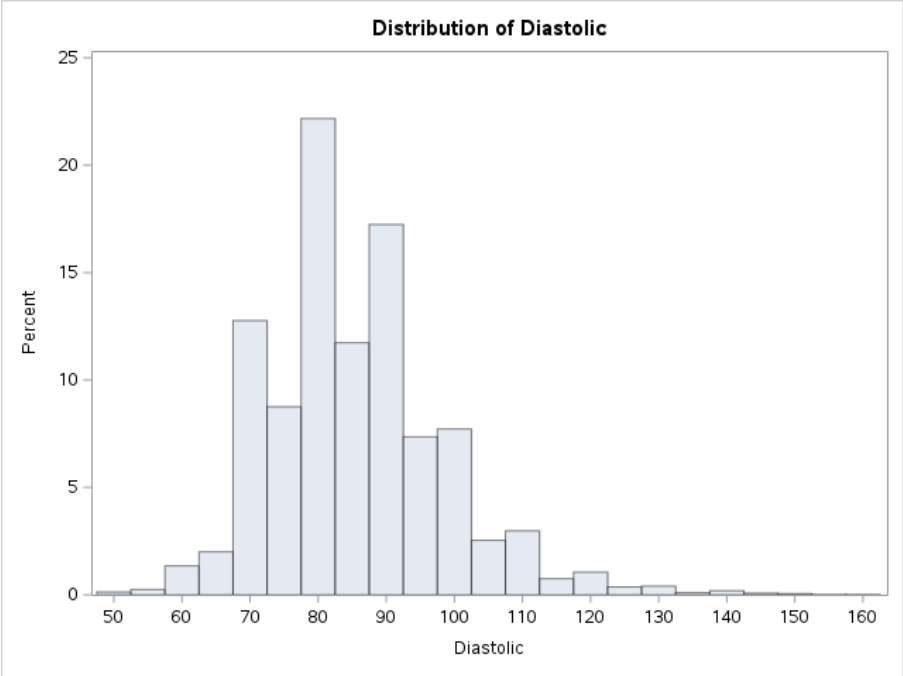
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



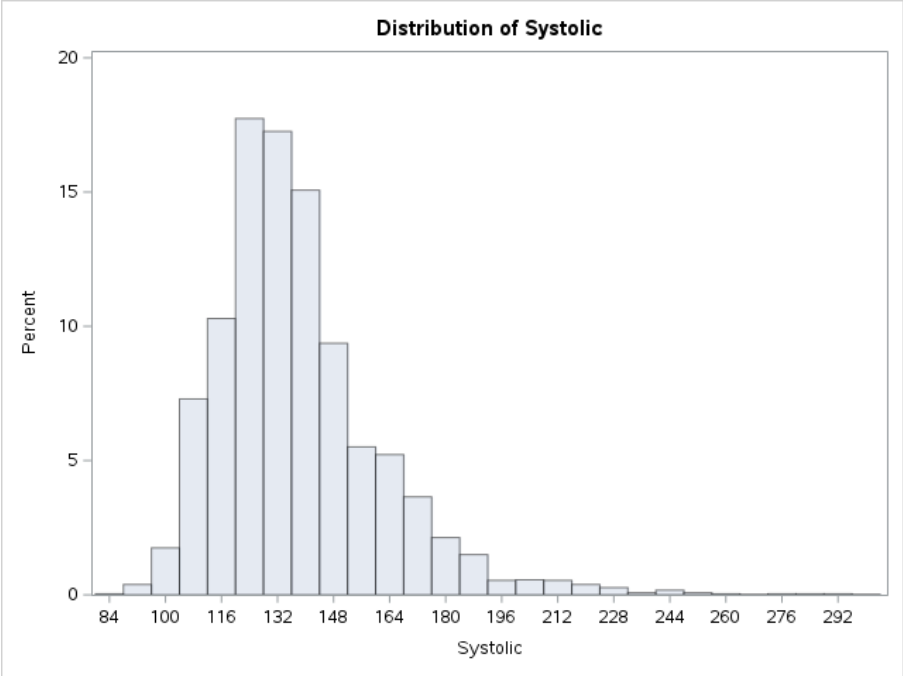
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



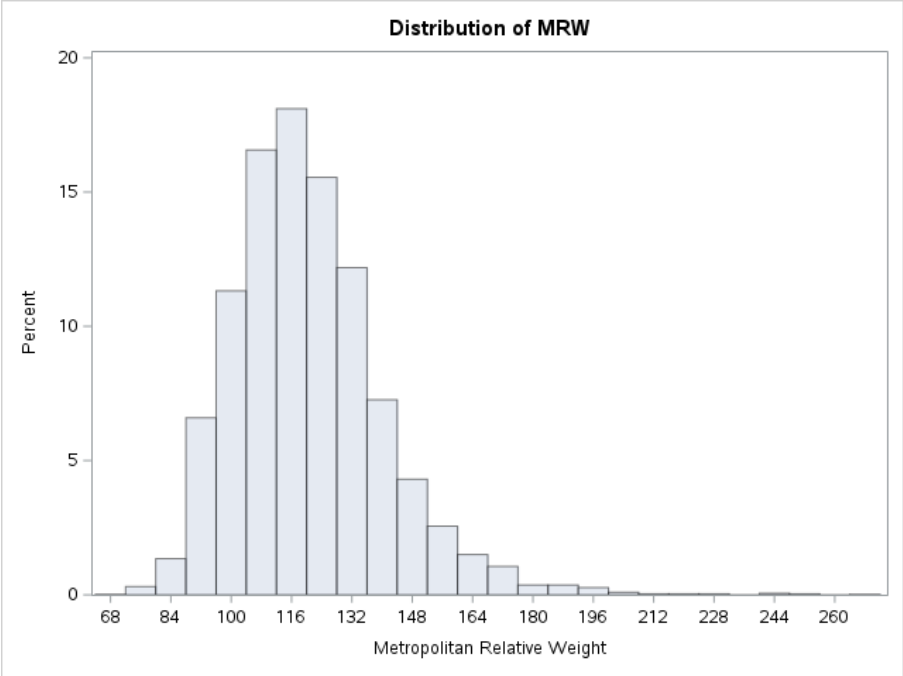
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



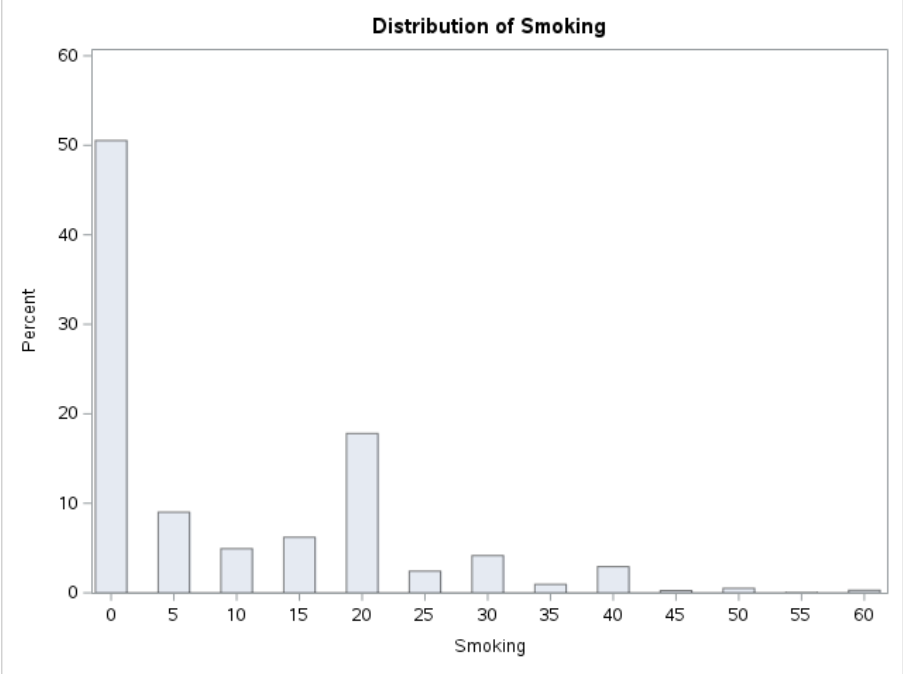
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



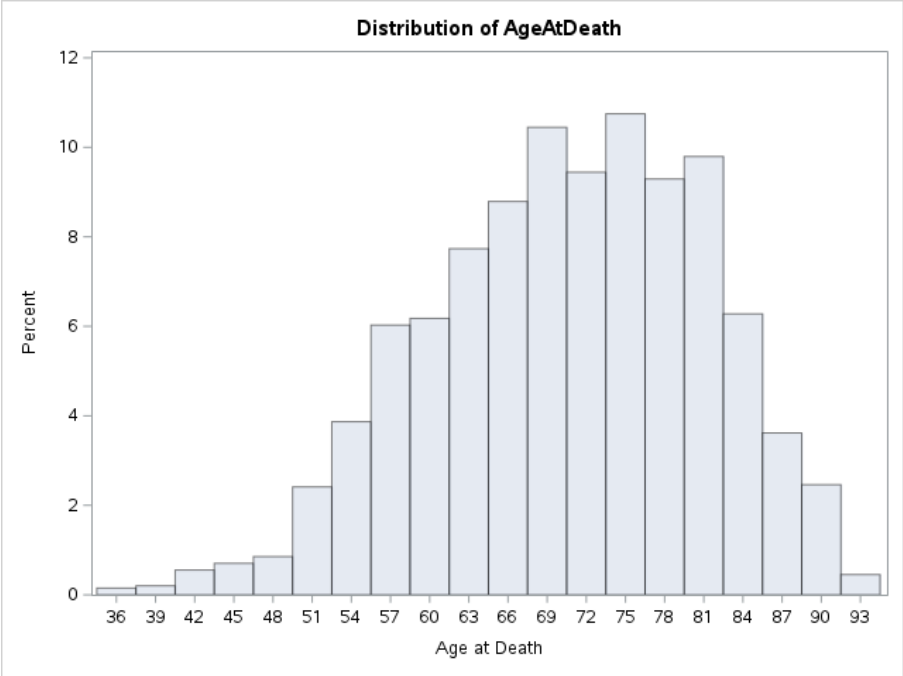
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



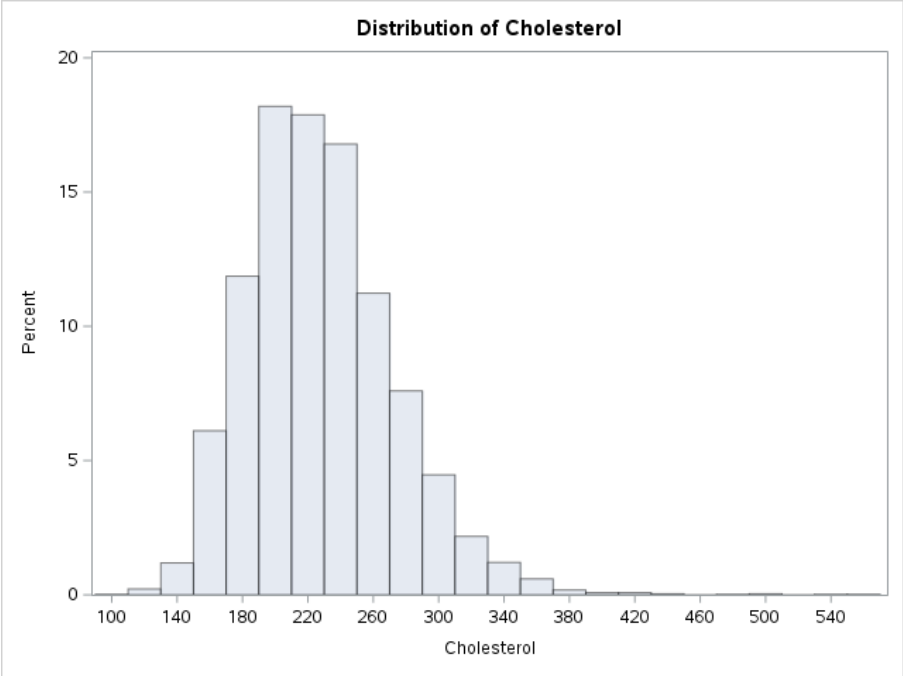
We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



We Observe The Shape,Spread and Central Tendency of The Data

To Visualize The Data Using Histogram (proc univariate)



We Observe The Shape,Spread and Central Tendency of The Data

To Find out Missing Values in Data (proc format,freq,delete)

Missing Data Frequencies

Status	Frequency	Percent
Non-missing	5209	100.00

Cause of Death		
DeathCause	Frequency	Percent
	3218	61.78
Non-missing	1991	38.22

Age CHD Diagnosed		
AgeCHDdiag	Frequency	Percent
.	3760	72.18
Non-missing	1449	27.82

Sex	Frequency	Percent
Non-missing	5209	100.00

Age at Start		
AgeAtStart	Frequency	Percent
Non-missing	5209	100.00

Height	Frequency	Percent
.	6	0.12
Non-missing	5203	99.88

Weight	Frequency	Percent
.	6	0.12
Non-missing	5203	99.88

Diastolic	Frequency	Percent
Non-missing	5209	100.00

Systolic	Frequency	Percent
Non-missing	5209	100.00

Metropolitan Relative Weight		
MRW	Frequency	Percent
.	6	0.12
Non-missing	5203	99.88

Smoking	Frequency	Percent
.	36	0.69
Non-missing	5173	99.31

Age at Death		
AgeAtDeath	Frequency	Percent
.	3218	61.78
Non-missing	1991	38.22

Cholesterol	Frequency	Percent
.	152	2.92
Non-missing	5057	97.08

Cholesterol Status		
Chol_Status	Frequency	Percent
	152	2.92
Non-missing	5057	97.08

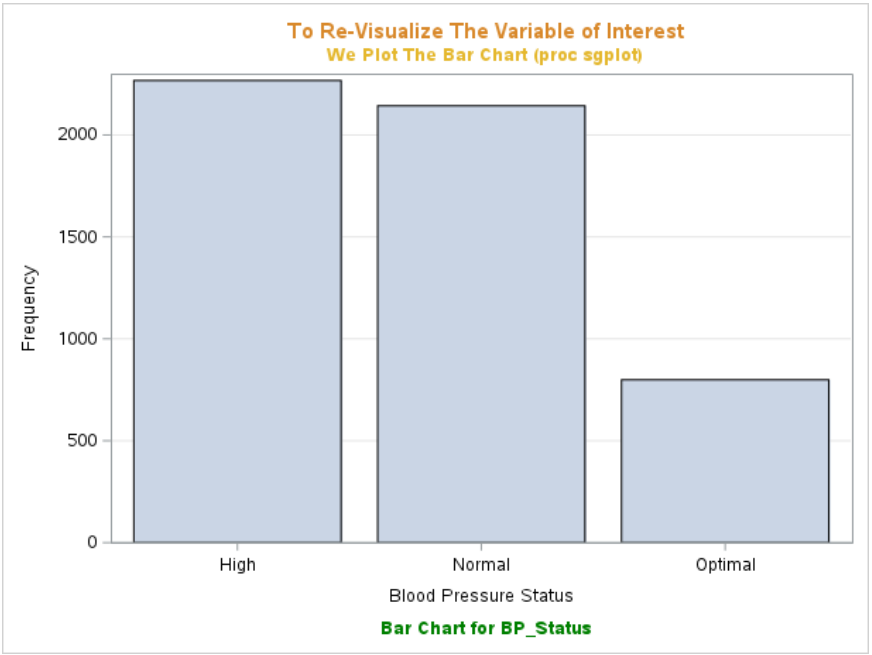
Blood Pressure Status		
BP_Status	Frequency	Percent
Non-missing	5209	100.00

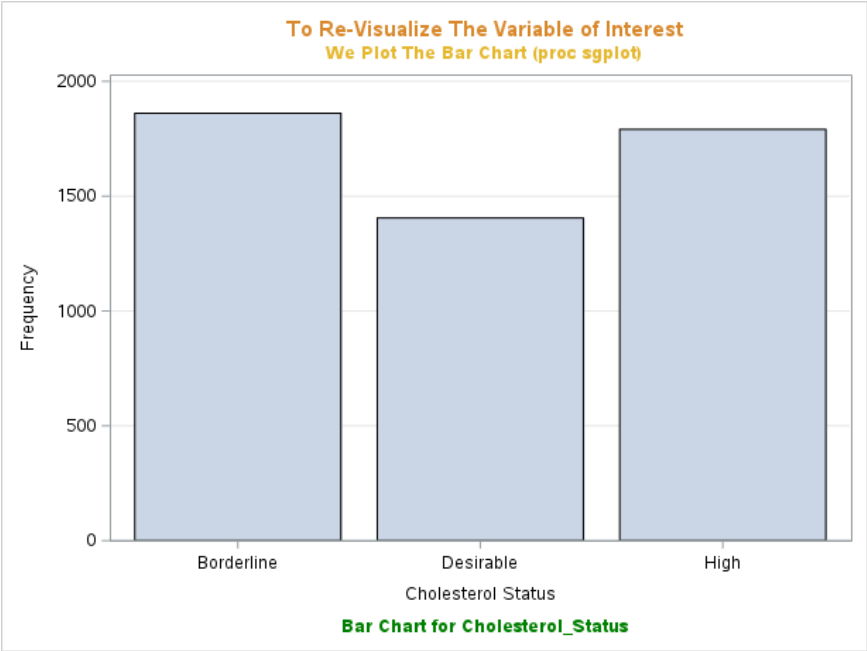
Weight Status		
Weight_Status	Frequency	Percent
	6	0.12
Non-missing	5203	99.88

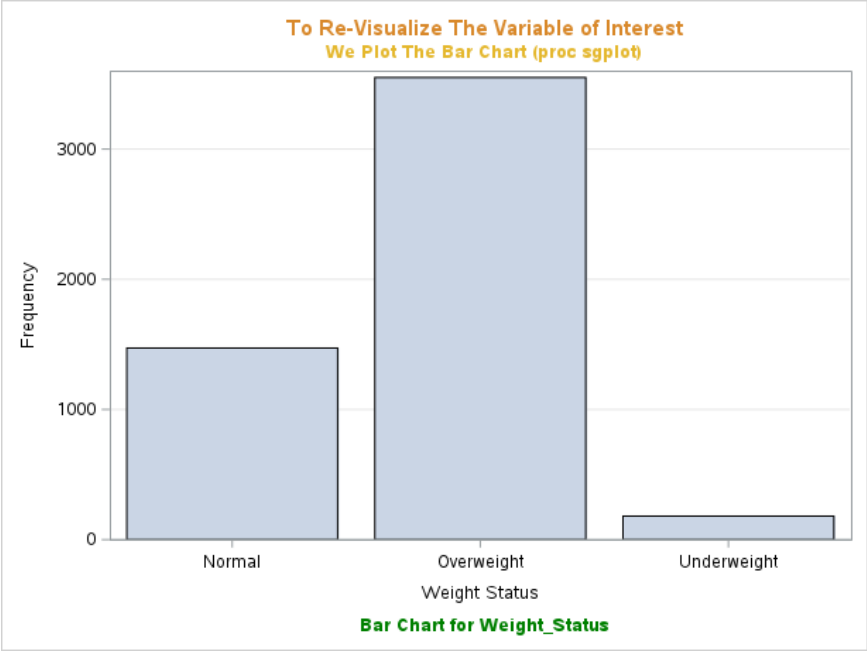
Smoking Status		
Smoking_Status	Frequency	Percent
	36	0.69
Non-missing	5173	99.31

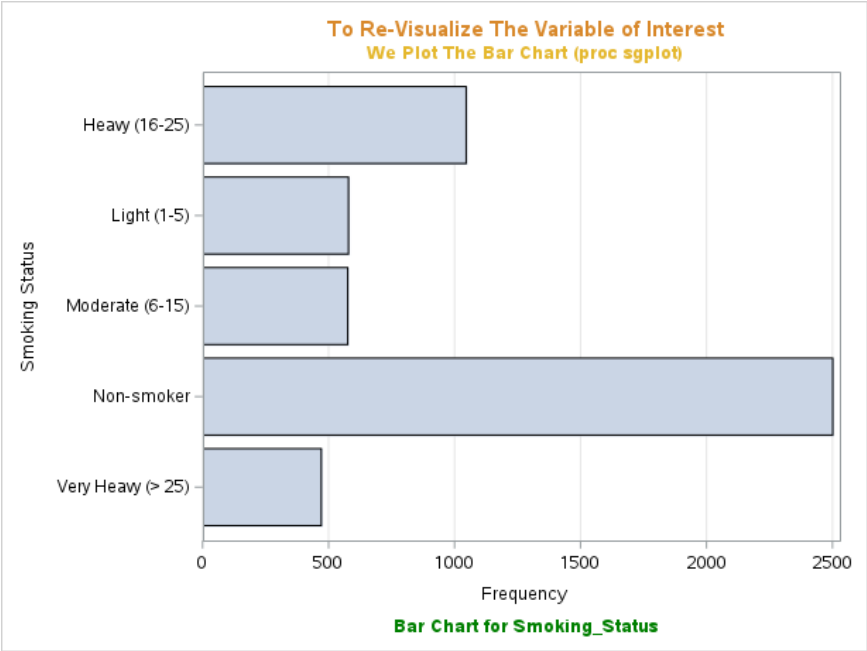
Missing Data Patterns across Variables

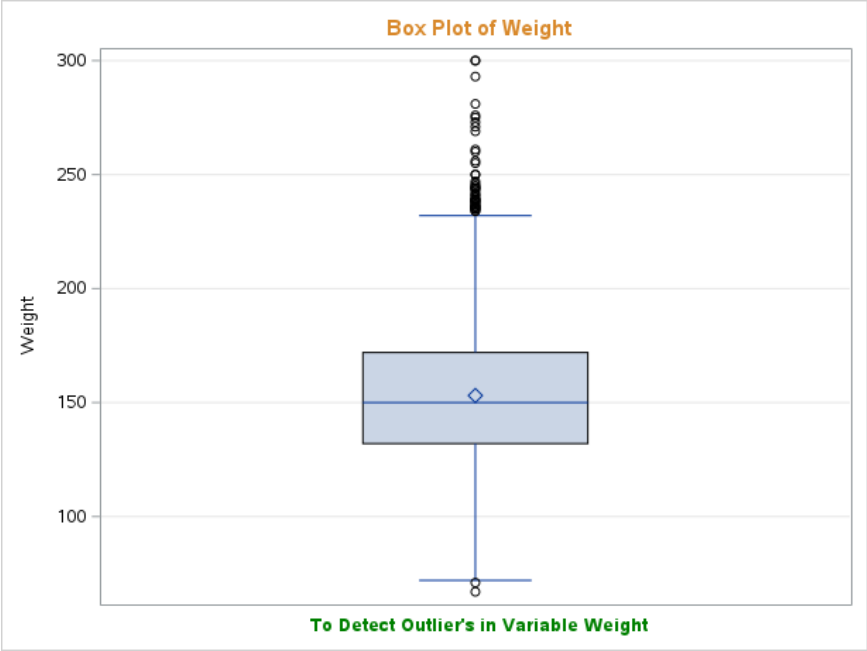
Status	Cause of Death	Age CHD Diagnosed	Sex	Age at Start	Height	Weight	Diastolic	Systolic	Metropolitan Relative Weight	Smoking	Age at Death	Cholesterol	Cholesterol Status	Blood Pressure Status	Weight Status	Smoking Status	Frequency	Percent
Non-missing		.	Non-missing	Non-missing	.	.	Non-missing	Non-missing	.	Non-missing	.	Non-missing	Non-missing	Non-missing		Non-missing	1	0.0192
Non-missing		.	Non-missing	Non-missing	Non-missing	.	Non-missing	Non-missing	.	Non-missing	.	Non-missing	Non-missing	Non-missing		Non-missing	2	0.0384
Non-missing		.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	.	.		Non-missing	Non-missing		7	0.1344
Non-missing		.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	.	Non-missing	Non-missing	Non-missing	Non-missing		6	0.1152
Non-missing		.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	.		Non-missing	Non-missing	Non-missing	62	1.1902
Non-missing		.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	2585	49.6256
Non-missing		Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	.	.		Non-missing	Non-missing		2	0.0384
Non-missing		Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	.	Non-missing	Non-missing	Non-missing	Non-missing		1	0.0192
Non-missing		Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	.		Non-missing	Non-missing	Non-missing	12	0.2304
Non-missing		Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	540	10.3667
Non-missing	Non-missing	.	Non-missing	Non-missing	.	.	Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing		Non-missing	1	0.0192
Non-missing	Non-missing	.	Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	3	0.0576
Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	Non-missing	.		Non-missing	Non-missing		13	0.2496
Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing		1	0.0192
Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.		Non-missing	Non-missing	Non-missing	29	0.5567
Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	1050	20.1574
Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	1	0.0192
Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.	Non-missing	Non-missing	.	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing		Non-missing	2	0.0384
Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.		Non-missing	Non-missing		6	0.1152
Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	.		Non-missing	Non-missing	Non-missing	21	0.4031
Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	Non-missing	864	16.5867

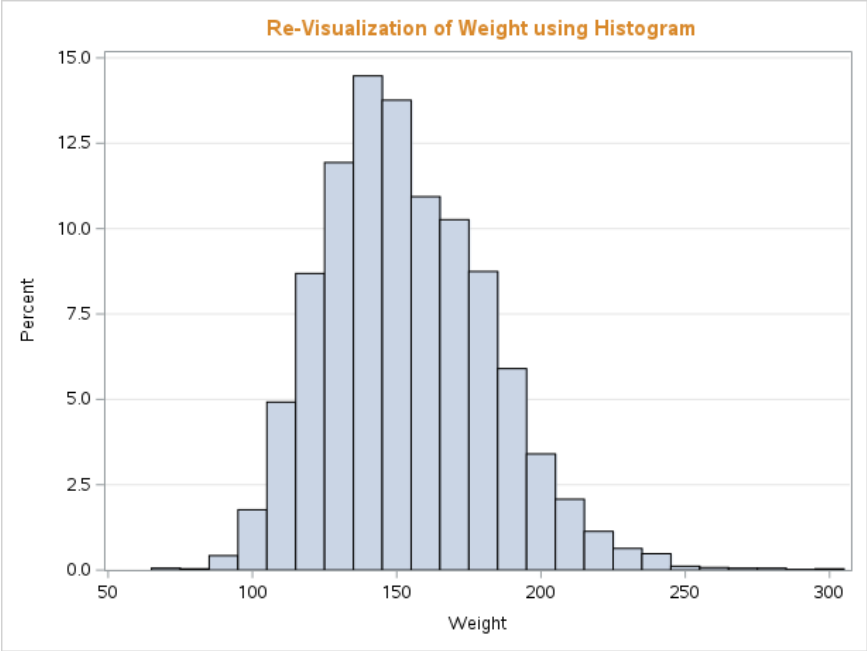



To Re-Visualize The Variable of Interest
We Plot The Bar Chart (proc sgplot)
2000
1500
1000
500
0
Frequency
Borderline
Desirable
High
Cholesterol Status
Bar Chart for Cholesterol_Status

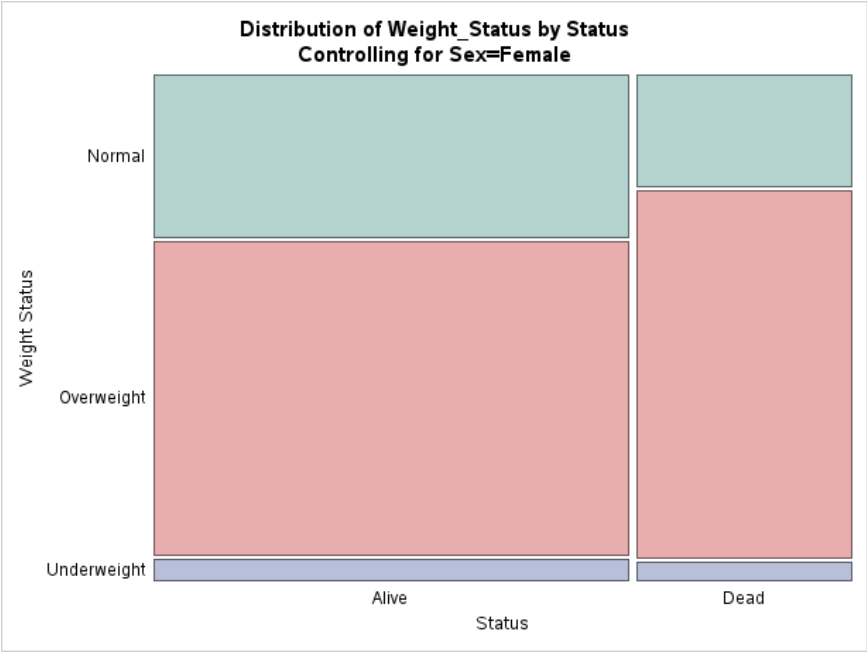

To Re-Visualize The Variable of Interest
We Plot The Bar Chart (proc sgplot)
3000
2000
1000
0
Frequency
Normal
Overweight
Underweight
Weight Status
Bar Chart for Weight_Status

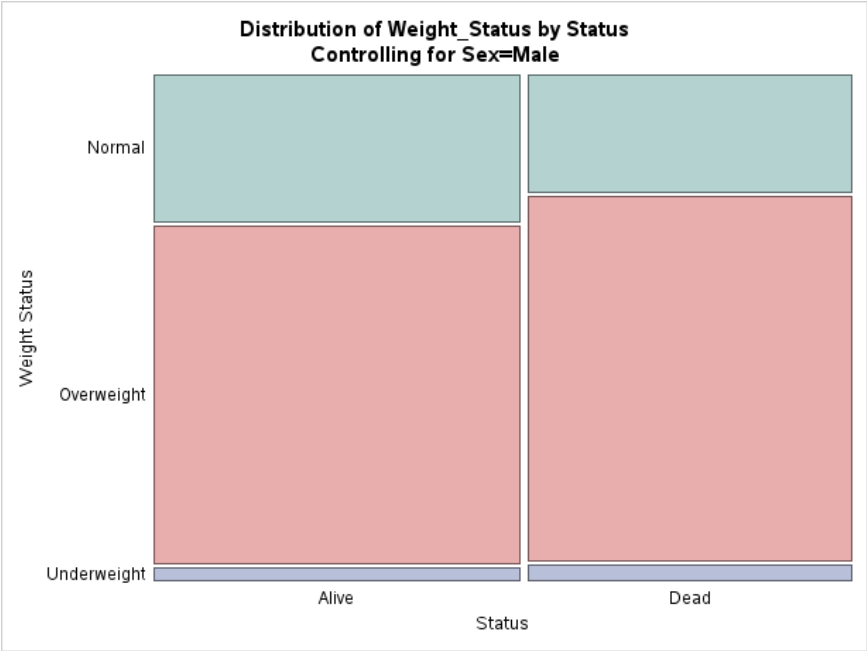

To Re-Visualize The Variable of Interest
We Plot The Bar Chart (proc sgplot)
Heavy (16-25)
Light (1-5)
Moderate (6-15)
Non-smoker
Very Heavy (> 25)
Smoking Status
0
500
1000
1500
2000
2500
Frequency
Bar Chart for Smoking_Status


Box Plot of Weight
300
250
200
150
100
Weight
To Detect Outlier's in Variable Weight



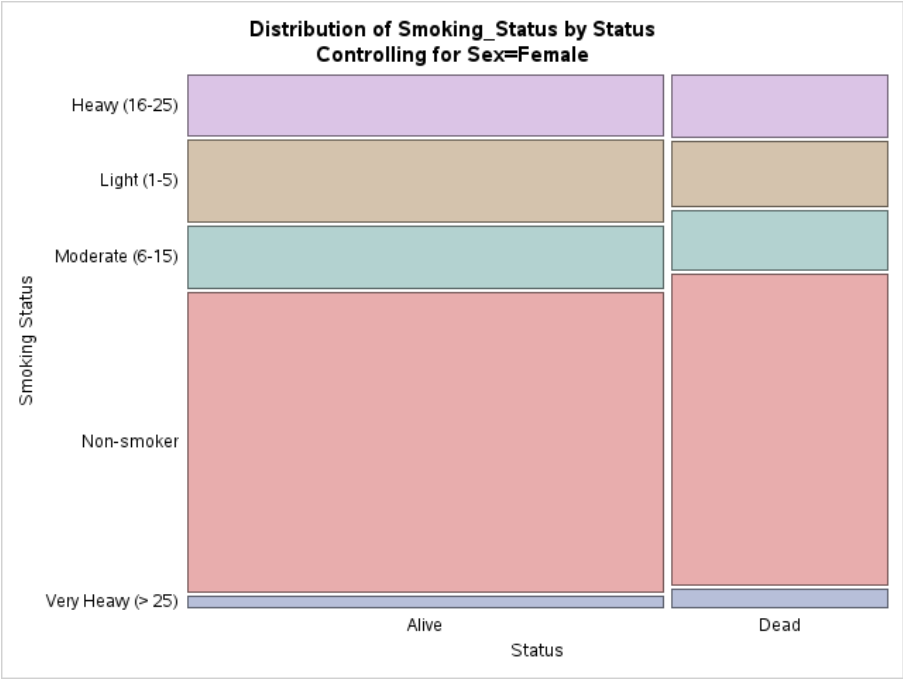
Mosaic Plot (proc freq)
Mosaic plots help show relationships and give a visual way to compare groups.

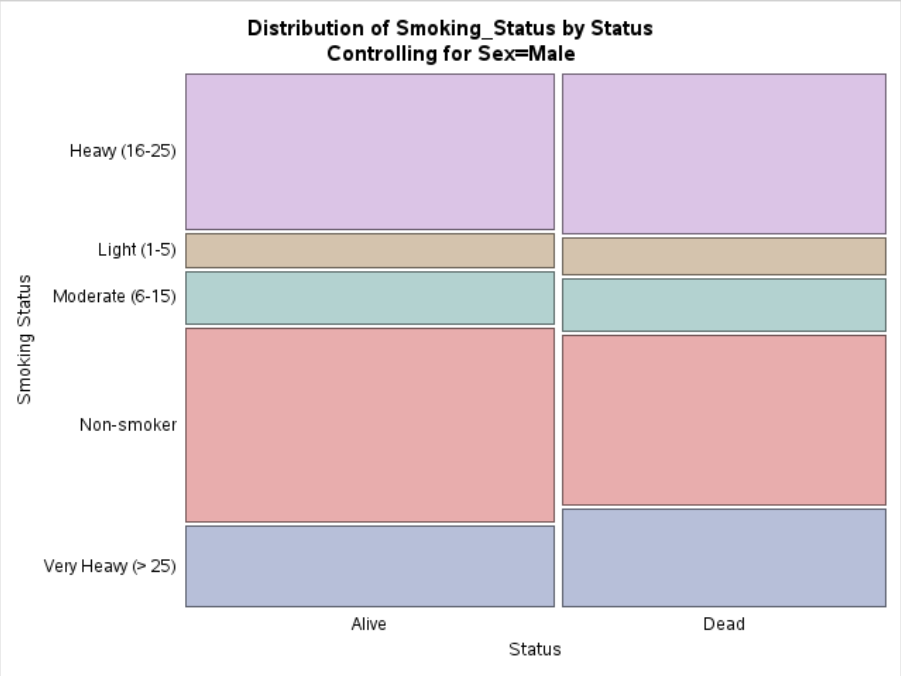




We visualize the Relationship Between Status,Gender and Weight_Status

---

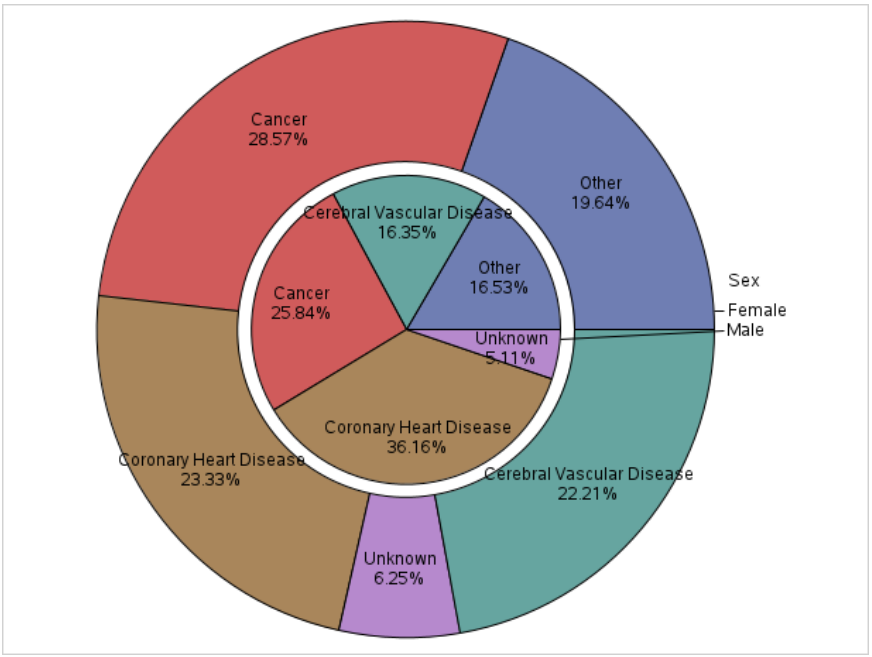


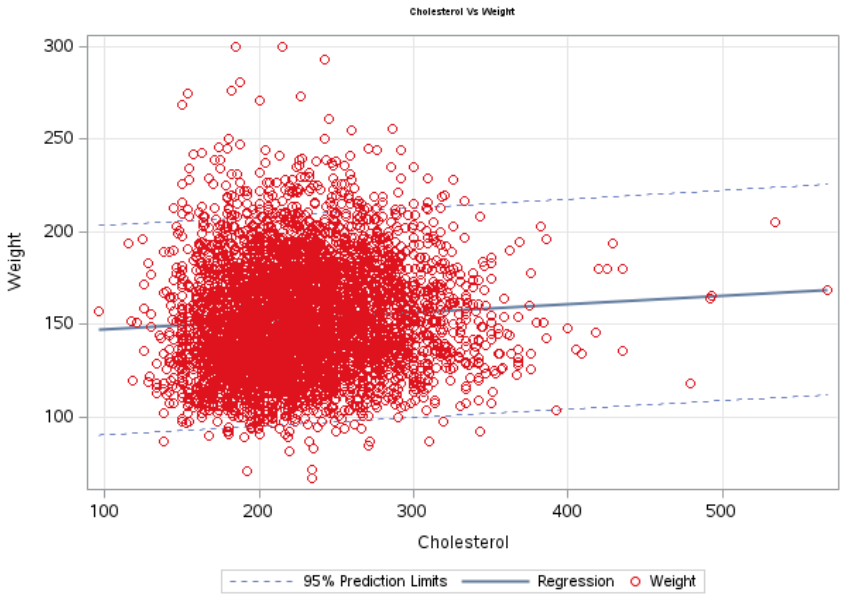


We visualize the Relationship Between Status,Gender and Smoking_Status

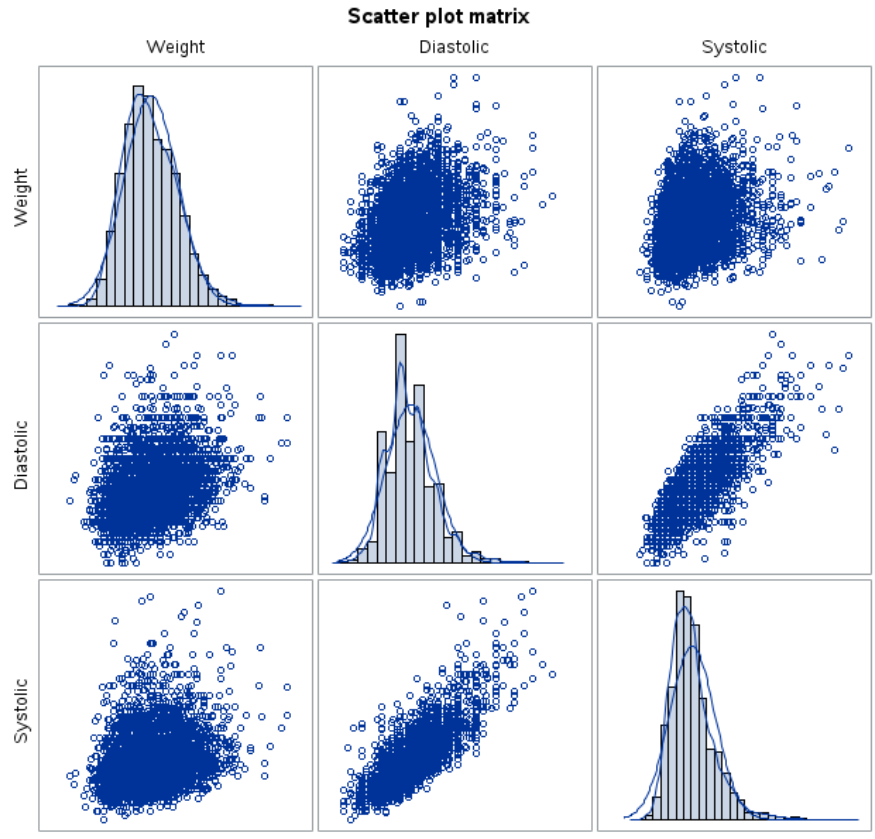
---

Pie Chart (proc template)
A pie chart helps organize and show data as a percentage of a whole.



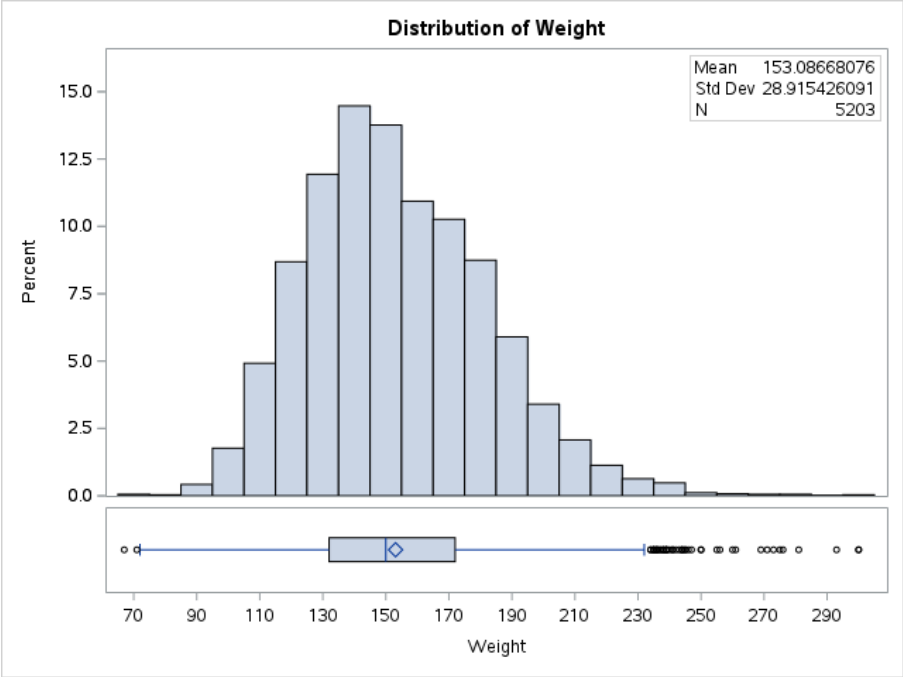


To check relation between Cholesterol and Weight



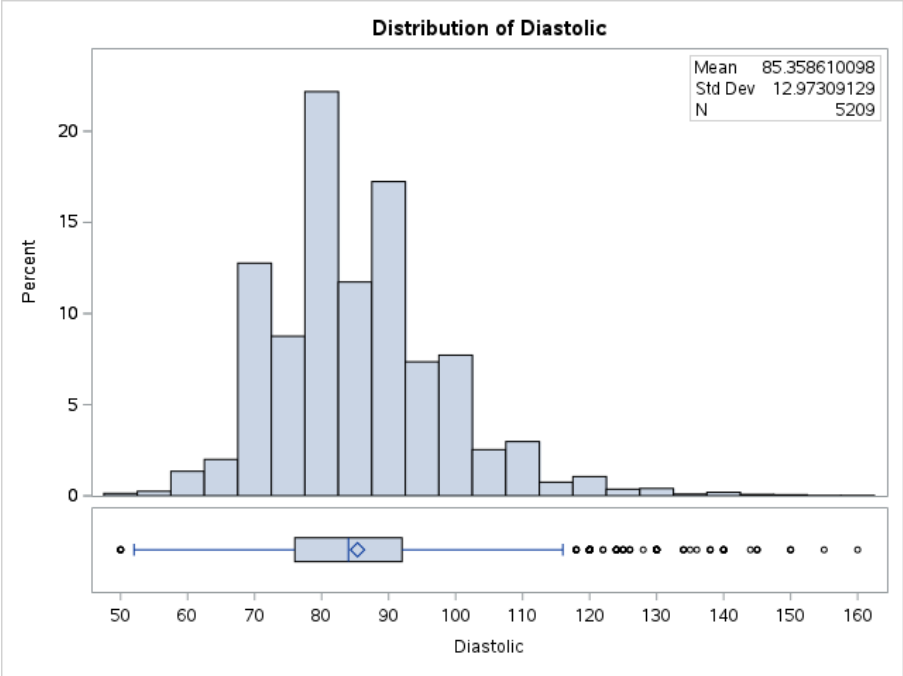
To visualize the relation between variable of intrest i.e. Weight ,Diastolic , Systolic

---



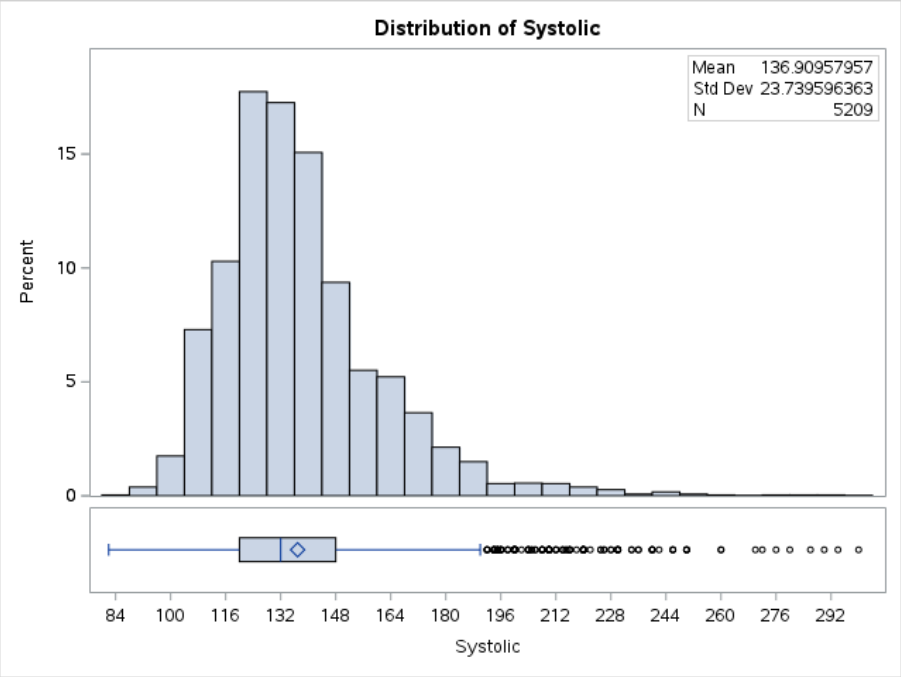
To visualize the relation between variable of intrest i.e. Weight ,Diastolic , Systolic

---



To visualize the relation between variable of intrest i.e. Weight ,Diastolic , Systolic

---



To visualize the relation between variable of intrest i.e. Weight ,Diastolic , Systolic

---

Summary Statistics of variable of interest

Variable	Mean	Std Dev	Minimum	Maximum	N
Weight	153.0866808	28.9154261	67.0000000	300.0000000	5203
Cholesterol	227.4174412	44.9355238	96.0000000	568.0000000	5057
Diastolic	85.3586101	12.9730913	50.0000000	160.0000000	5209
Systolic	136.9095796	23.7395964	82.0000000	300.0000000	5209

Mean of Cholesterol be 227.41 and Weight is 153.08

---

1 With Variables:	Cholesterol
1 Variables:	Weight

SSCP Matrix	
	Weight
Cholesterol	176293120.0

Covariance Matrix, DF = 5050	
	Weight
Cholesterol	94.05349981

Simple Statistics						
Variable	N	Mean	Std Dev	Median	Minimum	Maximum
Cholesterol	5051	227.40467	44.92447	223.00000	96.00000	568.00000
Weight	5051	153.06890	28.90566	150.00000	67.00000	300.00000

Pearson Correlation Coefficients, N = 5051 Prob > |r| under H0: Rho=0	
	Weight
Cholesterol	0.07243 <.0001

Spearman Correlation Coefficients, N = 5051 Prob > |r| under H0: Rho=0	
	Weight
Cholesterol	0.09730 <.0001

Pearson Correlation Statistics (Fisher's z Transformation)										
									H0:Rho=Rho0	
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		Rho0	p Value
Weight	Cholesterol	5051	0.07243	0.07256	7.17113E-6	0.07242	0.044932	0.099801	0	<.0001

Spearman Correlation Statistics (Fisher's z Transformation)										
									H0:Rho=Rho0	
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		Rho0	p Value
Weight	Cholesterol	5051	0.09730	0.09760	9.63325E-6	0.09729	0.069895	0.124531	0	<.0001

There is less correlation .

Variable: Weight

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.048465	Pr > D	<0.0100
Cramer-von Mises	W-Sq	3.05329	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	18.50518	Pr > A-Sq	<0.0050

The null hypothesis is that the data is normally distributed, while the alternative hypothesis is that it is not . If the p-value is less than the significance level, the null hypothesis is rejected and the data is not considered normally distributed .
The data will Right_tailed.