```
/*Print out the combined data set*/
PROC PRINT DATA = TITANIC;
RUN;

/*Convert Sex from male/female to a binary variable 1/0 */
DATA TITANIC;
        SET TITANIC;
        IF Sex = 'male' THEN Gender = 1;
        ELSE IF Sex = 'female' THEN Gender = 0;
        ELSE Gender = '.';
DROP Sex;
RUN;

/*Create child variable by defining the observations with age <=13 as a child */
DATA TITANIC;
        SET TITANIC;
        IF Age <= 13 THEN Child = 1;
        ELSE  Child = 0;
RUN;
```

Obs	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Gender	Child
1	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S	1	0
2	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C	0	0
3	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925		S	0	0
4	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S	0	0
5	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S	1	0
6	6	0	3	Moran, Mr. James	male	.	0	0	330877	8.4583		Q	1	1
7	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S	1	0
8	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S	1	1
9	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S	0	0
10	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C	0	0
11	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7	G6	S	0	1
12	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S	0	0
13	13	0	3	Saundercock, Mr. William Henry	male	20	0	0	A/5. 2151	8.05		S	1	0
14	14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.275		S	1	0
15	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542		S	0	0
16	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16		S	0	0
17	17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125		Q	1	1
18	18	1	2	Williams, Mr. Charles Eugene	male	.	0	0	244373	13		S	1	1
19	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31	1	0	345763	18		S	0	0
20	20	1	3	Masselmani, Mrs. Fatima	female	.	0	0	2649	7.225		C	0	1
21	21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26		S	1	0
22	22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698	13	D56	S	1	0
23	23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923	8.0292		Q	0	0
24	24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788	35.5	A6	S	1	0
25	25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909	21.075		S	0	1

```
/*Get number of missing values using NMISS*/
PROC MEANS DATA = TITANIC NMISS N MEAN MEDIAN MAX MIN;
        TITLE  "Titanic data";
RUN;
```

Titanic data

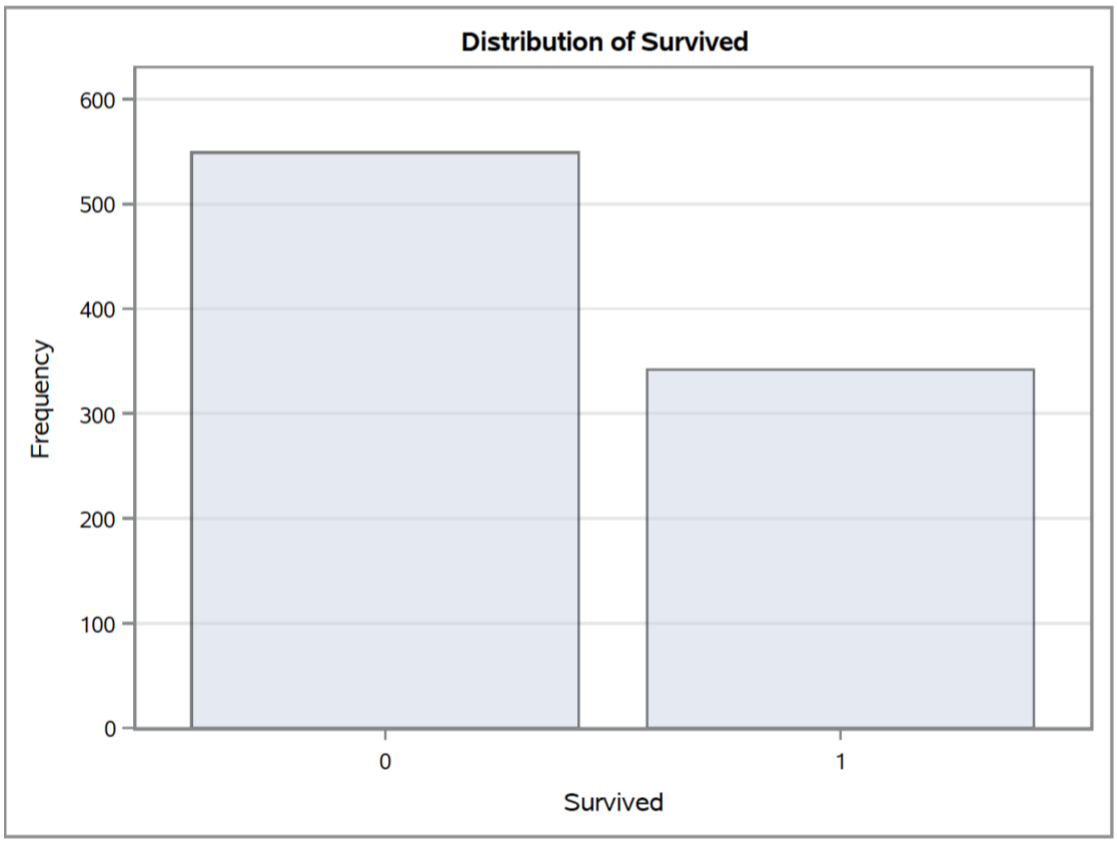
The MEANS Procedure

Variable	N Miss	N	Mean	Median	Maximum	Minimum
PassengerId	0	1309	655.0000000	655.0000000	1309.00	1.0000000
Survived	418	891	0.3838384	0	1.0000000	0
Pclass	0	1309	2.2948816	3.0000000	3.0000000	1.0000000
Age	263	1046	29.8811377	28.0000000	80.0000000	0.1700000
SibSp	0	1309	0.4988541	0	8.0000000	0
Parch	0	1309	0.3850267	0	9.0000000	0
Fare	1	1308	33.2954793	14.4542000	512.3292000	0
Gender	0	1309	0.6440031	1.0000000	1.0000000	0
Child	0	1309	0.2765470	0	1.0000000	0

```
/*Show the number of survivors and percentage*/
PROC FREQ DATA = TITANIC;
        TABLES survived / PLOTS=freqplot;
RUN;
```

The FREQ Procedure

Survived	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	549	61.62	549	61.62
1	342	38.38	891	100.00
Frequency Missing = 418				



```
/*Distribution of age variable*/
PROC UNIVARIATE DATA = TITANIC PLOT NORMAL;
        VAR age;
RUN;
```

The UNIVARIATE Procedure
Variable: Age

Moments			
N	1046	Sum Weights	1046
Mean	29.8811377	Sum Observations	31255.67
Std Deviation	14.4134932	Variance	207.748787
Skewness	0.40767456	Kurtosis	0.14694764
Uncorrected SS	1151052.46	Corrected SS	217097.482
Coeff Variation	48.2360925	Std Error Mean	0.44565974

Basic Statistical Measures			
Location		Variability	
Mean	29.88114	Std Deviation	14.41349
Median	28.00000	Variance	207.74879
Mode	24.00000	Range	79.83000
		Interquartile Range	18.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	67.04922	Pr > |t|	<.0001
Sign	M	523	Pr >= |M|	<.0001
Signed Rank	S	273790.5	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.979547	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.078928	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.306703	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	7.129863	Pr > A-Sq	<0.0050

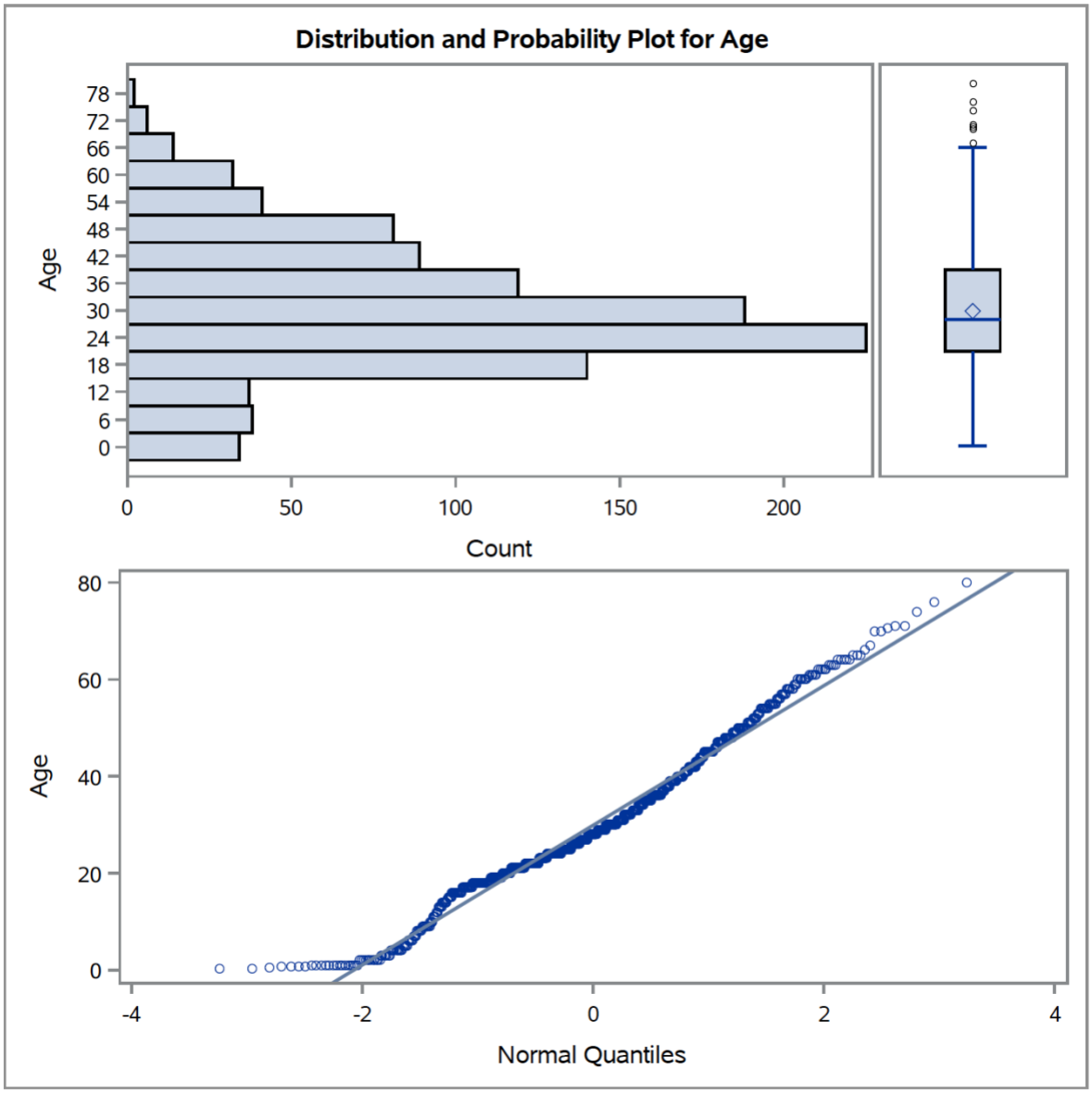
The UNIVARIATE Procedure
Variable: Age

Quantiles (Definition 5)	
Level	**Quantile**
10%	14.00
5%	5.00
1%	0.92
0% Min	0.17

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
0.17	1246	71	97
0.33	1093	71	494
0.42	804	74	852
0.67	756	76	988
0.75	1173	80	631

Missing Values			
		Percent Of	
Missing Value	**Count**	**All Obs**	**Missing Obs**
.	263	20.09	100.00

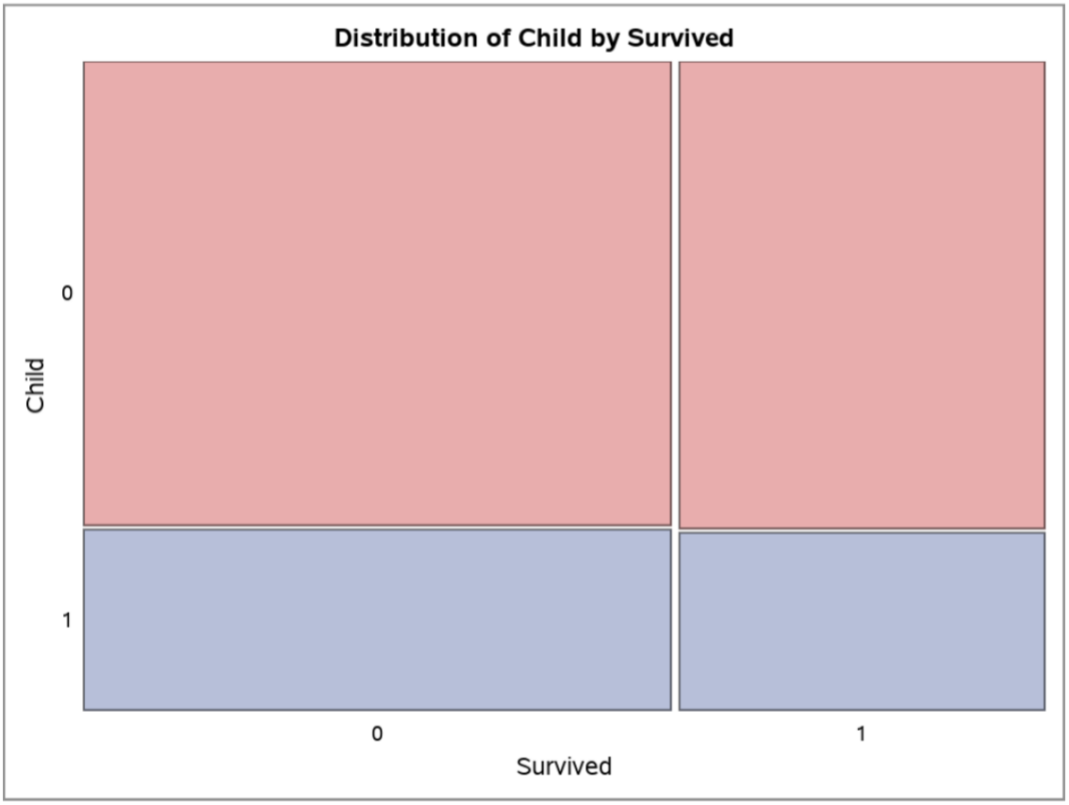
The UNIVARIATE Procedure



```sas
/*Show the number and percentage of child survivors compared to adults using a mosaicplot*/
PROC FREQ DATA = TITANIC;
        TABLES child*survived / PLOTS= mosaicplot;
RUN;
```

The FREQ Procedure

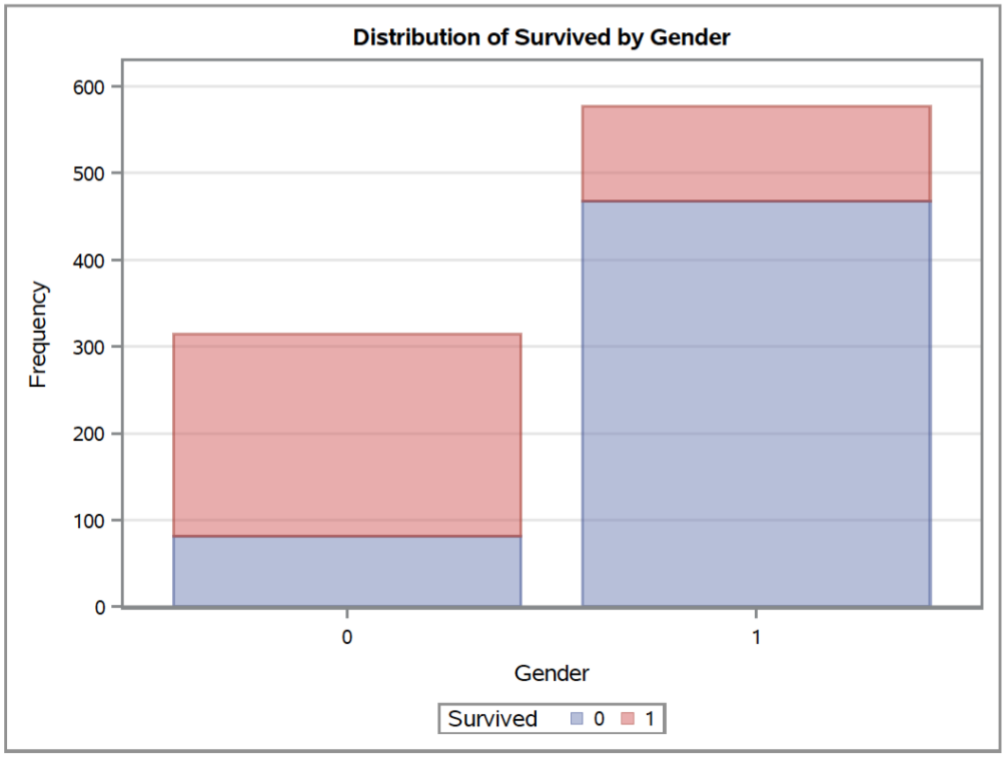
Frequency Percent Row Pct Col Pct	Table of Child by Survived		
	Survived		
Child	0	1	Total
0	395 44.33 61.43 71.95	248 27.83 38.57 72.51	643 72.17
1	154 17.28 62.10 28.05	94 10.55 37.90 27.49	248 27.83
Total	549 61.62	342 38.38	891 100.00
Frequency Missing = 418			

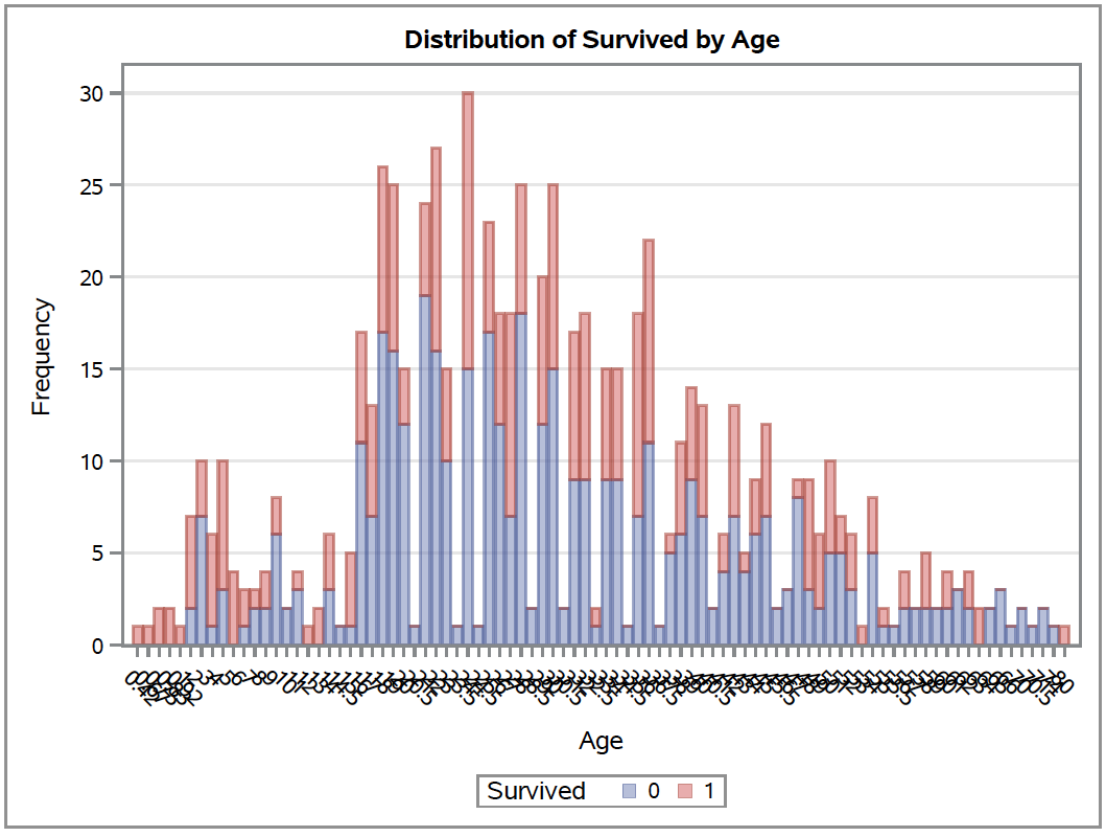


```
/*Show the number and percentage of survivors by gender and survivors by age*/
PROC FREQ DATA=TITANIC;
        TABLE survived*gender survived*age / PLOTS = freqplot(twoway = stacked);
RUN;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Survived by Gender		
	Gender		
Survived	0	1	Total
0	81 9.09 14.75 25.80	468 52.53 85.25 81.11	549 61.62
1	233 26.15 68.13 74.20	109 12.23 31.87 18.89	342 38.38
Total	314 35.24	577 64.76	891 100.00
Frequency Missing = 418			




Distribution of Survived by Age
Frequency
30
25
20
15
10
5
0
Age
Survived
0
1

```
/*The plot for age was difficult to read as it is not binned.*/
/*To solve the problem, proc sgpanel is used to display a histogram with age bined into 12 bins*/
PROC SGPANEL DATA = TITANIC;
        PANELBY survived;
        HISTOGRAM age / GROUP = Gender Scales = count nbins=12;
RUN;
```