```
title1 Bold color=black" VISHAL KISHOR KAPADNIS (2211)";
TITLE2 BOLD COLOR=RED" Exploratoory Data Analysis of Heart_Data from SASHelp_DataSets";


data heart;
   set sashelp.heart;
run;

/* proc print data=heart; */
run;
TITLE1;
TITLE2;


/* List Of Attributes in The Data */
title color=bio"List Of Attributes in The Data";
ods noproctitle;
ods select attributes variables;

proc datasets;
    contents data=WORK.HEART order=collate;
    footnote1 bold color=green" Here 17 Variables are Present and  observation size 5209";
quit;

title;
footnote1;

ods noproctitle;

/*** Analyze categorical variables ***/


title1 color=bio"To Analyze Categorical Variables  We use proc freq ";
title2 "Frequencies for Categorical Variables";
proc freq data=WORK.HEART;
    tables Status DeathCause Sex Chol_Status BP_Status Weight_Status
        Smoking_Status / plots=(freqplot);
    footnote1 bold color=green"Here we Plotted Simple Bar Graph and its Corresponding Freq..";
run;
title1;
title2;
footnote1;



/*** Analyze numeric variables ***/
title1 color=bio "Descriptive Statistics for Numeric Variables (proc means)";

proc means data=WORK.HEART n nmiss min mean median max std;
    var AgeCHDdiag AgeAtStart Height Weight Diastolic Systolic MRW Smoking
        AgeAtDeath Cholesterol;
        footnote1 bold color=green"We Analyze The Descriptive Statistics of The Data";
run;
title1;
footnote1;

title2 color=bioy" To Visualize The Data Using Histogram (proc univariate)";
proc univariate data=WORK.HEART noprint;
    histogram AgeCHDdiag AgeAtStart Height Weight Diastolic Systolic MRW Smoking
        AgeAtDeath Cholesterol;
        footnote2 bold color=green"We Observe The Shape,Spread and Central Tendency of The Data";
run;

title2;
footnote2;



/* Data Value Missing */
title1 color=bio"To Find out Missing Values in Data (proc format,freq,delete)";

ods noproctitle;

proc format;
    value _nmissprint low-high="Non-missing";
```

```
    value $_cmissprint " "=" " other="Non-missing";
run;

proc freq data=WORK.HEART;
    title3 "Missing Data Frequencies";

    format AgeCHDdiag AgeAtStart Height Weight Diastolic Systolic MRW Smoking
        AgeAtDeath Cholesterol _nmissprint.;
    format Status DeathCause Sex Chol_Status BP_Status Weight_Status
        Smoking_Status $_cmissprint.;
    tables Status DeathCause AgeCHDdiag Sex AgeAtStart Height Weight Diastolic
        Systolic MRW Smoking AgeAtDeath Cholesterol Chol_Status BP_Status
        Weight_Status Smoking_Status / missing nocum;
run;
title1;

proc freq data=WORK.HEART noprint;
    table Status * DeathCause * AgeCHDdiag * Sex * AgeAtStart * Height * Weight *
        Diastolic * Systolic * MRW * Smoking * AgeAtDeath * Cholesterol * Chol_Status
        * BP_Status * Weight_Status * Smoking_Status / missing out=Work._MissingData_;
    format AgeCHDdiag AgeAtStart Height Weight Diastolic Systolic MRW Smoking
        AgeAtDeath Cholesterol _nmissprint.;
    format Status DeathCause Sex Chol_Status BP_Status Weight_Status
        Smoking_Status $_cmissprint.;
run;

proc print data=Work._MissingData_ noobs label;
    title3 "Missing Data Patterns across Variables";

    format AgeCHDdiag AgeAtStart Height Weight Diastolic Systolic MRW Smoking
        AgeAtDeath Cholesterol _nmissprint.;
    format Status DeathCause Sex Chol_Status BP_Status Weight_Status
        Smoking_Status $_cmissprint.;
    label count="Frequency" percent="Percent";
run;

title3;

/* Clean up */
proc delete data=Work._MissingData_;
run;



/* Bar chart of BP_Status */
title1 color=bio" To Re-Visualize The Variable of Interest";
title2 color=bioy" We Plot The Bar Chart (proc sgplot)";
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.HEART;
    vbar BP_Status /;
    yaxis grid;
    footnote bold color=green" Bar Chart for BP_Status";
run;
footnote;
ods graphics / reset;

/* Bar chart of Chol_Status */
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.HEART;
    vbar Chol_Status /;
    yaxis grid;
    footnote bold color=green" Bar Chart for Cholesterol_Status";
run;
footnote;
ods graphics / reset;

/* Bar chart of Weight_Status */
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.HEART;
    vbar Weight_Status /;
    yaxis grid;
        footnote bold color=green" Bar Chart for Weight_Status";
run;
footnote;
```

```
ods graphics / reset;

/* Bar chart of Smoking_Status */

ods graphics / reset width=6.4in height=4.8in imagemap;
proc sgplot data=WORK.HEART;
    hbar Smoking_Status /;
    xaxis grid;
        footnote bold color=green" Bar Chart for Smoking_Status";
run;
footnote;
ods graphics / reset;
title1;
title2;


/* Box plot of weight */
title color=bio" Box Plot of Weight ";
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.HEART;
    vbox Weight /;
    yaxis grid;
    footnote bold color=green" To Detect Outlier's in Variable Weight";
run;
title;
footnote;
ods graphics / reset;


/* Histogram of weight */
title color=bio" Re-Visualization of Weight using Histogram";
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.HEART;
    histogram Weight /;
    yaxis grid;
run;
title;




/* Mosaic Plot of Status,Gender and WeightStatus   */
title1 color=bio"Mosaic Plot (proc freq)";
title2 color=bioy" Mosaic plots help show relationships and give a visual way to compare groups.";
ods noproctitle;

proc freq data=WORK.HEART;
    ods select MosaicPlot;
    tables Sex*Weight_Status*Status / plots=mosaicplot;
    footnote bold color=green" We visualize the Relationship Between Status,Gender and Weight_Status";
run;
title1;
title2;
footnote;
ods graphics / reset;



/* Mosaic Plot of Status,Gender and SmokingStatus*/
ods noproctitle;

proc freq data=WORK.HEART;
    ods select MosaicPlot;
    tables Sex*Smoking_Status*Status / plots=mosaicplot;
    footnote bold color=green" We visualize the Relationship Between Status,Gender and Smoking_Status";
run;
footnote;



/*Pie Chart of DeathCause and subcarecterized by Gender*/
/* Define Pie template */
title1 color=bio" Pie Chart (proc template)";
title2 color=bioy"  A pie chart helps organize and show data as a percentage of a whole.";
```

```
/* Define Pie template */
proc template;
    define statgraph SASStudio.Pie;
        begingraph;
        layout region;
        piechart category=DeathCause / group=Sex groupgap=2% stat=pct
            datalabellocation=inside;
        endlayout;
        endgraph;
    end;
run;

ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgrender template=SASStudio.Pie data=WORK.HEART;
run;
title1;
title2;


/*Scatter Plot of Weight and Chole*/

ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.HEART;
    title height=5pt "Cholesterol Vs Weight";
    footnote2 justify=left height=12pt
        "To check relation between Cholesterol and Weight";
    reg x=Cholesterol y=Weight / nomarkers cli alpha=0.05;
    scatter x=Cholesterol y=Weight / markerattrs=(color=CXdf131d);
    xaxis grid;
    yaxis grid;


run;

ods graphics / reset;
title;
footnote2;


/*Data Exploration of Weight,diastolic and systolic*/
options validvarname=any;
ods noproctitle;
ods graphics / imagemap=on;

/* Scatter plot matrix macro */
%macro scatterPlotMatrix(xVars=, title=, groupVar=);
    proc sgscatter data=WORK.HEART;
        matrix &xVars / %if(&groupVar ne %str()) %then
            %do;
                group=&groupVar legend=(sortorder=ascending) %end;
        diagonal=(histogram kernel normal);
        title &title;
    run;

    title;
%mend scatterPlotMatrix;

/* Histogram and box plot template */
proc template;
    define statgraph histobox;
        dynamic AVAR;
        begingraph;
        entrytitle "Distribution of " eval(catq('q', colname(AVAR)));
        layout lattice / rows=2 columndatarange=union rowgutter=0 rowweights=(0.75
            0.25);
        layout overlay / yaxisopts=(offsetmax=0.1) xaxisopts=(display=none);
        layout gridded / columns=2 border=on autoalign=(topright topleft);
        %let _lft = halign=left;
        %let _rgt = halign=right;
        entry &_lft "Mean";
        entry &_rgt eval(strip(put(mean(AVAR), best.)));
        entry &_lft "Std Dev";
        entry &_rgt eval(strip(put(stddev(AVAR), best.)));
        entry &_lft "N";
        entry &_rgt eval(strip(put(n(AVAR), best.)));
```

```
        endlayout;
        histogram AVAR /;
        endlayout;
        layout overlay /;
        BoxPlot Y=AVAR / orient=horizontal;
        endlayout;
        endlayout;
        endgraph;
    end;
    footnote bold color=green" To visualize the relation between variable of intrest i.e. Weight ,Diastolic , Systolic ";
run;


%scatterPlotMatrix(xVars=Weight Diastolic Systolic,
    title="Scatter plot matrix", groupVar=);

proc sgrender data=WORK.HEART template=histobox;
    dynamic AVAR='Weight';
run;

proc sgrender data=WORK.HEART template=histobox;
    dynamic AVAR='Diastolic';
run;

proc sgrender data=WORK.HEART template=histobox;
    dynamic AVAR='Systolic';
run;


/*Summary Statistics*/

ods noproctitle;
ods graphics / imagemap=on;
title color=bio" Summary Statistics of variable of interest ";
proc means data=WORK.HEART chartype mean std min max n vardef=df;
    var Weight Cholesterol Diastolic Systolic;
    footnote bold color=green" Mean of Cholesterol be 227.41 and Weight is 153.08  ";
run;
title;
footnote;



/*Correlation Analysis*/
ods noproctitle;
ods graphics / imagemap=on;

proc corr data=WORK.HEART pearson cov sscp spearman fisher(rho0=0 alpha=0.05
        type=TwoSided) nomiss plots=scatter(noinset ellipse=none nvar=2 nwith=2);
    var Weight;
    with Cholesterol;
    footnote bold color=green" There is less correlation . ";
run;
footnote;

/* Test for normality */
proc univariate data=WORK.HEART normal mu0=0;
    ods select TestsForNormality;
    var Weight;

footnote1 bold color=green"  The null hypothesis is that the data is normally distributed, while the alternative hypothesis i
footnote2 bold color=green" The data will Right_tailed.  ";
run;
footnote1;
footnote2;
```