

```

1  /*****
2  File : heart_descriptives.sas
3  Date: 7/18/2025
4  Brief : Clean SASHELP.HEART and generate descriptive statistics/plots
5  *****/
6
7  /* 1. Copy the built-in data to WORK so we can modify it safely */
8  proc sql;
9      create table work.heart_raw as
10     select *
11     from sashelp.heart;
12 quit;
13
14
15 /* 2. Initial inspection – how many rows / basic attribute info */
16 proc contents data=work.heart_raw order=varnum;
17     title "Variable Attributes - Raw Heart Data";
18 run;
19
20
21 proc freq data=work.heart_raw nlevels;
22     tables _all_ / nopercnt norow nocol;
23     title "Missing-value Scan - Raw Data";
24 run;
25
26
27 /* Choose a directory to save interim files if you wish */
28 %let projdir = %sysfunc(pathname(work));
29
30 /* ---- 3A. Filter and derive new fields ----- */
31 data work.heart_clean
32     work.heart_excluded;      /* keep the rejects to audit later */
33     set work.heart_raw;
34
35     /* Adult filter */
36     if AgeAtStart < 18 then do;
37         output work.heart_excluded;
38         return;
39     end;
40
41 /* Simple completeness rule:
42     Need (Height and Weight) OR (Systolic and Diastolic).
43     Missing either pair kicks the record out. */
44 if cmiss(Height, Weight) = 0 or cmiss(Systolic, Diastolic) = 0 then do;
45     /*---- Derived fields ----*/
46     if cmiss(Height, Weight) = 0 then bmi = (Weight / (Height**2)) * 703; /* height in inches */
47
48     if Systolic ne . and Diastolic ne . then do;
49         if Systolic < 120 and Diastolic < 80 then bp_cat = "Normal";
50         else if Systolic < 130 and Diastolic < 80 then bp_cat = "Elevated";
51         else if Systolic < 140 or Diastolic < 90 then bp_cat = "Stage 1";
52         else bp_cat = "Stage 2";
53     end;
54     output work.heart_clean;
55 end;
56 else output work.heart_excluded;
57 run;
58
59 /* ---- 3B. Quick sanity checks ----- */
60 title "Row Counts After Cleaning";
61 proc sql;
62     select "Kept rows" as Group, count(*) as N from work.heart_clean
63     union all
64     select "Excluded" as Group, count(*) from work.heart_excluded
65     union all
66     select "Original" as Group, count(*) from work.heart_raw;

```

```

67 quit;
68
69 title "Essential Descriptive Statistics - Clean Heart Data";
70 proc means data=work.heart_clean n nmiss mean std min p25 median p75 max;
71   class sex;
72   var AgeAtStart Weight Height Bmi Systolic Diastolic Cholesterol;
73 run;
74
75
76 title "Distribution of Blood-Pressure Category";
77 proc freq data=work.heart_clean;
78   tables bp_cat / missing nocum nopercnt;
79 run;
80
81 title "BMI by Sex";
82 proc sgplot data=work.heart_clean;
83   vbox bmi / category=sex;
84 run;
85
86 title "Systolic vs Diastolic (colored by BP category)";
87 proc sgplot data=work.heart_clean;
88   scatter x=systolic y=diastolic / group=bp_cat transparency=0.2;
89   refline 120 130 140 / axis=x lineattrs=(pattern=shortdash);
90   refline 80 90 / axis=y lineattrs=(pattern=shortdash);
91 run;
92
93 title; footnote;
94
95
96 /*****
97   Export for GitHub
98   *****/
99 proc export data=work.heart_clean
100   outfile="%projdir./heart_clean.csv"
101   dhms=csv

```