UCI HEART DISEASE DATA

Getting a Cleaned Dataset

Checking for missing values, unknown values, and incorrect data type

The MEANS Procedure

Variable	N Miss	N	Mean	Std Dev	Minimum	Maximum
id	0	920	460.5000000	265.7254222	1.0000000	920.0000000
age	0	920	53.5108696	9.4246852	28.0000000	77.0000000
trestbps	59	861	132.1324042	19.0660695	0	200.0000000
chol	30	890	199.1303371	110.7808104	0	603.0000000
thalch	55	865	137.5456647	25.9262765	60.0000000	202.0000000
oldpeak	62	858	0.8787879	1.0912262	-2.6000000	6.2000000
ca	611	309	0.6763754	0.9356530	0	3.0000000
num	0	920	0.9956522	1.1426934	0	4.0000000

Source: UCI Heart Data

UCI HEART DISEASE DATA

Exploring the Cleaned Dataset

The CONTENTS Procedure

Data Set Name	HEART.HEART_DISEASES	Observations	920
Member Type	DATA	Variables	15
Engine	V9	Indexes	0
Created	12/03/2024 22:18:24	Observation Length	144
Last Modified	12/03/2024 22:18:24	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Source: UCI Heart Data

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	3
First Data Page	1
Max Obs per Page	454
Obs in First Data Page	437
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	C:\Users\akhil\OneDrive\Desktop\SAS\Project\Heart\heart_diseases.sas7bdat
Release Created	9.0401M8
Host Created	X64_WIN+HOME
Owner Name	ADITYA\akhil
File Size	256KB
File Size (bytes)	262144

Source: UCI Heart Data

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
2	age	Num	8	BEST12.	BEST32.
13	ca	Num	8	BEST12.	BEST32.
7	chol	Num	8	BEST12.	BEST32.
5	cp	Char	15	$15.	$15.
11	exang	Char	5	$5.	$5.
8	fbs	Char	5	$5.	$5.
1	id	Num	8	BEST12.	BEST32.
15	num	Num	8	BEST12.	BEST32.
4	origin	Char	9	$9.	$9.
9	restecg	Char	14	$14.	$14.
3	sex	Char	6	$6.	$6.
12	slope	Char	11	$11.	$11.
14	thal	Char	17	$17.	$17.
10	thalch	Num	8	BEST12.	BEST32.
6	trestbps	Num	8	BEST12.	BEST32.

Source: UCI Heart Data

UCI HEART DISEASE DATA

Print Labelled and Formatted dataset

First 10 Observations of Heart Data

id	age	sex	cp	trestbps	chol	fbs	restecg	exang	thalch	slope	thal	ca	num
1	63	M	typical angina	145	233	T	lv hypertrophy	F	150	downsloping	fixed defect	0	No Disease
2	67	M	asymptomatic	160	286	F	lv hypertrophy	T	108	flat	normal	3	Medium chance of Disease
3	67	M	asymptomatic	120	229	F	lv hypertrophy	T	129	flat	reversable defect	2	Less chance of Disease
4	37	M	non-anginal	130	250	F	normal	F	187	downsloping	normal	0	No Disease
5	41	F	atypical angina	130	204	F	lv hypertrophy	F	172	upsloping	normal	0	No Disease
6	56	M	atypical angina	120	236	F	normal	F	178	upsloping	normal	0	No Disease
7	62	F	asymptomatic	140	268	F	lv hypertrophy	F	160	downsloping	normal	2	Highe chance of Disease
8	57	F	asymptomatic	120	354	F	normal	T	163	upsloping	normal	0	No Disease
9	63	M	asymptomatic	130	254	F	lv hypertrophy	F	147	flat	reversable defect	1	Medium chance of Disease
10	53	M	asymptomatic	140	203	T	lv hypertrophy	T	155	downsloping	reversable defect	0	Less chance of Disease

Source: UCI Heart Data

UCI HEART DISEASE DATA

Print Subset data using different variables

Subsetted data using age and cholestrol levels

Obs	id	age	sex	cp	chol	num
1	1	63	M	typical angina	233	No Disease
2	2	67	M	asymptomatic	286	Medium chance of Disease
3	3	67	M	asymptomatic	229	Less chance of Disease
4	6	56	M	atypical angina	236	No Disease
5	7	62	F	asymptomatic	268	Highe chance of Disease
6	8	57	F	asymptomatic	354	No Disease
7	9	63	M	asymptomatic	254	Medium chance of Disease
8	10	53	M	asymptomatic	203	Less chance of Disease
9	12	56	F	atypical angina	294	No Disease
10	13	56	M	non-anginal	256	Medium chance of Disease

Source: UCI Heart Data

UCI HEART DISEASE DATA

Dividing the formatted dataset

Dividing patients based on the age range

id	age	sex	origin	cp	trestbps	chol	fbs	restecg	thalch	exang	slope	ca	thal	num	age_group
1	63	M	Cleveland	typical angina	145	233	T	lv hypertrophy	150	F	downsloping	0	fixed defect	No Disease	Elder
2	67	M	Cleveland	asymptomatic	160	286	F	lv hypertrophy	108	T	flat	3	normal	Medium chance of Disease	Elder
3	67	M	Cleveland	asymptomatic	120	229	F	lv hypertrophy	129	T	flat	2	reversable defect	Less chance of Disease	Elder
4	37	M	Cleveland	non-anginal	130	250	F	normal	187	F	downsloping	0	normal	No Disease	Young
5	41	F	Cleveland	atypical angina	130	204	F	lv hypertrophy	172	F	upsloping	0	normal	No Disease	Middl
6	56	M	Cleveland	atypical angina	120	236	F	normal	178	F	upsloping	0	normal	No Disease	Middl
7	62	F	Cleveland	asymptomatic	140	268	F	lv hypertrophy	160	F	downsloping	2	normal	Highe chance of Disease	Elder
8	57	F	Cleveland	asymptomatic	120	354	F	normal	163	T	upsloping	0	normal	No Disease	Middl
9	63	M	Cleveland	asymptomatic	130	254	F	lv hypertrophy	147	F	flat	1	reversable defect	Medium chance of Disease	Elder
10	53	M	Cleveland	asymptomatic	140	203	T	lv hypertrophy	155	T	downsloping	0	reversable defect	Less chance of Disease	Middl

Source: UCI Heart Data

UCI HEART DISEASE DATA

Frequency table of Sex and Chest Pain type

Frequency Table of Sex and Chest Pain Type

The FREQ Procedure

Frequency	Table of sex by cp				
sex(Sex of Patient)	cp(Chest Pain Type)				
	asymptomatic	atypical angina	non-anginal	typical angina	Total
F	70	61	53	10	194
M	426	113	151	36	726
Total	496	174	204	46	920

Source: UCI Heart Data

UCI HEART DISEASE DATA

Frequency table of Age group and exercise-induced angina based on elder age category

Frequency Table of Age Group and Exercise-Induced Angina

The FREQ Procedure

age_group=Elder

Frequency Percent Row Pct	Table of age_group by exang		
	exang(Exercise-Induced Angina)		
age_group	F	T	Total
Elder	97 50.00 50.00	97 50.00 50.00	194 100.00
Total	97 50.00	97 50.00	194 100.00
Frequency Missing = 27			

Data reflects exercise response across age groups.

Source: UCI Heart Data

UCI HEART DISEASE DATA

Frequency table of Age group and Exercise-Induced Angina for Middle age category

Frequency Table of Age Group and Exercise-Induced Angina

The FREQ Procedure

age_group=Middl

Frequency Percent Row Pct	Table of age_group by exang		
	exang(Exercise-Induced Angina)		
age_group	F	T	Total
Middl	366 61.82 61.82	226 38.18 38.18	592 100.00
Total	366 61.82	226 38.18	592 100.00
Frequency Missing = 27			

Data reflects exercise response across age groups.

Source: UCI Heart Data

UCI HEART DISEASE DATA

Frequency table of Age group and Exercise-Induced Angina based on young age category

Frequency Table of Age Group and Exercise-Induced Angina

The FREQ Procedure

age_group=Young

Frequency Percent Row Pct	Table of age_group by exang		
	exang(Exercise-Induced Angina)		
age_group	F	T	Total
Young	65 82.28 82.28	14 17.72 17.72	79 100.00
Total	65 82.28	14 17.72	79 100.00
Frequency Missing = 1			

Data reflects exercise response across age groups.

Source: UCI Heart Data

UCI HEART DISEASE DATA

Summary Statistics for Age and Cholesterol variables

Summary Statistics for Age and Cholesterol

The MEANS Procedure

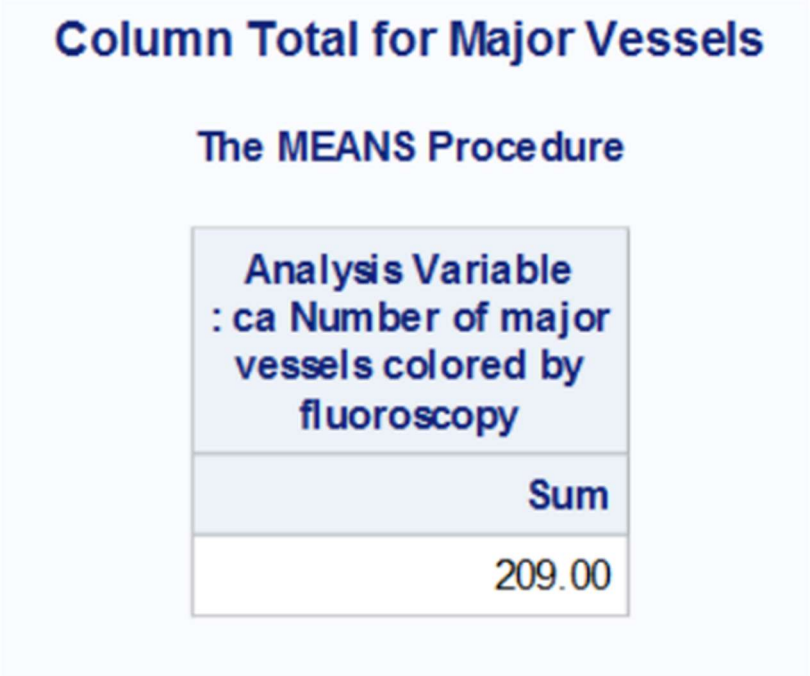
Variable	Label	N	Mean	Median	Std Dev	Variance	Minimum	Maximum
age	Age of Patient	920	53.51	54.00	9.42	88.82	28.00	77.00
trestbps	Resting Blood Pressure	861	132.13	130.00	19.07	363.52	0.00	200.00
thalch	Maximum Heart Rate Achieved	865	137.55	140.00	25.93	672.17	60.00	202.00
chol	Serum Cholesterol	890	199.13	223.00	110.78	12272.39	0.00	603.00

All values are rounded to 2 decimals for clarity.

Source: UCI Heart Data

UCI HEART DISEASE DATA

Column total for Major Vessels



Column Total for Major Vessels

The MEANS Procedure

Analysis Variable : ca Number of major vessels colored by fluoroscopy
Sum
209.00

Source: UCI Heart Data

UCI HEART DISEASE DATA

Uppercase transformation for Thalassemia results

Uppercase Transformation for Thalassemia Results

Obs	id	thal_upper
1	1	FIXED DEFECT
2	2	NORMAL
3	3	REVERSABLE DEFECT
4	4	NORMAL
5	5	NORMAL
6	6	NORMAL
7	7	NORMAL
8	8	NORMAL
9	9	REVERSABLE DEFECT
10	10	REVERSABLE DEFECT

Source: UCI Heart Data

UCI HEART DISEASE DATA

Rounded Cholesterol Values to near 10

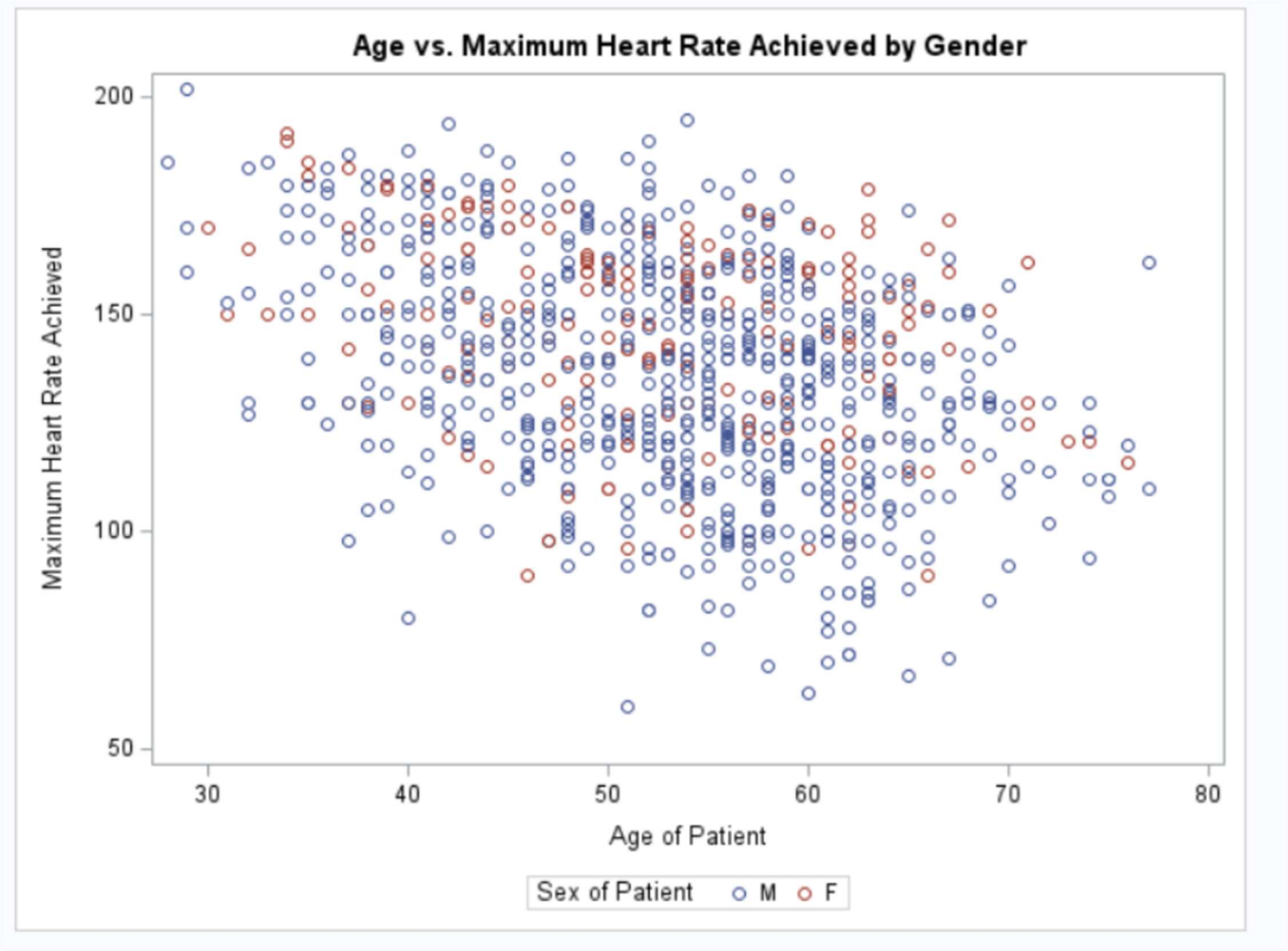
Rounded Cholestrol Values to near 10

Obs	id	chol	rounded_chol
1	1	233	230
2	2	286	280
3	3	229	220
4	4	250	250
5	5	204	200
6	6	236	230
7	7	268	260
8	8	354	350
9	9	254	250
10	10	203	200

Source: UCI Heart Data

UCI HEART DISEASE

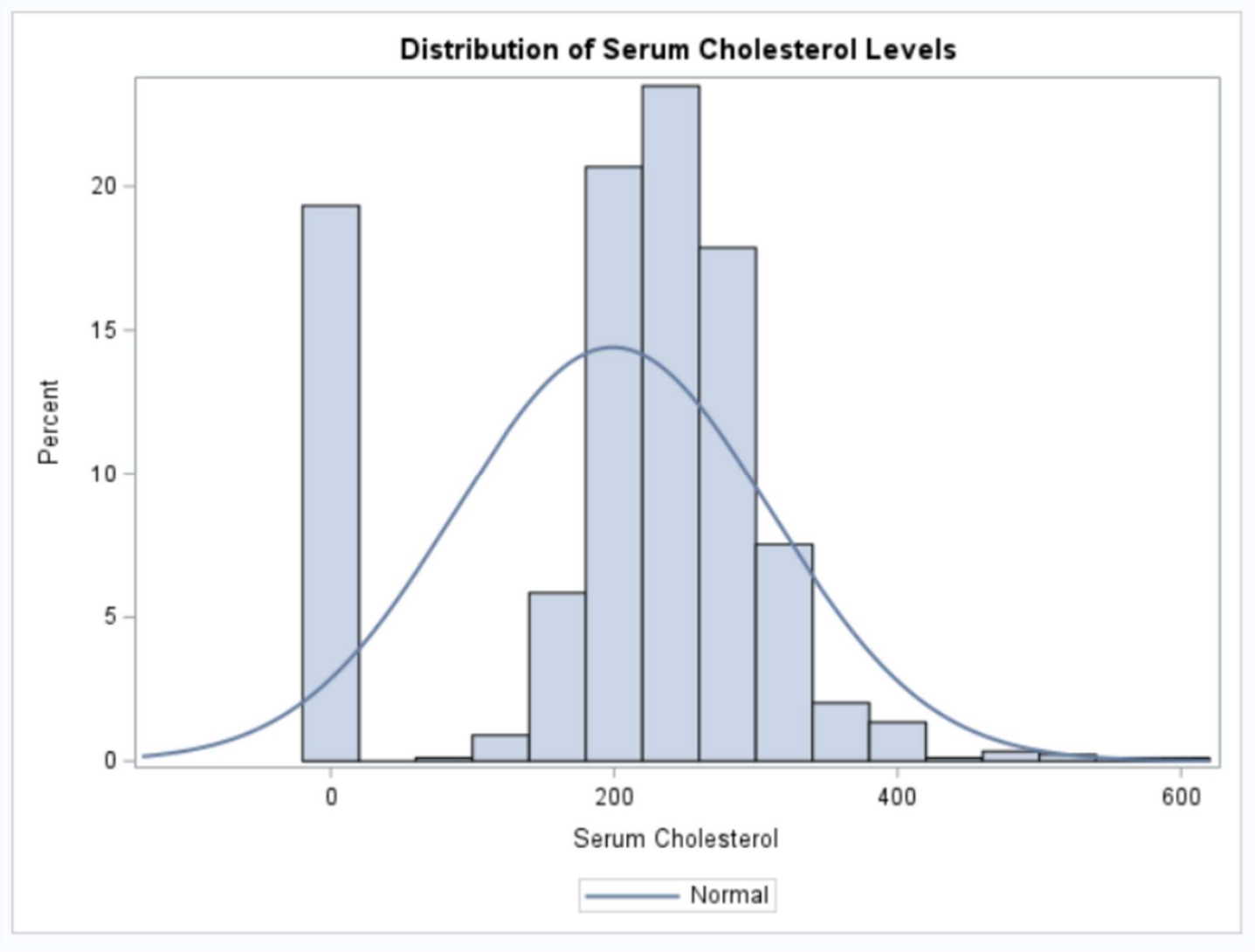
Plot Graph for Age and Maximum Heart Rate Achieved



Source: UCI Heart Data

UCI HEART DISEASE DATA

Normal Distribution Bar Graph of different of Serum Cholesterol Levels



Source: UCI Heart Data