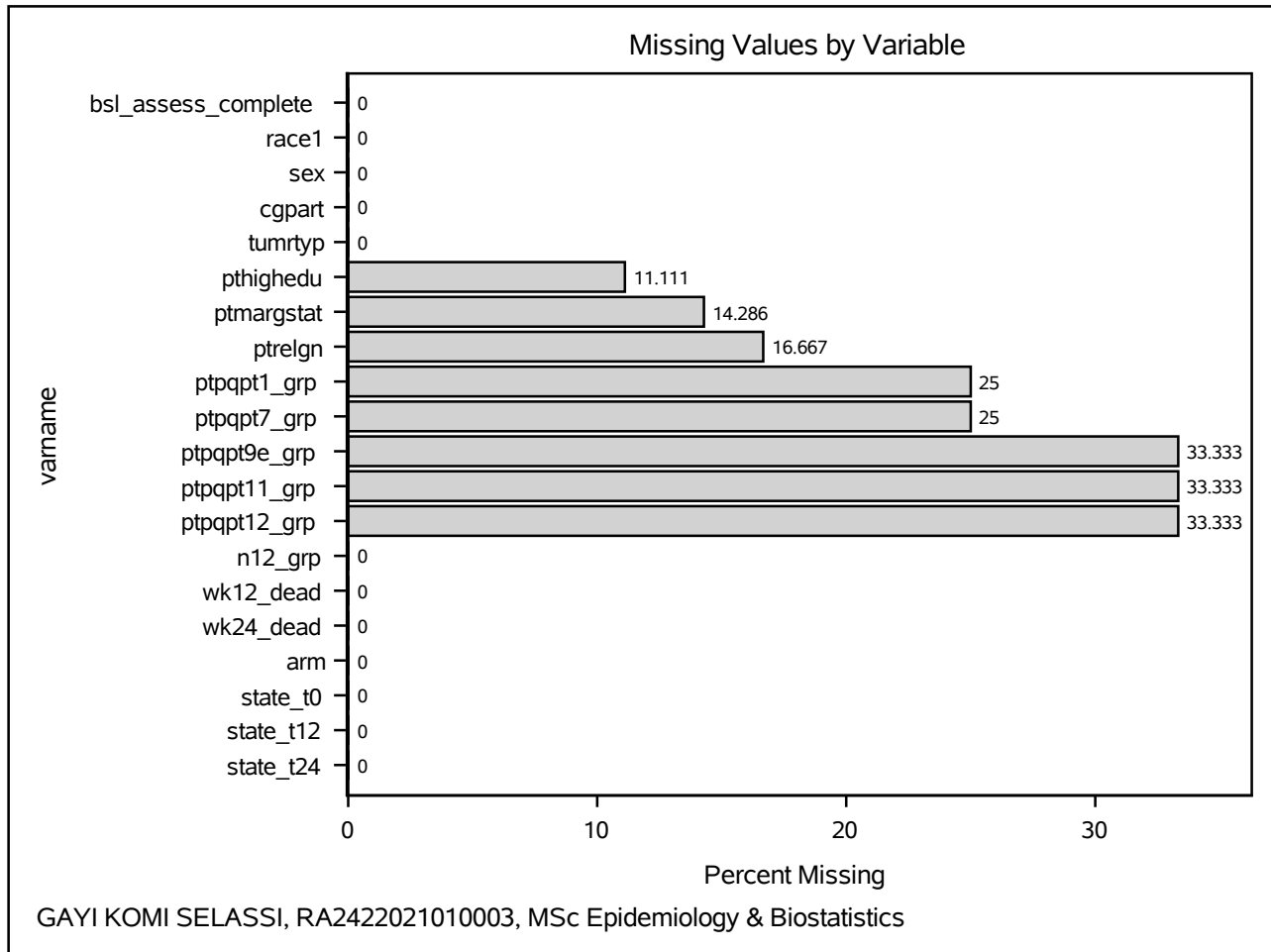


**Assignment 3 - Data Analysis**

Author : Gayi Komi Selassi

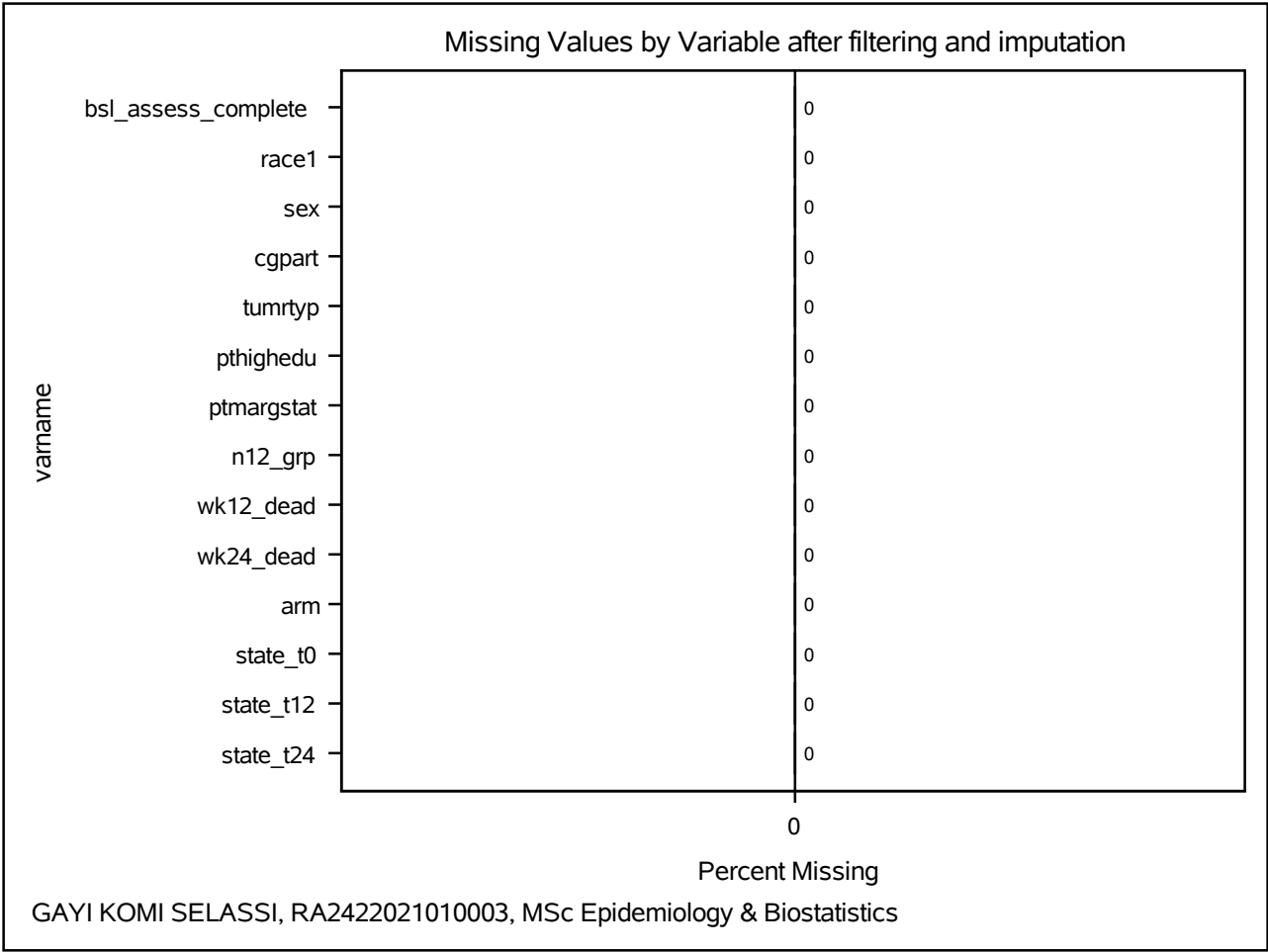
ID : RA2422021010003

Program : MSc Epidemiology &amp; Biostatistics

Portfolio: [Click Here](#)Access the GitHub Repository: [Click Here](#)**1. Visualization of missing value :**

Missing Values by Variable

2. Visualization of missing value after filtering and imputation:



**Missing Values by Variable after filtering and imputation****3. Information about the final dataset :****The CONTENTS Procedure**

Data Set Name	DATA_SP.DATA_MAIN_CLEAN_IMPUTED	Observations	352
Member Type	DATA	Variables	25
Engine	V9	Indexes	0
Created	05/09/2025 14:04:01	Observation Length	304
Last Modified	05/09/2025 14:04:01	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)		

**Engine/Host Dependent Information**

Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	430
Obs in First Data Page	352
Number of Data Set Repairs	0
Filename	/home/u64176007/sas_pratical_assignment/assignment_5_Summer_project/2.data_SP/data_main_clean_imputed.sas7bdat
Release Created	9.0401M8
Host Created	Linux
Inode Number	2293299073
Access Permission	rw-r--r--
Owner Name	u64176007
File Size	256KB
File Size (bytes)	262144

**Alphabetic List of Variables and Attributes**

#	Variable	Type	Len	Format	Informat
12	age	Num	8		
22	arm	Char	32		
2	bsl_assess_complete	Char	2	\$2.	\$2.
5	cgpart	Char	5	\$5.	\$5.
21	ecogps	Num	8		
13	fact_bl	Num	8		
14	fact_chg12	Num	8		
15	fact_chg24	Num	8		
16	hads_anx_chg12	Num	8		
17	hads_dep_chg12	Num	8		
18	hads_pt_anx_bl	Num	8		
19	hads_pt_dep_bl	Num	8		

**Missing Values by Variable after filtering and imputation****The CONTENTS Procedure**

<i>Alphabetic List of Variables and Attributes</i>					
#	Variable	Type	Len	Format	Informat
20	n12	Num	8		
9	n12_grp	Char	18	\$18.	\$18.
1	patref	Num	8	BEST12.	BEST32.
7	pthighedu	Char	31	\$31.	\$31.
8	ptmargstat	Char	24	\$24.	\$24.
3	race1	Char	48	\$48.	\$48.
4	sex	Char	3	\$3.	\$3.
23	state_t0	Char	1		
24	state_t12	Char	1		
25	state_t24	Char	1		
6	tumrtyp	Char	44	\$44.	\$44.
10	wk12_dead	Char	2	\$2.	\$2.
11	wk24_dead	Char	2	\$2.	\$2.

**Missing Values by Variable after filtering and imputation****4. Check frequency table of categorical variable and confirm recoding and imputation:****The FREQ Procedure**

<i>bsl_assess_complete</i>	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	352	100.00	352	100.00

<i>race1</i>	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Asian	14	3.98	14	3.98
Black or African American	42	11.93	56	15.91
Other	22	6.25	78	22.16
White	274	77.84	352	100.00

<i>sex</i>	Frequency	Percent	Cumulative Frequency	Cumulative Percent
f	156	44.32	156	44.32
m	196	55.68	352	100.00

<i>cgpart</i>	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	104	29.55	104	29.55
Yes	248	70.45	352	100.00

<i>tumrtyp</i>	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Esophageal/GEJ/Gastric	28	7.95	28	7.95
Hepatic/Biliary/Pancreatic/Unknown primary	111	31.53	139	39.49
Lung	213	60.51	352	100.00

<i>pthighedu</i>	Frequency	Percent	Cumulative Frequency	Cumulative Percent
8th grade or less	10	2.84	10	2.84
9-11th grade	24	6.82	34	9.66
Advanced degree	56	15.91	90	25.57
Associate degree/some college	59	16.76	149	42.33
Bachelor's degree	62	17.61	211	59.94
High school graduate/GED	116	32.95	327	92.90
Other	5	1.42	332	94.32
Vocational/technical school	20	5.68	352	100.00

<i>ptmargstat</i>	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Divorced	49	13.92	49	13.92
Domestic partnership	7	1.99	56	15.91
Married	215	61.08	271	76.99
Never married	31	8.81	302	85.80

**Missing Values by Variable after filtering and imputation****The FREQ Procedure**

<i>ptmargstat</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>Separated</i>	8	2.27	310	88.07
<i>Widowed</i>	42	11.93	352	100.00

<i>n12_grp</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>0 visits</i>	154	43.75	154	43.75
<i>1-2 visits</i>	83	23.58	237	67.33
<i>3-4 visits</i>	78	22.16	315	89.49
<i>5-6 visits</i>	28	7.95	343	97.44
<i>7 or more visits</i>	9	2.56	352	100.00

<i>wk12_dead</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>0</i>	294	83.52	294	83.52
<i>1</i>	58	16.48	352	100.00

<i>wk24_dead</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>0</i>	242	68.75	242	68.75
<i>1</i>	110	31.25	352	100.00

<i>arm</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>1</i>	169	48.01	169	48.01
<i>2</i>	183	51.99	352	100.00

<i>state_t0</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>A</i>	352	100.00	352	100.00

<i>state_t12</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>A</i>	192	54.55	192	54.55
<i>D</i>	58	16.48	250	71.02
<i>F</i>	102	28.98	352	100.00

<i>state_t24</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>A</i>	148	42.05	148	42.05
<i>D</i>	110	31.25	258	73.30
<i>F</i>	94	26.70	352	100.00