Q1: Import the Dataset 'Test_DataCore.csv' into your preferred statistical analysis program and include the code you used to import the dataset in your submitted code. (No *.doc* answer necessary for this question, only the code)

```
*Q1 import dataset;
options msglevel=I;
FILENAME REFFILE '/home/u39221714/sasuser.v94/Test_DataCore.csv';

PROC IMPORT DATAFILE=REFFILE
        DBMS=CSV
        OUT=core;
        GETNAMES=YES;
RUN;
```

Q2: How many patients are in the dataset?

80.

Q3: How many of the patients died?

4.

Q4: Vital statistics recently released the data on deaths in 2015. For H+H patients, these are available in the file 'Test_DataCore_VitalStats.csv'. Merge the missing dates of death received from vital statistics 'vitalstats' for the patients into the previous dataset. How many patients died in total?

9.

Q5: When was the last visit date for the patients who have died? Create a table which lists number of days from last visit date to date of death. Paste the output table in your answer.

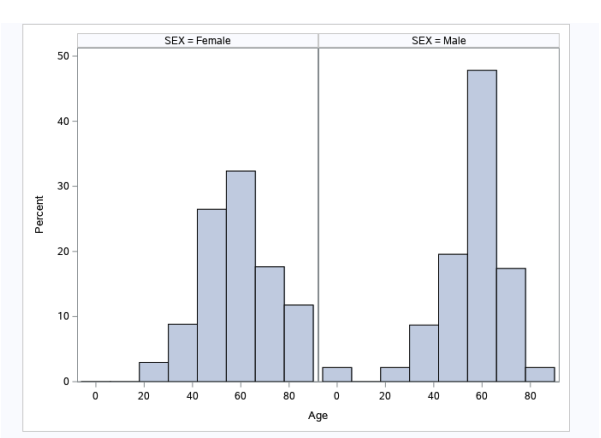
PATIENT_ID	length_of_days	COUNT
819392	1	1
15150240	18	1
28438144	11	1
28653184	15	1
30041536	9	1
30172352	3	1
44109184	51	1
53420640	47	1
63383264	160	1

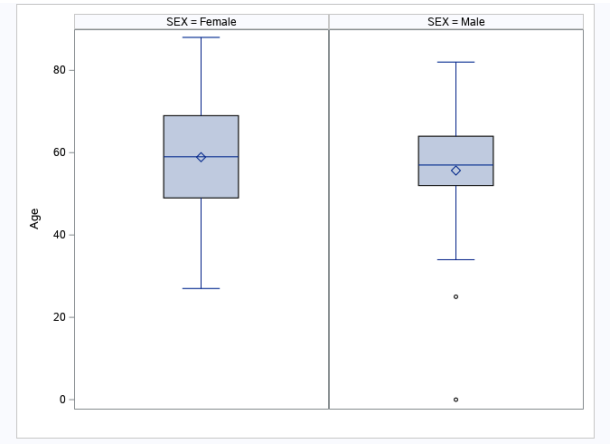
Q6: Create a table with the race distribution of patients in the dataset. Paste the output table in your answer.

The FREQ Procedure

RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
American	85	0.85	85	0.85
Black or	6181	61.81	6266	62.66
Hispanic	3295	32.95	9561	95.61
Other	261	2.61	9822	98.22
Unknown	13	0.13	9835	98.35
White	165	1.65	10000	100.00

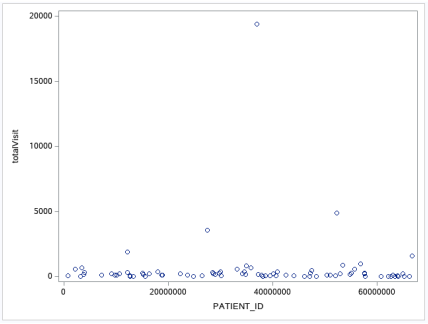
Q7: Create a visualization with a software/tool of your choice for age and age by gender distribution of patients in the dataset. Paste the visualizations in your answer. Also, briefly describe the distributions.

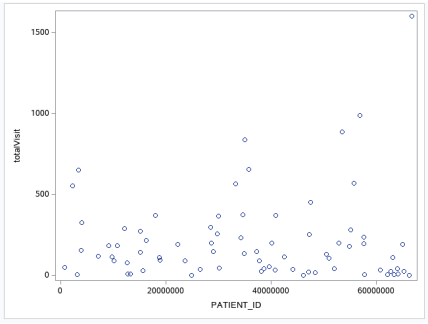


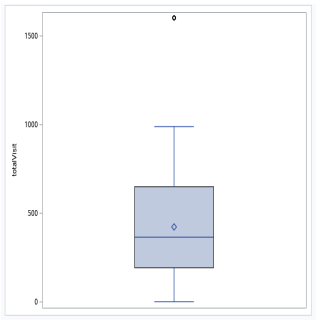


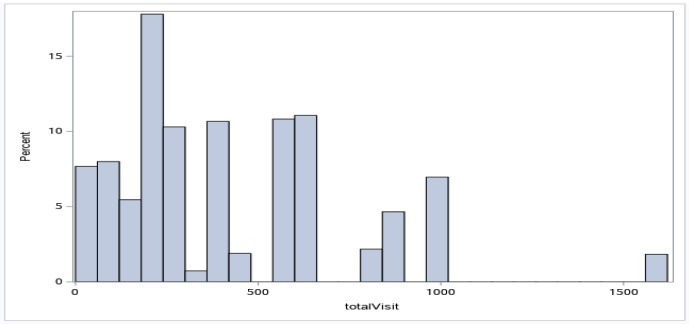
The age distribution seems to be normally distributed/slightly left skewed in females but not in the male subgroup. Female subjects had a wider range of age than males. Male subjects were mainly in their 60s, where there were outliers where age = 0. They should be deleted in future analysis.

Q8: Calculate total days spent at the hospital for each patient, visualize (with a software/tool of your choice), and describe the distribution. Paste the visualizations in your answer.









The total days spent at the hospital for each patient varied from 0 day to up to almost 20,000 days. I excluded over 5 years (1825 days) of hospital stays (long term care) and generated the last 3 graphs. Most patients spent around 300 days in hospitalization.

Q9: A) Investigate inpatient visits (patients who stayed for more than a day at the hospital) for each patient. Output a list of the top 10 patient_ids who spent maximum days in the hospital in the past year (2015). Paste the output table in your answer.

B) How many inpatient and outpatient visits did the previous list of patients have in 2015? Paste the output table in your answer.