Data Exploration – Project 2

In the data exploration process, some data caused concern upon examination.

* HOSPCD
  + No concerns
* Proced
  + Procedure 2??? This procedure is not mentioned in the research question.
  + RESOLVED: Spoke to investigators -> delete these observations
* ASA - 664 missing values (2.5%)
* WEIGHT -
  + Low value - 26.73, some weights in the 30s and 40s
    - This is not likely for a weight in lbs,
    - Check to see if a hospital recorded in kgs
      * Seems that sixmonth = 39, hospitals started recording weights in kgs.
      * Created sixmonth=39 dataset and converted kgs to lbs
      * NEED TO FIGURE OUT HOW TO REPLACE THESE VALUES INTO THE ORIGINAL DATASET
  + High value of 287.08 – ok
  + 702 missing values (2.65%)
* Height
  + No concerns
* BMI
  + Value of 2.438
    - Decimal error.
    - Change 2.4379 to 24.379994195 – RESOLVED
  + Value of 72.3
    - Entry error
    - Changed to calculated BMI value - RESOLVED
  + Value of 75.1
    - Entry error
    - Changed to calculated BMI value - RESOLVED
  + 702 missing values (2.65%)
* Albumin
  + 13,241 values missing (49.93%) missing
  + Death30
    - For those who did not die within 30 days of surgery (death30=0):
      * 50.11% were not missing albumin data
      * 49.89% were missing albumin data
    - For those who died within 30 days of surgery (death30 = 1):
      * 49.02% were not missing albumin data
      * 50.98% were missing albumin data
    - There does not seem to be difference in the distribution of missing albumin data among those who died within 30 days of surgery and those who did not. About 50% are not missing data and about 50% are missing albumin data.
  + Procedure
    - For those with missing a surgery code
      * 51.37% were not missing albumin data
      * 48.63% were missing albumin data
    - For those with valve surgery (proced = 0)
      * 50.11% were not missing albumin data
      * 49.89% were missing albumin data
    - For those with CABG surgery (proced = 1)
      * 50.03% were not missing albumin data
      * 49.97 % were missing albumin data
    - There does not seem to be a difference in the distribution of missing albumin data among those who had valve surgery and those who has CABG surgery. About 50% are not missing data and about 50% are missing albumin data.
  + ASA
    - For those missing an ASA code
      * 48.34% were not missing albumin data
      * 51.66% were missing albumin data
    - For those with an ASA code of 1
      * 47.62% were not missing albumin data
      * 52.30% were missing albumin data
    - For those with an ASA code of 2
      * 48.61% were not missing albumin data
      * 51.39% were missing albumin data
    - For those with an ASA code of 3
      * 50.09% were not missing albumin data
      * 49.91% were missing albumin data
    - For those with an ASA code of 4
      * 50.17% were not missing albumin data
      * 49.83% were missing albumin data
    - For those with an ASA code of 5
      * 65.45% were not missing albumin data (n = 36)
      * 34.55% were missing albumin data (n = 19)
    - There seems to be a difference in the distribution of missing data among those with an ASA code of 5. More people with an ASA score of 5 have albumin data than those with an ASA score of 5 that are missing albumin data. This could be due to the fact that these patients are sicker and needed more routine bloodwork than healthier patients. Among the other levels of ASA the distribution of missing albumin data seems to be 50/50.