Inan Khan
Assignment 10

1. The censoring of data works via filtering and censoring out data that is sensitive information. From what I understand the logic in this code is that is that whenever an analysis is done, not everyone will have an event in the system. The incomplete information is called censoring. If no one has had an event in N years then their time to when they will receive an event is greater than N years. Also the errors in the data are related to date, because subjects can age during the time it take to report the data.

2. In the data after analysis I would say that the variability in date results is the biggest factor in error. For example if we have something like 3 or four years of survival then you would also need to follow the subjects for an equivalent amount of years. Since the subjects that are being followed would age during that time that means that the treatment may be outdated during that entire time. It becomes a conflict between reporting ahead of time and doing long term studies

pair (ti , δi)
t= endpoint/ time, or previous follow up

δ is variable 0/1
The 0= subject censored (t)
1= subject event (t)

The code for R is survive(time,status)

And vtype is 'death' since vtype is a variable in the data