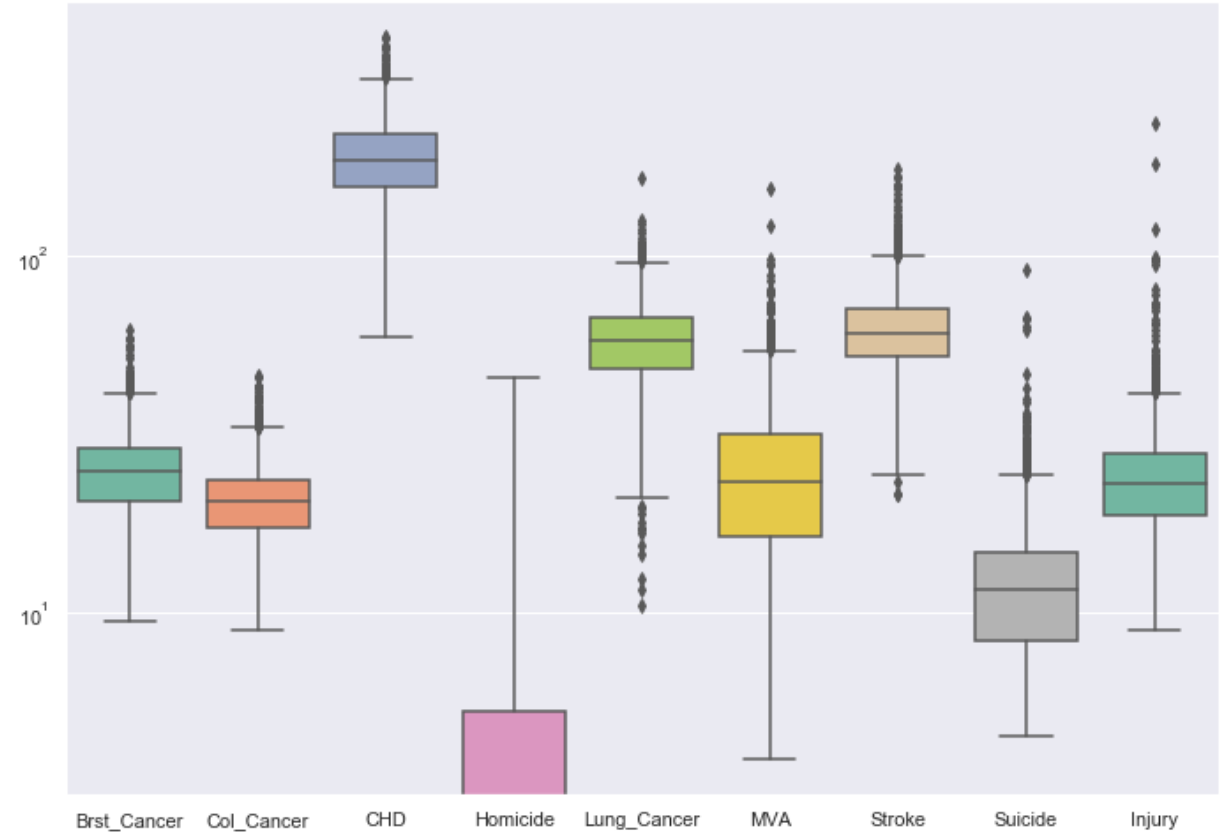
Granularity: Every record in the dataframe is record of one county in the US

In [13]:
```
ax = sns.boxplot(data=df_mbd[['Brst_Cancer','Col_Cancer','CHD','Homicide','Lung_C
ax.set_yscale('log')
```




```
Leading Causes of Death in order is: CHD(Cardiovascular Heart Disease), Lung
Cancer, Stroke, MVA(motor vehicle accidents), Breast Cancer, Injury, Colon
Cancer, Suicide, and then at last Homicide this is correlated to the data over
the states and deaths and negatively related to births
```

Data Cleaning: The dataset is huge and I picked measures of birth and death in the united states to study. I chose to use average columns given to do the visulization Therefore, study between different races, ages, and different causes were chosen as primary focus to create measure of births and death.

Conclusion: We see from the boxplot of leading causes of deaths in the United States is CHD. Also, looking at the birth data, women under age of 18 has higher chances of getting pregnant than women who are over 40. Age<18 for white women has more chances of pregnancy.

Hardest Part of the Project: The hardest part of this project was coming from no background in data analysis. It was challenging since this is something unique and doesn't focus more on development side but rather developing models to study data. Thanks to my teammates who helped me with every aspect.