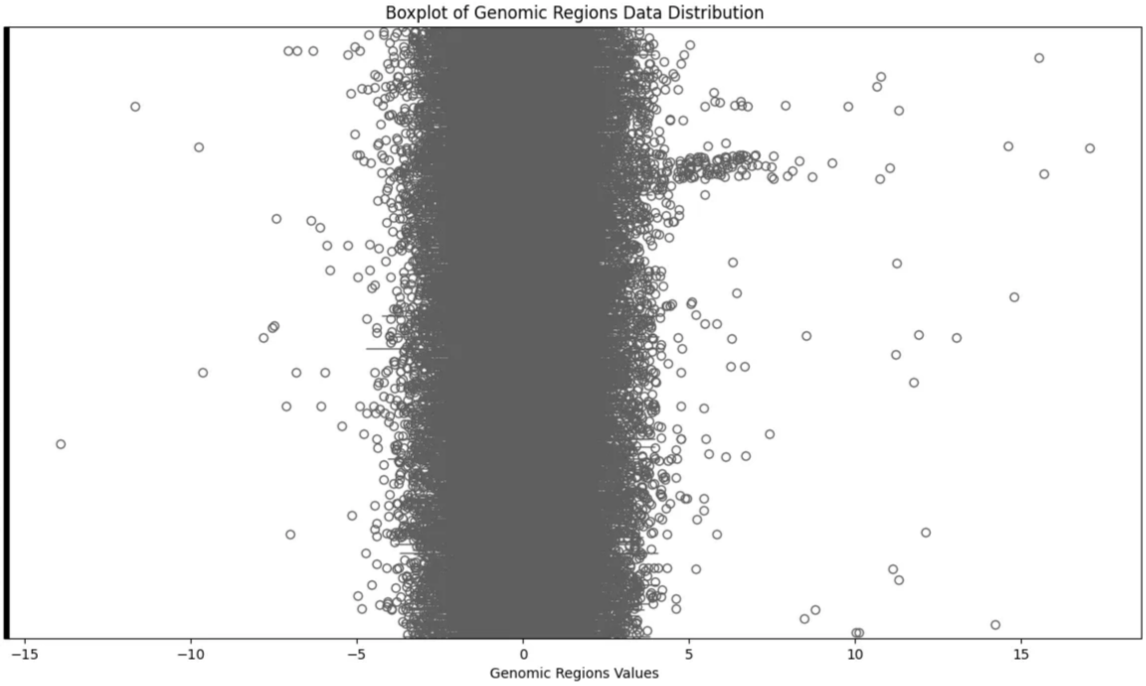
BoxPlot of dataset

Assignee	Ⓢ Sun Savannah
Status	Done
Due	@2024年5月26日
Project	🩸 BDD-Cancer detection
Priority	Medium
Tags	Data Analysis Visualization

Visualization of the dataset

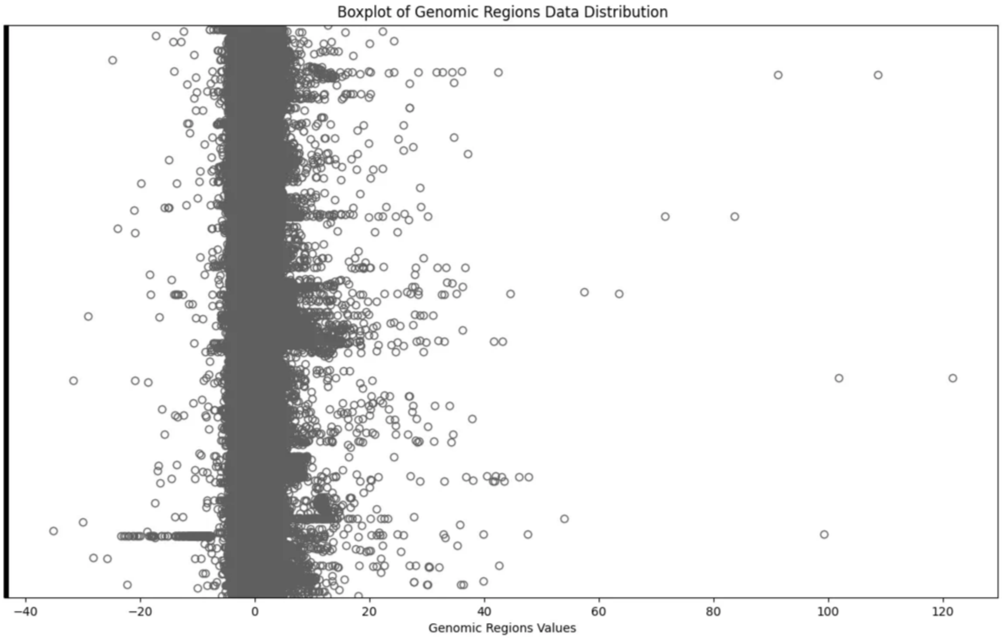
- X - concentration of CG after normalization
- Y - features (different part on chromosome)
- Figure 1: Boxplot of all the samples that are NOT having cancer (Negative samples)



There are less extreme value in this class.

Max: about 15, Min: about -14

- Figure 2: Boxplot of all the samples that are having cancer (Positive samples)



A lot of extreme values from -40 to 120

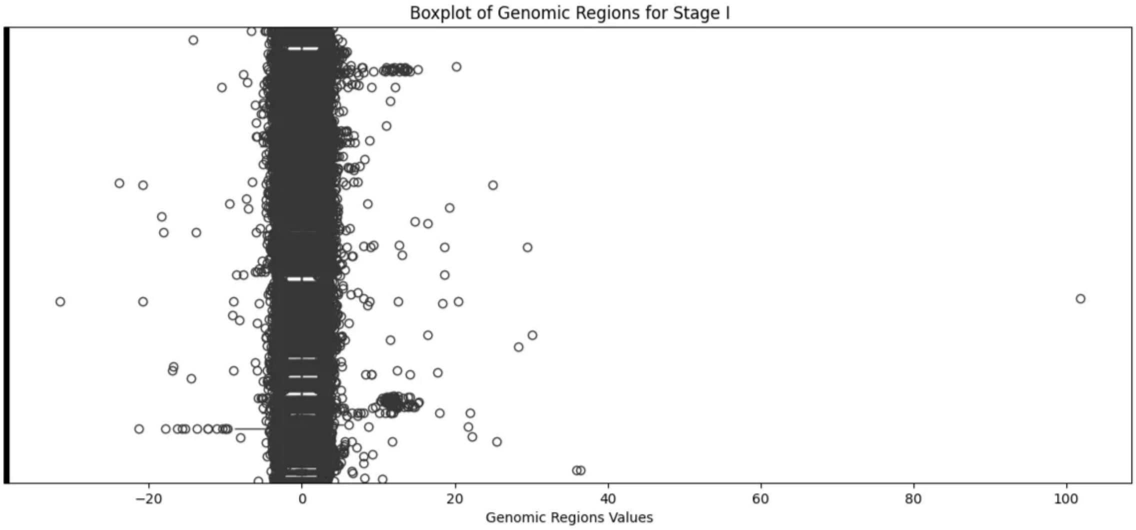
- Question: What is the meaning of extreme data?
- Significant Features or Bad Samples

Stage Analysis

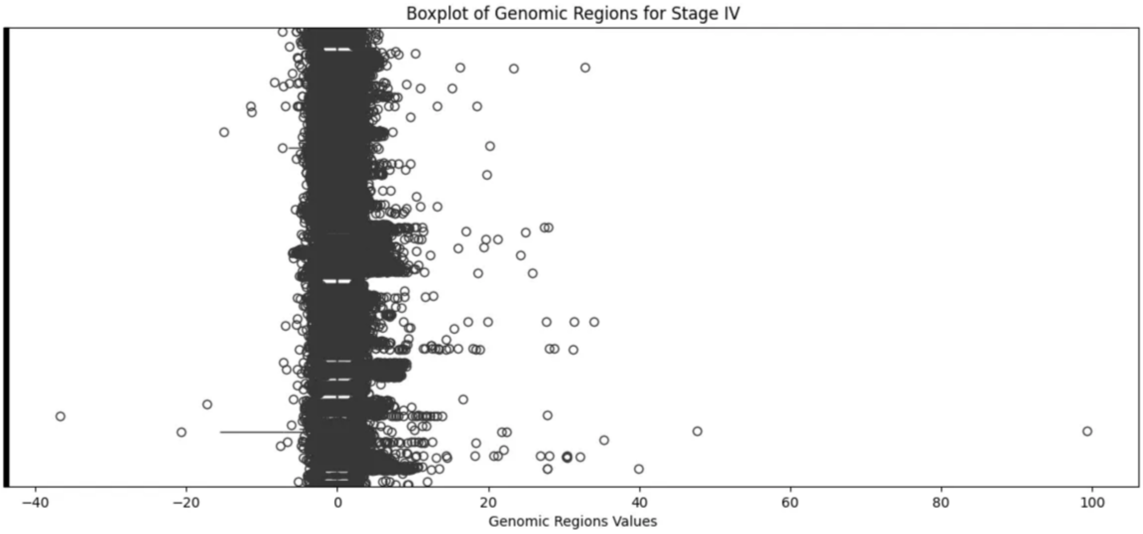
```
Stage
Normal                                642
IIB                                   262
IIA                                   123
IV                                    102
IIIC                                   75
IIIA                                   65
IIIB                                   64
IA                                     60
```

```
IB                                   55
III                                  41
I                                    34
II                                   25
IVA                                  19
IA2                                   9
IA3                                   9
IVB                                   8
IIC                                   7
IC                                    6
Not given                             4
IIIA1                                 3
IIIA2                                 2
0                                     2
0___TisN0M0                           2
biopsy only at time of blood draw     2
NaN                                   1
pT3N2Mx                               1
IA1                                   1
IIIV                                  1
IIIIA                                 1
0__TisN0M0                            1
0__Tis(2)N0M0                         1
```

- Stage I



- Stage IV



Stage IV CG’s concentration is much larger than stage I.