Assignment 2

Ceren Yıldız

2025-10-13

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

Dataset

The `BreastCancer` dataset contains cytological measurements of breast tissue samples, used to classify tumors as benign or malignant. It includes numerical features describing cell characteristics derived from microscopic examination.

To load the dataset to your environment, just run the code below:

```
install.packages("mlbench")
library(mlbench)
data("BreastCancer")
BreastCancer$Cl.thickness <- as.numeric(BreastCancer$Cl.thickness)
```

Drawing a plot for distribution

1. Please draw a plot to visualize the distribution of tissue thickness (`Cl.thickness`) of the patients according to `Class` indicating whether the tumor in the breast tissue is benign or malignant. (40 pts).

```
ggplot(BreastCancer,aes(x=Cl.thickness))+geom_histogram( )
```

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```
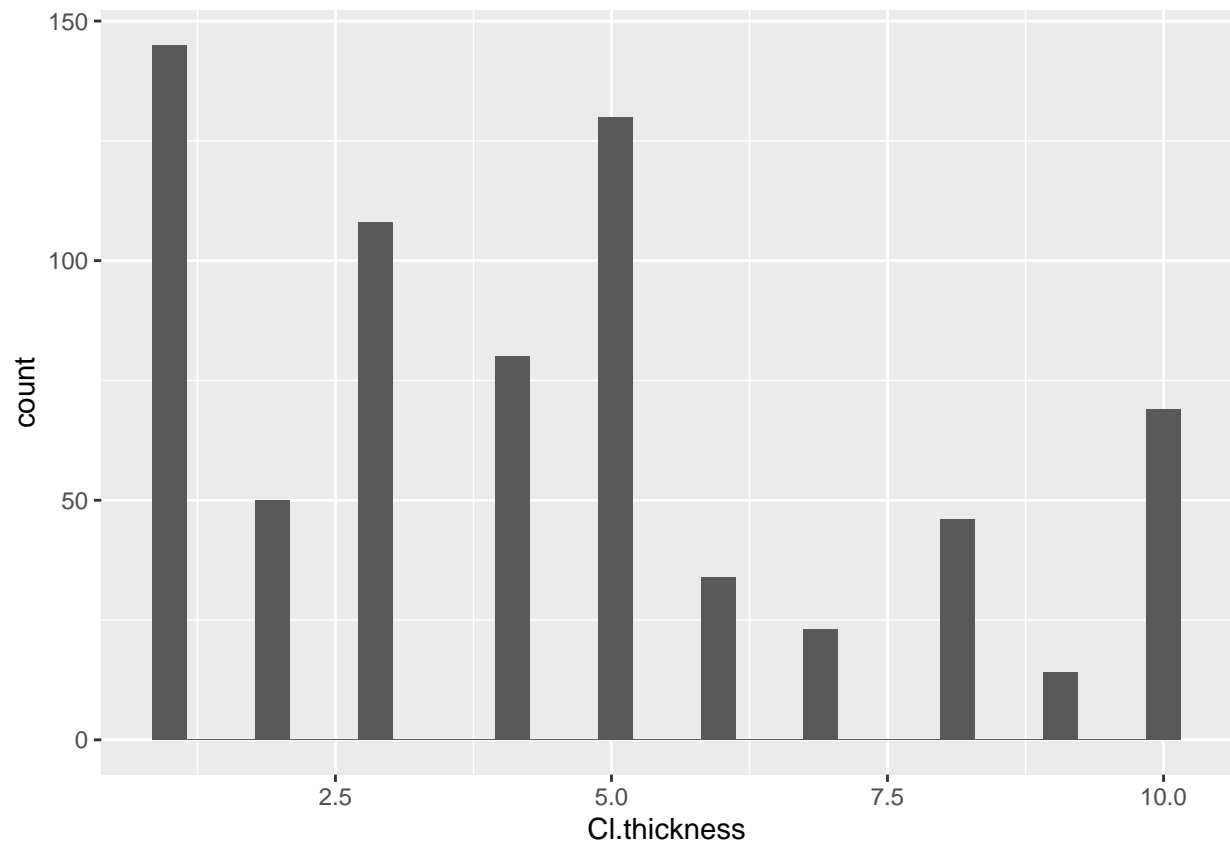
```
ggplot(BreastCancer,aes(x=Cl.thickness))+geom_boxplot()
```
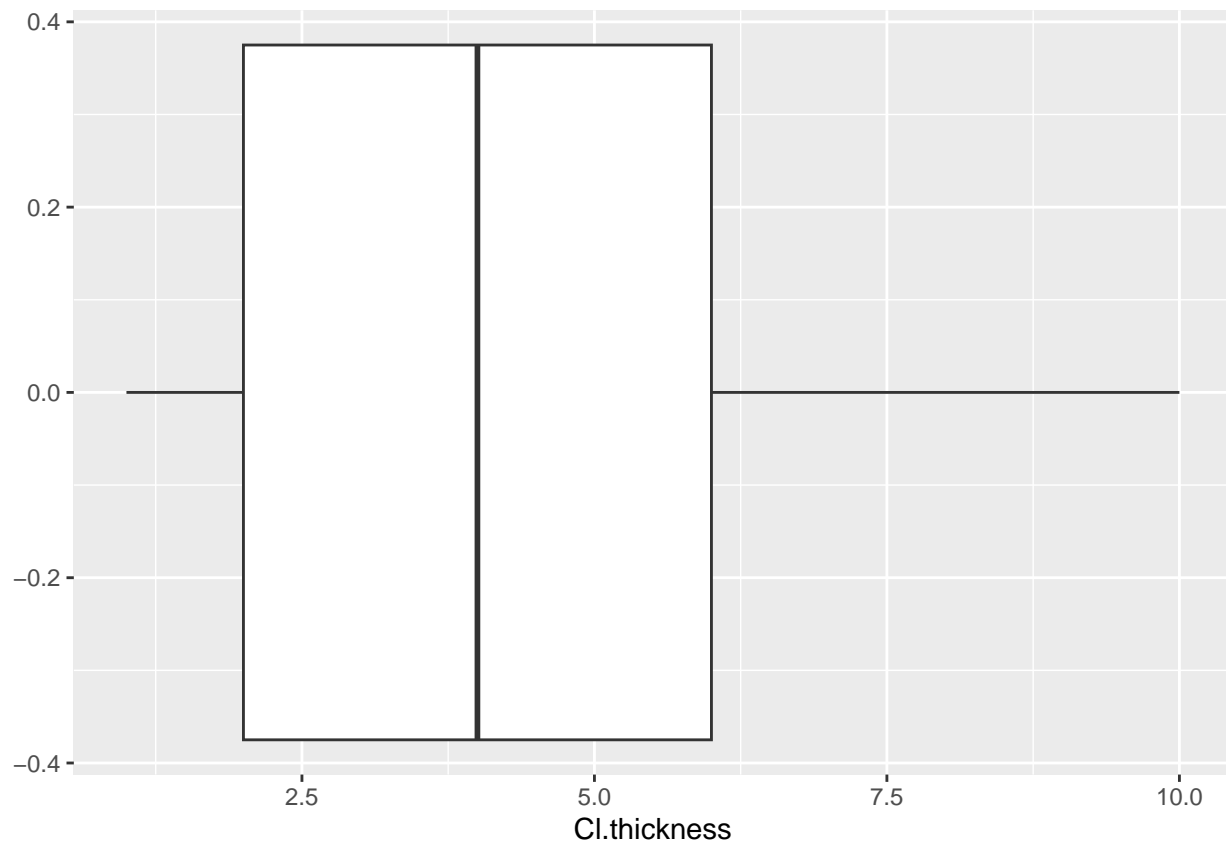
Cl.thickness

```
# Both graph are true. I think boxplot is good.
```

2. Please solve the problems in the plot and make it better by adding some information (30 pts). # I draw histogram.Maybe i can change binwidth = 5 but it is not good idea.

```
ggplot(BreastCancer,aes(x = Cl.thickness))+geom_histogram(binwidth = 0.5,color = "white", fill = "black"
```

Cl.thickness Distribution of BreastCanser
Histogram
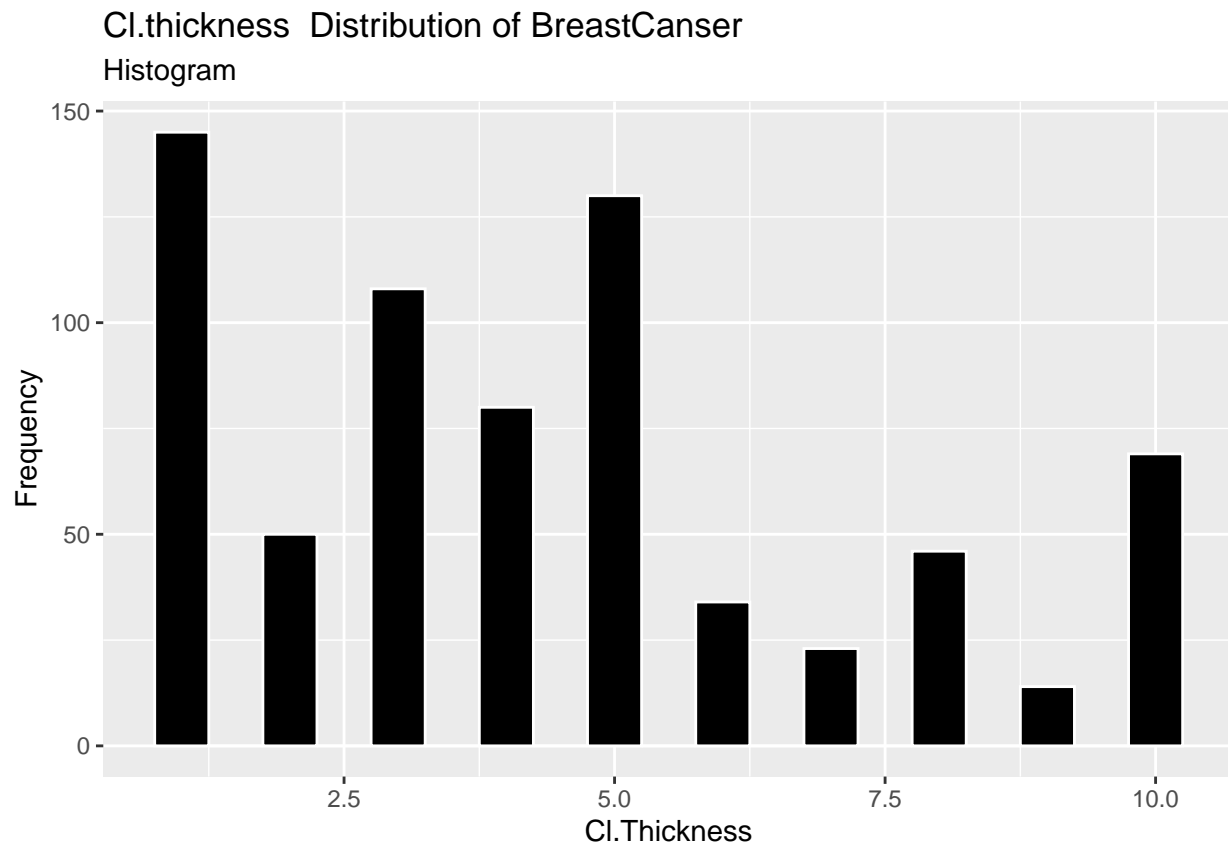
3. Interpret the plot (30 pts).

BreastCanser data has 11 variables,699 observation.One of them is Cl.thickness.We want to classify tumors as benign or malignant so you can use some graph.Maybe we can use histogram.Also we can use boxplot for distribution.You know that we find Q1,Q3 and median so we understand left skewed,right skewed.This graph is left skewed data.Also, $Q1 < 2.5$, $Q3 > 5$

. . .