

Assignment 2

Ramazan Karagöz

2025-10-13

```
# Installing necessary package(s)
install.packages("ggplot2")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
library(ggplot2)
```

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:

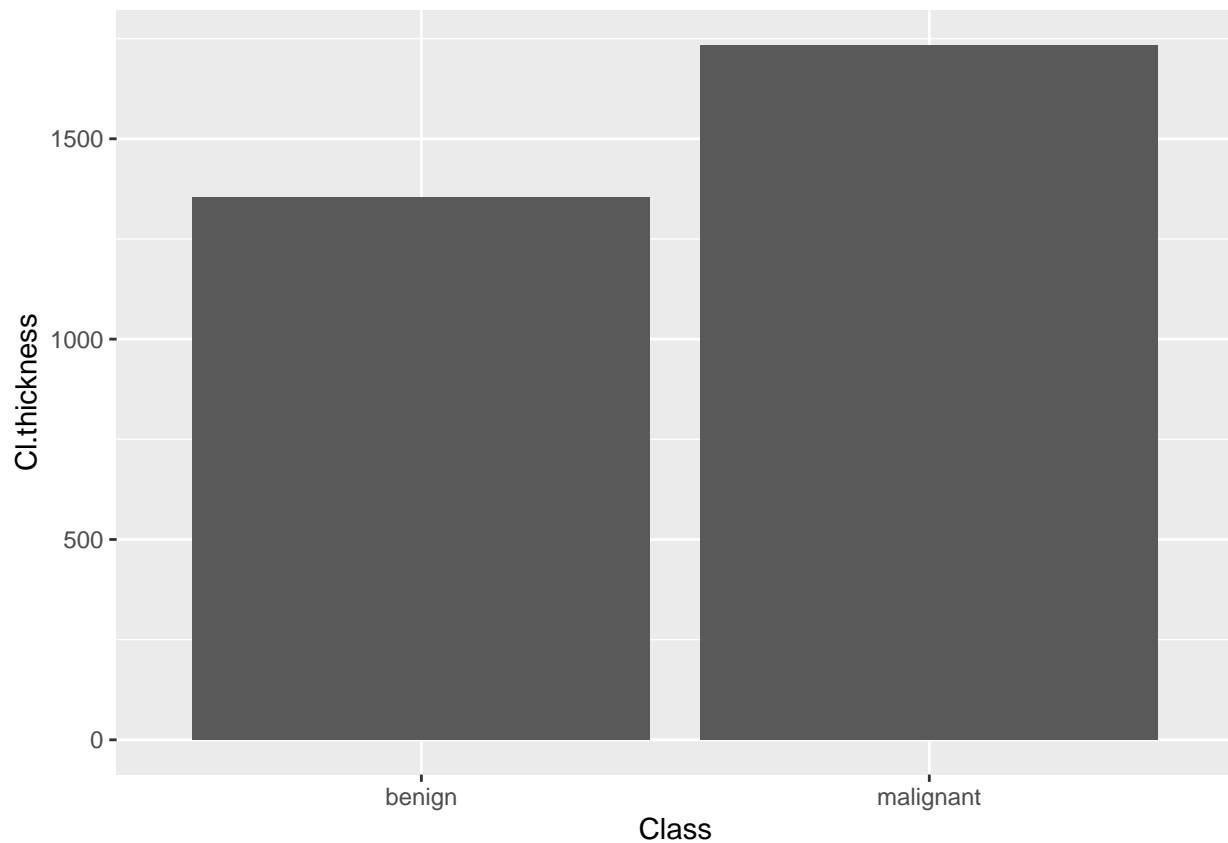
```
install.packages("mlbench")
library(mlbench)
data("BreastCancer")
BreastCancer$Cl.thickness <- as.numeric(BreastCancer$Cl.thickness)
```

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).

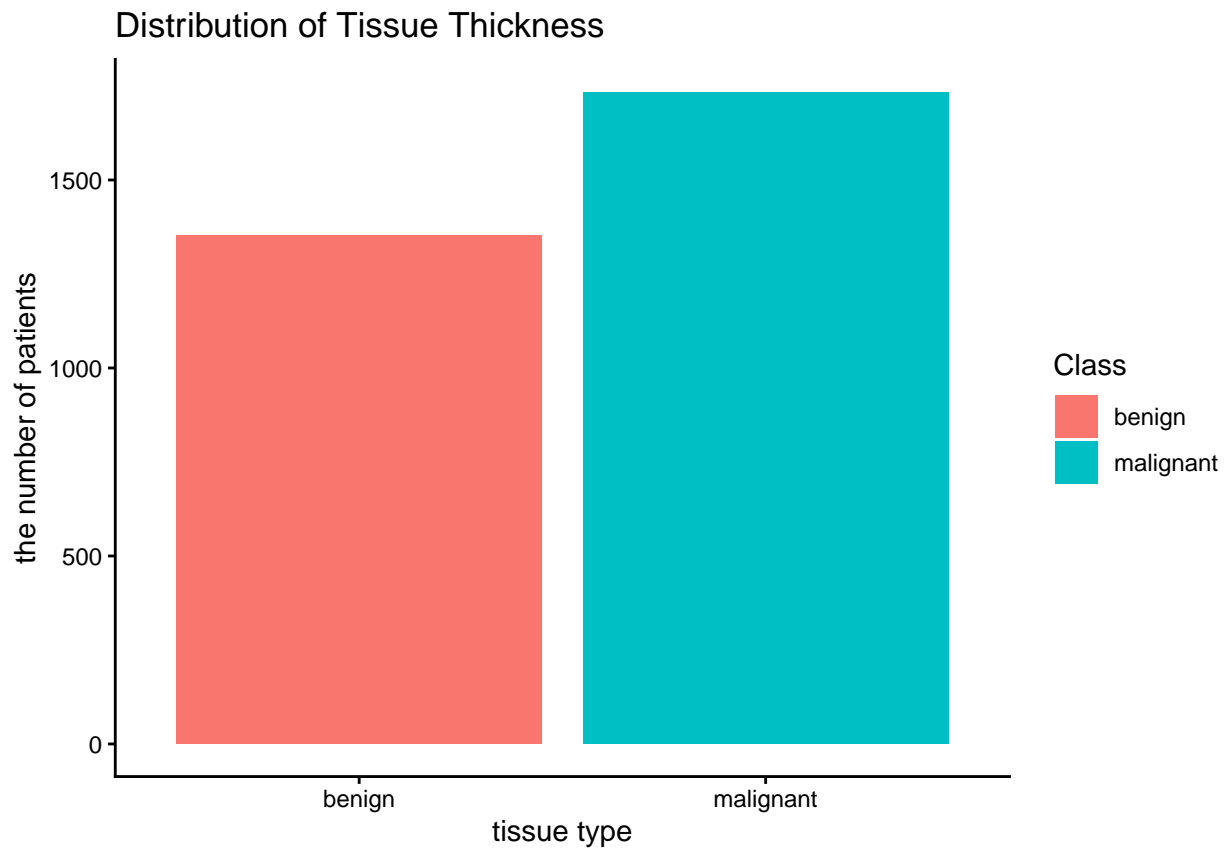
```
ggplot(data=BreastCancer, aes(x =Class,y=Cl.thickness)) +
  geom_histogram(stat="identity")
```

```
## Warning in geom_histogram(stat = "identity"): Ignoring unknown parameters:
## `binwidth` and `bins`
```



2. Please solve the problems in the plot and make it better by adding some information (30 pts).

```
ggplot(data=BreastCancer, aes(x =Class,y=Cl.thickness,fill=Class)) +
  geom_histogram(binwidth="",stat="identity")+
  labs(
    title="Distribution of Tissue Thickness",
    x="tissue type",
    y="the number of patients"
  )+
  theme_classic()
```



3. Interpret the plot (30 pts).

Malignant type tumors are found in greater numbers than benign type tumors ...