# Assignment 2

## Tuan Doğukan TEKEŞ

## 2025-10-13

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

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

```r
install.packages("dplyr")
```

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

```r
install.packages("ggplot2")
```

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

```r
library(DALEX)
```

```
## Welcome to DALEX (version: 2.5.2).
## Find examples and detailed introduction at: http://ema.drwhy.ai/
## Additional features will be available after installation of: ggpubr.
## Use 'install_dependencies()' to get all suggested dependencies
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:DALEX':
##
##     explain

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

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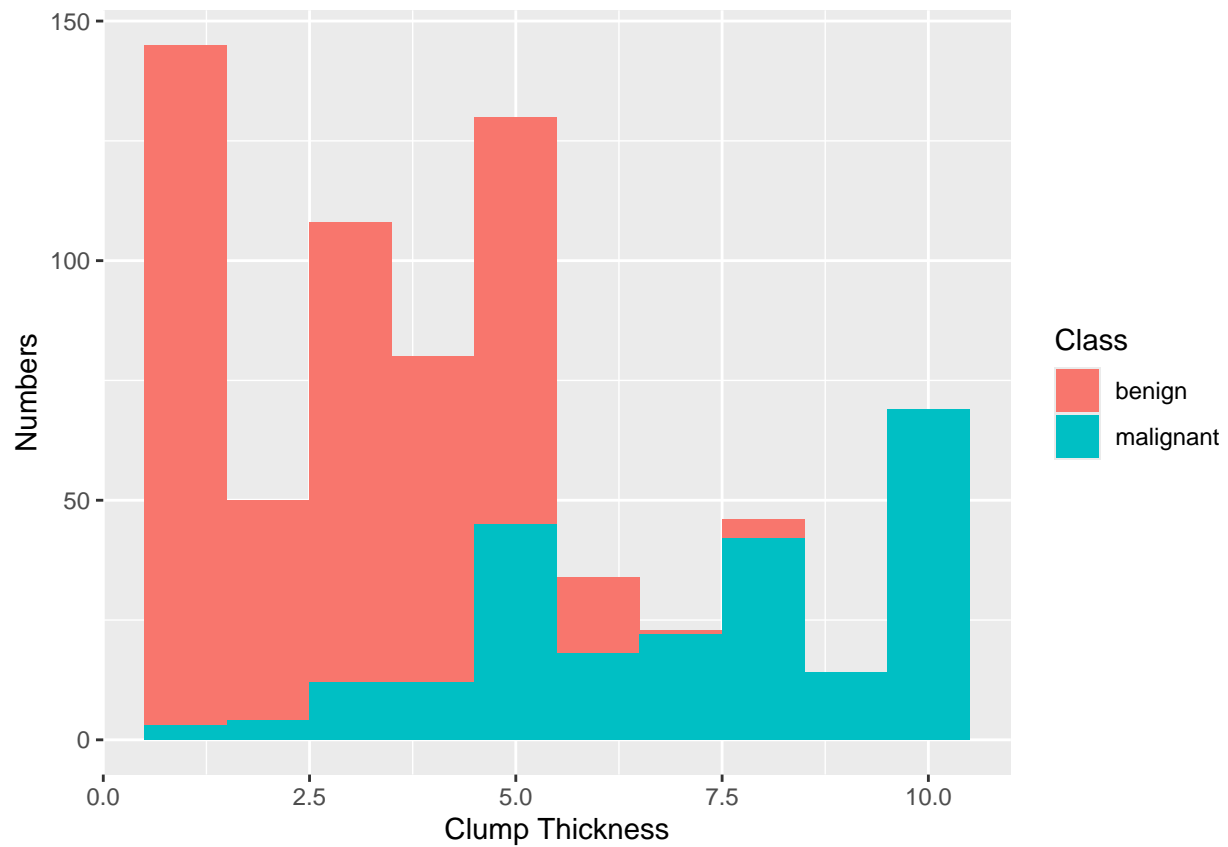
```
ggplot(BreastCancer, aes(x = Cl.thickness , fill = Class))+
geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value `binwidth`.
```
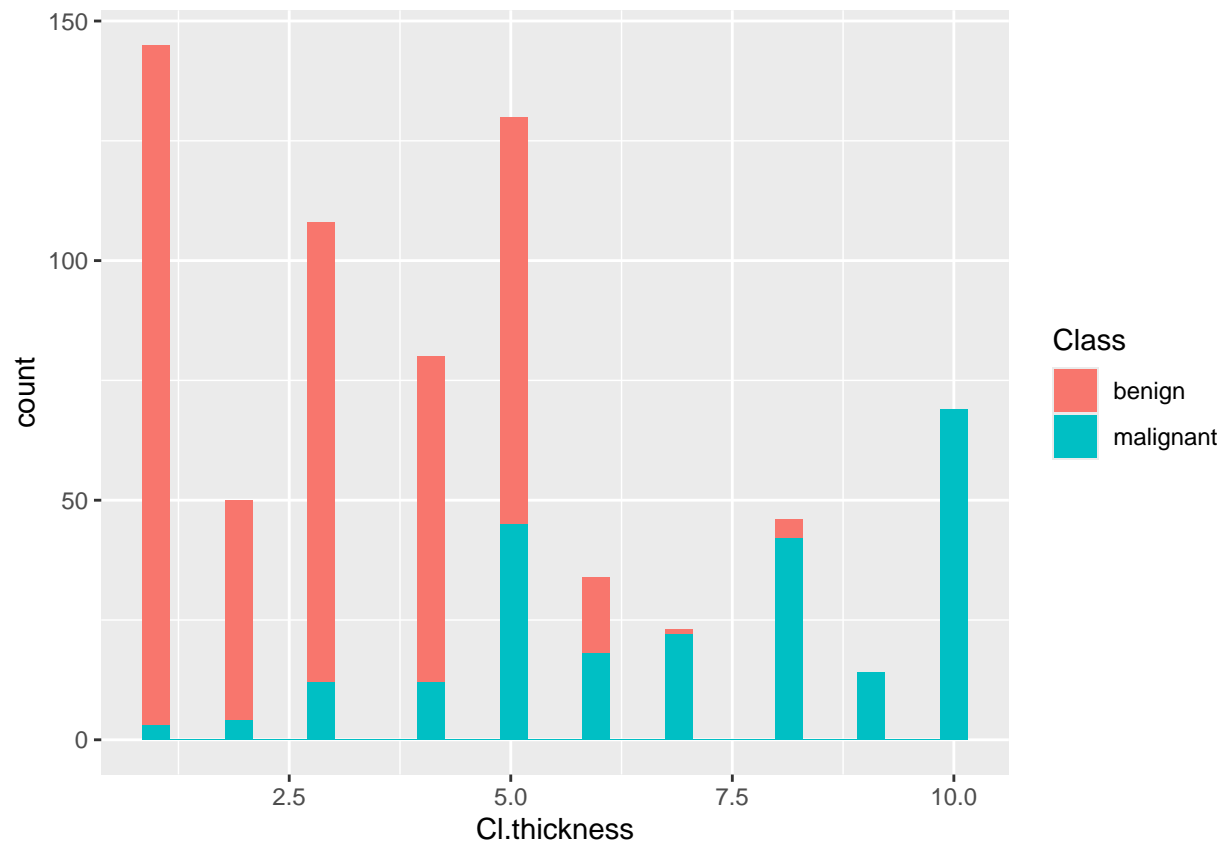


2. Please solve the problems in the plot and make it better by adding some information (30 pts).

```
ggplot(BreastCancer, aes(x = Cl.thickness, fill = Class))+
geom_histogram(binwidth = 1) +
  labs(x = "Clump Thickness" ,
       y = "Numbers")
```

```
ggplot(BreastCancer, aes(x = Cl.thickness, fill = Class)) +
  geom_density(alpha = 0.5) +
 labs(x = "Clump Thickness",
      y = "DNumbers",
      title = "Clump Thickness By Class",
      subtitle = "Kernel Density Estimation",
      fill = "Class") +
  scale_fill_discrete(labels = c("Benign", "Malignant"))
```

Clump Thickness By Class
Kernel Density Estimation
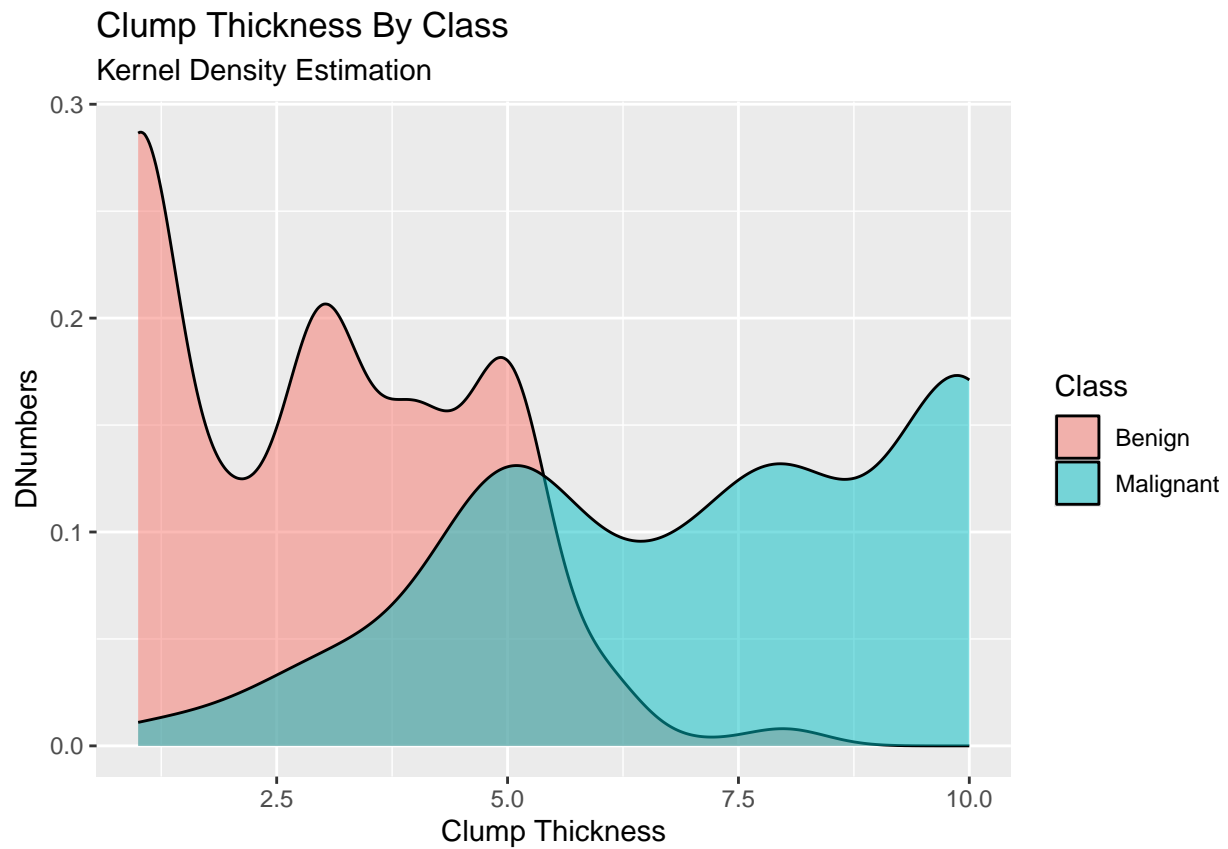
3. Interpret the plot (30 pts). The `BreastCancer` dataset contains cytological measurements of breast tissue samples, used to classify tumors as benign or malignant. In the plot there seems to be more benign cases in lower Clump Thickness ranges. More malignant cases in upper Clump Thickness. Malignant cases show a wider spread.

. . .