Assignment 2

Can Binay

2025-10-13

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.

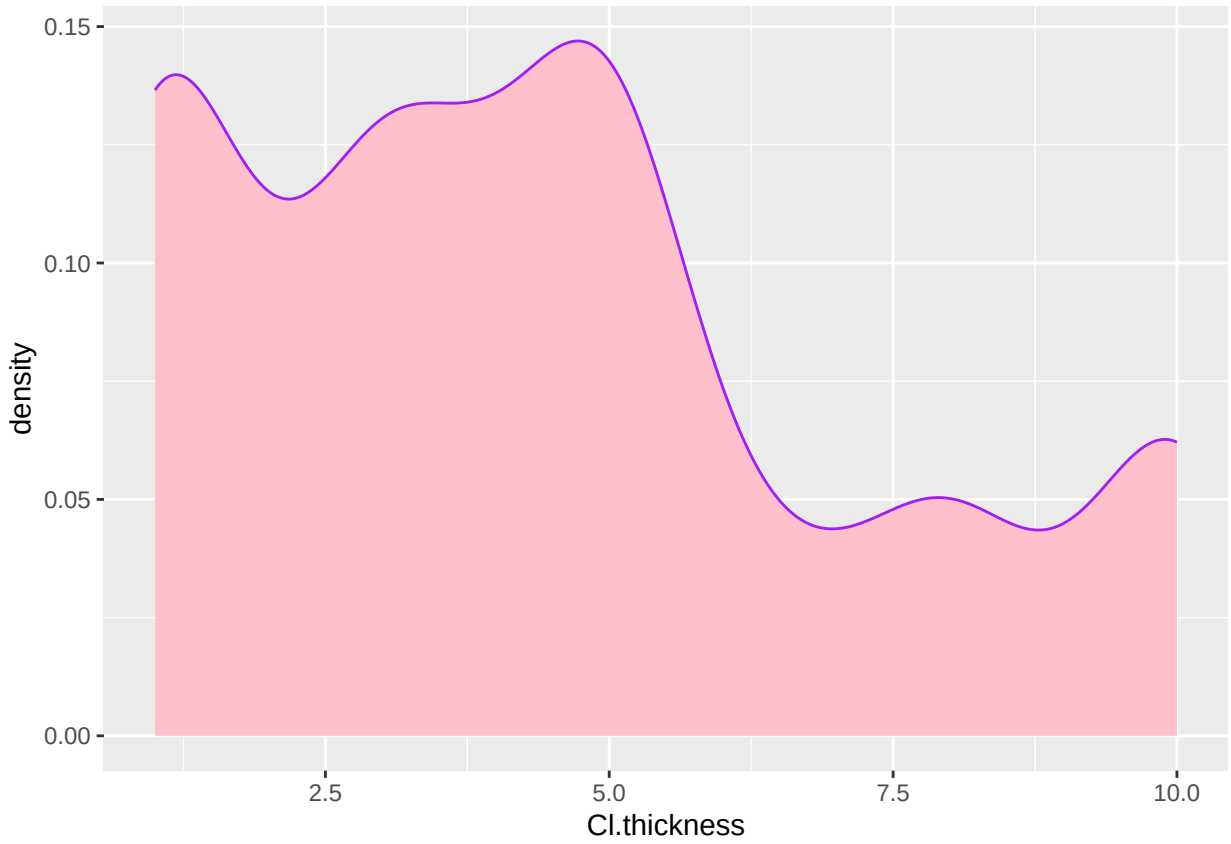
To load the dataset to your environment, just run the code below:

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

Drawing a plot for distribution

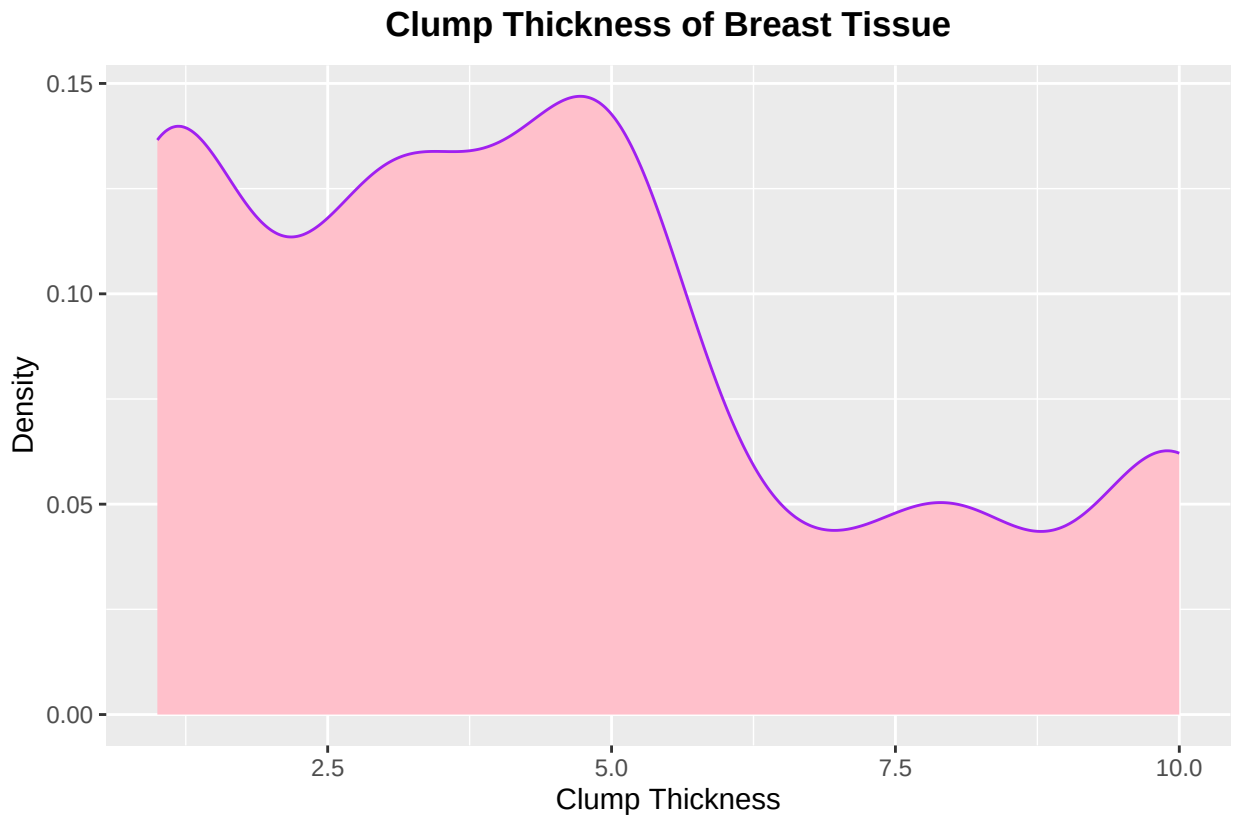
1. Please draw a plot to visualize the distribution of tissue thickness (`Cl.thickness`) of the patients according to `Class` indicating whether the tumor in the breast tissue is benign or malignant. (40 pts).

```
ggplot(BreastCancer, aes(x=Cl.thickness)) +
 geom_density(fill="pink", color="purple", alpha=1)
```



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

```
 ggplot(BreastCancer, aes(x=Cl.thickness)) +
  geom_density(fill="pink", color="purple", alpha=1) +
  labs(x = "Clump Thickness",
     y = "Density",
     title = "Clump Thickness of Breast Tissue",
     caption= "This breast cancer databases was obtained from the University of Wisconsin Hospitals, M
  theme(
  plot.title = element_text(hjust = 0.5, vjust = 2, face = "bold"),
  axis.title.x = element_text(hjust = 0.5),
  axis.title.y = element_text(hjust = 0.5)
)
```



ıst cancer databases was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg.

3. Interpret the plot (30 pts).

Interpretation

According to this data set, the size of the breast tumor in one of every two women was 3 times larger than the size of the tumor in the breast of another woman.