P8106 Final Codes

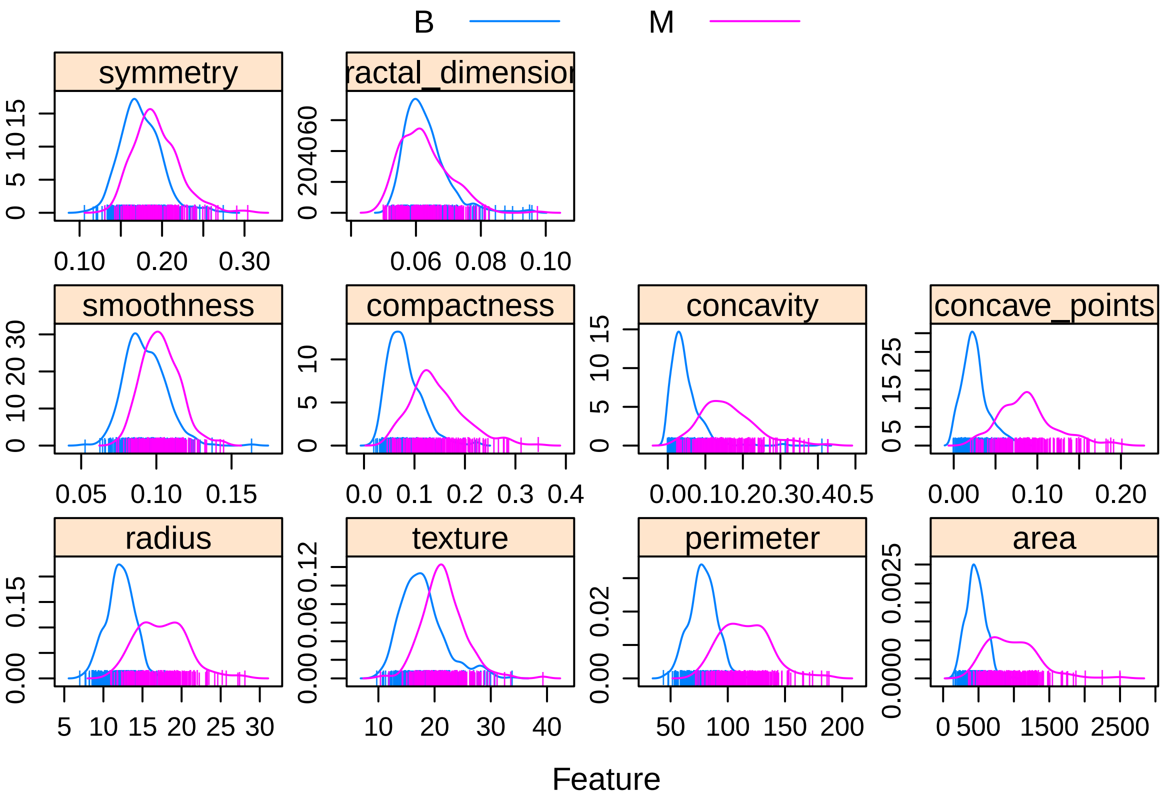
Yuxuan Chen | Yuan Meng | Paula Wu

```r
bc_df = read_csv("./data/breast-cancer.csv") %>%
  dplyr::select(-c(1, 33)) %>%
  janitor::clean_names() %>%
  # add extra row
  add_row(diagnosis = 'B', radius_mean = 7.76, texture_mean = 24.54,
          perimeter_mean = 47.92, area_mean = 181, smoothness_mean = 0.05263,
          compactness_mean = 0.04362, concavity_mean = 0,
          concave_points_mean = 0, symmetry_mean = 0.1587,
          fractal_dimension_mean = 0.05884, radius_se = 0.3857,
          texture_se = 1.428, perimeter_se = 2.548, area_se = 19.15,
          smoothness_se = 0.007189, compactness_se = 0.00466, concavity_se = 0,
          concave_points_se = 0, symmetry_se = 0.02676,
          fractal_dimension_se = 0.002783, radius_worst = 9.456,
          texture_worst = 30.37, perimeter_worst = 59.16, area_worst = 268.6,
          smoothness_worst = 0.08996, compactness_worst = 0.06444,
          concavity_worst = 0, concave_points_worst = 0,
          symmetry_worst = 0.2871, fractal_dimension_worst = 0.07039) %>%
  mutate(diagnosis =
           as.numeric(as.factor(recode(diagnosis, `M` = 1, `B` = 0))) - 1)
```

```r
# partitioning data
set.seed(31)
indexTrain <- createDataPartition(bc_df$diagnosis, p = 0.7, list = FALSE)
trainData = bc_df[indexTrain, ]
testData = bc_df[-indexTrain,]
```

```r
# very primitive EDA
bc_df_graph =
  bc_df %>%
  mutate(diagnosis = factor(recode(diagnosis, `1` = "M", `0` = "B"), level = c("B", "M")))
```

```r
cancer_mean = bc_df_graph[, 2:11] %>% as_tibble()
colnames(cancer_mean) = gsub("_mean", "", colnames(cancer_mean))
featurePlot(x = cancer_mean,
            y = bc_df_graph$diagnosis,
            scales = list(x = list(relation = "free"),
                          y = list(relation = "free")),
            plot = "density", pch = "|",
            auto.key = list(columns = 2))
```


B
M
symmetry
ractal_dimension
smoothness
compactness
concavity
concave_points
radius
texture
perimeter
area
0 5 10 15
0.10 0.20 0.30
0 20 40 60
0.06 0.08 0.10
0 10 20 30
0.05 0.10 0.15
0 5 10
0.0 0.1 0.2 0.3 0.4
0 5 10 15
0.0 0.1 0.2 0.3 0.4 0.5
0 5 15 25
0.00 0.10 0.20
0.00 0.15
5 10 15 20 25 30
0.00 0.06 0.12
10 20 30 40
0.00 0.02
50 100 150 200
0.0000 0.0025
0 500 1500 2500
Feature