

```
---
title: "R Notebook"
output: html_notebook
---
```

```
```{r}
x = read.csv('dataforanalysis.csv', header = TRUE)
Changing name of the first column
```

```
colnames(x)[1] = 'id'
dataset = data.frame(x[1:12])
```
```

```
```{r}
library(tidyverse)
library(lubridate)
library(tidylog)
library(ggplot2)
library(dplyr)
library(mlbench)
library(purrr)
library(ggplot2)
library(gridExtra)
```
```

```
```{r}
dataset %>%
 summary()
```
```

```
```{r}
colnames(dataset)
```
```

```
```{r}
dataset %>%
 group_by(diagnosis) %>%
 summarize(n = n())
```
```

Visualizing mean radius value in both benign and malignant conditions.

```
```{r}
dataset %>%
 group_by(diagnosis) %>%
 summarize(mean.radius.value = mean(radius_mean))
```
```

```
ggplot(data = dataset, mapping = aes(x = radius_mean)) +
  geom_histogram(binwidth = 5, color = "white")+
  facet_wrap(~diagnosis)+
  geom_vline(aes(xintercept= mean(radius_mean),
    col = 'red'),
    lwd= 0.5)
...

```

Visualizing mean texture value in both benign and malignant conditions.

```
```{r}
dataset %>%
 group_by(diagnosis) %>%
 summarize(mean.texture.value = mean(texture_mean))

ggplot(data = dataset, mapping = aes(x = texture_mean)) +
 geom_histogram(binwidth = 5, color = "white")+
 facet_wrap(~diagnosis)+
 geom_vline(aes(xintercept= mean(texture_mean),
 col = 'red'),
 lwd= 0.5)
...

```{r}
benign_mean = 17.91476
malignant_mean = 21.60491
...

```

```
```{r}
dataset %>%
 group_by(diagnosis) %>%
 summarize(mean.compactness.value = mean(compactness_mean))

ggplot(data = dataset, mapping = aes(x = compactness_mean)) +
 geom_histogram(binwidth = 0.05, color = "white")+
 facet_wrap(~diagnosis)+
 geom_vline(aes(xintercept= mean(compactness_mean),
 col = 'red'),
 lwd= 0.5)
...

```{r}
dataset = dataset %>% mutate(concavity.binary = if_else(dataset$concavity_mean < 0.06154,
0, 1))
...

```

```

```{r}
dataset %>%
 group_by(diagnosis) %>%
 summarize(mean.concavity.value = mean(concavity_mean))

ggplot(data = dataset, mapping = aes(x = concavity.binary)) +
 geom_histogram(binwidth = 0.05, color = "white")+
 facet_wrap(~diagnosis)+
 geom_vline(aes(xintercept= mean(concavity_mean),
 col = 'red'),
 lwd= 0.5)
...

```{r}
#dataset = dataset %>% mutate(concavity.binary = if_else(dataset$concavity_mean < 0.06154,
'typical', 'atypical'))

dataset = dataset %>%
  mutate(area.ordinal = case_when(
    area_mean < 420.3 ~ 'First Quartile',
    area_mean > 420.3 & area_mean < 551.1 ~ 'Second Quartile',
    area_mean > 551.1 & area_mean < 782.7 ~ 'Third Quartile',
    TRUE ~ 'Fourth Quartile'))
...

```{r}
dataset %>%
 group_by(diagnosis) %>%
 summarize(mean.perimeter.value = mean(perimeter_mean))

ggplot(data = dataset, mapping = aes(x = perimeter_mean)) +
 geom_histogram(binwidth = 0.05, color = "lightblue")+
 facet_wrap(~diagnosis)+
 geom_vline(aes(xintercept= mean(perimeter_mean),
 col = 'red'),
 lwd= 0.5)
...

```{r}
dataset %>%
  group_by(diagnosis) %>%
  summarize(mean.area.value = mean(area_mean))

```

```

ggplot(data = dataset, mapping = aes(x = area_mean)) +
  geom_histogram(binwidth = 0.05, color = "lightblue")+
  facet_wrap(~diagnosis)+
  geom_vline(aes(xintercept= mean(area_mean),
    col = 'red'),
    lwd= 0.5)
...

```

```

```{r}
dataset %>%
 ggplot(aes(x = texture_mean, color = concavity.binary, fill = concavity.binary)) +
 geom_histogram(position = 'stack', alpha = 0.5)+facet_wrap(~diagnosis)
...

```

```

```{r}
dataset %>%
  ggplot(aes(x = texture_mean, color = area.ordinal, fill = area.ordinal)) + geom_histogram(
    position = 'stack', alpha = 0.5)+facet_wrap(~diagnosis)
...

```