

Malaria Detector

Reading the date

The code reads a tab-delimited text file named "lancet_malaria.txt" located at "D:/BioHack/" into a dataframe 'df' in R, treating the first row as the header containing column names.

Hide

```
df <- read.table(file = "file:///D:/BioHack/lancet_malaria.txt",
                 header = TRUE, sep = "\t")
```

Preview the structure of the dataset

Hide

```
head(df)
```

| Review.Found | Author |
|--------------------|-----------------|
| <chr> | <chr> |
| 1 Whittaker et al. | Proietti |
| 2 Whittaker et al. | Souza |
| 3 Whittaker et al. | Souza |
| 4 Whittaker et al. | Souza |
| 5 Whittaker et al. | Atkinson |
| 6 Whittaker et al. | Pegha Moukandja |

6 rows | 1-3 of 23 columns

Hide

```
names(df)
```

```
[1] "Review.Found"      "Author"          "Title"           "Year"
[5] "Global.Region"    "Country"         "Location"        "PCR.N.Tested"
[9] "PCR.N.Positive"   "X..PCR..Positive" "Microscopy.N.Tested" "Microscopy.N.Positive"
[13] "X..Microscopy..Positive" "Historical.Transmission" "Current.Transmission" "Setting_20"
[17] "Setting_15"       "Setting_10"       "Setting_5"        "PCR_Method"
[21] "Microscopy_Fields" "Sampling_Season"  "Notes"
```

Scatter plot comparing PCR % against Microscopy %

Hide

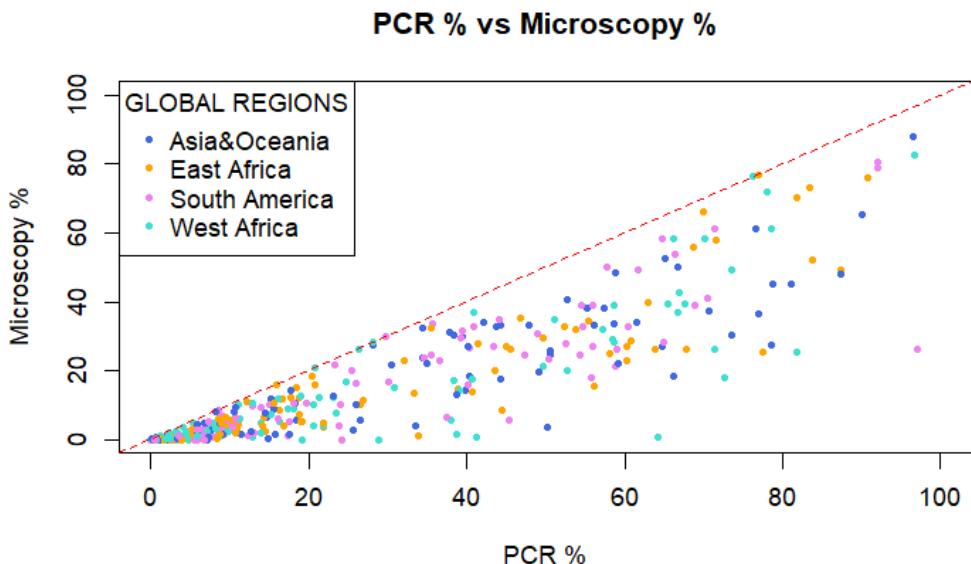
```
pallete = c("royalblue", "orange", "violet", "turquoise")

plot(
  x = df$X..PCR..Positive,
  y = df$X..Microscopy..Positive,
  xlab = "PCR %",
  ylab = "Microscopy %",
  main = "PCR % vs Microscopy %",
  xlim = c(0, 100),
  ylim = c(0, 100),
  pch = 20,
  col = pallete
)

abline(a = 0, b = 1, col = "red", lty = 2)
```

Hide

```
legend(
  "topleft",
  legend = levels(as.factor(df$Global.Region)),
  pch = 20, col=pallete,
  title = "GLOBAL REGIONS"
)
```



Calculate prevalence ratio

[Hide](#)

```
# Calculate prevalence ratio
df$Prevalence_Ratio <- df$Microscopy.N.Positive / df$PCR.N.Positive
```

PCR % and Microscopy % for different global regions

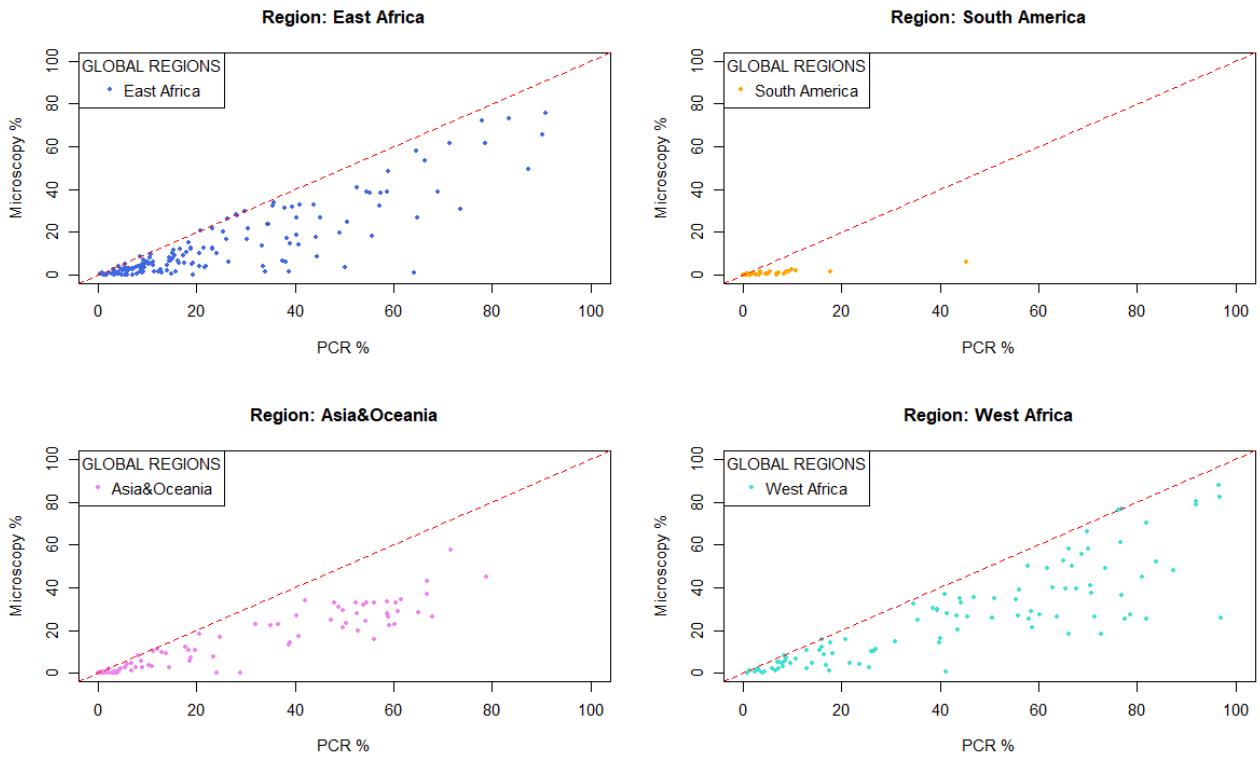
[Hide](#)

```
par(mfrow=c(2,2)) # 2x2 grid

# Filter data by global regions and plot
iteration_count <- 1
for (region in unique(df$Global.Region)) {
  print(region)
  subset_data <- df[df$Global.Region == region, ]
  plot(
    x=subset_data$X..PCR..Positive,
    y=subset_data$X..Microscopy..Positive,
    xlab = "PCR %",
    ylab = "Microscopy %",
    main = paste("Region:", region),
    xlim = c(0, 100),
    ylim = c(0,100),
    pch = 20,
    col = pallete[iteration_count]
  )
  abline(a = 0, b = 1, col = "red", lty = 2)

  legend(
    "topleft",
    legend = region,
    pch = 20, col=pallete[iteration_count],
    title = "GLOBAL REGIONS"
  )
  iteration_count <- iteration_count + 1
}
```

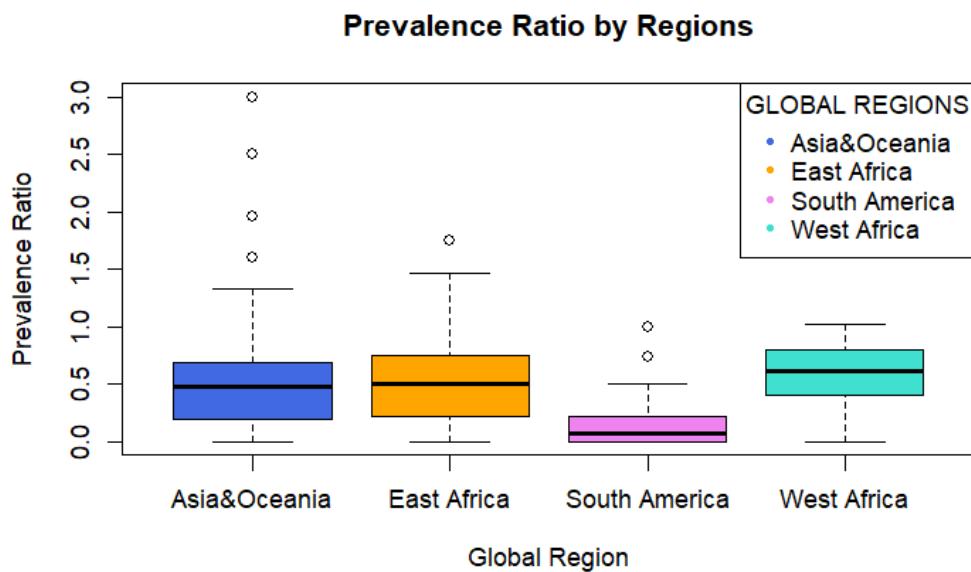
```
[1] "East Africa"
[1] "South America"
[1] "Asia&Oceania"
[1] "West Africa"
```



Boxplot for prevalence ratio by global regions

[Hide](#)

```
# Boxplot for prevalence ratio by global regions
boxplot(
  Prevalence_Ratio ~ Global.Region,
  data = df,
  xlab = "Global Region",
  ylab = "Prevalence Ratio",
  main = "Prevalence Ratio by Regions",
  col = pallete
)
legend(
  "topright",
  legend = levels(as.factor(df$Global.Region)),
  pch = 20, col=pallete,
  title = "GLOBAL REGIONS"
)
```



Interpretation:

It seems that in **South America**, the prevalence of malaria infections identified by microscopic methods is lower and more frequently infections identified by molecular methods Like PCR. It appears that submicroscopic infections in South America might have a higher density.