Report on Malaria Incidences in India from 2016 - 2019

Hemant Singh (Roll no. - 204161008)

14/04/2021

Subject : Data Visualization Laboratory (CS-595)

1. Motivation :

I personally feel that Health sector in India, is still an ignored topic for the masses. Amid the Covid-19 pandemic, the whole country realized on how important is to be healthy. As, the chance of more immune body works as a better self-defense mechanism for various diseases. This pandemic, also opened up eyes of the ignorant youth.

Corona-virus is just a single threat, there still are many diseases that disturbs the well-being of the masses. One such disease in **Malaria**. Malaria is a mosquito-borne infectious disease. Not only it affects the human body, it also affects other animals. The main symptoms include fever, tiredness, vomiting, and headaches. In severe cause it causes yellow skin, seizures, even coma, and in the worst-case scenario death. I personally feel motivated to visualize the trends shown by this disease, exploring such data may bring out techniques to tackle this disease.

Various visualizations will be done to show the effect of Malaria to the population of India. Also, to find out the trend of affected population by Malaria can further help in battling this disease. Having a geographical region to work upon this disease may lessen the toll it takes every year. So, this report shall visualize the important trends for Malaria over years, over all States and Union Territories. Also, the visualizations are done Region-wise, to have a more compact interpretation of the trends shown by this disease.

2. Objectives :

1. To analyze the Malaria Incidences in India, State-wise and Region-wise from year 2016 to 2019.
2. To compare different States and Union Territories, and different Regions of India and find out the top states and top regions prone to Malaria.
3. To see the distribution of Malaria over years 2016 - 2019, through various visualizations
4. To visualize how Malaria affects the Population of different regions of India.
5. To draw conclusion on how much Malaria as a disease, should be tackled in India.

3. Dataset :

I have taken the dataset from *www.data.gov.in*. The name of the Dataset is **State/UT-wise Malaria incidence from 2016 to 2019 (From: Ministry of Health and Family Welfare)**, the link of the dataset is *https://visualize.data.gov.in/?inst=ac16d8cb-286e-4b6f-b29d-898688904fa0&vid=101827*.

The dataset has total of 5 columns, these are, *State/UT*, 2016, 2017, 2018, 2019. The dataset has 37 rows with 36 States and Union Territories with an extra row for the Total value. One row contains the name of the State/UT, Malaria incidences in 2016, 2017, 2018, and 2019, respectively.

The code to see some rows of the dataset is stated below :-

3.1. Code :

```
malariaData <- read.csv("ac16d8cb-286e-4b6f-b29d-898688904fa0.csv")
malariaData$Region <- as.factor(c("Southern", "Northeastern", "Northeastern", "Eastern",
                                  "Central", "Western", "Western", "Northern",
                                  "Northern", "Northern", "Eastern", "Southern",
                                  "Southern", "Central", "Western", "Northeastern",
                                  "Northeastern", "Northeastern","Northeastern",
                                  "Eastern", "Northern", "Northern", "Northeastern",
                                  "Southern", "Southern", "Northeastern", "Northern",
                                  "Northern", "Eastern", "Islands", "Northern",
                                  "Western", "Western", "Northern", "Islands",
                                  "Southern", "None"))

colNames <- c("State.UT", "year.2016", "year.2017", "year.2018", "year.2019", "Region")
colnames(malariaData) <- colNames

totalOfDataset <- malariaData[nrow(malariaData),]
malariaData <- malariaData[1 : nrow(malariaData) - 1,]

head(malariaData)
```

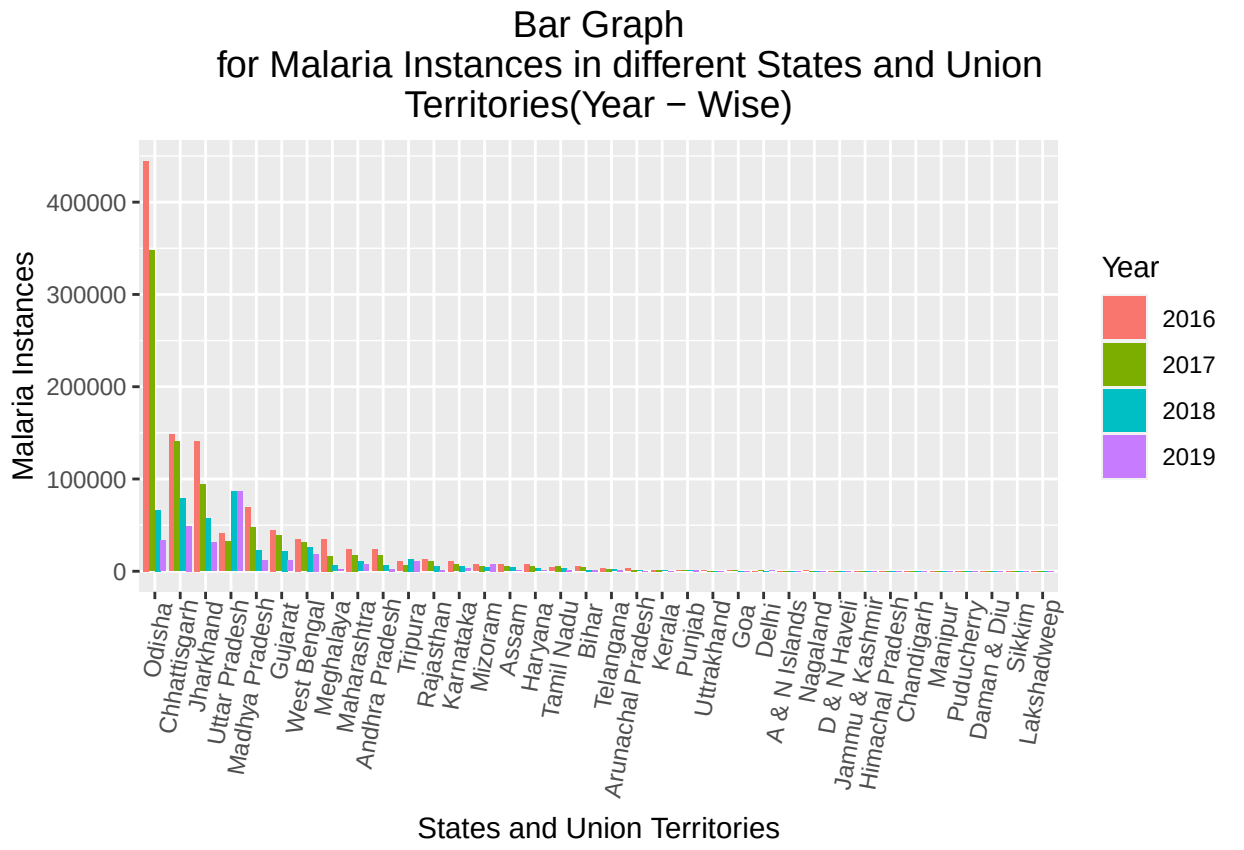
```
##            State.UT year.2016 year.2017 year.2018 year.2019       Region
## 1    Andhra Pradesh     23613     16972      6040      2623     Southern
## 2 Arunachal Pradesh      3128      1546       625       157 Northeastern
## 3             Assam      7826      5281      3816      1045 Northeastern
## 4             Bihar      5189      4020      1536      1302      Eastern
## 5      Chhattisgarh    148220    140727     78717     48774      Central
## 6               Goa       742       653       377       231      Western
```

4. Bar graph for all States and Union Territories showing Malaria incidences every year :

4.1. Code :

```
barData <- pivot_longer(malariaData[1 : 6], 2 : 5, names_to = "years",
                        values_to = "Instances")

barPlot <- ggplot(barData, aes(x = reorder(State.UT, -Instances), y = Instances)) +
  geom_bar(aes(fill = years), stat = "identity", position = "dodge") +
  scale_fill_discrete(name = "Year", labels = c("2016", "2017", "2018", "2019")) +
  labs(y = "Malaria Instances", x = "States and Union Territories",title = "Bar Graph
       for Malaria Instances in different States and Union \nTerritories(Year - Wise)") +
  theme(plot.title = element_text(hjust = 0.5, size = 14),
        axis.text.x = element_text(angle = 80, vjust = 1.0, hjust = 0.95))
```

4.2. Plot :



4.3. Conclusions :

1. Odisha have the most cases of Malaria with a huge margin over other States and UTs. As the peak of incidences took place in 2016. The trend did not seem to follow as till 2019, the cases got significantly reduced.
2. Next up in the most cases are Chattisgarh and Jharkhand. This result seems reasonable as these both states are geographically close. However, West Bengal, being another neighbor does not show this trend.

5. Radar graph for an yearly comparison of different Regions of India :

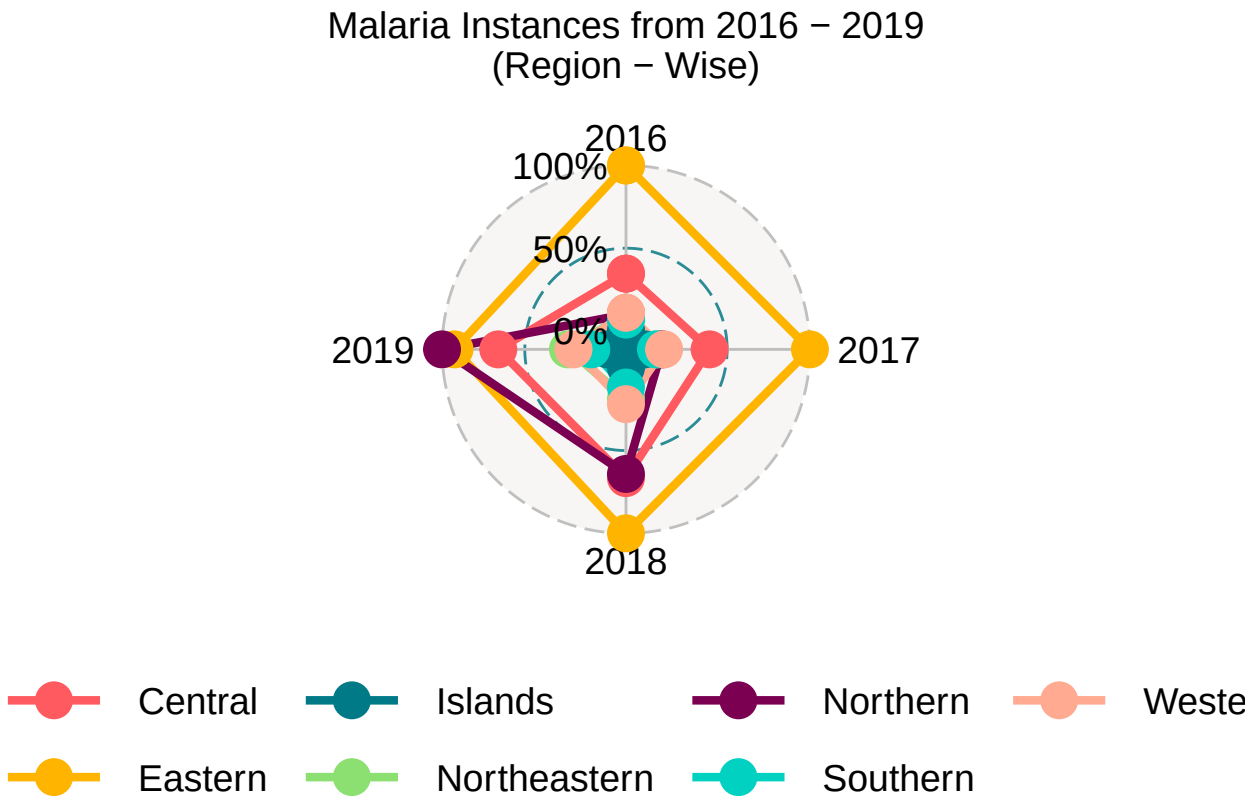
The Indian Regions are classified as follows :

1. Northern : Haryana, Himachal Pradesh, Jammu & Kashmir, Punjab, Rajasthan, Uttar Pradesh, Uttarakhand, Chandigarh, and Delhi.
2. North-Eastern : Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura.
3. Eastern : Bihar, Jharkhand, Odisha, and West Bengal.
4. Central : Chhattisgarh, and Madhya Pradesh.
5. Western : Goa, Gujarat, Maharashtra, Dadar & Nagar Haveli, and Daman & Diu.
6. Southern : Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Telangana, and Puducherry.
7. Islands : Andaman & Nicobar Islands, and Lakshadweep.

5.1. Code :

```
radarData <- ddply(malariaData[2 : 6], "Region", numcolwise(sum))
regions = radarData$Region
radarData <- radarData[2 : ncol(radarData)]
colnames(radarData) <- c("2016", "2017", "2018", "2019")
rownames(radarData) <- as.character(regions)
radarData <- radarData %>% mutate_each(funs(rescale))
radarData <- radarData %>% as_tibble(rownames = "Region") %>% mutate_at(vars(-Region),
                                                                        rescale)

 radarPlot <- ggradar(radarData) + labs(title = "Malaria Instances from 2016 - 2019\n(Region - Wise)") +
   theme(legend.position = "bottom", plot.title = element_text(hjust = 0.5, size = 14))
```

5.2. Plot :



5.3. Conclusion :

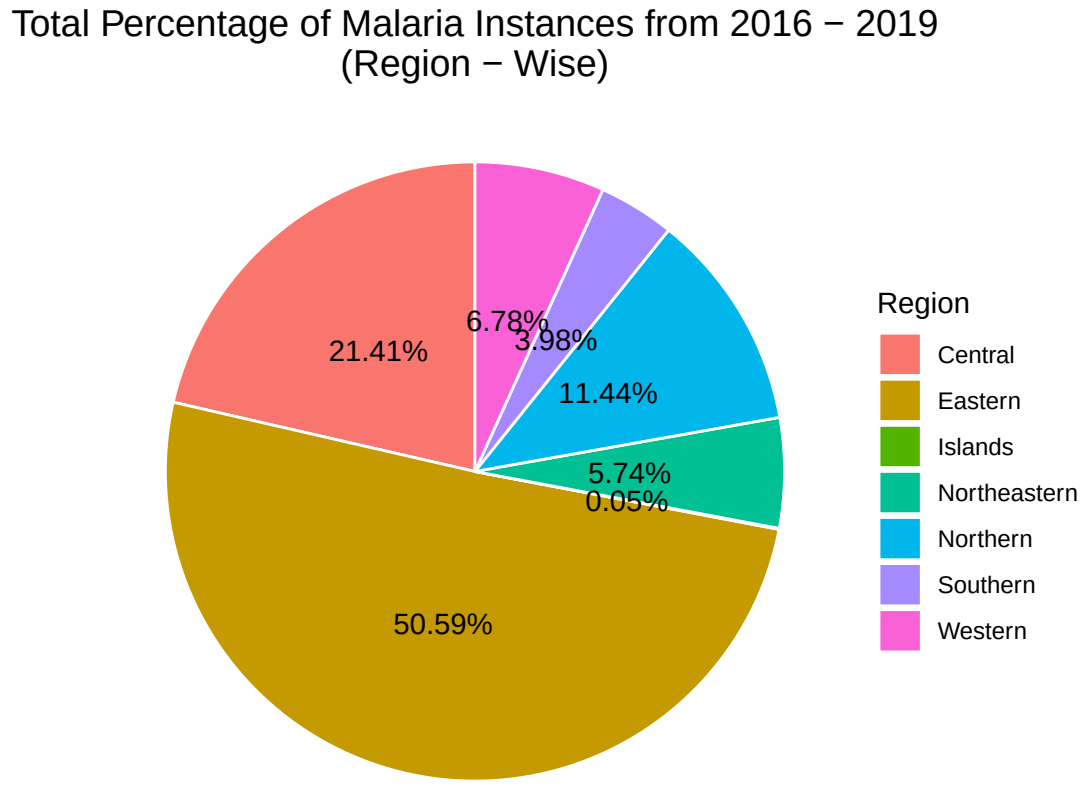
1. The Eastern states tends to show the maximum proportion of cases, but we did see the bar graph where Odisha has the maximum Malaria incidences and, It is in Eastern Region. Further analyzing the bar plot West Bengal did not show much incidences but is an Eastern State. So, it would not be fair to conclude that Eastern states have the mosty Malaria cases.
2. The North-Eastern States seem to increase their proportion on 2018 and 2019.
3. The Central shows the proportion to be close to 50% in all 4 years. The proportion is almost constant, compared to other regions.
4. For Western, Northern, Southern and Islands regions have very less proportion of Malaria cases.

6. Pie Chart to show the total Percentage of Malaria cases amongst different Regions :

6.1. Code :

```
pieData <- ddply(malariaData[2 : 6], "Region", numcolwise(sum))
pieData <- pieData %>%  mutate(Total = year.2016 + year.2017 + year.2018 + year.2019)
pieData <- pieData %>% mutate(Percentage = Total / sum(Total))

pieChart <- ggplot(pieData, aes(x = "", y = Percentage, fill = Region)) +
  geom_bar(stat = "identity", width = 1, color = "white") + coord_polar("y", start = 0) +
  geom_text(aes(label = paste0(round(Percentage * 100, 2), "%")), position = position_stack(vjust = 0.5)
  labs(x = NULL, y = NULL, title = "Total Percentage of Malaria Instances from 2016 - 2019\n(Region - W
  theme(plot.title = element_text(hjust = 0.5, size = 14))
```

6.2. Plot :



6.3. Conclusion :

1. Eastern Region shows a drastic amount of Malaria cases with a 50.59%. Still, knowing that the majority of cases were from Odisha, generalizing Eastern States is unfair. 2.Then, Central shows a significant amount of total cases, with a 21.41%
2. Rest all other regions seems to have a very less proportion of total Malaria cases in the period 2016 - 2019.
3. The Islands seems the best place to avoid Malaria. It's proportion is very less with a 0.05%.

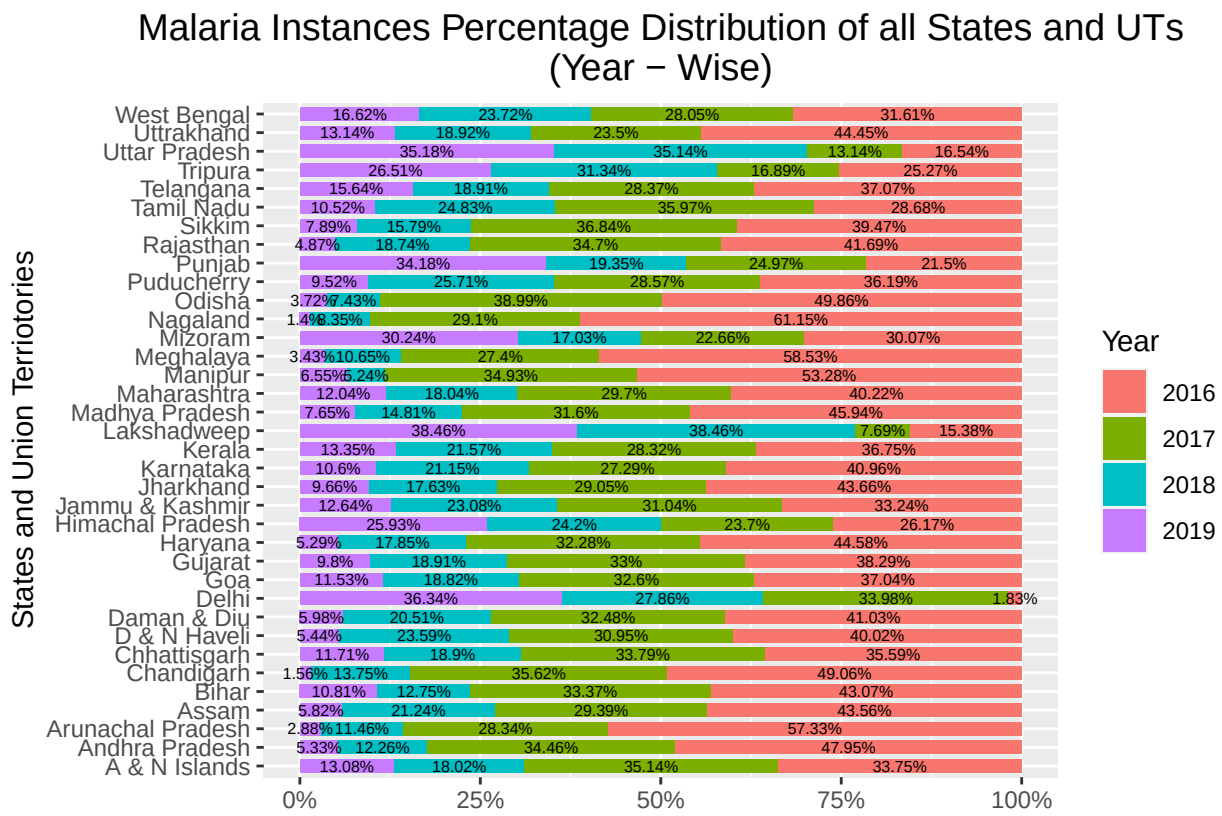
7. Stacked Bar Plot to show the percentage distribution of cases in these 4 years(2-16 - 2019), State-wise and Region-wise :

7.1 State - Wise Visualization :

7.1.1. Code :

```
stackedBarData <- malariaData %>%
  mutate(Total = year.2016 + year.2017 + year.2018 + year.2019)
stackedBarData <- stackedBarData %>%mutate(Percentage.2016 = year.2016 / Total,
                                           Percentage.2017 = year.2017 / Total,
                                           Percentage.2018 = year.2018 / Total,
                                           Percentage.2019 = year.2019 / Total)
stackedBarData <-  stackedBarData[, c("State.UT", "Percentage.2019", "Percentage.2018",
                                      "Percentage.2017", "Percentage.2016")]
stackedBarData <- pivot_longer(stackedBarData, 2 : ncol(stackedBarData), names_to = "Percentage.Year", v
stackedBarData$Percentage <- round(stackedBarData$Percentage * 100, 2)

stateWiseStackedBarPlot <- ggplot(stackedBarData, aes(x = State.UT, y = (Percentage / 100), fill = Perc
  geom_text(aes(label = paste0(Percentage, "%")), position = position_stack(vjust = 0.5), size = 2) +
  labs(x = "States and Union Terriotories", y = "", title = "Malaria Instances Percentage Distribution
  theme(plot.title = element_text(hjust = 0.5, size = 14)) +
  scale_fill_discrete(name = "Year", labels = c("2016", "2017", "2018", "2019")) +
  scale_y_continuous(labels=scales::percent) + coord_flip()
```

7.1.2. Plot :



7.1.3. Conclusion :

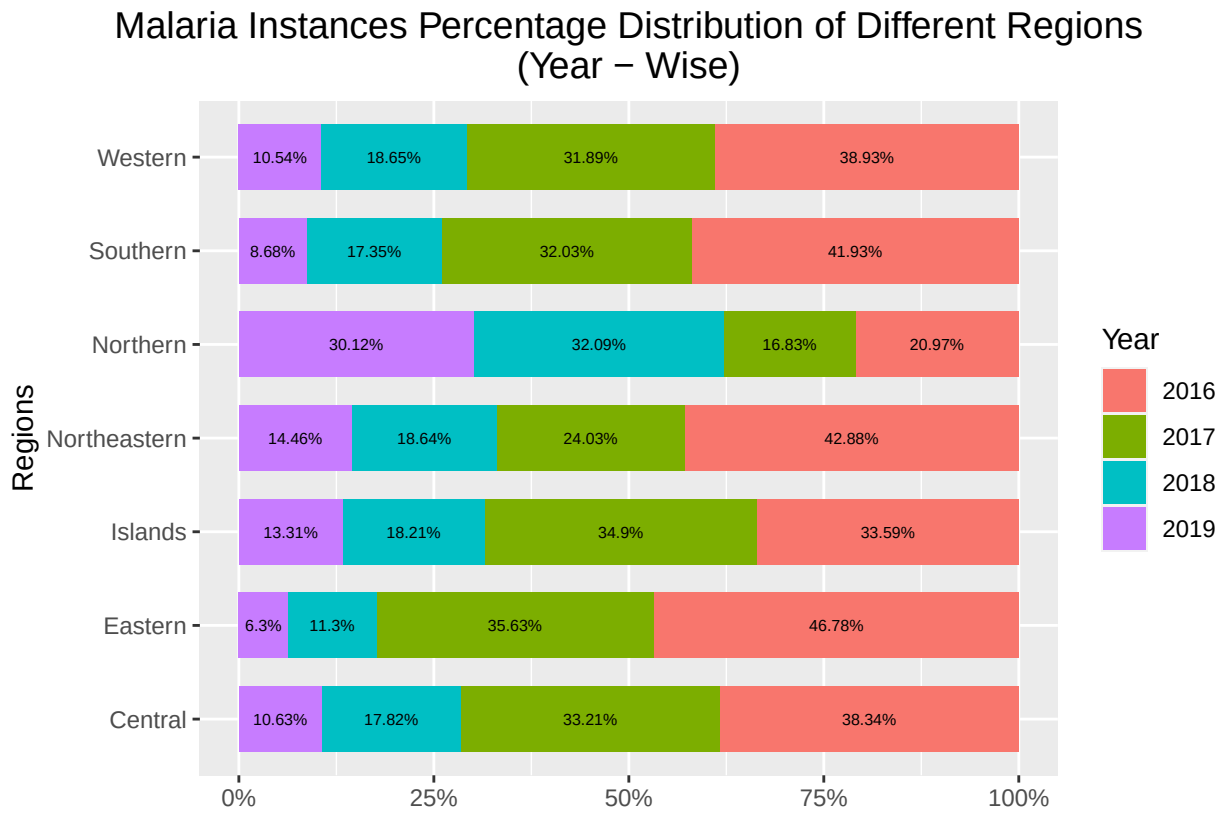
1. State-wise, the trend seems to follow, as most of the percent of cases of individual states are seen on 2016, with a percent values close to 40%.
2. Surprisingly, in 2017 the trend is even more stronger with most values close to 30%.
3. The proportion seems to decrease but the trend of same percent values still follow with percent value close to 20%.
4. By 2019, most of the States and UTs have a very less proportion of Malaria cases comparing their previous years.
5. As, most of the States and UTs shows a decrease in their Malaria cases, some States are doing completely opposite. These are Uttar Prades, Punjab, Lakshadweep, and Delhi.

7.2. Region - Wise Visualization :

7.2.1. Code :

```
stackedBar2Data <- ddply(malariaData[2 : 6], "Region", numcolwise(sum))
stackedBar2Data <- stackedBar2Data %>%
  mutate(Total = year.2016 + year.2017 + year.2018 + year.2019)
stackedBar2Data <- stackedBar2Data %>%
  mutate(Percentage.2016 = year.2016 / Total, Percentage.2017 = year.2017 / Total, Percentage.2018 = ye
stackedBar2Data <- stackedBar2Data[, c("Region", "Percentage.2016", "Percentage.2017", "Percentage.2018
stackedBar2Data <- pivot_longer(stackedBar2Data, 2 : ncol(stackedBar2Data), names_to = "Percentage.Year
stackedBar2Data$Percentage <- round(stackedBar2Data$Percentage * 100, 2)

regionWiseStackedBarPlot <- ggplot(stackedBar2Data, aes(x = Region, y = (Percentage / 100), fill = Perc
  geom_bar(position = position_stack(), stat = "identity", width = 0.7) +
  geom_text(aes(label = paste0(Percentage, "%")), position = position_stack(vjust = 0.5), size = 2) +
  labs(x = "Regions", y = "", title = "Malaria Instances Percentage Distribution of Different Regions\n
  theme(plot.title = element_text(hjust = 0.5, size = 14)) +
  scale_fill_discrete(name = "Year", labels = c("2016", "2017", "2018", "2019")) +
  scale_y_continuous(labels=scales::percent) + coord_flip()
```

7.2.2. Plot :



7.2.3 Conclusion :

1. Region-wise, all the region except Northern follows a trend of decreasing their proportions of malaria cases.
2. Only Northern Region seems to follow the opposite trend, with the increasing year the proportion of cases are rising.

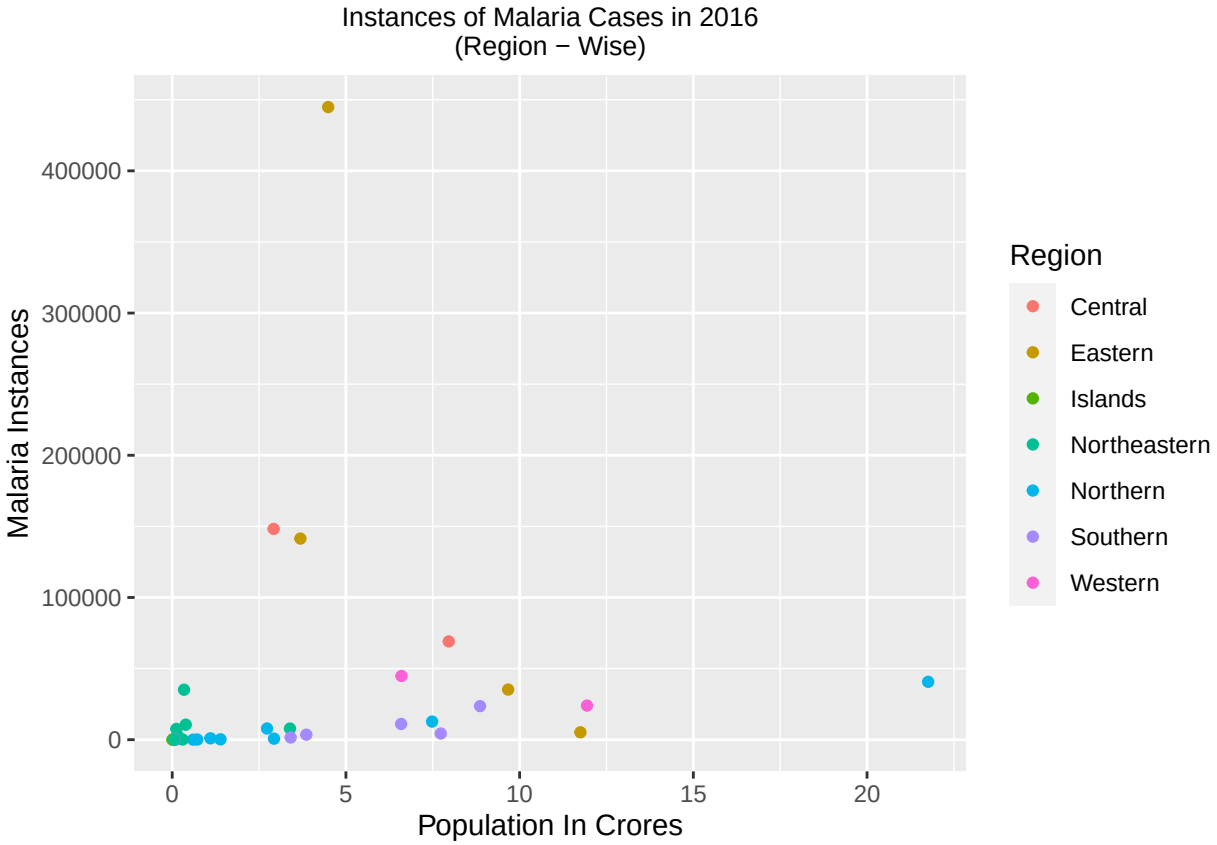
8. Scatter plots to show the distribution of cases of different Region in years 2016 - 2019 :

8.1 : Code :

```
#The population data is taken and inputted manually, since there was no particular data available
#anywhere, the values are taken from  www.census2011.co.in
populationData <- data.frame(
  State.UT = malariaData$State.UT,
  population.2016 = c(8.86, 0.15, 3.39, 11.75, 2.92, 0.15, 6.60, 2.73, 0.72, 1.39, 3.69,
                      6.59, 3.41, 7.96, 11.94, 0.30, 0.34, 0.12, 0.20, 4.49, 2.93, 7.48,
                      0.06, 7.73, 3.86, 0.39, 21.76, 1.10, 9.67, 0.04, 0.11, 0.04, 0.03,
                      .60, 0.006, 0.07),
  population.2017 = c(8.94, 0.16, 3.44, 12.03, 2.99, 0.15, 6.71, 2.77, 0.73, 1.42, 3.77,
                      6.68, 3.43, 8.09, 12.08, 0.32, 0.35, 0.12, 0.20, 4.54, 2.96, 7.60,
                      0.06, 7.86, 3.91, 0.40, 22.11, 1.12, 9.78, 0.04, 0.11, 0.04, 0.03,
                      2.80, 0.007, 0.08),
  population.2018 = c(9.02, 0.16, 3.50, 12.30, 3.07, 0.15, 6.82, 2.81, 0.74, 1.44, 3.85,
                      6.77, 3.44, 8.23, 12.22, 0.33, 0.36, 0.13, 0.21, 4.60, 2.99, 7.71,
                      0.06, 8.12, 3.96, 0.41, 22.46, 1.13, 9.88, 0.04, 0.11, 0.04, 0.03,
                      2.90, 0.007, 0.08),
  population.2019 = c(9.10, 0.16, 3.55, 12.57, 3.14, 0.15, 6.93, 2.85, 0.74, 1.47, 3.93,
                      6.87, 3.55, 8.37, 12.35, 0.34, 0.37, 0.13, 0.21, 4.65, 3.02, 7.84,
                      0.07, 8.24, 4.01, 0.41, 22.81, 1.15, 9.98, 0.04, 0.11, 0.04, 0.03,
                      3.0, 0.007, 0.08)
)

scatterData <- merge(malariaData, populationData, by.x = "State.UT")

scatterPlot2016 <- ggplot(scatterData, aes(x = population.2016, y = year.2016, color = Region)) +
  geom_point() +
  labs(x = "Population In Crores", y = "Malaria Instances", title = "Instances of Malaria Cases in 2016
  theme(plot.title = element_text(hjust = 0.5, size = 10))

scatterPlot2017 <- ggplot(scatterData, aes(x = population.2017, y = year.2017, color = Region)) + geom_
  geom_point() +
  labs(x = "Population In Crores", y = "Malaria Instances", title = "Instances of Malaria Cases in 2017
  theme(plot.title = element_text(hjust = 0.5, size = 10))

scatterPlot2018 <- ggplot(scatterData, aes(x = population.2018, y = year.2018, color = Region)) + geom_
  geom_point() +
  labs(x = "Population In Crores", y = "Malaria Instances", title = "Instances of Malaria Cases in 2018
  theme(plot.title = element_text(hjust = 0.5, size = 10))

scatterPlot2019 <- ggplot(scatterData, aes(x = population.2019, y = year.2019, color = Region)) +
  geom_point() +
  geom_point() +
  labs(x = "Population In Crores", y = "Malaria Instances", title = "Instances of Malaria Cases in 2019
  theme(plot.title = element_text(hjust = 0.5, size = 10))
```

8.2 Plots :

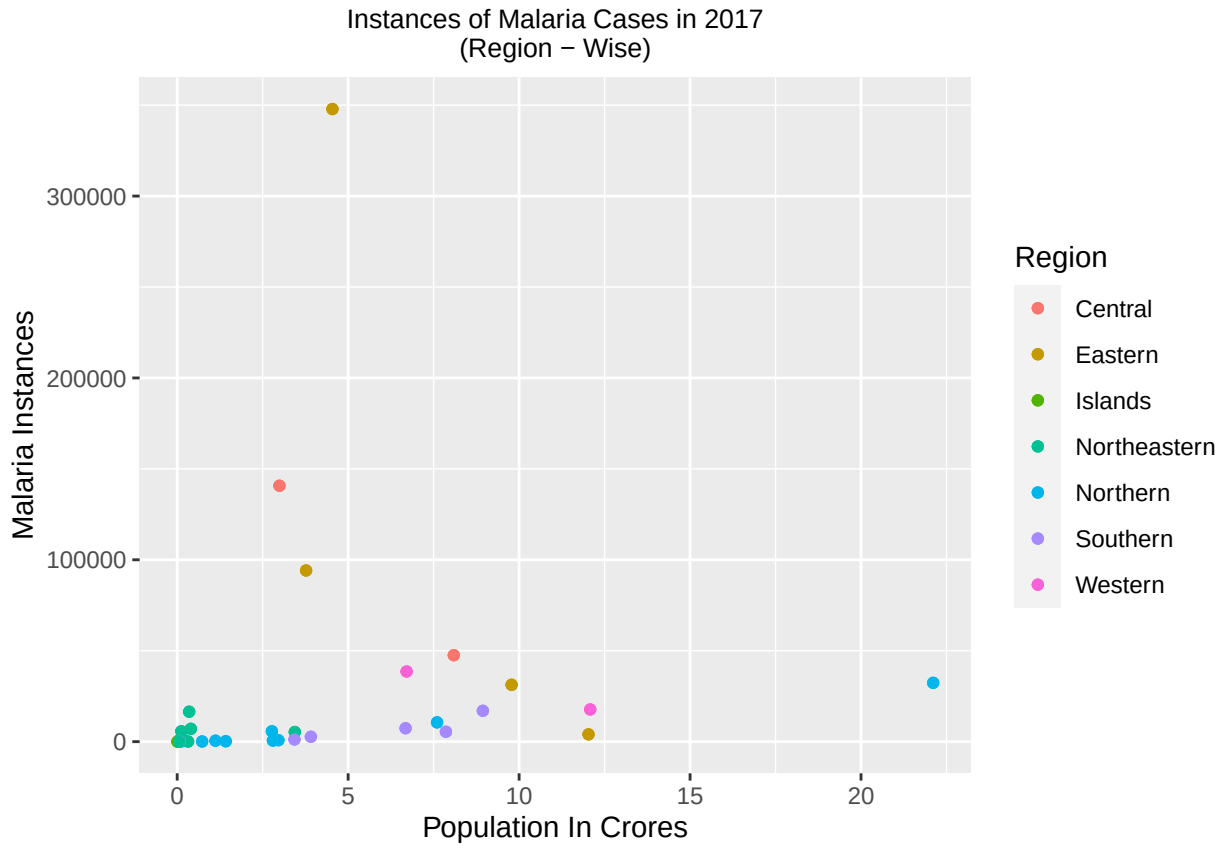
8.2.1. Plot for 2016 :



Conclusions for 2016 :

There is not a very good trend seen in the scatter plot. The only alarming data point is one data point belonging to Eastern Region. Rest of all have less malaria cases irrespective of their population.

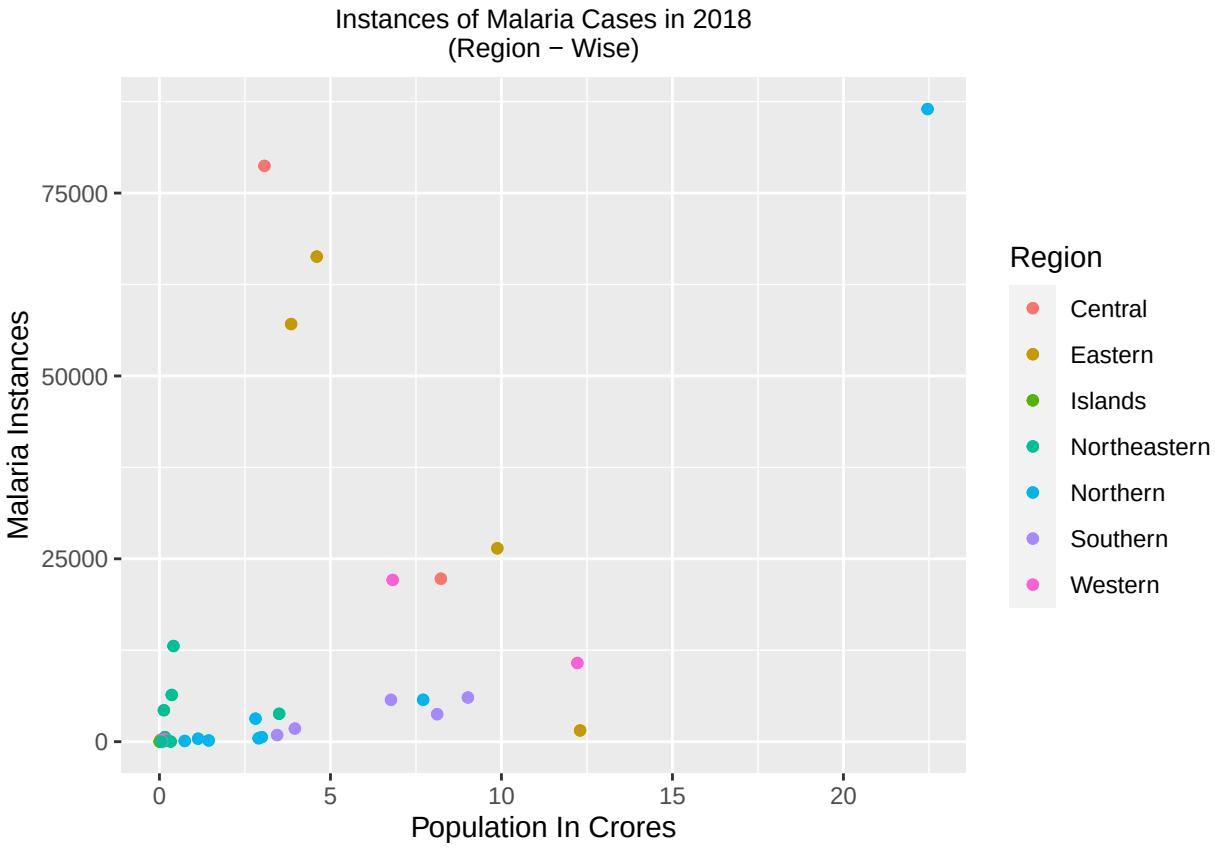
8.2.2. Plot for 2017 :



Conclusions for 2017 :

The trend is very similar to that of year 2016.

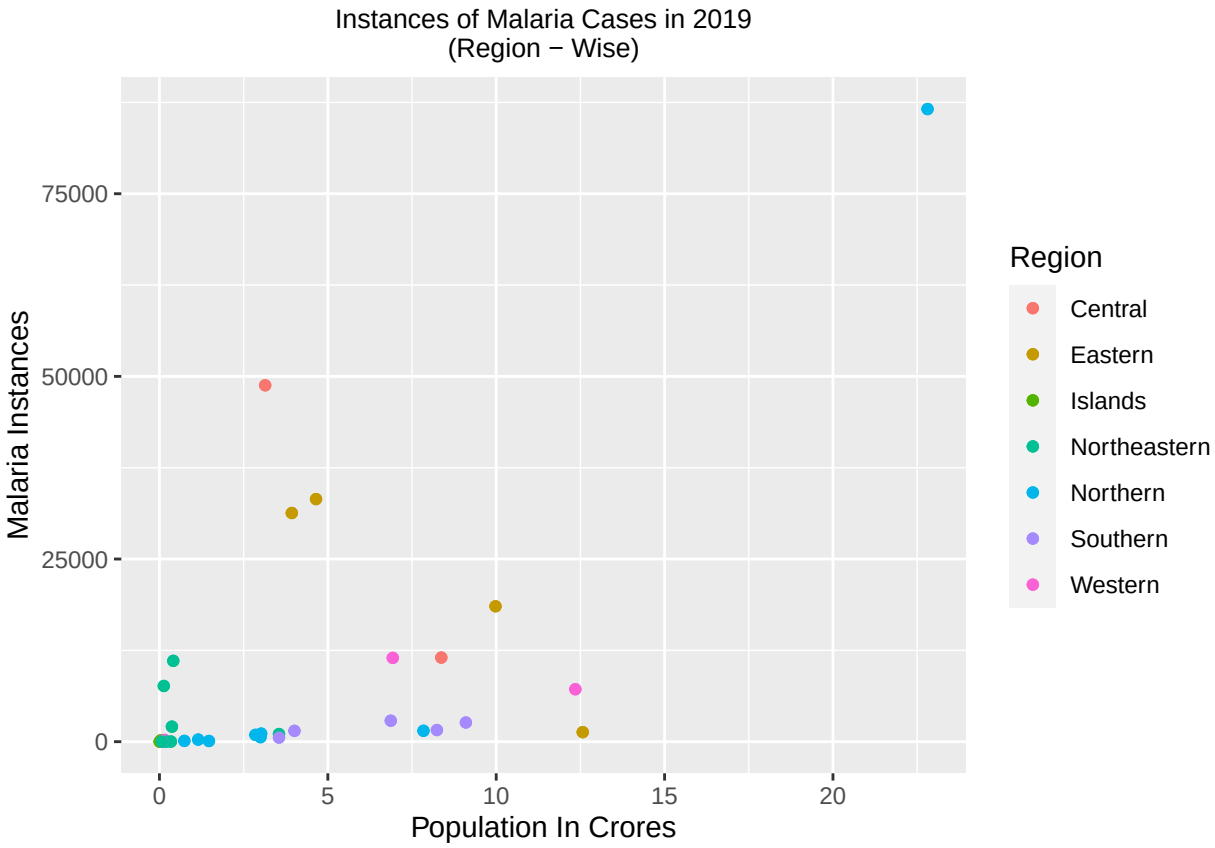
8.2.3. Plot for 2018 :



Conclusions for 2018 :

The trend has changed, Here some States are showing a significant hike in their values. These States are from Eastern and Central Regions.

8.2.4. Plot for 2019 :



Conclusions for 2019 :

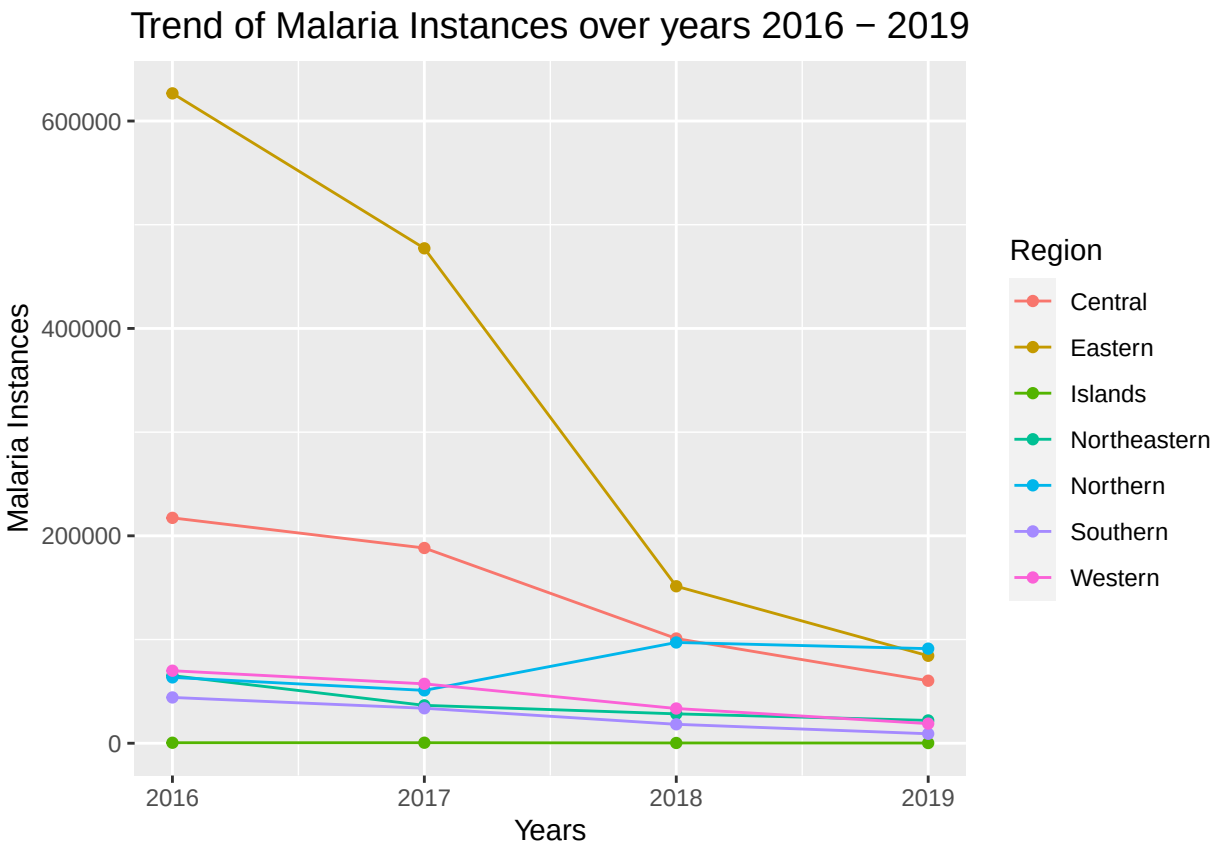
The trend has changed again, some more States or UTs of mainly Central and Eastern Region have started showing a hike in malaria cases. One of the Northern region has a vast number of malaria cases with a very large population. This State is Uttar Pradesh as we saw in the stacked bar chart, that Uttar Pradesh has increased malaria Cases over the subsequent years.

9. Line Chart to show the trend of Malaria cases over 2016 - 2019 in different Regions.

9.1. Code :

```
lineData <- ddply(malariaData[2 : 6], "Region", numcolwise(sum))
colnames(lineData) <- c("Region", "2016", "2017", "2018", "2019")
lineData <- pivot_longer(lineData, 2 : 5, names_to = "years", values_to = "Instances")

linePlot <- ggplot(lineData, aes(x = as.integer(years), y = Instances)) +
  geom_line(aes(color = Region)) + geom_point(aes(color = Region)) +
  labs(x = "Years", y = "Malaria Instances", title = "Trend of Malaria Instances over years 2016 - 2019
  theme(plot.title = element_text(hjust = 0.5, size = 14))
```

9.2 Plot :



9.3 Conclusion :

It seems that all the Regions were able to substantially decrease their malaria cases count as all of them are decreasing with subsequent years. Only Northern Region has an increase in the Malaria cases after 2017.

10. Proportional Symbol Map to show the intensity of total malaria struck cases in every State and UT.

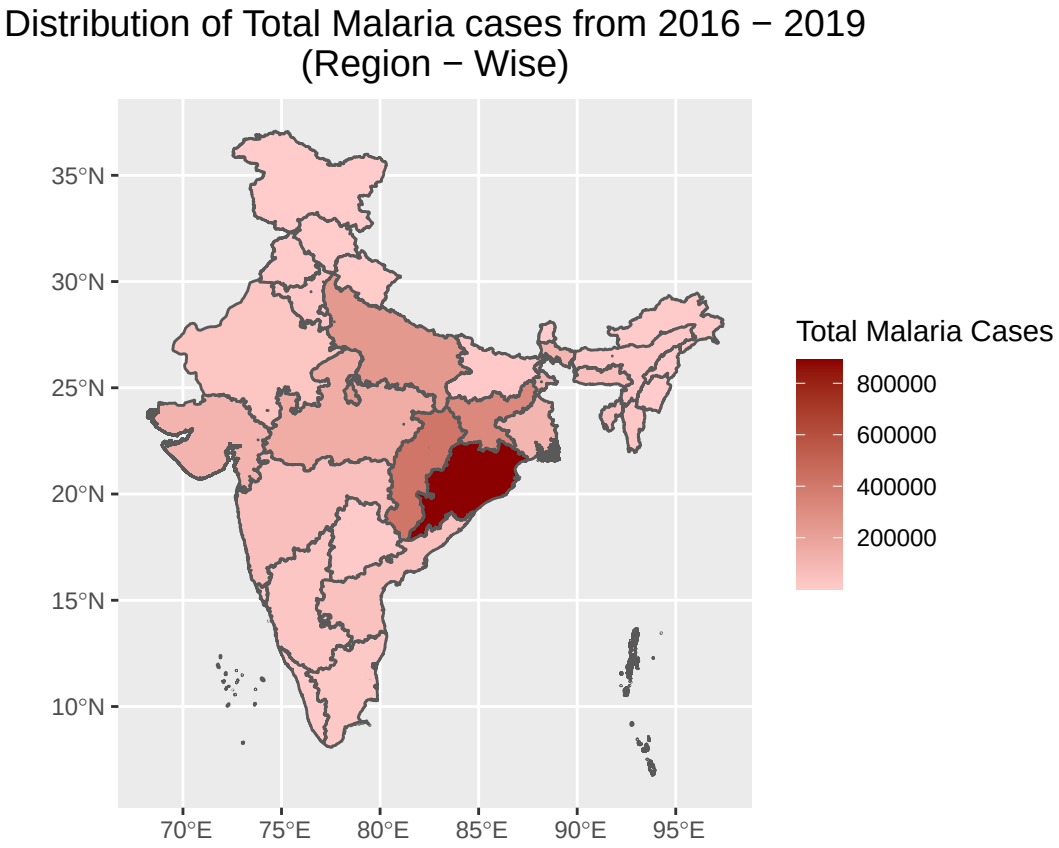
10.1. Code :

```
indianMap <- st_read("india_administrative_state_boundary/india_administrative_state_boundary.shp")
```

```
## Reading layer `india_administrative_state_boundary' from data source
##   `/home/devlc169/Academics/DataVisualisationLab/Assignments/Project for git/Code/india_administrati
##   using driver `ESRI Shapefile'
## Simple feature collection with 36 features and 4 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 68.18625 ymin: 6.755953 xmax: 97.41529 ymax: 37.07827
## Geodetic CRS:  WGS 84
```

```
indianMap$commonIndex = c(1 : 36)

malariaDataWithTotal <- malariaData %>% mutate(Total = year.2016 + year.2017 + year.2018 + year.2019)
indexesForStates <- c(36, 2, 3, 4, 6, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22,
                      23, 35, 26, 27, 28, 29, 30, 31, 32, 33, 34, 1, 5, 7, 8, 24, 17,
                      25)
malariaDataWithTotal$commonIndex <- indexesForStates

mapAndData <- inner_join(indianMap, malariaDataWithTotal, by = "commonIndex")

mapPlot <- ggplot(mapAndData, aes(fill = Total)) + geom_sf() +
  scale_fill_gradient(low = "#FFCCCB", high = "#8B0000", name = "Total Malaria Cases") +
  labs(title = "Distribution of Total Malaria cases from 2016 - 2019\n(Region - Wise)") +
  theme(plot.title = element_text(hjust = 0.5, size = 14))
```

10.2. Plot :



10.3 Conclusion :

1. Northern except Uttar Pradesh, Western Southern and North-Eastern States have around < 20000 cases in total of 4 years.
2. Central Region except Odisha, including Uttar Pradesh from Northern Region have cases in the range 20000 - 35000 in these 4 years.
3. Only Odisha alone has the maximum of total cases of > 80000 in the 4 years. From Other graphs it was shown that Odhisha managed a great deal of case reduction in year 2018 and 2019.