Vector Borne Diseases Mapped against Flooding and Population in USA

Final assignment for the course “Data Visualization” - SS22

Submitted by:

Anumula Naga Raviteja (12200983)

raviteja.anumula-naga@stud.th-deg.de

Under the supervision of

Prof. Dr Javier Valdes

Contents

List of Tables

List of Figures

LIST OF SYMBOLS, ABBREVIATIONS AND NOMENCLATURE

WNV : West Nile Virus

USA : United States of America

CA : California

MD : Maryland

EDA : Exploratory Data Analysis

TX : Texas

AZ : Arizona

CO : Colorado

FL : Florida

Chapter 1

OVERVIEW

In the United States, the most common vector-borne pathogens are transmitted by ticks or mosquitoes, including those causing Malaria, Rocky Mountain spotted fever, West Nile, dengue, and Zika virus diseases. Vector-borne diseases have been difficult to prevent and control[1]. In this report, we will see how the cases of three major vector-borne diseases namely Dengue, malaria and West Nile Virus has been increased in USA over the time period of 2013-2014. Also we will see how floods has affected the rising number of cases.

Keywords: Dengue, Malaria, West Nile Virus, Vector-borne diseases, Floods, Population

1.1 Introduction

Vectors are blood-feeding insects and ticks capable of transmitting pathogens between hosts. Wide varieties of pathogens have evolved to exploit vector transmission, including some viruses, bacteria, rickettsia, protozoa, and helminths. Dengue viruses are estimated to infect nearly 400 million persons worldwide each year, and malaria is a major cause of pediatric mortality in equatorial Africa[1]. In the United States, 16 vector-borne diseases are reported to state and territorial health departments. Among the diseases on the list that are caused by indigenous pathogens are Lyme disease (Borrelia burgdorferi), West Nile, dengue and Zika virus diseases, plague (Yersinia pestis), and Malaria. As a group, vector-borne diseases in the United States are notable for their wide distribution and resistance to control. A Food and Drug Administration–approved vaccine is available to prevent only one of the notifiable diseases, yellow fever.

Vector-borne disease epidemiology is complex because of environmental influences on the biology and behavior of the vectors. The longevity, distribution, biting habits, and propagation of vectors, which ultimately affect the intensity of transmission, depend on environmental factors such as rainfall, temperature, and shelter. Most vector-borne pathogens are zoonoses, often with wild animal reservoirs, such as rodents or birds, making them difficult or impossible to eliminate. Arthropod vectors can bridge the gap between animals and humans that would not ordinarily intersect, as happens in Lyme disease, plague, and WNV, facilitating the introduction of emerging animal pathogens to humans[1]. In this report we concentrate on particularly rainfall factor which in-turn causes floods.

The pace of emergence of new or obscure vector-borne pathogens through introduction or belated recognition appears to be increasing. in 2016, there were more than 20,000 cases of Dengue has been reported. And the cases of Malaria has been increasing over the time from 2013 - 2015. The highest number of Malaria cases were reported in the year 2017 i.e., 36,210. The number of cases of WNV are almost similar in every year around 14,000[2].

In this report we will examine the trends of reported vector-borne disease cases in the United States during 2013–2017 with respect to floods and population.

1.2 Problem Definition

The aim of this assignment is to analyze the data provided by the “Thailand Dengue Surveillance System” hosted on Project Tycho’s website for the effects of floods and population on the increase in number of vector-borne diseases. The floods data has been retrieved from the ”Global Active Archive of Large Flood Events” data-set collected by the “Dartmouth Flood Observatory”[3]. And population data has been retrieved from an official website of the USA government[4]. We will limit our exploration for the years 2013 and 2017 and for the provinces of “WYOMING”, ”GUAM”, ”DISTRICE OF COLUMBIA”, ”SOUTH DAKOTA” and “PUERTO RICO”. And also we are considering three major vector-borne diseases namely Dengue, Malaria and WNV. This limitation is because of data availability and conciseness of data exploration.

1.3 Objective

The main objective of this project is to perform Data Visualisation of the ”Project Tycho” data-sets and observe how floods and population has been an factors.This EDA will help researchers to compare and see if flooding caused the increase in number of vector-borne diseases in USA.

Chapter 2

Methodology

Various EDA has been performed on the data-sets by plotting graphs and charts in this assignment shown as follows:

2.1 Cases of Dengue, Malaria and WNV in USA(2013-2014)

Let's begin with visualising each vector-borne disease with the number of cases separately. Projected on a scale where less number of cases is ”white” and it gets ”dark blue” indicating the highest number of cases.

• Dengue Cases

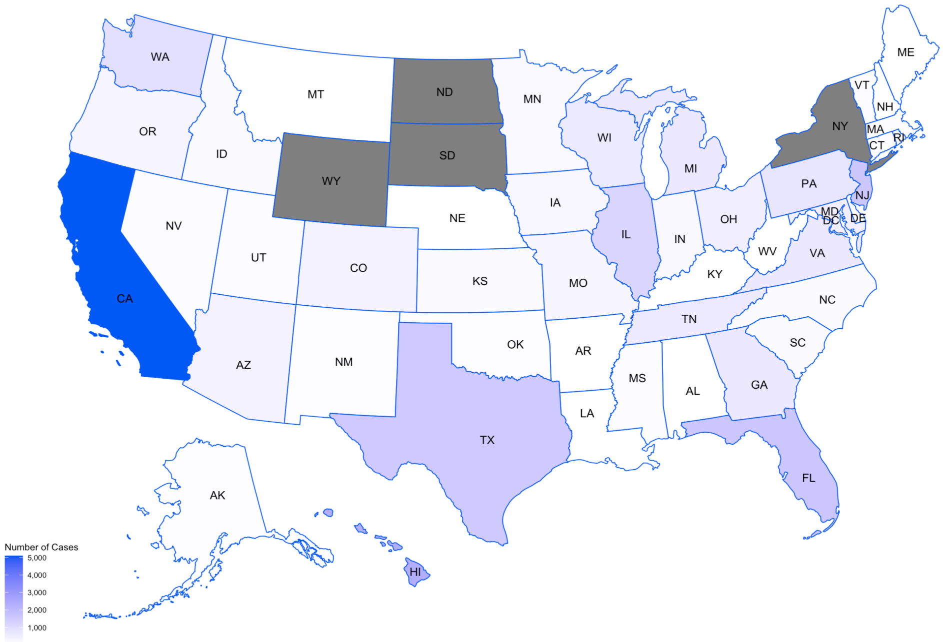

Figure 2.1: Dengue Cases over the years 2013-2017

We can observe that CA has been recorded with highest number of dengue cases followed by TX and FL.

- Malaria Cases

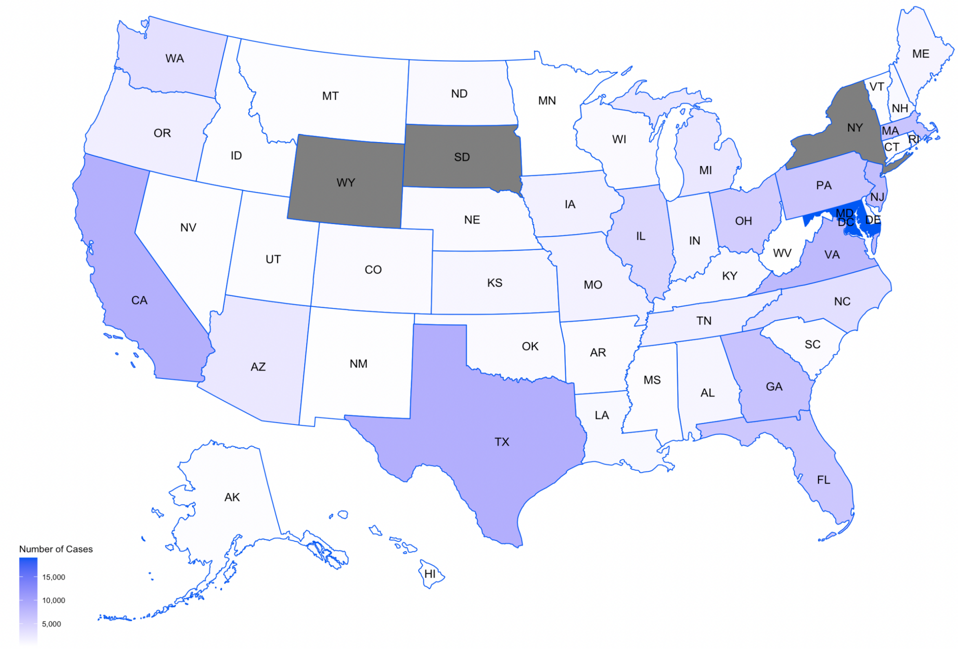

Figure 2.2: Malaria Cases over the years 2013-2017

Highest number of Malaria cases has been recorded with MD followed by its near by region. Apart from them CA and TX also have been affected severely.

• WNV Cases

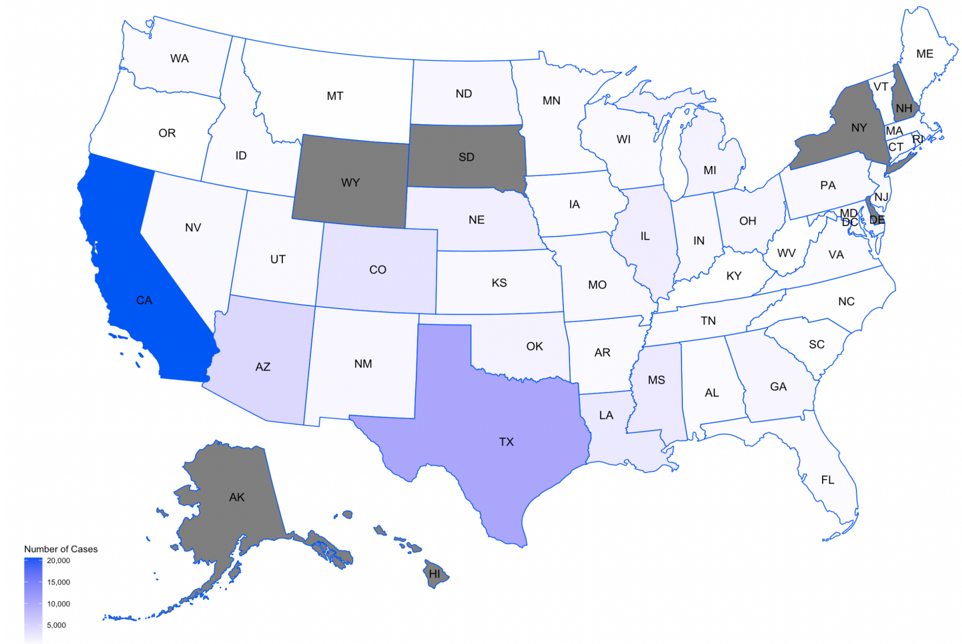

Figure 2.3: WNV Cases over the years 2013-2017

CA being the most affected by WNV followed by TX. Apart from them few cases were from AZ and CO.

2.2 Visualising Dengue, Malaria and WNV diseases combined in USA(2013-2014)

• Now we will see the table with all the cases over the years 2013-2017 in USA.

Table 2.1: Combined data of three diseases over the years 2013-2017

Year	Dengue Cases	Malaria Cases	WNV Cases
2015	3546	22237	17977
2016	23281	31125	14176
2017	9420	36210	18543
2013	0	7926	172
2014	0	30475	17424

From this we can see that the data of Dengue cases has been missing for the years 2013 and 2014.

- Let's visualize the above table using Box Plot and Grouped Bar plots -

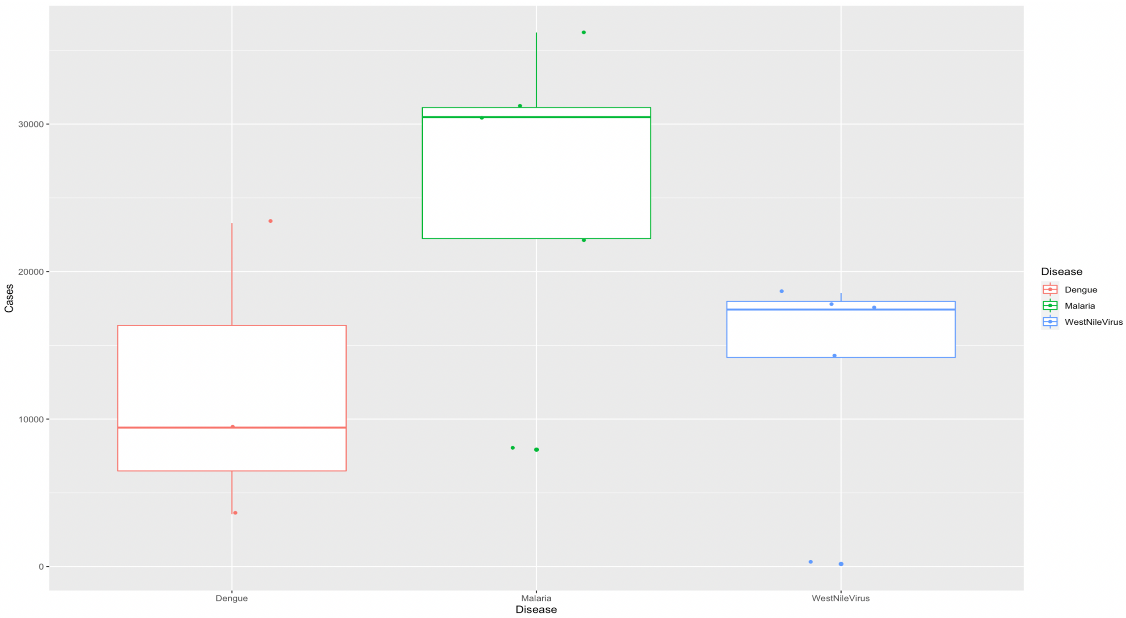

Figure 2.4: Dengue Vs Malaria Vs WNV Box Plot

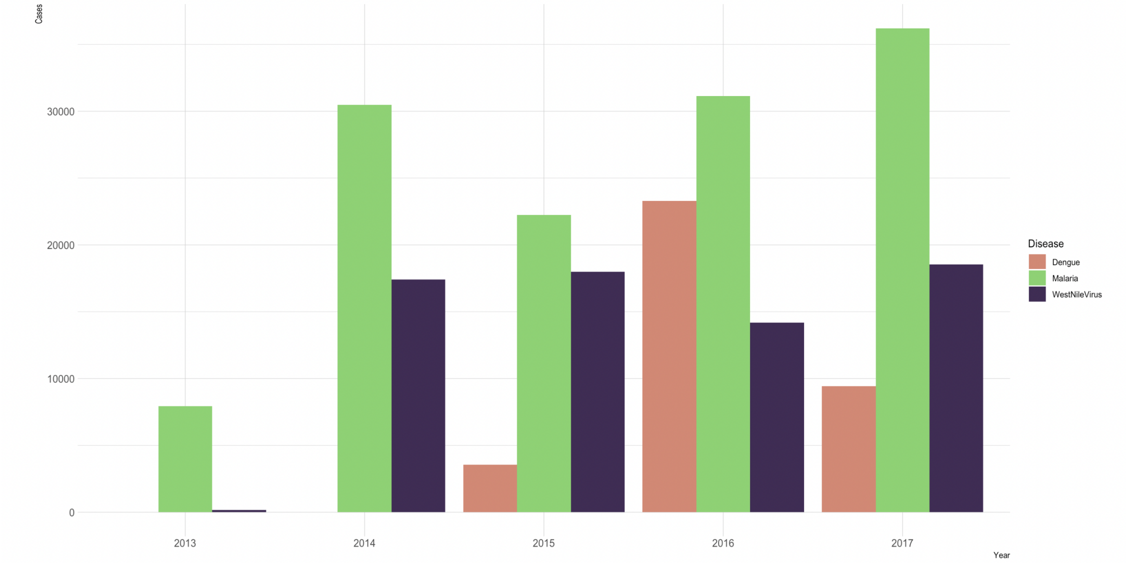

Figure 2.5: Dengue Vs Malaria Vs WNV Grouped Bar Plot

From the above two plots we can see that Malaria has the most number of cases among other vector-borne diseases.

2.3 Impact of Floods

Lets see if floods being a major factor in the increasing cases of vector-borne diseases.

- Dengue

Figure 2.6: Impact of Floods on increase in cases of Dengue

We can see that in 2015 there is no proof of floods in CA hence there are less number of Dengue cases. But in the next year 2016 CA was recorded with highest number of Dengue cases and also there is a indication that flood was occurred.

- Malaria

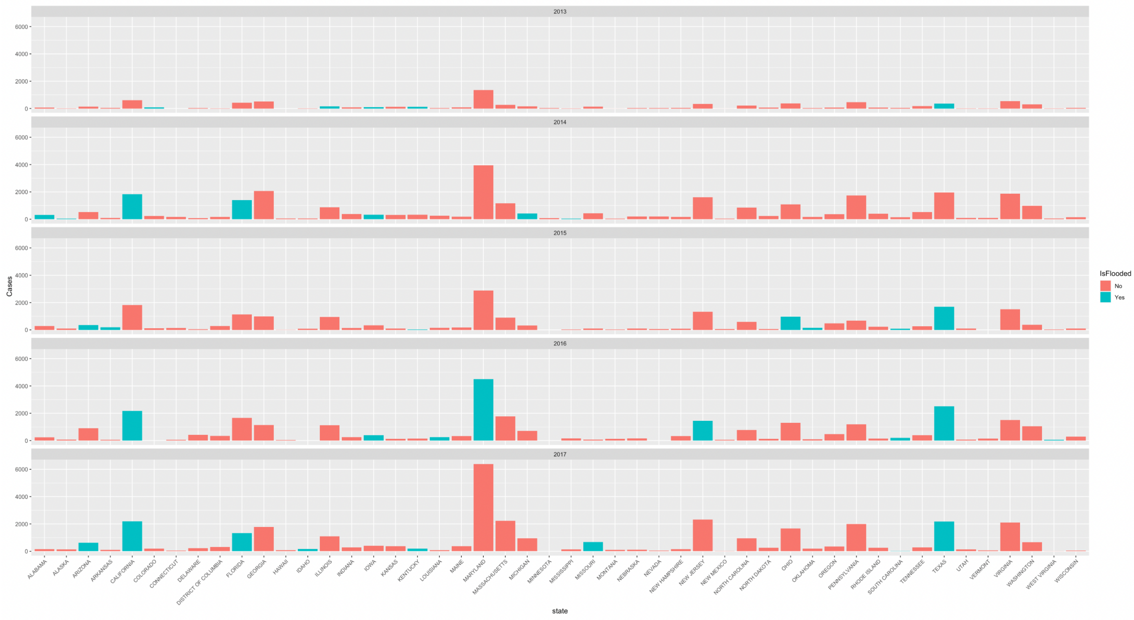

Figure 2.7: Impact of Floods on increase in cases of Malaria

We can see that due to floods in the year 2016 MD has recorded highest number of cases when due to floods. And in CA there is a clear indication of floods which can relate to the rising number of Malaria cases.

- WNV

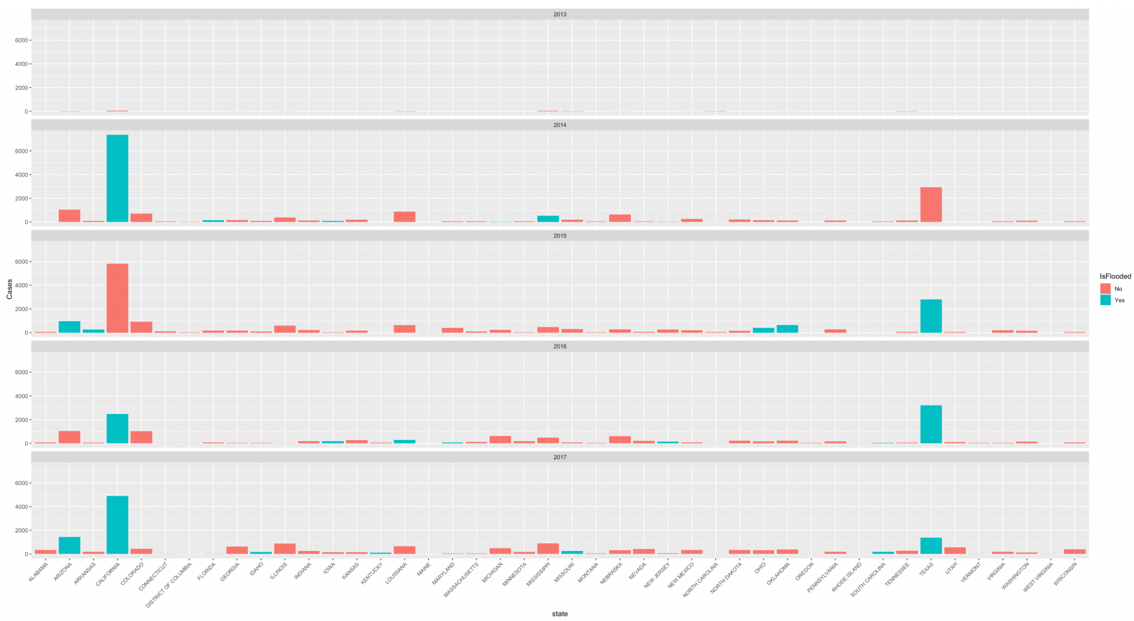

Figure 2.8: Impact of Floods on increase in cases of WNV

We can see that due to floods in CA and TX there is large number of WNV cases of WNV over the years 2013-2014

2.4 Impact of Population

Lets see how population can explain the number of vector-borne diseases cases in the states of USA.

- Dengue

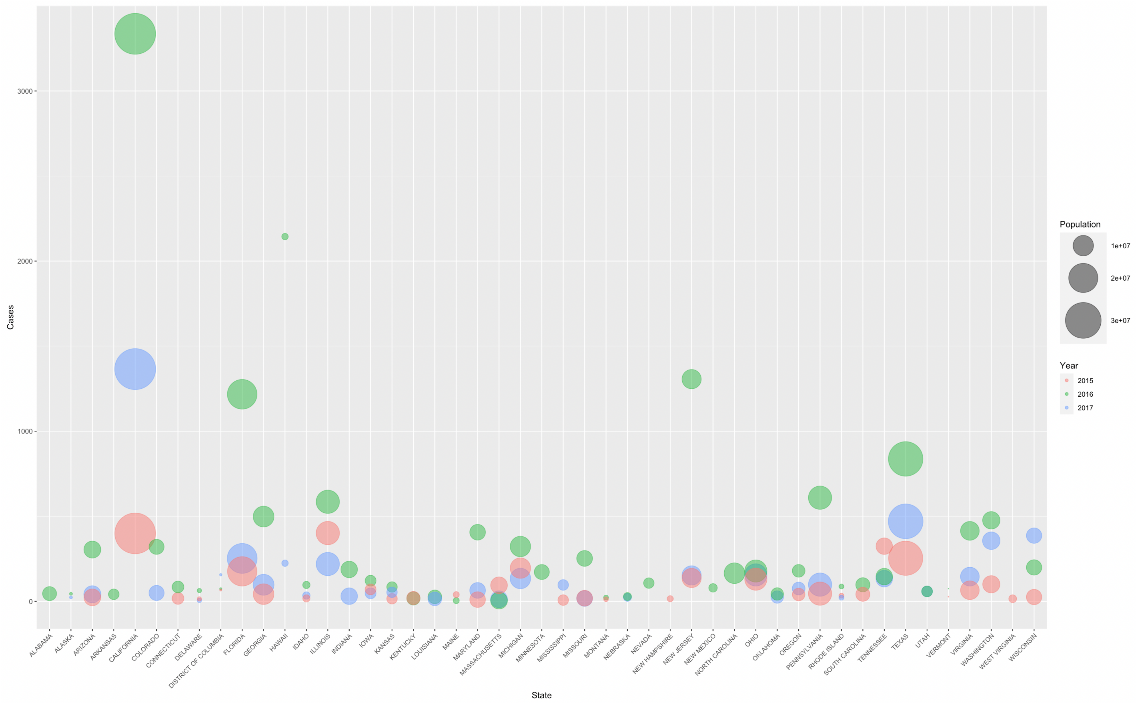

Figure 2.9: Population Vs Dengue

- Malaria

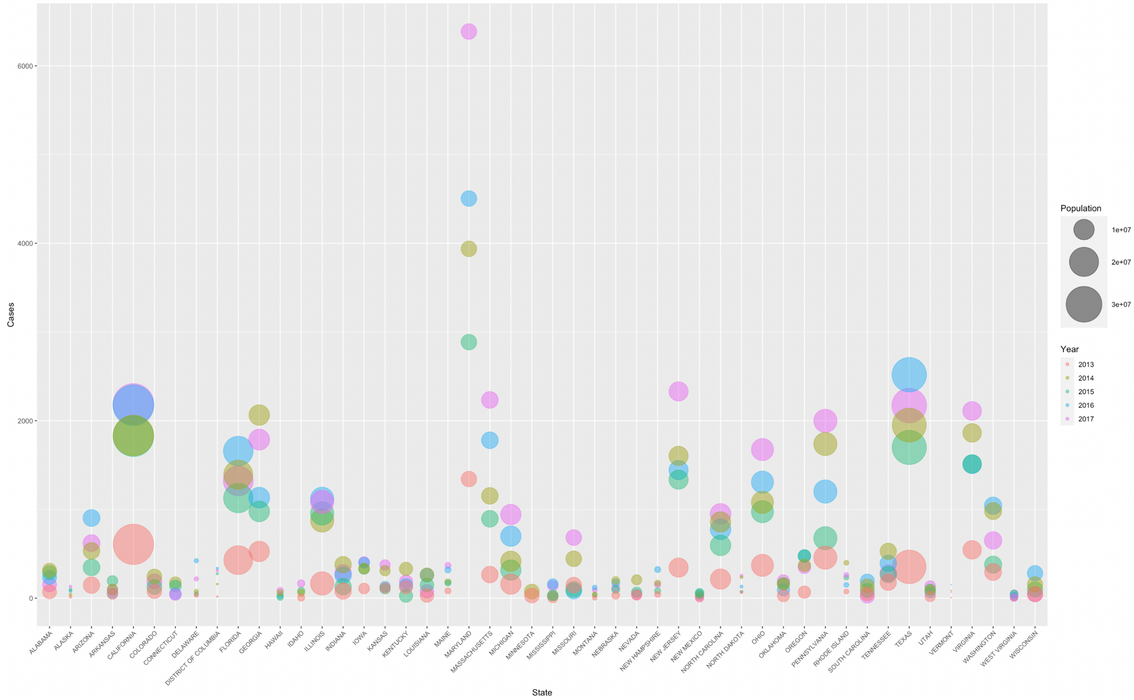

Figure 2.10: Population Vs Malaria

- WNV

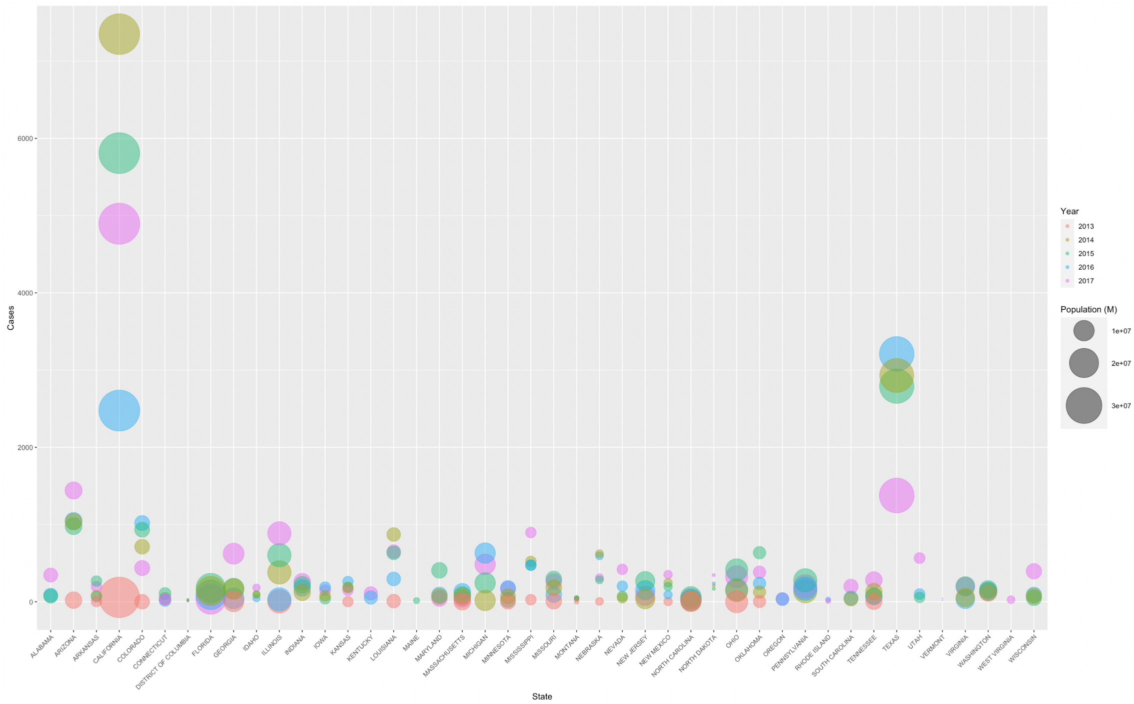

Figure 2.11: Population Vs WNV

From the three plots we can see that CA and TX has highest population and hence the number of people affected with vector-borne disease is also more.

Chapter 3

RESULTS

It can be observed that CA and TX has been reported highest number of vector-borne diseases and also we have seen that these states has been flooded as well as well populated countries.

We have observed that Malaria was recorded with the more number of cases over the other two diseases.

Chapter 4

DISCUSSION AND CONCLUSION

Although there seems to be a huge correlation between number of vector-borne diseases cases with floods and population but there are some exceptions. Such as -

- Even though there is no data of occurrence of floods in MD for the years 2013,2014,2015 and 2016 it has recorded highest number of Malaria cases.
- With the less population MD has recorded highest number of Malaria cases.

These exceptions maybe due to unavailability of data or there are other factors which will be affecting. More data is required in order to get more accurate correlation and results for future work.

Appendix

Source code for the plots

- US Map Chart

```
DengueCases <- DengueDF[c(1,4)]
DengueCases <- setNames(aggregate(DengueCases$Cases,
by=list(DengueCases$StateName),FUN=sum), c("state","Cases"))

plot_usmap(data = DengueCases, values = "Cases",
            color = "#0058F5",labels = TRUE) +
scale_fill_continuous(low = "white",
high = "#0058F5", name = "Number of Cases", label = scales::comma)+
labs(title = "Dengue Cases over the years 2013-2017") +
theme(panel.background=element_blank())
```

- Box Plot

```
ggplot(data = CombinedDF, aes(x=Disease,y=value, color=Disease)) +
  geom_boxplot()+
  scale_fill_brewer(palette="Green") +
  geom_jitter(shape=16, position=position_jitter(0.2))+
  labs(title = 'Dengue VS Malaria VS West Nile Virus',
       y='Cases',x='Disease')
```

- Grouped Bar Plot

```
ggplot(data = CombinedDF, aes(x=Year, y=value, fill=Disease )) +
 geom_bar(stat = "identity", position = "dodge") +
 theme_ipsum() + # Arial Narrow
  scale_fill_ipsum() +
  labs(title = 'Dengue VS Malaria VS West Nile Virus',
       y='Cases',x='Year')
```

- Bar Plot with facet-wrap()

```
ggplot(data = FloodVsDengueDF, aes(x=state, y=Cases, fill=IsFlooded )) +
  geom_bar(stat = "identity", position = "dodge") +
  theme(axis.text.x = element_text(angle = 45, size = 8, vjust = 1,
  hjust = 1))+
  labs(title = "Floods impact on Dengue disease")+
    facet_wrap(~Year,nrow = 5)
```

- Bubble Plot

```
ggplot(aes(x=State, y=Cases, size=Population, color=Year)) +
    geom_point(alpha=0.5) +
  theme(axis.text.x = element_text(angle = 45, size = 8,
  vjust = 1, hjust = 1))+
  labs(title = "Population Vs Dengue")+
    scale_size(range = c(.1, 24), name="Population")
```

Bibliography

[1] Vital Signs: Trends in Reported Vectorborne Disease Cases — United States and Territories, 2004–2016 industry. Link

[2] Project Tycho unlocks global health data to a rapidly growing user community of over 3,000 researchers, students, journalists, officials, and others in over 90 countries.Link

[3] G.R. Brakenridge. Global Active Archive of Large Flood Events. Dartmouth Flood Observatory, University of Colorado, USA. http://floodobservatory.colorado.edu/Archives/ (Accessed 1 October 2016).Link

[4] USA State Population Totals: 2010-2019. Link

[5] Vector Borne Diseases Mapped against Flooding (USA). Link