

Analysis of the zika virus

Project Proposal

1) Intro and Objectives

1.1 Introduction

For my assignment, I've decided to delve into the fascinating world of the Zika virus. This virus poses a significant public health concern, which has led to extensive efforts in gathering and disseminating information about its transmission and impact on different regions. Through my analysis, I aim to broader understande of the Zika virus by carefully examining data from reputable organizations like the European Centre for Disease Prevention and Control (ECDC) and the World Health Organization (WHO).

There are a couple of reasons why I've chosen to embark on this in-depth exploration. Firstly, I hope to uncover intriguing details about the virus's spread across different geographical areas. By studying its movement from place to place, we can gain a better understanding of how it spreads. Secondly, by closely examining the areas that have been affected, we can grasp the extent of the virus's reach and its implications for the people living in those regions.

1.2 Objectives

Within this project, I would like to explore the following:

- Country Transmission
 - Analyze the transmission of the Zika virus in different countries
 - Investigate the categorization of countries based on transmission risk by the World Health Organization (WHO)
- Brazil
 - Focus on Brazil and analyze the data related to the Zika virus.
 - Explore the number of Zika virus cases reported in Brazil.

For this project proposal, my goals are to:

1. Determine Data Requirements: I will assess the amount of data needed to effectively address the research questions and project goals. It's important to consider factors such as data availability, time constraints, and available resources to ensure a feasible and comprehensive analysis.
2. Data Collection: Employ web scraping techniques to collect the necessary data.
3. Data Cleaning: Perform data cleaning and transformation procedures to prepare the collected data for analysis.
4. Explore Data: Conduct exploratory data analysis to gain initial insights into the collected data.
5. Identify Key Findings: Summarize and interpret the key findings derived from the analysis.

1.3 Data

1.3.1 Data requirements

To analyze zika virus infections in various countries, including Brazil, I would need the following types of data:

- ECDC Local Transmission Countries Data: Information about the countries current WHO categories and most affected area in that exact country
- Brazil Transmission Data: Information about the total reported cases in each state of Brazil ##### 1.3.2 Choice of data I have chosen to use CSV files from BuzzFeed News' GitHub repository. These CSV files contain relevant data that I need for my research objectives. CSV format provides a structured and organized way to work with the data.
- <https://github.com/BuzzFeedNews/zika-data>

1.3.3 Limitations and constraints of the data

- Data Availability : It might be hard to find the proper data
- Data Accuracy : Data quality issues can arise, such as missing or incomplete records, inconsistencies, or errors in the data.

1.4 Ethics

1.4.1 Use of data text

- Public available data : The data is already in the public domain. The data is including any relevant author attribution and is publicly accessible.

1.4.3 Reuseage of data

In conducting the analysis of zika virus, considerations for the reuse of data were implemented, including:

- Ensuring compliance with data ownership and licensing terms.
- Respecting privacy and confidentiality by anonymizing sensitive information.
- Obtaining appropriate consent and permissions for data usage.
- Following best practices for data sharing and openness.
- Providing proper attribution and citation for data sources.
- Addressing ethical concerns, such as bias mitigation and harm prevention.

These measures aimed to ensure responsible and ethical data reuse, protecting privacy, upholding legal requirements, and promoting trust in the analysis.

1.4.1 Potential impacts of using data for the proposes analyses

The use of data for the proposed analyses in zika virus can have negative impacts:

- Privacy concerns if personal data is not properly anonymized or protected.
- Potential for biased analysis leading to inequitable outcomes.
- Dependence on data accuracy and quality, which may vary across sources.
- Risk of misinterpretation or misrepresentation of data leading to flawed conclusions.

2) Data Collection

2.1 Getting the data csv file

To collect the data CSV files for my Zika virus analysis, I followed a simple process:

- I started by downloading the entire folder of BuzzFeedNews related to the topic of zikavirus from GitHub. This folder contained various files and data related to the Zika virus.
- Once I had the folder on my computer, I went through its contents to locate the specific data CSV files that I needed for my project.
- I simply moved the data CSV files from the downloaded BuzzFeedNews folder into my midterm folder with my midterm directory.

2.2 Import csv files

I imported two CSV files: "brazil.csv" and "world.csv". To make them easier to work with and understand, I assigned them the variable names "bra" and "wor" respectively.

By assigning meaningful variable names to the imported CSV files, I made it simpler to access and manipulate the data throughout my project. This approach helps ensure clarity and enhances readability

```
In [92]: # Import libraries
import pandas as pd
import matplotlib.pyplot as plt
bra = pd.read_csv('brazil.csv')
wor = pd.read_csv('world.csv')
```

2.3 Test imported csv files head()

In order to verify that the imported CSV files, represented by the variables "bra" and "wor," were functioning correctly and displaying the expected data, I conducted a simple test.

To perform this test, I used the head() function in Python, which allows me to view the first few rows of the DataFrame.

```
In [93]: bra.head()
```

Out[93]:

	no	state	cases_under_investigation	cases_confirmed	cases_discarded	cases_re
0	1	Alagoas	158	0	0	
1	2	Bahia	471	35	27	
2	3	Ceará	218	4	7	
3	4	Maranhão	119	0	15	
4	5	Paraíba	497	31	181	

In [94]: `wor.head()`

Out[94]:

	country	areas	current_zika_transmission
0	American Samoa	NaN	Areas with interrupted transmission (WHO cat. 3)
1	Angola	NaN	Areas with virus transmission following virus ...
2	Anguilla	NaN	Areas with virus transmission following virus ...
3	Antigua and Barbuda	Areas with virus transmission following virus ...	NaN
4	Argentina	Argentina	Areas with virus transmission following virus ...

By conducting this test, I ensured that the imported CSV files were successfully loaded into the variables "bra" and "wor," and that the data within them was accessible for further analysis. This step allowed me to confirm that the data was properly imported and ready for exploration.

3) Data Cleaning

3.1 Remove Columns

I decided to remove three irrelevant columns from the Brazil dataset to create a more focused and clear dataset for my analysis. By eliminating these unnecessary columns, I can avoid confusion and stay focused on the essential variables that are directly relevant to my research objectives. This streamlined approach ensures that I work with the most meaningful data to derive valuable insights from the Brazil dataset.

In [95]: `bra = bra.drop(['no'], axis=1)`

```
In [96]: bra = bra.drop(['cases_under_investigation', 'cases_discarded'], axis=1)
```

After deleting the column, I checked the header of the data to confirm if the columns were successfully removed. This step was important to ensure that the unwanted columns were effectively eliminated from the dataset

```
In [97]: bra.head()
```

```
Out[97]:
```

	state	cases_confirmed	cases_reported_total
0	Alagoas	0	158
1	Bahia	35	533
2	Ceará	4	229
3	Maranhão	0	134
4	Paraíba	31	709

From the world dataset, I have made the decision to remove the "areas" column because it contains mostly empty values and is not necessary for my analysis.

```
In [98]: wor = wor.drop(['areas'], axis=1)
```

After deleting the column, I checked the header of the data to confirm if the columns were successfully removed. This step was important to ensure that the unwanted columns were effectively eliminated from the dataset

```
In [99]: wor.head()
```

```
Out[99]:
```

	country	current_zika_transmission
0	American Samoa	Areas with interrupted transmission (WHO cat. 3)
1	Angola	Areas with virus transmission following virus ...
2	Anguilla	Areas with virus transmission following virus ...
3	Antigua and Barbuda	NaN
4	Argentina	Areas with virus transmission following virus ...

3.2 Remove Rows

I decided to remove irrelevant rows from the Brazil dataset to create a more focused and clear dataset for my analysis. These rows have the data 'Sem registros', which mean no records in portuges.

```
In [100]: bra = bra[bra.apply(lambda row: 'Sem registros' not in row.values, axis=1)]
```

In [101]: `bra`

Out[101]:

	state	cases_confirmed	cases_reported_total
0	Alagoas	0	158
1	Bahia	35	533
2	Ceará	4	229
3	Maranhão	0	134
4	Paraíba	31	709
5	Pernambuco	138	1373
6	Piauí	0	91
7	Rio Grande do Norte	60	208
8	Sergipe	0	172
9	Espírito Santo	0	52
10	Minas Gerais	1	48
11	Rio de Janeiro	0	122
12	São Paulo	0	18
16	Pará	0	6
17	Rondônia	0	1
18	Roraima	0	5
19	Tocantins	0	82
20	Distrito Federal	0	14
21	Goiás	0	62
22	Mato Grosso	0	147
23	Mato Grosso do Sul	0	4
24	Paraná	0	10
25	Santa Catarina	0	1
26	Rio Grande do Sul	1	1

3.3 Missing Values

To ensure the quality of my Brazil and world datasets, I wanted to check if there were any missing values. To do this, I used the `isnull().sum()` command, which helped me identify which columns had missing values and how many were missing in each. This way, I could pinpoint the exact columns that required attention and understand the extent of missing data.

Brazil

```
In [102... bra.isnull().sum()
```

```
Out[102]: state          0
cases_confirmed      0
cases_reported_total  0
dtype: int64
```

My Brazil dataset is complete and does not contain any missing values.

World

```
In [103... wor.isnull().sum()
```

```
Out[103]: country          0
current_zika_transmission  10
dtype: int64
```

I found that my world dataset has some missing values specifically in the "current_zika_transmission" column. To be precise, there are 10 missing values in this column.

Now, my next step is to identify the specific countries that do not have a classification for current Zika transmission from the World Health Organization (WHO). This will allow me to understand which countries lack this crucial information and consider how it may impact my analysis.

```
In [104... missing_values_df = wor[wor['current_zika_transmission'].isnull()]
missing_values_df
```

```
Out[104]:
```

	country	current_zika_transmission
3	Antigua and Barbuda	NaN
46	British Virgin Islands	NaN
53	Central African Republic	NaN
62	Dominican Republic	NaN
112	Lao People's Democratic Republic	NaN
148	Papua New Guinea	NaN
169	Saint Kitts and Nevis	NaN
172	Saint Vincent and the Grenadines	NaN
181	Trinidad and Tobago	NaN
182	Turks and Caicos islands	NaN

I have made the decision to remove these rows. Although there are missing values, their absence will not significantly impact the majority of my analysis.

```
In [105... wor = wor.dropna(subset=['current_zika_transmission'])
```

Now, I proceeded to check if there are any remaining missing values in my world dataset.

```
In [106... wor.isnull().sum()
```

```
Out[106]: country                0
current_zika_transmission      0
dtype: int64
```

There are no more missing values in my world dataset.

3.4 Edit Values

I have decided to make some changes to certain values.

Firstly, I intend to modify the column name "current_zika_transmission" to "WHO_Category." This alteration will result in a more descriptive and meaningful column name that better represents the information it contains. By using "WHO_Category," it becomes easier to understand the purpose and significance of this column in the dataset.

```
In [107... wor = wor.rename(columns={'current_zika_transmission': 'WHO_Category'})
wor.head()
```

```
Out[107]:
```

	country	WHO_Category
0	American Samoa	Areas with interrupted transmission (WHO cat. 3)
1	Angola	Areas with virus transmission following virus ...
2	Anguilla	Areas with virus transmission following virus ...
4	Argentina	Areas with virus transmission following virus ...
5	Argentina	Areas with virus transmission following virus ...

Secondly, I would like to modify the text in the "WHO_Category" column to display only the category number provided by the World Health Organization (WHO). This adjustment aims to simplify the data representation, making it more concise and focused on the essential information.

```
In [108... wor['WHO_Category'] = wor['WHO_Category'].str.extract('(\d+)').astype(float)
wor['WHO_Category'] = wor['WHO_Category'].astype(int)
```

```
In [109... wor.head()
```

```
Out[109]:
```

	country	WHO_Category
0	American Samoa	3
1	Angola	1
2	Anguilla	1
4	Argentina	1
5	Argentina	1

4) Explore Data

4.1 World

As I embark on my analysis, I am excited to delve into the Zika virus data within my world dataset. This initial exploration holds the key to uncovering valuable insights about the countries that are more susceptible to the Zika virus.

4.1.1 WHO Category

The World Health Organization (WHO) categorizes areas with virus transmission into three groups:

- Category 1: Areas where there is active virus transmission following a new or reintroduced virus. These areas pose a significant public health concern as the virus is actively spreading.
- Category 2: Areas where virus transmission occurred in the past but has now been interrupted or temporarily stopped. The risk of active transmission is lower in these areas due to successful control measures.
- Category 3: Areas where virus transmission has been successfully interrupted and there is currently no active circulation. The risk of new infections is minimal in these areas.

This classification helps us understand the status of virus transmission in different areas and guides appropriate public health measures and interventions to prevent further spread.

4.1.2 WHO Category Analysis

I am curious to find out how many countries around the world are currently categorized into each of the three classifications (Category 1, Category 2, and Category 3) established by the World Health Organization (WHO).

To accomplish this, I will thoroughly examine the available data and identify the countries falling into Category 1. These countries are currently experiencing active virus transmission following the introduction or reintroduction of the virus. I will also determine the number of countries falling into Category 2, which refers to areas where virus transmission had occurred previously but has now been successfully interrupted. Lastly, I will identify the countries falling into Category 3, representing areas with no active virus transmission due to successful interruption measures.

```
In [110]: category_counts = wor['WHO_Category'].value_counts()  
category_counts
```

```
Out[110]: 1    127  
          2     52  
          3      4  
          Name: WHO_Category, dtype: int64
```

From the counts, there are 127 countries classified in Category 1, indicating active virus transmission. Additionally, there are 52 countries in Category 2, representing areas with interrupted virus transmission. Lastly, there are 4 countries in Category 3, where virus transmission has been successfully interrupted.

Now, I will proceed to identify the countries that fall under Category 3, where virus transmission has been successfully interrupted.

```
In [111]: category_3_countries = wor[wor['WHO_Category'] == 3]['country']  
category_3_countries
```

```
Out[111]: 0        American Samoa  
          56         Cook Islands  
          168    Saint Barthelemy  
          190           Vanuatu  
          Name: country, dtype: object
```

Among the countries that have successfully interrupted virus transmission, we have American Samoa, Cook Islands, Saint Barthélemy, and Vanuatu. These four countries demonstrate remarkable achievements in effectively stopping the spread of the virus.

I will now check which countries fall under Category 2, where virus transmission has been interrupted.

```
In [112]: category_2_countries = wor[wor['WHO_Category'] == 2]['country']
category_2_countries

Out[112]: 11      Bangladesh
18      Brazil
19      Brazil
20      Brazil
47      Burkina Faso
48      Burundi
49      Cambodia
50      Cameroon
58      Côte d'Ivoire
68      Gabon
74      Haiti
76      India
78      Indonesia
79      Indonesia
80      Indonesia
81      Indonesia
82      Indonesia
83      Indonesia
84      Indonesia
85      Indonesia
86      Indonesia
87      Indonesia
88      Indonesia
89      Indonesia
90      Indonesia
91      Indonesia
92      Indonesia
93      Indonesia
94      Indonesia
95      Indonesia
96      Indonesia
97      Indonesia
98      Indonesia
99      Indonesia
100     Indonesia
101     Indonesia
102     Indonesia
103     Indonesia
104     Indonesia
105     Indonesia
106     Indonesia
107     Indonesia
108     Indonesia
109     Indonesia
110     Indonesia
113     Malaysia
145     Nigeria
166     Philippines
174     Senegal
179     Thailand
183     Uganda
192     Vietnam
Name: country, dtype: object
```

I noticed the presence of recurring countries. To eliminate duplicate entries, I will undertake the necessary steps to clean up the dataset and remove these duplicate country records.

```
In [113... wor = wor.drop_duplicates(subset='country')
```

I recognized the need to reevaluate the count of countries in each category. This will provide accurate information and allow me to determine the precise number of countries in each category, as defined earlier.

```
In [114... category_counts = wor['WHO_Category'].value_counts()
category_counts
```

```
Out[114]: 1    52
          2    17
          3     4
          Name: WHO_Category, dtype: int64
```

After the reassessment, I see that there are now 17 countries in Category 2. Additionally, there are still 4 countries remaining in Category 3.

```
In [115... category_2_countries = wor[wor['WHO_Category'] == 2]['country']
category_2_countries
```

```
Out[115]: 11    Bangladesh
          18    Brazil
          47    Burkina Faso
          48    Burundi
          49    Cambodia
          50    Cameroon
          58    Côte d'Ivoire
          68    Gabon
          74    Haiti
          76    India
          113   Malaysia
          145   Nigeria
          166   Philippines
          174   Senegal
          179   Thailand
          183   Uganda
          192   Vietnam
          Name: country, dtype: object
```

The following countries fall into Category 2, indicating interrupted virus transmission. These countries have experienced a reduction in virus transmission

I will now check which countries fall under Category 1, which signifies areas with ongoing virus transmission.

```
In [116... category_1_countries = wor[wor['WHO_Category'] == 1]['country']
category_1_countries
```

```
Out[116]: 1          Angola
           2          Anguilla
           4          Argentina
           8          Argentina
           9          Aruba
          10          Bahamas
          12          Barbados
          13          Belize
          14          Bolivia
          15  Bonaire, Sint Eustatius and Saba
          51          Cape Verde
          52          Cayman Islands
          54          Chile
          55          Colombia
          57          Costa Rica
          59          Cuba
          60          Curaçao
          61          Dominica
          63          Ecuador
          64          El Salvador
          65          Fiji
          66          France
          69          Grenada
          70          Guadeloupe
          71          Guatemala
          72          Guinea-Bissau
          73          Guyana
          75          Honduras
          77          Indonesia
         111          Jamaica
         114          Maldives
         115          Marshall Islands
         116          Mexico
         142          Micronesia
         143          Montserrat
         144          Nicaragua
         146          Palau
         147          Panama
         149          Paraguay
         165          Peru
         167          Puerto Rico
         170          Saint Lucia
         171          Saint Martin
         173          Samoa
         175          Singapore
         176          Sint Maarten
         177          Solomon Islands
         178          Suriname
         180          Tonga
         184          United States of America
         189          United States Virgin Islands
         191          Venezuela
Name: country, dtype: object
```

The following countries fall into Category 3, indicating areas with ongoing virus transmission.

I will now create a visualization that displays the categories, from 1 to 3, alongside the corresponding countries. This visual representation will provide a clear and comprehensive overview, allowing us to easily identify which country falls into each category.

```
In [117]: category_1_countries.reset_index(drop=True, inplace=True)
category_2_countries.reset_index(drop=True, inplace=True)
category_3_countries.reset_index(drop=True, inplace=True)

#Change column text from country to Category
country_1 = pd.concat([category_1_countries], axis=1)
country_1 = country_1.rename(columns={'country': 'Category 1'})

country_2 = pd.concat([category_2_countries], axis=1)
country_2 = country_2.rename(columns={'country': 'Category 2'})

country_3 = pd.concat([category_3_countries], axis=1)
country_3 = country_3.rename(columns={'country': 'Category 3'})

world = pd.concat([country_1, country_2, country_3], axis=1)
world
```

```
Out[117]:
```

	Category 1	Category 2	Category 3
0	Angola	Bangladesh	American Samoa
1	Anguilla	Brazil	Cook Islands
2	Argentina	Burkina Faso	Saint Barthelemy
3	Argentina	Burundi	Vanuatu
4	Aruba	Cambodia	NaN
5	Bahamas	Cameroon	NaN
6	Barbados	Côte d'Ivoire	NaN
7	Belize	Gabon	NaN
8	Bolivia	Haiti	NaN
9	Bonaire, Sint Eustatius and Saba	India	NaN
10	Cape Verde	Malaysia	NaN
11	Cayman Islands	Nigeria	NaN
12	Chile	Philippines	NaN
13	Colombia	Senegal	NaN
14	Costa Rica	Thailand	NaN
15	Cuba	Uganda	NaN

16	Curaçao	Vietnam	NaN
17	Dominica	NaN	NaN
18	Ecuador	NaN	NaN
19	El Salvador	NaN	NaN
20	Fiji	NaN	NaN
21	France	NaN	NaN
22	Grenada	NaN	NaN
23	Guadeloupe	NaN	NaN
24	Guatemala	NaN	NaN
25	Guinea-Bissau	NaN	NaN
26	Guyana	NaN	NaN
27	Honduras	NaN	NaN
28	Indonesia	NaN	NaN
29	Jamaica	NaN	NaN
30	Maldives	NaN	NaN
31	Marshall Islands	NaN	NaN
32	Mexico	NaN	NaN
33	Micronesia	NaN	NaN
34	Montserrat	NaN	NaN
35	Nicaragua	NaN	NaN
36	Palau	NaN	NaN
37	Panama	NaN	NaN
38	Paraguay	NaN	NaN
39	Peru	NaN	NaN
40	Puerto Rico	NaN	NaN
41	Saint Lucia	NaN	NaN
42	Saint Martin	NaN	NaN
43	Samoa	NaN	NaN
44	Singapore	NaN	NaN
45	Sint Maarten	NaN	NaN
46	Solomon Islands	NaN	NaN
47	Suriname	NaN	NaN
48	Tonga	NaN	NaN
49	United States of America	NaN	NaN

50	United States Virgin Islands	NaN	NaN
51	Venezuela	NaN	NaN

4.2 Brazil

4.2.1 Bubble Chart

As I delve into my analysis of Brazil, I begin by creating a bubble chart to visually depict the confirmed cases and reported cases in each state. This visualization will provide a clear and intuitive representation, enabling me to grasp the extent and distribution of the Zika virus impact throughout different states in Brazil.

- My first step is to convert the columns to numeric values.

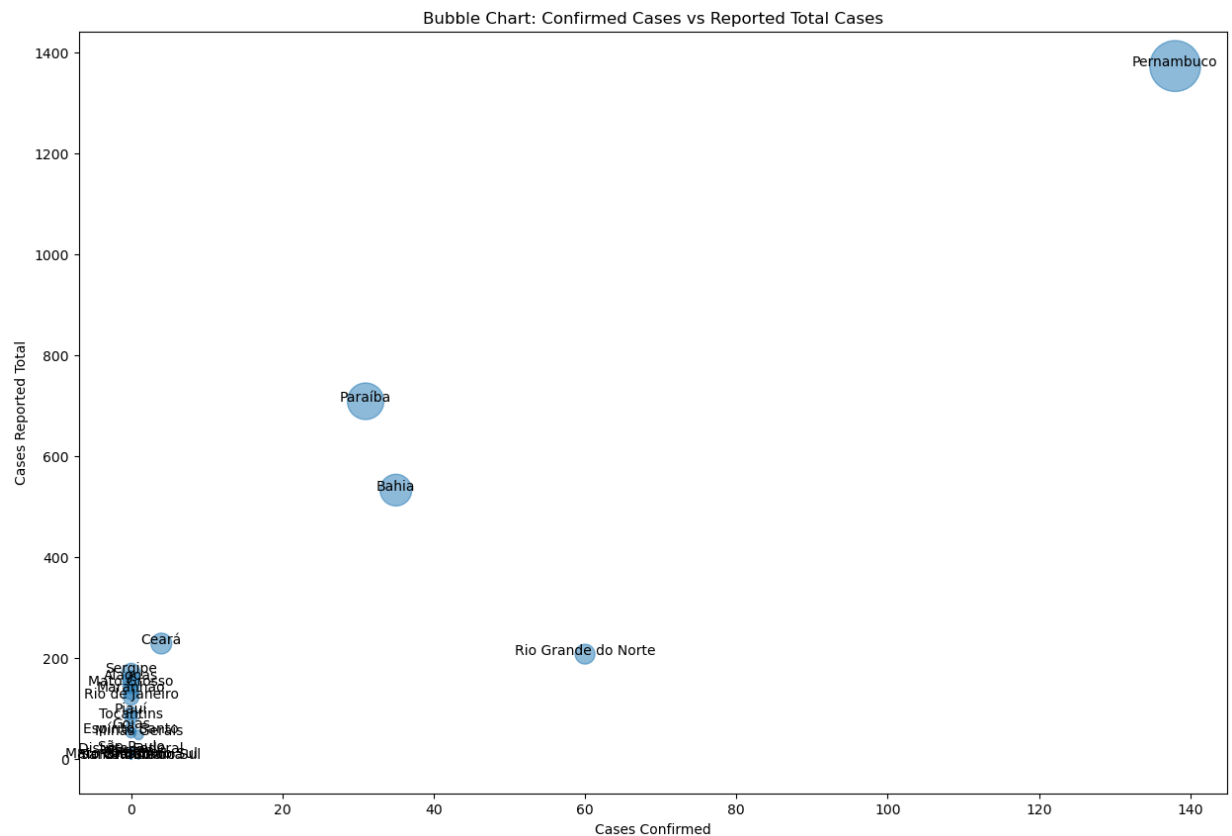
```
In [118... bra['cases_confirmed'] = pd.to_numeric(bra['cases_confirmed'], errors='co
bra['cases_reported_total'] = pd.to_numeric(bra['cases_reported_total'],
```

- My second step is to create the Bubble Chart

```
In [119... plt.figure(figsize=(15, 10))
plt.scatter(bra['cases_confirmed'], bra['cases_reported_total'], s=bra['c

for i in range(bra.shape[0]):
    plt.text(bra['cases_confirmed'].iloc[i], bra['cases_reported_total'].

plt.xlabel('Cases Confirmed')
plt.ylabel('Cases Reported Total')
plt.title('Bubble Chart: Confirmed Cases vs Reported Total Cases')
plt.show()
```



Upon analyzing the bubble chart, it becomes apparent that not all states have reported any confirmed cases. It is quite striking to see that a majority of the states have yet to confirm any cases. This observation raises concerns about the effectiveness of testing efforts in the country, suggesting a potential lack of widespread testing coverage.

Moreover, the bubble chart highlights Pernambuco state as having the highest number of both reported and confirmed cases. It is worth noting that the bubble representing Pernambuco is visibly larger compared to other states, indicating a greater magnitude of cases. This visual correlation between bubble size and reported cases suggests that a higher number of reported cases often corresponds to a higher likelihood of confirmed cases.

By closely examining the bubble chart, we gain valuable insights into the distribution and relationship between reported and confirmed cases across different states. These findings not only raise questions about testing strategies and coverage but also provide a basis for further investigation and understanding of the Zika virus situation in Brazil.

4.2.2 Cases

To better comprehend which states have the highest number of cases, I will split the "cases-confirmed" and "cases-reported-total" columns into separate tables. By arranging the data in ascending order, I can easily identify the states with the most cases.

- Total Confirmed Cases in Brazil

```
In [120...] bra_sorted_con = bra.sort_values(by='cases_confirmed', ascending=False)
```

```
In [121...] bra_sorted_con = bra_sorted_con.drop(['cases_reported_total'], axis=1)
bra_sorted_con = bra_sorted_con[bra_sorted_con['cases_confirmed'] != 0]
```

```
In [122...] bra_sorted_con
```

```
Out[122]:
```

	state	cases_confirmed
5	Pernambuco	138
7	Rio Grande do Norte	60
1	Bahia	35
4	Paraíba	31
2	Ceará	4
26	Rio Grande do Sul	1
10	Minas Gerais	1

Here you can see the confirmed zika virus cases in brazil in different states.

```
In [123...] bra_sorted_con_sum = bra_sorted_con['cases_confirmed'].sum()
bra_sorted_con_sum
```

```
Out[123]: 270
```

Now the total confirmed case in brazil is 270.

- Total Reported Cases in Brazil

```
In [124...] bra_sorted_rep = bra.sort_values(by='cases_reported_total', ascending=False)
bra_sorted_rep = bra_sorted_rep.drop(['cases_confirmed'], axis=1)
```

```
In [125...] bra_sorted_rep
```

Out [125]:

	state	cases_reported_total
5	Pernambuco	1373
4	Paraíba	709
1	Bahia	533
2	Ceará	229
7	Rio Grande do Norte	208
8	Sergipe	172
0	Alagoas	158
22	Mato Grosso	147
3	Maranhão	134
11	Rio de Janeiro	122
6	Piauí	91
19	Tocantins	82
21	Goiás	62
9	Espírito Santo	52
10	Minas Gerais	48
12	São Paulo	18
20	Distrito Federal	14
24	Paraná	10
16	Pará	6
18	Roraima	5
23	Mato Grosso do Sul	4
17	Rondônia	1
25	Santa Catarina	1
26	Rio Grande do Sul	1

Here you can see the reported zika virus cases in brazil in different states.

```
In [126.. bra_sorted_rep_sum = bra_sorted_rep['cases_reported_total'].sum()
bra_sorted_rep_sum
```

Out [126]: 4180

Now the total reported case in brazil is 4180.

5) Summary

5.1 Conclusion

After thoroughly examining the Zika virus data, I have reached some important conclusions regarding its impact on the world and Brazil:

Testing Discrepancy: It is notable that there is a significant difference between the total reported Zika virus cases (4,180) and the total confirmed cases (270) in Brazil. This suggests a potential issue with testing, as the positivity rate appears to be relatively low. Further investigation is necessary to understand and address this testing discrepancy, with the aim of improving the accuracy of case detection.

Identification of Hotspot: Pernambuco has emerged as a prominent hotspot for Zika virus transmission in Brazil. It stands out with the highest number of reported and confirmed cases, indicating a higher risk of contracting the virus in this particular region. This highlights the importance of targeted interventions and dedicated public health efforts to effectively manage the situation in areas experiencing a higher incidence of infection.

Efficiency in Testing: The state of Rio Grande do Norte exhibits a relatively higher positivity rate for confirmed cases, suggesting more efficient testing and diagnostic capabilities, particularly in urban regions. This underscores the significance of having robust testing facilities and sufficient resources in promptly identifying and confirming Zika virus cases.

In conclusion, it is crucial to improve testing protocols, implement targeted interventions in high-risk areas, and allocate adequate resources to enhance testing capabilities across Brazil. By addressing these aspects, we can enhance the accuracy of case detection, effectively manage the spread of the Zika virus, and ensure timely care for those affected.

5.2 Summary of data

5.2.1 World

In summary, the analysis reveals the following significant insights:

- Category 1: There are 51 countries experiencing active virus transmission. This highlights the widespread nature of the virus and its impact on various regions.
- Category 2: Notably, 16 countries have successfully interrupted virus transmission and significantly reduced infection rates. This underscores the effectiveness of control measures and interventions in combating the virus.
- Category 3: Four countries have successfully eliminated the virus, showcasing the potential for complete eradication. However, it is worth noting that these countries are relatively small in size. This suggests that completely wiping out the virus can be a challenging endeavor, yet it instills a sense of hope and serves as a testament to the possibility of victory.

These findings underscore the global efforts to control and combat the Zika virus. While challenges persist, the successes achieved in interrupting transmission and eliminating the virus in certain regions provide inspiration and encouragement to persevere in the fight against this infectious disease.

5.2.2 Brazil

After analyzing the Brazil dataset, several significant findings have come to light. The total reported Zika virus cases in Brazil amount to 4,180, while only 270 cases have been confirmed. This stark contrast indicates a potential issue with testing, as the positivity rate is relatively low. It raises concerns about the accuracy and effectiveness of the testing process.

A closer examination of the "confirmed" and "reported" tables reveals that Pernambuco is the region most affected by the Zika virus, serving as a hotspot for its transmission in Brazil. The data also highlights that Rio Grande do Norte has a higher positivity rate for confirmed cases compared to other states. This could be attributed to better testing facilities and resources in urban areas, allowing for more accurate identification and confirmation of Zika virus cases.

These findings emphasize the importance of swift and accurate testing to promptly diagnose Zika virus cases. By doing so, it enables timely care and interventions for patients, potentially saving lives. Improving testing capabilities across the country is crucial to effectively manage the spread of the virus and provide appropriate healthcare to those affected.

6) Resources

6.1 Resources

Data

<https://github.com/BuzzFeedNews/zika-data>