

**TEAM 11**

* 20BDA05 RESHMA DUA
* 20BDA21 JOEL BHARAT MONIS
* 20BDA35 NIDHI TERESA GEORGE
* 20BDA47 RAKSHITH KN

THE GLOBAL TERRORISM DATABASE

Table of Contents

[1 Introduction 2](#_Toc57670113)

[2 Characteristics Of The Global Terrorism Database Dataset. 2](#_Toc57670114)

[3 Problem Statement 2](#_Toc57670115)

[4 Data Collection Methodology Used 2](#_Toc57670116)

[5 Data Preprocessing 3](#_Toc57670117)

[6 Visualisation 3](#_Toc57670118)

[7 Conclusion. 14](#_Toc57670119)

8. Appendix………………………………….…………………………………………………………….15

[Bibliography 19](#_Toc57670120)

# 1 iNTRODUCTION

The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 180,000 attacks. For each event, a wide range of information is available, including the date and location of the incident, the weapons used, nature of the target, the number of casualties, and – when identifiable – the group or individual responsible. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland and makes the GTD available online in an effort to increase understanding of terrorist violence so that it can be more readily studied and defeated.

# 2 Characteristics of the Global Terrorism Database Dataset.

* Contains information on over 180,000 terrorist attacks
* Currently the most comprehensive unclassified database on terrorist attacks in the world
* Includes information on more than 95,000 bombings, 20,000 assassinations, and 15,000 kidnappings and hostage events since 1970
* Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables
* More than 4,000,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2017 alone

# 3 PROBLEM STATEMENT

Analysis and visualisation of the Global terrorism events to interpret occurrences of terrorist activities.

# 4 DATA COLLECTION METHODOLOGY USED

(a) The ongoing collection has made significant improvements to the methodology used to compile the database since Apr 2012, balancing the strengths of artificial and human intelligence. The process begins with a more diverse set of news media sources from around the world for identifying and documenting the incidents that are included in the GTD—more than two million articles published daily. Natural language processing, named entity extraction, and machine learning models facilitate the identification and organization of news articles that include information about terrorist attacks. The GTD team developed a proprietary Data Management System that allows analysts to identify unique attacks, record the details of each event, and update records for previously recorded events as new information becomes available.

(b) Key factors that impact the content of datasets like the GTD include definitions, sources, and workflows. While the GTD team has applied a single definition of terrorism over the full span of the database, access to source materials and the efficiency of workflows have varied over time. For example, the availability of source materials was best at times when the data collection had the shortest lag behind real time. Improvements in technology and the expansion of the internet have enhanced both the accessibility of source materials and the efficiency of workflows. In general, users should interpret trends over time with caution. In particular, differences in levels of attacks and casualties before and after January 1, 1998; April 1, 2008; and January 1, 2012 may be partially explained by shifts in data collection. Likewise, the GTD is not directly comparable to other sources of data on terrorism, as differences in definitions, sources, and workflows will produce misleading inferences.

# 5 Data Preprocessing

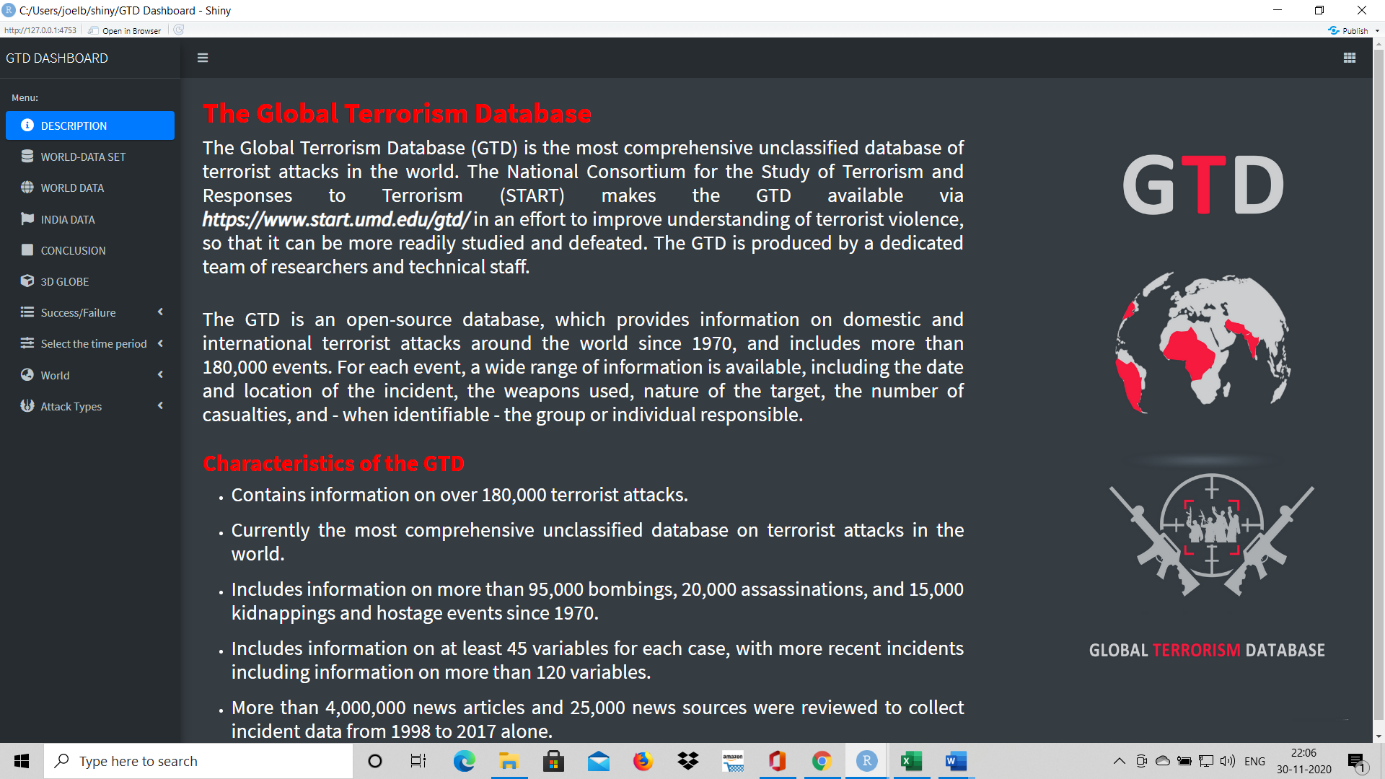
Of the 165 fields in the database, there were many which did not have data in most or all the records. There were 70 such fields. These fields were removed. The list of these fields and the number of absent records is placed at Appendix A. The remaining data was studied to explore the methods of analysis and visualisation.

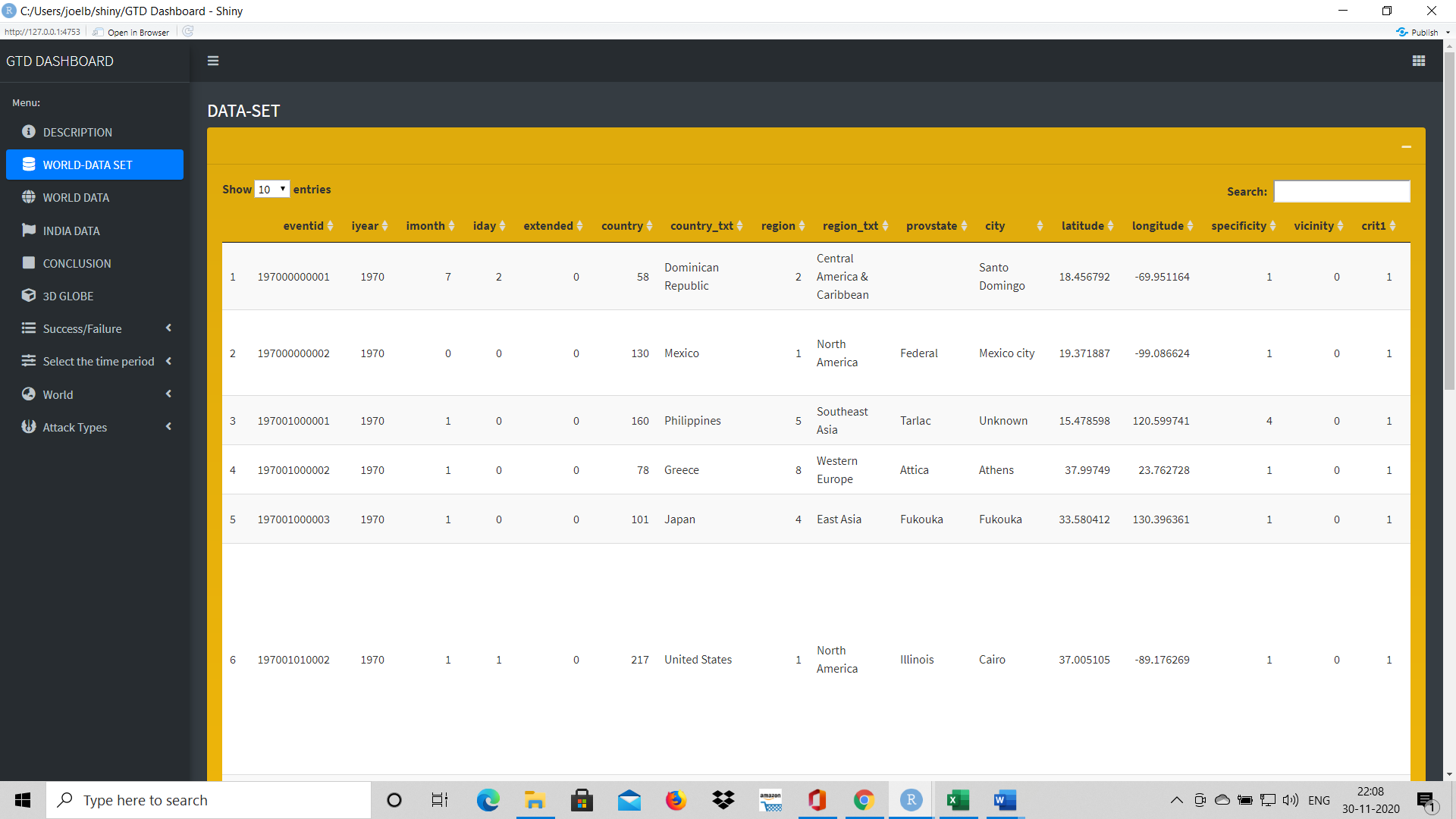
# 6 visualisation

(a) Visualisation was broadly divided into two categories – World and India. Data on terrorism was represented for these categories. The variables for which visualisation was prepared are the time duration, success/failure of attacks, selection of worldwide zones, Types of attacks, Methods /Weapons used and Terrorist groups.

(b) The visualisations were organized into dashboards and presented in a manner which would augur logical deductions. Screenshots of the dashboards are given in the subsequent paragraphs.

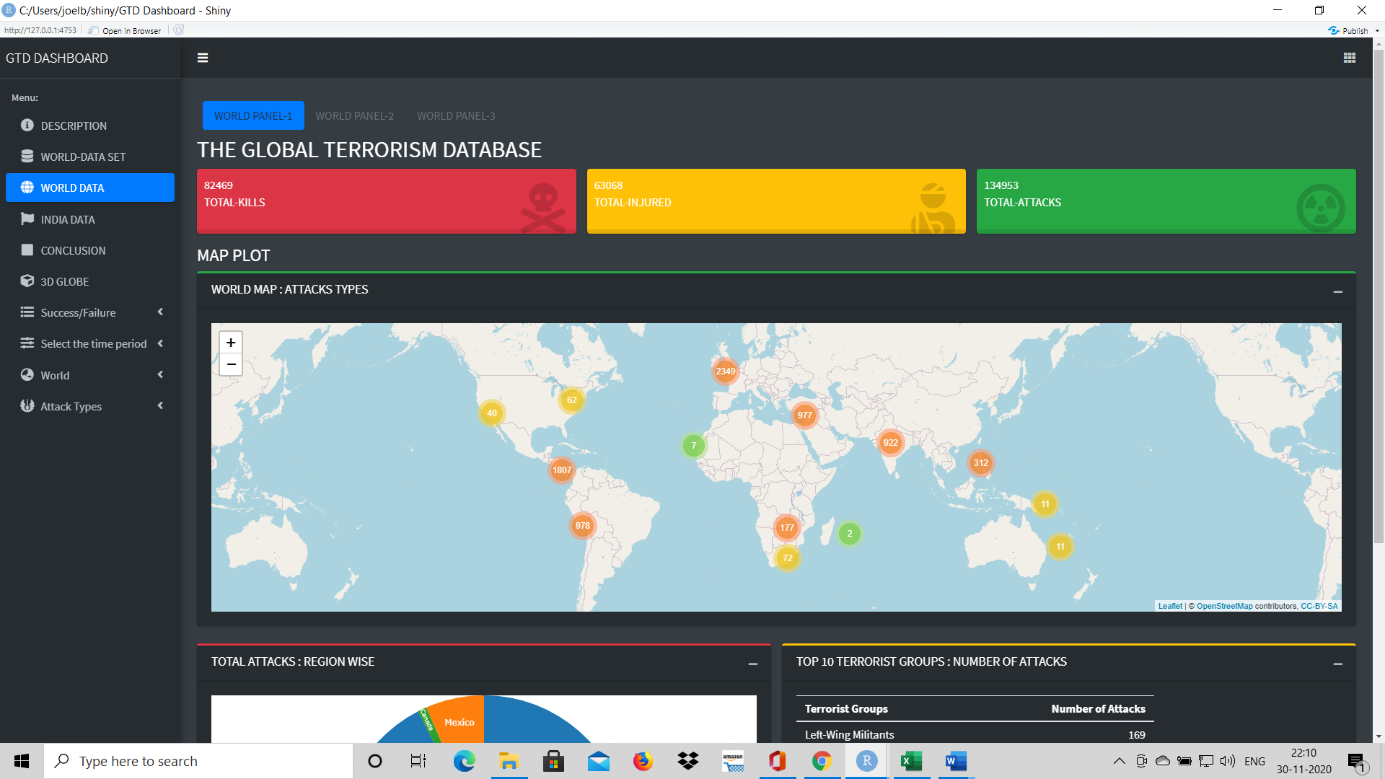
(c) Opening Page.



(d) The World Data Set.

(e) World Data Panel 1.

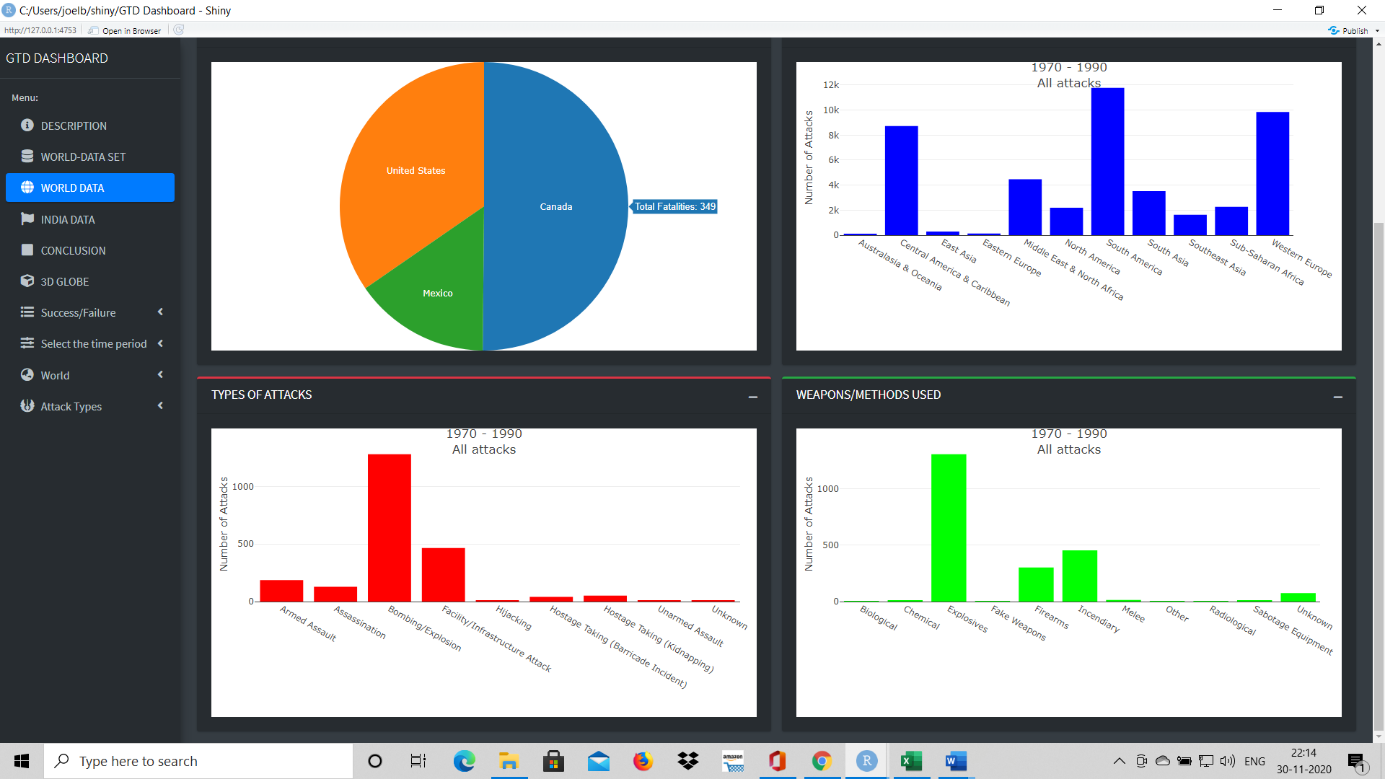
1. Part 1.

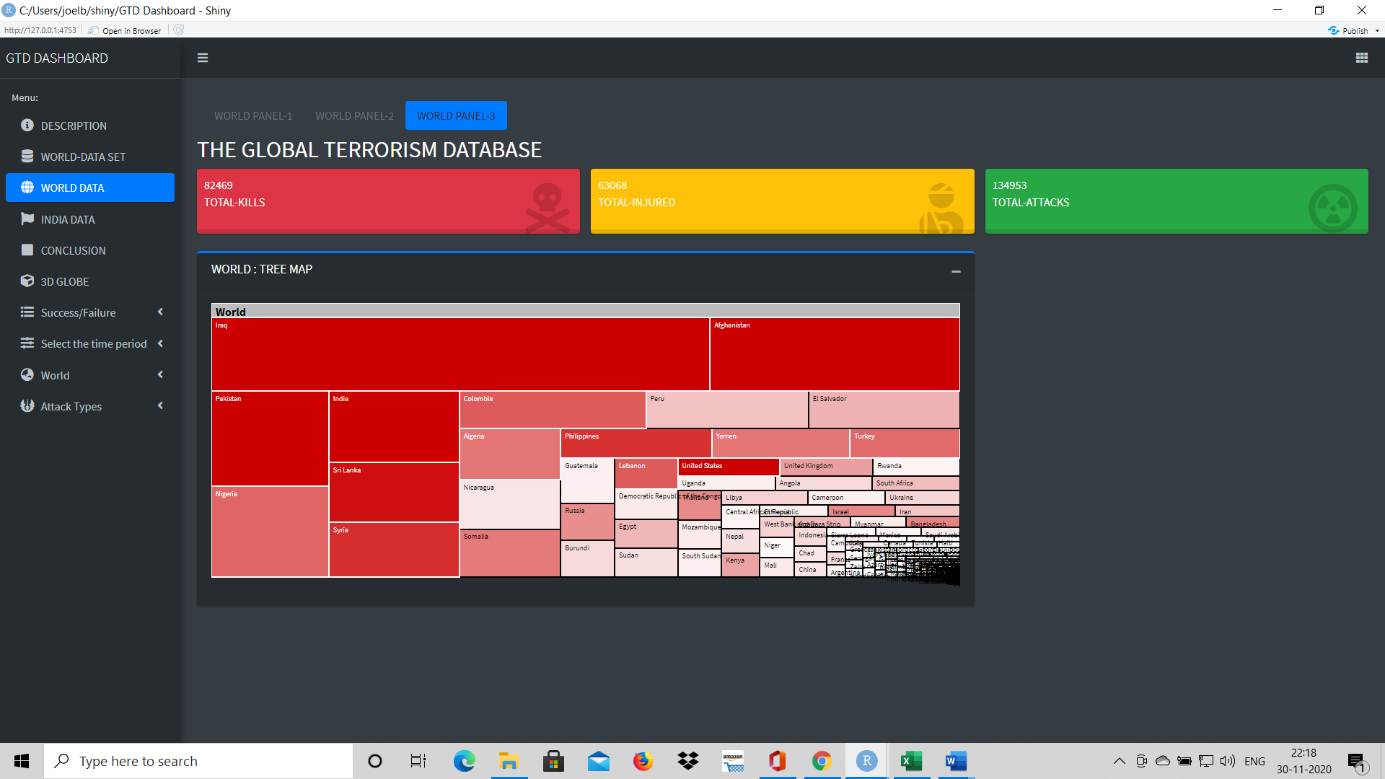


1. Part 2.

(f) World Data Panel 2

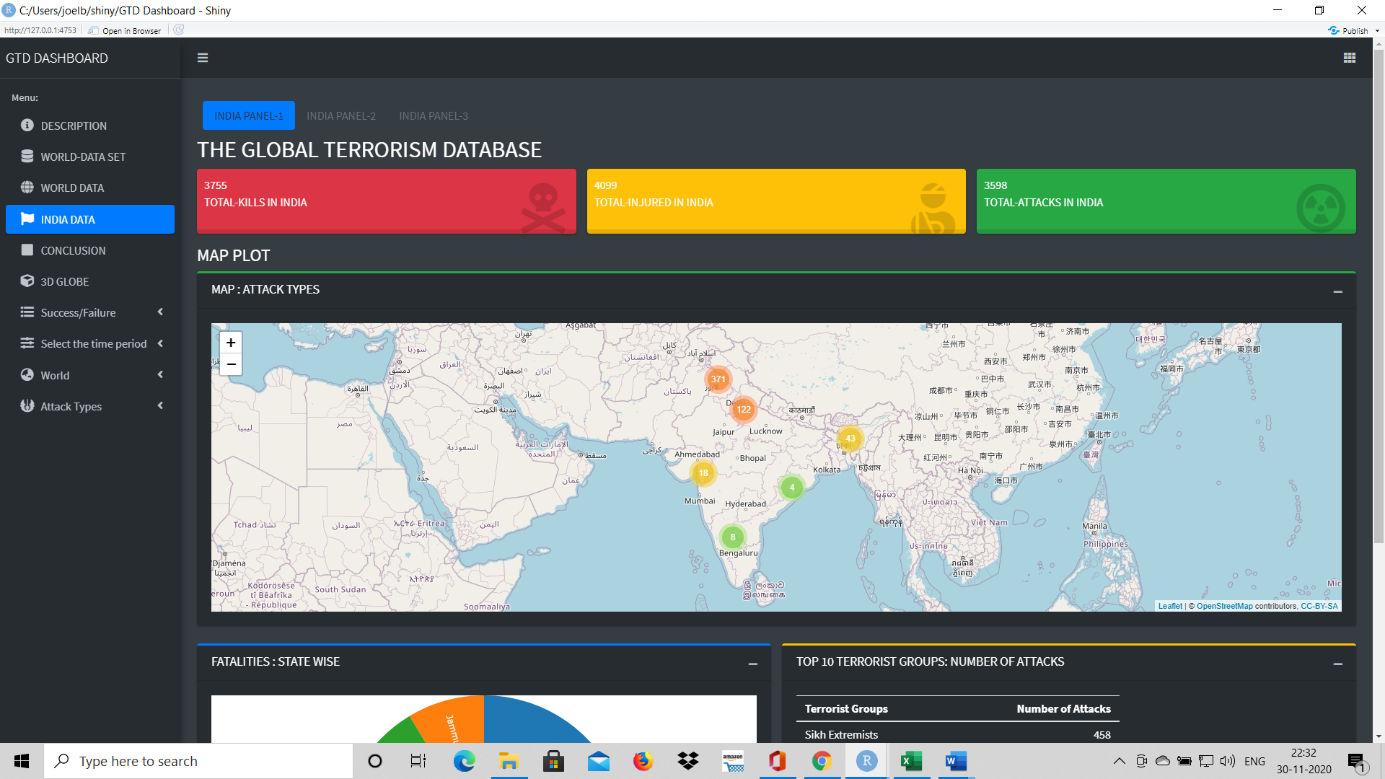
1. Part 1.
2. Part 2.



(g) World Data Panel 3.

(h) India Data : Panel 1.

1. Part 1.

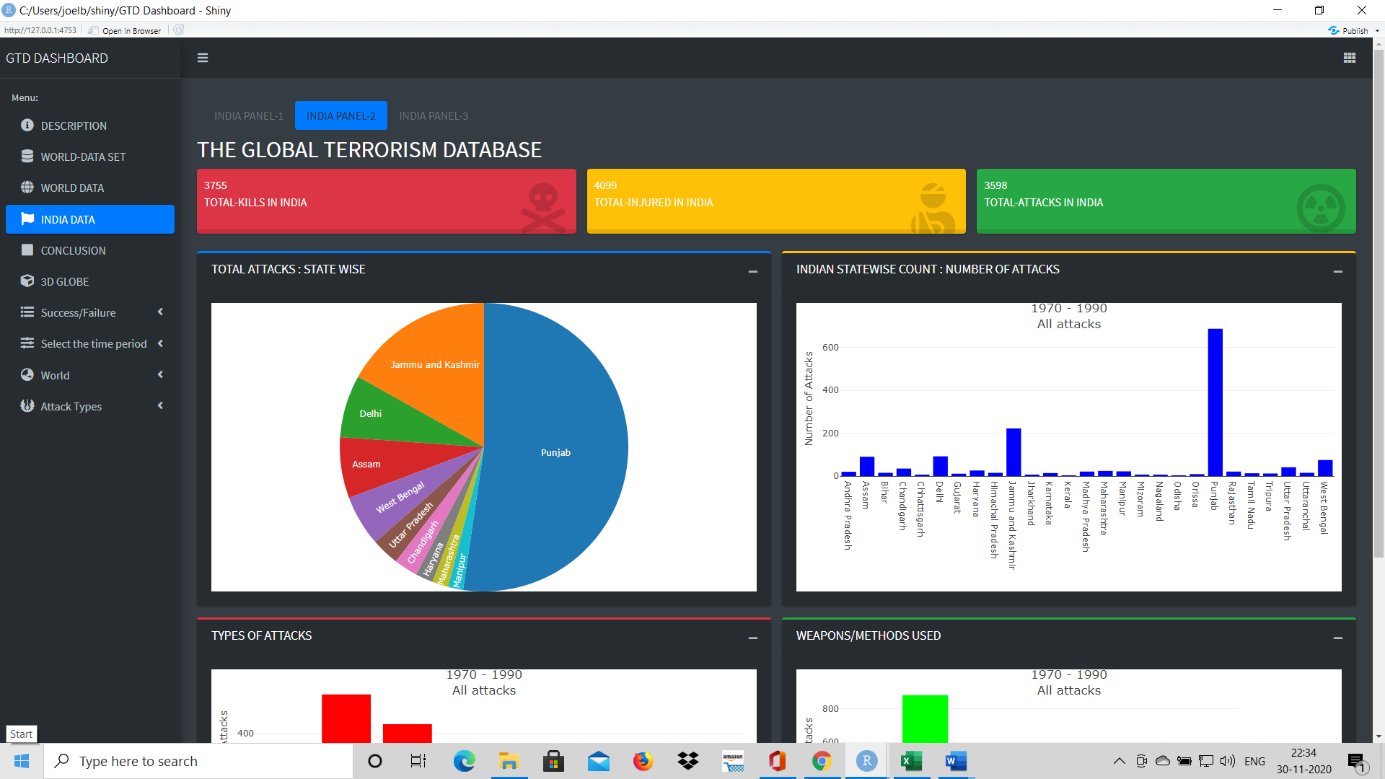


1. Part 2.

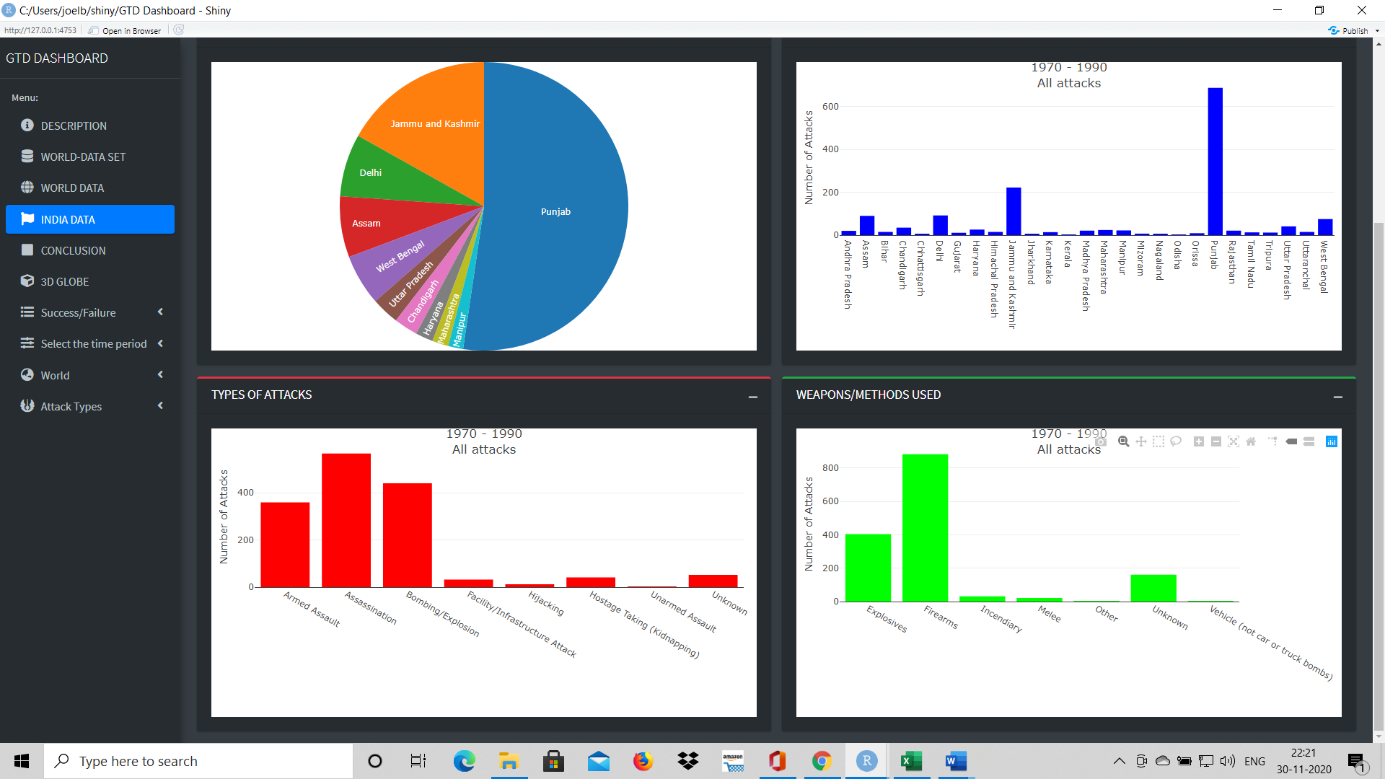


(j) India Data : Panel 2.

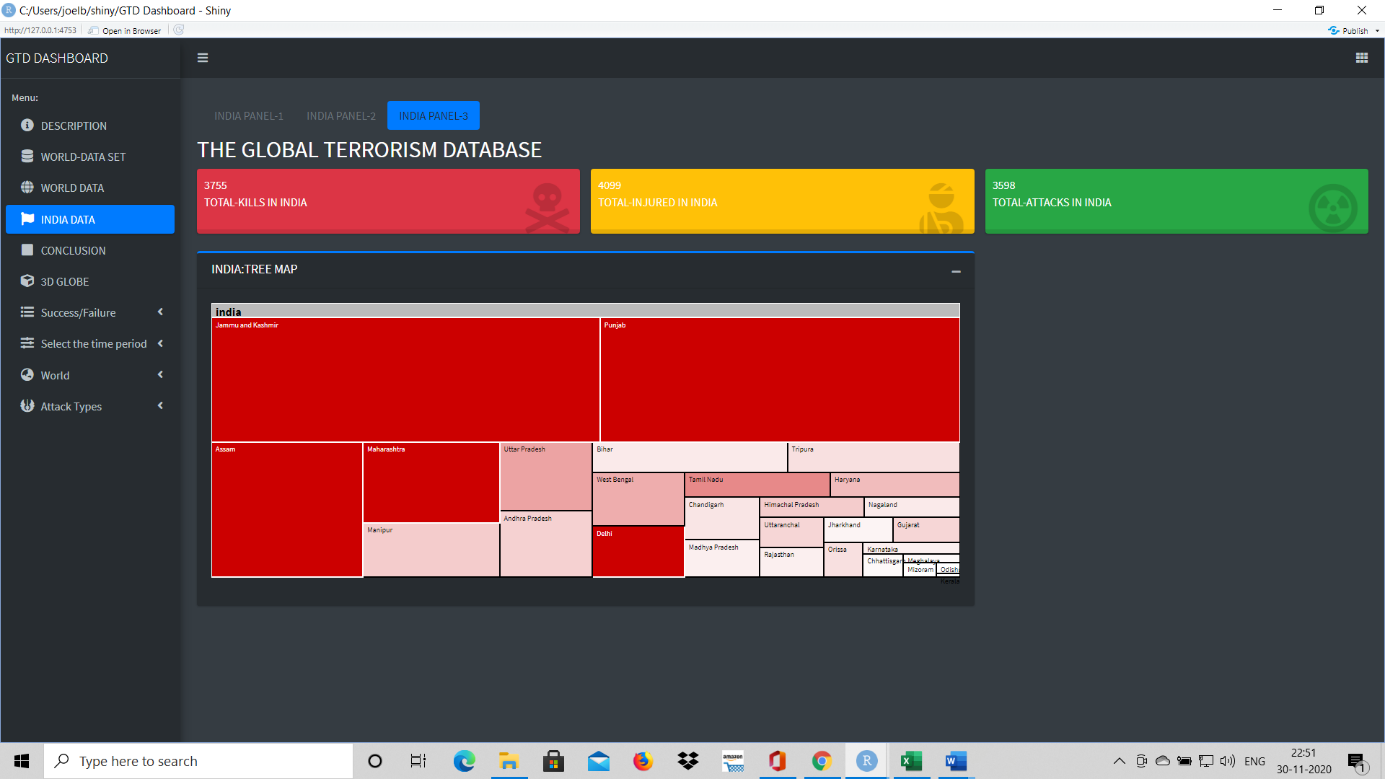
1. Part 1.



1. Part 2.



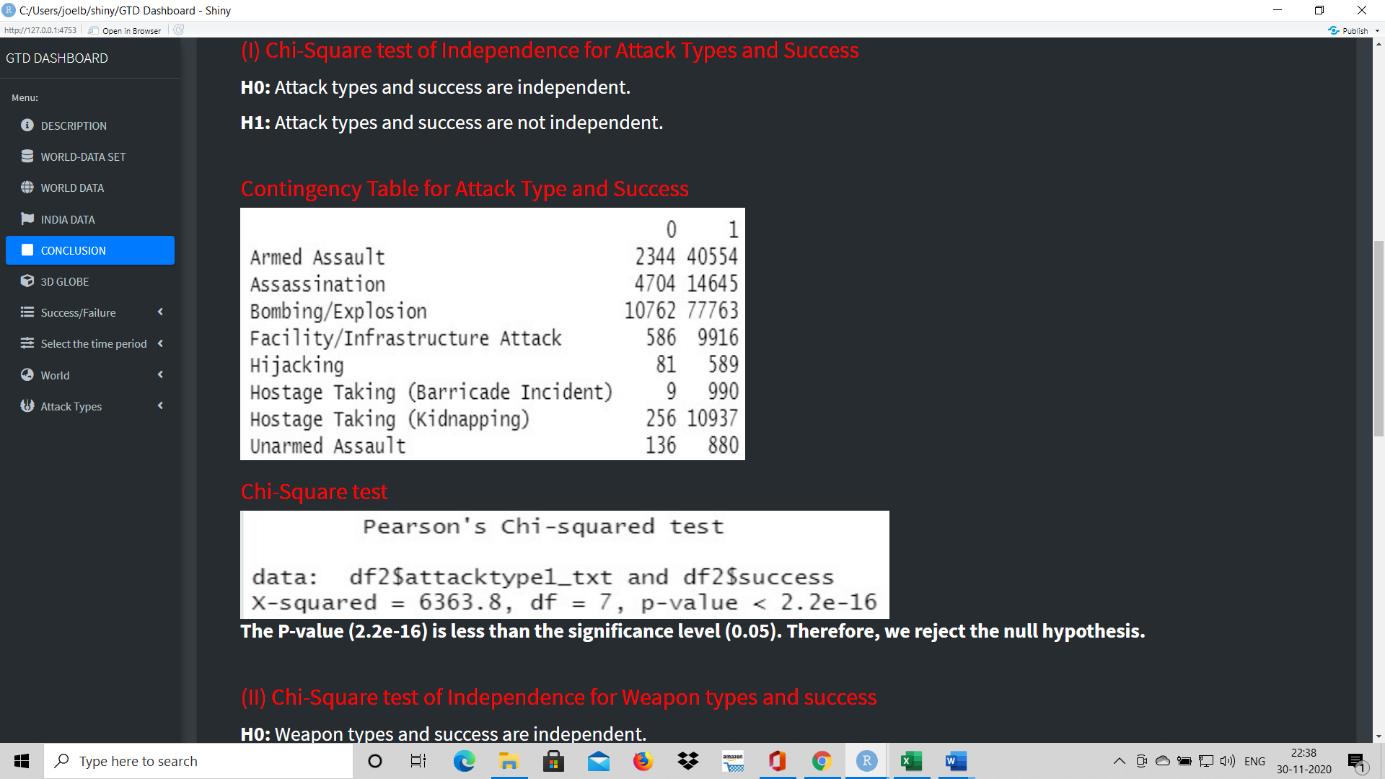
(k) India Data : Panel 3.

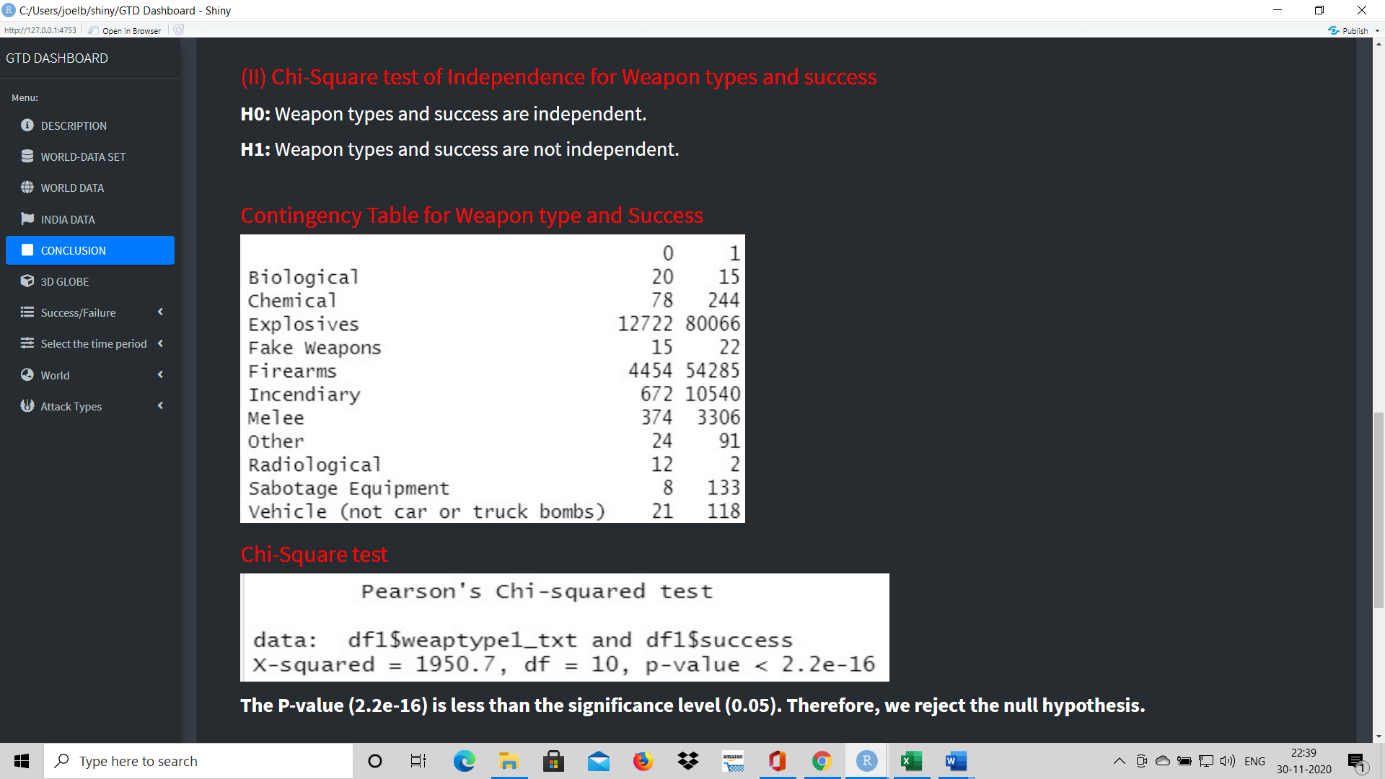


(l) Conclusion.

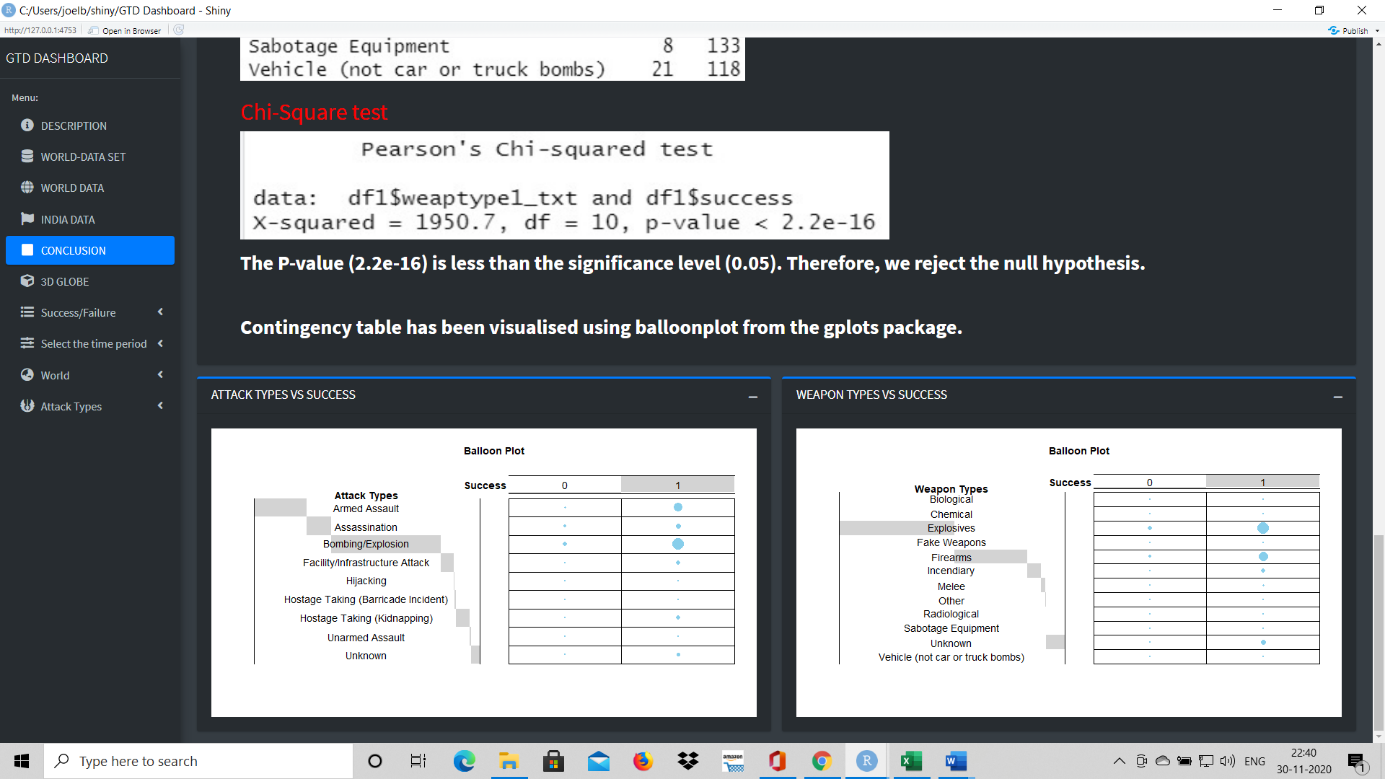
1. Part 1.



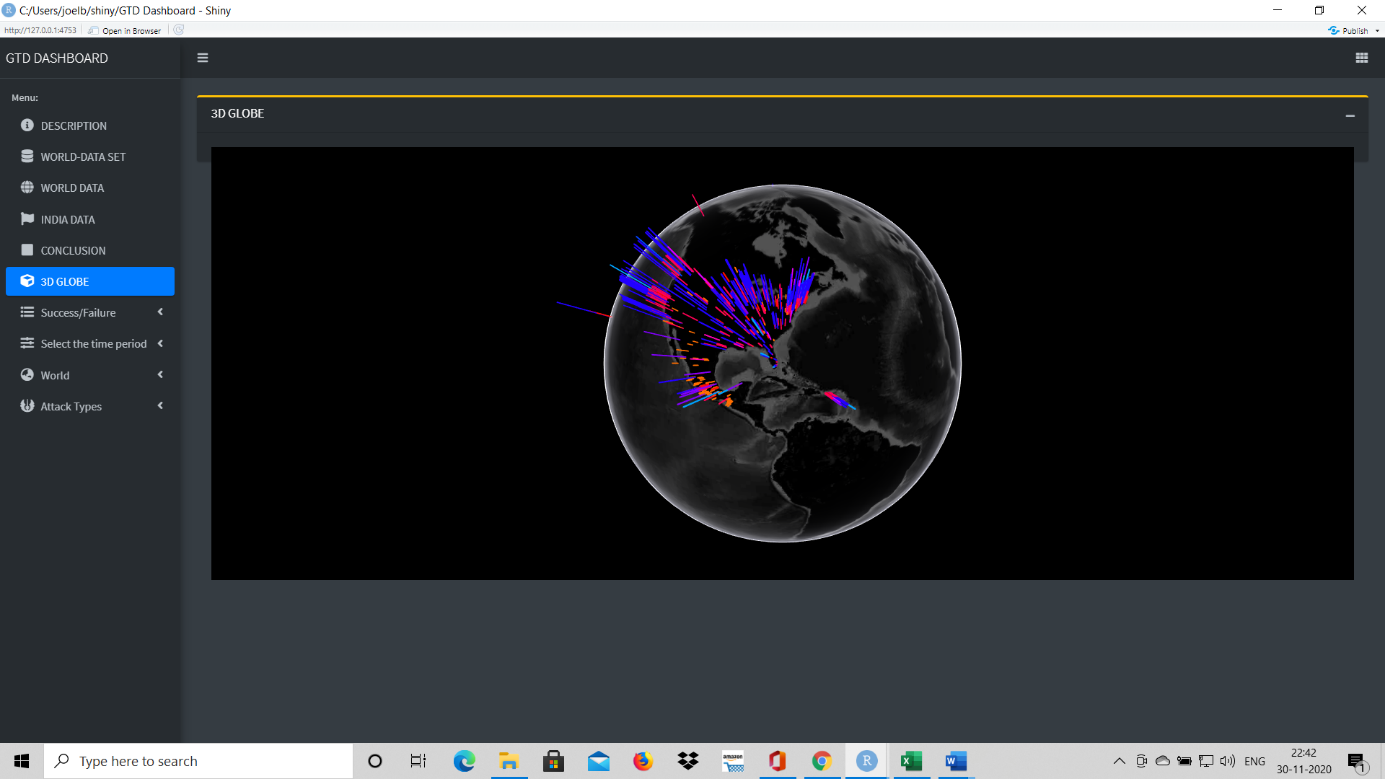
1. Part 2.
2. Part 3.



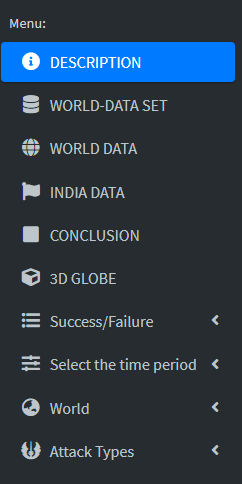
1. Part 4.

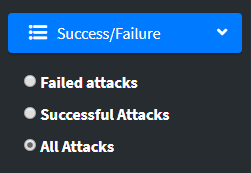


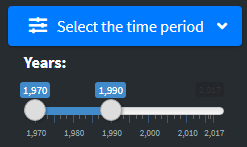
(l) 3D Globe.

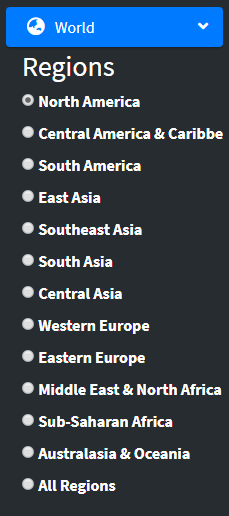


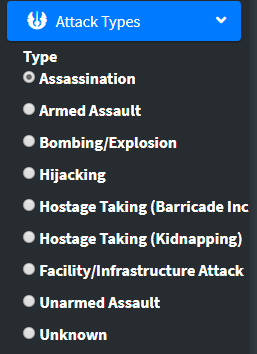
(m) Side Panels.

1. Main Side Panel.
2. Success/Failure Panel.



1. Time Period Selection Panel.
2. World Regions Selection Panel.



1. Attack Type Selection Panel.

# 7 CONCLUSION.

(a) From the analysis of Terrorist attacks with the World in general and India in particular, we wanted to see if there was a relationship between the key factors and the success of attacks. Therefore, Chi square tests were carried out on the key factors of type of attack and the weapons used. What we can draw from these chi square tests is that there is a relationship between the types of attack and success and weapon types and success. This was also visible in the barplots which we saw in the world and India panels.

(b) To further visualise these relations we used the balloon plot. In these visuals, each cell contains a dot whose size reflects the relative magnitude of association of the corresponding components. Here, we could clearly see the prominent association of armed assault and bombing / explosion with success and the association of success with explosives and firearms.

(c) Going a step further, we could also see that terrorists prefer to use explosives and firearms over other methods. Therefore, in order to address a large proportion of means of terrorist attacks, these sources need to be dried out. At this stage, it augurs a thought as to if the sources of production of these elements of terrorism could be rooted out or if there would be strict governmental control and regulation on the materials used for production of these explosives and firearms, one of the elephants in the room could be addressed to a large extent.

(d) Stemming the flow of illegal sources of arms and explosives would require such curbs to be enforced strictly. Therefore, Governmental legislation and curbs need to be strengthened to eliminate, suppress or control the birth and proliferation of the sources of terrorist attacks.

**Appendix A**

**LIST OF DELETED FIELDS AND THE NUMBER OF RECORDS NOT HAVING DATA**

1. $approxdate 172452
2. $resolution 179471
3. $location 126196
4. $summary 66129
5. $alternative 152680
6. $alternative\_txt 152680
7. $attacktype2 175377
8. $attacktype2\_txt 175377
9. $attacktype3 181691
10. $attacktype3\_txt 181691
11. $targtype2 170547
12. $targtype2\_txt 170547
13. $targsubtype2 171006
14. $targsubtype2\_txt 171006
15. $corp2 171574
16. $target2 70671
17. $natlty2 170863
18. $natlty2\_txt 170863
19. $targtype3 180515
20. $targtype3\_txt 180515
21. $targsubtype3 180594
22. $targsubtype3\_txt 180594
23. $corp3 180665
24. $target3 180516
25. $natlty3 180544
26. $natlty3\_txt 180544
27. $gsubname 175801
28. $gname2 179678
29. $gsubname2 181691
30. $gname3 181691
31. $gsubname3 181691
32. $guncertain2 179736
33. $guncertain3 181691
34. $claimmode 162608
35. $claimmode\_txt 162608
36. $claim2 179801
37. $claimmode2 181691
38. $claimmode2\_txt 181691
39. $claim3 181691
40. $claimmode3 181691
41. $claimmode3\_txt 181691
42. $compclaim 181691
43. $weaptype2 168564
44. $weaptype2\_txt 168564
45. $weapsubtype2 170149
46. $weapsubtype2\_txt 170149
47. $weaptype3 179828
48. $weaptype3\_txt 179828
49. $weapsubtype3 179998
50. $weapsubtype3\_txt 179998
51. $weaptype4 181691
52. $weaptype4\_txt 181691
53. $weapsubtype4 181691
54. $weapsubtype4\_txt 181691
55. $nhostkid 168119
56. $nhostkidus 168174
57. $nhours 177628
58. $ndays 173567
59. $divert 181367
60. $kidhijcountry 178386
61. $ransomamt 180341
62. $ransomamtus 181691
63. $ransompaid 180917
64. $ransompaidus 181691
65. $ransomnote 181691
66. $hostkidoutcome 170700
67. $hostkidoutcome\_txt 170700
68. $nreleased 171291
69. $addnotes 153402
70. $related 156653

## bibliography

* [**https://www.start.umd.edu/gtd/**](https://www.start.umd.edu/gtd/)
* [**https://www.kaggle.com/START-UMD/gtd**](https://www.kaggle.com/START-UMD/gtd)
* [**https://datascience42.shinyapps.io/terrorism/**](https://datascience42.shinyapps.io/terrorism/)
* [**https://hahmedsid.shinyapps.io/Terrorism\_Dashboard/**](https://hahmedsid.shinyapps.io/Terrorism_Dashboard/)
* [**https://sarthakdasadia.shinyapps.io/GlobalTerrorism/**](https://sarthakdasadia.shinyapps.io/GlobalTerrorism/)
* [**https://github.com/pranavpandya84/masters\_thesis\_on\_global\_terrorism**](https://github.com/pranavpandya84/masters_thesis_on_global_terrorism)