Data Wrangling

Final Project

Ding, Shuochen

5/4/2018

Bombed German City During World War II and Their Population Recovery

Introduction

During World War II(WWII), Germany had been an area that suffered from bombed by The United States. Especially during 1942 to 1945, many towns, factories, airports, armaments, railroads were destroyed by one of the most terrifying weapons during the decades------bomb. Because of the brutal war, many people were bombed out during the WWII, and huge percentage of them, including soldiers and innocent citizen, were killed by the flames of war. According to online source, about 7 million people, which is around 8.8 percent of the German at the time were lost by WWII. So, in this project, I will research on how Germany’s population and industries was bombed by the United States. Furthermore, combining the population of Germany in 2015, the population recover will also be unveiled in this research------whether the most bombed city recovered or not in terms of its population; are this cities with highest attacking priority still the city of most importance for people to live? And this could also prove Germany’s tremendous vitality in terms of technology.

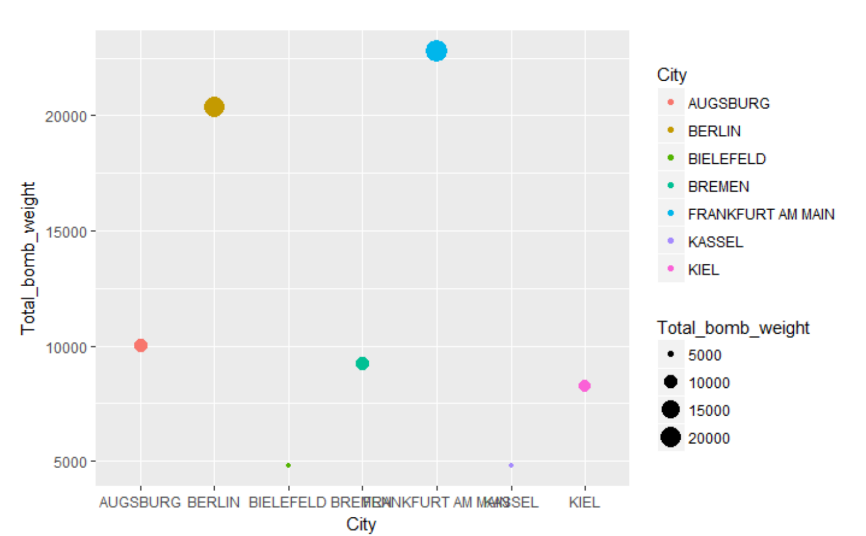
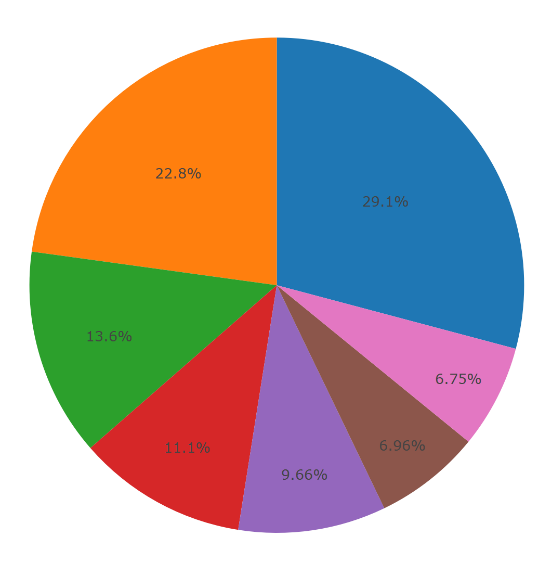
In this project, the dataset I selected consists of digitized paper mission reports from WWII. Each record includes the date, conflict, geographic location, and other data elements to form a live-action sequence of air warfare from 1939 to 1945. The records include U.S. and Royal Air Force data, in addition to some Australian, New Zealand and South African air force missions. But only U.S.’s attack records will be used in the project. The combining dataset being chosen is from Wikipedia called “German-city”, which shows the Germany’s population by city.

Data Processing and Analysis

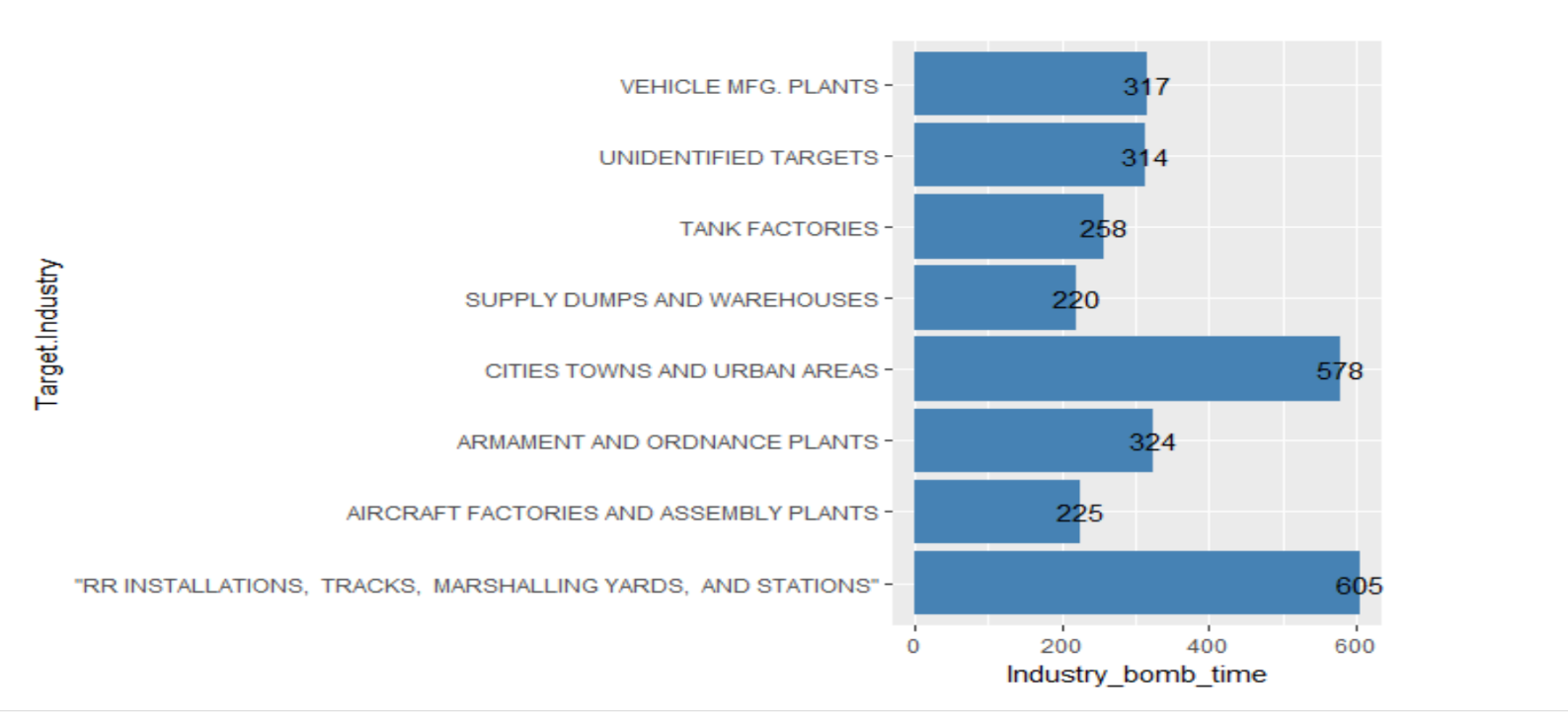
The record of bomb is contained in the document “operations”, which is huge table, including more than 171 thousand of record. So I used filter to choose out USA as my country who released bomb. Then I selected the columns I need from both tables, and with a left\_join operation, I obtain the first usable table to generate my plots and tables. However, more issues came out as my research went deeper.

From this point, I generated several interesting plots. With ggplot, I found the times of bomb attacks by each air force. Not surprisingly, 8AF and 9AF was most frequently used.

Moreover, the following two pictures shows the most frequently bombed city(the plotly package gives us a better view in RStudio), Berlin is the most frequently attacked city, which had 712 bomb attack during 1942-1945. Frankfurt is the second, Bremen is the third and Augsburg is the fourth. This four city combined around 75 percent of bomb attacks among all 25 cities based on record. By doing simple calculation, it shows how inhuman the war is: the dataset record about 26 consecutive months from 1942-1945, which is around 26\*30 =780 days, therefore, in Berlin, in average, there was almost one bomb attack a day for 26 months. It indicates an extremely long nightmare being produced by WWII in Berlin, Germany. Later, we can focus on how this city had been recovered from WWII in terms of population in 2015.



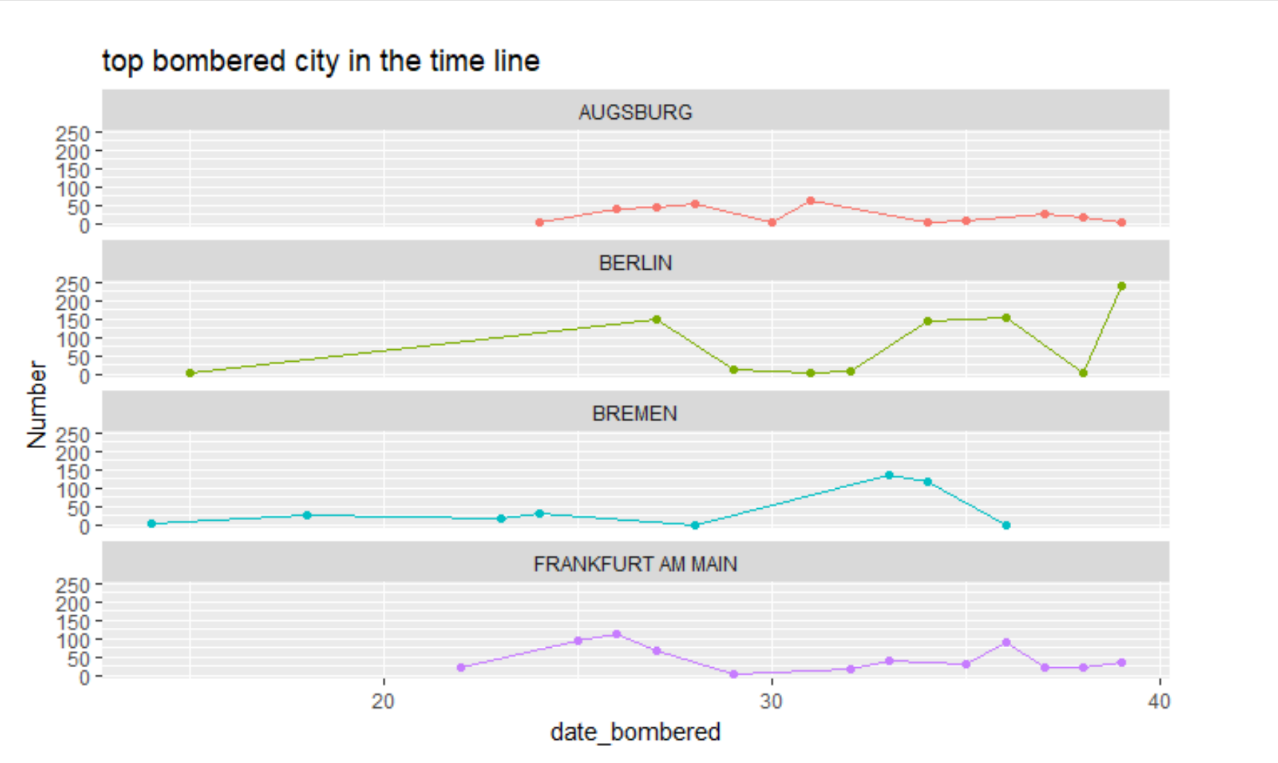
The total weight of bomb release on those top cities also indicates that it could be only blood, dead bodies, flames of war and craters in the city. However, there is something cannot be forgotten-----industries. Because bomb was not only used for killing enemy, it was also for destroying local industries, such as airports, armaments, railroads. The result shows that as the following picture:

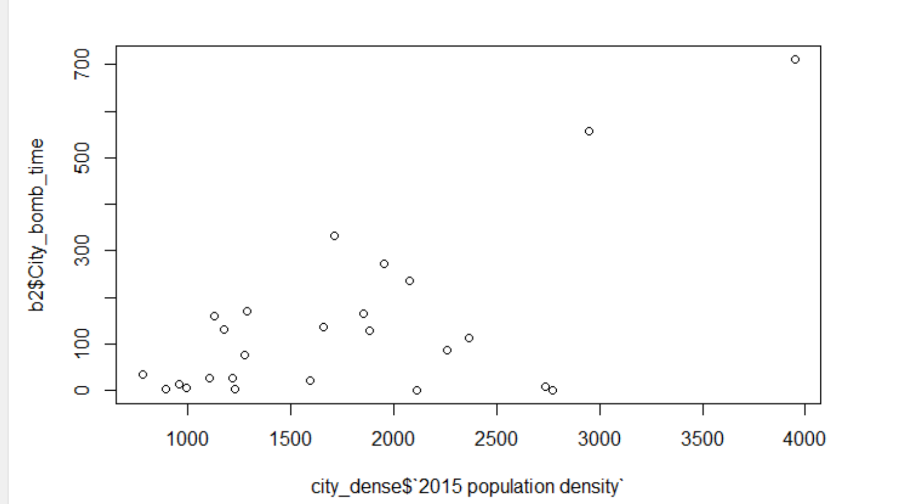


the most vital industry in German during WARII is “rail road installations, tracks, marshalling yards and stations”, the number is 605 during the time period. The second is “city towns and urban areas” with 578 times. The number of this two type are as many as two times of the rest industry, which includes “tank factory”, “armament and ordnance plants” and so on. As far as I’m concerned, destroying local transportation cut off everything including delivering equipment, food and other supply, especially information exchange. Since the techniques for sharing information was not advanced as today, and accuracy of information in a war is the most crucial thing. Also, the huge number for attacking city towns and urban areas, again, reflects the brutality of war------killing everyone and destroying where most people dwelled.

Beside the total times of bomb on each industry, I am more interested in priority level that military of the united states gave to each industry during each attack. So I generate different tables for each industry, and plot out pies for the percentage of their priority------what percentage an industry was put as first priority, and what percentage an industry was put as second priority and so on. So in this way, I could further interpret what industries was usually picked as priority and why. The result is astonishing as I plot them in pie. The lowest percentage for first priority is 0.4, which is “supply dumps and warehouses”; the highest is “aircraft factories and assembly plants”------85.3 percent of time U.S. chose to attack first. For the rest, the percentage of first priority is usually surpass 60%. Looking at the overall percentage, 54.8 percent of attack is priority, 29.0 percent is second, 7.7 percent is third and 4.8 percent for fourth. These numbers show the fact that in most invasions, as long as a warplane captain saw an industry, 83.8 percent of possibility the captain would find first and second priority and choose to release bomb. Again, and again, it reveals brutality of war.

For keeping track of the frequency of bomb in the timeline, some methods learned from course need to be used. Because the mission date given by the database was set as class “factor”, so I usde str\_split from stringr package to split month, day and year with “/”, and then created another variable to paste on this date information and finally pasted it back to the original table. Also, I set the first month after December 1941 as 1, following as 2 and so on. Then, I plotted and listed four top bombed cities in the time line vertically with same x-coordinate as below. In this way , I could easily found on which month, in which city, the bomb released most comparing each line vertically.



Furthermore, I want to look at how the most bombed city recovered in terms of population. so I did a model, the x is population density in 2015, y is number of bomb attack during WWII. As we can see in the picture below, an obvious positive correlation shows that the cities once was bombed most have been recovered. Therefore, these big cities, such as Berlin or Augsburg, preserve their vitality and importance during WWII, and they are still booming city in the world.

Conclusion

War is a frightful calamity, but the will of people is more horrifying. Based on our findings in the project, during WWII, the warring factions always attacked opponent’s vital part to make direct damage and casualties. However, in the meantime, will of people is also the power of dragging this country out of wars of blood and ruin. It was the people in Germany that have been building their hometown restlessly. As a result, those city which suffered most bomb attack had recovered quickly, and came back to the list of top cities in the world.

Citation:

1. *TopForeignStocks.com*, topforeignstocks.com/2016/04/19/chart-world-war-ii-casualties-as-a-percentage-of-each-countrys-population/.
2. <https://en.wikipedia.org/wiki/List_of_cities_in_Germany_by_population>
3. “List of Cities in Germany by Population.” *Wikipedia*, Wikimedia Foundation, 3 May 2018, en.wikipedia.org/wiki/List\_of\_cities\_in\_Germany\_by\_population.