Individual Project Documentation
Ella Kim
March 14th, 2022
CSC-324
I/we confirm that the above list of sources is complete AND that I/we have not talked to anyone else about the solution to this problem.

Purpose

The Sinking of Sewol-Ho (Sewol Ferry) is prominently known as one of the largest accidents in Korea for the past several decades. Killing over 304 out of 476 people, and most of them being young high school students, the incident led to major outrage against the government for its lack of care at the time. The main purpose of this project is to investigate and find correlations between passengers' information and their attributes such as occupation, gender, or type of passenger. In detail, it lends way to understanding whether there were certain attributes shared between passengers that led to higher survival or death rate. It also looks into what type of reaction the incident spurred amongst the citizens of South Korea, by looking into the protests and demonstrations against the government thereafter. Another possibly more significant motive is to remember the victims of this appallling accident of the Sewol-Ho, while giving valid recognition to why so many people were outraged.

Data Description / Collection

This project utilizes two datasets—one regarding passengers' information, and the other on protests:

1. Sewol-Ho Passengers' Information Dataset:
https://www.kaggle.com/ljaem10/the-sinking-of-mv-sewol-south-korea

	X	Category.1	Category.2	Category.3	floor	location	Raw	gender	age
1	0	sailor	Deck	Captain	5	front	survival	male	69
2	1	sailor	Deck	1st voyage	5	front	survival	male	42
3	2	sailor	Deck	1, etc.	5	front	survival	male	34
4	3	sailor	Deck	2nd saver	5	front	survival	male	47
5	4	sailor	Deck	3,	5	front	survival	female	26
6	5	sailor	Deck	Streets	5	front	survival	male	56
7	6	sailor	Deck	Streets	5	front	survival	male	60
8	7	sailor	Deck	Streets	5	front	survival	male	58
9	8	sailor	Agency	warden	5	front	survival	male	54
10	9	sailor	Agency	First institution	3	back	survival	male	58
11	10	sailor	Agency	3rd institutions	5	front	survival	female	26
12	11	sailor	Agency	Premature	3	back	survival	male	61
13	12	sailor	Agency	Early	stokehold	back	survival	male	56
14	13	sailor	Agency	Early	stokehold	back	survival	male	59
15	14	sailor	Agency	Early	3	back	survival	male	62

The above dataset is a collection of passengers' information related to the Ferry accident, including information such as type of Passenger (Student, Sailor, Normal), gender, location, floor, age, and whether they survived. This dataset was the foci of project, as there were many points to make analyses on and derive initially unnoticable patterns until data was made visual.

2. Data on Protests against President Park and Government

Protest	Date	Protestors.Number	Police.Number	Location
1st	2016/10/29	50000	4800	Seoul
2nd	2016/11/05	300000	17600	Seoul
3rd	2016/11/12	1060000	25000	Seoul
4th	2016/11/19	960000	20000	Seoul
5th	2016/11/26	1900000	25000	Seoul
Strike	2016/11/30	60000	8000	Daegu
6th	2016/12/03	2320000	30000	Seoul
7th	2016/12/10	1043400	18200	Seoul
8th	2016/12/17	771750	18240	Seoul
9th	2016/12/24	702000	14700	Seoul
10th	2016/12/31	1104000	18400	Seoul
11th	2016/01/07	643380	15920	Seoul
12th	2017/01/14	146700	14700	Daegu
13th	2017/01/21	353400	15500	Busan
14th	2017/02/04	425500	14600	Busan
15th	2017/02/10	806270	16000	Gwangju

The above dataset is a collection of data on protests led from citizens' frustration with government's actions regarding Sewol-ho. It was found initially through a Korean news article, which lend reference to the above data published on the link above. It is well-known that the sinking of Sewol-Ho resulted in widespread social and political reaction within South Korea. The administration of President Park Geun-hye was especially criticized for her response to the disaster and attempts to downplay government culpability. This information on protests connects directly to the accident as it helps further understand the impact Sewol-Ho had on the country as a whole.

Visualization Audience

The users/audience of this visualization is targeted towards those who have interest in understanding more about the details of the Sinking of Sewol-Ho incident. It is an ambiton of this project that the data visualizations and implications derived from it can be useful for further research or analyses on this topic, or any related topics to accidents on sea.

Additionally, this visualization targets the average people who should take attention. Considering the incident's degree of tragedy and calamity, it seems that it is less highlighted in countries

outside of Korea. Therefore, it is also a goal for this project to be able to teach and reach more people through simpler, easy-to-understand interactive plots.

Questions

1. What patterns/relationships are there between certain attributes (i.e. type of passenger, gender, location) and death or survival?
2. Why was this incident was more problematic than any other accidents over the decades?
3. To what degree did the incident impact the country of South Korea based on data on protests?

Insights

Although the incident is known mostly as one that killed students, it was notable that there were also "normal" passengers (as in those who were traveling to Jeju—the destination of the ferry) for their personal reasons. Through the survival vs. death rate chart, however, it is recognizable that more student passengers died compared to normal passengers (% vs. % death rate). What clearly upset citizens more however, was the fact that most sailors/staffs of the ferry survived. The staffs were irresponsible for the lives of all other passengers and got out of the ferry by themselves.

Investigating what possibilities led to the drastic death rate of students, there were a few relations that were significant. Firstly, checking the first interactive bar chart lets us know most students were located on the fourth floor of the ferry. This contradicts most people's initial intuition that, since the fourth floor is the higher floors on the ship, they would have more time to survive or get out of the ship. This showcases the first point of anger in citizens: they had time and potential to all survive as long as the adults had taken the responsible actions for their lives.

The second interactive bar chart clearly demonstrates to users that the number of students that could not survive are drastically higher compared to any other categories of passengers. To look into this in more depth and accuracy, the bar chart on the third sidebar (named "Survival Rate") shows the comparison of survival vs. death rate of normal, sailor, and student passengers. The subject of indignation is the fact that the survival rate is higher than death rate for staffs of ship, whereas the survival rate of students is nowhere near the survival rate of the staffs.

Improvement

There is room for improvement with the map visualization; firstly, the map does not distinguish which protest it was (1st, 2nd, 3rd…) but only shows the number of participants depending on the circle you hover over. Therefore, improvements in these details could be revised. Additionally, I want to improve my app to address further information regarding the incident and the effects it brought about. Using this first dataset as the basis I want to build onto the app with more information about the aftereffects of this incident.

References

[Kaggle Sewolho passenger information Dataset]
[Korean map source of reference]
[Protests against President Park after Dataset]
[Video on Sewol-ho Tragedy]
[Demonstrations after incident information]
[Information Regarding Sewol-ho]

Process and Development

First, read in dataset "sewol.csv" through read.csv. In other words, import dataset and name it:

```
sewol <- read.csv("sewol.csv")
```

Because the dataset contains names that can be ambiguous, inconsistent capitalization, or misinterpreted in the process of translating from Korean to English, convert these to be understandable as the following:

```
sewol <- sewol %>%
  rename(
    Passenger.Type = Category.1,
    Occupation = Category.2,
    Passenger.Type.Detail = Category.3
  )
sewol$Occupation <- replace(sewol$Occupation, sewol$Occupation=="task", "Worker")
sewol$Occupation <- replace(sewol$Occupation, sewol$Occupation=="teacher", "Teacher")
sewol$Occupation <- replace(sewol$Occupation, sewol$Occupation=="In", "Sailor")
sewol$Occupation <- replace(sewol$Occupation, sewol$Occupation=="student", "Student")
sewol$Occupation <- replace(sewol$Occupation, sewol$Occupation=="move", "Mover")
sewol$Passenger.Type <- replace(sewol$Passenger.Type, sewol$Passenger.Type=="Normal", "normal")
sewol$location <- replace(sewol$location, sewol$location=="In", "in")
```

In order to answer the question regarding rate of survival vs. death amongst different types of passengers, we must work with a new dataset that specifically addresses these categories and mathematically calculate the rates. We can do this by the following:

Make data frame based on the fact that there are a total of 476 passengers on the ferry, set it as tmp_data:

```
tmp_data <- data.frame(total1 = c(476,476,476,476,476,476))
```

Make data fram of the number of people in each category, 104 normal, 33 sailors, and 339 students, set it as tmp_data2:

```
tmp_data2 <- data.frame(total2 = c(104,104,33,33,339,339))
```

Bind them:

```
sewol_tmp <- cbind(sewol_tmp, tmp_data2)
```

Make the rates of each category by the total number of people in each cateogry:

```
sewol_tmp$rate1 <- (sewol_tmp$count/sewol_tmp$total1)*100
```

```
sewol_tmp$rate2 <- (sewol_tmp$count/sewol_tmp$total2)*100
```

For the map visualization, we import the other dataset "demonstration.csv" that contains information regarding each protest against government after the Sewol-ho incident:

Again, first Import dataset of protests against government after Sewol Sinking, name variable:

```
protest <- read.csv("demostration.csv")
```

Make data frame, which will soon thereby act as our map visualization, and save each location where protests occured, their latitudes and longitudes for each individual province:

```
korea <- data.frame(
  name = c("Seoul", "Gwangju", "Busan","Daegu"),
  lat = c(37.5604,35.1595,35.1064,35.8714),
  lng = c(126.9800,126.8526,129.0361,128.6014),
  col = c("blue","red","blue","blue"))
```

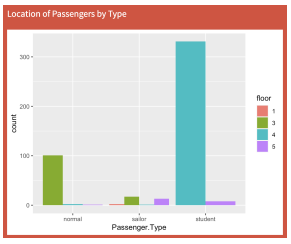
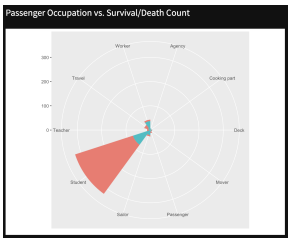
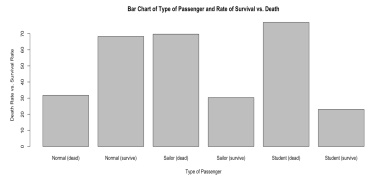
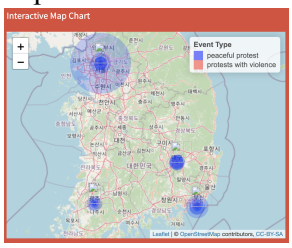
Make this into a map plot using the leaflet library function, adding markers to each province location with a popup of their names :

```
korea <- korea %>%
  leaflet()  %>%
  addTiles()  %>%
  addMarkers(popup=korea$name)  %>%
  addLegend(labels = c("peaceful protest","protets with violence"),
            colors = c("blue", "red"), title = "Event Type")
```

Add Circle Markers based on the size of the protests (number of people attending it) to visualize the size of protests:

```
addCircleMarkers(
  korea,
  radius = (sqrt(protest$Protestors.Number)/70),
  stroke = FALSE,
  color = "blue",
  weight = 5,
  opacity = 0.5,
  fill = TRUE,
  fillColor = "blue",
  fillOpacity = 0.2,
  options = pathOptions(),
  clusterOptions = NULL,
  clusterId = NULL,
  data = getMapData(korea))
```

Design Decision

Chart	What	Why	How
Bar Chart 	The interactive section of the bar chart shows the count (number) of people that died vs. survived depending on different attributes (i.e. Passenger Type).	Users can use this interactive bar chart section to find out whether certain categorized people died more than others.	The heights of each bar clearly informs users on which survived vs. died significantly more or less. Since it is also a side-by-side chart, it also allows users to make direct comparisons.
Polar Plot 	The polar chart shows which type of specific passenger survived or died. The color red represents number of "dead" and blue as "survived."	Users can use this polar chart to find out which category of people died the most or which category of people survived more.	Users can make use of polar plots to see the clear distinction by the color of the plots and size of each circular arc for each category.
Bar Chart 	This bar chart is based on the calculated death and survival rate by number of people in each 'Passenger Type' category.	Users can use this bar chart to compare between the normal, sailor, and student passengers survival rate.	Since this is a chart based on the rate out of the number of people within their category, users can directly compare between students' death rate vs. sailors' death rate.
Map 	This map visualization shows the size of protests after the Sinking of Sewol-Ho, based on the number of people that attended the protests.	Users can use this map visualization to understand degree of significance the incident held. Users can also hover over each circle to see number of people	The size of the circles and event type being all "peaceful protests" shows that progress was made without any use of violence from the citizens attending these consecutive protests over the different provinces in

		that attended the protest.	the country.