

## Preparation of Project 3

### Abstract

*This document recorded how we chose the data and some initial design sketches.*

### Data Selection

We first chose 8 datasets and listed their advantages and limitations.

No.	Name	Advantages	Limitations	Source
1	<b>Students Performance in Exams</b>	— easy to understand — no null value	— little amount of data (about 1000 rows) — too simple	<a href="https://www.kaggle.com/spscientist/students-performance-in-exams">https://www.kaggle.com/spscientist/students-performance-in-exams</a>
2	<b>Used Cars information</b>	— enough data (more than 37,000 rows) — abundant attribute types(ordinal, quantitative, categorical)	— no geographical attributes — needs translation (dataset is in German)	<a href="https://www.kaggle.com/orgesleka/used-cars-database">https://www.kaggle.com/orgesleka/used-cars-database</a>
3	<b>Video Game Sales</b>	— enough data ( about 17,000 rows) — ten available attributes	— the topic is boring	<a href="https://www.kaggle.com/gregorut/videogamesales">https://www.kaggle.com/gregorut/videogamesales</a>
4	<b>Filipino Family Income and Expenditure</b>	— enough data (about than 40,000 rows) — has overview and detail — abundant attribute types(ordinal, quantitative, categorical)	— no "year" attribute, cannot show changing trend	<a href="https://www.kaggle.com/grosvenpaul/family-income-and-expenditure">https://www.kaggle.com/grosvenpaul/family-income-and-expenditure</a>
5	<b>Seattle Crisis Data</b>	— interesting topic — enough data (about 40,000 rows) — only a few null values	— no overview and detail — no significant trend	<a href="https://www.kaggle.com/city-of-seattle/seattle-crisis-data">https://www.kaggle.com/city-of-seattle/seattle-crisis-data</a>
6	<b>Gun Violence Data</b>	— show trend & geographical information — enough data (about 26,000 rows)	— need to deal with null value — exact incident geographical information is shown in longitude and latitude & Need to type it manually	<a href="https://www.kaggle.com/jamesiko/gun-violence-data">https://www.kaggle.com/jamesiko/gun-violence-data</a>
7	<b>Netflix Movies and TV Shows</b>	— abundant attribute types(ordinal, quantitative, categorical)	— need to deal with Multi-Valued Attribute	<a href="https://www.kaggle.com/shivamb/netflix-shows#netflix_titles.csv">https://www.kaggle.com/shivamb/netflix-shows#netflix_titles.csv</a>
8	<b>World University Ranking</b>	— show university ranking in many ways — easy to understand	— decide on which aspect we want to focus — is the ranking authoritative?	<a href="https://www.kaggle.com/mylesoneill/world-university-rankings#school_and_country_table.csv">https://www.kaggle.com/mylesoneill/world-university-rankings#school_and_country_table.csv</a>

## Data Description

After careful comparison and voting, we finally chose dataset number 6 Gun Violence Data. The example of the data is shown in figure 1 and figure 2. Because of the high number of the attributes, the example is divided into two figures. We will not use all attributes in our design.

# gun-violence-data_0... incident_id	date	gun-violence_d... state	gun-violence-data_01-20... city_or_county	# n_killed	# n_injured	# congressional_dist... gun_stolen	Abc gun_type	Abc gun-violence-data_01-2013_0... incident_characte...	Abc gun-violence-data_01-2013_0... latitude	# n_guns_involved	Abc gun-violence-data_01-2013_0... participant_age	Abc gun-violence-data_01-2013_0... participant_age_gr...	
461,105	01.01.2013	Pennsylvania	McKeesport	0	4.0000	14	null	Shot - Wounded/Injur...	40.3467	-79.8559	null	0:20	0:Adult 18+  1::Adul...
460,726	01.01.2013	California	Hawthorne	1	3.0000	43	null	Shot - Wounded/Injur...	33.9090	-118.3330	null	0:20	0:Adult 18+  1::Adul...
478,855	01.01.2013	Ohio	Lorain	1	3.0000	9	0::Unknown  1::Unkn...	0::Unknown  1::Unkn...	41.4455	-82.1377	2	0:25  1::31  2:33  ...	0:Adult 18+  1::Adul...
478,925	05.01.2013	Colorado	Aurora	4	0.0000	6	null	Shot - Dead (murder, ...	39.6518	-104.8020	null	0:29  1::33  2:56  ...	0:Adult 18+  1::Adul...
478,959	07.01.2013	North Carolina	Greensboro	2	2.0000	6	0::Unknown  1::Unkn...	0::Handgun  1::Hand...	36.1140	-79.9569	2	0:18  1::46  2:14  ...	0:Adult 18+  1::Adul...
478,948	07.01.2013	Oklahoma	Tulsa	4	0.0000	1	null	Shot - Dead (murder, ...	36.2405	-95.9768	null	0:23  1::23  2:33  ...	0:Adult 18+  1::Adul...
479,363	19.01.2013	New Mexico	Albuquerque	5	0.0000	1	0::Unknown  1::Unkn...	0:22 LR  1::223 Rem...	34.9791	-106.7160	2	0:51  1::40  2:9  3...	0:Adult 18+  1::Adul...
479,374	21.01.2013	Louisiana	New Orleans	0	5.0000	2	null	Shot - Wounded/Injur...	29.9435	-90.0836	null	null	null
479,389	21.01.2013	California	Brentwood	0	4.0000	9	null	Shot - Wounded/Injur...	37.9656	-121.7180	null	null	0:Teen 12-17  1::Tee...
492,151	23.01.2013	Maryland	Baltimore	1	6.0000	7	null	Shot - Wounded/Injur...	39.2899	-76.6412	null	0:15	0:Teen 12-17  1::Ad...

Figure1: First Part of the Attributes

Abc gun-violence-data_01-2013_0... participant_gender	Abc gun-violence-data_01-2013_0... participant_name	Abc gun-violence-data_01-2013_0... participant_relatio...	Abc gun-violence-data_01-2013_0... participant_status	Abc gun-violence-data_01-2013_0... participant_type	Abc gun-violence-data_01-2013_0... sources	# gun-violence-data_01-2013_0... state_house_district	# gun-violence-data_01-2013_0... state_senate_distr...
0:Male  1::Male  3::...	0:Julian Sims		null	0::Arrested  1::Injur...	0::Victim  1::Victim  ...	http://pittsburgh.cbs...	null
0:Male	0:Bernard Gillis		null	0::Killed  1::Injured  ...	0::Victim  1::Victim  ...	http://losangeles.cbs...	62
0::Male  1::Male  2::...	0:Damien Bell  1::De...		null	0::Injured, Unharmed...	0::Subject-Suspect  ...	http://www.morningj...	56
0::Female  1::Male  ...	0:Stacie Philbrook  ...		null	0::Killed  1::Killed  2...	0::Victim  1::Victim  ...	http://denver.cbsloca...	40
0::Female  1::Male  ...	0:Danielle Imani Ja...	3::Family		0::Injured  1::Injured...	0::Victim  1::Victim  ...	http://myfox8.com/2...	62
0::Female  1::Female...	0:Rebeika Powell  1::...		null	0::Killed  1::Killed  2...	0::Victim  1::Victim  ...	http://www.kjrh.com...	72
0::Male  1::Female  ...	0:Greg Griego  1::Sa...	5::Family		0::Killed  1::Killed  2...	0::Victim  1::Victim  ...	http://www.cbsnews...	10
0::Male  1::Male  2::...				0::Injured  1::Injured...	0::Victim  1::Victim  ...	http://www.huffingt...	93
0::Male  1::Male  2::...				0::Injured  1::Injured...	0::Victim  1::Victim  ...	http://www.contraco...	11
0::Male	0:Deshawn Jones		null	0::Killed  1::Injured  ...	0::Victim  1::Victim  ...	http://articles.baltim...	null

Figure 2: Second Part of the Attributes

This dataset makes a record of more than 26,000 gun violence incidents in the US between January 2013 and March 2018. This data is downloaded from <https://www.kaggle.com/jameslko/gun-violence-data>, originally from [gunviolencearchive.org](http://gunviolencearchive.org). Gun violence Archive is a non-profitable corporation, who provides free online public access to information about gun-related violence in the US (GVA). This data contains detailed information about each incident, for example location, number of people injured or killed, guns involved, and participants' information.

We thought many ways this dataset could be put to good use. To start with, we considered attributes types and some open-ended questions, for instance, how the number of gun violence incidents

changes in the past five years and what is the relationship between the number of injured people and the number of killed people? We wrote down our questions and sketched possible visualisations. Then, we picked up one best visualisation for each question and developed chosen visualisations in Tableau to see accurate data change.

## Reference

1. Gun Violence Archive (GVA). [Online]. <http://www.gunviolencearchive.org/>. [Accessed on 23.March 2020].

## Appendix A : Sketches

To Start With : Attribute Types & Possible Questions

- Attribute Types

c: categorical    o: ordinal    Q: quantitative

incident\_id : C

date : c

state : c

n\_killed : Q

n\_injured : Q

n\_injured : Q

gun\_stolen : Qtc

participant\_age\_group : o

participant\_name : c

participant\_gender : c

- Possible Questions

- How the number of gun violence incidents changes over past 5 years?

- What is the relationship between the number of injured people & the number of killed people?

- Which state has the highest number of gun violence incidents?

- Where are the relatively dangerous places?

- When is the most dangerous time?

- Would the states with more guns tend to be more dangerous?

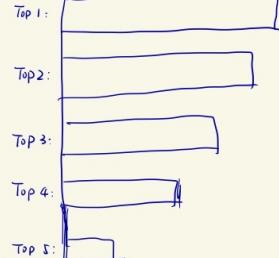
## The U.S. Gun Violence Data From 2013~

Question 1: Which state has the highest number of gun related violence?

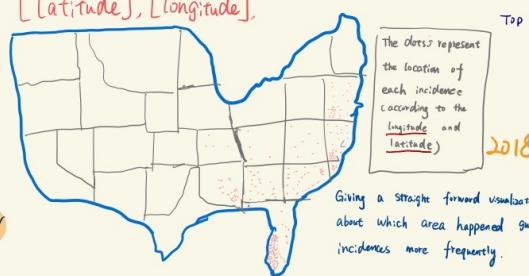
[State], count [incident-id]



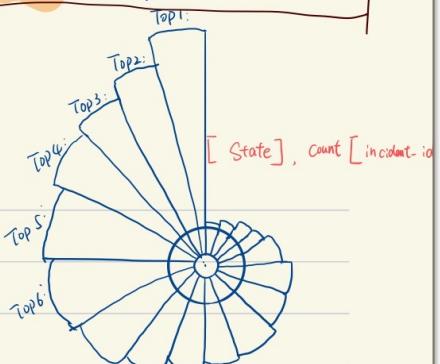
Top dangerous states with highest number of gun related violences.



② Located by  
'incidents'  
(location or)

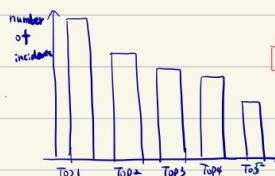
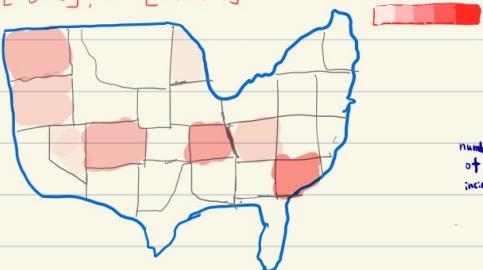


Top cities with highest number of gun violence



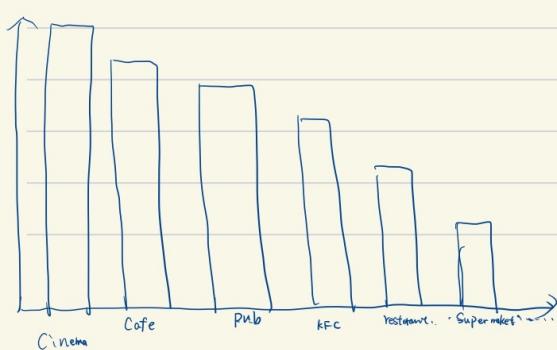
③ Located by  
States.

[State], count [incident-id]



Q2: The dangerous places where have gun related cases a lot.

Word cloud.



Cinema

Pub.

Supermarket.

cafe

KFC

Restaurant.

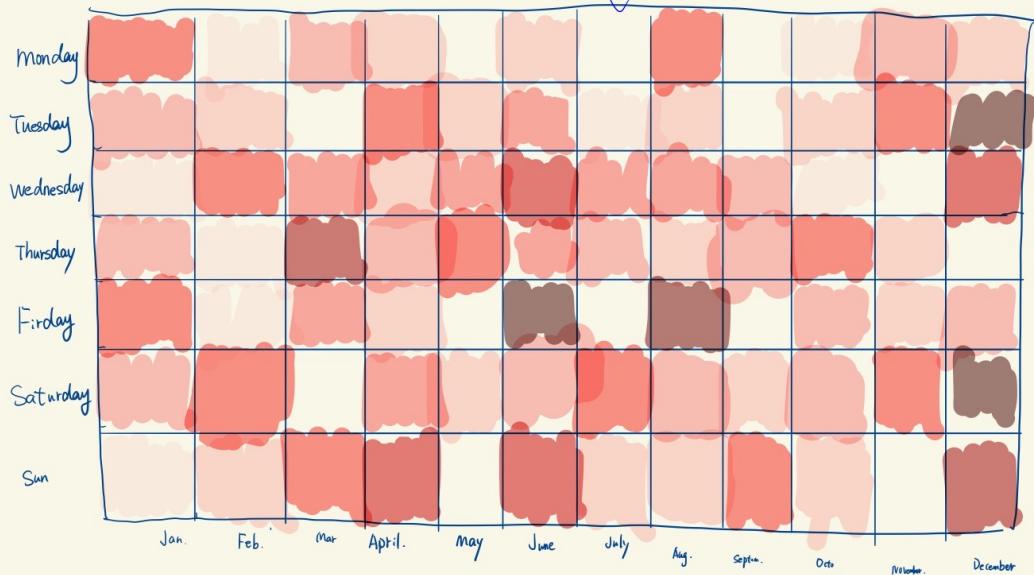
Q3: Based on the data from 2013 ~ 2018,

[ date ] Count[ incident\_id ]

When is it more likely to happen

Gun violence incidents?

use hue to express different level of quantity.

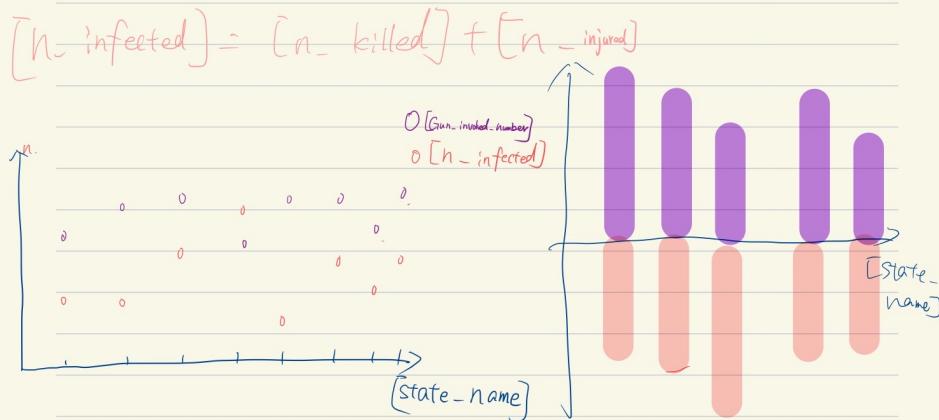


Question 4: The overall Situation of gun violence in the U.S. from 2013~2018?



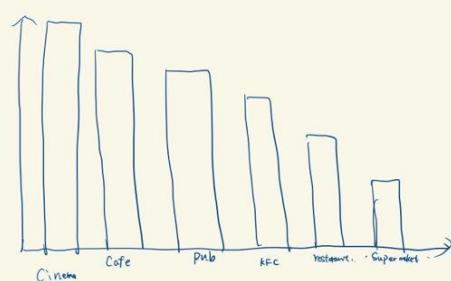
Q5:

Would the states with more guns be more dangerous (more people infected by gun related incident?)



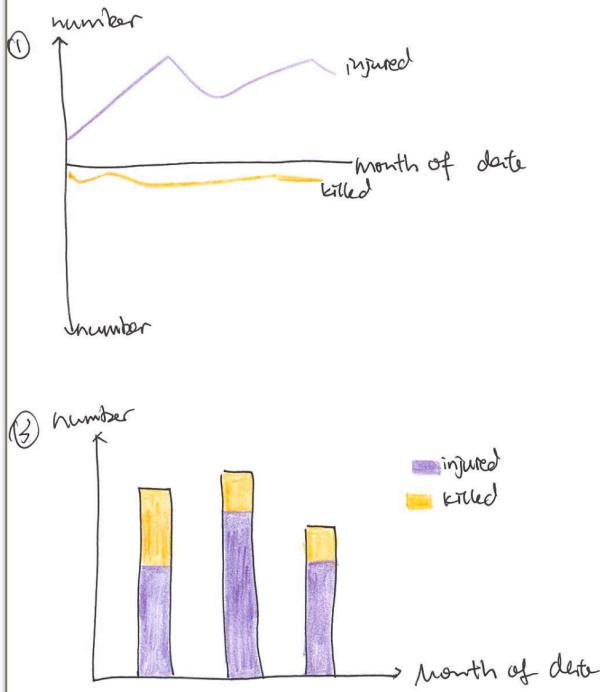
Q6: The dangerous places where have gun related cases a lot.

Word cloud.

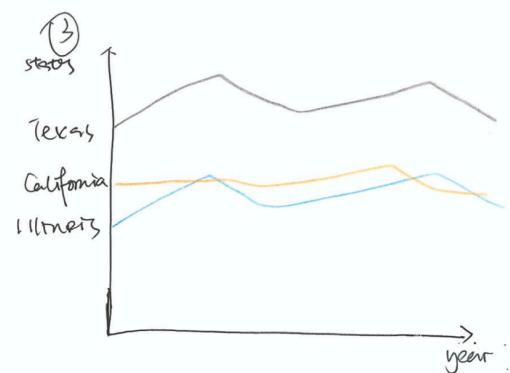
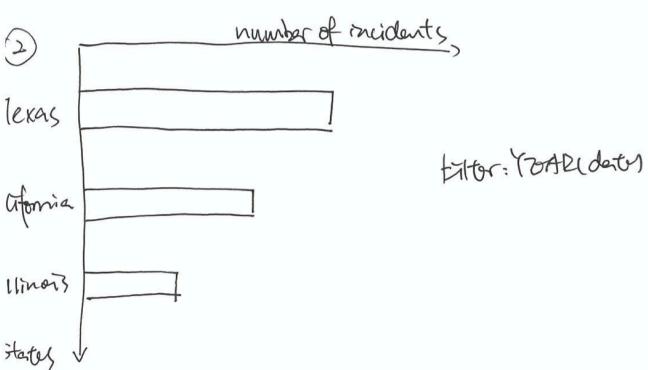
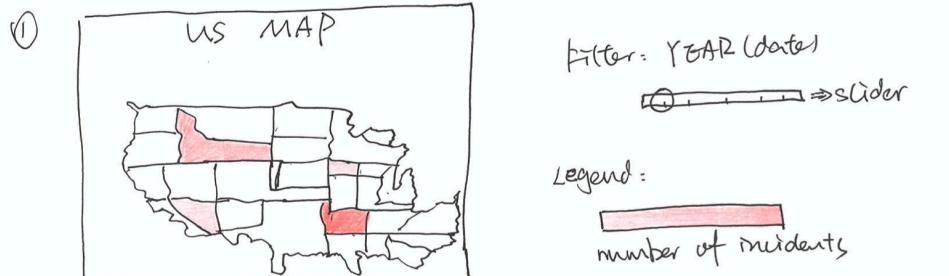


Cinema      Pub.  
Supermarket.  
KFC  
Restaurant.  
cafe

Question 7: What is the relationship between the number of injured people & the number of killed people in gun violence incidents?



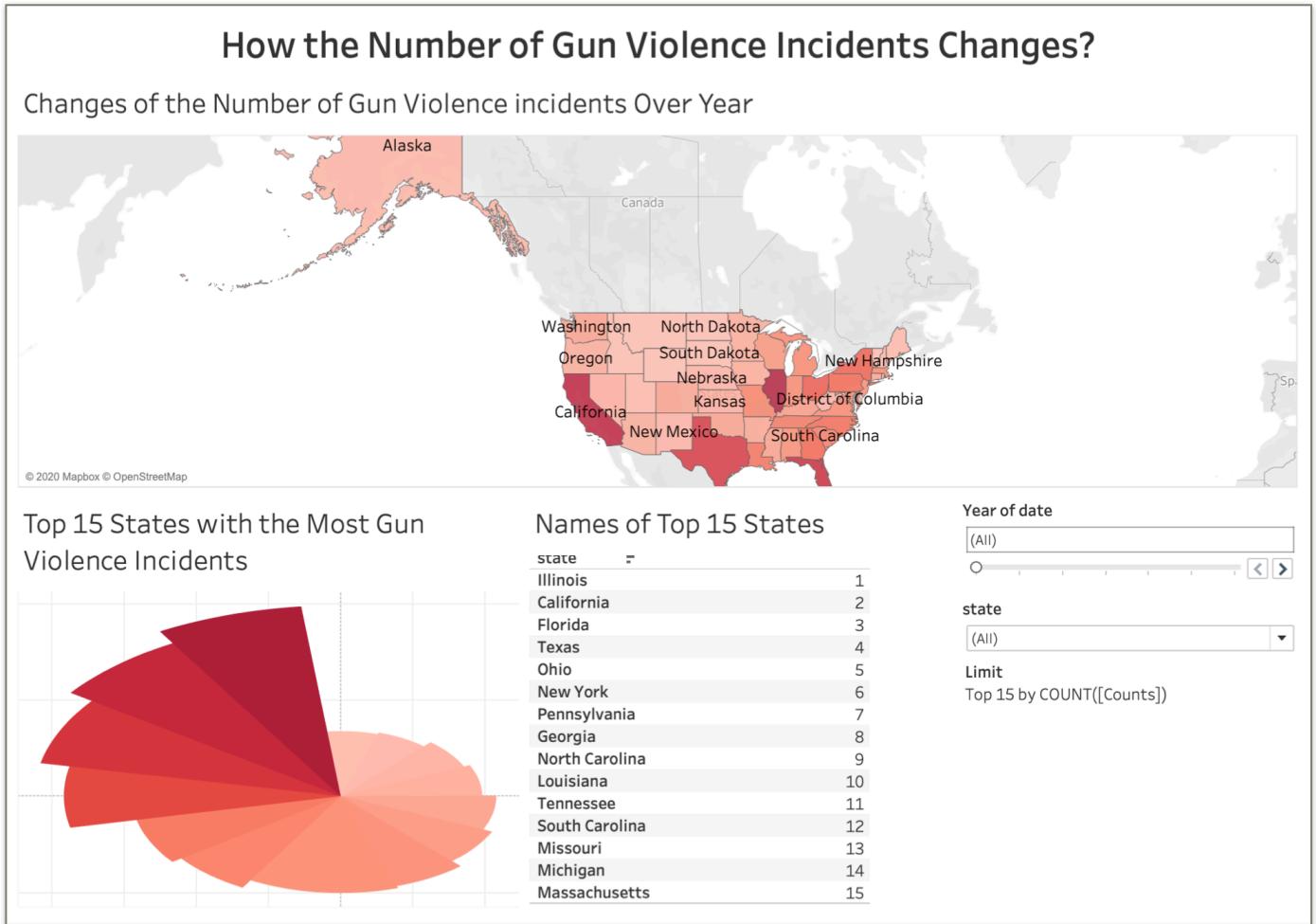
Question 8: How the number of gun violence incidents changes in the past 5 years?



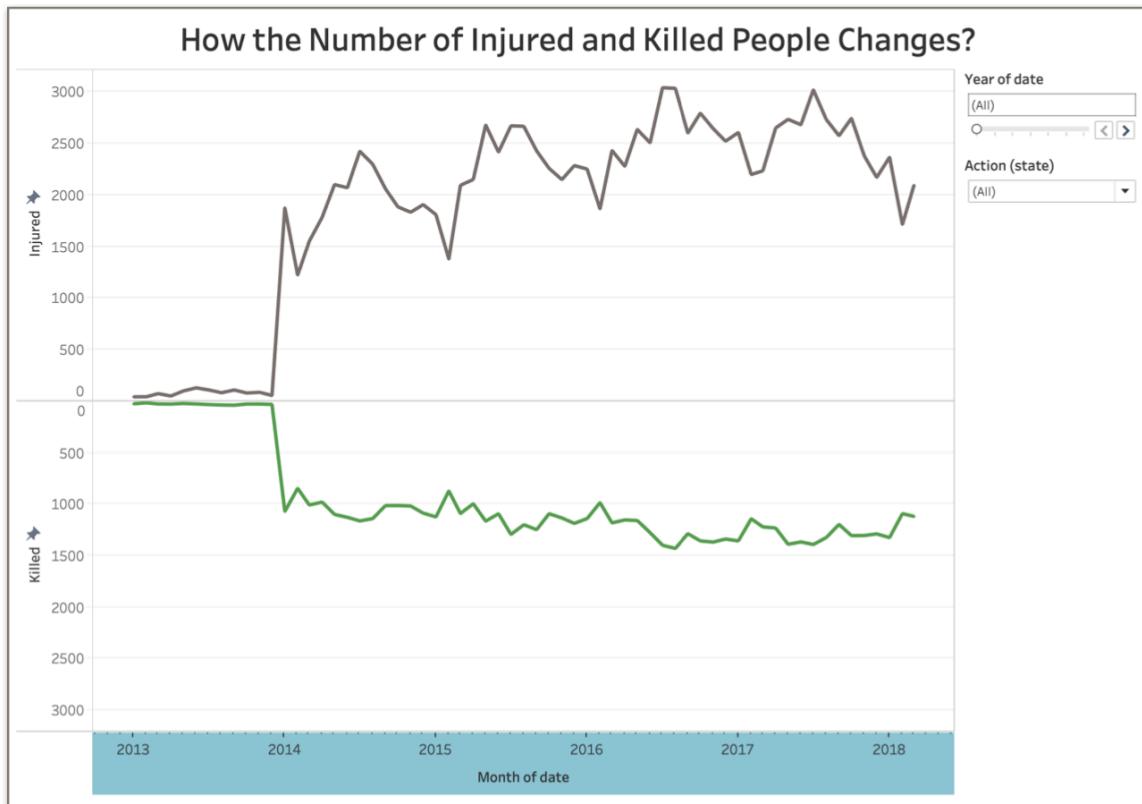
## Appendix B : Sketches in Tableau

— Screenshots of the dashboards in sketches\_in\_tableau.twbx

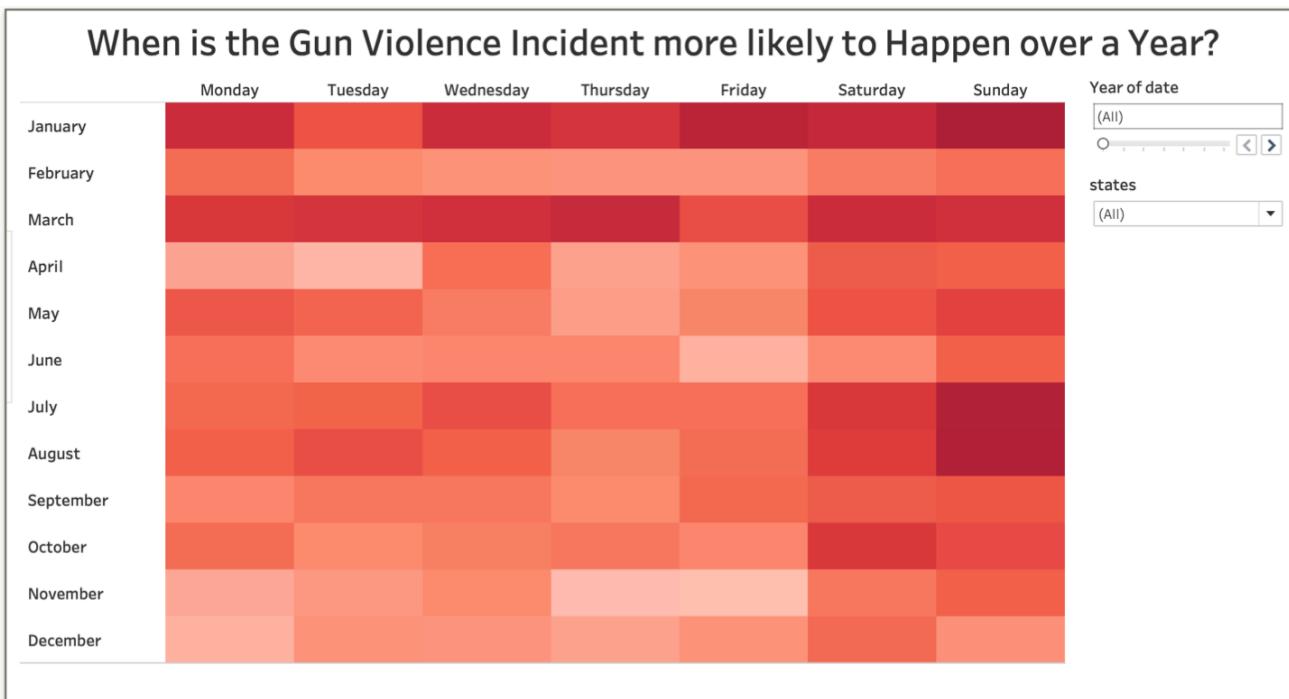
### — Dashboard 1



## — Dashboard 2



## — Dashboard 3



## — Dashboard 4

