UNIVERSITÀ degli STUDI di CATANIA | DIPARTIMENTO di ECONOMIA e IMPRESA

FATAL POLICE SHOOTING IN THE US

Do the Black people have any reason to protest?

TEACHING : BIG DATA ANALYTICS

PROFESSOR : ORAZIO TOMARCHIO

STUDENTS : GIANLUIGI MAZZAGLIA

DAVIDE SAITTA

Sommario

Introduction and Dataset Description

The report is based on the analysis of the shooting made from the police to the US people. The Database is available on https://www.kaggle.com/kwullum/fatal-police-shootings-in-the-us?select=PoliceKillingsUS.csv and we replace the table Police_killings_US with the table “fatal-police-shootings-data” taken from https://github.com/washingtonpost/data-police-shootings and that

describes every single fatal shooting in the US by a police from 2015 up to now. The Dataset contains 5 tables, 3 of them are taken from the US census on poverty rate, high school graduation rate and median household income (in 2015) with the goal to discover potential correlation with the shooting. The main table is a tracking of any person shoot with some characteristic of them as race, age, gender and so on, while the last table describes the percentage of any race per city.

fatal-police-shootings-data
id
name
date
manner of death
armed
age
gender
race
state
City
signs_of_mental_illness
threat_level
flee
longitude
latitude

ShareRaceByCity
Geographic Area
City
share_white
share_black
share_hispanic
share_native_american
share_asian

PercentagePeopleBelowPovertyLevel
Geographic Area
City
Poverty_rate

PercentOver25CompletedHighSchool
Geographic Area
City
Poverty_rate

MedianHouseholdIncome2015
Geographic Area
City
Median Income

Business Questions

Question 1 : General Information

a) What is the rate of total shooting in USA from 2015 to 2020?
b) Which gender is most involved in?
c) What is the age range with the higher deaths?
d) How are distributed the people shot by the police all around the US?

Question 2 : Any race more Shot?

a) Is there any race particularly involved in this context?
b) What is the rate of unarmed people shot?
c) Is there any state where any race are particularly spread?
d) What is the trend of shooting by race per year?

Question 3 : Socio economic factors

a) How to the povery rate by state affects the shooting?
b) And, Median Income?
c) Is the Education Level is relevant in the Local Shooting?

General Question

Is there any real consideration that brought the Black race to the recent spot “Blacks live matter”?

Data Fact Model

The **Dimensional fact model** is a graphical conceptual model for data marts, proposed by Golfarelly & Rizzi. A **fact** is a concept relevant to decision-making processes, A **measure** is a numerical propery of a fact and describes a quantitative aspect that is relevant to analysis, a **Dimension** is a fact property with a finite domain and describes an analysis axis of the fact. In our final dataset we identified Shoot_by_Police as the fact, and the Number of shot people is the only measure. The **dimension**, that determine the minimum granularity adopted to represents fact, are identified in **Location**, **Date**, **Death**. The relationships among the dimensional attributes are expressed by hierarchies.

For **Date** :

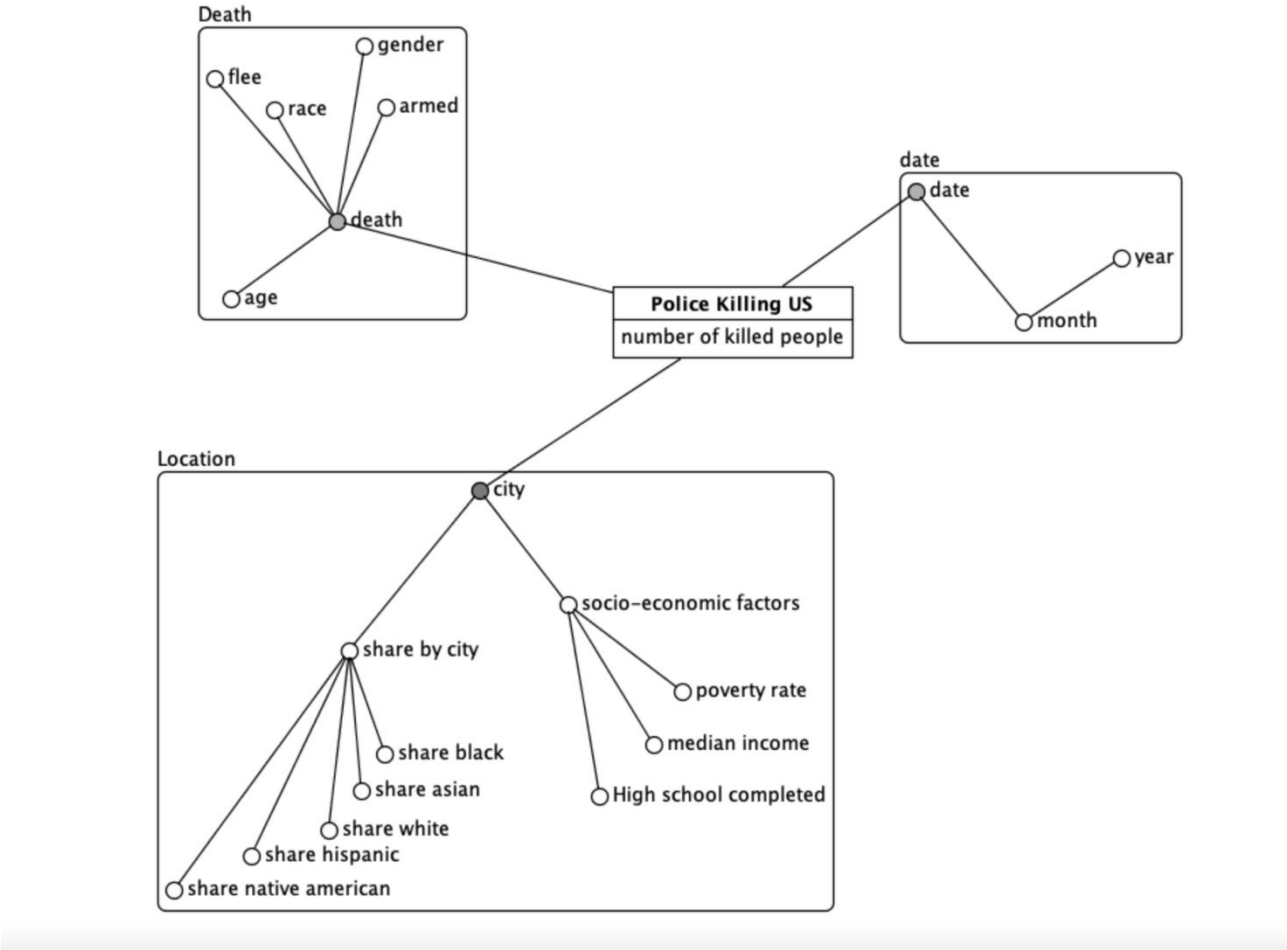
- Date -> month -> year

For **Location** :

- City -> socio-economic-factors -> poverty rate
- City -> socio-economic-factors -> median income
- City -> socio-economic-factors -> High School completed
-
- City -> Share bycity -> share black
- City -> Share bycity -> share white
- City -> Share bycity -> share asian
- City -> Share bycity -> share native american
- City -> Share bycity -> share hispanic

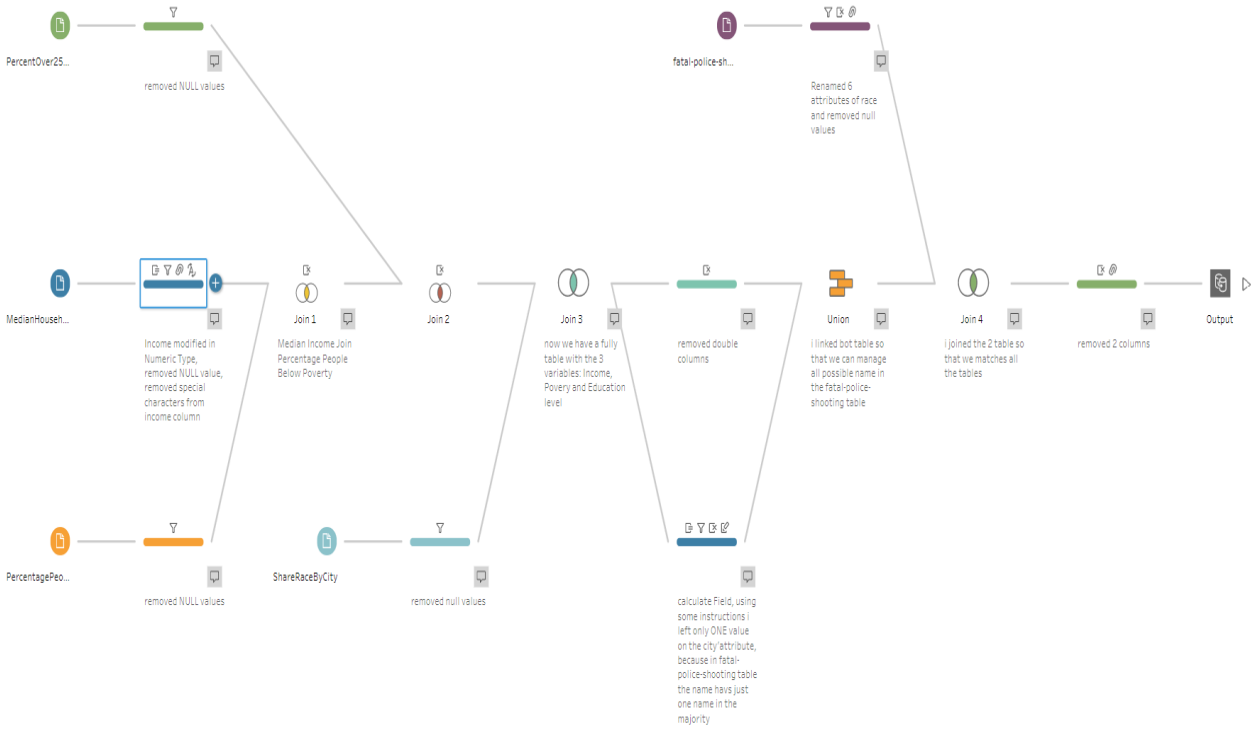
For **Death** :

- death -> race
- death -> flee
- death -> gender
- death -> armed
- death -> age

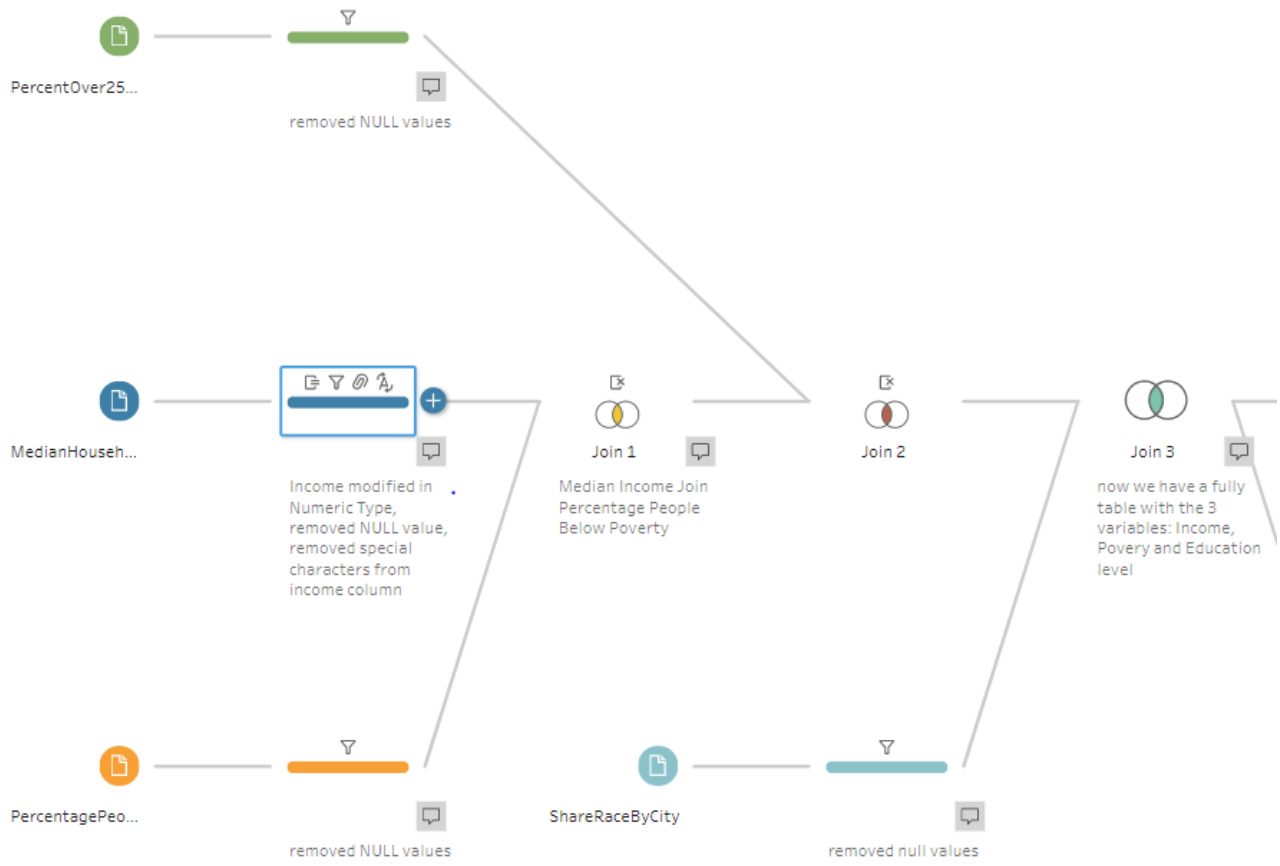


ETL Process

The first step before proceeding with the analysis is the ETL Process that consisting using the Tool Tableau Data Prep in Cleaning , Create some calculate fields, joining the resulting in a final datasets.


PercentOver25...
removed NULL values
fatal-police-sh...
Renamed 6 attributes of race and removed null values
MedianHouseh...
Income modified in Numeric Type, removed NULL value, removed special characters from income column
Join 1
Median Income Join Percentage People Below Poverty
Join 2
Join 3
now we have a fully table with the 3 variables: Income, Povery and Education level
removed double columns
Union
i linked bot table so that we can manage all possible name in the fatal-police-shooting table
Join 4
i joined the 2 table so that we matches all the tables
removed 2 columns
Output
PercentagePeo...
removed NULL values
ShareRaceByCity
removed null values
calculate Field, using some instructions i left only ONE value on the city'attribute, because in fatal-police-shooting table the name havs just one name in the majority

Cleaning and Join



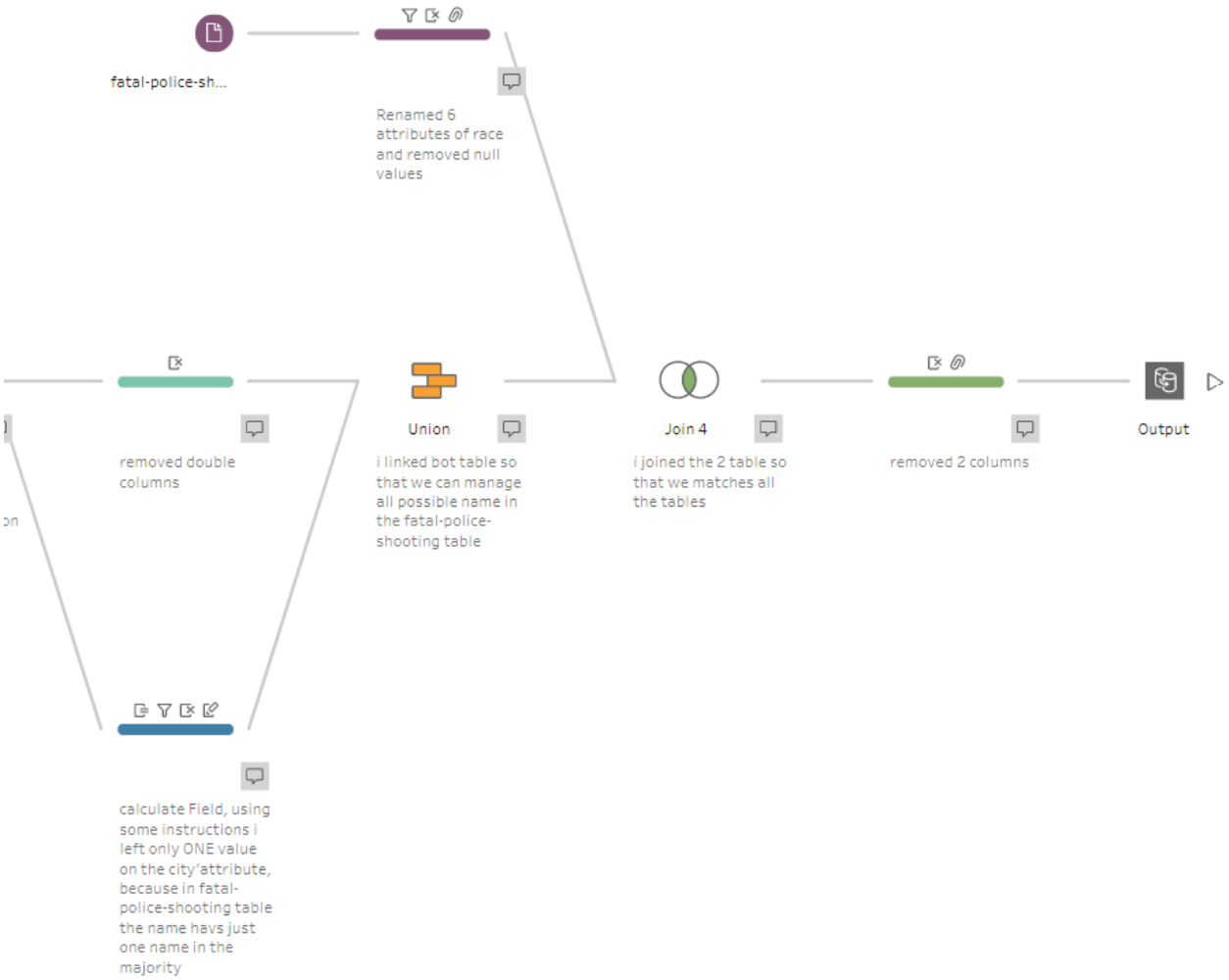
Median Household Income : On this table we have modified the Data Type of "Median Income" from String to Numerical Value, cleaning all the rows with special characters, letters and space.

PercentagePeopleBelowPoverty : On this table we removed NULL values

PercentOver25CompletedHighSchool : On this table we removed NULL values

All of 3 tables differes just for the last column, so after cleaning we joined all of 3 tables so that we create a single table that we join with the 4th table :

ShareByCity : Here we have a represantion for each city of the share of people by race, we removed the NULL values, so now after the Inner Join we a complete table that include all the variables by city.



We duplicate the last table obtained, adding a **calculate field**, using the following instructions :

```
IF(CONTAINS([City],"CDP")) THEN
REPLACE([City],"CDP","")
ELSEIF(CONTAINS([City],"city")) THEN
REPLACE([City],"city","")
ELSEIF(CONTAINS([City],"town")) THEN
REPLACE([City],"town","")
ELSEIF(CONTAINS([City],"City")) THEN
REPLACE([City],"City","")
ELSEIF(CONTAINS([City],"village")) THEN
REPLACE([City],"village","")
ELSEIF(CONTAINS([City],"Village")) THEN
REPLACE([City],"Village","")
ELSEIF(CONTAINS([City],"municipality")) THEN
REPLACE([City],"municipality","")
ELSEIF(CONTAINS([City],"borough")) THEN
REPLACE([City],"borough","")
ELSEIF(CONTAINS([City],"government")) THEN
REPLACE([City],"government","")
ELSEIF(CONTAINS([City],"county")) THEN
REPLACE([City],"county","")
ELSEIF(CONTAINS([City],"County")) THEN
REPLACE([City],"County","")
END
```

In this way, considering that we have taken another table from another source, the only columns in common were State and City, but the last one contained some more details that couldn't afford to let do the join. So this because the field City contained some more information as "borough", "town", "municipality" that weren't present in the other records. So by the way this information wasn't the same for all records, first we duplicate the table (considering some rows matched without any cleaning) then we did an **UNION** between both the table and then we did an INNER JOIN with the last table **FATAL_SHOOTING_POLICE** that contains all the record of people killed by police in a specific State, City. Removing the double columns, the final Output :

id	state	city	date	age	gender	race	armed	flee	percent_completed_hs	Median Income	poverty_ra...
6.438	California	Los Angeles city	19/12/2020	32	Male	Hispanic	gun	Not fleeing	75,5	50.205	22,1
6.238	California	Los Angeles city	16/10/2020	25	Male	Black, non-Hispanic	gun	Foot	75,5	50.205	22,1
6.370	California	Los Angeles city	15/10/2020	47	Male	Black, non-Hispanic	gun	Foot	75,5	50.205	22,1
6.165	California	Los Angeles city	10/09/2020	41	Male	Hispanic	gun	Not fleeing	75,5	50.205	22,1
6.134	California	Los Angeles city	31/08/2020	29	Male	Black, non-Hispanic	gun	Foot	75,5	50.205	22,1

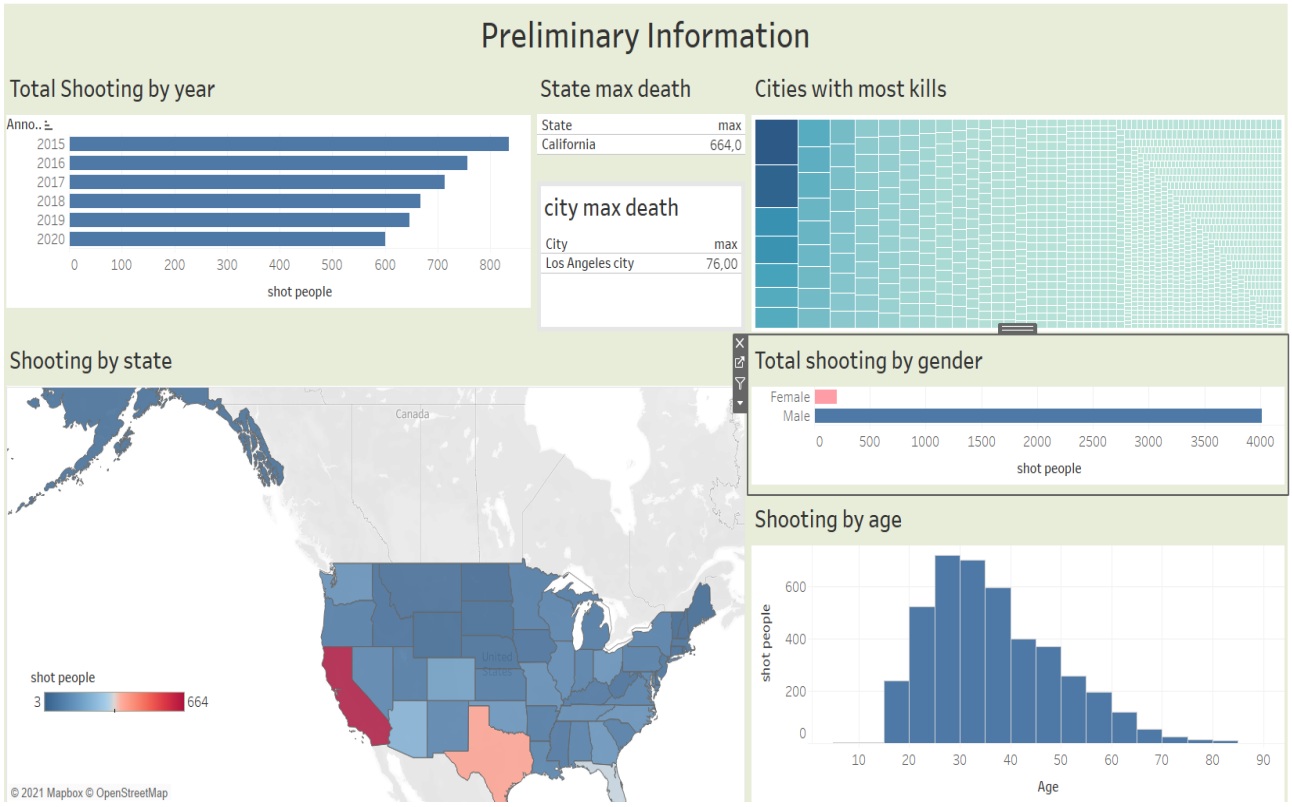
share_white	share_black	share_native_american	share_asian	share_hispanic	latitude	longitude
49,8	9,6	0,7	11,3	48,5	33,942	-118,353
49,8	9,6	0,7	11,3	48,5	33,923	-118,265
49,8	9,6	0,7	11,3	48,5	33,927	-118,292
49,8	9,6	0,7	11,3	48,5	33,892	-118,203
49,8	9,6	0,7	11,3	48,5	33,936	-118,296
49,8	9,6	0,7	11,3	48,5	33,937	-118,294

DASHBOARDS

In order to answer our Business Questions we built 3 Dashboards and we organized it within a story, here in detail we see every worksheet.

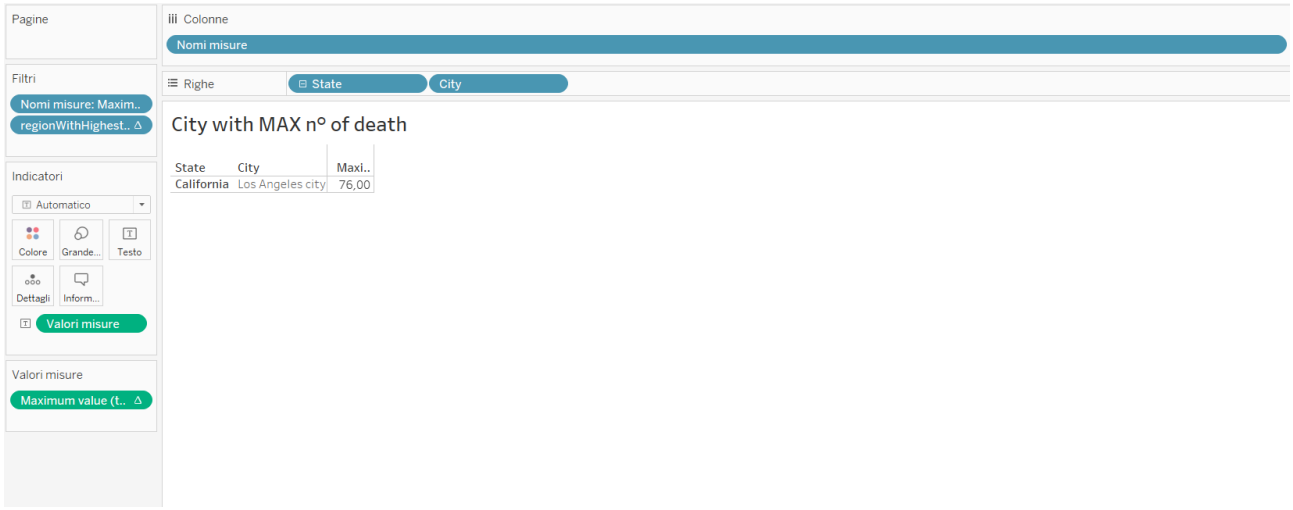
Dashboard 1

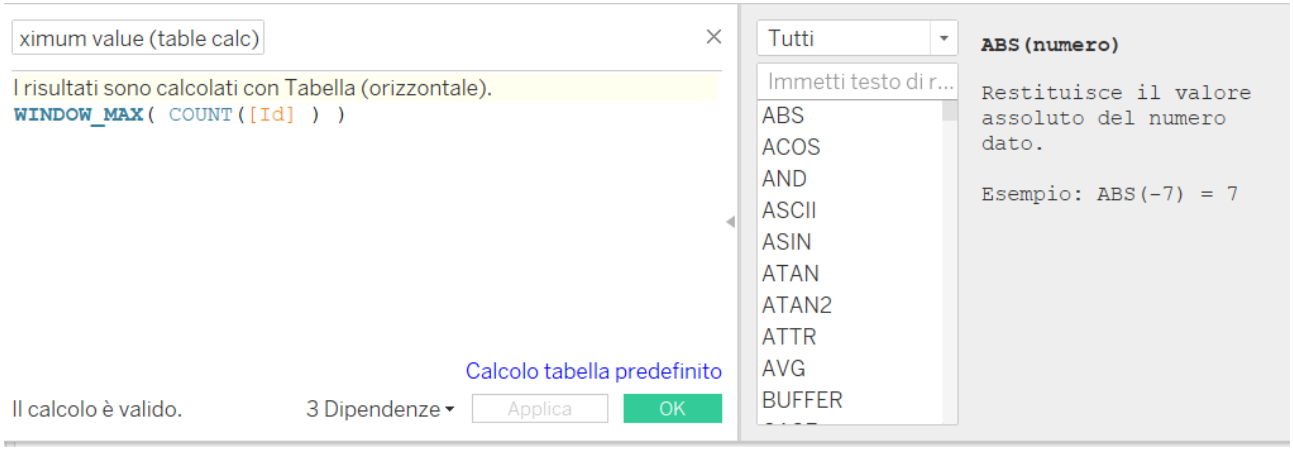
The first dashboard aims to analyse general information, which will be useful for the following Dashboards. It contains 5 sheets: Total shooting by year, shooting per month by year, shooting by state, Total shooting by gender, Shooting by age, we also calculated the State and the city with the highest number of death .

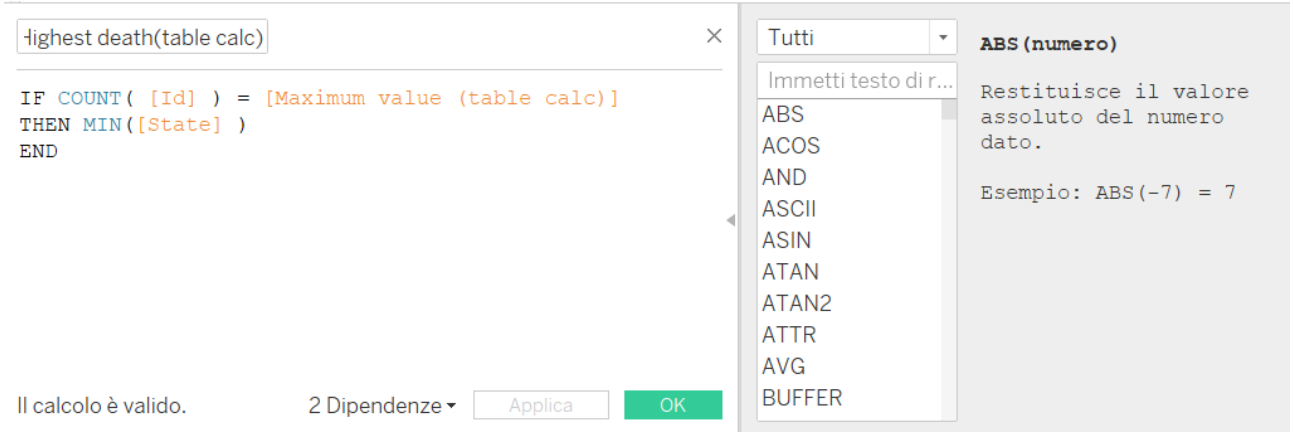


More in depth, the single sheets are composed as follows:

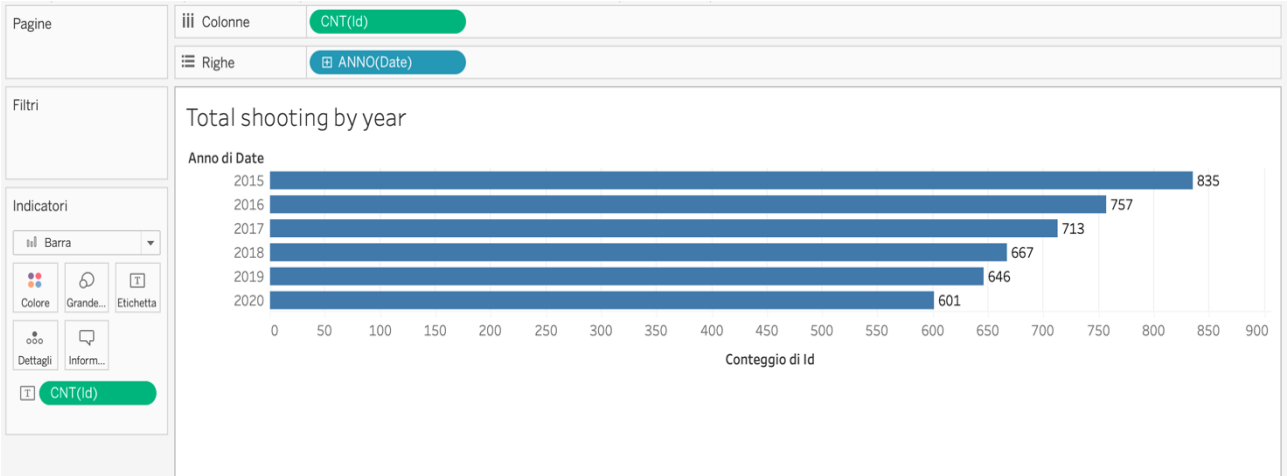
First of all, we calculated the city with the Max number of death, to do that we did a calculated field:


Pagine
Colonne
Nomi misure
Filtri
Nomi misure: Maxim..
regionWithHighest..
Righe
State
City
City with MAX n° of death
State City Maxi..
California Los Angeles city 76,00
Indicatori
Automatico
Colore
Grande...
Testo
Dettagli
Inform...
Valori misure
Valori misure
Maximum value (t..


ximum value (table calc)
I risultati sono calcolati con Tabella (orizzontale).
WINDOW_MAX(COUNT([Id]))
Calcolo tabella predefinito
Il calcolo è valido.
3 Dipendenze
Applica
OK
Tutti
Immetti testo di r...
ABS
ACOS
AND
ASCII
ASIN
ATAN
ATAN2
ATTR
AVG
BUFFER
ABS(numero)
Restituisce il valore assoluto del numero dato.
Esempio: ABS(-7) = 7


highest death(table calc)
IF COUNT([Id]) = [Maximum value (table calc)]
THEN MIN([State])
END
Il calcolo è valido.
2 Dipendenze
Applica
OK
Tutti
Immetti testo di r...
ABS
ACOS
AND
ASCII
ASIN
ATAN
ATAN2
ATTR
AVG
BUFFER
ABS(numero)
Restituisce il valore assoluto del numero dato.
Esempio: ABS(-7) = 7

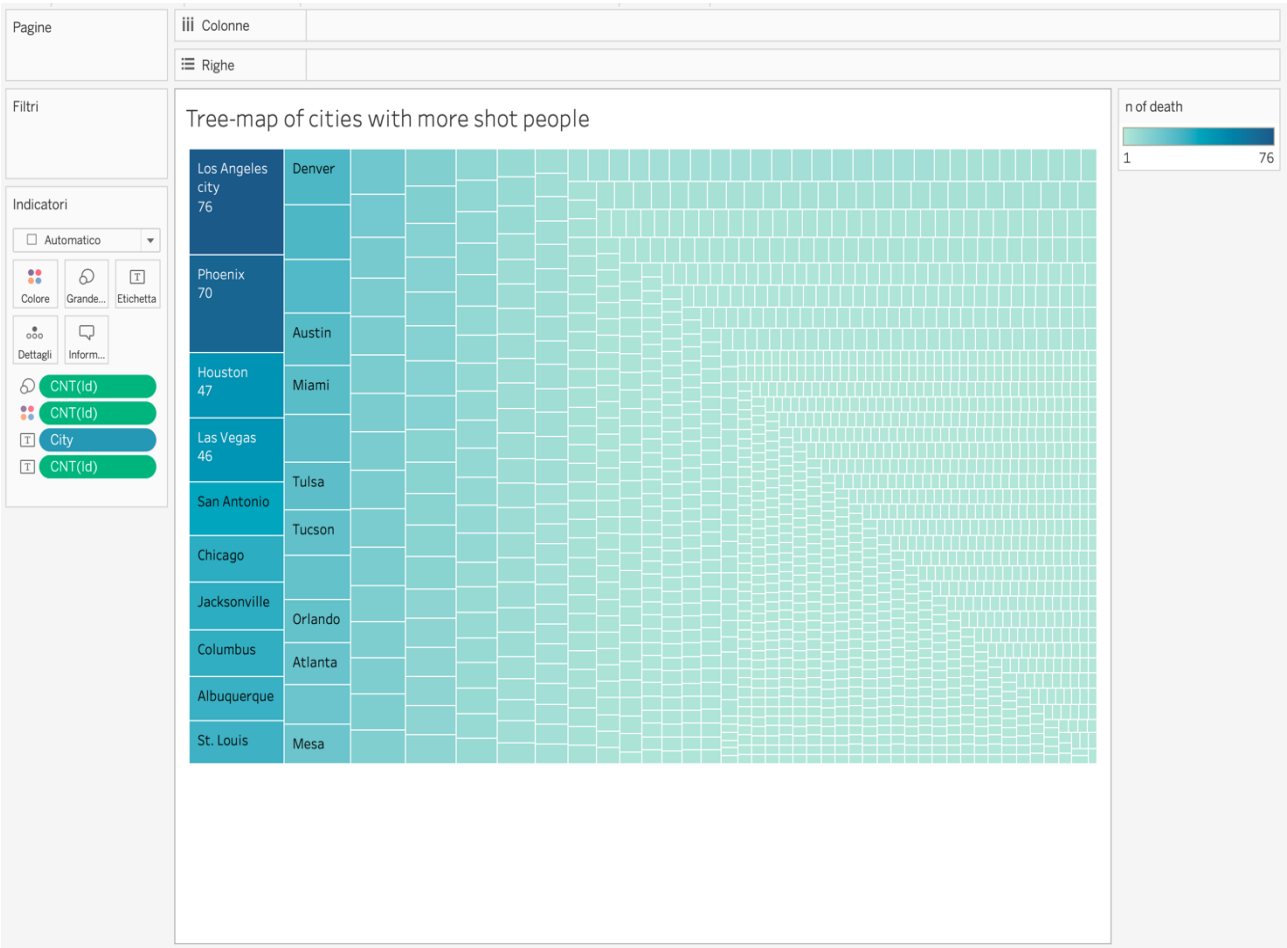
Total shooting by year



Total shooting by year sheet, shows a bar-chart in which in the columns there are CNT (id) and in the rows, the years information.

From this sheet is pretty evident that, shot people decrease almost constantly each year.

Cities with most kills

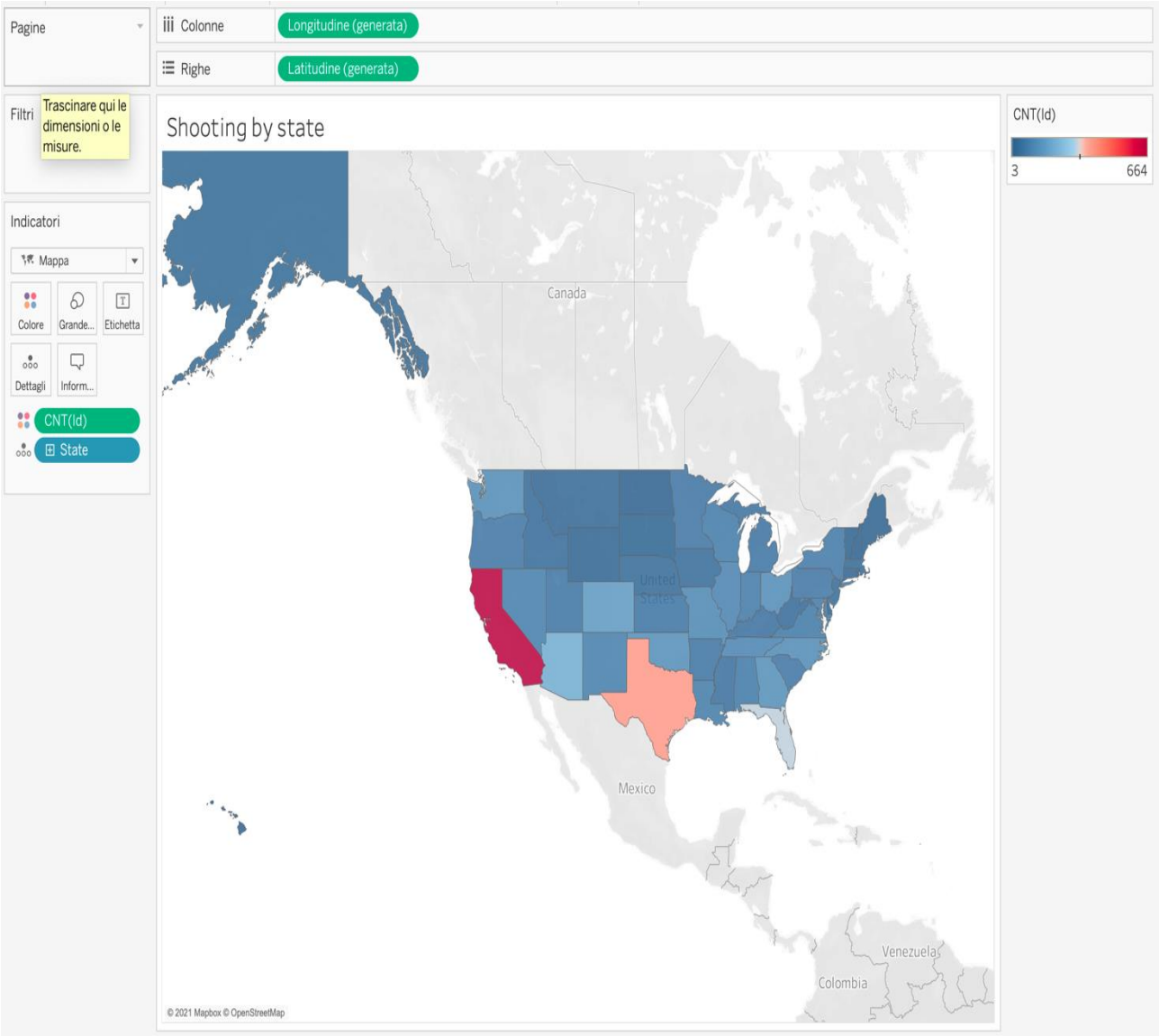


Tree-map of cities with more shot people sheet, shows a tree-map of shot people by city.

Most of the cities has a low number of shot people.

The only cities with more than 50 shot people are Los Angeles with 76 and Phoenix with 70, followed by Houston and Las Vegas respectively with 47 and 46.

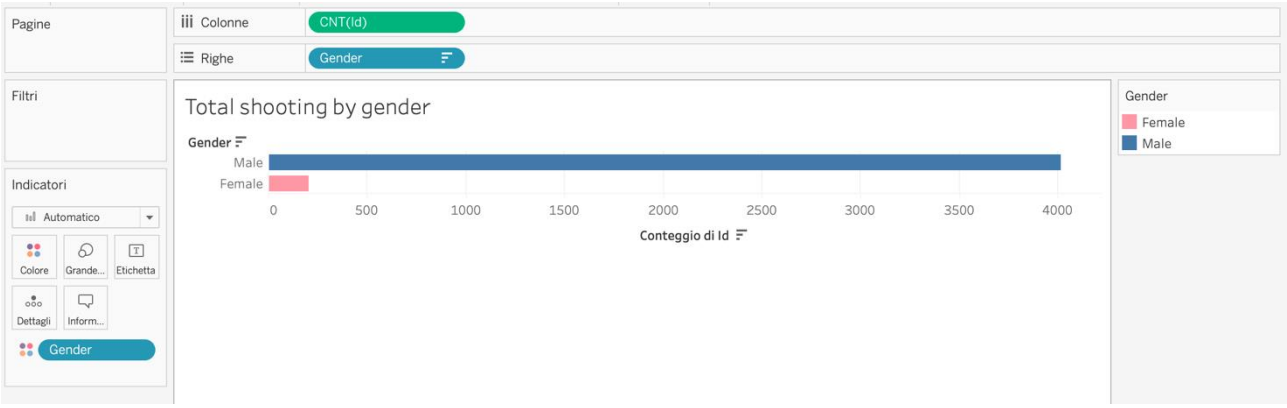
Shooting by state



Shooting by state sheet, shows a map of all US states filled using a red-blue colour indicator, giving the number of shoot people by state.

The state with most shot people is California, followed by Texas and Florida.

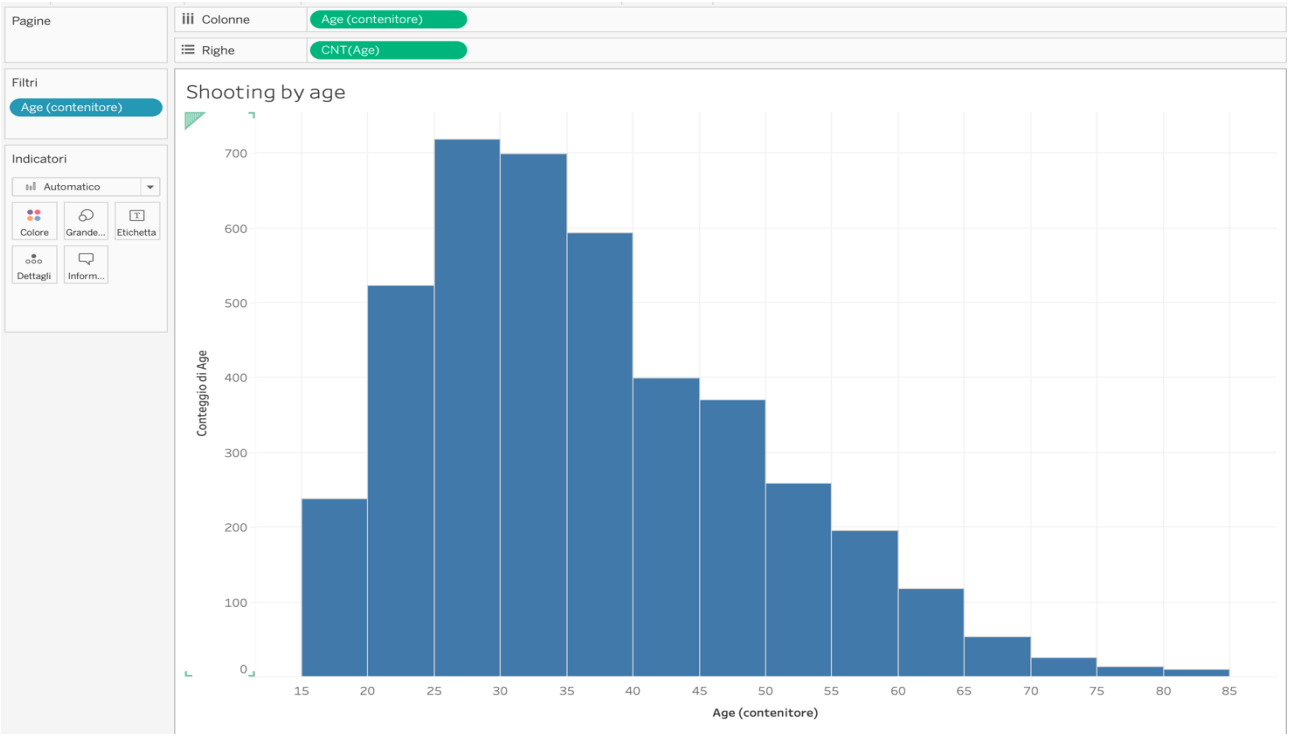
Total shooting by gender



Total shot by gender sheet, shows a bar graph with gender differentiation.

From this graph we can assume that the number of women is almost irrelevant when compared to the number of men shot.

Shooting by age

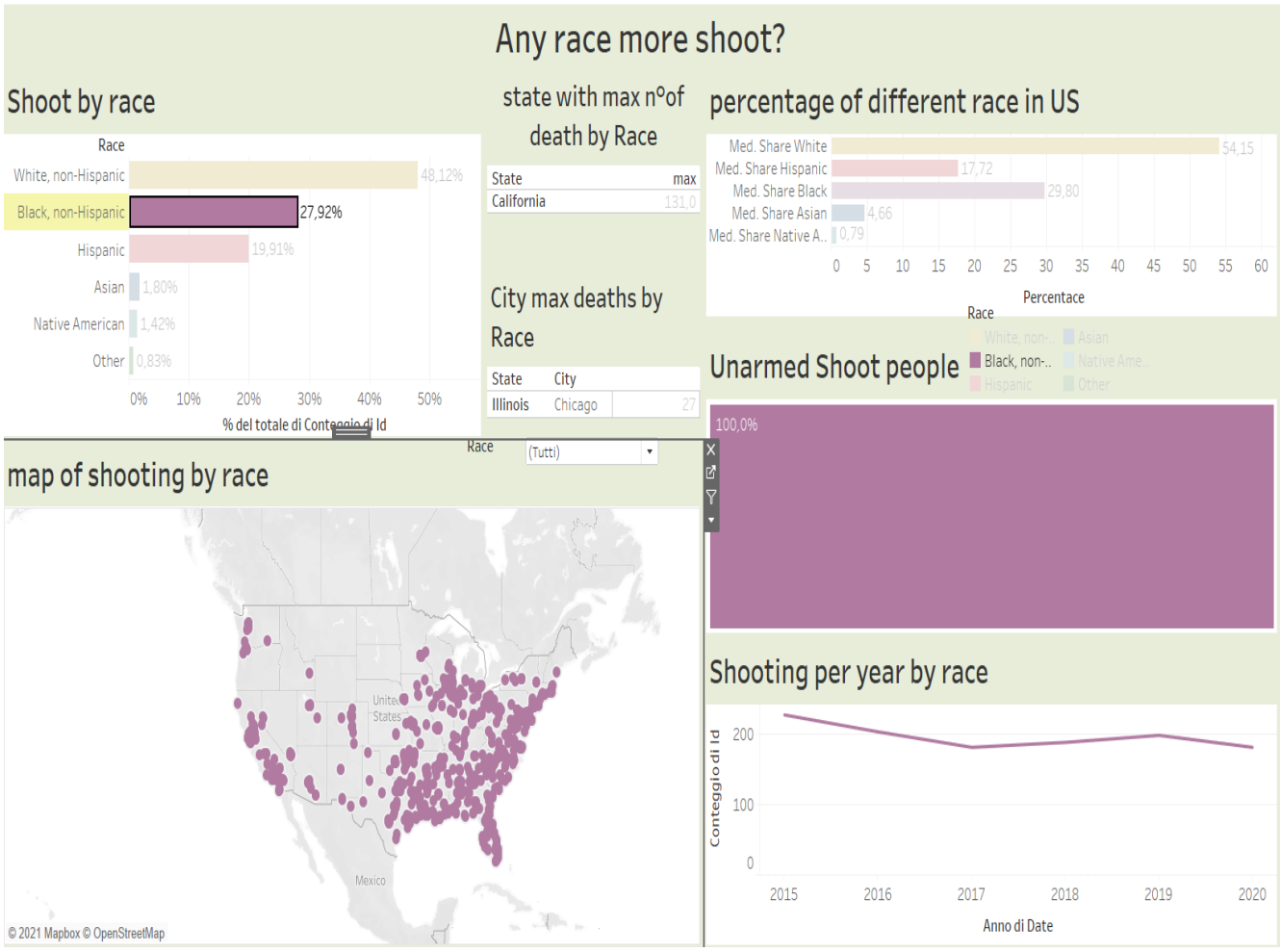


Total shot by age sheet, shows a histogram that gives us information on the age ranges of the people shot.

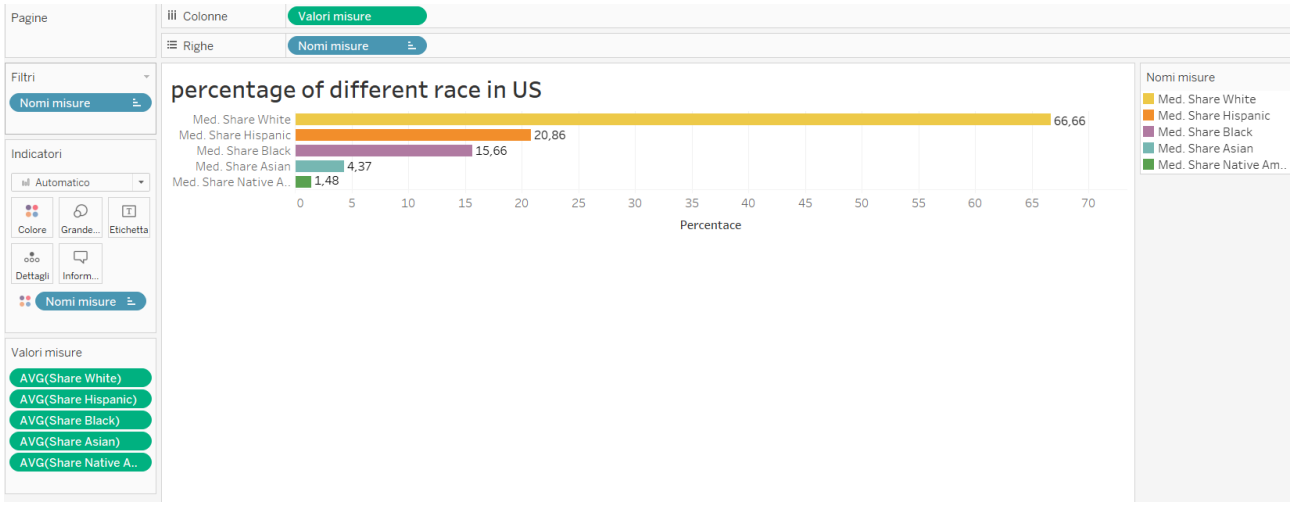
It is clear that, age groups most involved are those between 20 and 50 years of age, with a peak between 25 and 35.

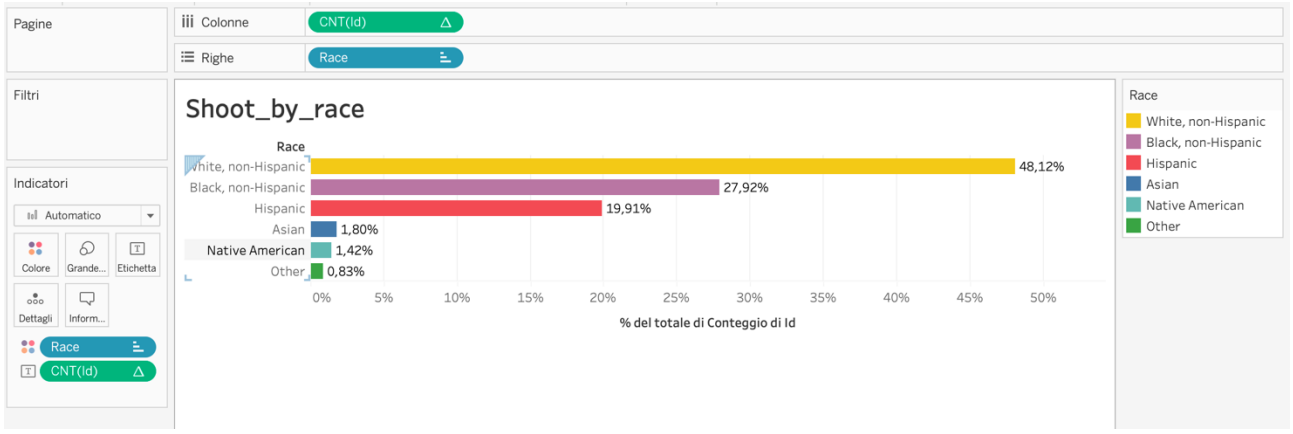
Dashboard 2

We added 2 calculate field and determine the State and the city with the highest number of shot people by race.



What is the rate of any race in US?

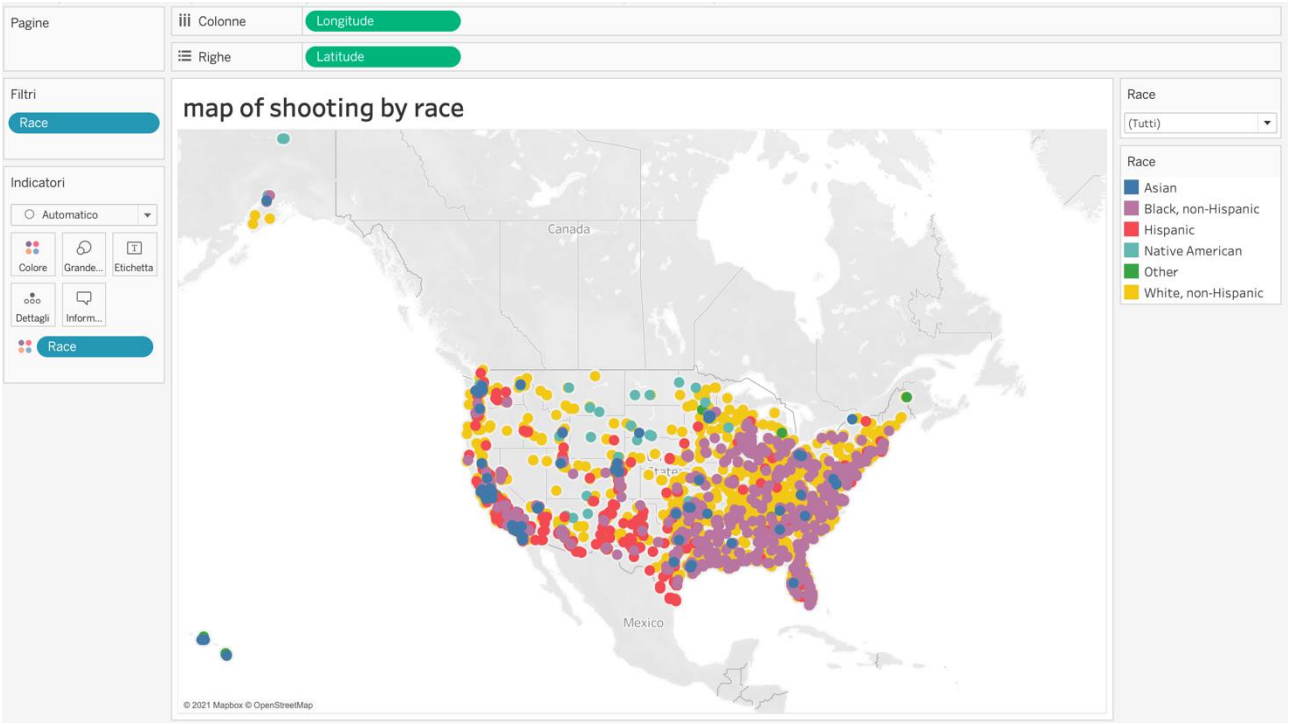




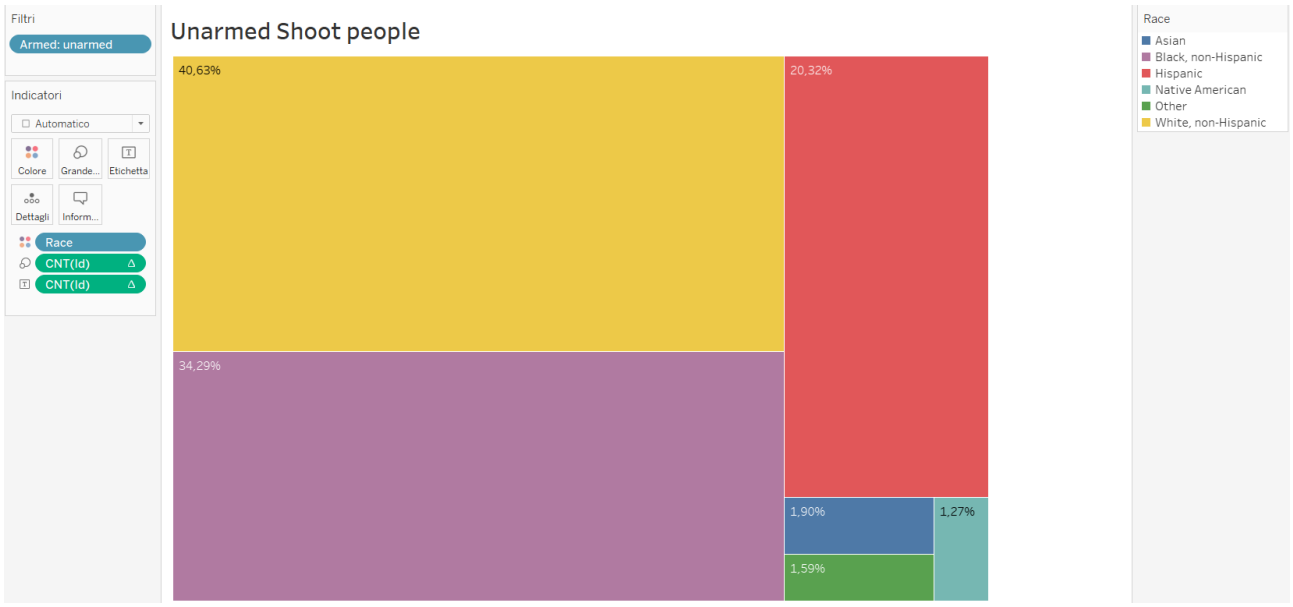
As we can see in the worksheets above built using Tableau Desktop, we can see that 66.6 % of total population (330 million of people) is White non-Hispanic on Average, 20.8 % is Hispanic , 15.6 % is Black while the rest rispectivery in lower percentage are Asian (4.30 %) , native American (1.48 %).

This means that considering the 3 major races, **the white race is 4 times higher than the Black race** and **it is 3 times higher than Hispanic race**. According to that we can

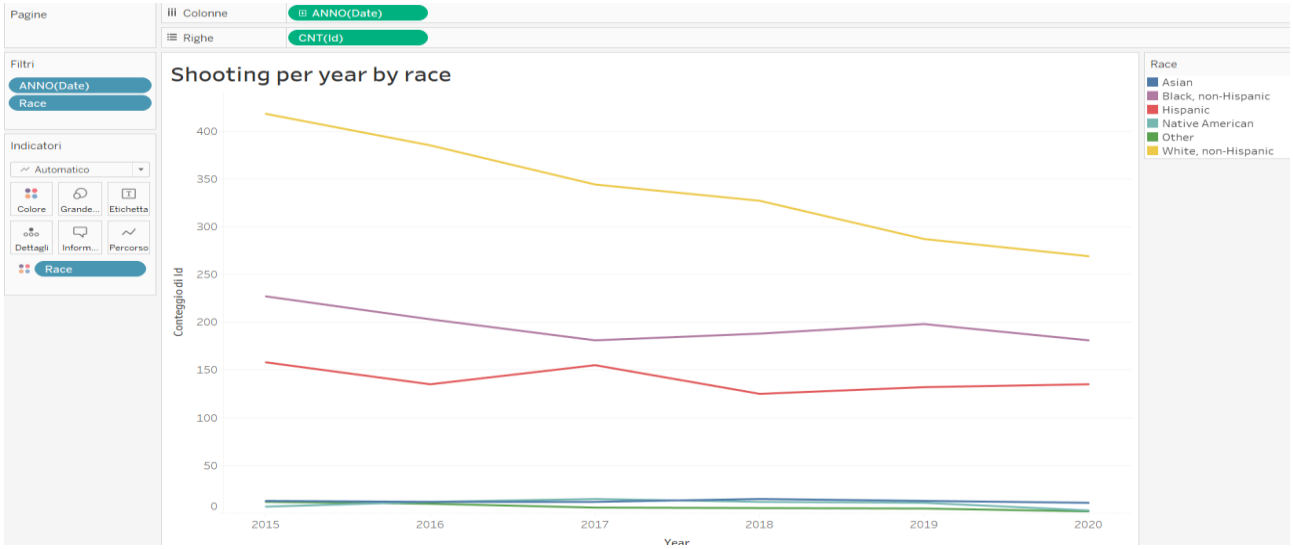
see that 48% of total shot (in total **4219 record from 2015 to 2020**) are White, almost 28 % of Black have been shot by the police and almost 20 % of Hispanic have been shot. Considering the different race in US and his own rate of presence all in the country, they have been shoot **28 Black per 1 million**, **10.4 White per 1 million** and **21.5 Hispanic per 1 million** of people.



I want to show the fact that the Hispanic people shoot are more concentrated in the border with the Mexico (where there are Hispanic race people), while the central part of US (North and South Dakota, Montana, Nebraska, Wyoming,…) there are just a low number of cases, expecially Native American and White non-Hispanic, but generally this part is that one with less shot by the police.



Another factor we want to show is the fact that according to the different races, the rate of people shot by the police that were **Unarmed** is almost similar for White and Black, this considering the fact of the previous analysis (n° of White is 4 times the n° of black) means that for every white shot there are 4 blacks shot by the police. If we consider the Hispanic we see also that every white shot there are 2 Hispanic shoot by the police.

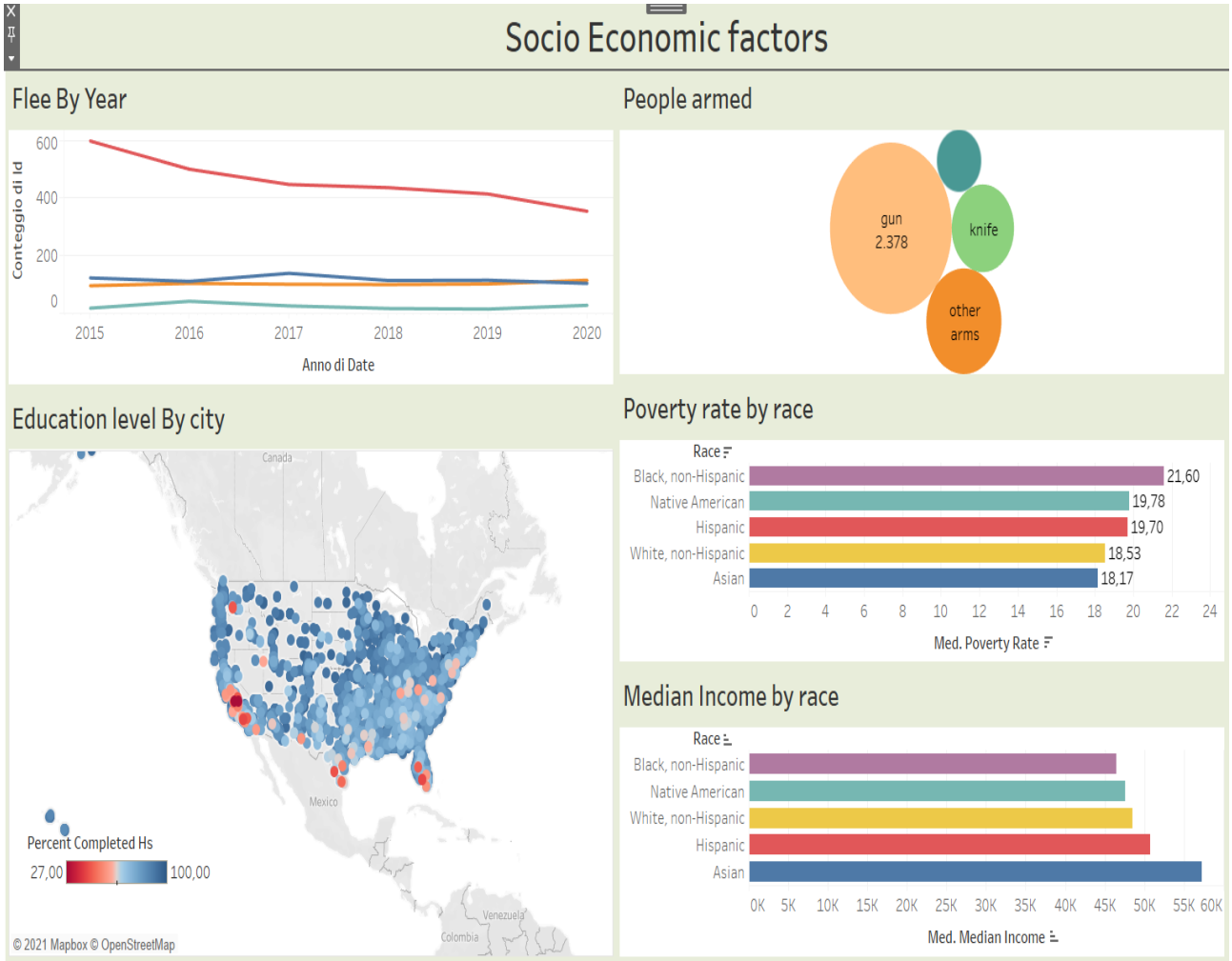


From the 2015 to 2020 we see that the number of shot people every year still costant, except to the White race that is the only to show a decreasing trend. The other 2 main race (Black and Hispanic) still maintain the same amount of people shot by the police in the last 6 years. We can see that in the last year in the black race, there is a reduction in terms of death , probably due to the pandemy.

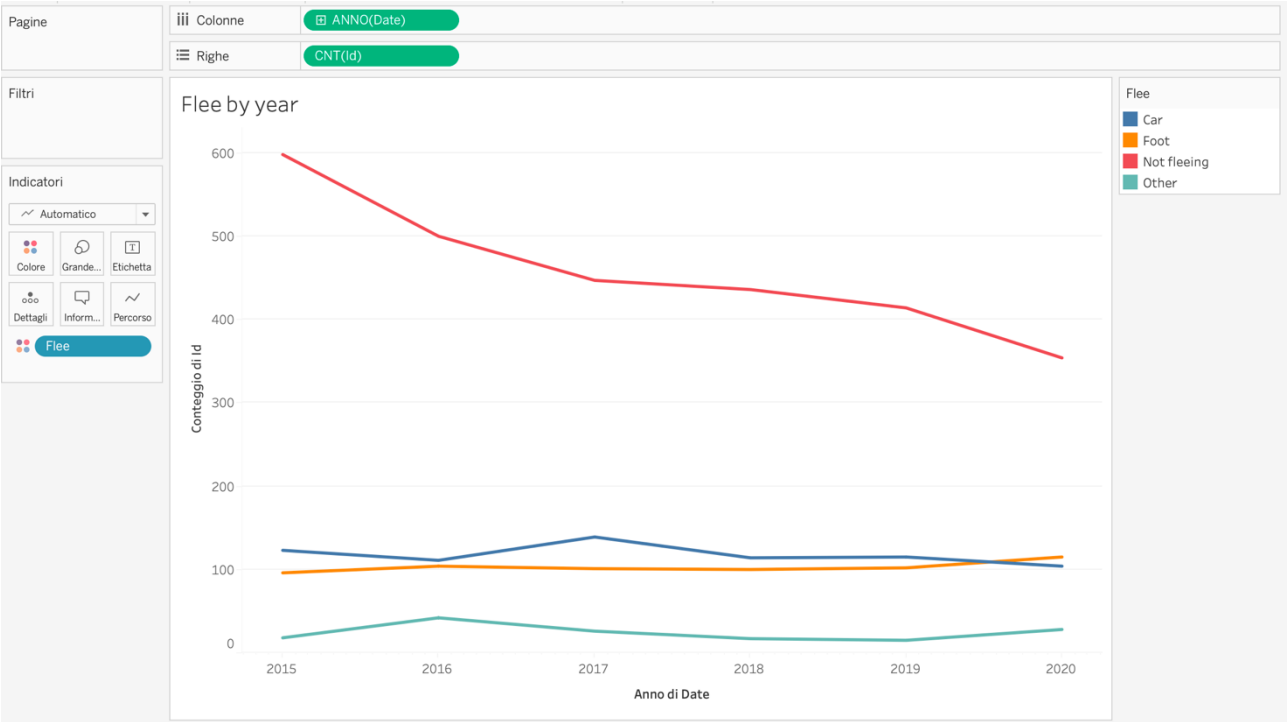
Dashboard 3

This final dashboard is focus in the analysis of socio-economic factors.

It contains 5 sheets: Flee by year, People armed, Map of education level by city, People shot by poverty rate, People shot by income.



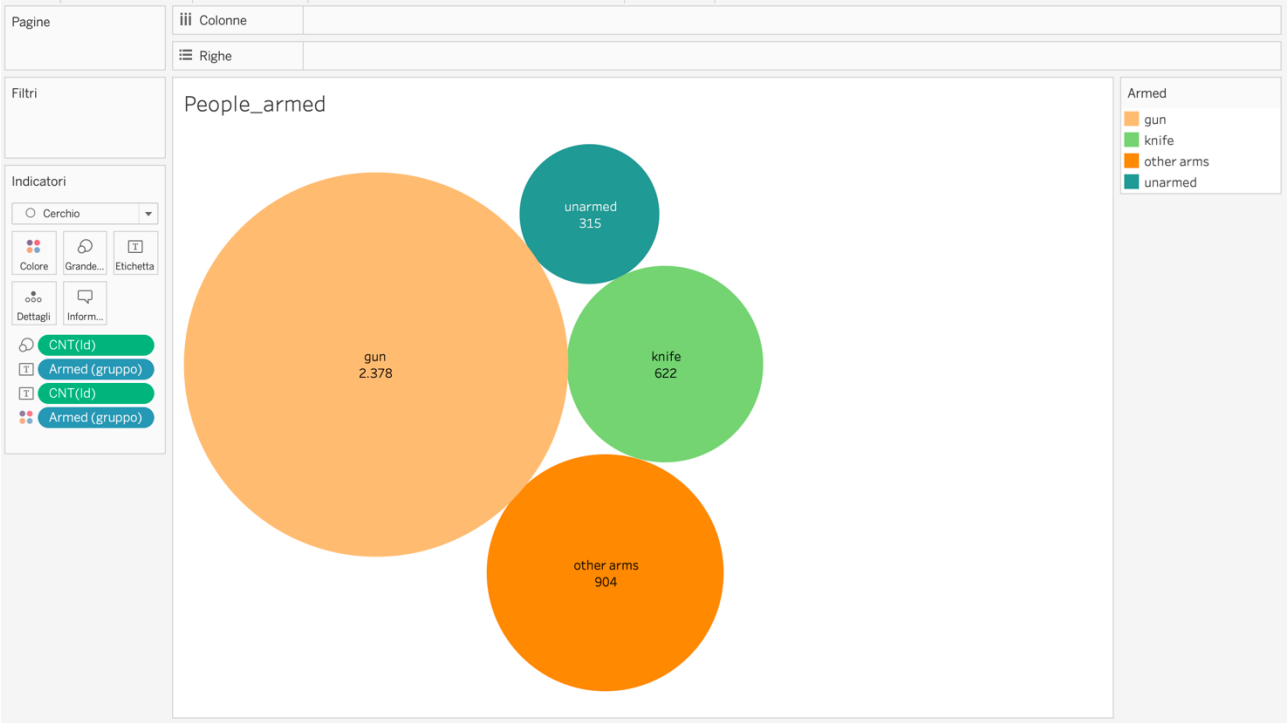
Flee by year



Flee by year sheet, shows a line-chart within the rows there are CNT (id) and in the column the date(year).

From this worksheet, it is clear that, the number of victims who do not flee is decreasing, while, in general, those who flee remains constant.

People armed

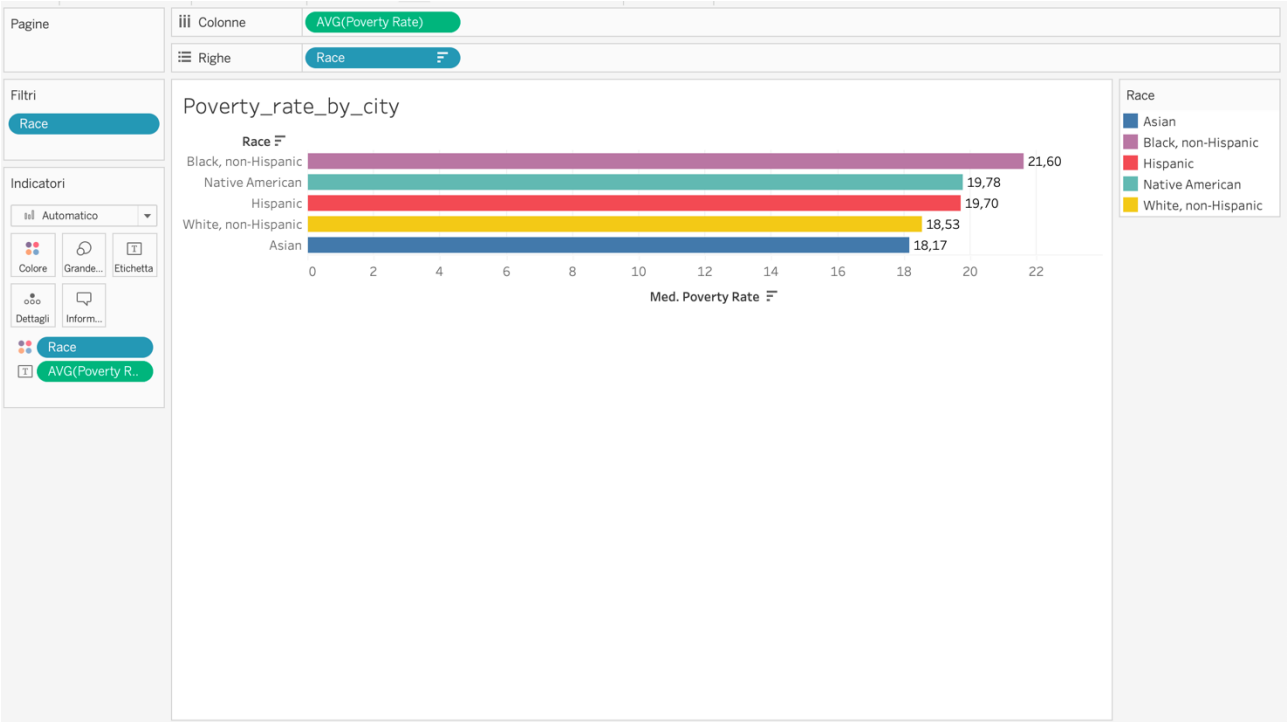


People armed sheet, shows a babble-chart that gives us information about the arms**.**

Specifically, we distinguish **unarmed** people from people armed with a **gun**, a **knife**, or **other weapons**.

It appears that most of the people were armed with the gun being the most used weapon followed by the knife.

Poverty rate by race



Average Poverty rate by race sheet, shows a bar-chart that gives us information on the average poverty level by race.

Listing them by breed from the poorest to the least we find:

Black, Native American, Hispanic, White, and Asian, with blacks deviating from other breeds in the obvious mine.

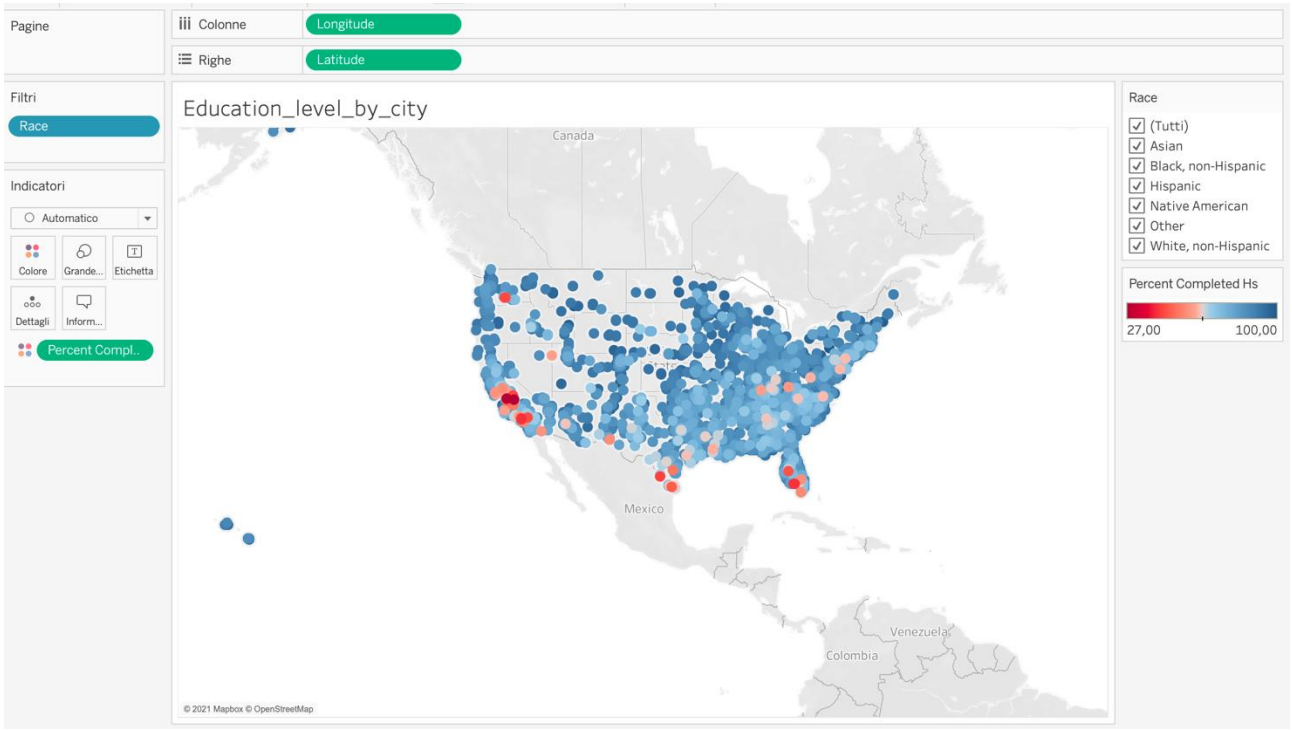
Median Income by race

Average income by race sheet, shows a bar-chart that gives us information on the average income by race.

Listing them from the lowest income to the highest we find:

Black, Native American, White, Hispanic and Asian.

Education level by city

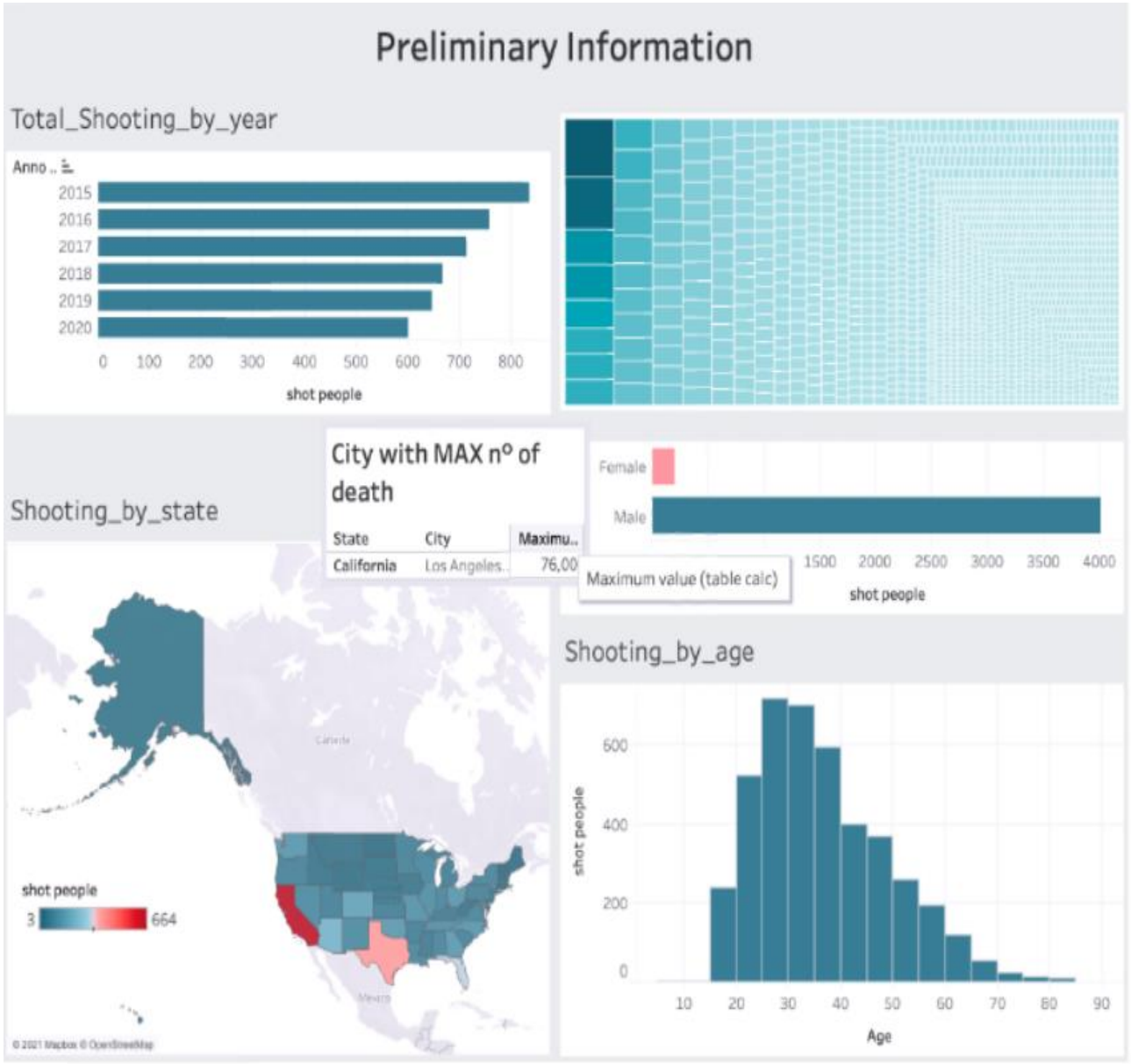


Map of education level by city sheet, shows a map of US cities filled using a red-blue colour indicator, giving the % of high school graduation rate for people over 25 in US cities.

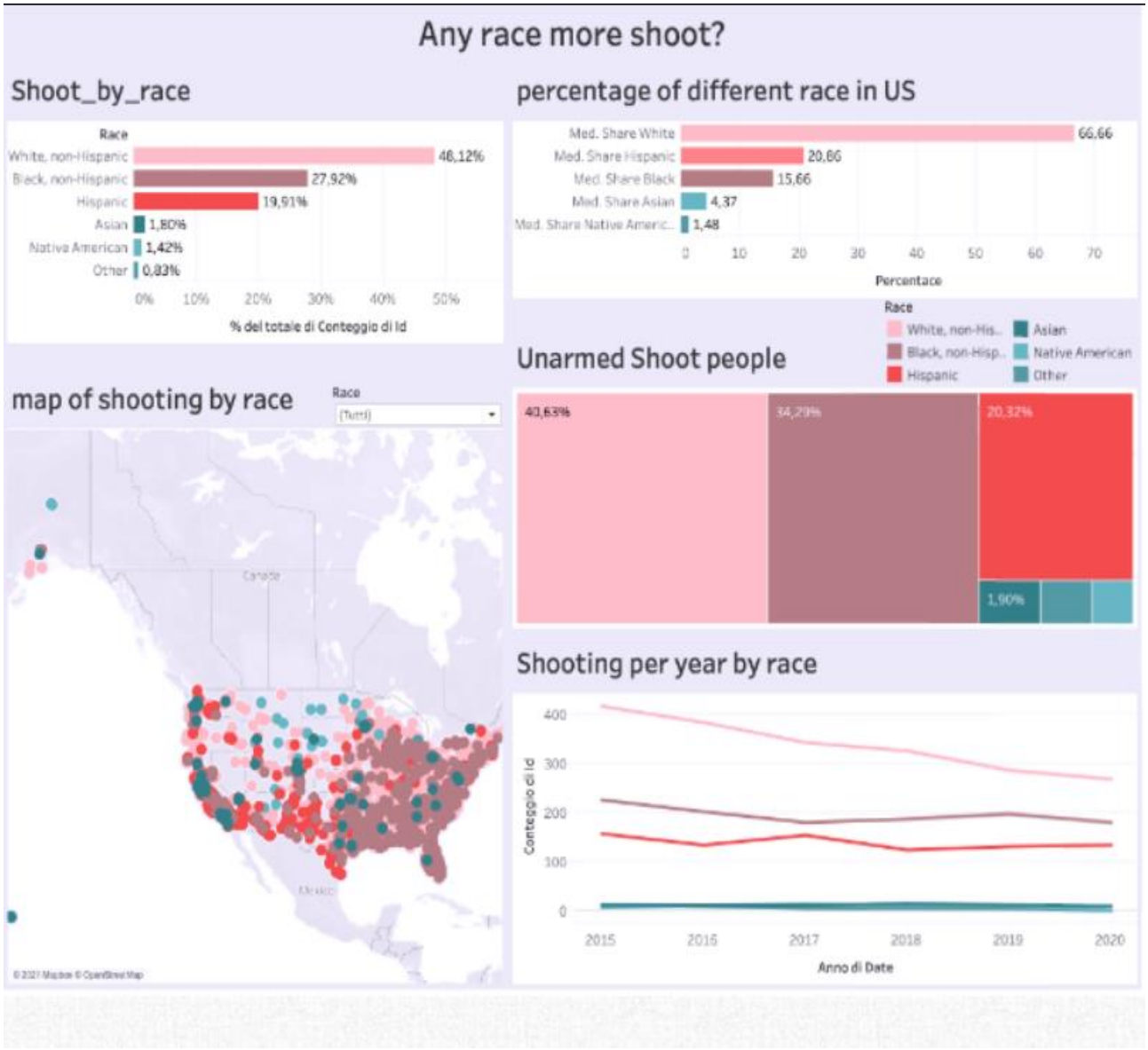
Color Blindness Test

For each dashboard, the colour blindness test is carried out, in order to evaluate whether the dashboards achieved can be easily analysed by people who have anomalies in the perception of colours. The **Coblis** simulator was used (https://www.color-blindness.com/coblis-color-blindness-simulator/).

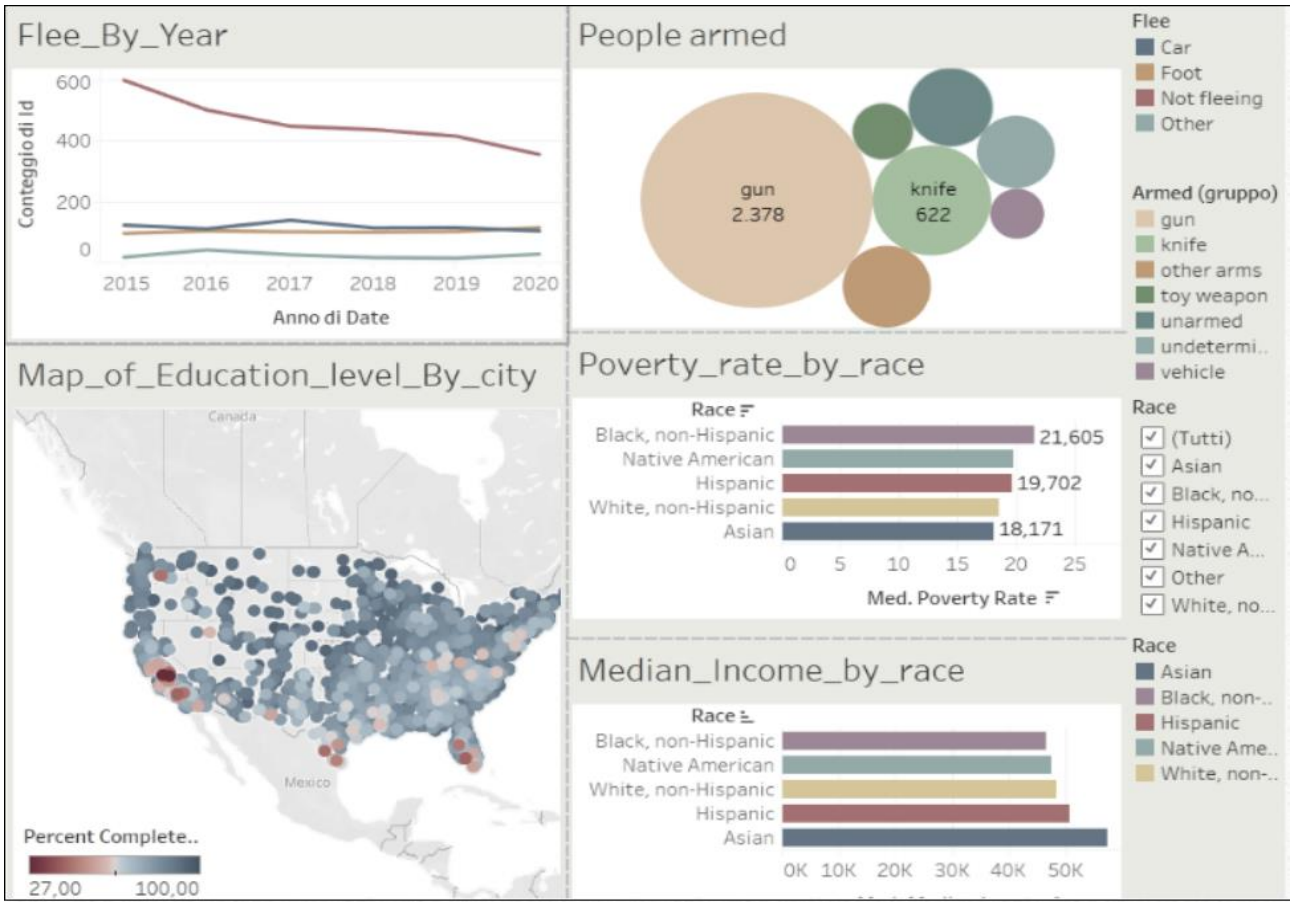
Blue -Weak / Tritanomaly



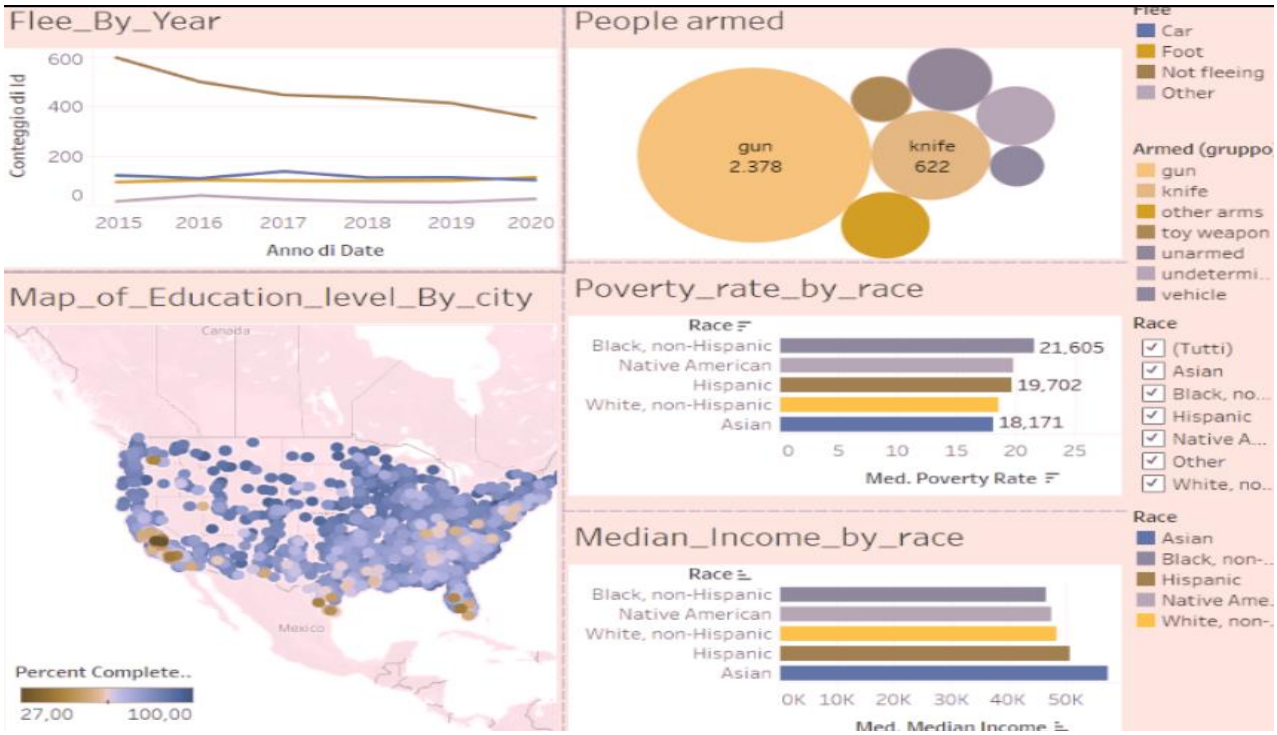
Blue- Blind / Tritanopia



Blue Cone – Monochromacy



Green / Blind Deuteranopia



CONCLUSION AND INSIGHTS

Many considerations can be made based on the resulting we have talked about, expecially to answer the main Question: " Is there any real consideration that brought the Black race to the recent spot "Black lives matter"? in effect there exist some evidence that show as the black race is that one more shot by the police under the considerations made.

In particular we have seen that even if the number of death each year in general is decreasing, this is true only for the White race non-Hispanic, while for all the other races the trand is still costant or evenly increasing. We have also seen that talking on the Black race, the main states where they have been shot are California, Florida and Texas (sort by decreasing order).

Another consideration that has been topic of discussion in the last years is the use of weapons, and expecially we have seen that, expecially the young generation, has his own gun, so this could be a relevant variable to generate a fire conflict.

Knowing that US has a population of around 330 Millions of people (66,6 % of White, 15,6 % of Black, and 20,8% of Hispanic), we can see that any million of people there are :

- **28** Black killed per **1 million**
- **21,5** Hispanic shot per **1 million**
- **10,4** White killed per **1 million**

Whe have also seen that the most poverty rate all around the US is the Black race, that corresponds to the race most shot, seems clear that there is a correlation between the economic condition and rate of death. The same analysis is made on the Median Income, and the result show as always the black race is that with the lowest rate.

Therefore, in general, according to our analysis appear clear that, the manifestation of the Black lives metter, can have reason to exist; in proportion the black race is undoubtedly the race most shot by the police.