Violence and death

The problems discussed in this part are:

- Gun Vs vehicle death rates
- Gun deaths and race relation
- Homicide vs other death causes
- Suicide correlation with age and education

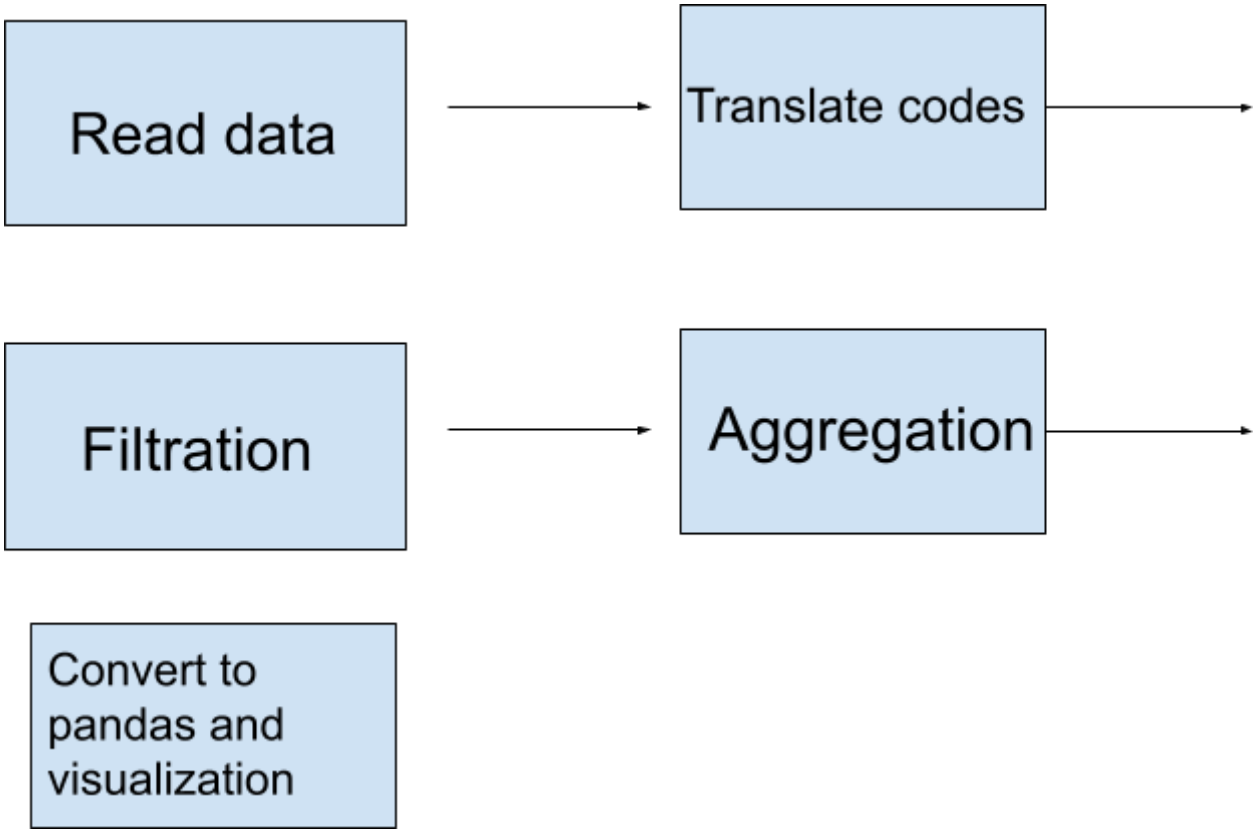
Analysis:

Because data frames in spark are optimized in time and plans it was used directly in all the analysis.

General analysis steps:

- First step in every analysis was constructing columns that translate codes used in the dataset to its values.
  Sometimes the codes were complex and required more than one column to be translated.
- Second step was filtering the required data (suicide deaths, gun death, … etc)
- Third step is aggregation (mean, percentage, count).
- Final step is results visualization using pandas.

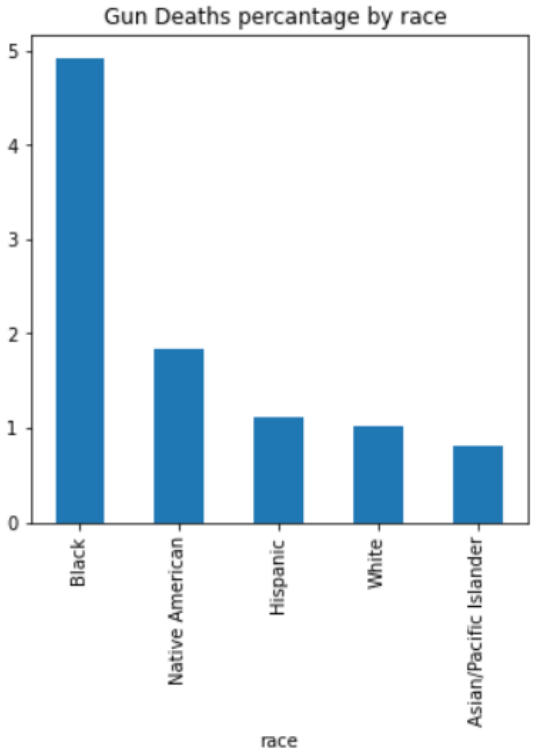
PipeLine:



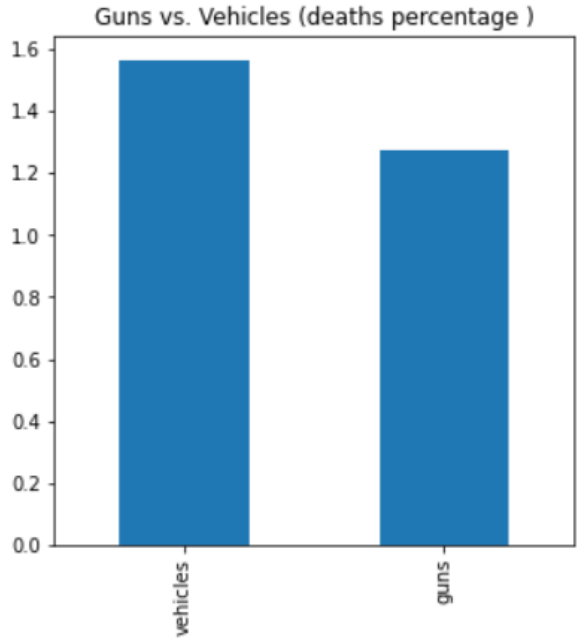
Trials not included:

- I tried to add the translation column using df.withcolumn but it didn't work with custom code translation functions; Instead I used rdds to construct a new dataframe with the new column which worked with no problems.
- There were two educational codes which complete each other, so integrating them in one column was hard and required many trials. This was treated finally using custom code to map similar education levels to one education interval.

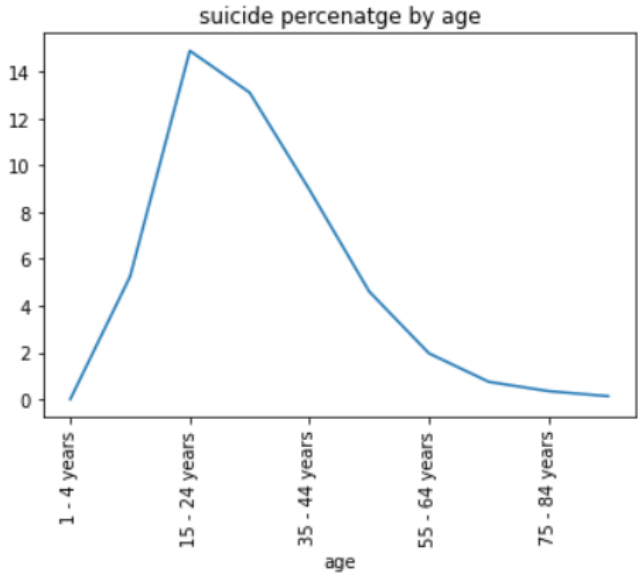
Results and evaluation:



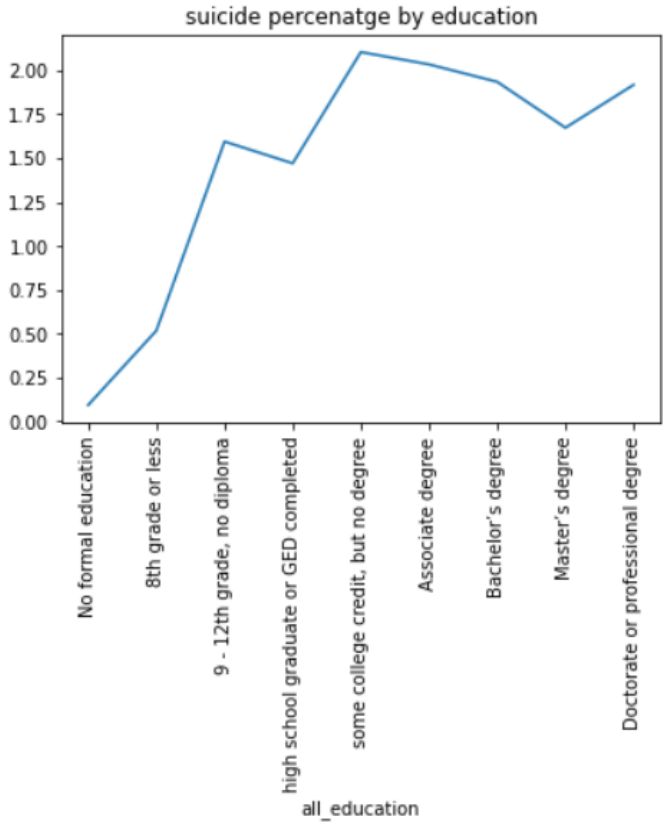
> Black are the dying with guns more than other races and then comes native americans and hispanics.



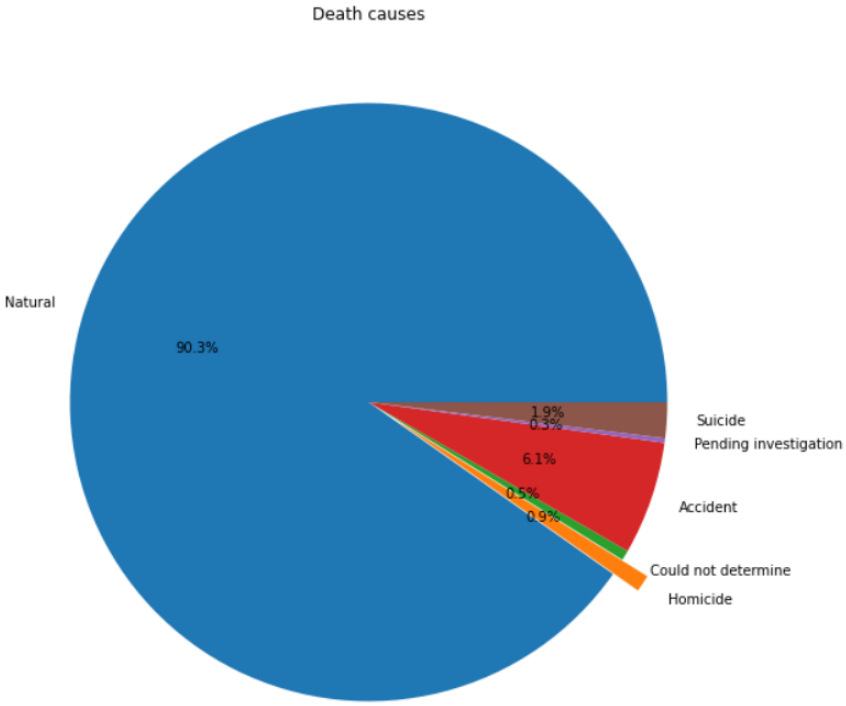
> vehicles is higher in death percentage than gun deaths by small pecentage (around .2%)



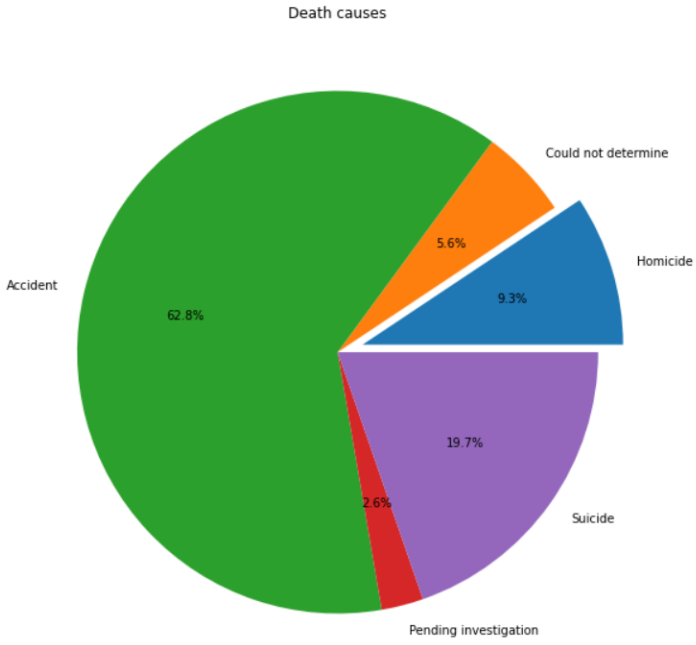
> The graph shows a peak in the 15-24 years youth for about 17.5% and then gets low again.



> It can be observed that the highest suiscide percentage is among college students with no degree
>
> This high rate may be related to the age range for college students (20-24).
>
> The rates remain high with higher degrees although they are olderin lower suicide rates.



> Natural causes of death make 90.5% of the total death cases followed by accidents for another 6.5% and the remaining 3% is for other death causes



> Excluding natural reasons for death homicide is 9.3% of the overall death cases, as accidents are the most common death cause with 62.8% percentage.

Enhancement and future work:

- Some enhancement could be done by exploring more points of view for different proposed questions to test our hypothesis on the observations.
- Statistical tests (like p test) may be carried to test the hypotheses proposed too.
- Analyzing different relevant dataset to explore more points of view and to see if the observations will be consistent.