Video Games vs Crime in the USA

Manuel Pinto
IST - MEIC
75973
manuel11.pinto11@gmail.com

Rodrigo Simas
IST - MEIC
81536
rodrigo_simas10@hotmail.com

Rodrigo Santos
IST - MEIC
94036
rodrigo.g.santos@tecnico.ulisboa.pt

INTRODUCTION

The purpose of this work is to address a common theme that has been talked throughout the years, especially on the media in the US – How violent crime, as well as mass shootings, are linked to the increasing popularity of videogames, specially of the violent genres. This hypothesis has been talked by many public figures and been overexposed to a point where some people claim it as an essential truth, while others see it as complete nonsense. In the end, they're mostly theories, as there hasn't been conclusive evidence of this link. This is what we are trying to do with this work, we are trying to find if there's any link between these two phenomena and provide a visualization so that people can draw their own conclusions. More important than just trying to prove it, is to make sure people can understand what's going on and generate their own logical conclusion on this thematic.

We are approaching this at a macro level, which is, we aren't trying to do a behavioral study or a psychological profile of these situations but doing it with pure data on the USA's demographics, and combining them with data about the popularity of videogames.

Our main goal with this visualization, is not only to provide new answers, but mainly to educate the public. We want to provide a simple way to draw conclusions with consistent visualizations, backed by true and consistent data that was carefully researched and handled.

With this, we hope to be able to deliver a consistent and true visualization that, not only provides people with customizability, but also answers some smaller questions that people might have.

Here below are a couple of examples of questions that people can find their answer, through our visualization:

- What were the most popular genres of videogames, throughout the years?
- How does a genre's popularity correlate with crime?
- Did shooter games lead to a change in violent crime?
- Are video games referenced as a cause for mass shootings?
- What state has the most incidence of mass shootings?
- Which years had the most popularity of violent genres, such as Shooter or Action games? Did that lead to a change in crime?
- Do spikes in violent videogame popularity lead to spikes in crime?

RELATED WORK

After our motivation was found, the main goal was to find how other visualizations in this area did this approach, and gather some inspiration on them, with hopes to improve on them. The conclusion to this search was that there was much more scientific research on this subject, supported by data, as well visualization techniques on it. Here we found four essential problems:

1. Research was mainly focused on one specific topic (e.g., "Effects of violent video games in critical thinking" [1])
2. The connections of these two themes (being violent crime and violent video games) are mostly subject to academic scrutiny and are, in a way, inviable as a method of showing the "casual audience", the conclusion of this subject.
3. Instead of visualizations, this subject is commonly talked throughout the media, so multiple articles talk about this subject, with little to no references, and also provide very little visual reference on the conclusions that are drawn, even when they're valid.
4. Whenever visualizations were present, they were mostly one or two idioms at max, and not a unified view. These were used mainly to support arguments in either articles or research, and not as a tool to educate people on the subject.

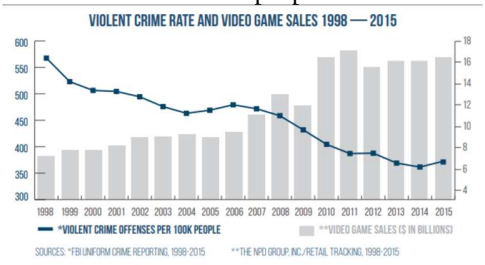

Figure 1. Example of an idiom on violent crime rate vs videogames during 1998-2015.

By diving each problem into a sub-category, we can better analyze how a full-on visualization can correctly tackle these problems.

Research is ultra-focused on one specific topic

Our main findings were on very specific topics about our current theme, so the main goal here was to give users a much broader view. With this in mind, the best way to approach to approach this issue would be to give the users as many information about the subject, without being too cluttered. So, it is a good idea to show, per example, multiple types of crimes and video game genres, so that users can do their own correlations based on multiple topics, without being constrained into one domain.

Research being too "academic" and not friendly to a casual audience

The simple act of providing a visualization with an intuitive interface is much better for the casual user than an academic report. While we can't offer the in-depth knowledge that such research includes, this is not our goal. This also offers the chance to contribute as a supporting act that confirms this research and provides a way to visualize the same conclusions in a much more friendly way.

Most of the information on this subject is in the form of media articles

These represent a very informal way of researching on the topic, and mainly provide a signal on how the public reacts to this subject, as well as the main questions that they have. By knowing these questions, we can adapt our visualization to provide better answers to them, and specially making them easier to see. While being very challenging to get all the questions answered, we believe that it's possible to provide answers to the most frequently asked ones.

Existing visualizations are made of one or two idioms at max and represent a fragmented view

Our main improving is essentially here. While a good visualization composed by one or two idioms can answer some great questions, it doesn't represent a full view on the matter, simply because it lacks the space and information to provide a lot of the answers for the many questions that users have. By providing multiple idioms we can better tackle this subject without leaving out a lot of important sub-subjects to talk about.

Relevant Research

Besides the problems referred previously, multiple research on this subject has been found that provide an interesting view on this issue. One of our main motivators was a research entitled:

"Understanding the Effects of Violent Video Games on Violent Crime" [2]

This research dives deep on the inner psychological phenomena that links video games and violent behavior together. The results are both surprising and not surprising. Our main expectation was for very little to no correlation to be found, but the main conclusion actually points to the fact that violent videogames actually have a positive effect in the crime rate, which means, violent video games popularity rose throughout the times while and overall violent crime decreased. This not only disproves what the traditional media publicizes, but also utterly contradicts it.

So, the challenge here is to take research like this, which already has proven data that shows this behavior in a visualization with multiple vectors of sight and enabling the users to see what is really going on, leveraging on facts provided by real-world data.

At the design level, our inspiration mainly came from examples on the subreddit [3] *r/dataisbeautiful,* as well as examples of the best work made by students of this course (Information Visualization at MEIC-IST), mostly in the Hall of Fame section.

THE DATA

As important as the design of the visualization is, its fuel is data. With that in mind, data becomes the single most important thing about this visualization, so our approach was to leverage on data that was very consistent and based on official input. Since our main demographic is to show this study specifically in the USA, we made sure to gather the demographic data from official governmental sources, while with videogame data we leveraged on sources from the community of the website Kaggle[4] , as we found this approach sufficient for our needs.

Our idea with Video Game data was to focus on sales as a main indicator for popularity, as this is a great way to show the cultural impact of such videogames and its overlying genres, resulting in more consistency and truth within our visualization.

We ended up with three main datasets from which we could derive from:

- Crime in the U.S – A dataset derived from the Summary Reporting System[5] (SRS). This reflects the estimates the FBI has traditionally used for its annual publications. This comprises of data between 1976 and 2016.
- Mass Shootings in the U.S – This data has been crowdsourced from many individuals and every row can be retraceable from a simple Google search. Contains 308 mass shootings from 1966 to 2017, with their descriptions, casualties, location per coordinates and many other useful data.
- Video games sales – A dataset gathered by the community of Kaggle that contains over 55000 videogames with sales, critic and user scores.

Each dataset was carefully cleaned and treated in very specific ways to make sure that data was consistent and ready to be imported into D3.js. Every single file was also in the CSV format, separated by commas, which was our preferred approach since the original datasets were already in this format, without requiring standardization into other ones. It was also decided that the time frame in which we would analyze the data would be between 1976 and 2016, as this represents the period of time in which we have enough data to draw conclusions from.

Crime in the U.S

This dataset was very consistent and didn't have any missing data. What we found were essentially a couple of rows that weren't needed, such as "caveats", so we simply removed those and added a unique identifier, similar to the already existent ones from each state in the U.S, but pointing to the overall data. This allowed us to represent it as a "per state" approach, as well as the U.S as a whole, and help with the identification of specific types of crime that have a better correlation with videogame sales. We transformed this data using Pentaho's Spoon interface, into the same CSV format, and imported it to D3.js, sorted by year and state.

Mass Shootings in the U.S

The nature of this dataset is very aggregate, but also has a lot of information. Most of it wasn't needed, but it also shaped our ideas into what we could use for visualization purposes, such as the description of a mass shooting, to see if there were any signs of behavior modelled to videogames.

On a first stage, we found data inconsistency, as some of the states were missing from the original dataset, so we had to search for the location of that specific mass shooting via the coordinates that were given by the dataset and add the state in which they occurred.

After dealing with that situation, we realized that some columns were also pointing to null values for injuries or race/ethnic information, but since we didn't plan to use most of those values, we either replaced them with N/A or with 0, depending on the situation. In this process, we didn't eliminate any rows (keeping the same volume of data), but around 70% of the columns were eliminated.

Video game sales

We leveraged on a dataset that showed sales of videogames that was crowdsourced from Kaggle. One of the biggest problems with this was that there was a lot of missing data, and this truly crippled what we could do with our work. As we realized that we couldn't work with ESRB ratings to find games that were potentially violent (due to the number of null rows), we decided to scrap them and remove them from the database.

With the idea to provide users with more customizability, we decided to use genres as a way of differentiating the sales metric. When showing genres vs sales, we hoped for users to understand how a genres popularity evolved throughout the years in comparison to crime data. This was also a way to work around the lack of data in some specific columns, while using the ones that were more complete. This ended up shaping part of our visualizations as well as the information we are offering. From this dataset, we also joined a couple of tables with sales numbers, as some had global sales, while others were separated with regional data (e.g., NA sales, EU sales).

Generalized calculations on the datasets and derivations

We tried to derive some questions from the datasets. Here's a couple of examples of derivations and why we decided, or not, to use them:

- **Number of video game sales per year** – This ended up being very useful to line up data for a linear time read. We categorized these into the specific genres that were available. We realized that mainstream games (with big sales numbers) had a huge presence in this database, and while they're very good to individual trending games, it fails to predict an industry's trend, so we decided to scrap and use another metric instead.
- **Number of different videogames sold per genre** – Our preferred metric to evaluate the trend the video game market had that specific year. When a specific genre is popular, a large amount of different games of that specific genre are made (e.g., around 2008-2012 shooter games featuring zombies were massively popular and lots of different titles were made in according to that trend). With that idea in mind, we decided to rely on this metric as we believe it accurately predicts cultural phenomenon better than straight up sales.
- **Ratio of video games sold per year with ESRB rating Mature or Above, or Teen and Below** – We decided to scrap this metric since, as noted above, we had too much missing data to make it reliable as a metric, so we didn't want to sacrifice the validity of the rest of the work.
- **Crime per 1000 capita** – This was important to get a better visualization out of the data. By using a ratio such as "per capita", we can compare the different types of crime, and better spot their trends, even if they're very different in size, related to one another. With this we can restrict the axis in which we show the data and provide a much better scale to compare trends and types of crime.

With these strategies, we were able to provide a visualization that was concise and true. While we didn't derive a lot of data, the datasets were good enough to draw conclusions straight from them raw data, shifting our focus into perfecting the design for the solution ahead.

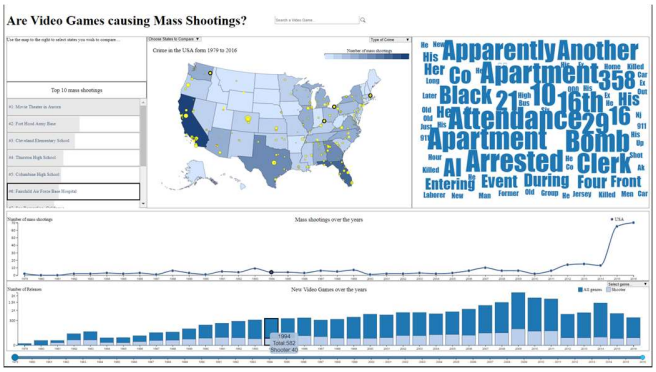

Figure 2. Sample of our whole Visualization.

VISUALIZATION

After sketching a paper prototype, and improving on it, our final solution was made leveraging on HTML5 + D3.js. As an important note, this visualization was developed by using Chrome and Windows 10 as our test environment, so best results are achieved with this browser.

Overall Description and Rationale

On the beginning of this page, Figure 2 represents our proposed solution for this context.

This solution is divided in two parts, from top to bottom:

- On the top three quarters, we have the crime section, with a map of the United States, a state selector for you to interact with the idioms, a list of the top mass shootings, a word cloud with the most frequent words from descriptions of mass shootings, and finally, a line chart that shows the types of crime per 1000 capita, dividing per state and type of crime.
- On the bottom we have the videogame section, composed by a bar chart that compares the number of new video games per year, and divides the bars into two ones, with the first (smaller one) being the genre of video game that we selected, and the second all representing all videogames.

On the bottom of everything we have a timeline slider. This slider is responsible for changing the scope of the time we want to analyze. Sliding it changes all the data from the idioms we mentioned above.

We are going to describe the functionality of each of the idioms, as well as their interactions with other idioms.

Choropleth Map

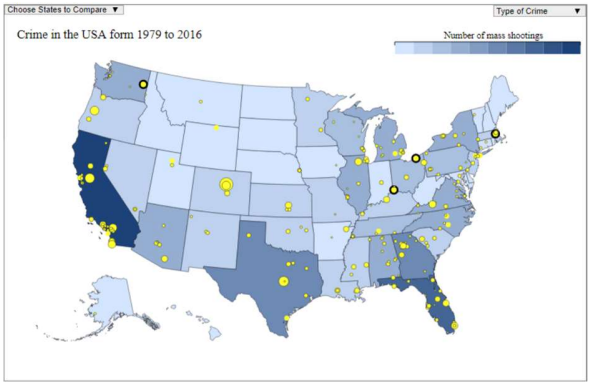

Figure 3. Map of the United States showing incidences of Mass Shootings, either their location as well as their distribution per state

As mentioned in Figure 3, the Choropleth Map of the U.S shows information related to mass shootings, however, it has some features that provide interactivity over other elements. Clicking on a state will select it for comparison, and alter the information in the line chart, as well as the word cloud as the descriptions will be different.

We decided to use a continuous color scheme that relates also to the overall theme of the visualization, predominantly different shades of blue. For the locations to pop out, we decided to use yellow as it pops out extremely well within blue. The size of the location is also different depending on the casualties. These markers can be hovered and the year where the mass shooting happened will pop out in the line and bar chart, as well as in the Top 10 Mass Shootings List.

This visualization was possible due to the completeness of our mass shootings data, as it included the exact location of each occurrence. Scalability is an issue once we go past a huge occurrence of mass shootings, but since the list stops at around 300 cases, this isn't a problem.

Top 10 Mass Shootings List

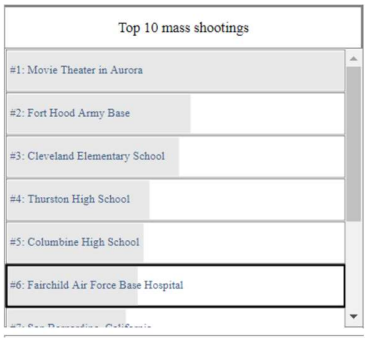

Figure 4. Map of the United States showing incidences of Mass Shootings, either their location as well as their distribution per state

Here we can represent a list of the top 10 mass shootings that happened between the years represented by the timeline slider. Whenever we hover on one of the mass shootings, its

location appears on the choropleth map, their words pop out at the word cloud and the year in which it happened pops out in the line chart and bar chart. We believe this can correctly tell users where each of the occurrences happened and if it can be linked to the popularity of videogames.

Again, the mass shootings dataset was complete enough to provide this, but this doesn't represent the most complex idiom, being just one simple curiosity.

Word Cloud – Mass Shootings Descriptions

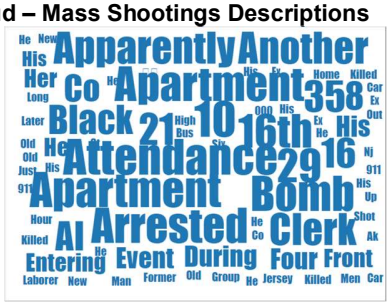

Figure 5. Word cloud of the descriptions of mass shootings related to selected years and states

We used the mass shootings dataset to provide with a visualization regarding their descriptions, into a word cloud, where the most used words are made bigger and bolder. With this we hope to try to find clues if video game market trends, and crime trends can correlate part of this behavior. Per example, a high focus on the word "bomb" with a peak of the popularity of "action" game genres, as well as violent crime such as larceny, can be a correlation.

The color scheme is consistent with the visualization's design, and we decided to use the same color as we realized that simply showing a size difference would be sufficient for people to realize the message.

Selecting different states will make different words with different colors, regarding the color attributed to the state, to show in the word cloud.

Line Chart – Crime

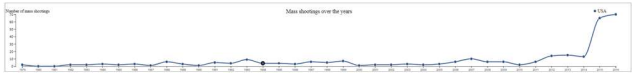

Figure 6. Line chart with the type of crime, as well as its trendline throughout the years

By leveraging on the U.S. Crime dataset, we did a line chart that provides a visualization on how the trend on specific types of crime evolved. By using the state selector, you can pick which states you would like to see on this area.

The colors are attributed by the state selector. Hovering on each of the dots in the line chart will give you the exact number, as well as the state in which it is attributed. It will also make the corresponding state and year to pop out in the other idioms.

We believe this quantitative approach reaches our objectives, either as a metric or as a trend visualization, as it can be constrained by time with timeline chart and easily manipulated within the visualization, to provide a lot of answers.

Bar Chart – Video Games

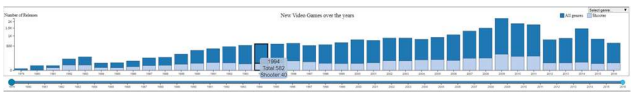

Figure 7. Bar chart with the trend on video games throughout the years

The bar chart takes the information on the videogames database and derives the new video games per year, selected by a relative measure on the whole spectrum vs the genres. With that we can find a good relative measure on how the videogames qualify as a trend on popularity and compare with the data on the upper part of the visualization. By hovering on the bars, we get the concrete numbers on the new videogames, as well as a pop-up event that shows the exact number of new video games sold on that specific year.

Our color usage is two different colors for the whole year of new video games, and the separation per genre. We used a special logarithmic scale to make sure that these trends can be correctly visible, without jeopardizing the overall quantification of the view.

DEMONSTRATION – USE CASES

We are demonstrating a couple of use cases, depending on the questions that we set, that demonstrate the full potential of our visualization. Our main goal was to find out if some of the saying of mass media on how video games have been provoking crime are true or not, so with that in mind we selected some questions that can be correlated with the data that we gathered, as visualized with our solution.

Did Shooter Games lead to a change in Violent Crime?

Let's try selecting the Shooter Games genre, and see a trend around the year 2005 and 2016, as this is where Shooter Games reached much more popularity.

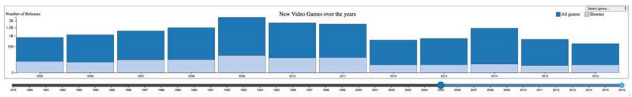

Figure 8. Trend of new video games sales, between 2005 and 2016, on the Shooter Genre

As we can see here, video games as a whole had a huge popularity boost around 2008, with the Shooter genre joining peak popularity in there as well. While hovering on the bar of that specific year, we can try to spot the trend on the line chart below and selecting "Violent Crime" as the one we want to visualize.

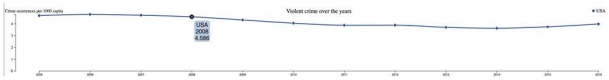

Figure 9. Line chart trend with Violent Crime for the whole U.S selected

It's actually surprising, but we can easily spot that Violent Crime suffered a great drop around the 2005-2010 time, while video games went through peak popularity, specially the Shooter Genre. With this in mind, we can see that Violent Video Games might actually have a positive correlation, which means, the more popular shooter genres are, the more violent crime decreases.

Are video games responsible for a spike in mass shootings?

To do this, we will try to rely on the whole visualization. Let's try to see how mass shootings have evolved throughout the years.

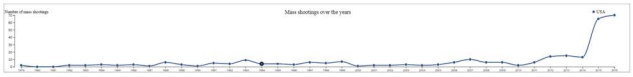

Figure 10. Line chart trend with number of Mass Shootings for the whole U.S selected

Sadly, this suggests a huge spike in mass shootings around the year 2011, and even more around 2014. If we compare it with the data from Figure 7, we see that around 2009, the number of new videogames in the Shooter genre has actually dropped, and the same thing actually occurs in other games that are considered of a violent nature.

Let's see if we can spot a pattern in the upper part of our visualization.

By looking at figure 3, we see that this phenomenon is very prevalent in the California area, and not spread out like video game usage. This suggest that this situation isn't really correlated, and it makes no sense due to the data that we have.

Let's come with a conclusion by looking at the word cloud in Figure 5.

After careful analysis, there's very little things we can relate to video games. Maybe the word "bomb", as these are very normal around videogames, but no direct mention to them occurs. There seems to be no direct connection, so we conclude that video games are not responsible for mass shootings.

IMPLEMENTATION DETAILS

Our implementation speed was severely cut due to our unfamiliarity with JavaScript, but especially due to the learning curve of D3.js. Every interactivity we had was described through an event on JavaScript, to make sure all the idioms popped out for the correct cases, so that was quite simplistic, except to the fact that getting that pattern right takes a lot of time learning it.

Playing with CSV data was also a challenge as we didn't correlate the fact that we need a key for every entry file on the CSV, and how it is very strict on data not being incoherent. Doing the state selector, as well as the list of mass shootings was a bit straight forward as it relied on mostly HTML5, we only had a careful approach when iterating it through the D3.js visualizations to be sure that the events were carefully handled. We relied on a lot of Case-Switch semantics for the states, as well as the type of crime.

Our view on the visualization didn't change drastically on the original plans. We had a very strict view on what we wanted to do, so the problem was really just learning JavaScript and D3.js. It took a lot of iterations, and much more time than we thought previously, but we believe that we were able to do it by relying more on project members with better overall coding skills.

One last problem was marking the coordinates on the map. Even if we appended the points after drawing the map, the map is such o complex item to render that it would take longer to render and appear in front of the points. A delay mechanism had to be implemented to force the points to be drawn only a few seconds after the map finished rendering.

CONCLUSION

After the conclusion of our work, we think we made a good job on joining these two big themes and provide a lot of answers to the questions we were looking for. We believe we used the correct metrics that we were looking for, however, our works reflects much heavier on the criminality and mass shootings theme, than on videogames. With this in mind, we think that things could be improved by relying on more metrics on video games to correlate with criminality data. By doing this we would have a more solid approach on the problem, while answering a couple more questions that were left (e.g., ESRB ratings vs crime, video game industry profit vs crime).

Our main issue was with making sure that D3.js correctly accepted our data and create the necessary events. The Choropleth Map also relied on a lot of research on how to correctly implement it, so we believe we had a huge bottleneck in there, that took many more work hours than previously thought.

In the end, we presented a very cohesive image. We believe that maybe color could be improved, but our initial choice in color palette was correct enough to make this work and stand out. In the future, maybe adopting a dark theme would suffice.

It's safe to say that all idioms suffered small alterations from the initial idea to better accommodate the data and our purposes, but with this process we managed to improve the quality of the visualization, so this ended up being a win-win situation.

As per the results of our findings, they not only correctly assumed that video games are not causing any sort of spike in criminality, but also that there's a slight correlation in

which the usage of video games, specially violent genres, can be attributed to a decrease in violent crime, which is actually a slightly surprising result, as we expected to be no correlation.

Future Work

After much deliberation, our main issue with the visualization is due to not exploring the video games domain in a way that we are happy with. So, with many more resources and more time, we'd like to rebuild the whole video games dataset, as well as finding and joining better datasets to accommodate our approach. We believe that an extra idiom on the video games area would bring a whole level of interaction and answering that we don't have yet.

As an example, profit for videogames genres would be an interesting metric to work it, as well as the density of sales in videogames per area, so that we could bring down the domain by looking at each state's adoption of certain genres of videogames. This wasn't possible due to our resources, but we are confident that with many more iterations, this idea can be better approached and improved to answer any virtual question on this theme.

REFERENCES

1. Shaw, Julia & Crosby, Kimberly & Porter, Stephen. (2015). The Impact of a Video Game on Criminal Thinking. Simulation & Gaming. 45. 786-804. 10.1177/1046878115574018
2. Cunningham, Anthony & Engelstätter, Benjamin & Ward, Michael. (2011). Understanding the Effects of Violent Video Games on Violent Crime. 10.2139/ssrn.1886419.
3. https://www.reddit.com/r/dataisbeautiful
4. https://www.kaggle.com/
5. https://crime-data-explorer.fr.cloud.gov/downloads-and-docs