

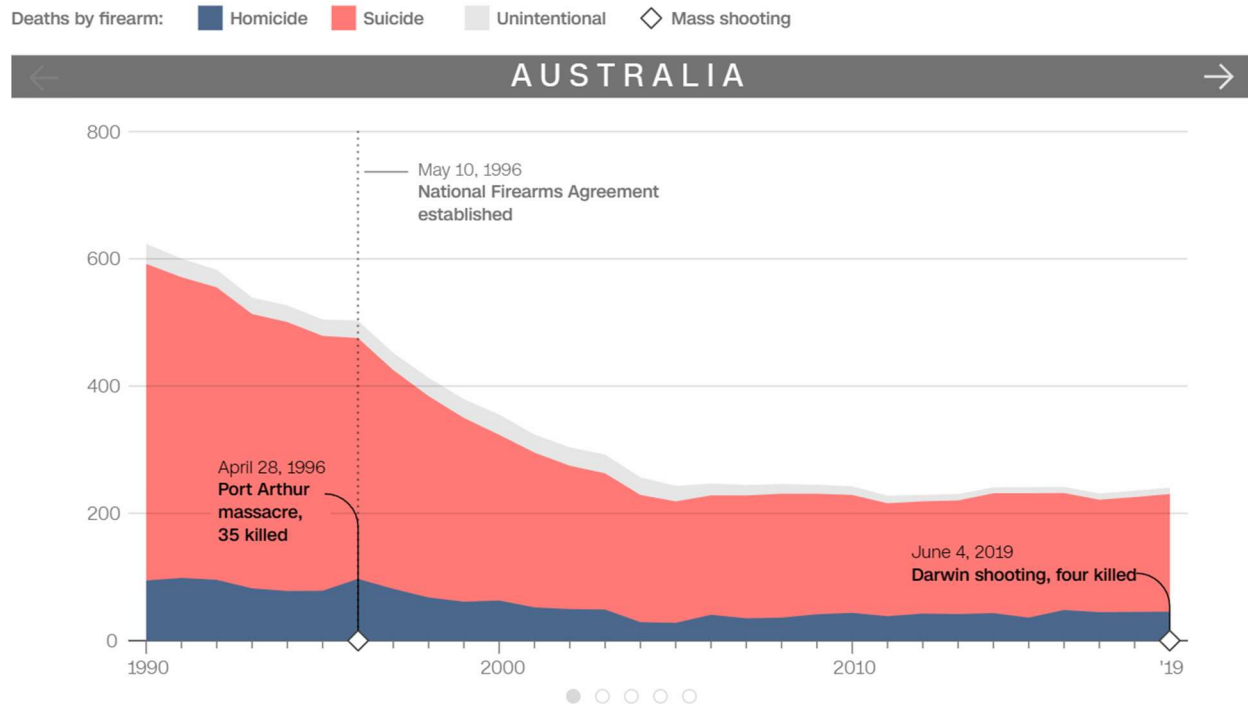
Commentary on the News Article [United States vs the world: Gun culture, laws and mass shootings | CNN](#)

This commentary focuses data science tasks (e.g., data cleaning, exploratory data analysis (EDA) and data visualization) discussed during the Data601 class and applicability to the referenced article. The commentary in this essay are the views of the Data 601 instructor Felix Gonzalez and are used for educational purposes and in the context of various topics within the Data601 class. This article discusses gun culture, laws, and ownership across the world and compares to that of the United States. As an author of an article or analysis, an author needs to decide which data and visualization types to show. As author you will need to select the visualization that you and your reviewers believe has highest impact on your story and message. The reference article documents its assumptions at the end, these include but is not limited to definitions utilized throughout the article as well as data sources.

In the Data601 class we have talked about data cleaning. One of the tasks within data cleaning is to identify and address outlier datapoints. Depending on the type of analysis how outliers are addressed (keep vs. remove) will depend on the use of the data and the message of a story. In this case, the main point of the article is to point that the US is an outlier and this is the most important data point and needs to be kept.

Presumably, the author had to perform some form of exploratory data analysis (EDA) at some point during creation of the article. In this EDA he probably explored many visualizations, of which they probably selected only those they believed made the highest impact to message they wanted to transmit. The majority of the visualizations in the article focus on showing how the US is an outlier.

The article also uses dynamic plots to provide interactivity to the reader. Note that these dynamic plots also have limits and specific decision were made on what interactivity is provided. For example, the plot on Countries licensing rules and effects on gun related deaths has only plots for five countries (e.g., Australia, South Africa, United Kingdom, Finland, and Germany). See Figure 1 for an example of the plots.



Note: Additional gun legislation may have been passed that is not visualized here. Deaths reported in mass shootings exclude the perpetrator.

Figure 1: Australia Licensing Rules and Effects on Gun Related Deaths (Fox, Shveda, Crooker, Chacon, & Pettersson, 2024)

Source: [Institute for Health Metrics and Evaluation \(Global Burden of Disease 2019\)](#), [Parliament of Australia](#), [Australian Attorney-General's Department](#), Jason R. Silva

The article's discussion on countries legislation attempts to show how some countries that introduced gun control legislation resulted in lowering of the trend in gun deaths. However, for all countries except South Africa, the trend of gun deaths was already going down prior to the introduction of the legislation. From a data analysis standpoint, it is difficult to conclude what was the effect of the legislation given the prior negative trend. As a data scientist I would have also liked to see the trend of the US related deaths in this same plot too and not only in the five countries. I would have also liked to see the longer date range trends if the data was available.

There are a few things to consider when creating conclusions out of data. These includes but is not limited to the following:

- Correlation and causation: correlation does not necessarily imply causation.
- Multicollinearity: predictors in a model are linearly dependent or strongly correlated.
- Ethical considerations when showing data

In the Data601 we discussed various data visualization libraries such as Matplotlib and Seaborn. In the referenced article it looks like the dynamic figures were deployed in the website using SVG. To explore the code, you can use the right click mouse click, select "Inspect" and explore the Frame source and elements. Python seems to have an SVG

library (i.e., svgutils), however, svgutils library is NOT included in the Anaconda Data Science Distribution and needs to be installed separately. The SVG Library website can be found in the following URL: <https://pypi.python.org/pypi/svgutils>.

It is important to recognize that every data will have its limitations

References

Fox, K., Shveda, K., Crooker, N., Chacon, M., & Pettersson, H. (2024, 2 15). *CNN*. Retrieved from CNN: <https://www.cnn.com/2021/11/26/world/us-gun-culture-world-comparison-intl-cmd/index.html>

University of Maryland Baltimore County (UMBC). (2023, 4 13). *Academic Integrity Plagiarism Education*. Retrieved from UMBC: <https://academicconduct.umbc.edu/plagiarism/>