

Visualising human rights data

Izzy Stewart, Data Visualisation and the web

Introduction

In this project I investigate the way in which LGBT global rights are understood visually. Using the Data collected by the International Lesbian and Gay Association (ILGA), a global human rights lobbying organisation, I seek to explore the following research questions:

1. What legal protections and discriminations do LGBT people face globally?
2. In what ways can different analysis approaches affect our understanding of human rights situations?

This is achieved using a number of data visualisations that approach and interrogate the data from numerous angles. It is through these different approaches that I reveal how different data presentations can produce differing empathetic reactions and understandings.

I wanted to look at both the legal protections, such as workplace or marriage equality, and discriminations LGBT people face. This discrimination takes the form of, what are often colonial era laws, such as the Buggary Act or Indecently Laws. Whilst these laws affect the entire population of a country, they are the ones identified by ILGA and others as used to police and criminalise Lesbian, Gay, Bi and Trans lifestyles and expressions.

It is worth noting that due to various political reasons ILGA does not have data for some countries such as North Korea and multiple African nations. These gaps should be considered when reading the visualisations. Finally, I think it is important to remember that this data is taken from their 2017 report (the latest published) and so some of the laws in each Country will have changed, for example the recent lifting of the Buggary Act in India.

Data overview

For both research question I identified a set of concepts that were important to the research. From these concepts I could then define the variables needed for analysis. The table below shows my results.

Question	Concept	Variable	Type
One	Area	Country	Nominal
One	Area	Continent	Nominal
One	LGBT protection laws	Protection: Constitution	Nominal
One	LGBT protection laws	Protection: Employment	Nominal
One	LGBT protection laws	Protection: Other	Nominal
One	LGBT protection laws	Protection: Hate Crime	Nominal
One	LGBT protection laws	Protection: Incitement	Nominal
One	LGBT protection laws	Protection: Conversion therapy ban	Nominal
Both	LGBT protection laws	Legal: same-sex relations	Nominal
Both	LGBT discrimination laws	Maximum prison sentence	Ordinal
Both	LGBT discrimination laws	Illegal: male	Nominal
Both	LGBT discrimination laws	Illegal: female	Nominal
Both	LGBT discrimination laws	Penalising text: against nature	Nominal
Both	LGBT discrimination laws	Penalising text: against buggery	Nominal
Both	LGBT discrimination laws	Penalising text: against indecency	Nominal
Both	LGBT discrimination laws	Penalising text: sexual act	Nominal
Both	LGBT discrimination laws	Penalising text: sodomy	Nominal
Two	Population	2017 population per Country	Ratio

The data I have used has come from two sources. The first is taken from a dataset found on 'data.world' created by 'dataforacause' as part of a data analysis competition. This data is taken from the '2017 State Sponsored Homophobia' report published by the International Lesbian and Gay Association (ILGA). I found this dataset when researching this area and looking for an appropriate dataset. Although ILGA actually provide the same data on their website, I found this dataset more useful as much of the data had already been cleaned and pre-processed. However, as you can see in my code file I do use the ILGA data to check for potential errors and when comparing with their 2016 dataset (which 'dataforacause' do not provide). The second dataset is taken from the 'data.worldbank.org' website and provides data for population data for 2017 per Country and Continent. Both datasets are csv file format.

Although the 'dataforacause' dataset had already been partly pre-processed and cleaned from the original ILGA dataset, some further pre-processing and cleaning techniques were necessary for the analysis I had planned. For some plot types I needed to change the 'maximum sentence variables' from a group of nominal variables to one ordinal variable. I also wanted to compare the 2016 dataset with the 2017 one. This involved recoding variables into binary 0-1 format (updating all text values to 1 and all NaN values to 0).

The second part of my analysis required me to combine the population and ILGA datasets. I did this using pandas join capabilities. This however proved difficult where some countries weren't listed in ILGAs report, as well as some Countries having slightly different names. If I was to do this process again I would want to try a different technique, allowing a cleaner result.

A consideration throughout this project was the reliability of the data and the potential for it to be misinterpreted. Particularly where I would combine the two datasets, this meant having to drop

some countries where I did not have the population data. This also meant talking about 'global population' where in reality it is only showing the Countries where ILGA has collected the data.

Exploratory analysis and visualisations

What legal protections and discrimination effected LGBT people in 2017?

The first section gives an overview of the analysis I have taken to help explore the first research question. In figure 1 below you can see a pie chart depicting the percentage of Countries used in the ILGA report where same-sex relations are 'Legal' or 'Illegal.' The variables used for this analysis were: 'Country' and 'Legal: All Genders', taken from the ILGA 2017 dataset. The 'Legal: All Genders' variable was a nominal variable with the values 0 or 1. This variable is set to '1' where same-sex relations are legal for both genders in that Country. The chart shows that in over 60% of countries, same-sex relations are legal. For this chart I have considered opponent colour theory, where I have used blue and yellow as contrasting colours. For consistency, I will use these two colours throughout the report where I describe/compare two variables or values in a variable.

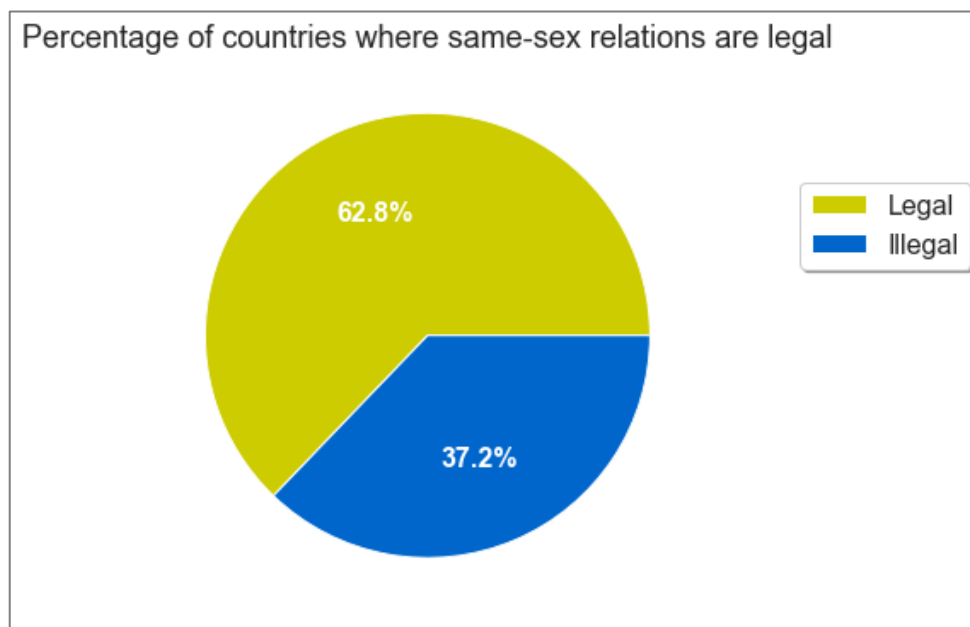


Figure 1 Pie chart showing percentage of countries where same-sex relations are legal and illegal

After looking at the total number of Countries where same-sex relations are legal/illegal, I wanted to compare the results per Continent (figure 2). To do this I used the 'crosstab' feature in pandas to create a normalised table of values giving the percentage of countries per Continent that where same-sex relations are illegal. In this analysis I used the variables 'illegal male' and 'illegal female' which are both nominal variables that tell you if same-sex relations are illegal if you are a male or female respectively. Looking at the results I found it interesting to see that, apart from Europe, all Continents had a higher percentage of Countries where it is illegal for male same-sex relations than female. Particularly in Oceania, where the number of Countries where same-sex relations are illegal for females is lower than in Americas, where for males its much higher. I chose a bar chart where each Continent has two side-by-side bars of opposing colours, to highlight the difference between these laws for these two genders.

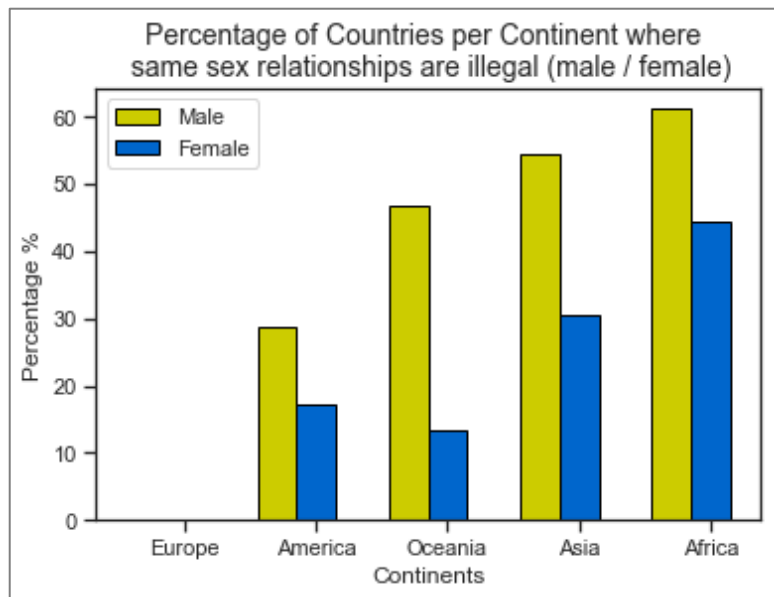


Figure 2 Bar chart showing the percentage of Countries per Continent where same sex relations are illegal comparing male and female

	continent	illegal_male	illegal_female
3	EU	0.000000	0.000000
1	AM	28.571429	17.142857
4	OC	46.666667	13.333333
2	AS	54.347826	30.434783
0	AF	61.111111	44.444444

Figure 3 Table used to create bar chart in Figure 4

Figure 4 uses a bar chart to show the lengths of prison sentences for same-sex relations in Countries where these sentences apply. This uses the 'Maximum sentence' variables, in this case I treat as separate nominal variables but in later charts I use as one ordinal scale (going from the 1 month-2 years and ending with the death sentence). For this chart I used a hue of blue where the colour becomes stronger as the sentence increases. The darkest blue is on the death sentence, highlighting this as the most shocking value. This uses the perception theory of increasing contrast to emphasis importance. It is interesting to see the bar chart below follows a normal distribution.

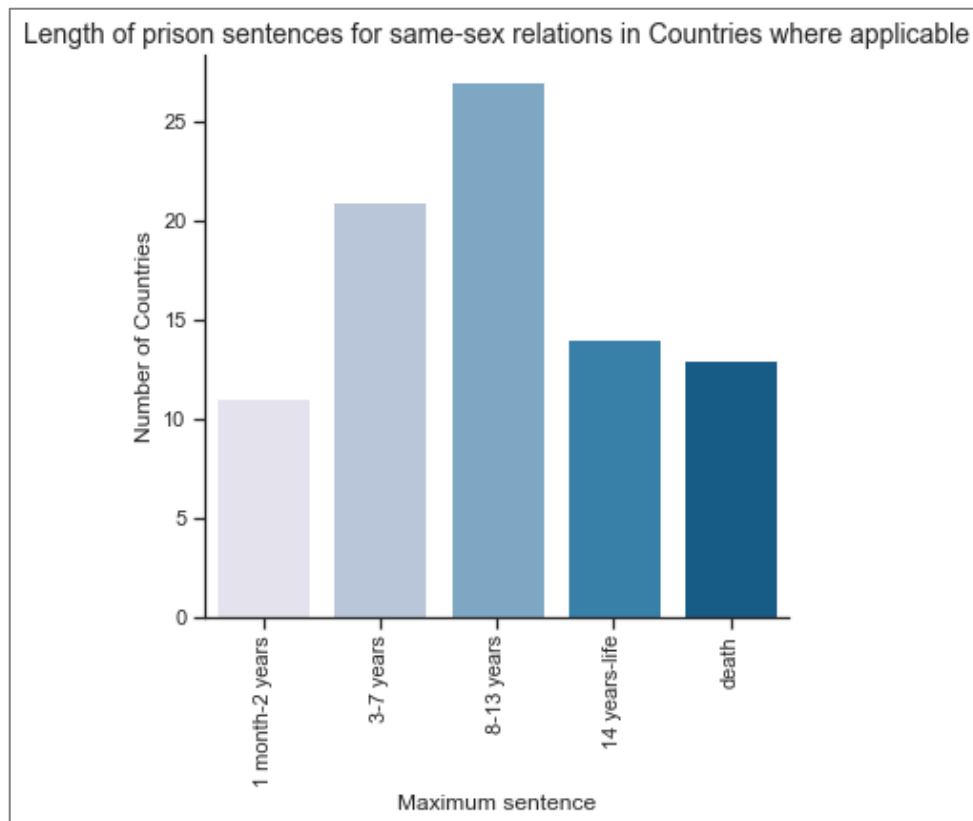


Figure 4 Bar chart representing the length of prison sentence for same-sex relations in Countries where it applies

	continent	country	legal	2 years	3 to 7 years	8 to 13 years	14 to life	death
0	AF	Algeria	0	0	1	0	0	0
1	AF	Angola	0	0	1	0	0	0
2	AF	Benin	1	0	0	0	0	0
3	AF	Botswana	0	0	1	0	0	0
4	AF	Burkina Faso	1	0	0	0	0	0

Figure 5 First five rows from a table used to create Figure 2 bar chart

Figure 6 uses a heatmap to look at the maximum prison sentences given for same-sex sexual activity, but this time comparing the percentage of Countries per Continent. I chose this chart as it both gives you an idea of what sentences are used largely in each Continent as well as showing you how consistent a Continent is with its laws across its Countries. For example, you can see that in Oceania, 40% of its Counties use the 8-13-year maximum sentence. Whereas in Asia the maximum sentence used is much more spread evenly across Countries.

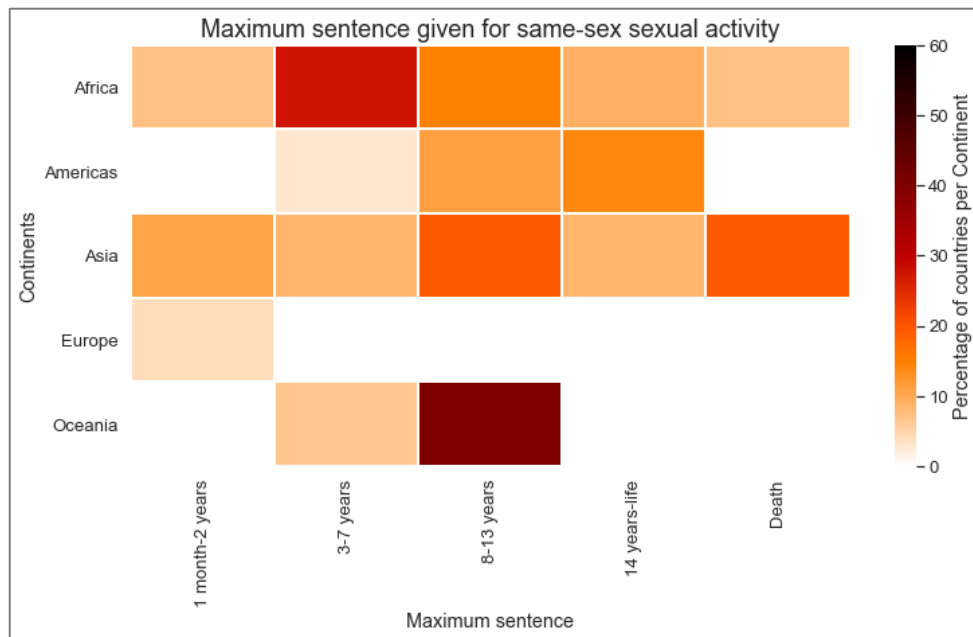


Figure 6 Heatmap displaying the percentage of Countries per Continent using the following maximum prison sentences

For the next part of my analysis I wanted to compare the different laws in place in each Country that both protect and discriminate against a person's sexual orientation. Figure 7 depicts a scatter graph where each point represents a different law and the percentage of countries per continent that have that law in place. The blue and orange colours divide the laws into where ILGA have defined them as 'Protection' and 'Penalising text.' I have outlined the description below of each law (description taken from the ILGA report). In this graph I intentionally anonymised each law giving you only if it is a 'Protection' or 'Penalising' law. This allows you to compare each Continent by the percentage of Countries with each law type they have in place. As you can see Europe has only protection laws in place, where some are used across a high percentage of Countries. Whereas the other Continents all have a mix of both where some Continents have a higher percentage of countries with penalising laws than protection.

Protection laws

- Constitution: Constitutional prohibition of discrimination based on sexual orientation
- Conversion Therapy Ban: Ban on conversion therapies
- Employment: Prohibition of discrimination in employment based on sexual orientation
- Hate Crime: Hate crimes based on sexual orientation considered an aggravating circumstance
- Incitement: Incitement to hatred based on sexual orientation prohibited
- Other: Various other non-discrimination provisions based on sexual orientation

Penalising text

- Against nature: Actual offence that generates maximum sentences: against nature
- Buggery: Actual offence that generates maximum sentences: buggery
- Indecency other: Actual offence that generates maximum sentences: indecency or other
- Sexual act: Actual offence that generates maximum sentences: sexual act
- Sodomy: Actual offence that generates maximum sentences: sodomy

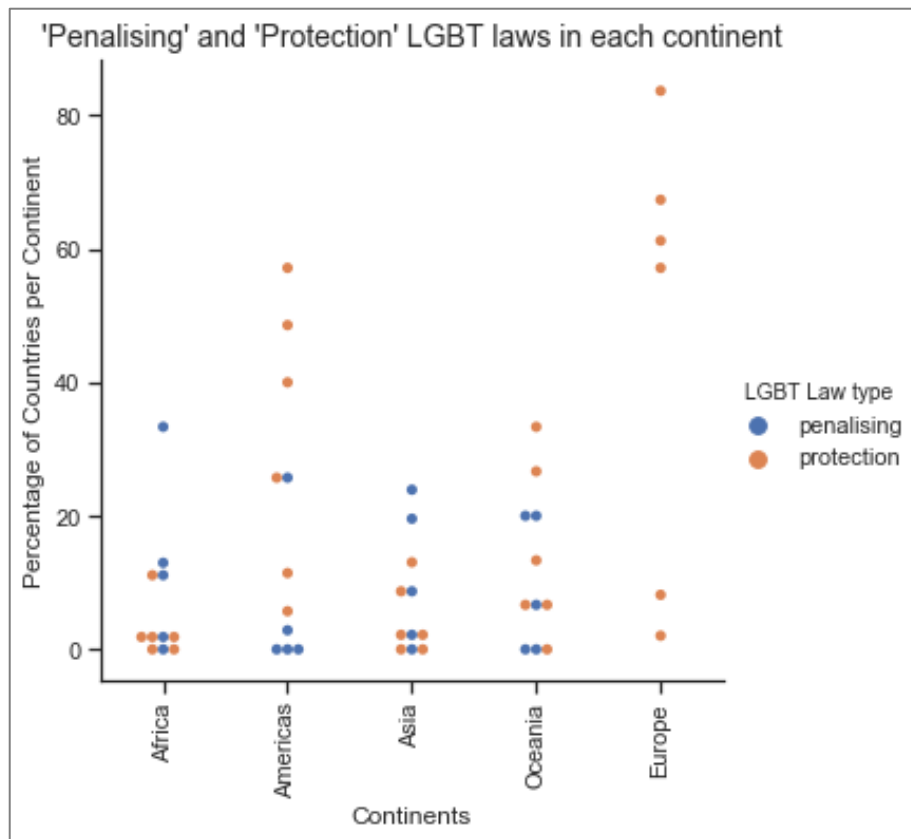


Figure 7 Scatter graph depicting the percentage of Countries per Continent with Penalising a Protection laws

I wanted to look further at the laws defined as discriminating against sexual orientation and find a way of combining these into one variable to measure the level of discrimination per Country and Continent. To do this I created a new 'score' variable that gave each Country a total score based on the discriminating laws they had in place. In the table below, I have outlined the variables used to create the score and the number attributed to each. As you can see all variables are given a score of one, except the 'maximum' sentence variable where I increase the number as the sentence increases (up to death with a score of 5). This allowed me to use these nominal and ordinal variables as a ratio variable. I intentionally left out the same-sex relations legal/illegal variable from this score as I wanted to use this to compare with the discrimination score variable in my analysis.

Variable	Score
Age of consent: Unequal	1
Penalising text: Sexual act	1
Penalising text: Sodomy	1
Penalising text: Against nature	1
Penalising text: Buggery	1
Penalising text: Indecency/other	1
Promotion or Morality: Penal code	1
Promotion or Morality: Morality code	1
Max Sentences: 1 Month to 2 Years	1
Max Sentences: 3 Years to 7 Years	2
Max Sentences: 8 Years to 13 Years	3
Max Sentences: 14 Years to Life	4
Max Sentences: Death	5
Arrests in past 3 years - Yes	1

Once I had created the new 'score' variable, I then produced two interactive boxplots, looking at the discrimination scores per continents (Figure 8) and comparing same-sex legislation (Figure 9). Figure 8 gives you a good indication of the distribution of the discrimination scores across Countries in each Continent. For example, you can see that approximately 50% of the Countries in Asia have a discrimination score between 0-6, whereas in Africa 50% score between 1-5. I also found this plot very useful in highlighting the outliers. Looking at the outlier with a score of 3 in the Europe plot, I found out this Country was Russia. This was helpful to explain why it was an outlier as, although ILGA had defined Russia as in Europe, it is often associated as in Asia (where its score would sit near the median value. To highlight this I created custom annotations that point to Russia as well as the highest outlier Nigeria (with a score of 12). I have used primary colours, along with green and orange. These colours are used throughout the report for consistency in clearly representing Continents.

Figure 9 gave me a good insight into the distribution of discrimination scores in Countries where same-sex sexual activity is by law legal and illegal. Unsurprisingly, where a Country is set to 'legal', the score is mainly set to zero. However, as with Figure 8 this box plot also proved useful in highlighting outliers that may have skewed the data in other visualisations. For example, where I have annotated the outlier 'Egypt.' Noting that Egypt had a discrimination score of 5, led me to look for more information on this Country within the ILGA report. Interestingly the report stated 'Sexual relations between consenting adult persons of the same sex in private are not prohibited in Egyptian law. However, as recorded the Law on the Combating of Prostitution, and the law against debauchery have been used liberally to imprison gay men in recent years.'

Both box plots have an interactive element, where hovering on each plot reveals the statistics for that group. For example, in Figure 9, looking Countries where same-sex relations are illegal, I can see they have a median score of 5, an upper quartile of 6 and a lower quartile of 4. It also tells you the minimum score is 1 and the maximum is 12.

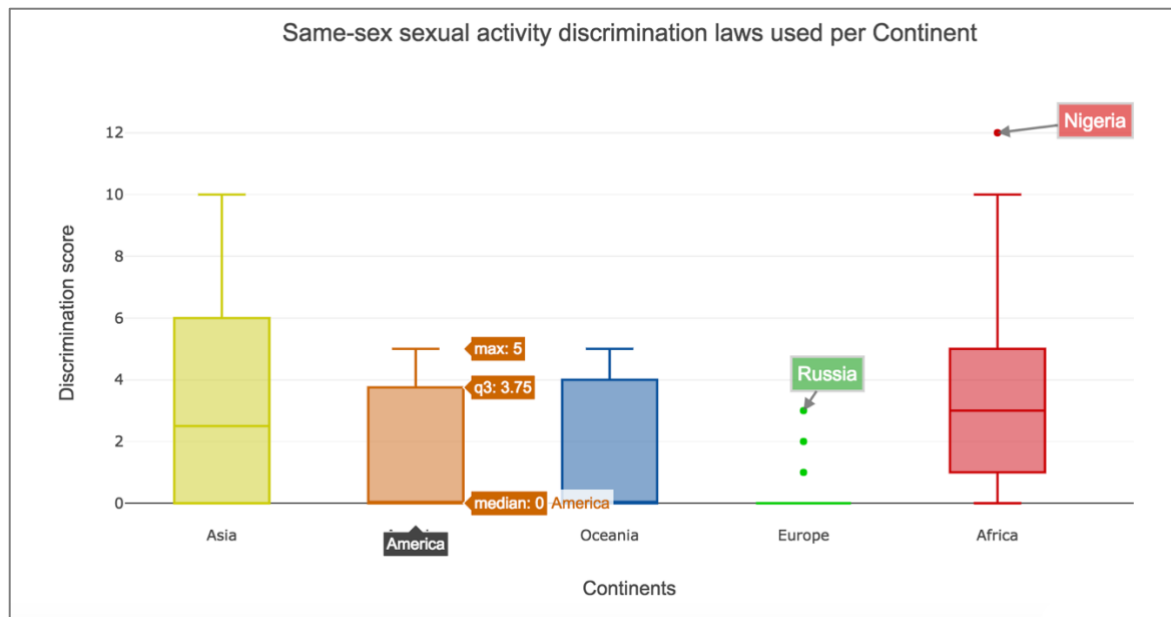


Figure 8 Interactive box plot displaying the same-sex sex relations discrimination laws used per continent

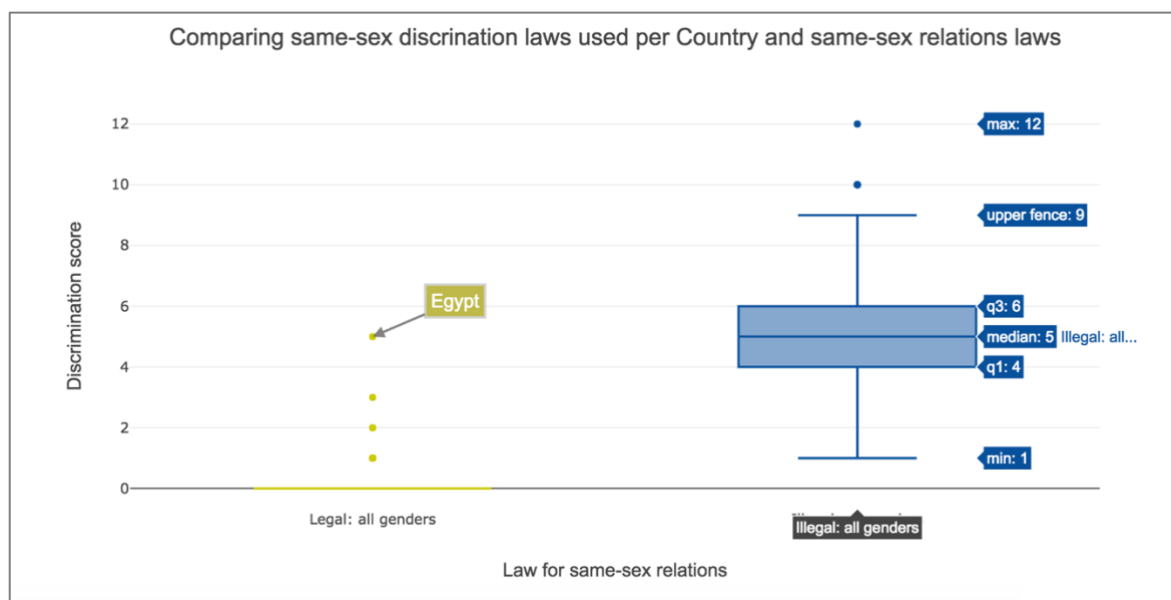


Figure 9 Interactive box plot comparing same-sex discrimination laws with same-sex legalisation

In what ways different analysis approaches affect our understanding of human right situations?

This section combines 2017 population data from 'worldbank.org' with the ILGA report data, exploring my second research question. Figures 10, 11 and 12 use horizontal bar charts to show the total number of individuals globally that; are protected by legality of same-sex relations, live in Countries with prison sentences for same-sex relations and live in Countries with 'penalising text' discrimination laws. I feel that looking at the numbers in terms of people rather than Continents gives you a deeper connection to the data, where you can see the population of people affected by these laws rather than separated by Countries. In particular, looking at Figure 11, the death sentence stands out as a very large number compared to Figure 4 where you would not know how many people are affected.

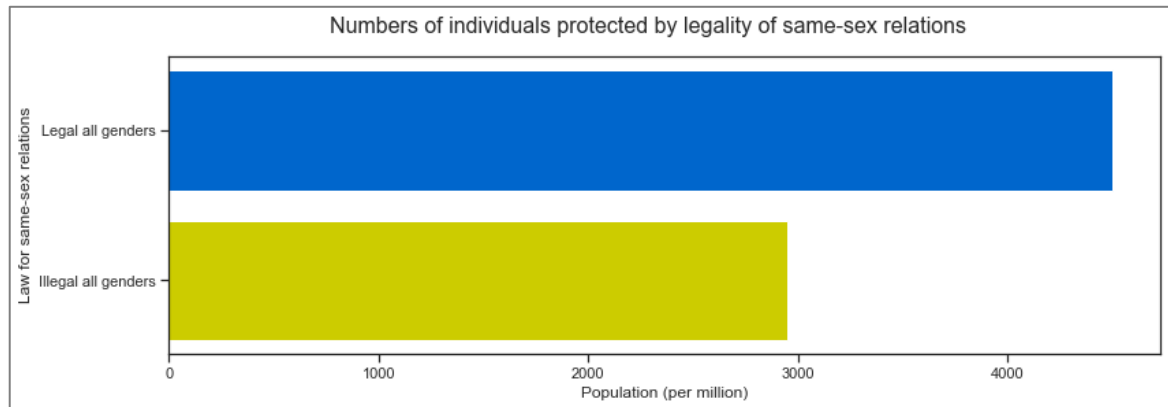


Figure 10 Horizontal bar chart depicting the number of individuals globally protected by legality of same-sex relations

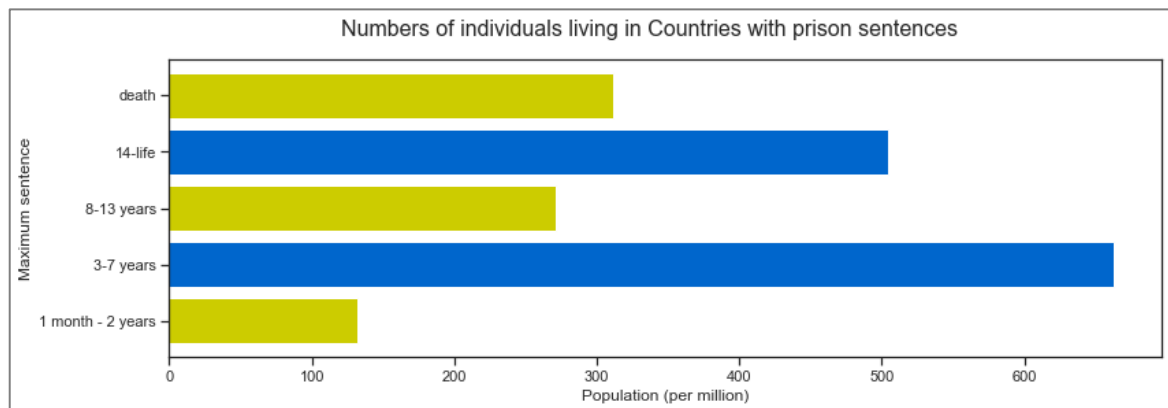


Figure 11 Horizontal bar chart depicting the number of individuals globally living in Countries with maximum prison sentences for same-sex relations

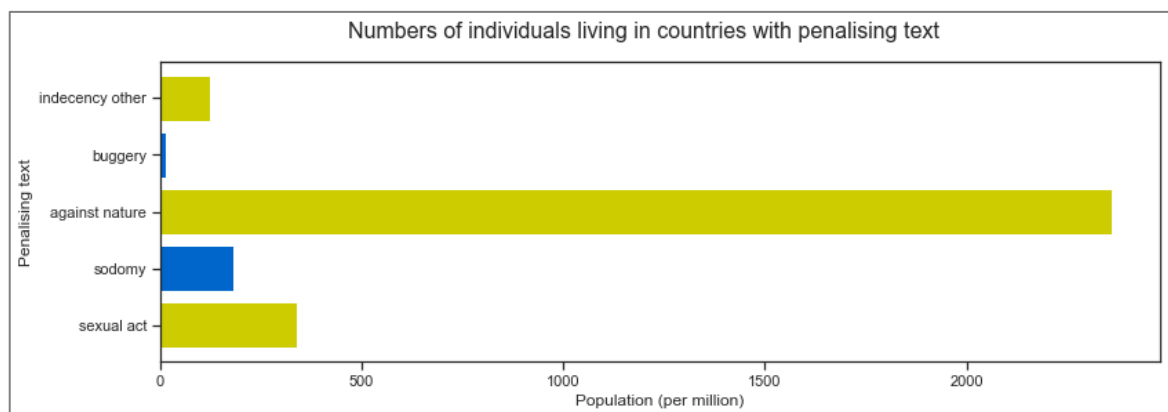


Figure 12 Horizontal bar chart depicting the number of individuals globally living in Countries with penalising text

The final visualisation (Figure 13) uses an interactive filled area plot to show the population distribution of prison sentences for same-sex relations per Continent. The interactive element lets you hover on each sentence type to see what percentage comes from each Continent.

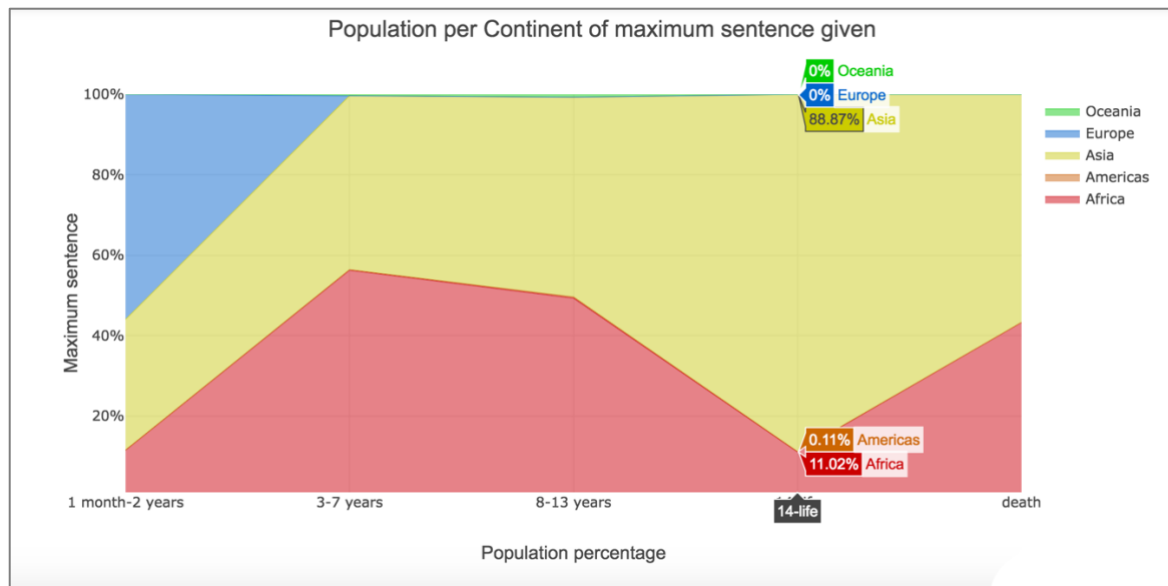


Figure 13 Interactive filled area plot showing the population per Continent of different maximum prison sentences given for same-sex sexual relations

Conclusion

In looking at the legal protections and discriminations that LGBT people face, I found it particularly shocking to see the number of Countries that, in 2017, had high prison sentences or worse a death penalty for same-sex sexual activity. Another insight that I was not aware of before starting this analysis was that many Countries have different laws for same-sex sexual activity depending on if you are a male or female. In looking at how different analysis approaches affect our understanding of human right situations, using the population data as a way to group the LGBT laws rather than looking by Country gave me a better understanding of the number of people effected globally. This could be used as a useful empathy building tool for activists and campaigns.

Overall, I am happy with the visualisations I have produced; however, I feel the data I used was lacking more ratio variables which would have allowed me to produce a wider range of plots. I aimed to rectify this by creating the 'discrimination score' variable from a set of nominal variables. If I could take this project further, I would extend this process by using these nominal variables to create clusters to better understand the similarities and differences in LGBT rights per Country.

Further improvements I would like to have introduced would be a timeseries element looking at how LGBT laws have changed over a number of years. Unfortunately, ILGA only have data from 2016 and 2017. In my analysis I compare some of the law changes between these years but did not find enough had changed to produce effective visualisations.

This project helped me develop my knowledge of python visualisation packages and modules such as matplotlib, seaborn, bokeh and plotly. In particular I developed my knowledge of making plots interactive using plotly, which I will find useful in future projects.