Rebecca Reilly

INTRODUCTION:

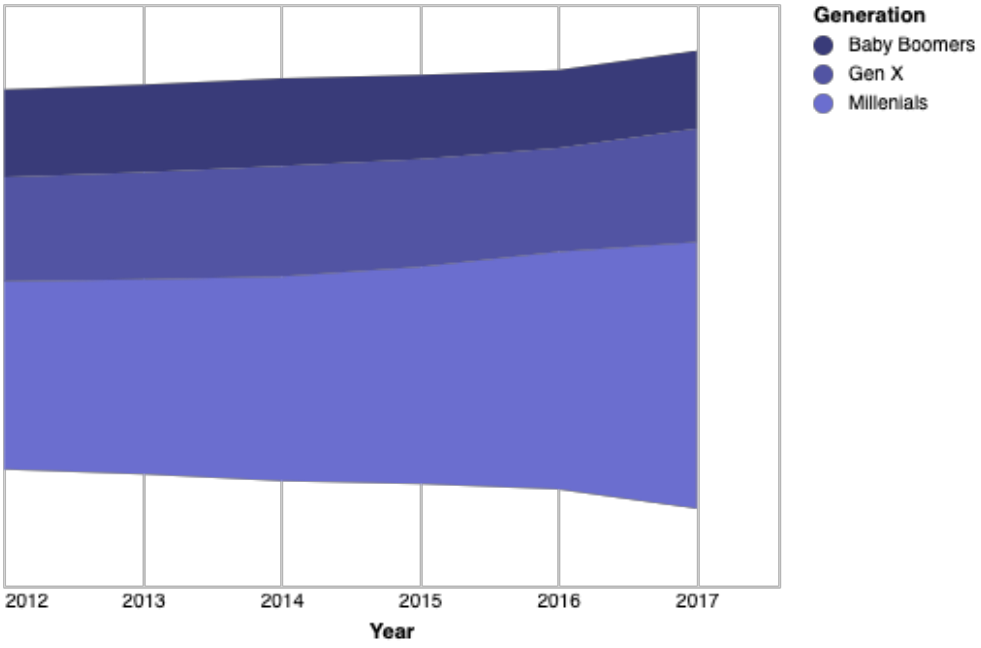
For this project I used various datasets related to the LGBTQ+ community– LGBT population, LGBT related laws by US State, State Equality Index information, Google Trends, Workplace Discrimination, and State Equality data. The sources for these are the Human Rights Campaign, the Gallup Poll, The Center for Employment Equity, lgbtmap.org, equaldex.com, and google trends. I chose to explore this topic, and thus the various related datasets, because as a member of this community, safety is a subject I always need to consider when traveling, and I was curious about being able to rank states in this manner. The questions that I posed were around what the population looks like, and safety in terms of what protections states have in place, what LGBT people are reporting for discrimination (available in workplace data), as well as how many negative bills have been introduced in states, assuming that would mean there are anti-LGBT politicians trying to combat the protections in place, if any. Aside from looking at legislative bills, and state ratings, I was also interested in finding out if there was information I could link to the less ideal states for LGBT people to indicate an existing hypocrisy.

SUMMARY OF DATA:

* caveat: any self-report data is bound to not tell the whole story, so I acknowledge that numbers may be under representative especially in more conservative states where people won't feel as comfortable disclosing or reporting

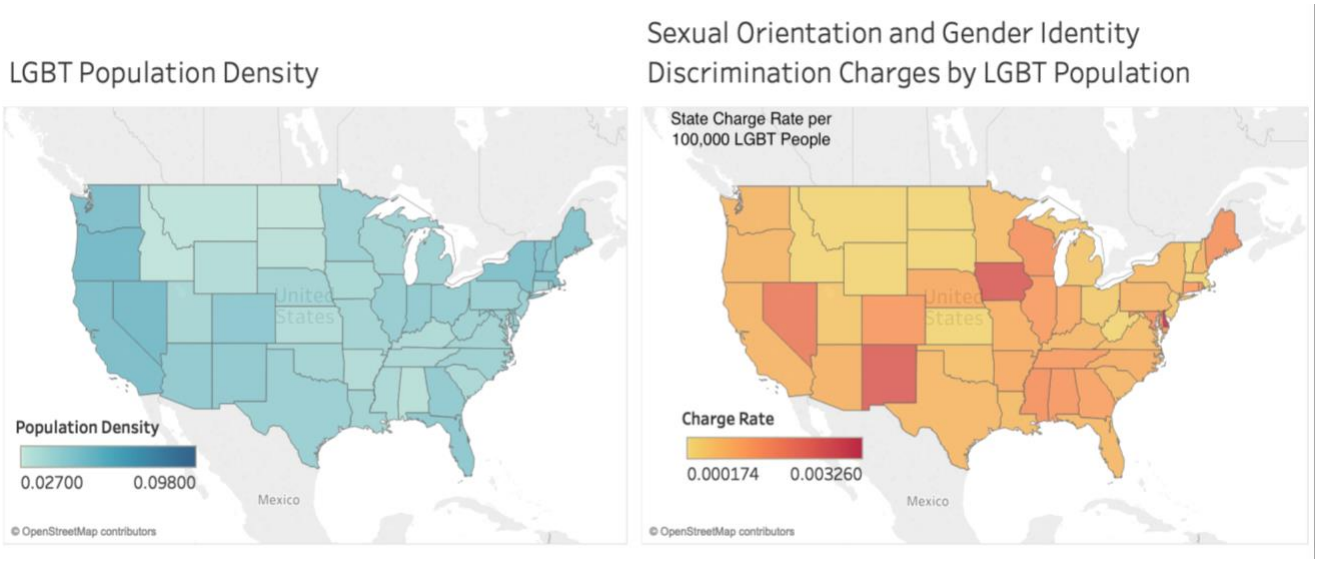
i. How has the community population been changing by generation?

- Data only collected since 2012; we can see general increases but significantly heavier identification by younger people



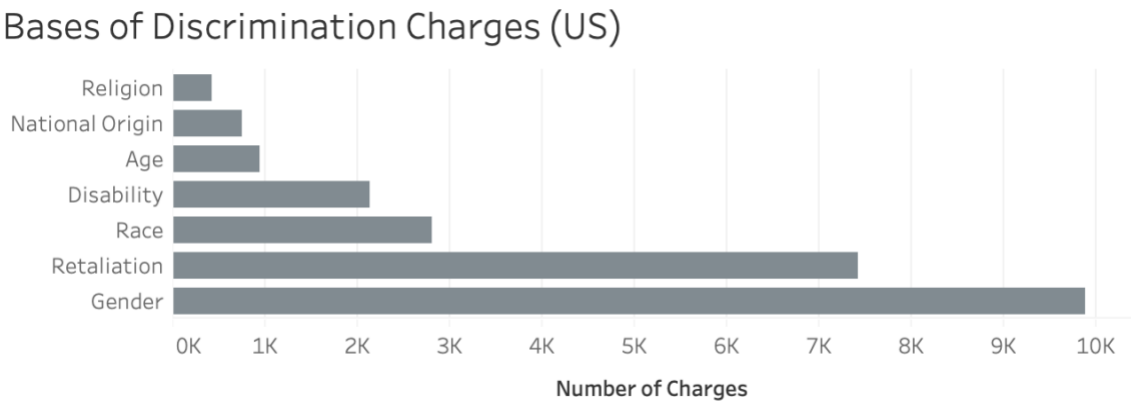
ii. How many people identify as a part of the LGBT community?

- The first step in looking into this community is understanding how much of the population they make up. Additionally important is understanding where, in terms of reported charges by LGBT people, there is the most workplace discrimination.



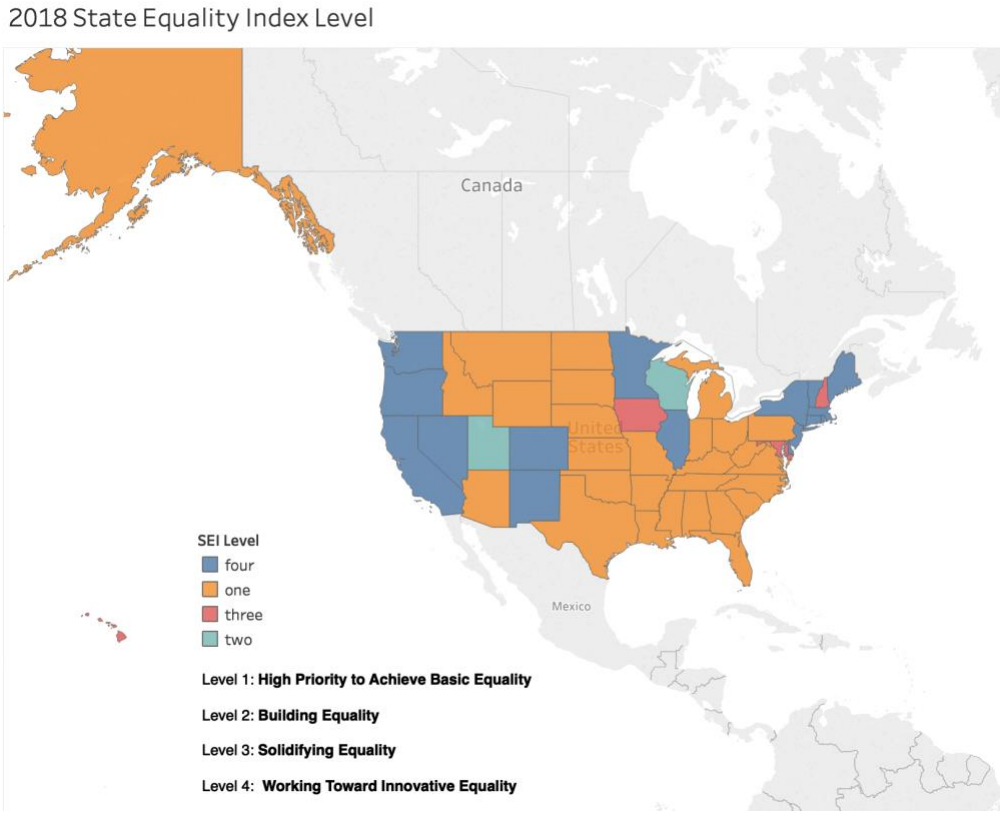
iii. In term of discrimination, on what bases are LGBT people reporting?

- Overall, it is clear that LGBT people are indication 'Gender' as the main basis for their reported discrimination. This indicates that across the US, the community feels that their Gender Identity is being attacked.

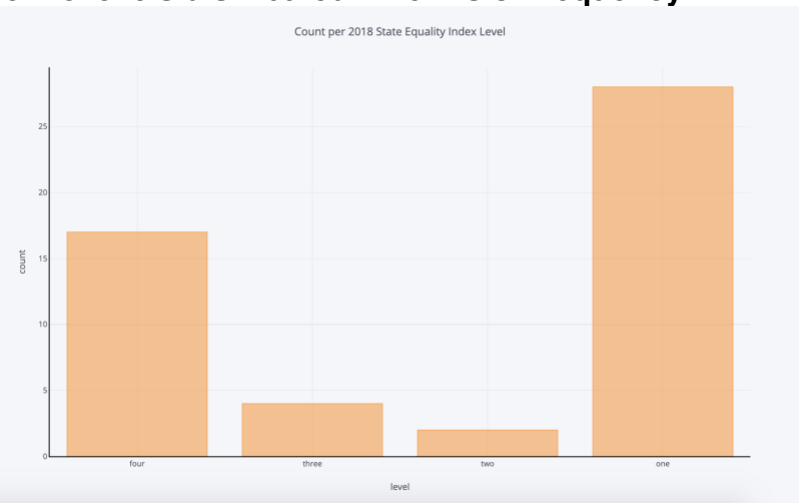


iv. If this discrimination exists at higher rates in certain states, are those states that the HRC considers at the lowest level of their equality ranking?

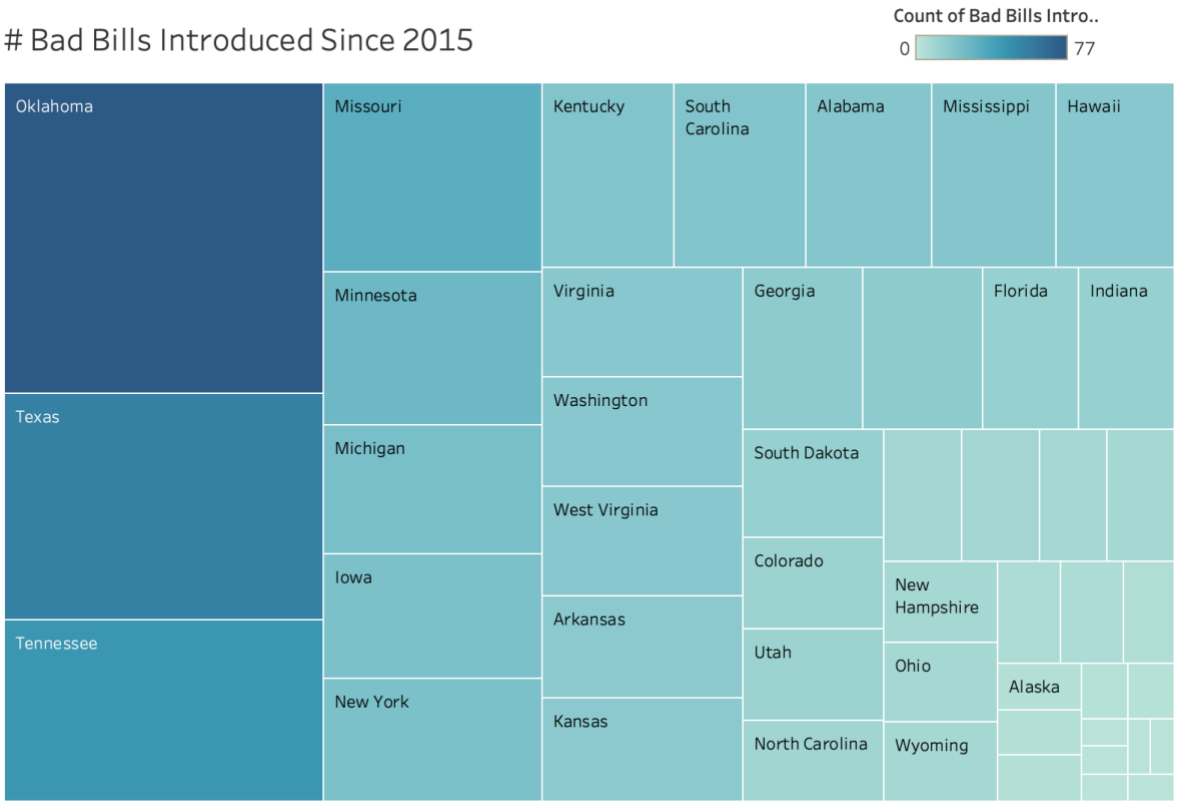
- It is not surprising that middle and southern states receive the lowest level. It is surprising that states with higher levels like Nevada or New Mexico have higher discrimination charge rates.



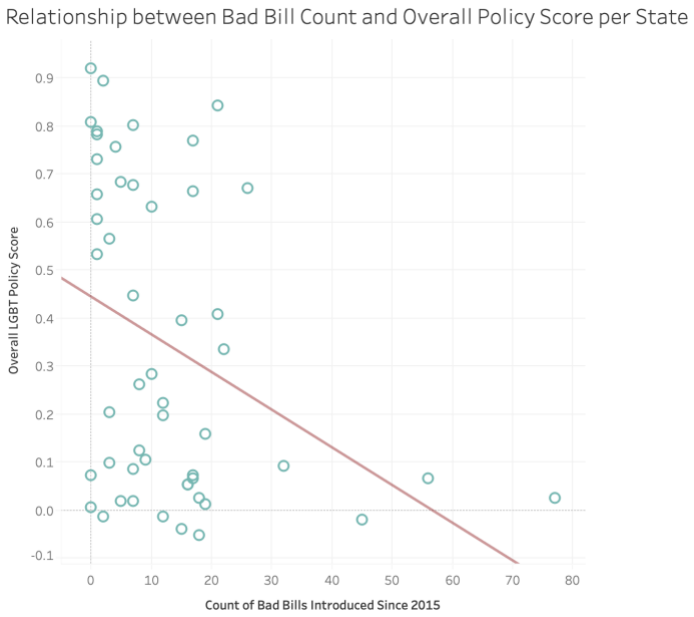
v. How are the levels distributed in terms of frequency?



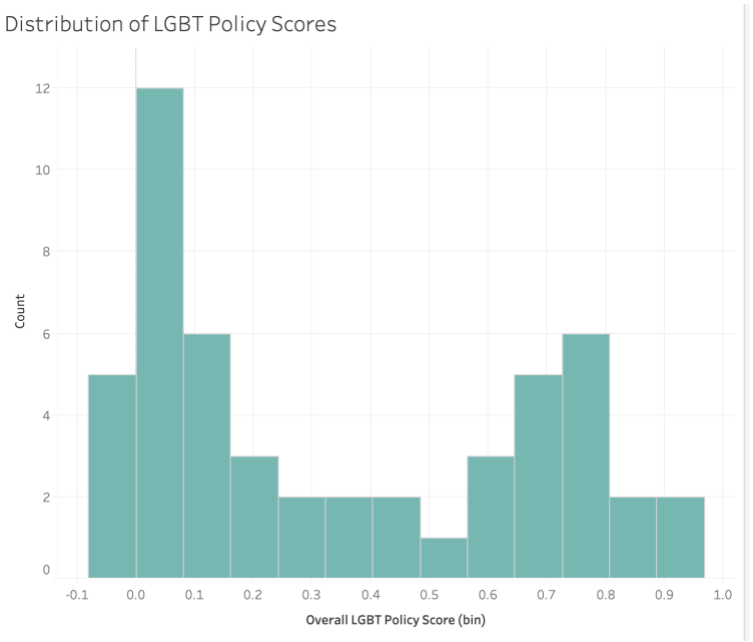
vi. Given we now know how HRC ranks states, let's look at which states introduced, since 2015 (legalization of gay marriage) "bad"/anti-lgbt bills



vii. The Movement Advancement Project gives every state a rating based on their overall equality policies, what is the relationship to number of bad bills introduced?

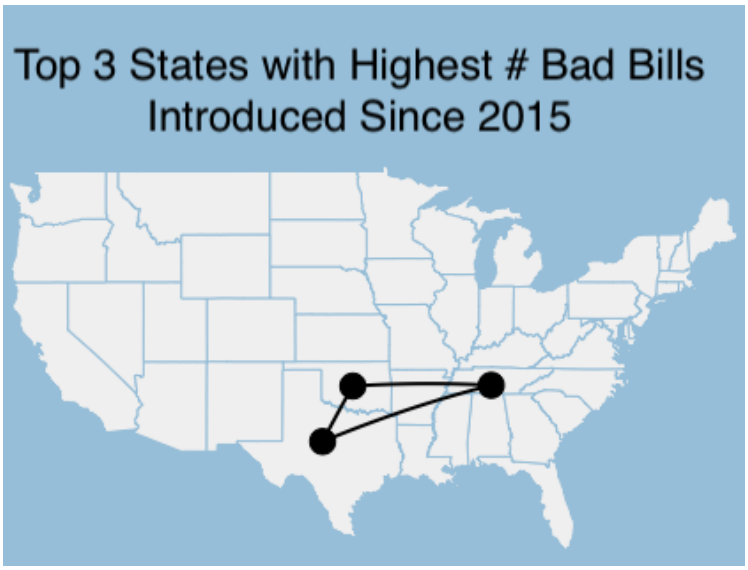


viii. What is the distribution for the overall policy score?



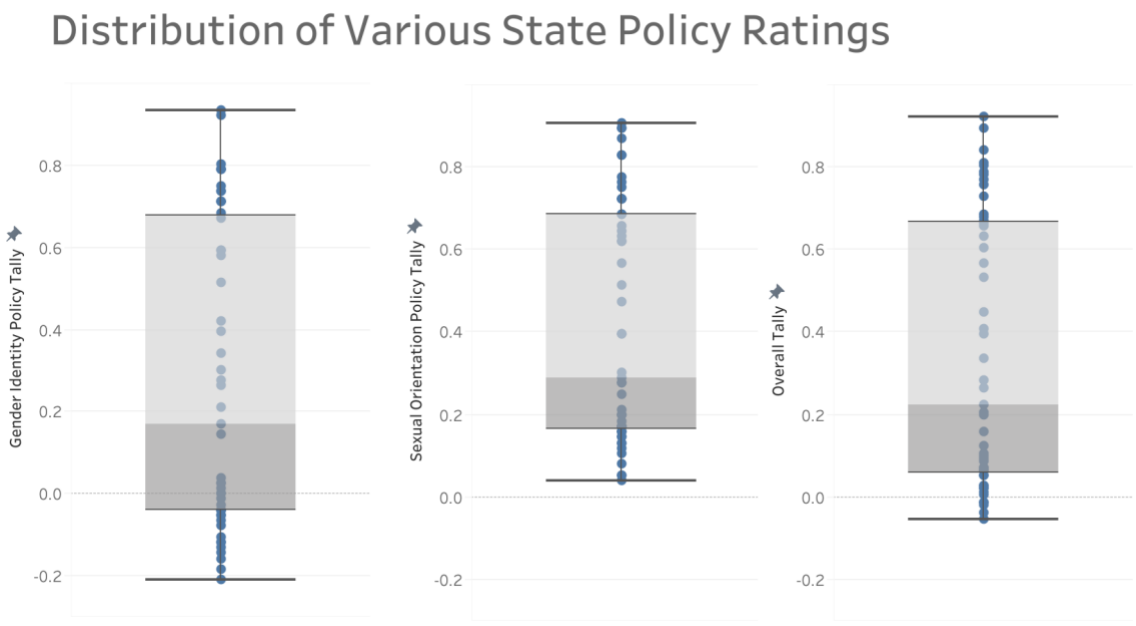
ix. What are the top 3 states with the most bad bills introduced?

- Texas, Tennessee, Oklahoma

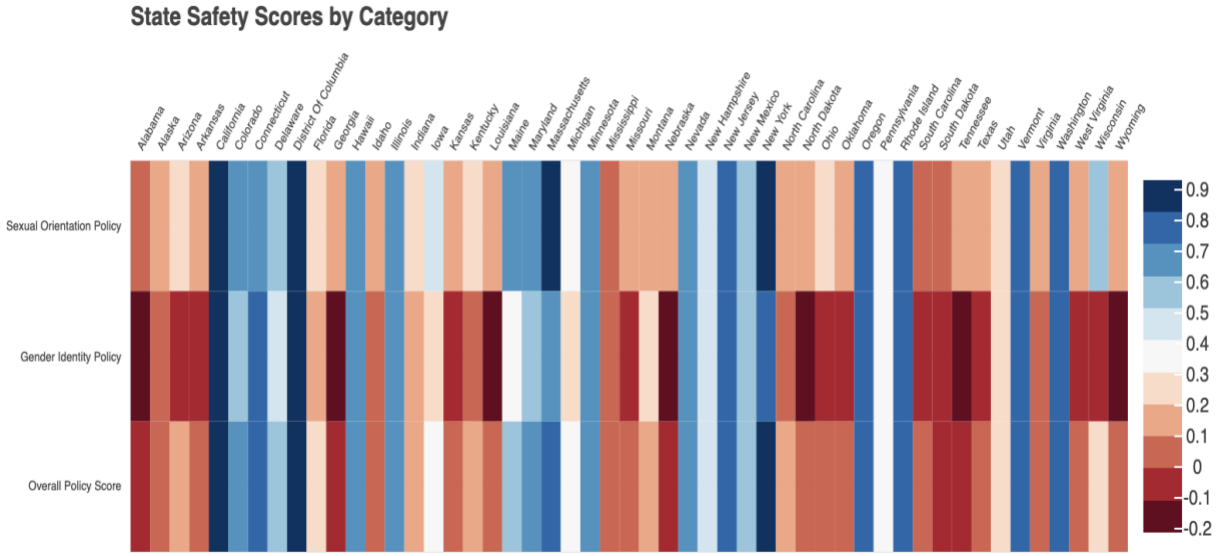


x. **The Movement Advancement Project gives every state a rating on overall policies, sexual orientation policies, and gender identity policies, how do the scores compare?**

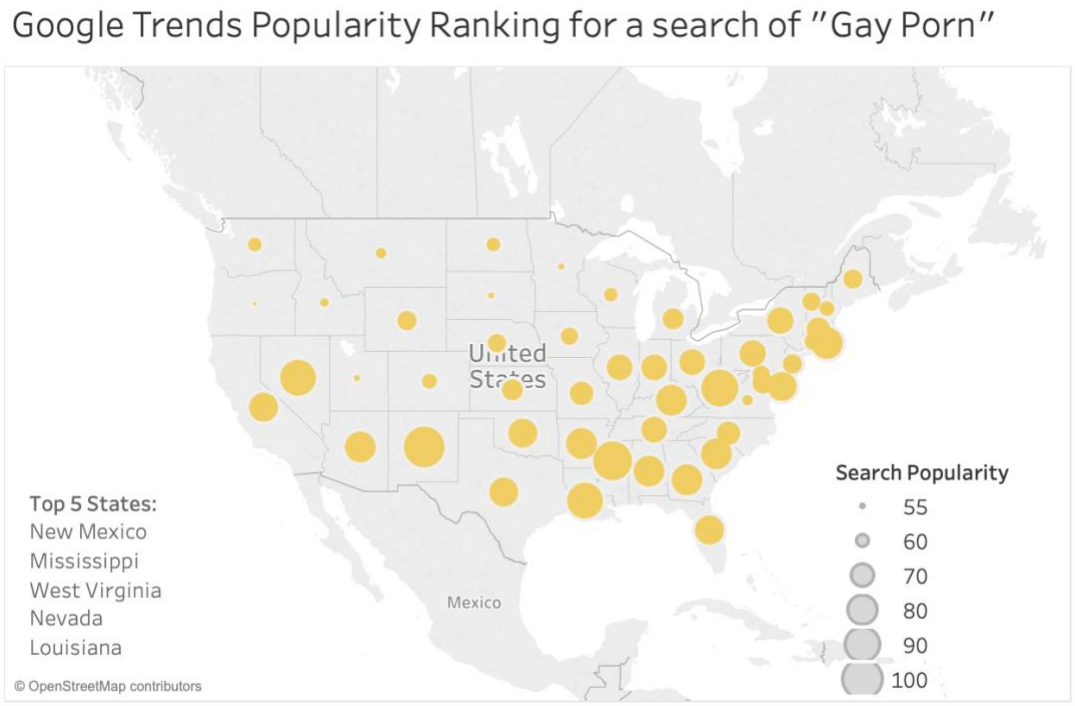
- Gender Identity policies are the most widely lacking



xi. Visualizing States by their normalized equality scores



xii. Hypocrisy: States Ranked in terms of popularity for searching 'gay porn' on Google

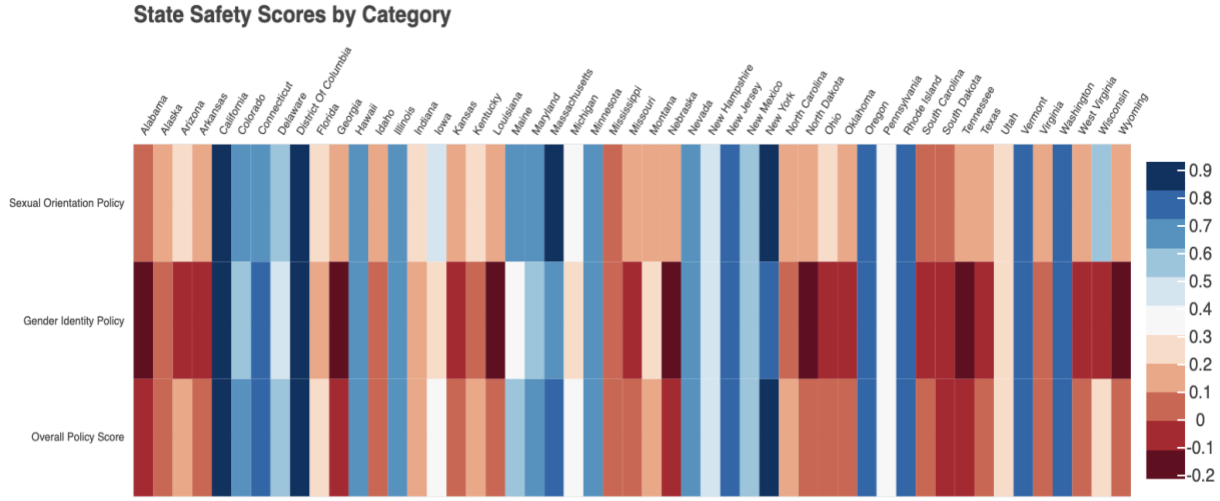


Rebecca Reilly

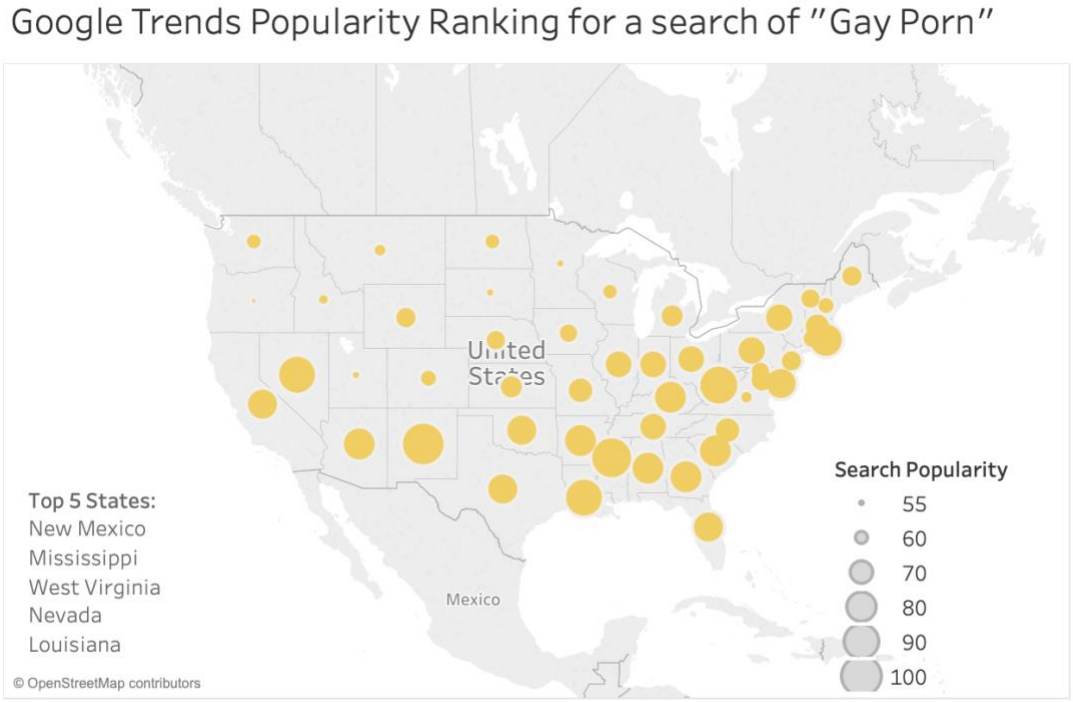
STORY:

WHICH STATES ARE SAFEST? WHICH ARE HYPOCRITES?

After visualizing a heatmap depicting from red to blue the most dangerous to the safest for the LGBT community in terms of laws:



We can then contrast this analysis of their safety for LGBT people with their interest in searching for gay pornography



From the heatmap, it is obvious that the safest and most dangerous states are in line with how the HRC has classified each state. The safest are California, New York, D.C., and Massachusetts, while the most dangerous are Alabama, Georgia, South Dakota, and South Carolina. These are not surprising results, given the blue-assigned states are known for being more liberal, while the red-assigned are conversely known for being conservative. What is surprising is that when considering all factors analyzed – LGBT workplace discrimination rates, legislation – the states that have the highest rates of searching for gay pornography are actually those with the higher discrimination rates (Nevada, New Mexico) and/or states with consistently negative ratings for their policies affecting LGBT people. Of course a caveat to this is that in more conservative states, this data could reflect less outward identification as LGBT and more incentive to search pornography.

GITHUB:

https://github.com/r-reilly/data_visualizations

CITATIONS:

Population Data:
http://www.lgbtmap.org/equality-maps/lgbt_populations
https://news.gallup.com/poll/234863/estimate-lgbt-population-rises.aspx

State Equality Data:
http://www.lgbtmap.org/equality_maps/profile_state/MT
https://assets2.hrc.org/files/assets/resources/SEI-2018-Categories.pdf?_ga=2.67154065.121360372.1556733446-1172111344.1555530504

Discrimination Data:
https://www.umass.edu/employmentequity/diversity-analytics/data/lgbt

Law Data:
https://www.equaldex.com/

Code Appendix

- Population Chloropleth and SEI Category Map: Tableau
- Scatterplot: Tableau
- Bar Chart Bases: Tableau
- Histogram: Tableau
- Treemap: Tableau
- Bubblemap: Tableau
- Boxplot: Tableau
- Bar Chart SEI Level: Plotly

```
# plotly standard imports
import plotly.graph_objs as go
import plotly.plotly as py
# Cufflinks wrapper on plotly
import cufflinks
# Data science imports
import pandas as pd
import numpy as np
# Options for pandas
pd.options.display.max_columns = 30
# Display all cell outputs
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = 'all'
from plotly.offline import iplot
cufflinks.go_offline()
# Set global theme
cufflinks.set_config_file(world_readable=True, theme='pearl')

sei.groupby('level').count().iplot(
    kind='bar', yTitle='level', xTitle='count', linecolor='black',
    title='Count per 2018 State Equality Index Level')
```

- Connection Map: R

```
par(mar=c(0,0,0,0))
map('state', col="#f2f2f2", fill=TRUE, bg="#A6CAE0",  mar=rep(0,4), border=0)

## oklahoma, texas, tennessee
oklahoma=c(-97.5164, 35.4676)
texas=c(-99.9018,31.9686)
tennessee=c(-86.5804, 35.5175)
data=rbind(oklahoma, texas, tennessee) %>% as.data.frame()
colnames(data)=c("long","lat")


points(x=data$long, y=data$lat, col="black", cex=3, pch=20)


inter <- gcIntermediate(oklahoma,  texas, n=50, addStartEnd=TRUE, breakAtDateLine=F)
```

```
lines(inter, col="black", lwd=2)

inter <- gcIntermediate(texas,  tennessee, n=50, addStartEnd=TRUE, breakAtDateLine=F)
lines(inter, col="black", lwd=2)

inter <- gcIntermediate(tennessee,  oklahoma, n=50, addStartEnd=TRUE, breakAtDateLine=F)
lines(inter, col="black", lwd=2)
```

- Heatmap: Bokeh

```
import pandas as pd

from bokeh.io import output_file, show
from bokeh.models import BasicTicker, ColorBar,
ColumnDataSource, LinearColorMapper, PrintfTickFormatter
from bokeh.plotting import figure
from bokeh.transform import transform

output_file("state_ratings.html")

data = bills.set_index('State')
data.drop(['count_since_2015',  'count_fully_protective_laws',  'score'],
          axis=1, inplace=True)
data.columns = ['Sexual Orientation Policy', 'Gender Identity Policy',
                'Overall Policy Score']
data.columns.name = 'Score'

# reshape to 1D array or rates with a month and year for each row.
df = pd.DataFrame(data.stack(), columns=['score']).reset_index()

source = ColumnDataSource(df)


colors = ['#67001f', '#b2182b', '#d6604d', '#f4a582', '#fddbc7',
          '#f7f7f7', '#d1e5f0', '#92c5de',
          '#4393c3', '#2166ac', '#053061']
mapper = LinearColorMapper(palette=colors, low=df.score.min(), high=df.score.max())

p = figure(plot_width=800, plot_height=300, title="State Safety Scores by Category",
           x_range=list(data.index), y_range=list(reversed(data.columns)),
           toolbar_location=None, tools="", x_axis_location="above")

p.rect(x="State", y="Score", width=1, height=1, source=source,
       line_color=None, fill_color=transform('score', mapper))

color_bar = ColorBar(color_mapper=mapper, location=(0, 0),
                     ticker=BasicTicker(desired_num_ticks=len(colors)))

p.add_layout(color_bar, 'right')

p.axis.axis_line_color = None
p.axis.major_tick_line_color = None
p.axis.major_label_text_font_size = "5pt"
p.axis.major_label_standoff = 0
```

```
p.xaxis.major_label_orientation = 1.0

show(p)
```

- Streamgraph: Altair

```
knitr::opts_chunk$set(echo = TRUE, cache = TRUE)

pop = {'year':['2012','2013','2014','2015','2016','2017'],
  'Millenials':[5.8,6.0,6.3,6.7,7.3,8.2],
       'Gen X':[3.2,3.3,3.4,3.3,3.2,3.5],
      'Baby Boomers':[2.7,2.7,2.7,2.6,2.4,2.4]}
pop = pd.DataFrame(pop)
df = pop.reset_index()
pop = pd.melt(df, id_vars='year', value_vars=['Millenials', 'Baby Boomers', 'Gen X'])
pop.columns = ['Year', 'Generation', '% LGBT']

alt.Chart(source).mark_area().encode(
    alt.X('Year:T',
        axis=alt.Axis(format='%Y', domain=False, tickSize=0)),
    alt.Y('% LGBT:Q', stack='center', axis=None),
    alt.Color('Generation:N',
        scale=alt.Scale(scheme='category20b'))).interactive()
```