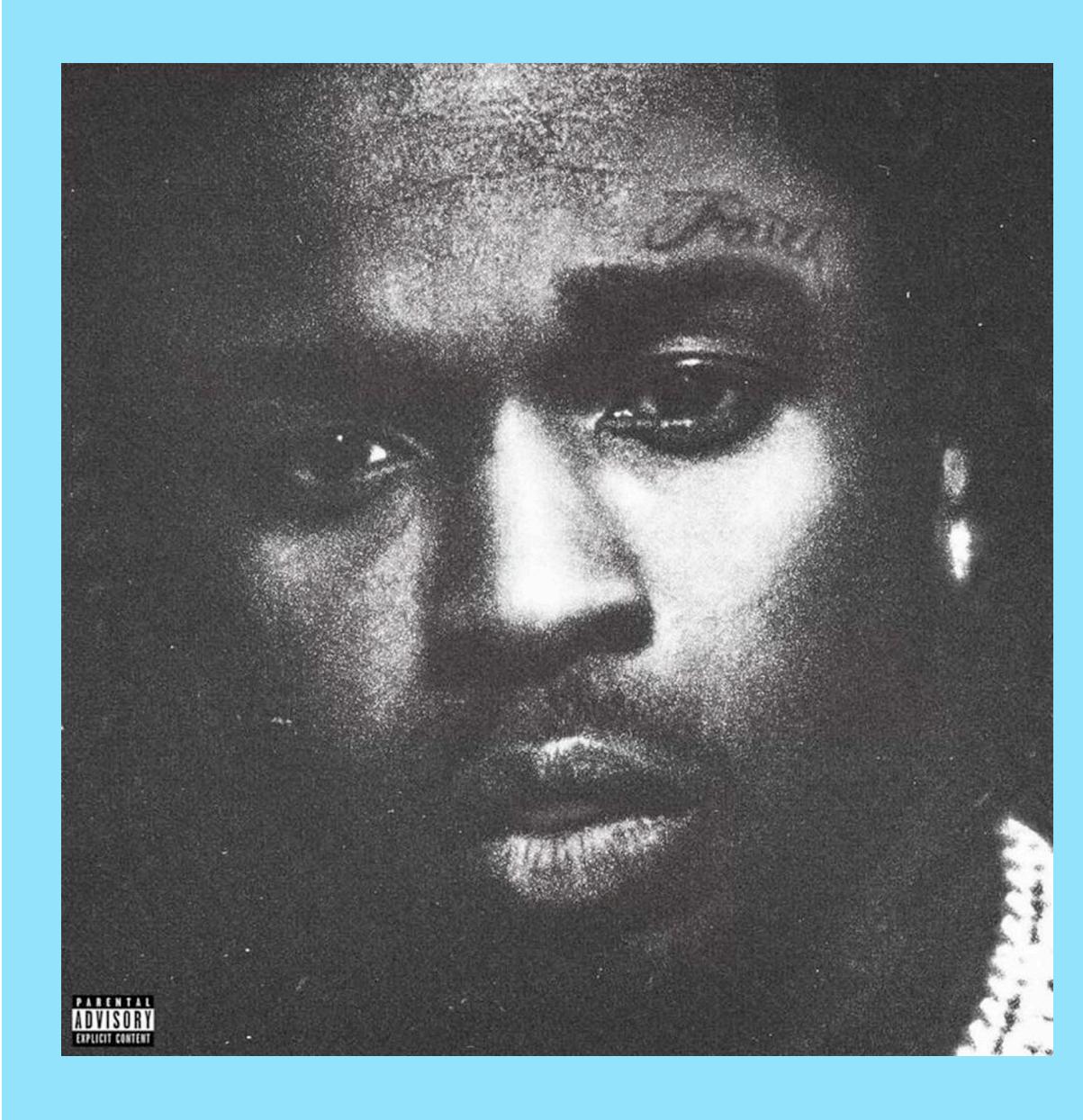




# Final Project

## Data Analytics Bootcamp 2021

# 0: Introduction & Ideation



#82 worldwide

7.632.507 Follower

27.432.331 Monthly Listeners

Most listened to in...

London, GB

Los Angeles, USA

Chicago, USA

New York City,  
USA

Sydney, AU



#87 worldwide

31.049.851 Follower

27.035.390 Monthly Listeners

Most listened to in...

Los Angeles, USA

Chicago, USA

Dallas, USA

Houston, USA

Atlanta, USA



#320 worldwide

11.286.798 Follower

13.531.983 Monthly Listeners

Most listened to in...

Los Angeles, USA

Sydney, AU

London, GB

Chicago, USA

Dallas, USA

# 0: Introduction & Ideation



**Date of death**  
February 19, 2020

**Age**  
**20**

**Place of death**  
Los Angeles, USA

**Cause of death** Shot and killed



**Date of death**  
June 18, 2018

**Age**  
**20**

**Place of death**  
Deerfield Beach,  
USA

**Cause of death** Shot and killed

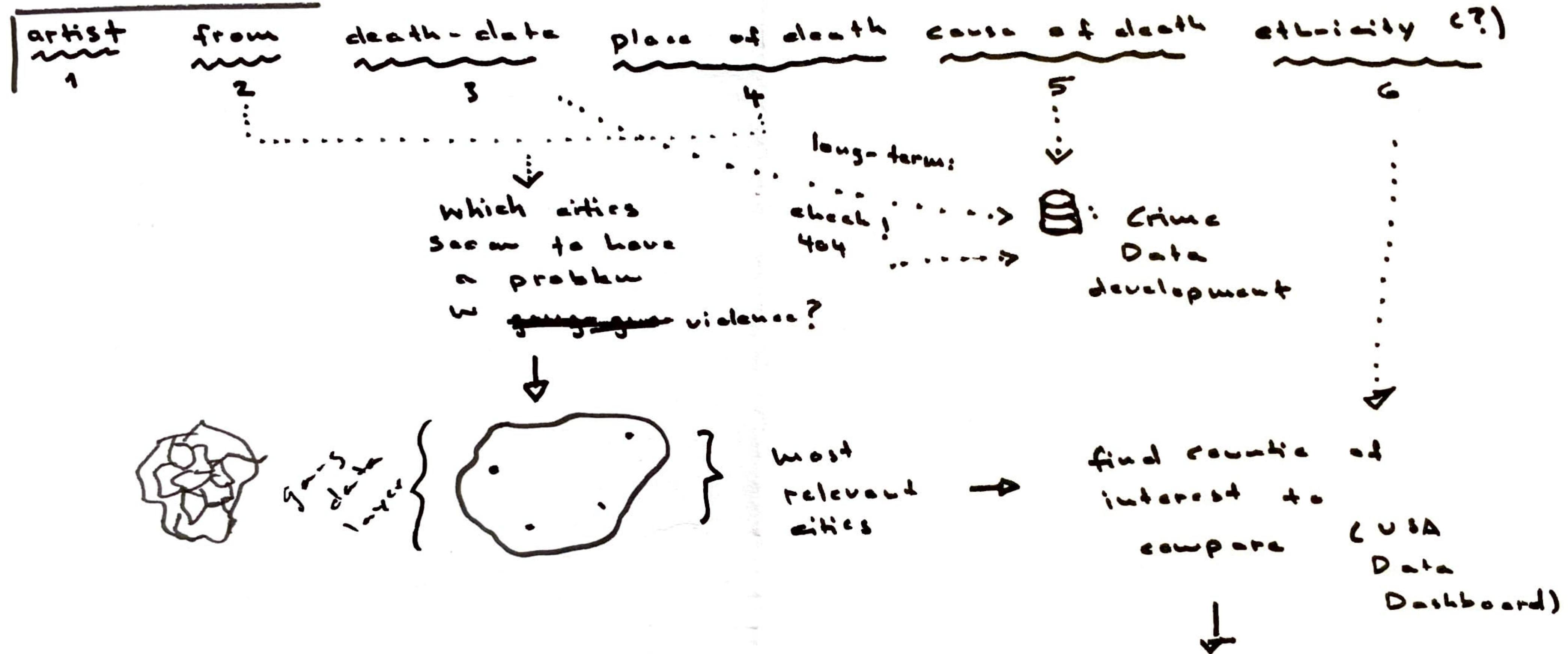


**Date of death**  
September 13, 1996

**Age**  
**25**

**Place of death**  
Las Vegas, Nevada

**Cause of death** Shot and killed



Acknowledgments  
Biases &  
Improvements  
in Approach

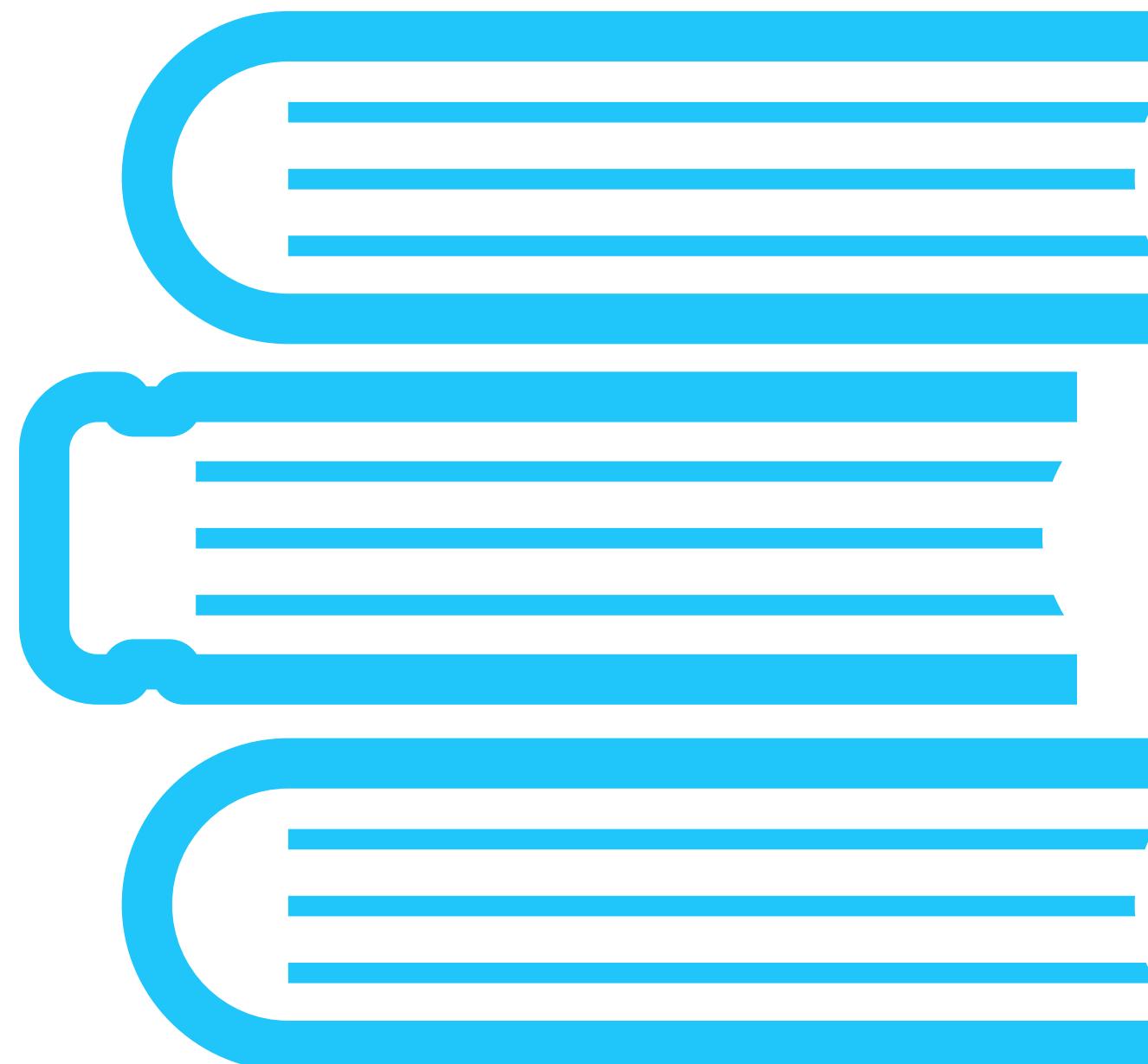
Quality  
problems  
Cross-Data  
Sources

Contextualize  
countries (mention  
policy)  
w/ w/ obsoletions  
& NLP

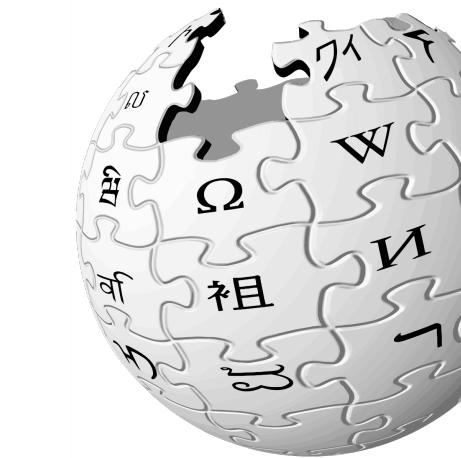
#1:

# Name Gathering

# 1: Name Gathering



wikipedia\_table



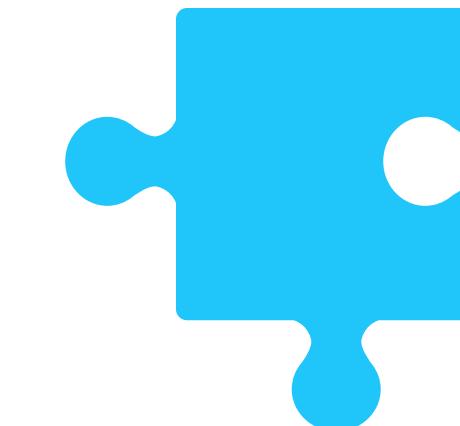
„List of murdered  
hip hop  
musicians“

.....  
.

634

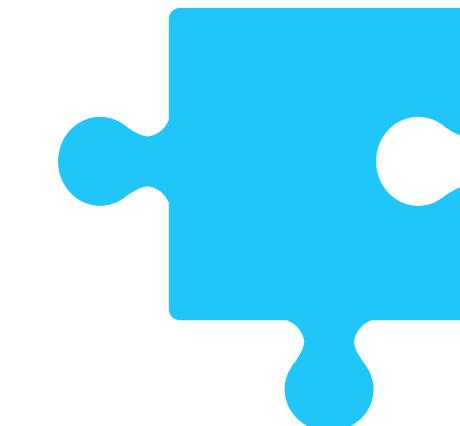
backlog: Further validation necessary

name\_list\_1



[rap.fandom.com](http://rap.fandom.com)

name\_list\_2



[hiphopdatabase.fandom.com](http://hiphopdatabase.fandom.com)

1273

names after cleaning



wiki\_ping\_filtering

639

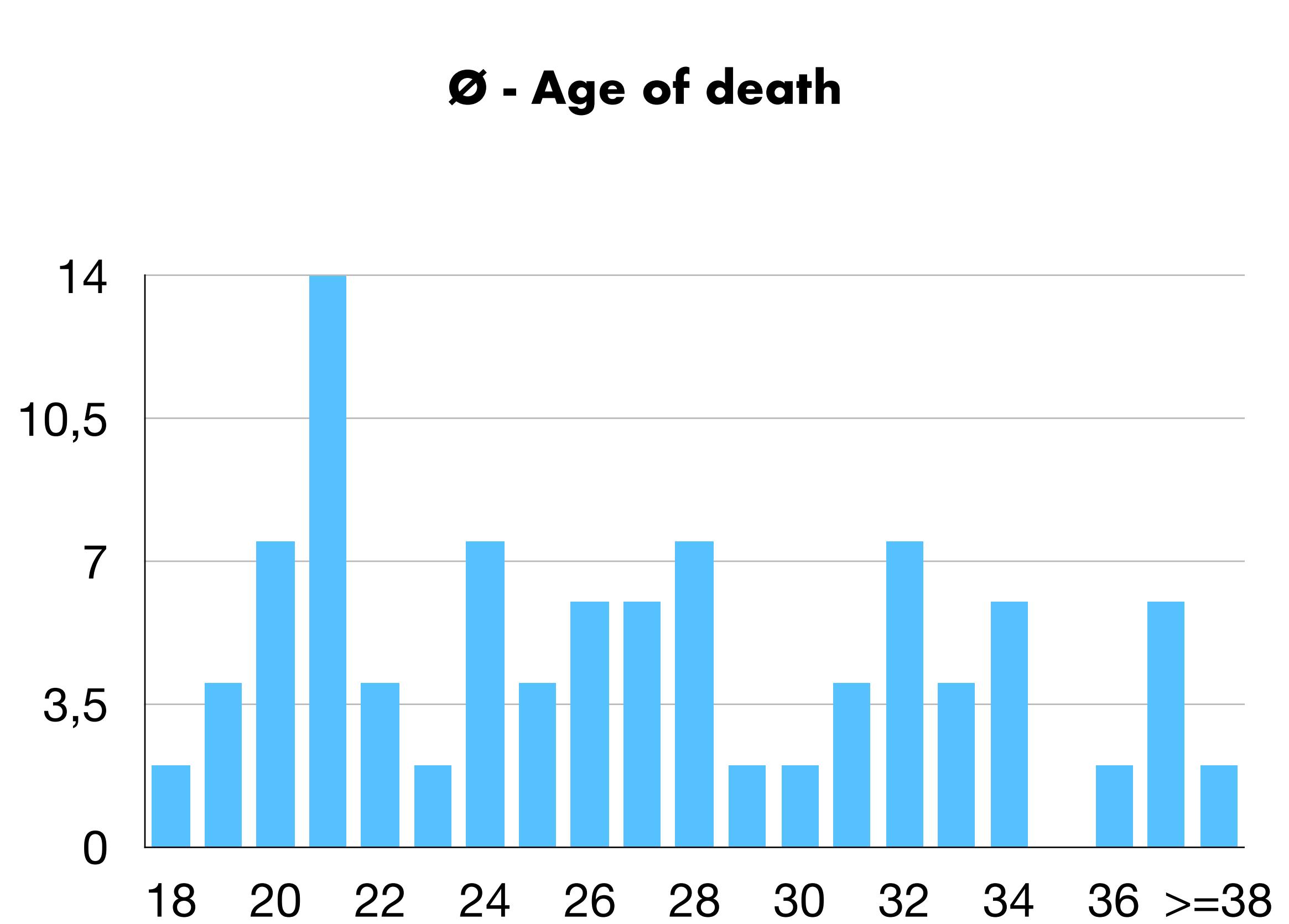
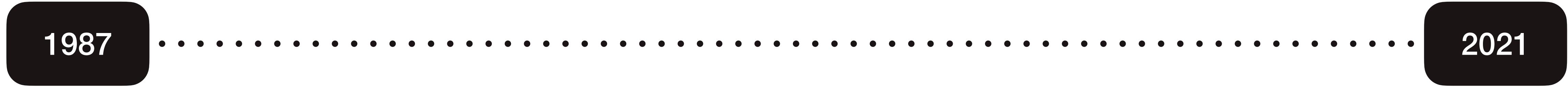
names after wiki ping filtering

# #2: Understanding the data

# 2: Understanding the data



## 2: Understanding the data



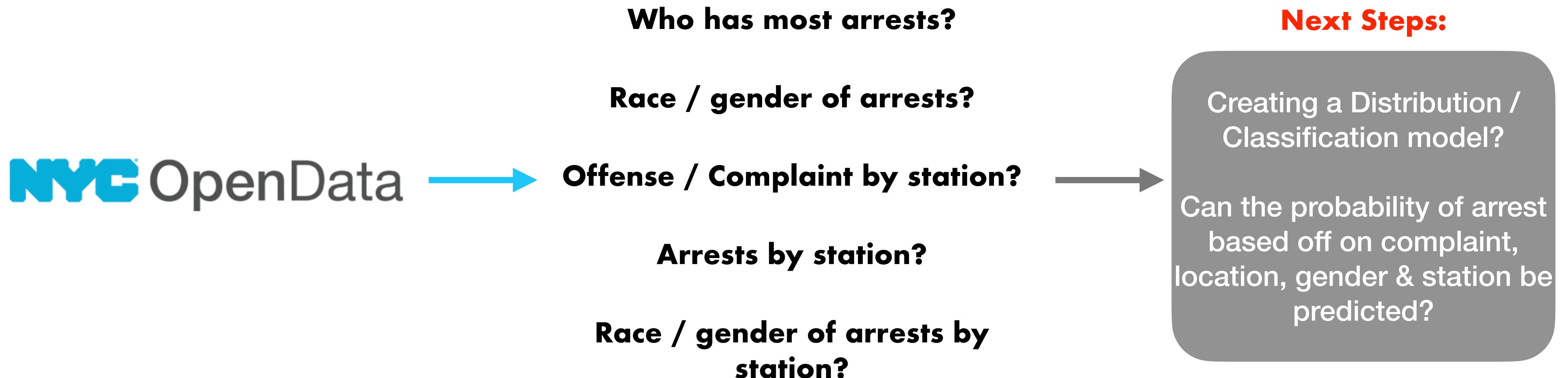
**Top 3 Cities USA**



#3:

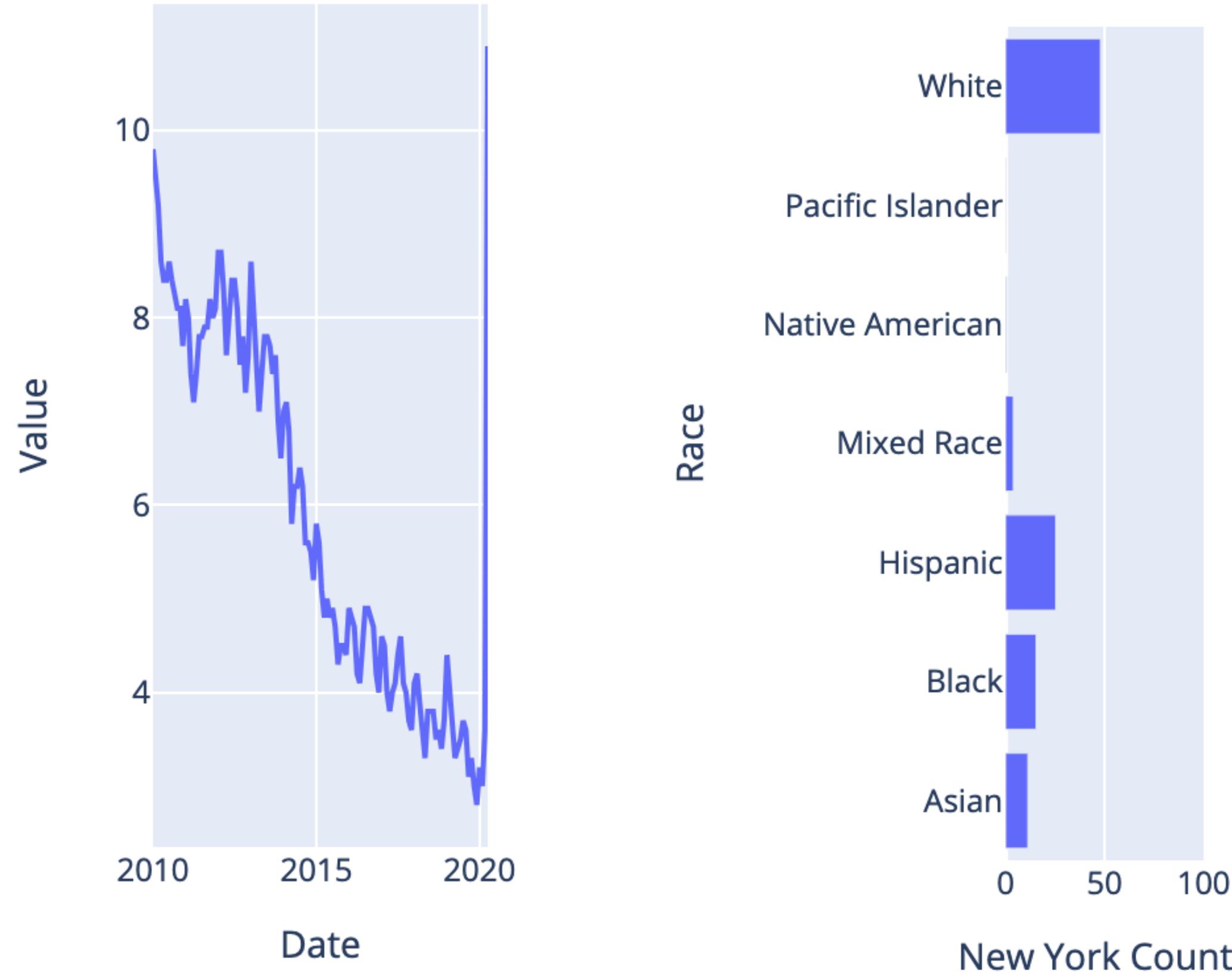
# NYC Crime Data Analysis

# 3: NYC Crime Data Analysis



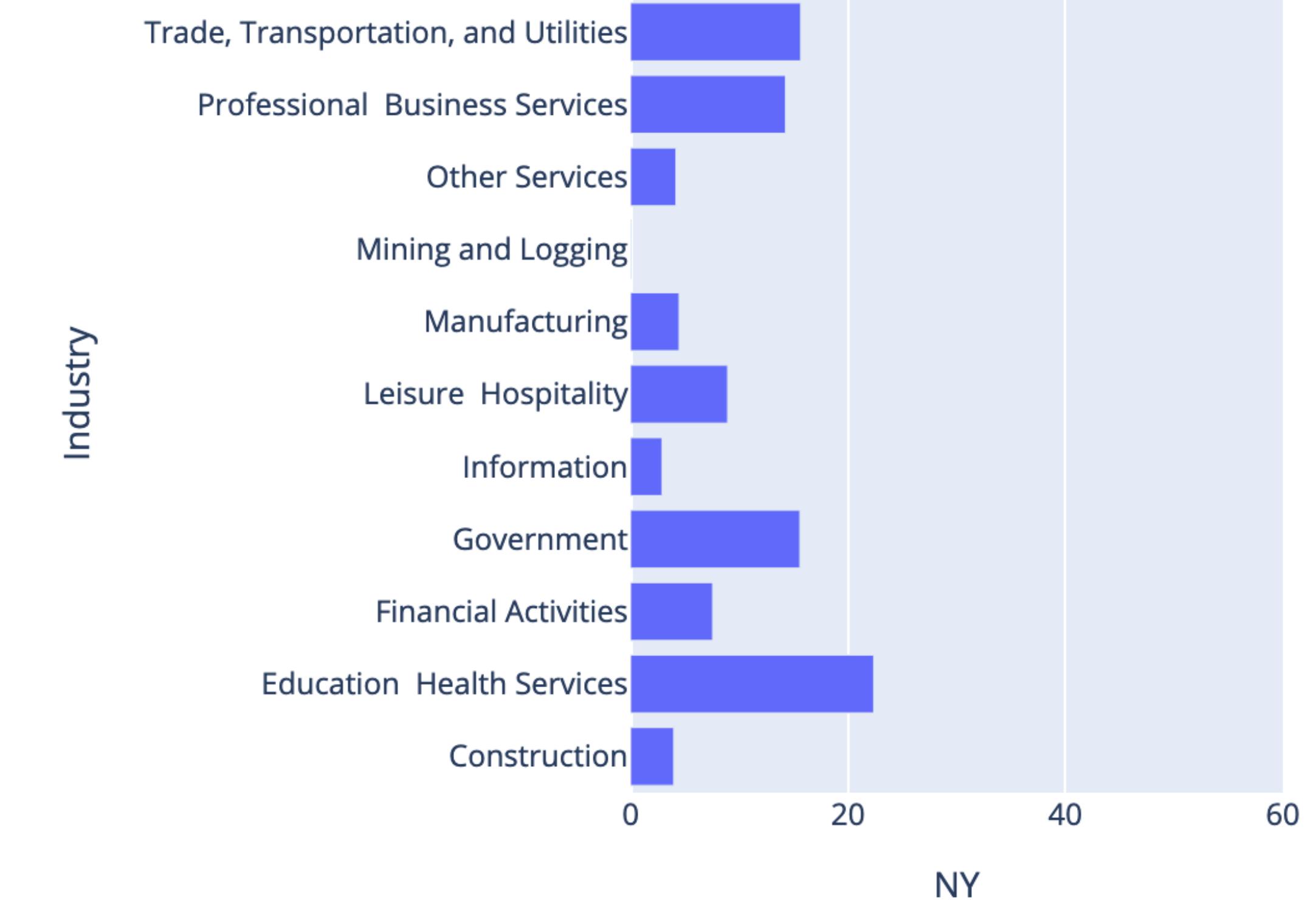
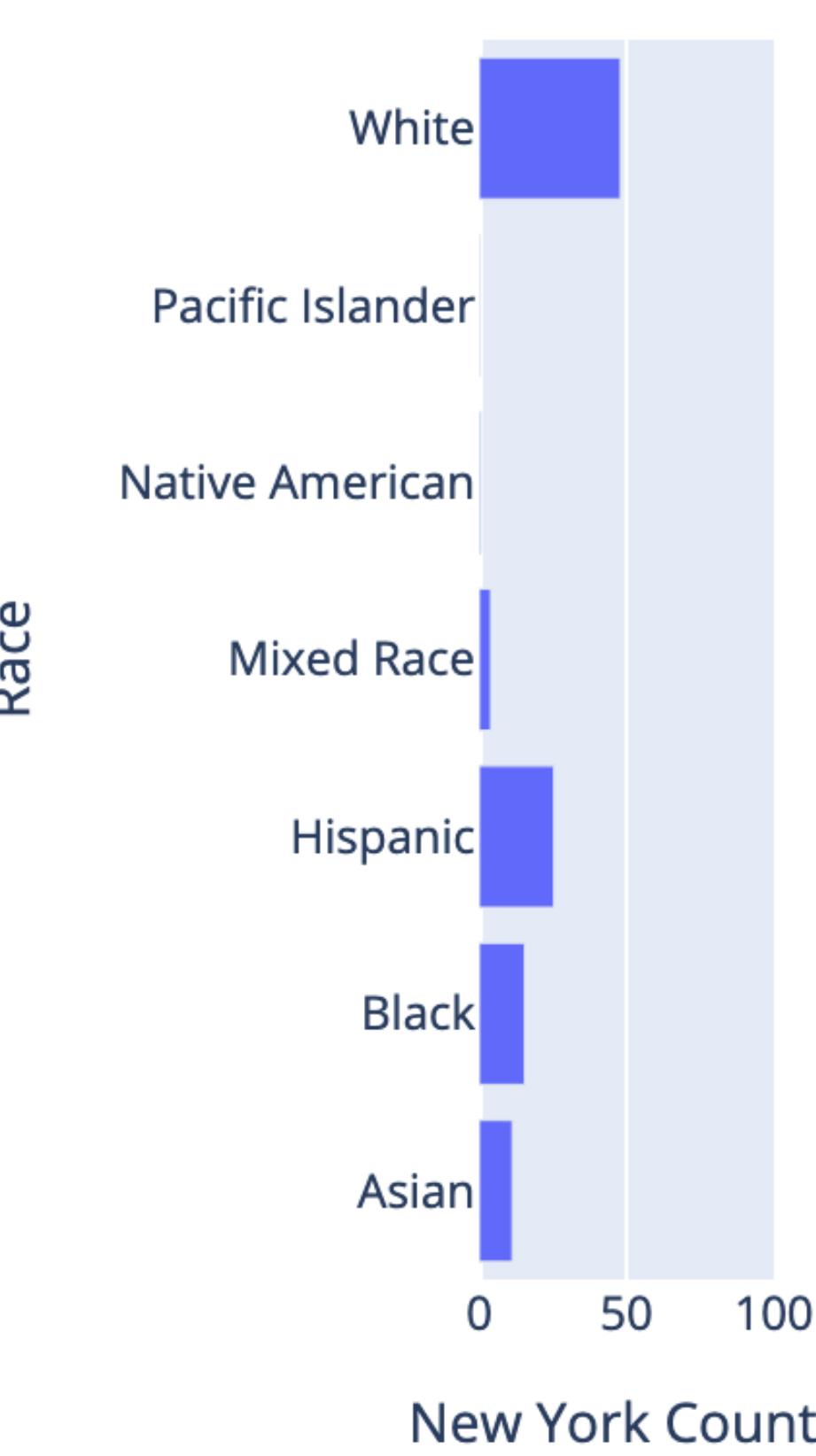
# #4: **Social Inequality Studies**

# 4: Social Inequality Studies



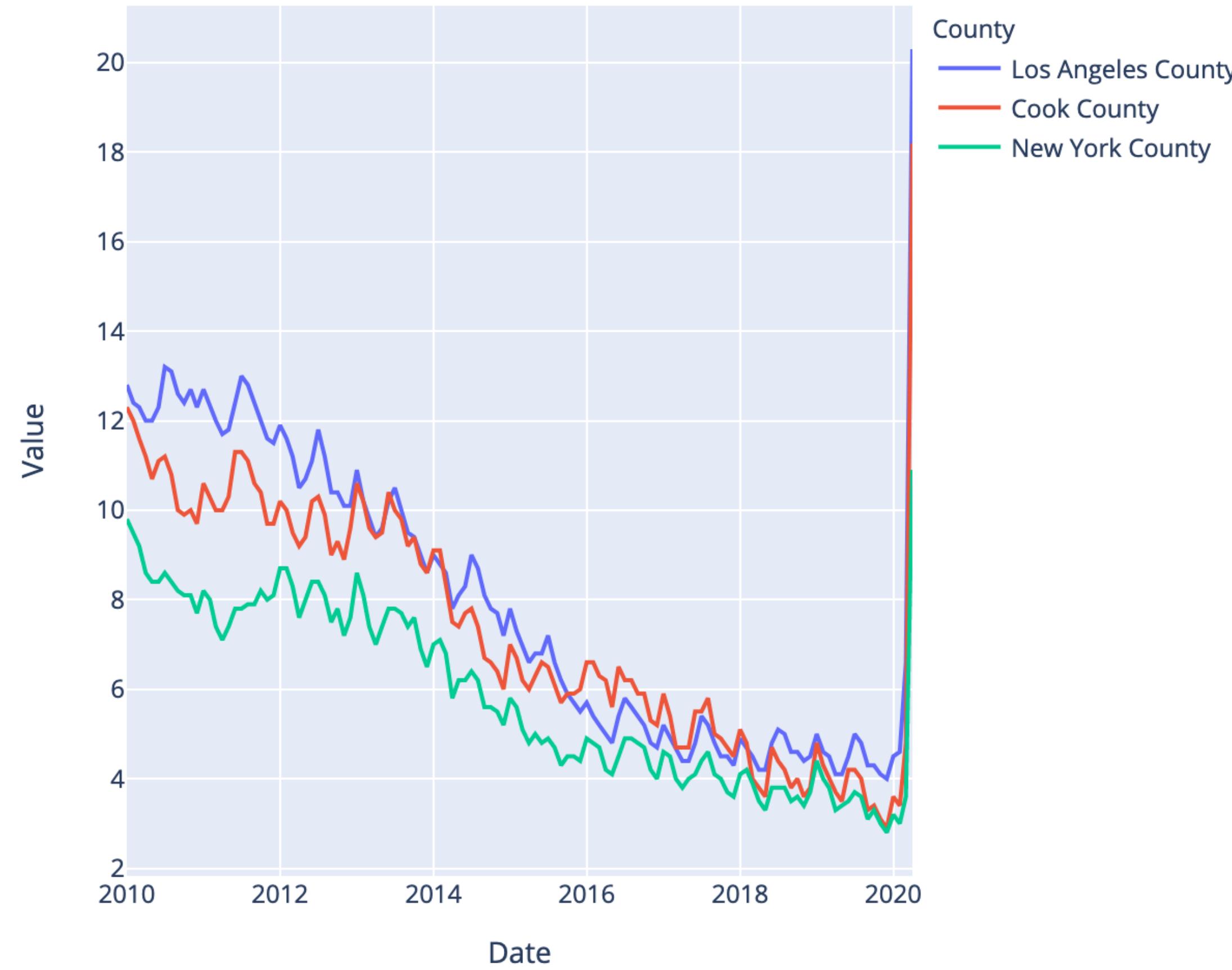
**NYC Unemployment Rate**

**Percent Population by race**



**Employment by Industry**

# 4: Social Inequality Studies

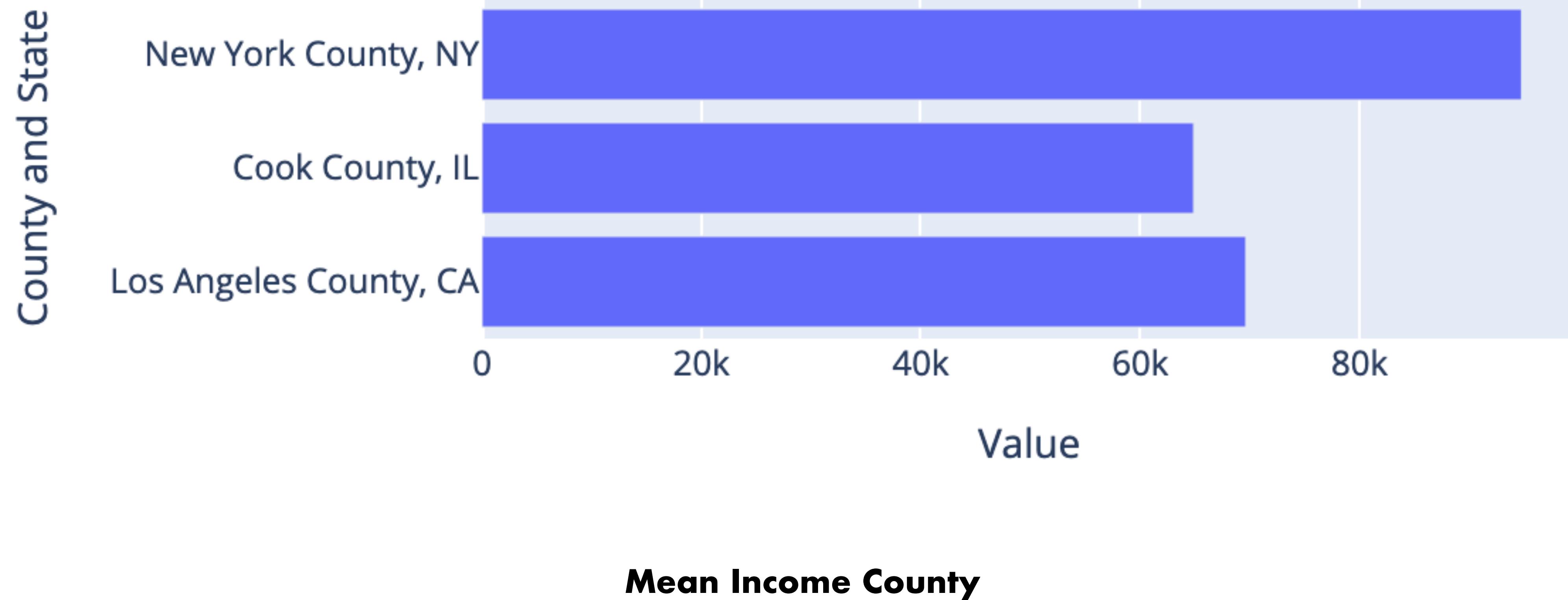


**Unemployment Rate History**

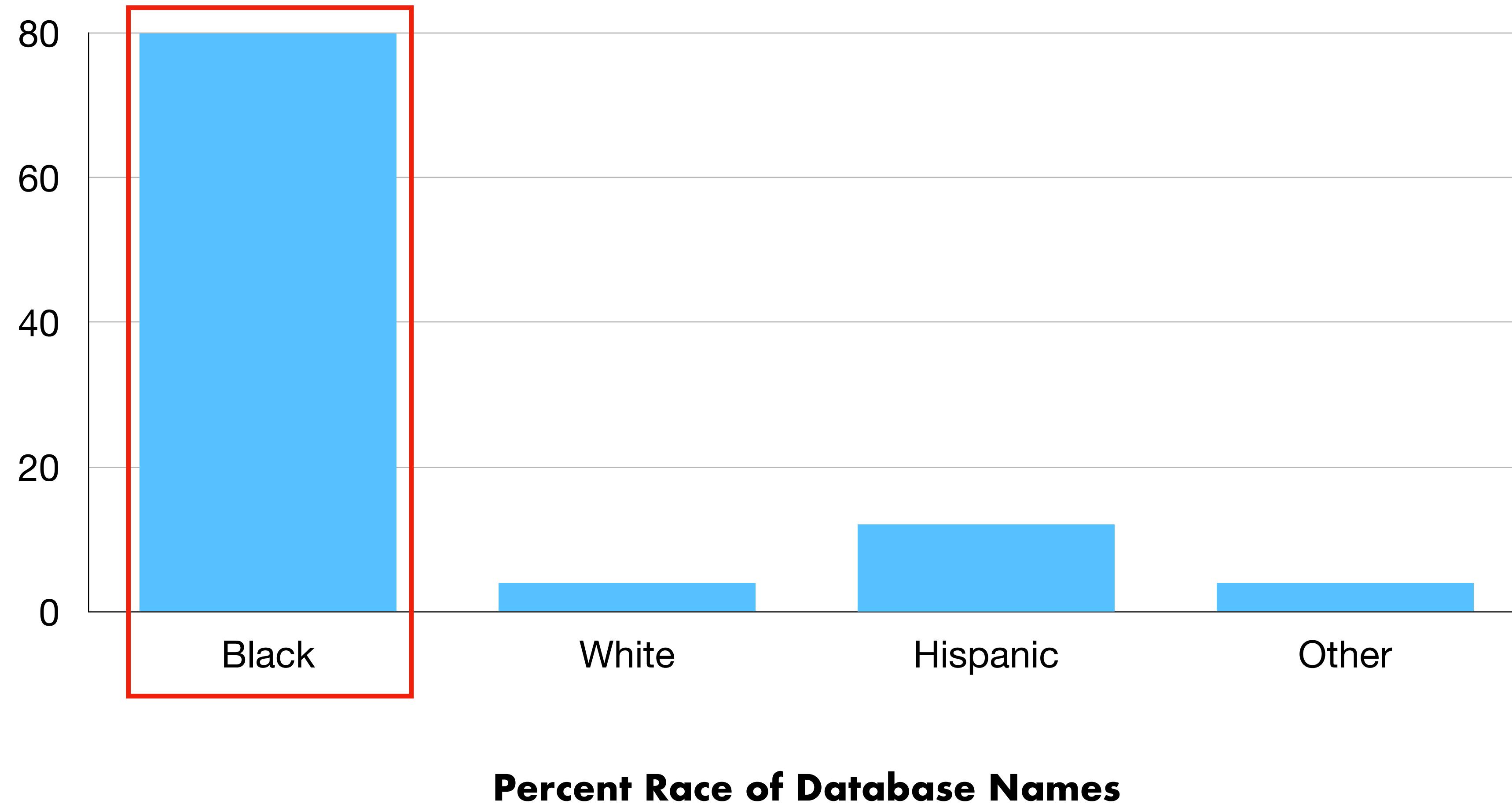


**Percent Race of County Population**

## 4: Social Inequality Studies - Top 3



## 4: Social Inequality Studies - Top 3



# #5: **Challenges Black Communities face in the US**

# 5: Challenges Black Communities face in the US

A  
Find Counties of interest to compare

B  
Investigating the County through policy & media

C  
Quantify Problem

## Next Steps:



**Acknowledge Biases & Improvements in Approach**

**Using US  
data.gov**

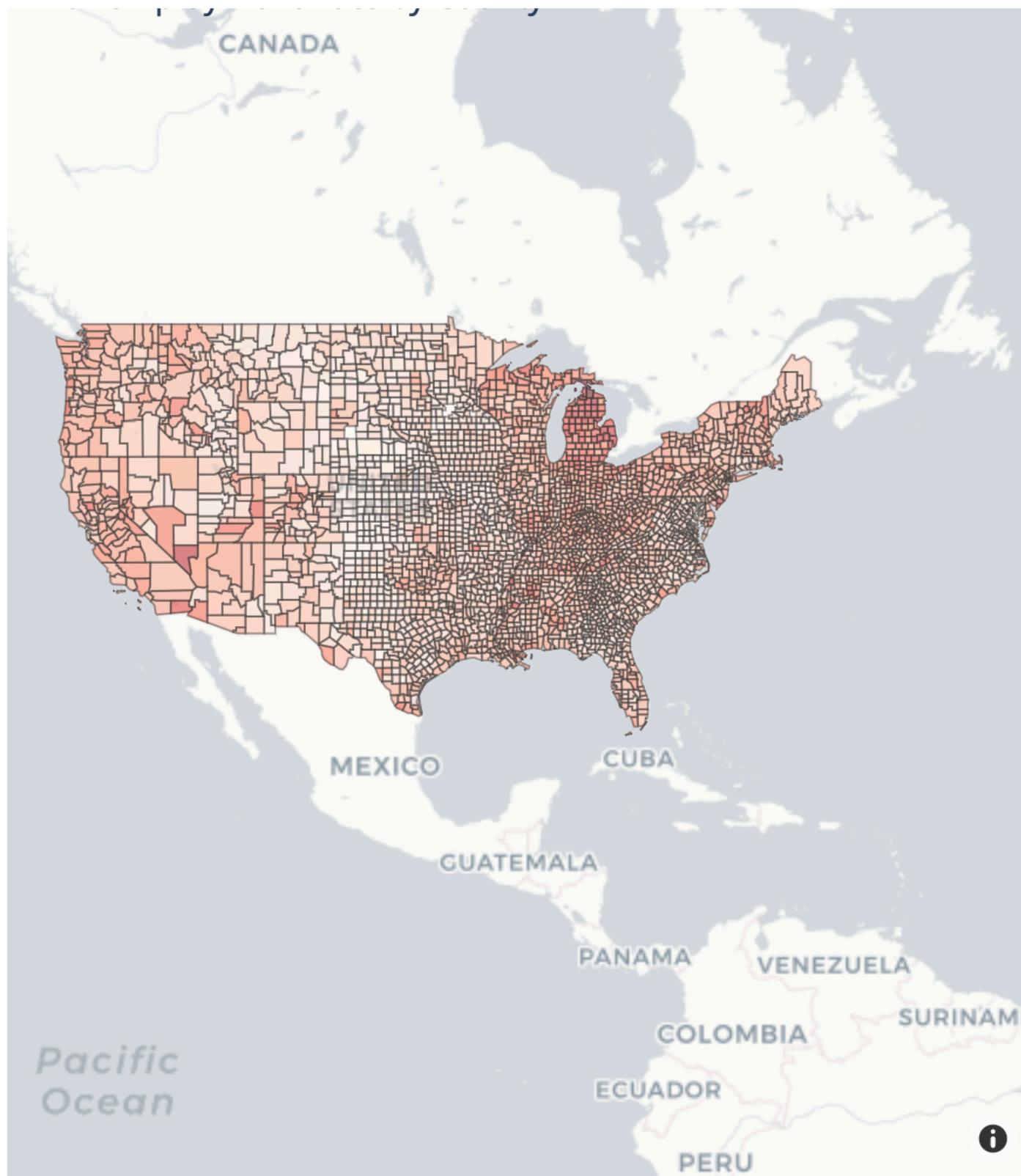
**Using  
Webscraping &  
NLP**

**Using Open Data  
Sources**

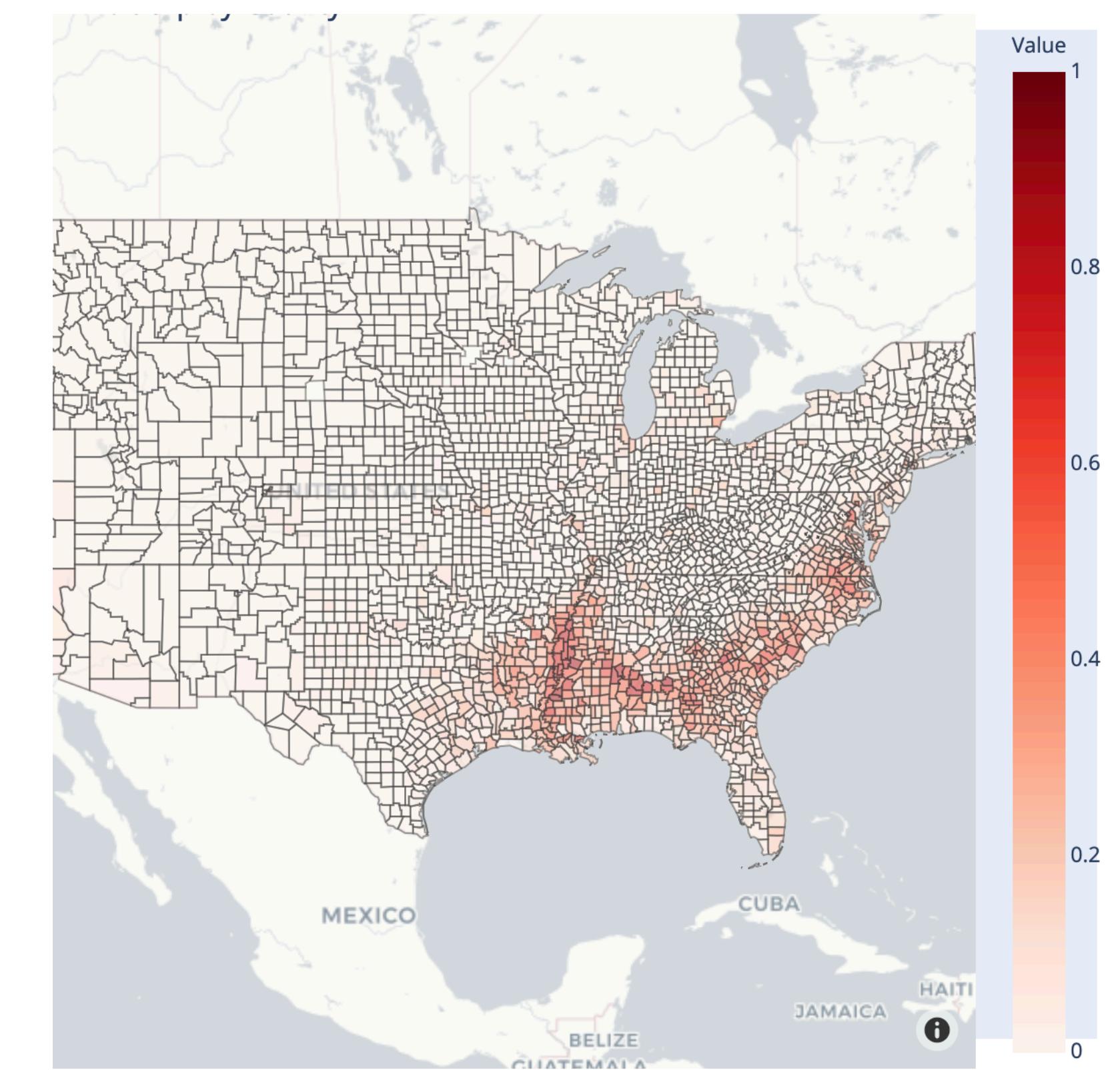
**#5A:**  
**Find Counties of interest to**  
**compare**

## 5A: Find Counties of interest to compare

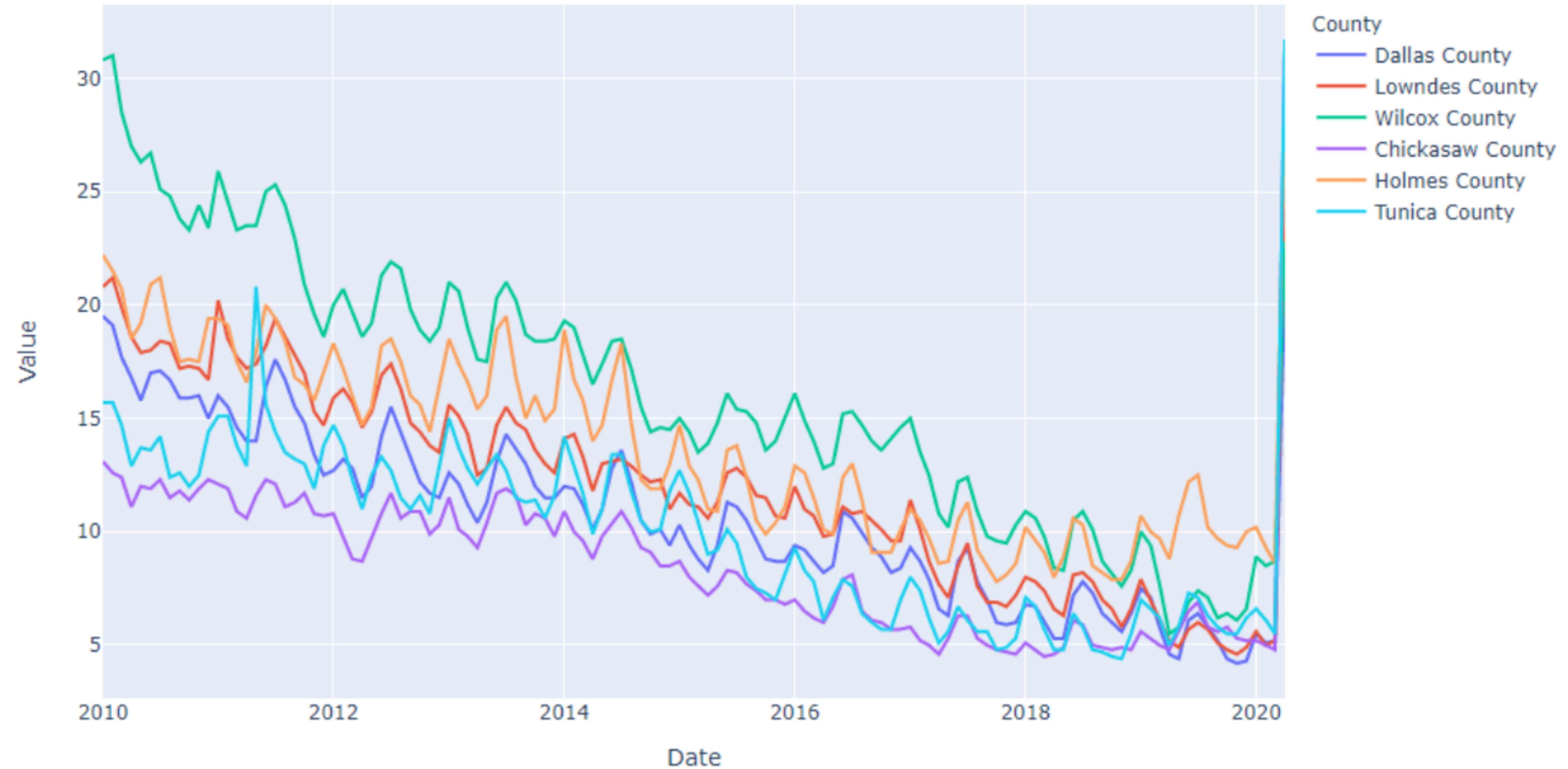
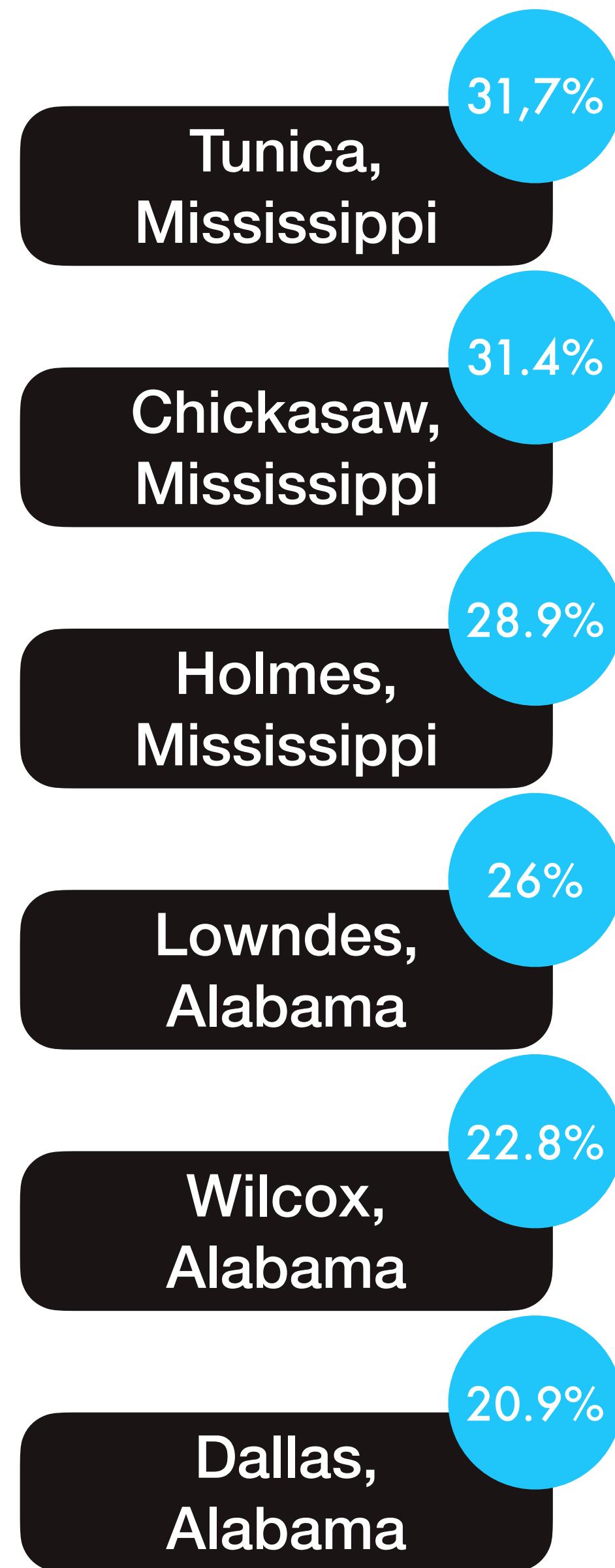
Search for the counties with highest unemployment rate



Use Census data to find areas with highest black population

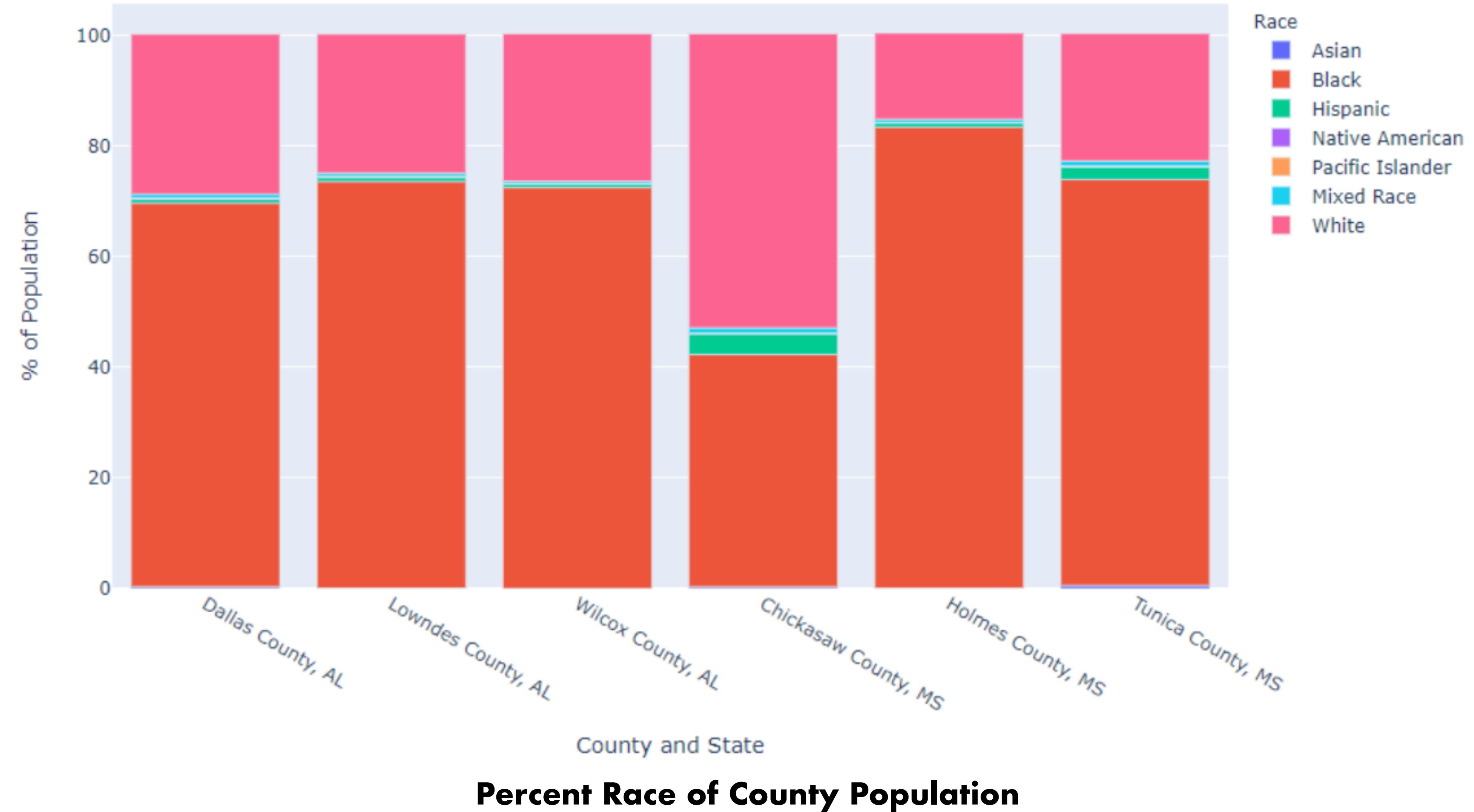


## 5A: Find Counties of interest to compare

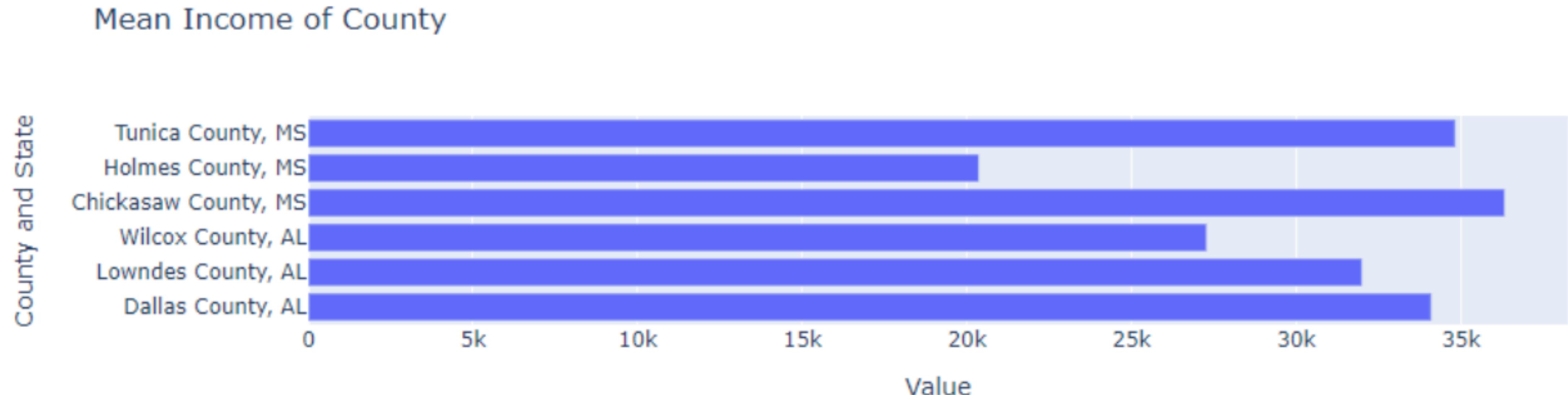


**Unemployment Rate History**

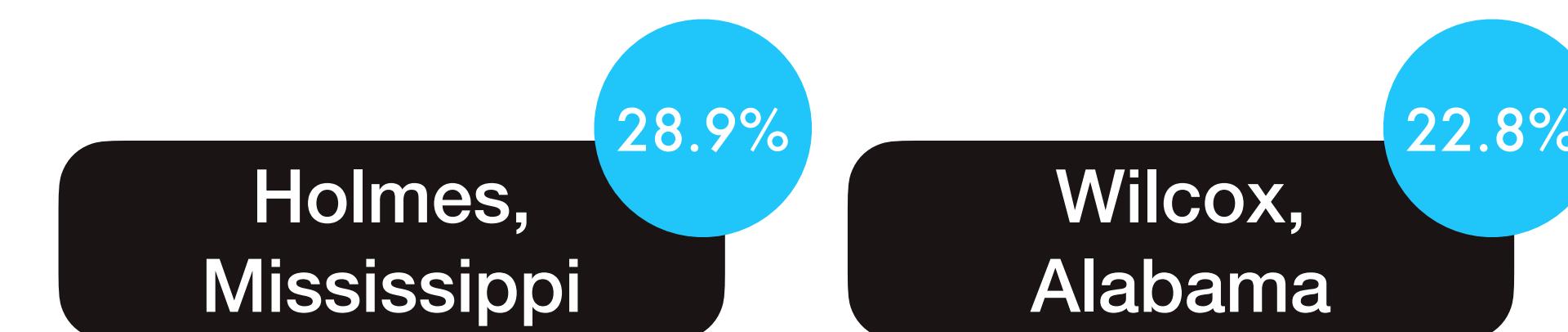
## 5A: Find Counties of interest to compare



## 5A: Find Counties of interest to compare

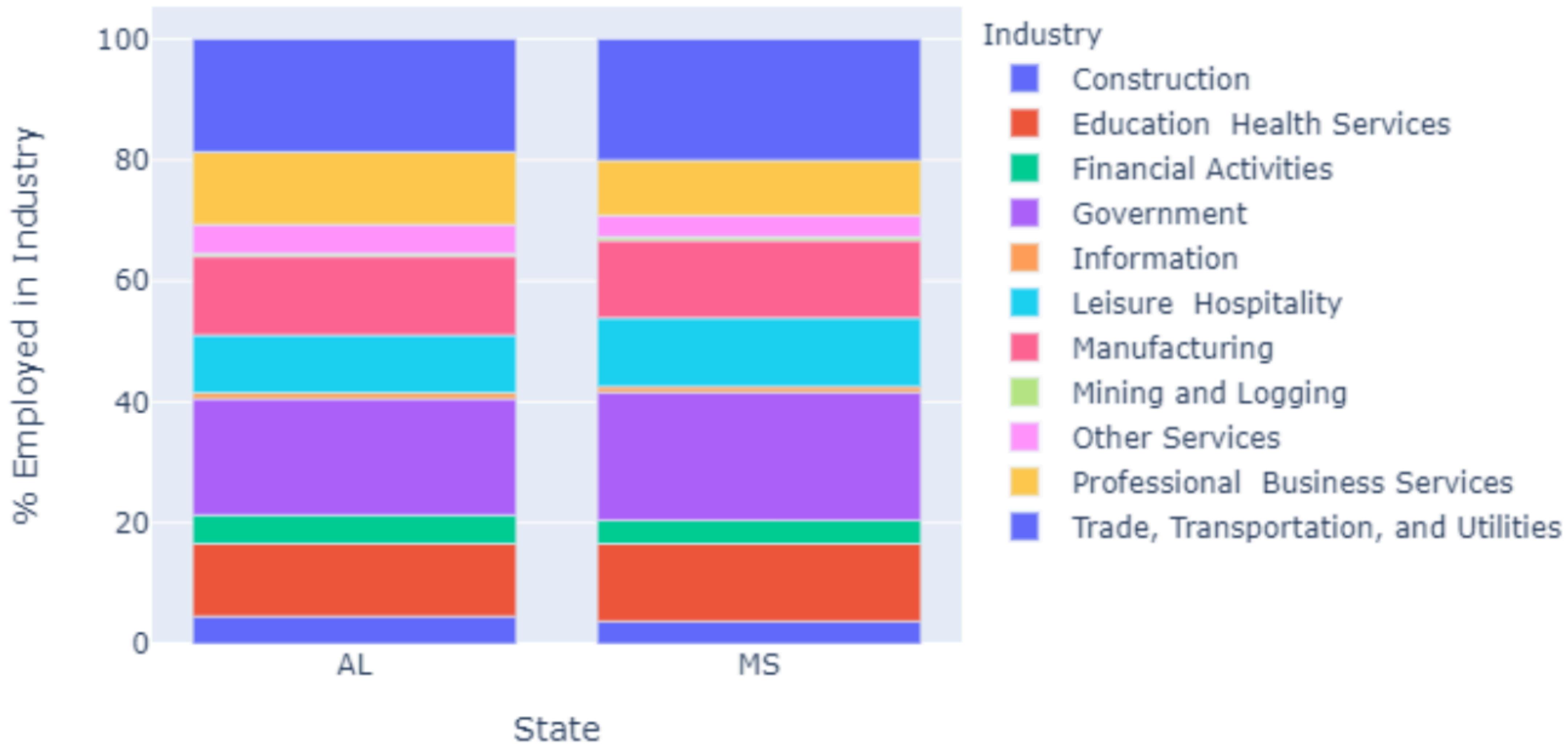


### Mean Income County



## 5A: Find Counties of interest to compare

### Percentage Employment by Industry



# #5B: Investigating the County through policy & media

## 5B: Investigating the County through policy & media

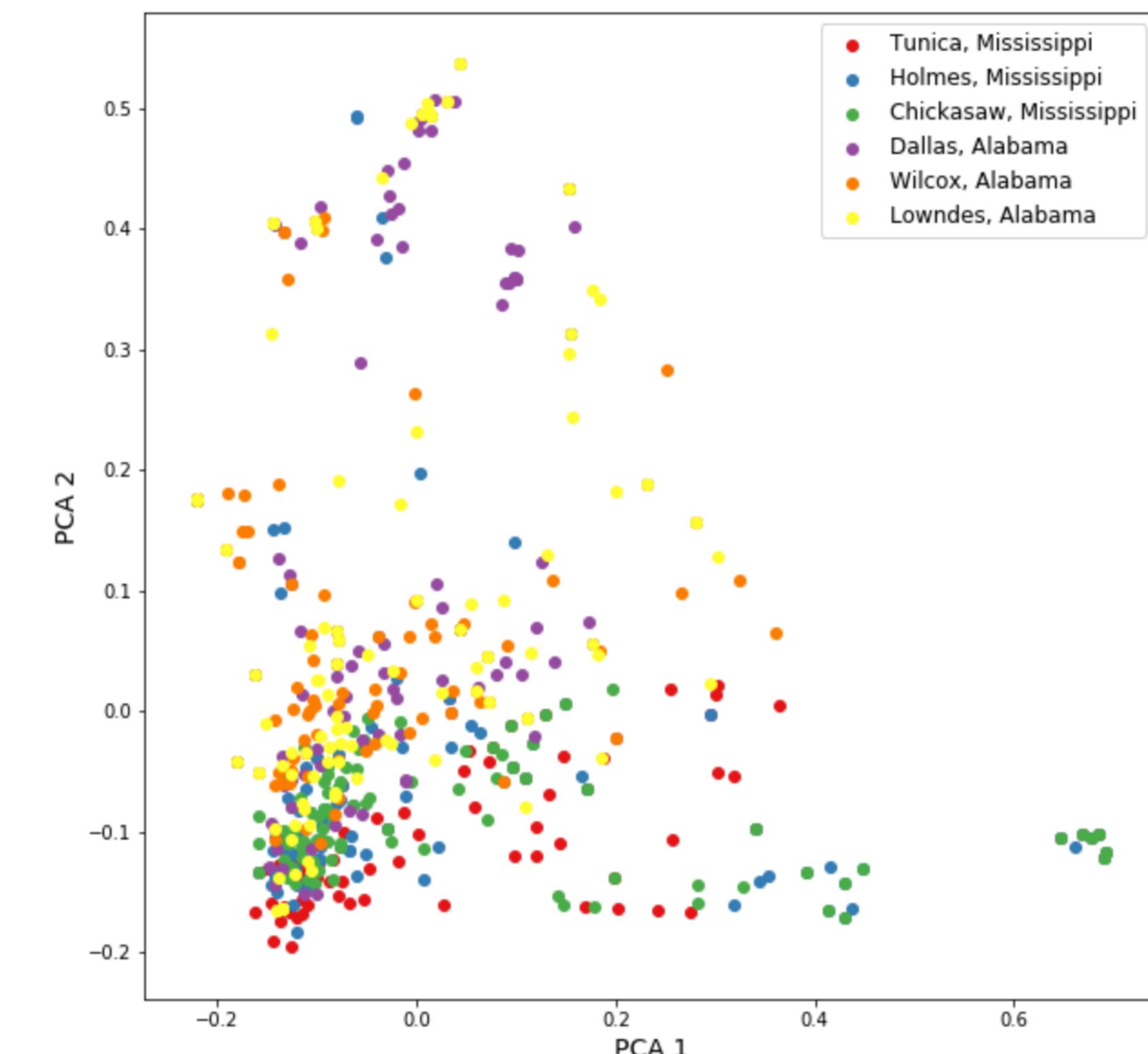
start by understanding media coverage and public policies in the area

NLP preprocessing/visualization methods

web scrape articles from the search  
“county\_name + ‘economy’” using a  
Google news query format ([https://news.google.com/rss/search?  
q={}](https://news.google.com/rss/search?q={}))

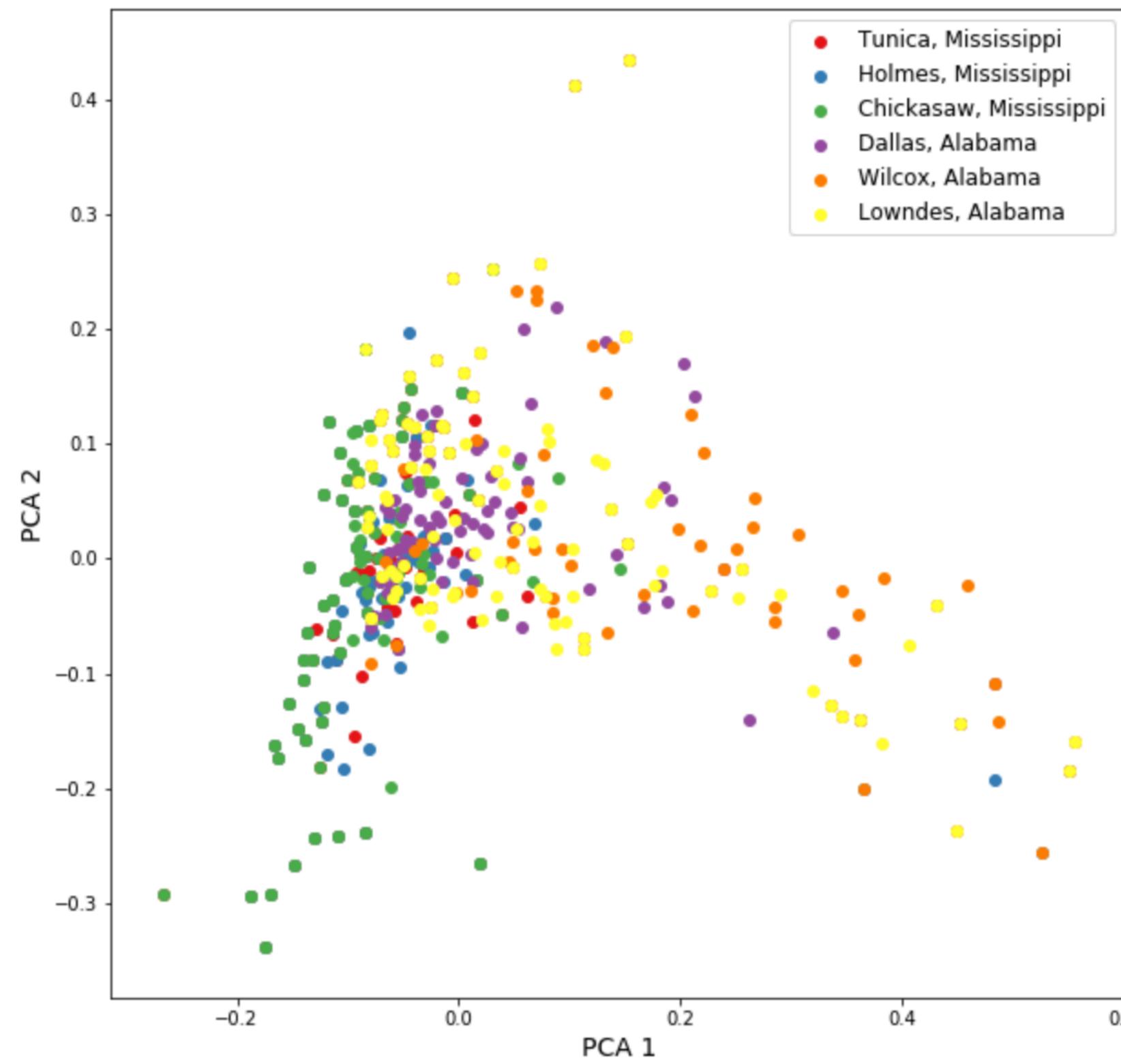
tokenized the words (using nltk to remove stop words and lemmatize) in each article using TF-IDF (Scikit-learn)

PCA analysis (Scikit-learn) for all the articles, plotting the first two components

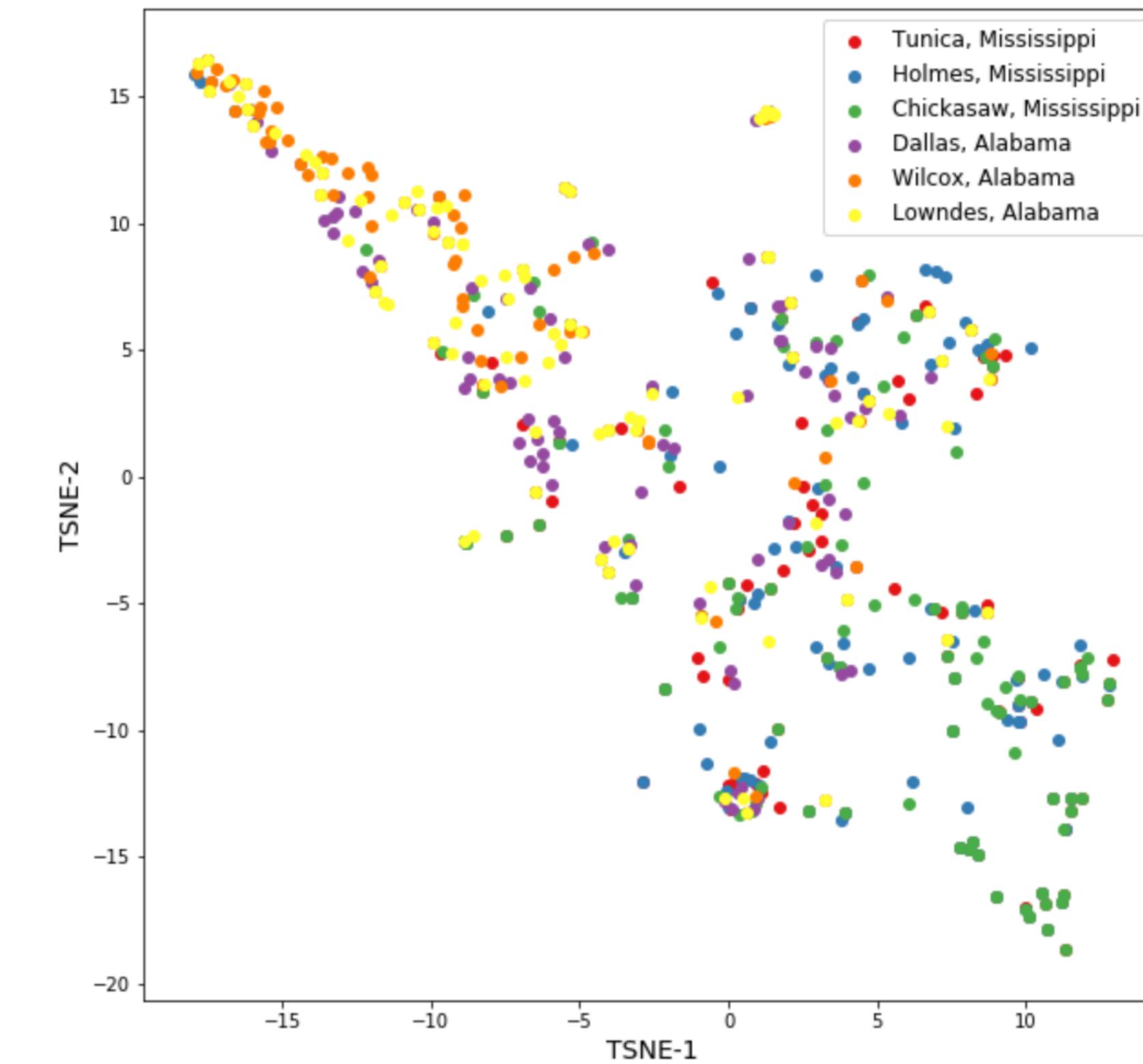


## 5B: Investigating the County through policy & media

PCA analysis (Scikit-learn) for all the articles, plotting the first two components on different time-frame

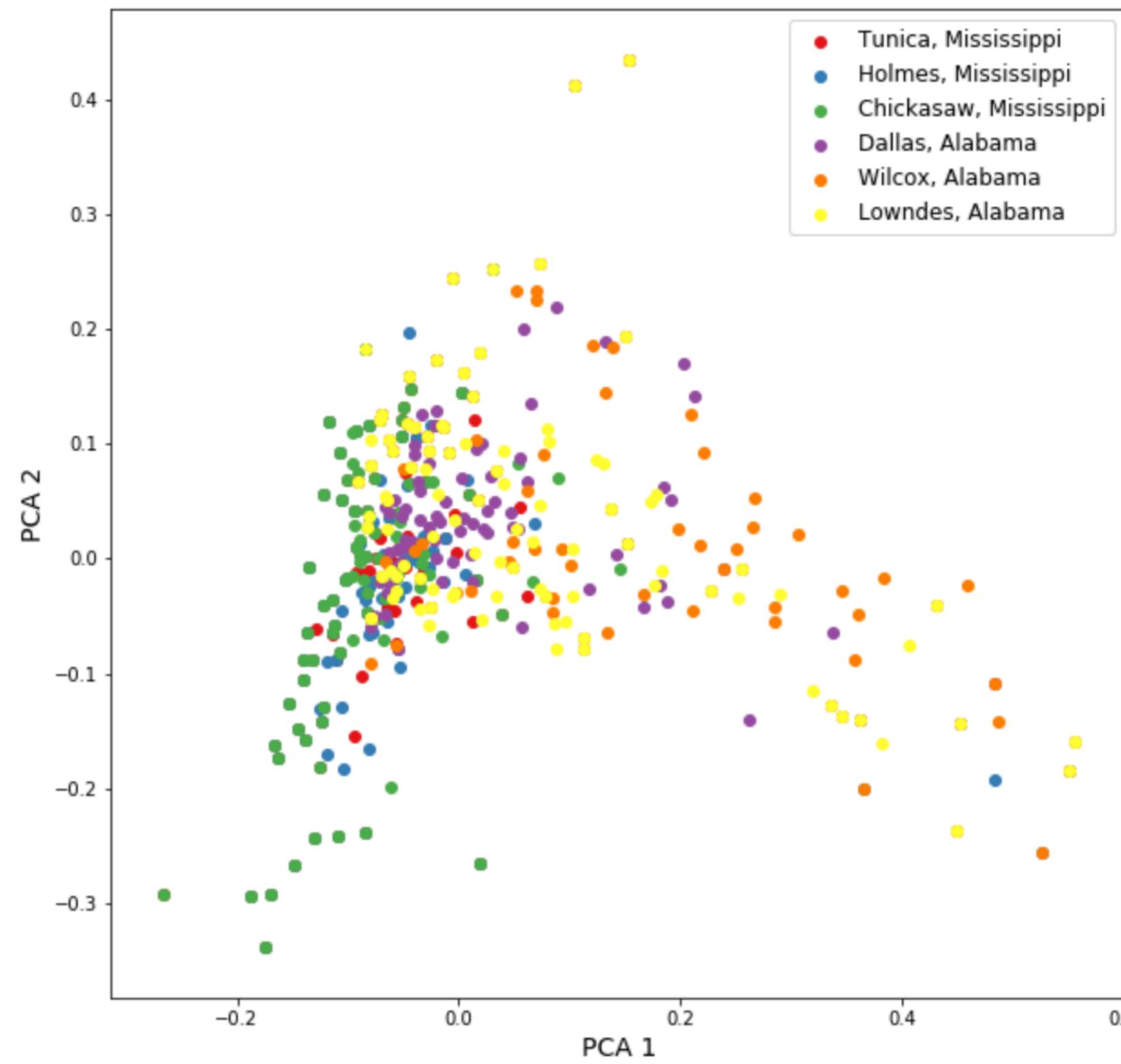


TSNE as well on the PCA components

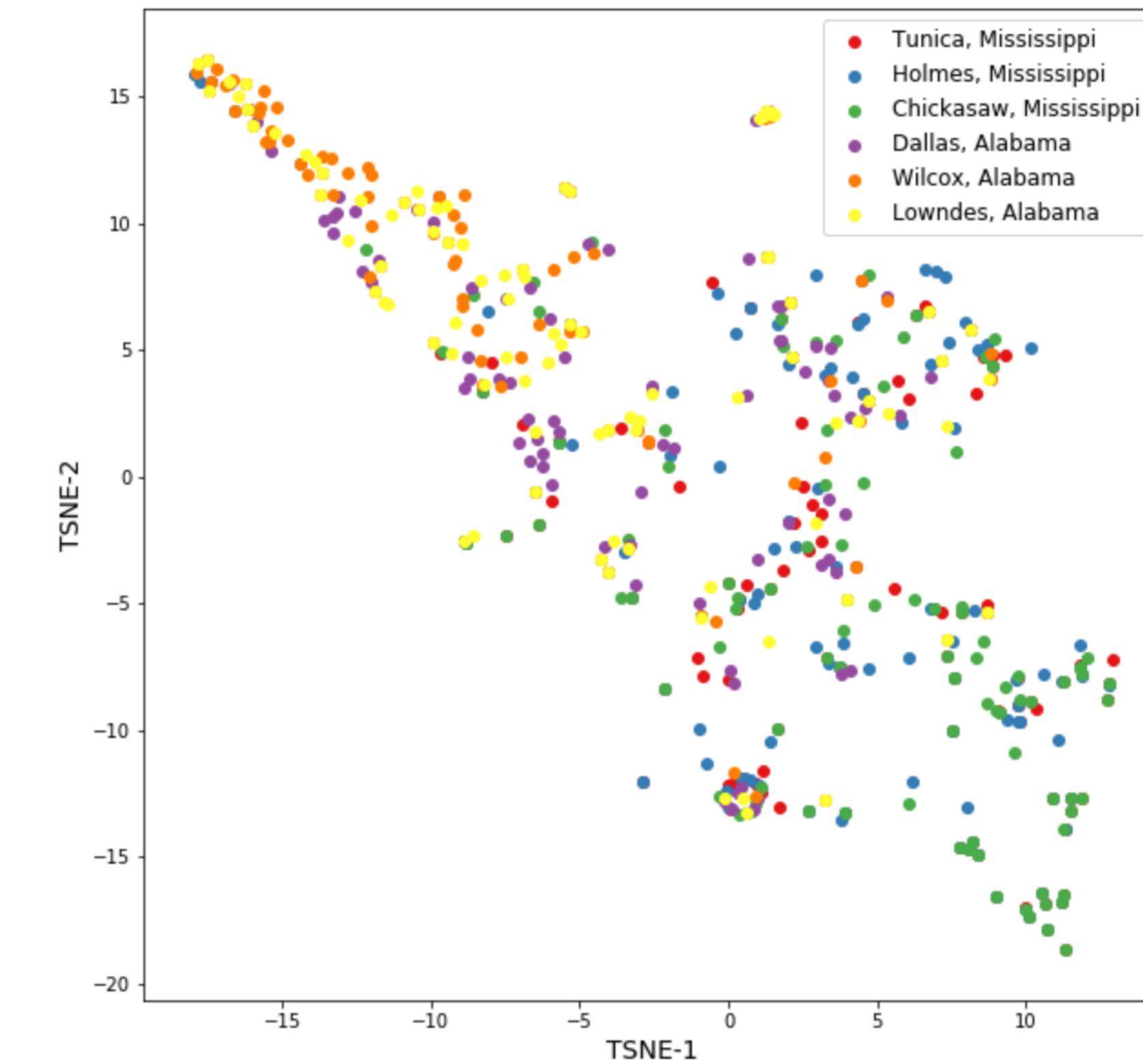


## 5B: Investigating the County through policy & media

PCA analysis (Scikit-learn) for all the articles, plotting the first two components on different time-frame



TSNE as well on the PCA components





# Final Project

## Data Analytics Bootcamp 2021