# Unsupervised Machine Learning - Final Project
## Working Group

Hanning Luo, Nikki Ting, Naiyu Jiang

May 11, 2021

# Overview

1. Introduction

2. Preliminary Results

# Project Objectives

- While various studies examine networks of social and political communication by exploring the content and link of users' tweets, few projects directly point to the tweeting habits of the American public.

- In order to understand the activity of U.S. Twitter users, we employ the 2018 Twitter Survey data co-conducted by the Pew Research Center and the Ipsos Public Affairs to further explore the underlying patterns on social activities and political polarization.

- Based on existing observations, our project aims to explore
  - 1 ) the social behavior and communication habits of American Twitter users
  - 2 ) Twitter's role in contributing to partisan polarization.

# Methods

- Exploratory Data Analysis
- Association Rules
- Dimension Reduction + Clustering

# Sample Survey Questions of Our Interest

**For social trust questions:**

SOCTRUST2: Generally speaking, would you say that. . .

1 Most people can be trusted
2 Most people cannot be trusted

GSSTRUST2: Do you think most people...

1 Would try to take advantage of you if they got a chance
2 Would try to be fair no matter what

TRUSTCONGe: In your view, how much of the time do MEMBERS OF CONGRESS admit mistakes and take responsibility for them

1 Almost all the time
2 Some of the time
3 Only a little of the time
4 None of the time

# Sample Survey Questions of Our Interest

**For political activity questions:**

NATPROBS: How much of a problem do you think each of the following are in the country today?

The affordability of health care:
1 A very big problem; 2 A moderately big problem; 3 A small problem

FAIRTRT: Overall, in our country today, would you say that. . .

1 Blacks are treated less fairly than whites
2 Whites are treated less fairly than blacks
3 Both are treated about equally

POL1DT: Do you approve or disapprove of the way Donald Trump is handling his job as President?

1 Approve
2 Disapprove

# Sample Survey Questions of Our Interest

**Measurement of polarization:**

REPANTIP (for Republicans): Thinking for a moment about people who consider themselves Democrats, which of these comes closer to your views about them?

1 They feel differently than I do about politics, but they probably share many of my other values and goals (less polarized)
2 They feel differently than I do about politics, and they probably DON'T share many of my other values and goals, either (highly polarized)

DEMANTIP (for Democrats): Thinking for a moment about people who consider themselves Republicans, which of these comes closer to your views about them?

1 They feel differently than I do about politics, but they probably share many of my other values and goals (less polarized)
2 They feel differently than I do about politics, and they probably DON'T share many of my other values and goals, either (highly polarized)
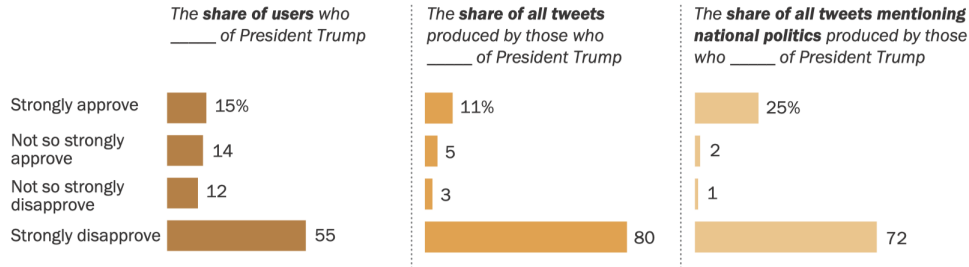
# Sample Survey Questions of Our Interest

```r
survey <- dataset %>%
  select(SOCTRUST2, COMATTACH, GSSTRUST2, GSSTRUST3, # social trust questions (group)
         TRUSTCONGa, TRUSTCONGb, TRUSTCONGc, TRUSTCONGd, TRUSTCONGe,  # legislator responsiveness (group)
         TWITTER_USE, POLTWEET, # twitter usage
         POL1DT, FAIRTRT, WOMENOPPS, IMMCULT2, ECONFAIR2, POLCRCT,
         NATPROBSa, NATPROBSb, NATPROBSc, NATPROBSd, NATPROBSe, NATPROBSf,
         NATPROBSg, NATPROBSh, NATPROBSi, NATPROBSj, # attitudes towards various issues (group)
         THERMOa, THERMOb, THERMOc, THERMOd, THERMOe, THERMOf, THERMOg, THERMOh, # feeling thermometer (group)
         QBELIEF3, QBELIEF4, TWKNOW, # friends on twitter
         PARTY, REPANTIP, DEMANTIP, # party affiliation (layer)
         CIVIC_ENG_ACTYRa, CIVIC_ENG_ACTYRb, CIVIC_ENG_ACTYRc, # physical activities
         RELIMP, # religion identity (layer)
         ) %>%
```

**Those who strongly approve or disapprove of Donald Trump generate majority of public tweets from U.S. adults that mention national politics**

*Among U.S. adults with public Twitter accounts ...*

| | *The **share of users** who _____ of President Trump* | *The **share of all tweets** produced by those who _____ of President Trump* | *The **share of all tweets mentioning national politics** produced by those who _____ of President Trump* |
|---|---|---|---|
| Strongly approve | 15% | 11% | 25% |
| Not so strongly approve | 14 | 5 | 2 |
| Not so strongly disapprove | 12 | 3 | 1 |
| Strongly disapprove | 55 | 80 | 72 |

Note: Tweets about national politics include those that reference national politicians, political groups or institutions, or political behaviors such as voting.
Source: Survey of 2,427 U.S. adult Twitter users with public accounts conducted Nov. 21-Dec. 17, 2018. Tweets collected via Twitter API, June 10, 2018-June 9, 2019.
"National Politics on Twitter: Small Share of U.S. Adults Produce Majority of Tweets"

**PEW RESEARCH CENTER**

# Social Trust and Polarization: Associational Rules

```{r}
trust <- subset(trust_rules,
                lhs %in% "high trust" &
                rhs %in% c("high polarization", "low polarization") &
                lift > 1.25); inspect(sort(trust[1:5], by = "confidence"))
```

| | lhs<br><chr> | | rhs<br><chr> | support<br><dbl> | confidence<br><dbl> |
|---|---|---|---|---|---|
| [1] | {high trust,never tweet,unfair society} | => | {low polarization} | 0.1311358 | 0.6245734 |
| [2] | {high trust,mixed politics followed,never tweet} | => | {low polarization} | 0.1608742 | 0.6201657 |
| [3] | {high trust,never tweet,social assistance} | => | {low polarization} | 0.1372268 | 0.6177419 |
| [4] | {high trust,mixed politics followed,other affiliations} | => | {low polarization} | 0.1099964 | 0.6055227 |
| [5] | {high attachment,high trust,never tweet} | => | {low polarization} | 0.1375851 | 0.6018809 |

5 rows | 1-6 of 8 columns

```{r}
attachment <- subset(trust_rules,
                     lhs %in% "high attachment" &
                     rhs %in% c("high polarization", "low polarization") &
                     lift > 1.25); inspect(sort(attachment[1:5], by = "confidence"))
```

```
    lhs                       rhs                    support confidence coverage     lift count
[1] {high attachment,
     high trust,
     never tweet,
     social assistance}    => {low polarization} 0.1003225  0.6263982 0.1601576 1.301770  280
[2] {high attachment,
     high trust,
     mixed politics followed,
     never tweet}          => {low polarization} 0.1128628  0.6262425 0.1802221 1.301447  315
[3] {high attachment,
     never tweet,
     social assistance}    => {low polarization} 0.1103547  0.6135458 0.1798638 1.275061  308
[4] {high attachment,
     high trust,
     never tweet,
     no rally}             => {low polarization} 0.1218201  0.6082290 0.2002866 1.264011  340
[5] {high attachment,
     high trust,
     never tweet}          => {low polarization} 0.1375851  0.6018809 0.2285919 1.250819  384
```

# Twitter Use and Polarization: Associational Rules

```{r}
twitter_use <- subset(social_rules,
                      rhs %in% c("low polarization") &
                      lhs %in% c("never tweet", "frequent tweet", "infrequent use", "frequent use") &
                      lift > 1.2); inspect(sort(twitter_use[1:5], by = "confidence"))
```

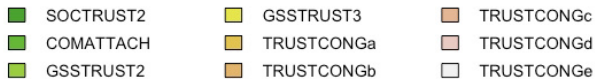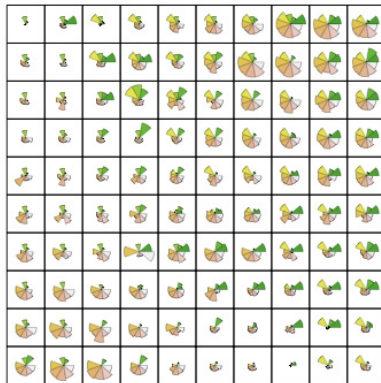|  | lhs <chr> | | rhs <chr> | support <dbl> | confidence <dbl> |
|---|---|---|---|---|---|
| [1] | {healthcare concern,never tweet,other affiliations} | => | {low polarization} | 0.1092798 | 0.5980392 |
| [2] | {drug concern,never tweet,other affiliations} | => | {low polarization} | 0.1049803 | 0.5967413 |
| [3] | {college cost concern,never tweet,other affiliations} | => | {low polarization} | 0.1028305 | 0.5954357 |
| [4] | {never tweet,other affiliations} | => | {low polarization} | 0.1142960 | 0.5863971 |
| [5] | {never tweet,not xenophobic} | => | {low polarization} | 0.1644572 | 0.5802781 |

5 rows | 1-6 of 8 columns

```r
trump_1 <- subset(social_rules,
            rhs %in% c("offensive", "defensive") &
            lhs %in% "Trump adversary" &
            lift > 1.5); inspect(sort(trump_1[1:5], by = "confidence"))
```

|     | lhs<br><chr> | | rhs<br><chr> | support<br><dbl> | confidence<br><dbl> |
|-----|------|---|-----|-----|-----|
| [1] | {frequent tweet,not xenophobic,Trump adversary} | => | {defensive} | 0.1017556 | 0.6810552 |
| [2] | {frequent tweet,Trump adversary,unfair for black} | => | {defensive} | 0.1006808 | 0.6787440 |
| [3] | {frequent tweet,immigrants treatment concern,Trump adversary} | => | {defensive} | 0.1006808 | 0.6738608 |
| [4] | {frequent tweet,Trump adversary} | => | {defensive} | 0.1042637 | 0.6583710 |
| [5] | {Democrat,Trump adversary} | => | {defensive} | 0.2311000 | 0.6096408 |

5 rows | 1-6 of 8 columns

```r
trump_2 <- subset(social_rules,
            rhs %in% c("offensive", "defensive") &
            lhs %in% "Trump supporter" &
            lift > 1.5); inspect(sort(trump_2[1:5], by = "confidence"))
```

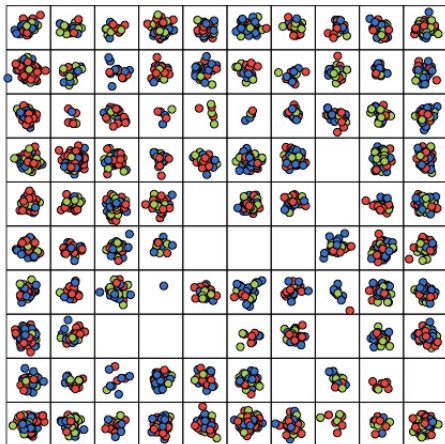|     | lhs<br><chr> | | rhs<br><chr> | support<br><dbl> | confidence<br><dbl> | coverage<br><dbl> |
|-----|------|---|-----|-----|-----|-----|
| [1] | {racism apathy,Trump supporter} | => | {offensive} | 0.1343604 | 0.9101942 | 0.1476173 |
| [2] | {gun apathy,Trump supporter} | => | {offensive} | 0.1246865 | 0.8969072 | 0.1390183 |
| [3] | {immigrants treatment apathy,Trump supporter} | => | {offensive} | 0.1866714 | 0.8967298 | 0.2081691 |
| [4] | {income gap apathy,Trump supporter} | => | {offensive} | 0.1343604 | 0.8949881 | 0.1501254 |
| [5] | {sexism apathy,Trump supporter} | => | {offensive} | 0.1780724 | 0.8827709 | 0.2017198 |

5 rows | 1-7 of 8 columns

# Social Trust: Self-Organizing Map

**Codes plot**



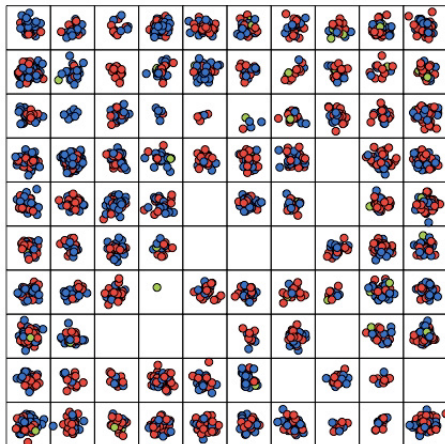| | | |
|---|---|---|
| ■ SOCTRUST2 | ■ GSSTRUST3 | ■ TRUSTCONGc |
| ■ COMATTACH | ■ TRUSTCONGa | ■ TRUSTCONGd |
| ■ GSSTRUST2 | ■ TRUSTCONGb | □ TRUSTCONGe |

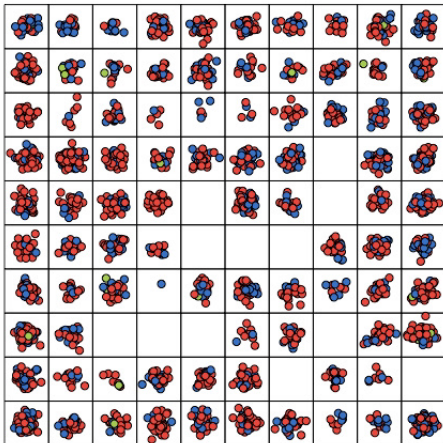# Social Trust: Self-Organizing Map



Coloring by Party

Coloring by Polarization

# Social Trust: Self-Organizing Map



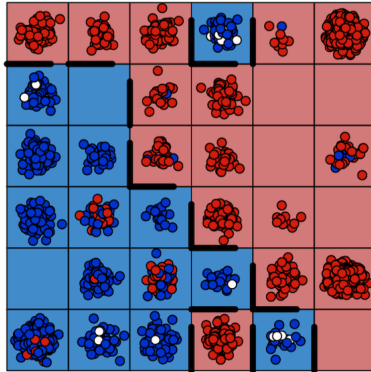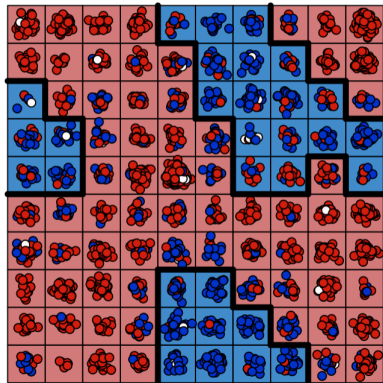**Coloring by Trump supporter or adversary**

**Coloring by Twitter Usage**

**Attitudes towards the equality and diversity of communities, coloring by whether people approve of the way Donald Trump is handling his job as president**
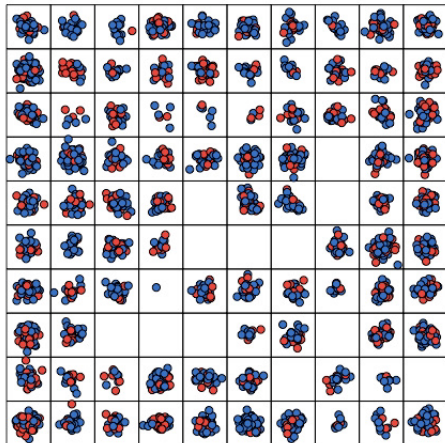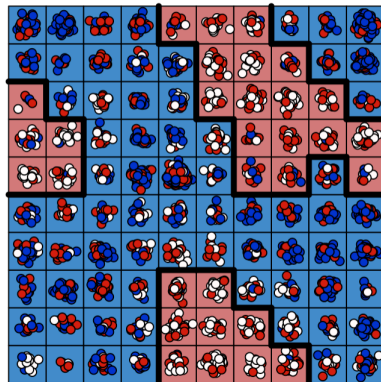
# Political Activities: Self-Organizing Map

**Attitudes towards salient social issues**
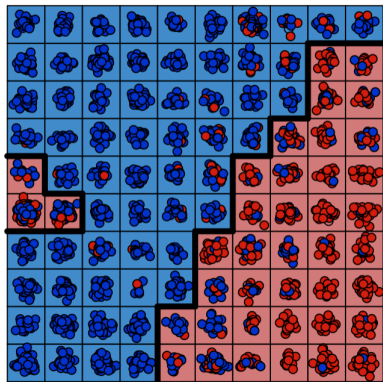


(a) coloring by Trump

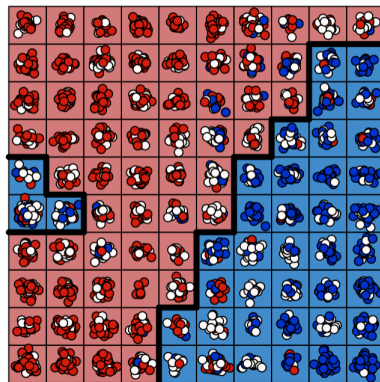**Attitudes towards salient social issues**



(b) coloring by party

# Political Activities: Self-Organizing Map



feeling thermometer by Trump

(c) coloring by Trump



feeling thermometer by party

(d) coloring by party

Q&A Session