

CCC Data Summary

Asa Gold

4/30/2022

Basic Summary

Take a quick look at the Crowd Counting Consortium's GitHub page and their website.

The Crowd Counting Consortium (CCC), updated weekly, currently contains ~100k observations between 2017-01-20 and 2022-12-10.

Below is a glimpse at all 69 variables.

```
## Rows: 100,836
## Columns: 69
## $ date          <date> 2017-01-20, 2017-01-21, 2017-01-21, 2017-01-21, ~
## $ locality      <chr> "Seattle", "Adak", "Anchorage", "Bethel", "Cordov~
## $ state         <chr> "WA", "AK", "AK", "AK", "AK", "AK", "AK", "AK", "~
## $ location_detail <chr> "University of Washington campus", NA, NA, NA, NA~
## $ online        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ type          <chr> "protest; demonstration", "march", "march", "marc~
## $ title         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ macroevent    <chr> NA, "20170121-womensmarch", "20170121-womensmarch~
## $ actors        <chr> "general protestors", NA, NA, NA, NA, NA, NA, NA,~
## $ organizations <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ participants  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ claims        <chr> "anti-Yiannopolous; anti-Breitbart; anti-Trump", ~
## $ valence       <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ issues        <chr> "executive", "women's rights", "women's rights", ~
## $ size_text     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ size_low      <int> 200, 10, 2000, 40, 108, 2000, 105, 150, 900, 1000~
## $ size_high     <int> 700, 10, 4000, 100, 114, 2000, 105, 170, 900, 100~
## $ size_mean     <dbl> 450, 10, 3000, 70, 111, 2000, 105, 160, 900, 1000~
## $ size_cat      <int> 2, 1, 3, 1, 2, 3, 2, 2, 2, 3, 2, 2, 1, 1, 2, 1, 1~
## $ arrests       <chr> "1", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ arrests_any   <int> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ injuries_crowd <chr> "1", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ injuries_crowd_any <int> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ injuries_police <chr> "0", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ injuries_police_any <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ property_damage <chr> "1", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ property_damage_any <int> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ chemical_agents <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ participant_measures <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ police_measures <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ participant_deaths <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

```

## $ police_deaths      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_1          <chr> "http://infoweb.newsbank.com.du.idm.oclc.org/reso~
## $ source_2          <chr> "http://www.seattletimes.com/seattle-news/northwe~
## $ source_3          <chr> "http://www.seattletimes.com/seattle-news/educati~
## $ source_4          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_5          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_6          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_7          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_8          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_9          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_10         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_11         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_12         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_13         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_14         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_15         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_16         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_17         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_18         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_19         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_20         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_21         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_22         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_23         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_24         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_25         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_26         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_27         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_28         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_29         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ source_30         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ notes             <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ lat               <dbl> 47.60621, 51.88000, 61.21806, 60.79222, 60.54241, ~
## $ lon               <dbl> -122.3321, -176.6581, -149.9003, -161.7558, -145.~
## $ resolved_locality <chr> "Seattle", "Adak", "Anchorage", "Bethel", "Cordov~
## $ resolved_county   <chr> "King County", "Aleutians West", "Anchorage", "Be~
## $ resolved_state     <chr> "WA", "AK", "AK", "AK", "AK", "AK", "AK", "AK", "~
## $ fips_code         <int> 53033, 2016, 2020, 2050, 2261, 2090, 2105, 2100, ~

```

Several are particularly relevant.

- Descriptive Variables

- **location_detail**: non-standardized descriptions of the place of the protest. Ex: Downtown Dayton, OSU Campus, Comcast Company Headquarters; City Hall
- **actors**: Who engaged in the protest. Slightly more standardized than location_detail but still highly variable (25252 unique values). Ex: MAGA March, Counter-protesters; socialist Party USA, Democratic Club of the High Desert, Colorado MAGA March
- **claims**: main grievances, goals, or rallying cries of the protestors. Ex: support LGBTQ, Pro Trump, oppose anti-Semitism
- **valance**: political leaning of the protest; 0 = neutral; 1 = left; 2 = right.

- Location Variables

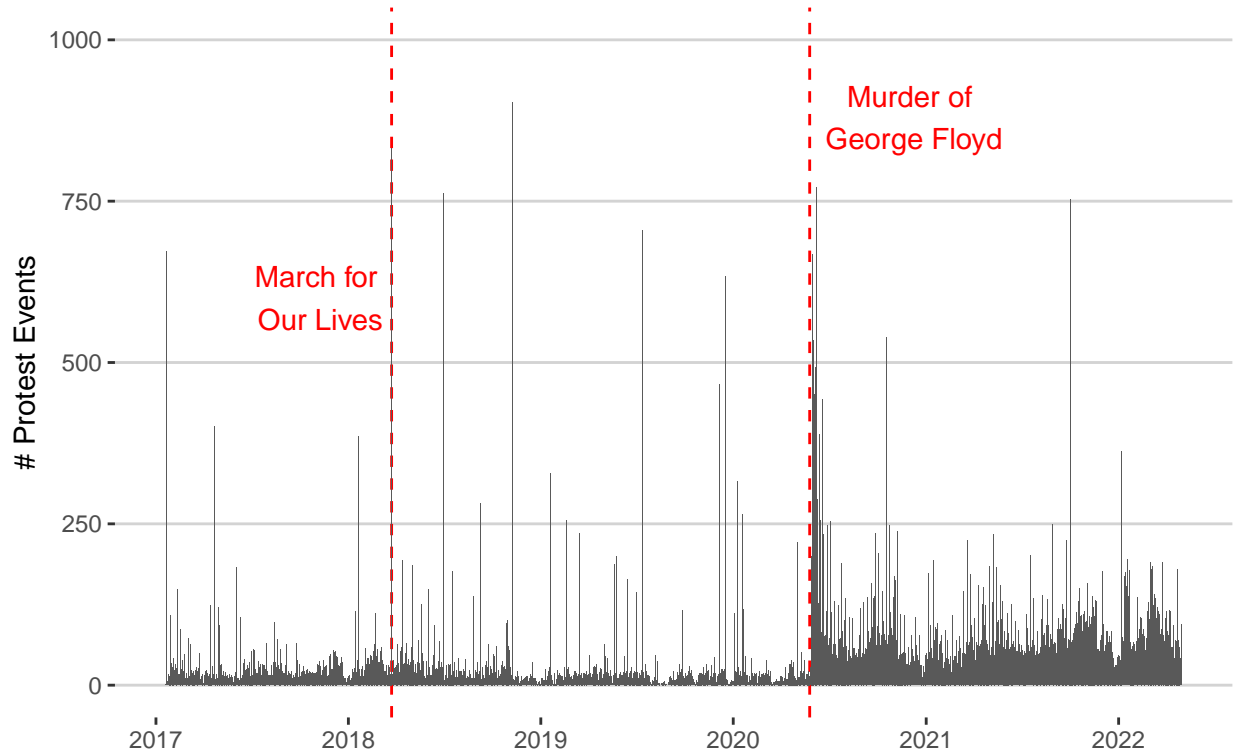
- **locality**: normally a city in or around which the protest event took place. Ex: Lancaster, Meadville, Philadelphia, Pittsburgh.

– **fips_code**: county code associated with the locality

It is important to note that the relative frequency of protest events have ebbed and flowed in response to major nation-wide events. The plot below reports daily count of protest events between Jan 2017 and present. Note the spikes during and directly after several headline-making national events, like March for Our Lives (gun control) and George Floyd’s murder (racism and police violence). This plot does not indicate the relative *size* of the protests; for more on size estimates, see **size__[low/mean/high]** variables

CCC Protests Over Time

Excluding several March for Our Lives days with >3000 events

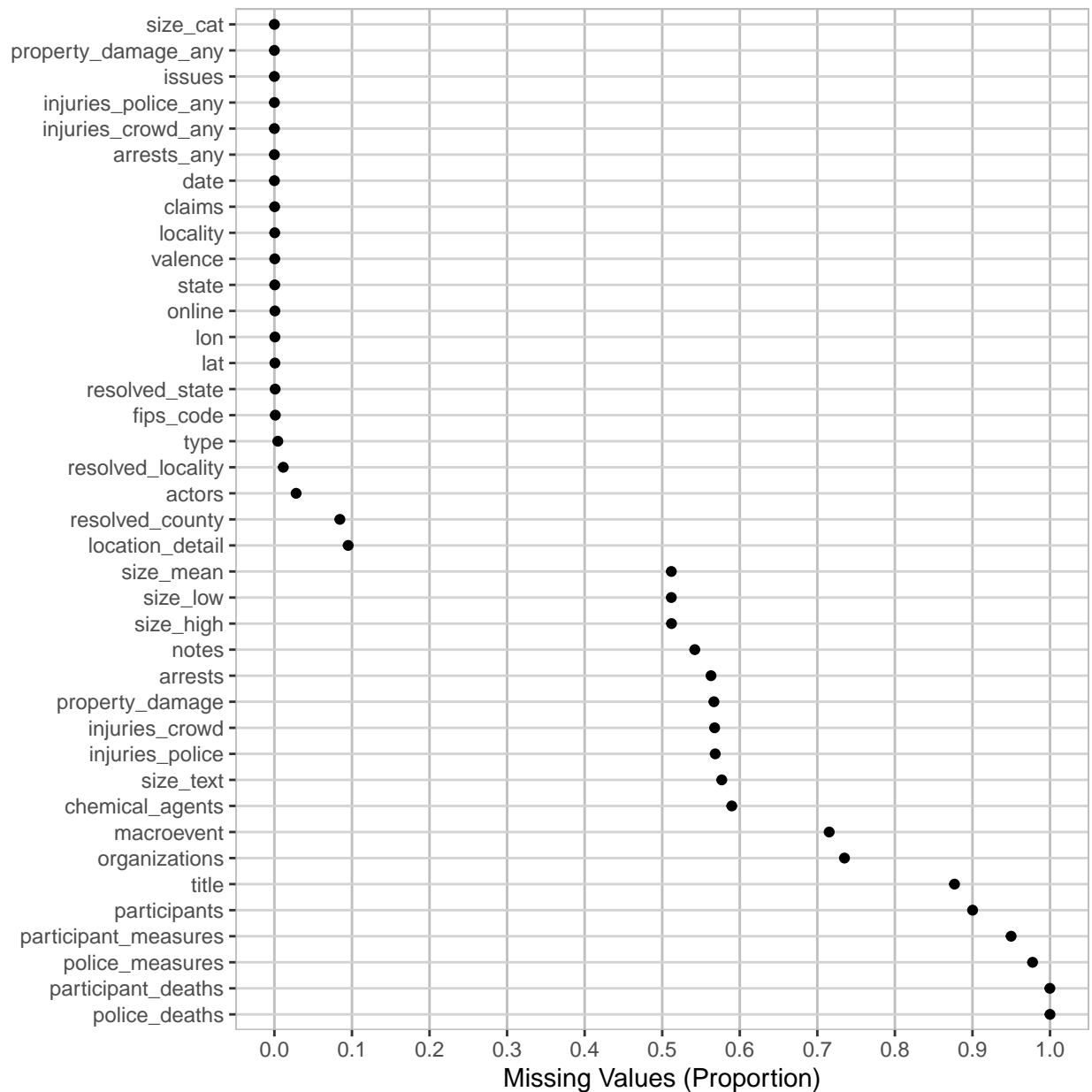


Missing Values

Below are proportions of missing values for all variables, except the *source_n* variables, which are largely missing.

CCC Missing Values

Not shown: source_n (mostly missing)



Filtering

To restrict our dataset to only include protests occurring on college or university campuses or organized by college students, we use **location_detail** and **actors**. The code below filters the master dataset such that only those observations that reference university or college in *either* **location_detail** or **actors**

```
# NOTE: 'ccc' is the master dataset
college_filtered <- ccc |>
```

```

filter(
  str_detect(actors,
    regex('(universit|college)', ignore_case = TRUE)
  ) |
  str_detect(location_detail,
    regex('(universit|college)', ignore_case = TRUE))
) |>

filter(date <= Sys.Date())

```

With this initial, naive filtering, we are left with 5159 observations. Below, inspect several `location_detail` and `actors` observations.

```

##                                location_detail
## 1                University of Washington campus
## 2                Dartmouth College
## 3                Dartmouth College
## 4      University of Colorado; Watershed School
## 5 Oklahoma State University; outside of President Hargis' office
## 6                University of Colorado
##
##                                actors
## 1      general protestors
## 2                students
## 3 mostly veterans; community members
## 4                students
## 5 African American Student Alliance
## 6      mostly students

```

While this initial filtering is highly effective, manual filtering may still be necessary. For instance, one March for Our Lives protest occurred near “College Lakes Recreation Center.” Such observations cannot be excluded by naive filtering alone.

The below plot indicates daily protest occurrence in college or university campuses between Jan 2017 and present.

Protests Over Time: College/University Only

Excluding several March for Our Lives days with >50 events

