

# Difficult dialogues: communicating data analyses effectively

Jo Hardin, Pomona College

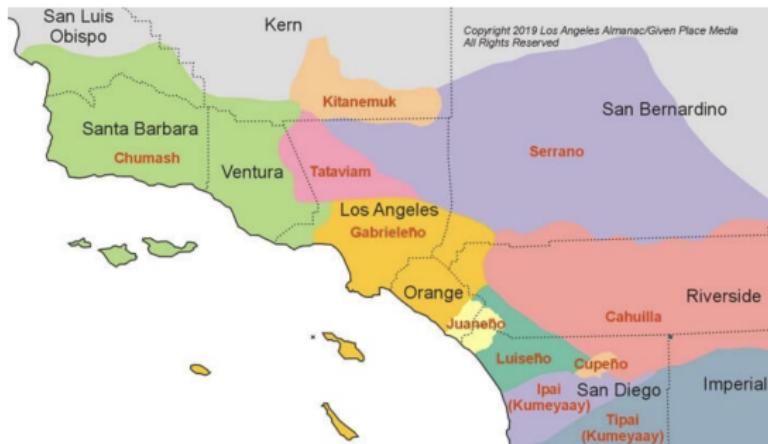
8.5.2020



# Land Acknowledgement

*With our deepest respects to the Tongva and Serrano Peoples, past, present, and emerging.*

## Original People of Los Angeles County



Map of territories of Original Peoples with county boundaries in Southern California, Los Angeles Almanac, 2019.

Information sources: *Handbook of North American Indians, Vol. 8, California*, William C. Sturtevant (Gen. Editor) & Robert F. Heizer (Vol. Editor), 1978, Smithsonian Institute, and Dr. E. Gary Stickel, Ph.D. (UCLA), Tribal Archeologist, Kizh Nation / Gabrieleño Band of Mission Indians.

# Stanford Policing Data

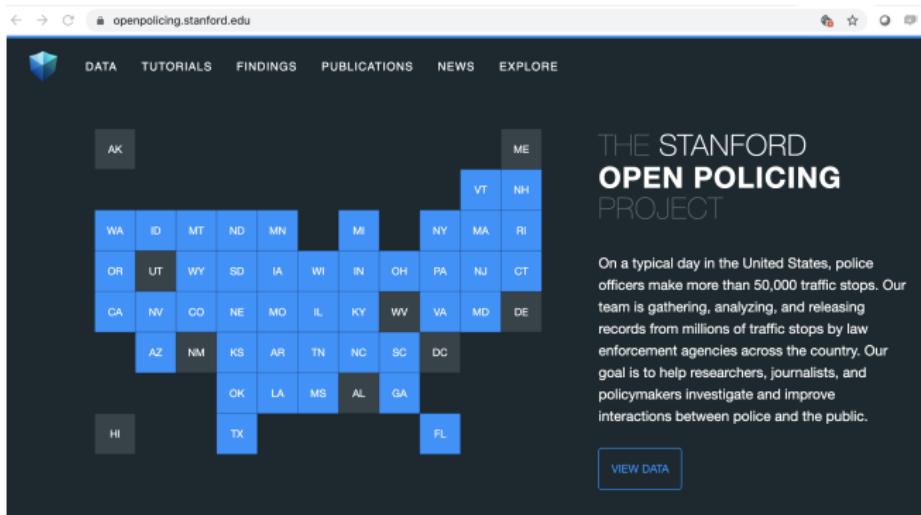


Figure 1: <https://openpolicing.stanford.edu/>

# Stanford Policing Data

The screenshot shows a web browser displaying the 'openpolicing.stanford.edu/data/' page. The top navigation bar includes links for DATA, TUTORIALS, FINDINGS, PUBLICATIONS, NEWS, and EXPLORE. Below the navigation is a search bar and a dropdown menu. The main content area is a grid of 11 rows, each representing a dataset for a different location in California. The columns represent various data points: State, Download, Stops, Time Range, Stop Date, Stop Time, Stop Location, Driver Race, Driver Sex, Driver Age, Search Conducted, Contraband Found, Citation Issued, Warning Issued, Frisk Performed, Arrest Made, Reason for Stop, and Violation.

State	Download	Stops	Time Range	Stop Date	Stop Time	Stop Location	Driver Race	Driver Sex	Driver Age	Search Conducted	Contraband Found	Citation Issued	Warning Issued	Frisk Performed	Arrest Made	Reason for Stop	Violation
CA																	
Anaheim		87,876	Dec 2011 - Mar 2017														
Bakersfield		189,685	Mar 2008 - Mar 2018														
Long Beach		365,924 <sup>2</sup>	Dec 2007 - Dec 2017														
Los Angeles		5,418,400 <sup>2</sup>	Dec 2009 - Jun 2018														
Oakland		133,405 <sup>2</sup>	Mar 2013 - Dec 2017														
San Bernardino		90,523	Dec 2011 - Sep 2017														
San Diego		382,844	Dec 2013 - Mar 2017														
San Francisco		905,070	Dec 2006 - Jun 2016														
San Jose		152,833 <sup>2</sup>	Aug 2013 - Mar 2018														
Santa Ana		46,268 <sup>2</sup>	Jun 2014 - Apr 2018														
State Patrol		31,778,515	Jun 2009 - Jun 2016														
Stockton		41,629	Dec 2011 - Dec 2016														

Figure 2: Just under 100 datasets.  
<https://openpolicing.stanford.edu/data/>

## Why this project?

- ▶ Engaging questions (for me and for students)
- ▶ Goldilocks level of data wrangling
- ▶ Each student can work with a different dataset
- ▶ Ability (need!) to work with SQL

## Step 1: get data

```
con <- dbConnect(  
  MySQL(), host = "XXX", user = "XXX",  
  password = "XXX", dbname = "XXX")  
  
raleigh_df <- DBI::dbGetQuery(con, "SELECT * FROM NCraleigh")
```

## Step 2: data viz

```
raleigh_df %>%  
  
  # remove missing data  
  filter(!is.na(sex) & !is.na(race)) %>%  
  
  # use group_by and summarize to count number of stops per  
  group_by(sex, race) %>%  
  summarize(count = n()) %>%  
  ungroup() %>%  
  
  # find the percentage of age/race stops  
  mutate(percentage = round(prop.table(count), digits = 2))
```

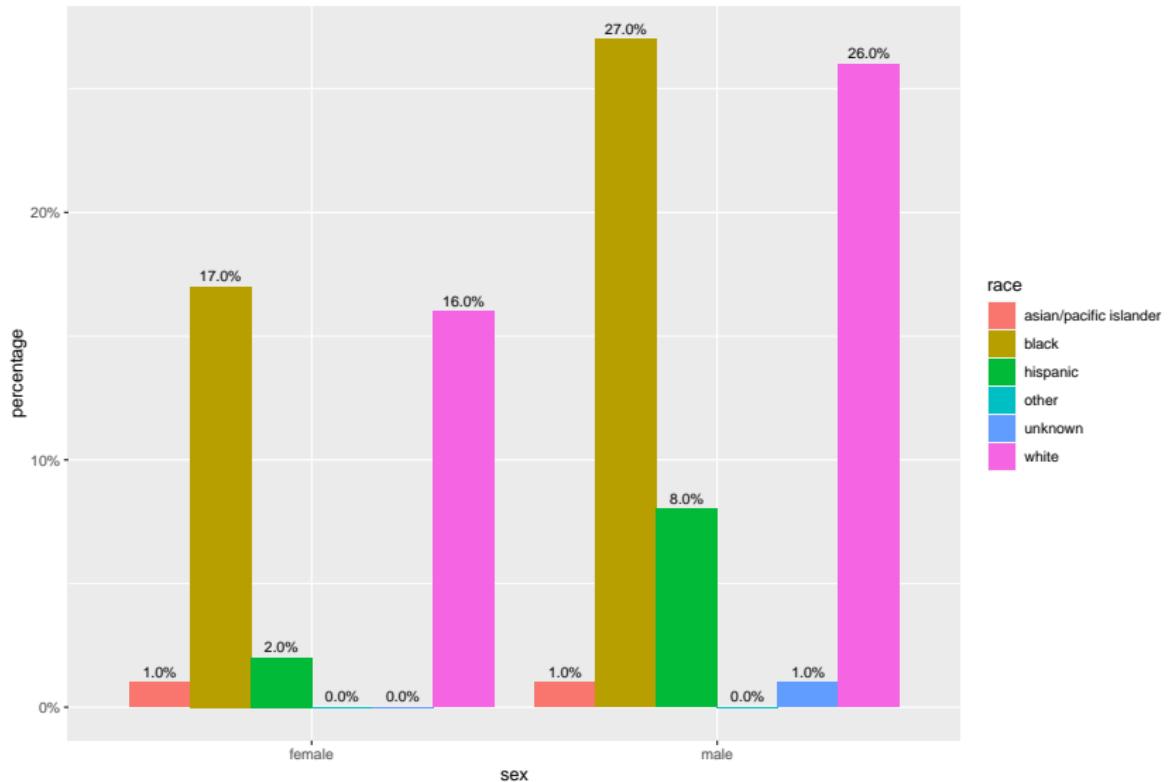
## Step 2: data viz

... continued

```
# plot percentages
ggplot(mapping = aes(x = sex, y = percentage,
                      fill = race,
                      label = scales::percent(percentage)))
  geom_bar(position = "dodge", stat = "identity") +
  
# adjust labels
  geom_text(position = position_dodge(width = .9),
            vjust = -0.5,
            size = 3) +
  scale_y_continuous(labels = scales::percent) +
  
# provide labels
  ggtitle("Race & gender breakdown, % out of total")
```

## Step 2: data viz

Race & gender breakdown, % out of total



## Step 2: advanced data viz

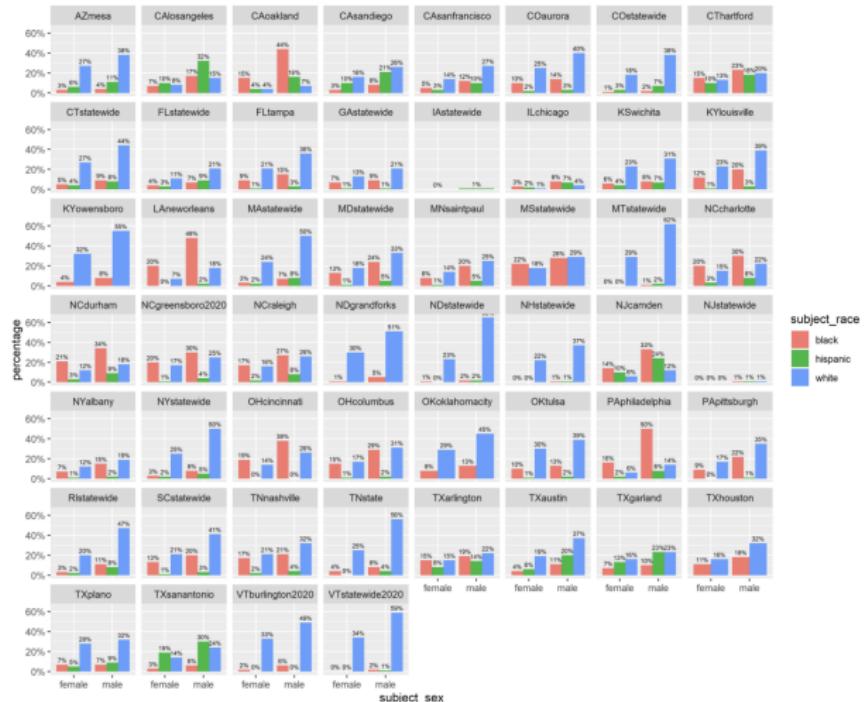


Figure 3: Accessing, joining, and wrangling all the datasets.

## Step 2: advanced data viz



Figure 4: With facet\_geo

## Step 3: modeling search

n.b., **all** the observations were traffic stops, so we can't model demographics of who was pulled over (model search instead).

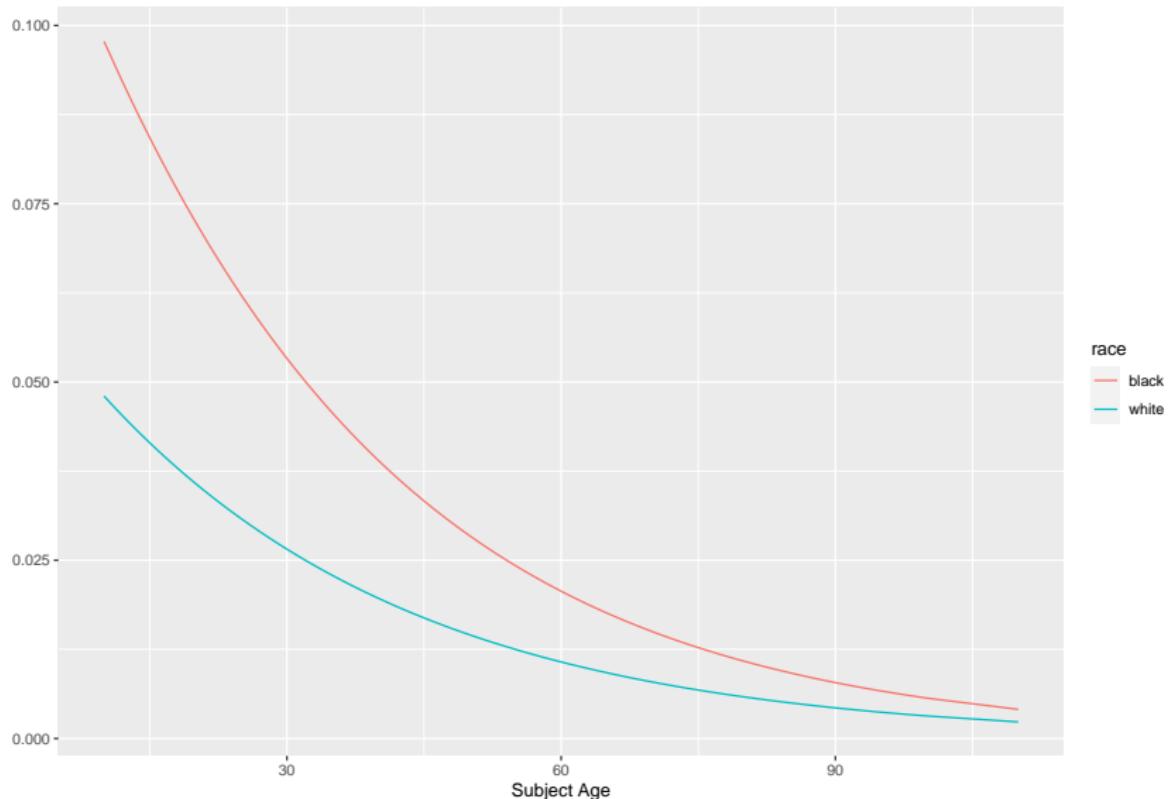
```
raleigh_search <- glm(formula = search ~ age * race,  
family = "binomial", data = raleigh_df,  
subset = (race %in% c("black", "white")))
```

```
raleigh_search %>% tidy()
```

```
## # A tibble: 4 x 5  
##   term            estimate std.error statistic p.value  
##   <chr>          <dbl>     <dbl>      <dbl>    <dbl>  
## 1 (Intercept) -1.90      0.0240     -78.9    0.  
## 2 age         -0.0327    0.000759     -43.1    0.  
## 3 racewhite   -0.785     0.0402     -19.5  5.39e-85  
## 4 age:racewhite 0.00200   0.00125      1.60 1.10e- 1
```

## Step 3: data viz of model

Probability of being searched from logistic model, broken down by race and age.



# Language around sensitive data

Table 2.1 From data ethics to data justice	
<b>Concepts That Secure Power</b> Because they locate the source of the problem in individuals or technical systems	<b>Concepts That Challenge Power</b> Because they acknowledge structural power differentials and work toward dismantling them
Ethics	Justice
Bias	Oppression
Fairness	Equity
Accountability	Co-liberation
Transparency	Reflexivity
Understanding algorithms	Understanding history, culture, and context

Figure 5: **Data Feminism** <https://datafeminism.io/>

# Equity vs Equality

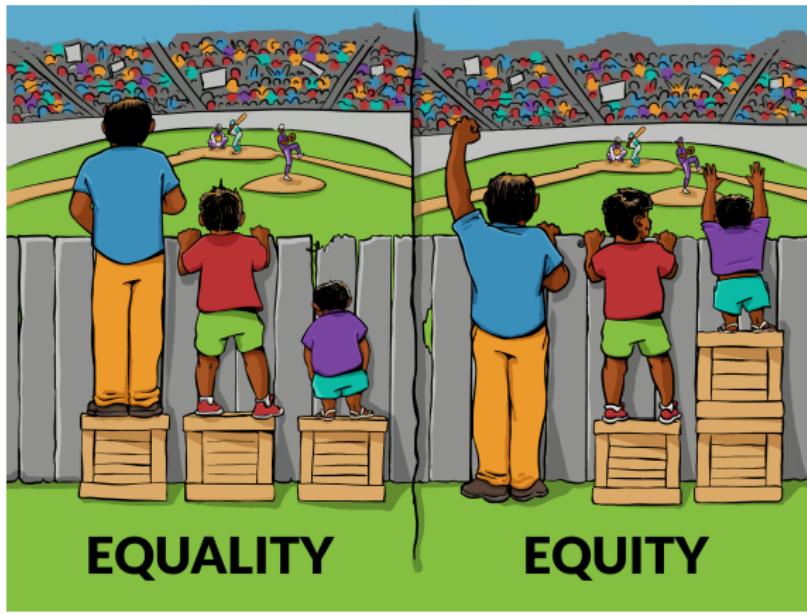


Figure 6: Image credit: Interaction Institute for Social Change | Artist: Angus Maguire.

## Traffic stops

What are the systems that have created structural racial discrepancies in the US?

- ▶ Redlining: systematic exclusion of people of color from obtaining mortgages (Federal lending programs)
- ▶ Home ownership is lower for people of color
- ▶ Housing discrimination continues today and creates racial differences in where people **live** and **work**.

## Traffic stops

Pierson et al. "A large-scale analysis of racial disparities in police stops across the United States" (Nature Human Behaviour, 2020)

*Our results indicate that police stops and search decisions suffer from persistent racial bias and point to the value of policy interventions to mitigate these disparities.*

While the research shows that stops are neither equal nor equitable, a discussion on equity belongs in the conversation around racial discrepancies in traffic stops.

## 8am weekly zoom calls

The undergraduates who did the complete analysis and wrote the report are Pomona College students: Amber Lee ('22), Arm Wonghirundacha ('22), Emma Godfrey ('21), Ethan Ong ('21), Ivy Yuan ('21), Oliver Chang ('22), and Will Gray ('22).



Thank you!

Jo Hardin

[jo.hardin@pomona.edu](mailto:jo.hardin@pomona.edu)

[@jo\\_hardin47](https://twitter.com/jo_hardin47) 

<https://github.com/hardin47>



<http://research.pomona.edu/johardin/>