Difficult dialogues:

communicating data analyses effectively

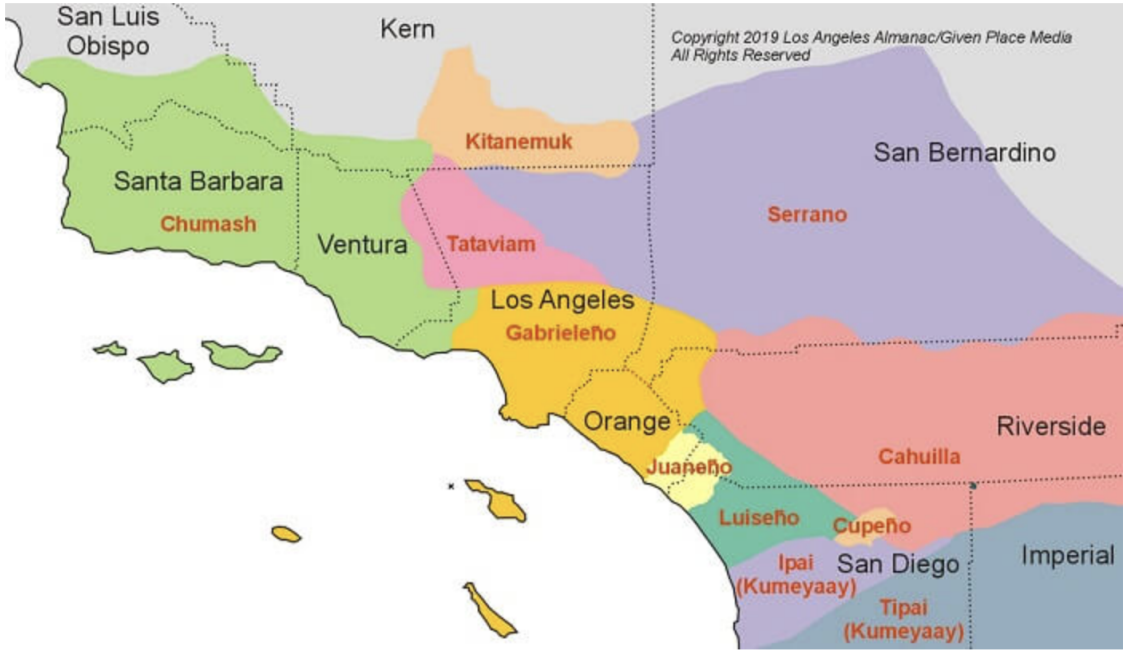
Jo Hardin, Pomona College

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 Orignal 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 / Gabrieleno Band of Mission Indians.

Stanford Policing Data

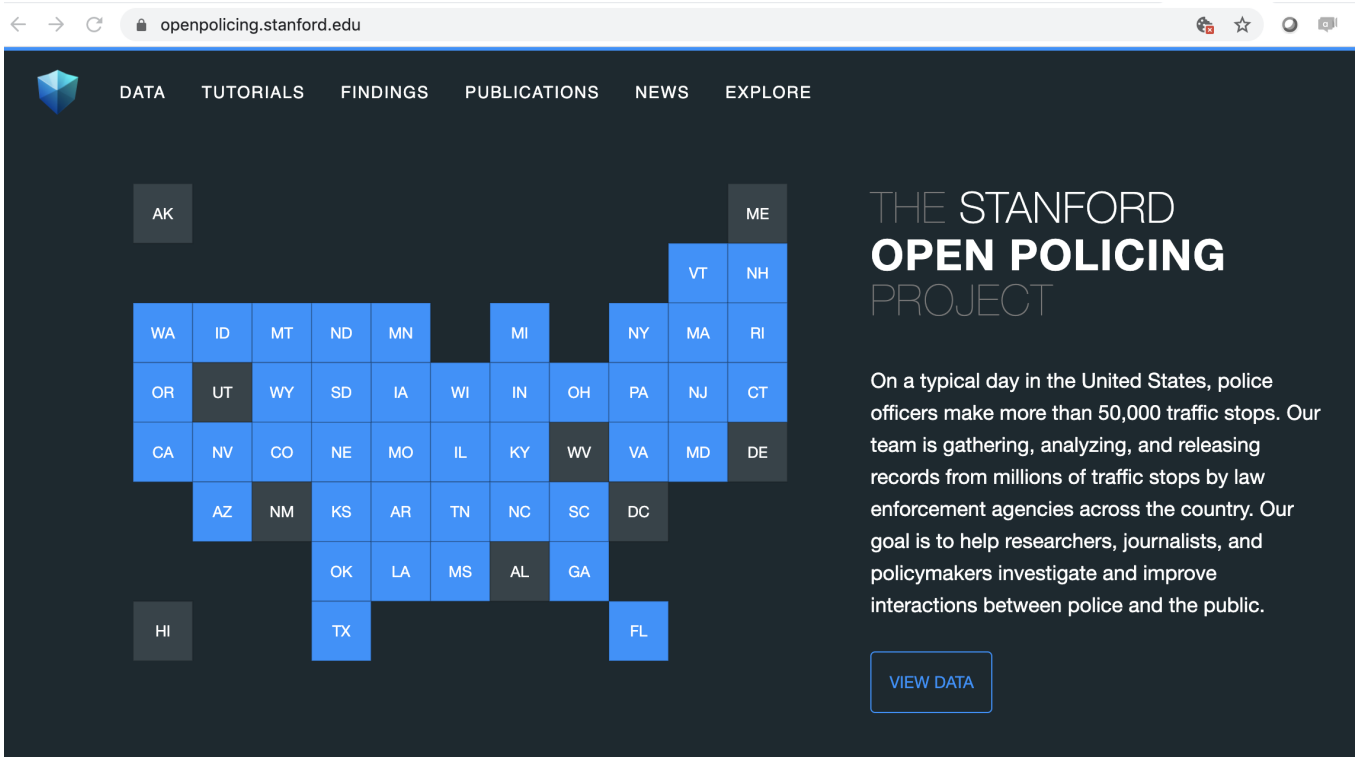
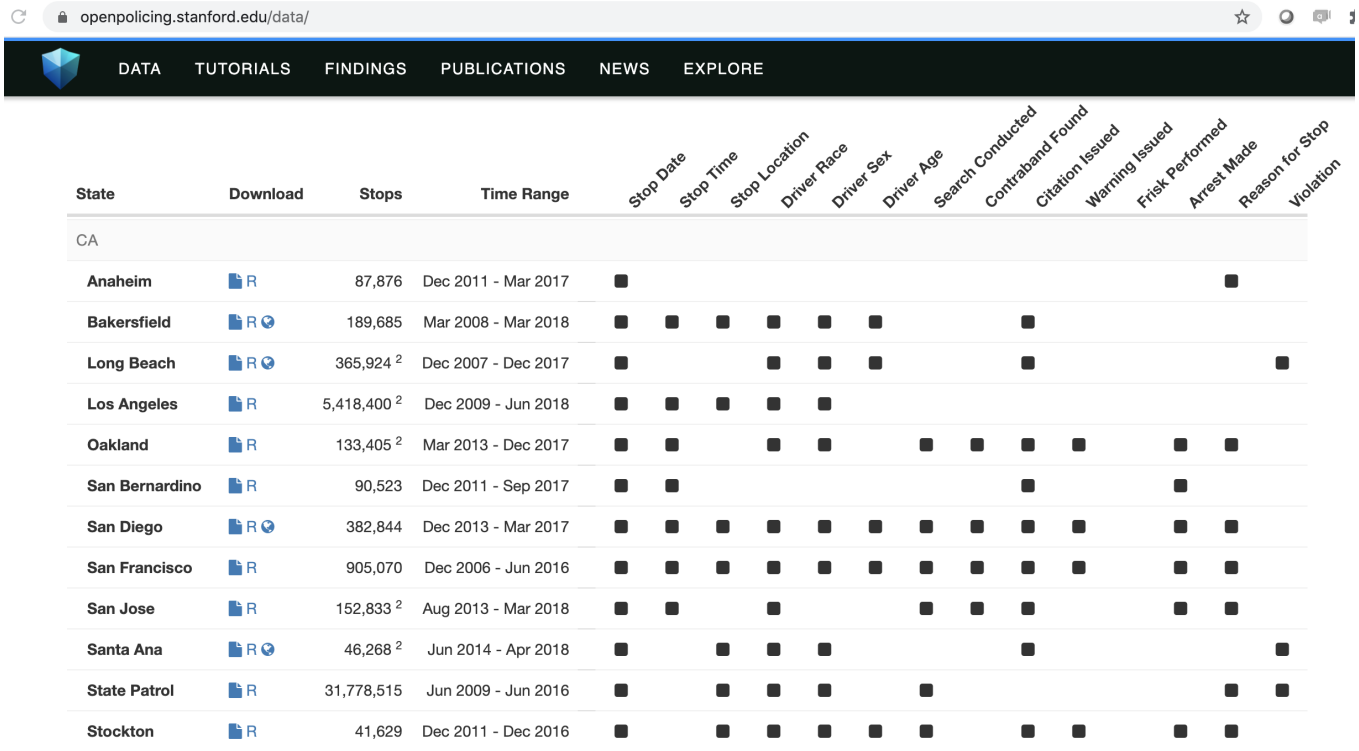


Figure 1: https://openpolicing.stanford.edu/

Stanford Policing Data



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	R	87,876	Dec 2011 - Mar 2017	■											■		
Bakersfield	R	189,685	Mar 2008 - Mar 2018	■	■	■	■	■	■		■						
Long Beach	R	365,924 [2]	Dec 2007 - Dec 2017	■			■	■	■		■						■
Los Angeles	R	5,418,400 [2]	Dec 2009 - Jun 2018	■	■	■	■	■									
Oakland	R	133,405 [2]	Mar 2013 - Dec 2017	■	■		■	■		■	■	■	■		■	■	
San Bernardino	R	90,523	Dec 2011 - Sep 2017	■	■							■	■		■		
San Diego	R	382,844	Dec 2013 - Mar 2017	■	■	■	■	■	■	■	■	■	■		■	■	
San Francisco	R	905,070	Dec 2006 - Jun 2016	■	■	■	■	■	■	■	■	■	■		■	■	
San Jose	R	152,833 [2]	Aug 2013 - Mar 2018	■	■		■			■	■	■			■	■	
Santa Ana	R	46,268 [2]	Jun 2014 - Apr 2018	■		■	■	■				■					■
State Patrol	R	31,778,515	Jun 2009 - Jun 2016	■		■	■	■		■					■	■	
Stockton	R	41,629	Dec 2011 - Dec 2016	■		■	■	■	■	■		■	■		■	■	

Figure 2: Just under 100 datasets, ~1 TB of data.
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 pe
  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

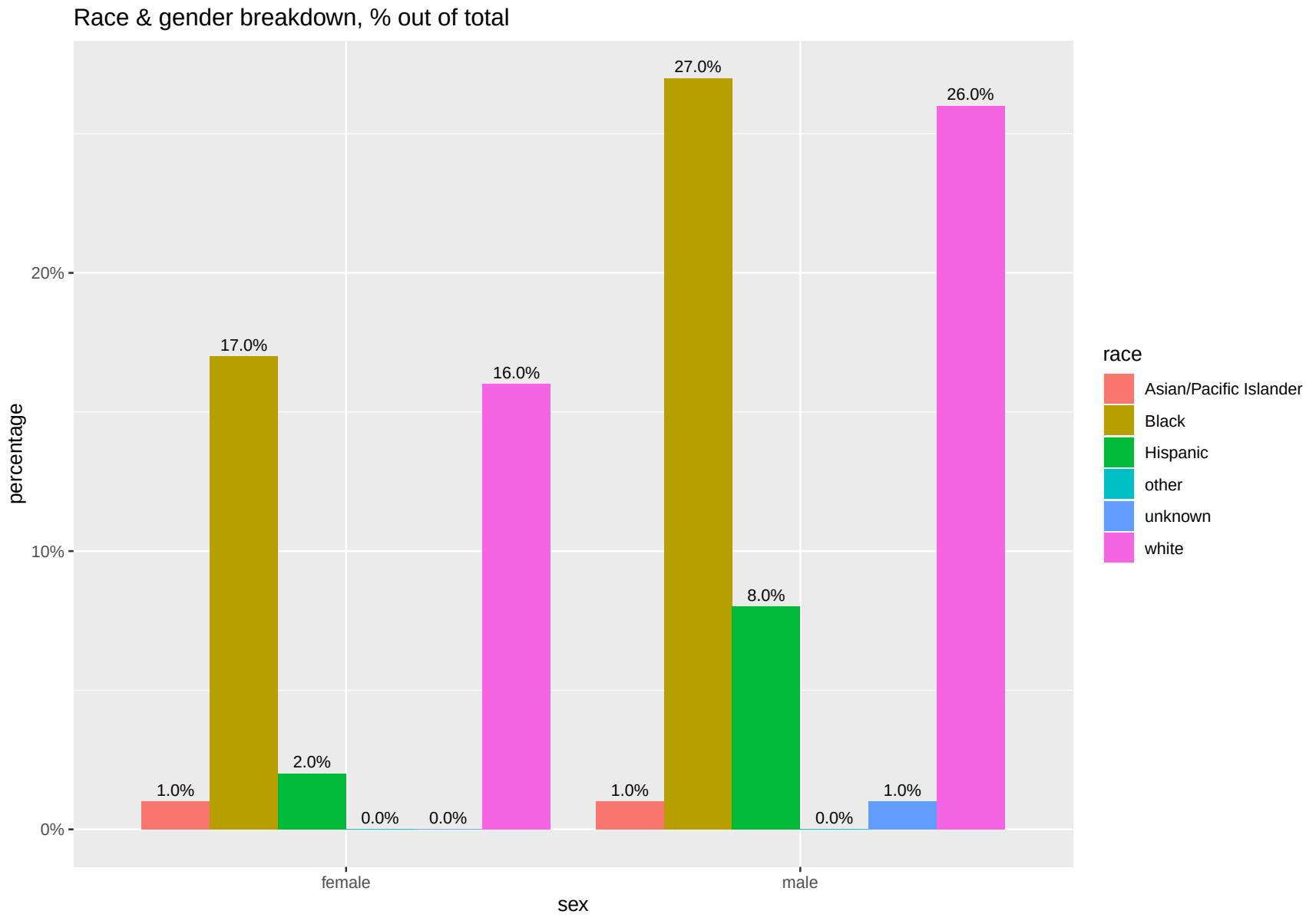
. . . continued

```
# 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



Step 2: advanced data viz

Accessing, joining, and wrangling all the datasets.

Step 2: advanced data viz

With `facet_geo`

Step 3: modeling search

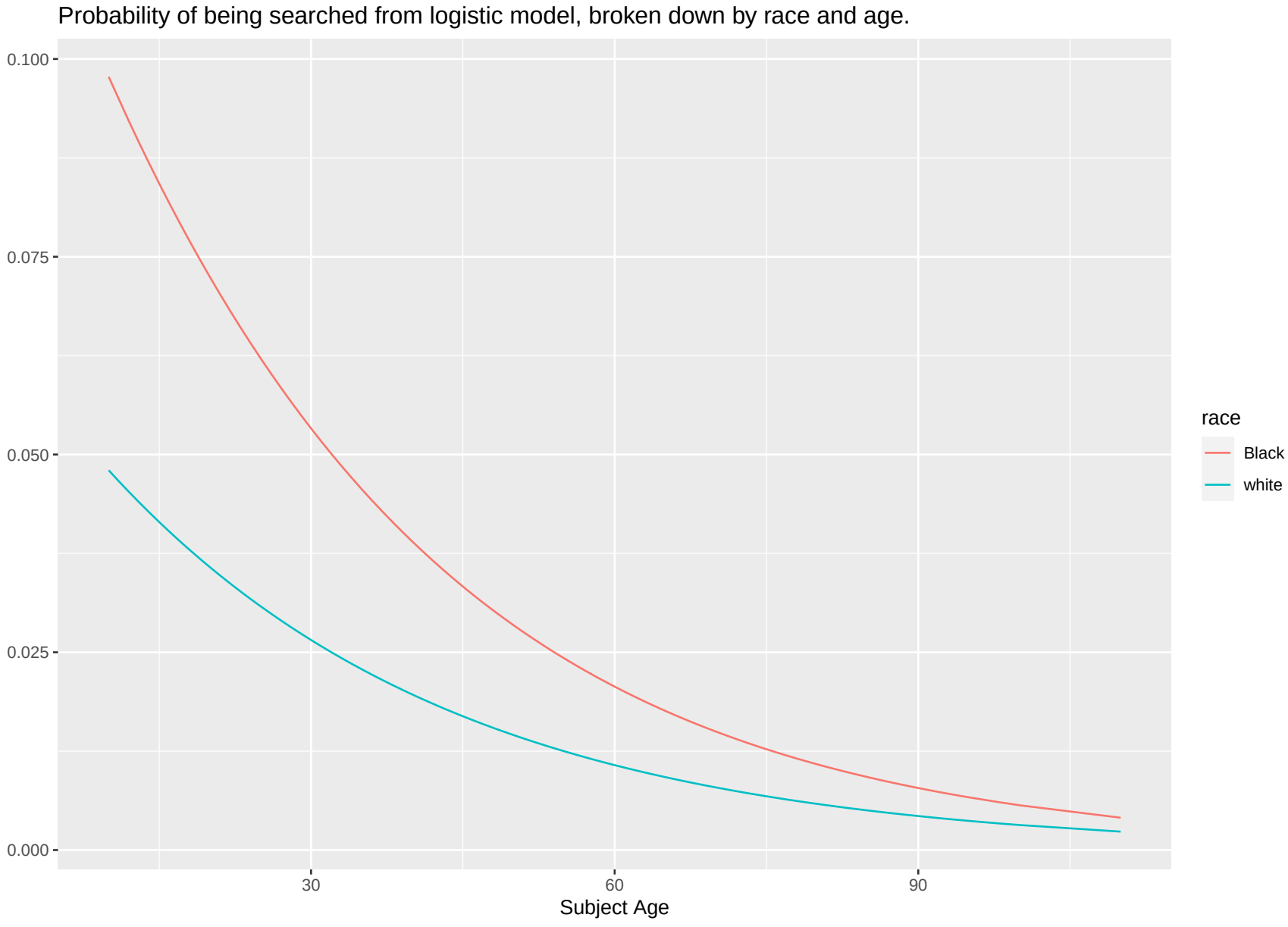
n.b., **all** the observations were traffic stops, so we can't model demographics of who was pulled over (model `search` instead).

```
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



Language around sensitive data

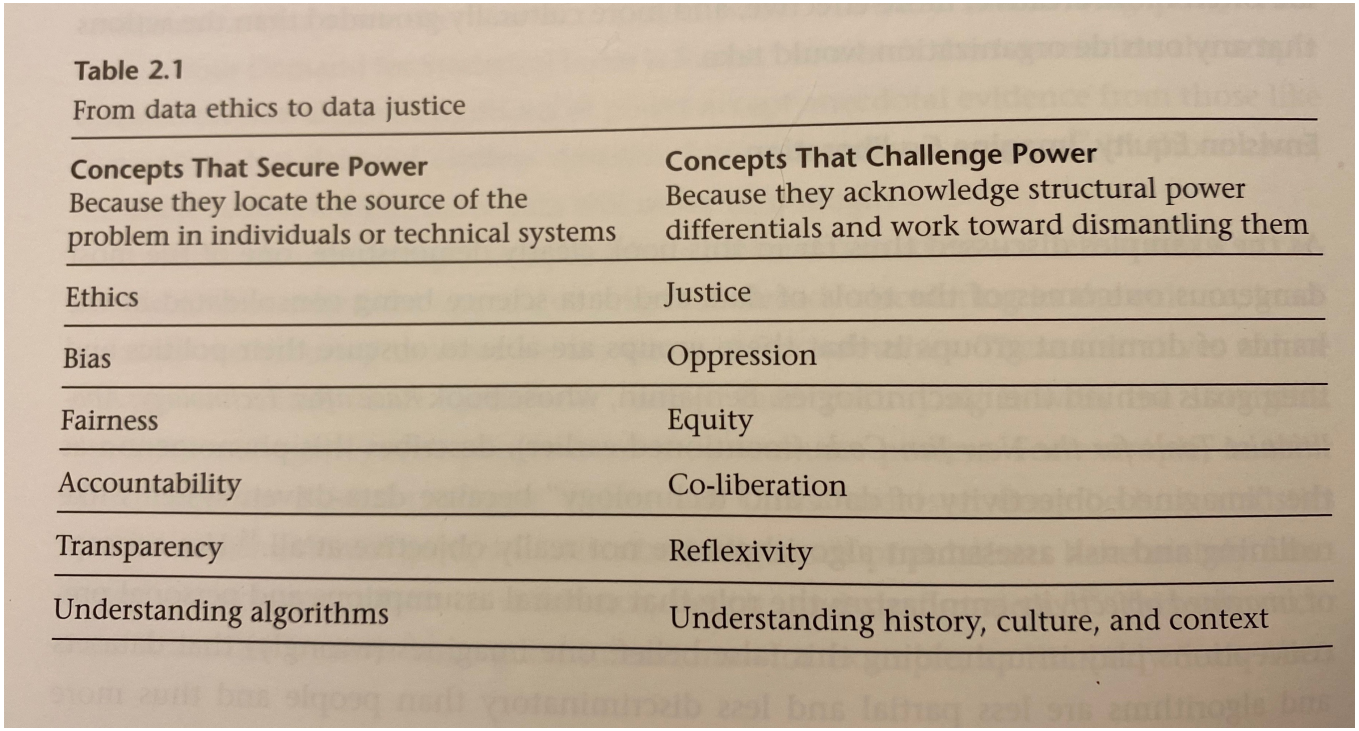


Table 2.1
From data ethics to data justice

Concepts That Secure Power Because they locate the source of the problem in individuals or technical systems	**Concepts That Challenge Power** 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 3: **Data Feminism** by Catherine D'Ignazio & Lauren F. Kleinhttps://datafeminism.io/

Bias or Oppression?

Figure 4: **Machine Bias** ProPublica
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Language around sensitive data

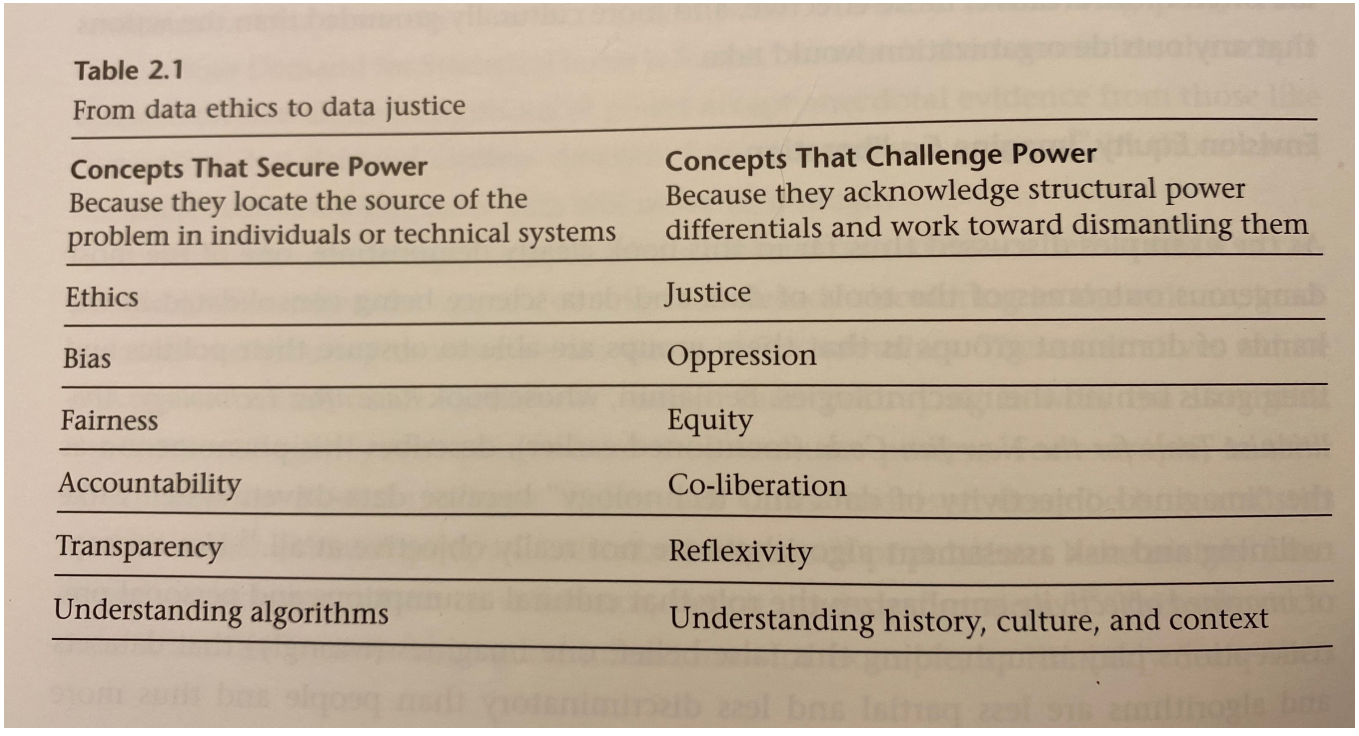


Table 2.1
From data ethics to data justice

Concepts That Secure Power Because they locate the source of the problem in individuals or technical systems	**Concepts That Challenge Power** 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** by Catherine D'Ignazio & Lauren F. Klein, https://datafeminism.io/

Equity vs Equality

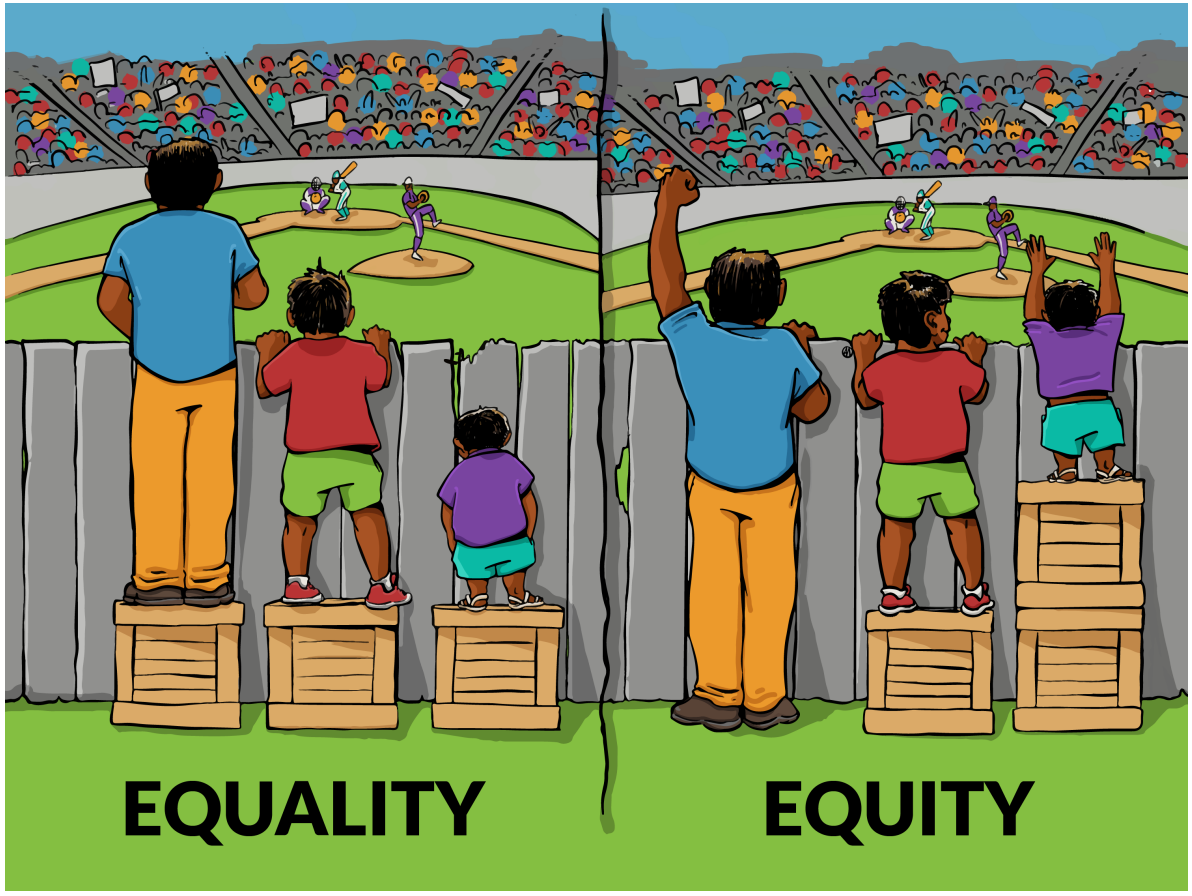


Figure 6: Image credit: Interaction Institute for Social Change | Artist: Angus Maguire.

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

> *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.

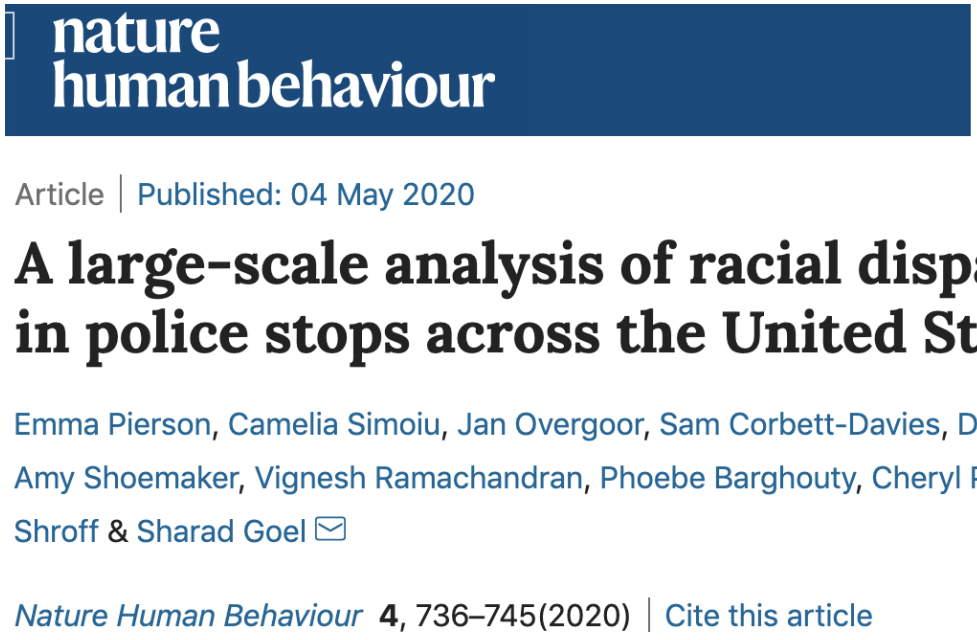


Figure 7: Evidence of racial disparities in traffic stops.

Significance Magazine (August 2020)

Measures for determining racial disparities:

- outcomes test (was contraband found?)
- veil of darkness (is there a difference just before vs just after sunset?)

ANALYSIS

Racial disparities in police stops in US cities

Following the death of George Floyd, and protests against systemic racism, **Roberto Rivera** and **Janet Rosenbaum** assess evidence for racial disparities in police traffic stop data in San Diego and San Francisco

On 25 May 2020 George Floyd, an African American man, died in Minneapolis, Minnesota, while in police custody. Floyd was confronted by four police officers and was handcuffed on the ground, and one officer – Derek Chauvin, a white police officer – pressed his knee to Floyd's neck

being found are dependent on a driver's race, racial bias may be occurring.

This is what we set out to investigate.

Stops and searches

Some cities that share police traffic stop data in open data portals include Cincinnati,

Figure 1 summarises vehicle stops in San Diego and San Francisco by driver's race and whether the driver was searched. Racial groups were ordered by frequency of vehicle stops, and search incidences were stacked by driver's race. Stacked bar charts help visualise the relative frequency of an event based on

which is demographically expected. However, a greater proportion of black and Hispanic drivers are searched than are drivers of other races. Specifically, for black drivers, the "search" bars cover more of the entire "race" bars than for white drivers. The same can be said for Hispanic drivers.

Search disparity between

actionsports/Bigstock.com

Figure 8: GitHub repo: https://github.com/bakuninpr/traffic-stops-and-racial-disparity

Traffic Stop

Motorcycle with Montana plates vs. Van with Colorado plates, traffic stop on Sunday, Aug 2, 2020.

Figure 9: Family with young children handcuffed.

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

@jo_hardin47

https://github.com/hardin47

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