Data Schema Descriptions

Officer Roster

- **birth_year**: Birth year of the officer.
- **appointed_month**: The month and year the officer was made an officer in YYYY-MM-DD format. The day is always the first day of the month.
- **officer_id**: Unique identifier for each officer.
- **officer_race**: Race of the officer.
- **officer_gender**: Sex of the officer.
- **spanish**: Does the officer speak Spanish or not?
- Uniquely identified by **officer_id**. The unit of observation is an officer.
- **Number of officers**: 33645

Shift Assignments

- **officer_id**: Unique identifier for each officer.
- **month**: Month of the shift in YYYY-MM-DD format. The day is always the first day of the month.
- **rank**: Rank of the officer assigned to the shift.
- **unit**: Unit of the officer assigned to the shift.
- **date**: Date of the shift in YYYY-MM-DD format.
- **shift**: The shift the officer is assigned to.
- **start_time**: Hour start time of the shift in military time.
- **end_time**: Hour end time of the shift in military time.
- **weekday**: Day of the week of the shift.
- **beat_assigned**: The beat the officer is assigned to.
- **appointed_month**: The month and year the officer was made an officer. **Dropped as it's redundant**.
- **months_from_start**: The number of months between the officer's appointment date and their shift date.
- **months_from_start_sq**: The number of months between the officer's appointment date and their shift date, squared. **Dropped as it can be recreated**.
- **duration**: Length of the shift in hours.
- **shift_id**: Unique identifier for each shift assignment. **Created by me**.
- Uniquely identified by **officer_id** and **date** or **shift_id**. The unit of observation is a specific shift for a specific officer.
- **Number of shift assignments**: 3519518

Stops

- **stop_id**: Identifier for each stop.
- **time**: Time of the stop in YYYY-MM-DD HH:MM:SS format.
- **date**: Date of the stop in YYYY-MM-DD format.
- **district**: Police district where the stop took place.
- **po_first**: Was the focal officer the first to respond to the scene?
- **stop_type**: What was the type of the stop?
- **contact_type**: Collapsed version of **stop_type** (less categories).
- **civ.race**: Race of the civilian.
- **civ.gender**: Sex of the civilian.

- **civ.age**: Age of the civilian at the time of the stop.
- **lat**: Latitude of the stop.
- **lon**: Longitude of the stop.
- **officer_id**: Unique identifier for the officer.
- **month**: Month of the stop in YYYY-MM-DD format. The day is always the first day of the month.
- **civilian_race_short**: Collapsed version of **civ.race** (less categories).
- **hour**: Hour of the day when the stop took place rounded to the nearest hour in military time.
- **stop_officer_id**: Unique identifier for each entry. **Created by me**.
- The unit of analysis is a unique officer involved in a stop. Each row can be uniquely identified by **officer_id** and **stop_id** or by **stop_officer_id**.
- Each stop involves only one civilian, but they can involve multiple officers. It is possible multiple stops are all a part of one larger incident involving multiple civilians. This can be investigated by examining stops that took place in the same location at the same time involving the same officers.
- **Number of rows**: 1703158
- **Number of unique stops**: 946912

Arrests

- **date**: Date of the arrest in YYYY-MM-DD format.
- **hour**: Hour of the day when the arrest took place rounded to the nearest hour in military time.
- **crime_code**: The suspected crime type causing the arrest.
- **statute_description**: More detailed categories describing what specific statute was suspected to have been violated.
- **lat**: Latitude of the arrest.
- **lon**: Longitude of the arrest.
- **district**: Police district where the arrest took place.
- **civ.race**: Race of the civilian.
- **civ.gender**: Sex of the civilian.
- **civ.age**: Age of the civilian at the time of the stop.
- **arrest_id**: Identifier for each arrest.
- **officer_id**: Unique identifier for each officer.
- **month**: Month of the arrest in YYYY-MM-DD format. The day is always the first day of the month.
- **civilian_race_short**: Collapsed version of **civ.race** (less categories).
- **arrest_officer_id**: Unique identifier for each entry. **Created by me**.
- The unit of analysis is a unique officer involved in an arrest. Each row can be uniquely identified by **officer_id** and **arrest_id** or by **arrest_officer_id**.
- Each arrest involves only one civilian, but they can involve multiple officers. It is possible multiple arrests are all a part of one larger incident involving multiple civilians. This can be investigated by examining arrests that took place in the same location at the same time involving the same officers.
- **Number of rows**: 321872
- **Number of unique arrests**: 164802

Uses of Force

- **date**: Date of the use of force in YYYY-MM-DD format.
- **time**: Time of the use of force in YYYY-MM-DD HH:MM:SS format.
- **district**: Police district where the use of force took place.
- **lat**: Latitude of the use of force.
- **lon**: Longitude of the force.
- **civ.race**: Race of the civilian.
- **civ.gender**: Sex of the civilian.
- **civ.age**: Age of the civilian at the time of the stop.
- **civ.injured**: Was the civilian injured?
- **force_id**: Unique identifier for each use of force incident.

- **officer_id**: Unique identifier for each officer.
- **month**: Month of the arrest in YYYY-MM-DD format. The day is always the first day of the month.
- **civilian_race_short**: Collapsed version of **civ.race** (less categories).
- **hour**: Hour of the day when the use of force took place rounded to the nearest hour in military time.
- The unit of analysis is a use of force incident. Only one police officer is listed for each use of force incident. Each row can be uniquely identified by **force_id**.
- **Number of uses of force**: 9293

```
## Warning: Predicate functions must be wrapped in `where()`.
##
##   # Bad
##   data %>% select(is.character)
##
##   # Good
##   data %>% select(where(is.character))
##
## i Please update your code.
## This message is displayed once per session.
```

```
## Warning: Removed 839 rows containing non-finite values (stat_density).
```

```
## Warning: Removed 839 rows containing non-finite values (stat_boxplot).
```

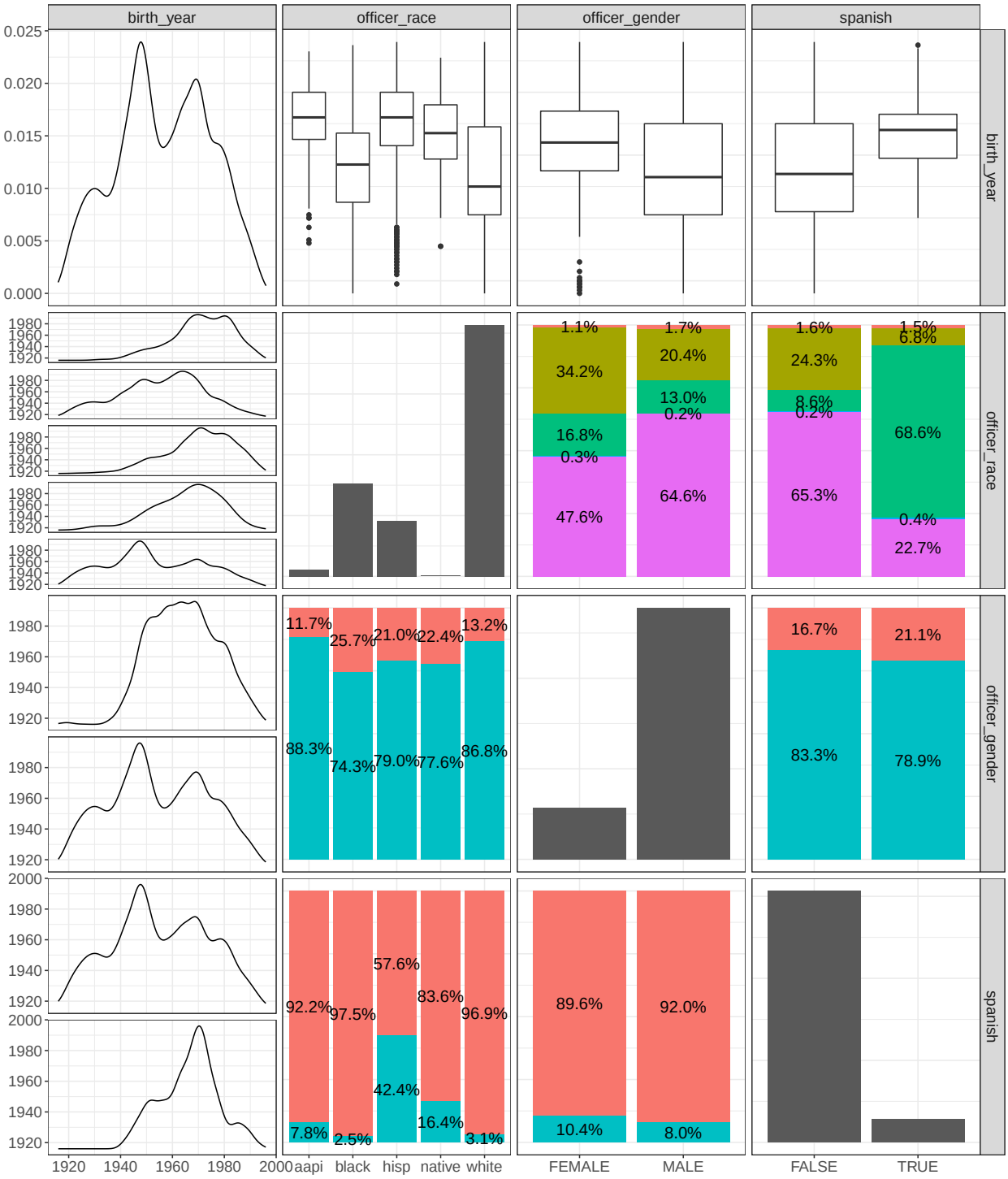
```
## Warning: Removed 839 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 839 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 839 rows containing non-finite values (stat_density).
```

```
## Warning: Removed 839 rows containing non-finite values (stat_density).
```

```
## Warning: Removed 839 rows containing non-finite values (stat_density).
```


birth_year
officer_race
officer_gender
spanish
birth_year
officer_race
officer_gender
spanish
0.025
0.020
0.015
0.010
0.005
0.000
1980
1960
1940
1920
2000
1.1%
34.2%
16.8%
0.3%
47.6%
1.7%
20.4%
13.0%
0.2%
64.6%
1.6%
24.3%
8.6%
0.2%
65.3%
1.5%
6.8%
68.6%
0.4%
22.7%
11.7%
88.3%
25.7%
74.3%
21.0%
79.0%
22.4%
77.6%
13.2%
86.8%
16.7%
83.3%
21.1%
78.9%
92.2%
7.8%
97.5%
2.5%
57.6%
42.4%
83.6%
16.4%
96.9%
3.1%
89.6%
10.4%
92.0%
8.0%
aapi
black
hisp
native
white
FEMALE
MALE
FALSE
TRUE