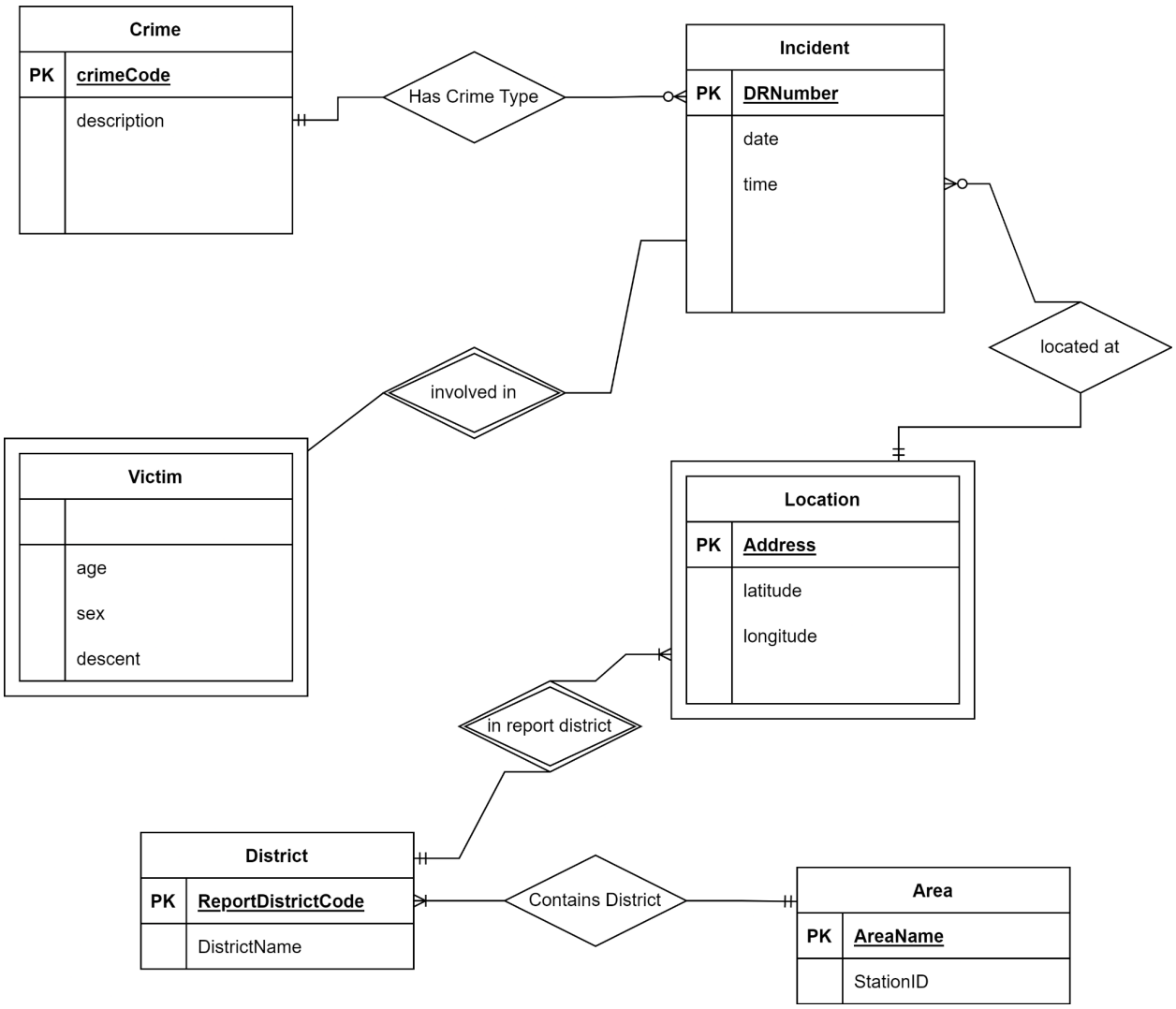
# Stage 2:

Jacob Stolker, Tianshun Gao, Alex Capps, Alex Ostrowski

## ER Diagram




## Assumptions we've made for each entity and relationship.

- Crime (1…1) and Incident (0…)
  - Each incident has exactly one crime (we ignore the 2nd, 3rd, and 4th crime codes from the original data, since most of these are empty and irrelevant to the main crime.)
  - Each crime type can correspond to 0 or many incidents.
- Incident (1…1) and Victim (1…1)
  - Each victim is only identifiable based on the incident they were involved in, since age/sex/descent are not unique to any one person.

- Incident (0…) and Location (1…1)
    - Each incident has exactly one location.
    - Each location can be attributed to zero or many incidents.
- Location (1…) and District (1…1)
    - Each location is entirely within one district
    - Each district can have many locations within it, but has to have at least one.
    - Since we are using Address as private key for Location, and Address is only street name and number, we assume that Address is only identifiable by the District it's in. (since street names/numbers can be repeated in other districts.)
- District (1…) and Area (1…1)
    - Each district must belong to exactly one Area, but each Area can have 1 to many districts.
    - We assume each district is entirely within one area (with one area code) and not split between two areas.

## Relational Schema

- Crime(Code:INT [PK] , Desc:VARCHAR(50))

- Area(Name:VARCHAR(20) [PK], StationID:INT)

- Location(Address:VARCHAR(50) [PK], Latitude:REAL, Longitude:REAL)

- Incident(DRNumber:INT [PK], Time:VARCHAR(20), Date:DATE)

- District(ReportDistrictCode:INT [PK], DistrictName:VARCHAR(20))

- Victim(IncidentDRNumber:INT [FK to Incident.DRNumber], Age:INT, Sex:VARCHAR(2), Descent:VARCHAR(20))