CITS5504 DATA WAREHOUSING

Project 2 - GRAPH DATABASE

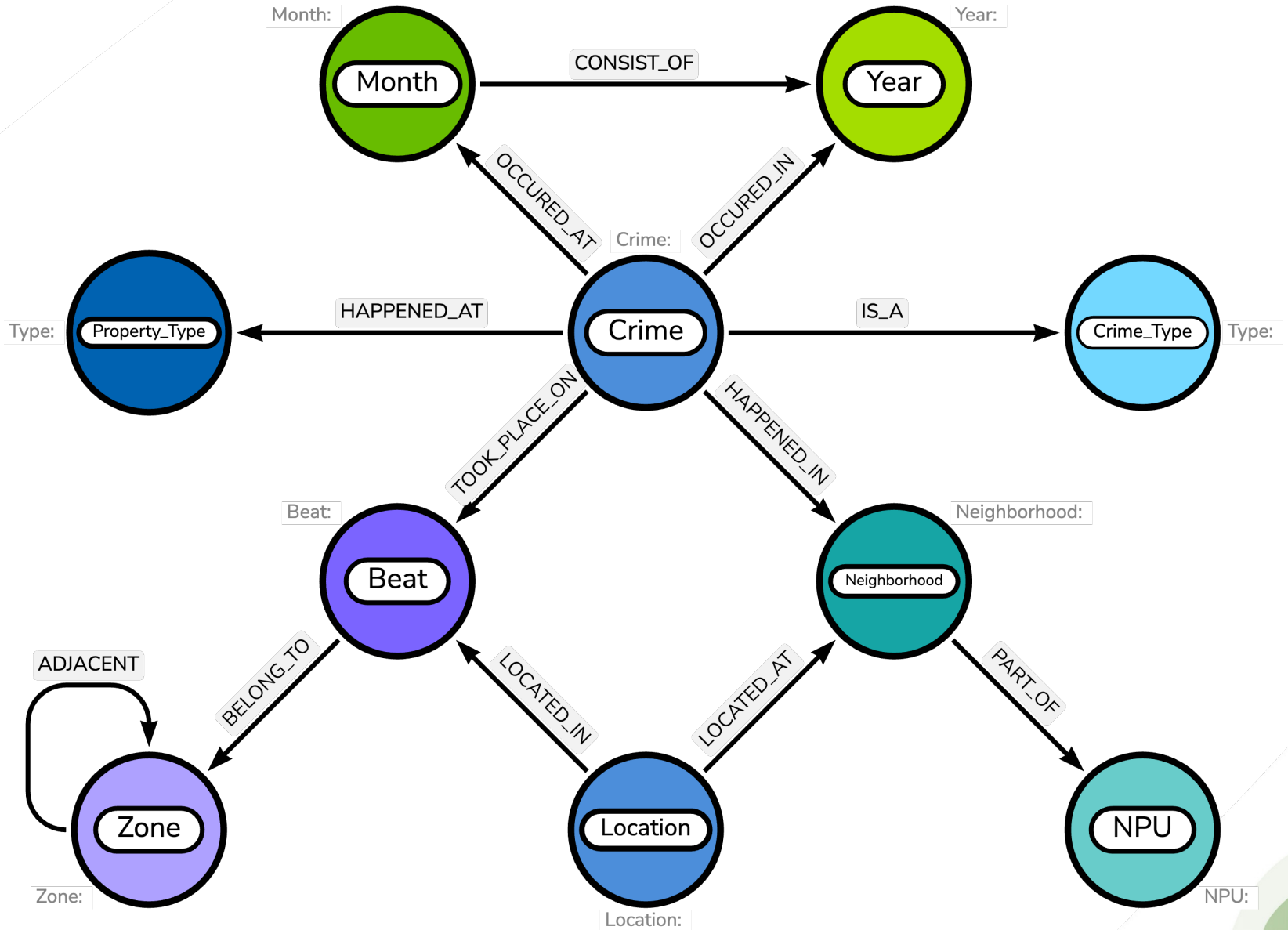
Jiaze Li

Lei Chen

CONTENT

- Graph Database Design
- Nodes Design & Relation Design
- Questions answering
- Arrow tool import
- Graph to answer useful queries
- Answer the queries

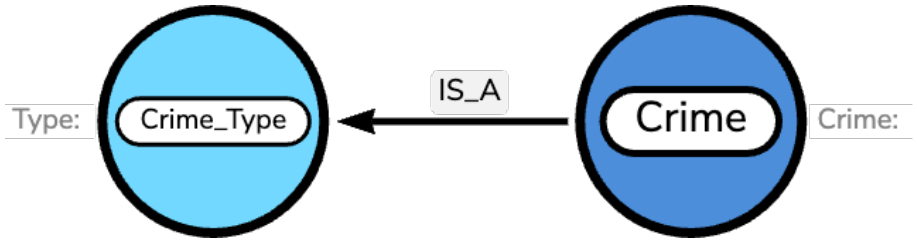
Graph Database Design



Nodes Design & Relationship Design

Nodes: Crime
Property Type

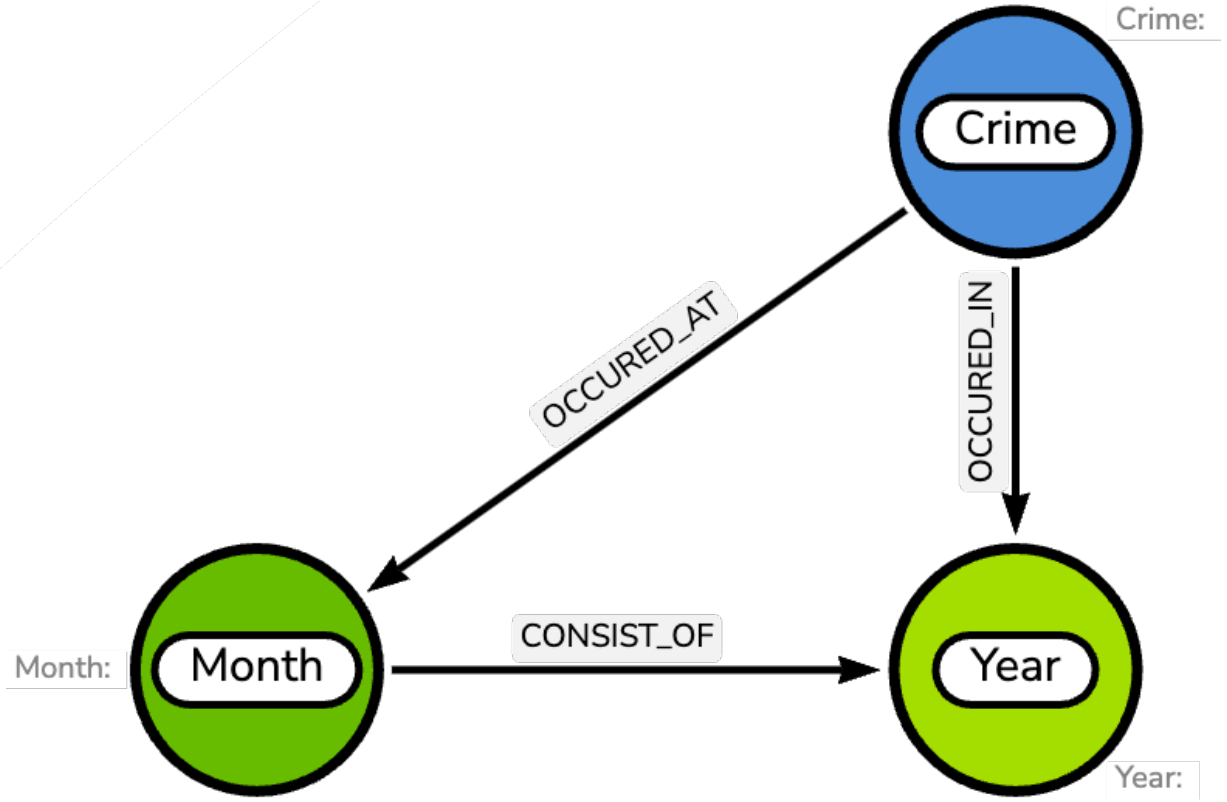
Relationship: HAPPENED AT



Nodes: Crime
Crime Type

Relationship: IS A

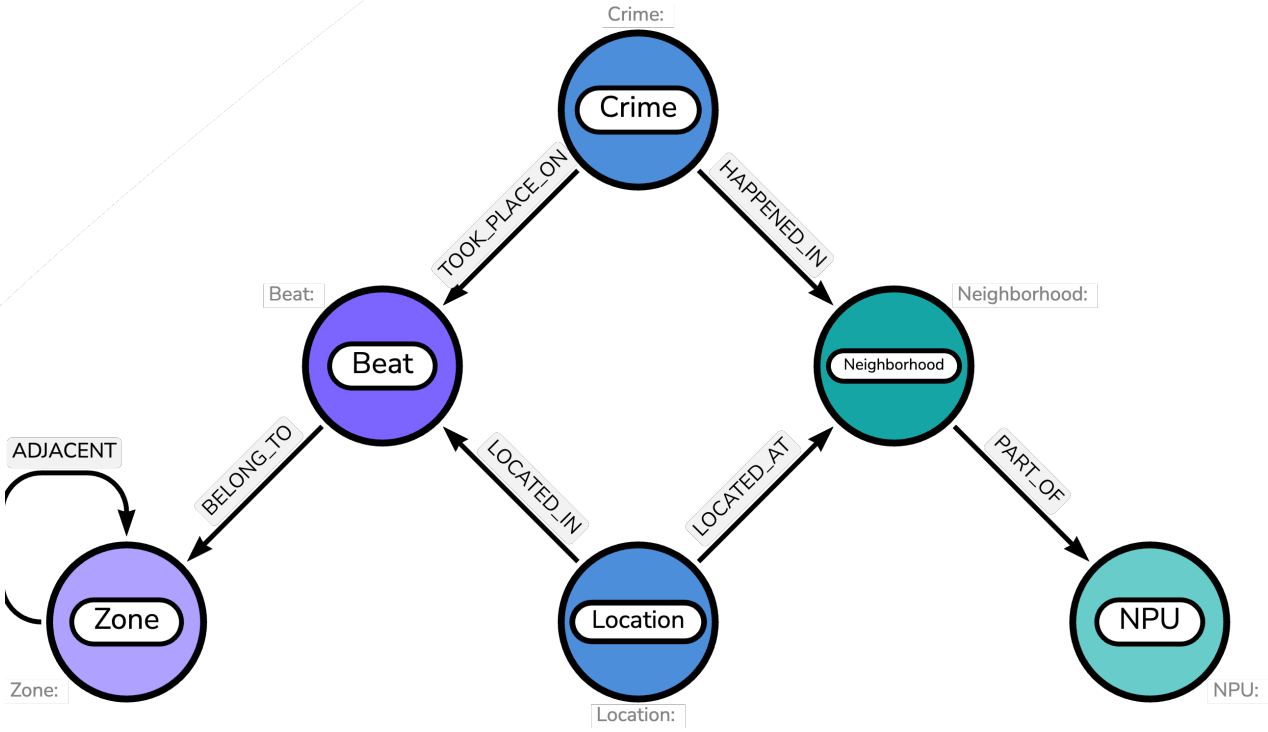
Nodes Design & Relation Design



Nodes: Crime
Month
Year

Relationship: OCCURED AT
OCCURED IN
CONSIST OF

Nodes Design & Relation Design



Nodes:

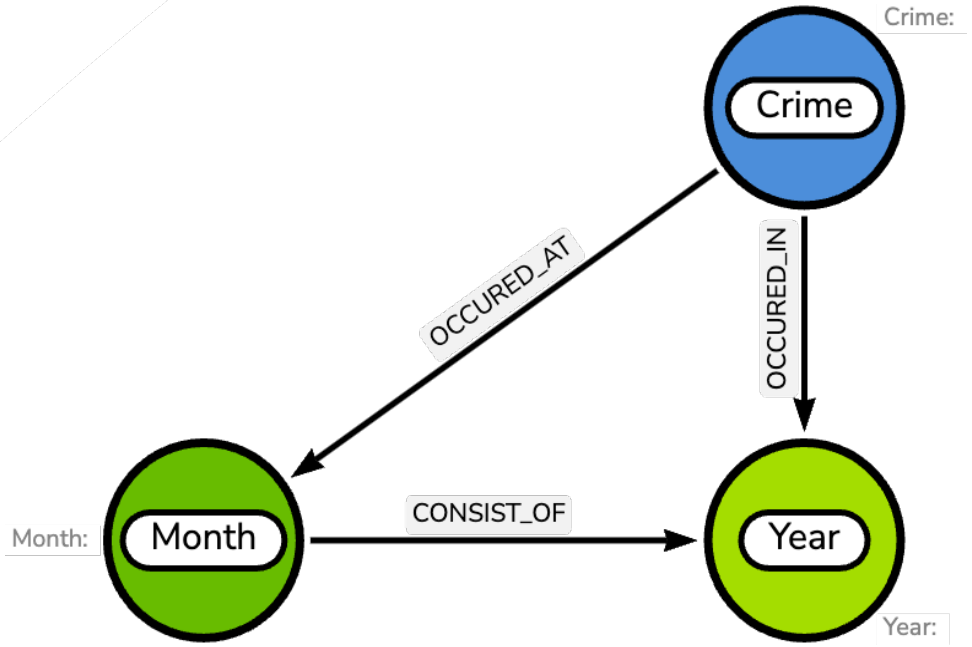
- Crime
- Neighborhood
- NPU
- Beat
- Zone
- Location

Relationship:

- HAPPENED AT
- PART OF
- TOOK PLACE ON
- BELONG TO
- ADJACENT
- LOCATED AT
- LOCATED IN

When to use a property rather than a node

Data Complexity & Query requirements



Or

?

When to add properties to a relation

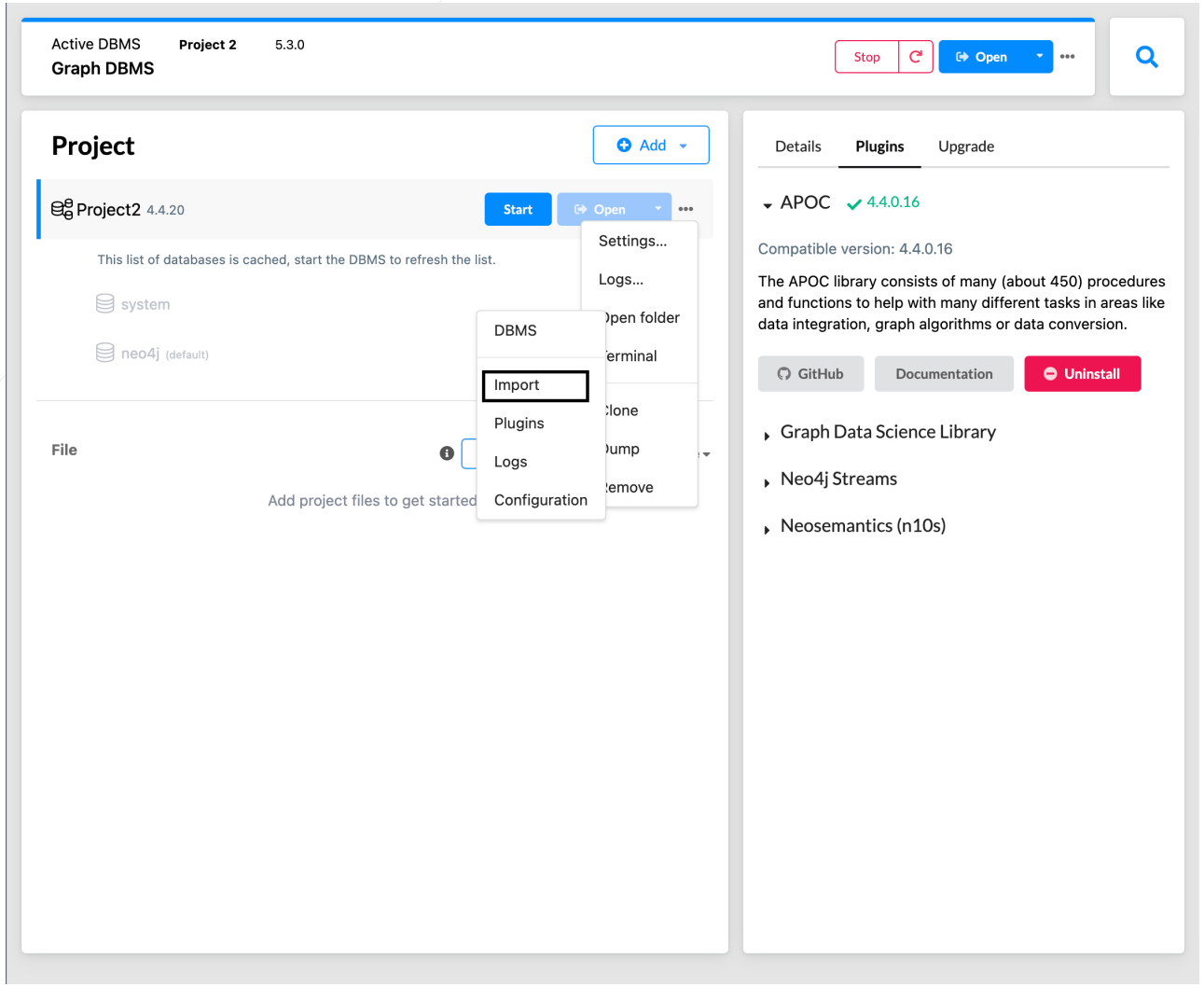
- **Detailing the Relationship**
- **Temporal Context**
- **Quantifying the Relationship**
- **Status or State**

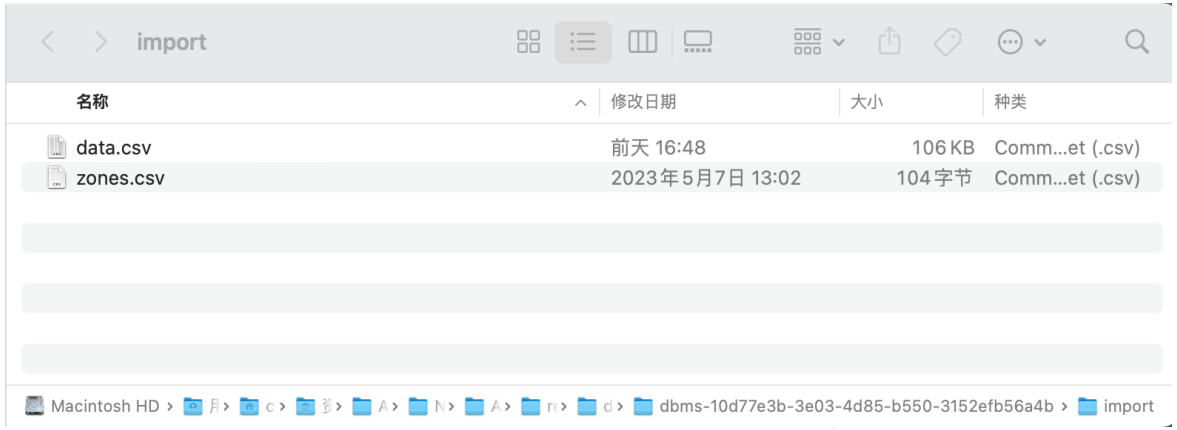
When to add extra nodes and for what type of queries

- Increase Query Efficiency
- Additional Details/Complexity
- Expand Scope of Analysis
- Handling Hierarchical Data
- Enable New Types of Relationships

Arrow tool import

a.Load the data into the graph database





Arrow tool import

b. Zone table and data table showing

zones.csv

ZoneKey	AdjacentZoneKey
1	2
1	3
1	4
1	5
2	1

data.csv

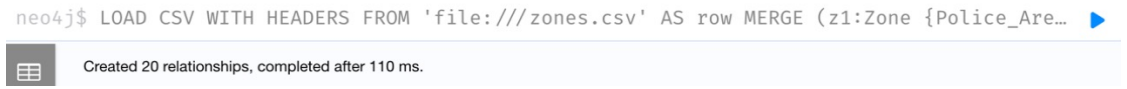
CrimeType	Location	Beat	Neighborhood	NPU	PropertyType	Year	Month	Zone	Crime	ZoneKey
BURGLARY-RESIDENCE	242 ANDERSON AVE NW	108	Dixie Hills	J	house_number	2011	10	Zone1	1	1
AUTO THEFT	1270 CAROLINE ST NE	609	Edgewood	O	amenity	2012	10	Zone6	2	6
LARCENY-FROM VEHICLE	3180 ARDLEY RD SW	406	East Ardley Road	I	house_number	2016	10	Zone4	3	4
AGG ASSAULT	738 PRINCE PL NW	109	Center Hill	J	neighbourhood	2016	10	Zone1	4	1
AUTO THEFT	2909 CAMPBELLTON RD SW	409	Southwest	R	house_number	2010	10	Zone4	5	4

Arrow tool import

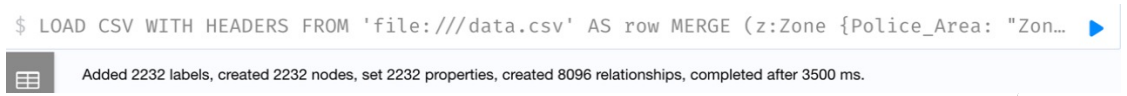
c. Creating nodes and establishing relationships

Create

```
LOAD CSV WITH HEADERS FROM 'file:///zones.csv' AS row
MERGE (z1:Zone {Police_Area: "Zone" + row.ZoneKey})
MERGE (z2:Zone {Police_Area: "Zone" + row.AdjacentZoneKey})
MERGE (z1)-[:ADJACENT]->(z2)
```

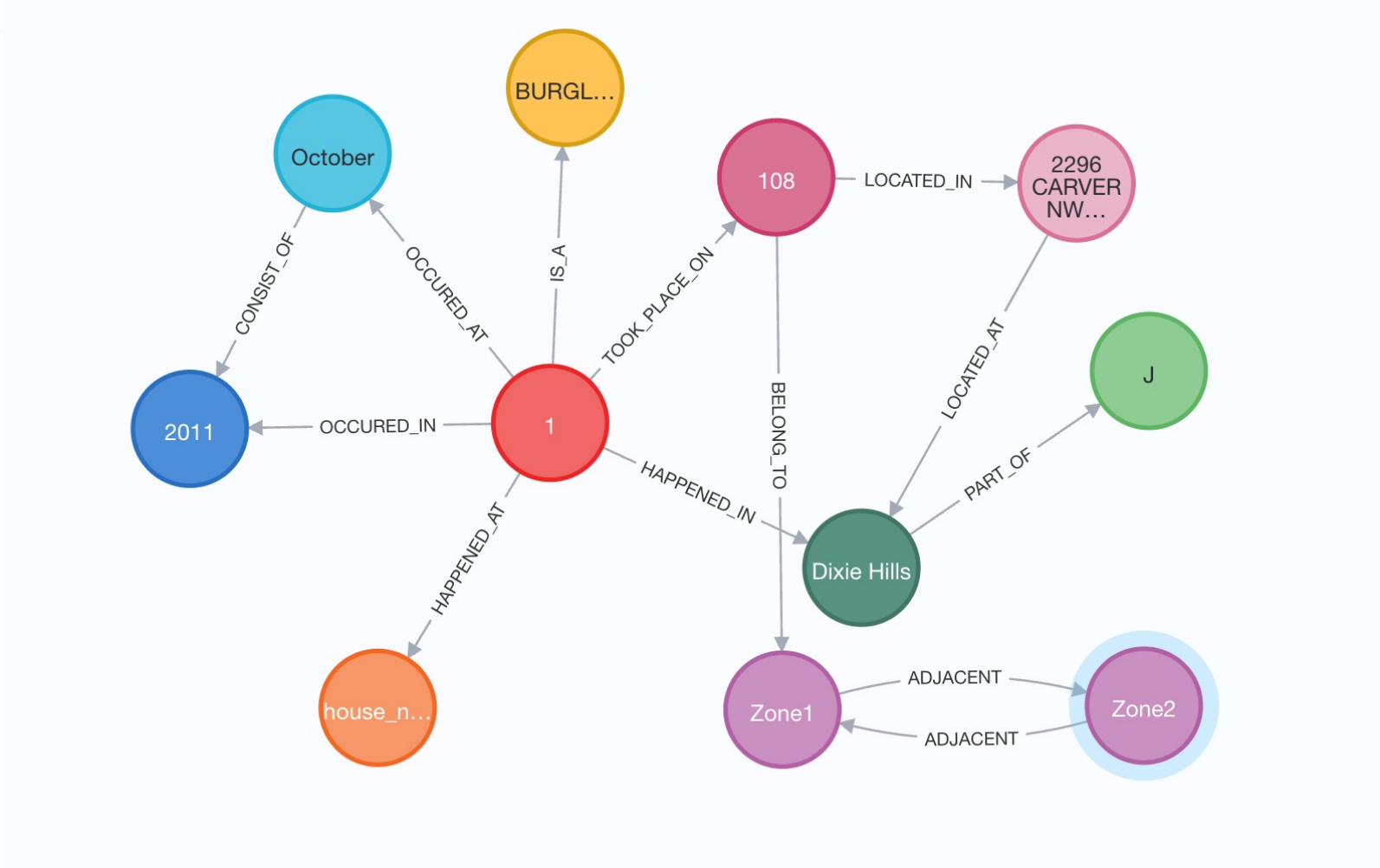



```
LOAD CSV WITH HEADERS FROM 'file:///data.csv' AS row
MERGE (z:Zone {Police_Area: "Zone" + row.ZoneKey})
MERGE (b:Beat {Beat: row.Beat})
MERGE (n:Neighborhood {Neighborhood: row.Neighborhood})
MERGE (npu:NPU {NPU: row.NPU})
MERGE (p:Property_Type {Type: row.PropertyType})
MERGE (m:Month {Month: row.Month})
MERGE (y:Year {Year: row.Year})
MERGE (ct:Crime_Type {Type: row.CrimeType})
MERGE (l:Location {Location: row.Location})
MERGE (c:Crime {Crime: row.Crime})
MERGE (b)-[:BELONG_TO]->(z)
MERGE (c)-[:TOOK_PLACE_ON]->(b)
MERGE (c)-[:HAPPENED_IN]->(n)
MERGE (n)-[:PART_OF]->(npu)
MERGE (c)-[:HAPPENED_AT]->(p)
MERGE (c)-[:OCCURED_AT]->(m)
MERGE (c)-[:OCCURED_IN]->(y)
MERGE (m)-[:CONSIST_OF]->(y)
MERGE (c)-[:IS_A]->(ct)
MERGE (l)<-[:LOCATED_IN]-(b)
MERGE (l)-[:LOCATED_AT]->(n)
```



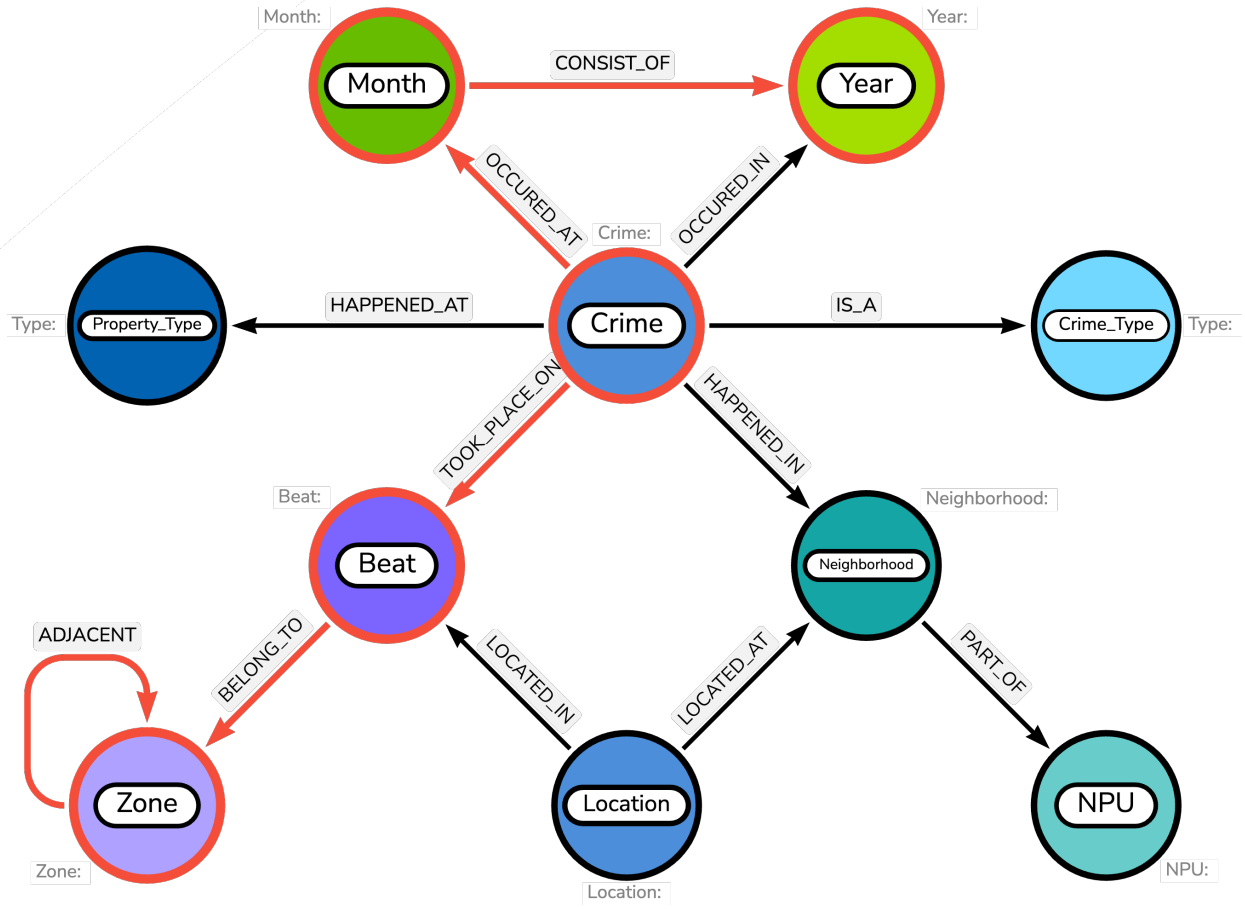
Arrow tool import

d. Example display



Graph to answer useful queries.

Purpose: **To find the zones that have are adjacent and sharing the same high crime months.**



a) **Create a relationship(ADJACENT) table**
b) **Find the desired nodes**
c) **Find the answer to a query by joining the nodes and relationships.**

Answer the queries.

Purpose:

To find the zones that have are adjacent and sharing the same high crime months.

```
MATCH (z:Zone)-[:ADJACENT]→(az:Zone)
MATCH (z)←[:BELONG_TO]-(b:Beat)←[:TOOK_PLACE_ON]-(c:Crime)-[:OCCURED_AT]→
(m:Month), (m)-[:CONSIST_OF]→(y:Year)
WITH z.Police_Area AS Zone, m.Month AS Month, y.Year AS Year, COUNT(c) AS
CrimeCount
ORDER BY CrimeCount DESC
WITH Zone, Month, Year, MAX(CrimeCount) AS MaxCrimeCount
WITH COLLECT({Zone: Zone, Month: Month, Year: Year, CrimeCount: MaxCrimeCount}) AS
ZoneData
MATCH (z:Zone)-[:ADJACENT]→(az:Zone)
WITH ZoneData, z.Police_Area AS Zone, az.Police_Area AS AdjacentZone
UNWIND ZoneData AS zd1
```

Table | Text | Code

	Zone	AdjacentZone	SharedHighCrimeMonth	Year
1	"Zone1"	"Zone2"	"10"	"2009"
2	"Zone1"	"Zone3"	"10"	"2009"
3	"Zone1"	"Zone4"	"10"	"2009"
4	"Zone1"	"Zone5"	"10"	"2009"
5	"Zone2"	"Zone1"	"10"	"2009"
6	"Zone2"	"Zone5"	"10"	"2009"
7				

Started streaming 180 records after 45 ms and completed after 245 ms.

a) **Matches all adjacent area pairs**
b) **Calculate the number of crimes in each area in each month.**
c) **Sorting the number**
d) **Return the answer**

THANKS