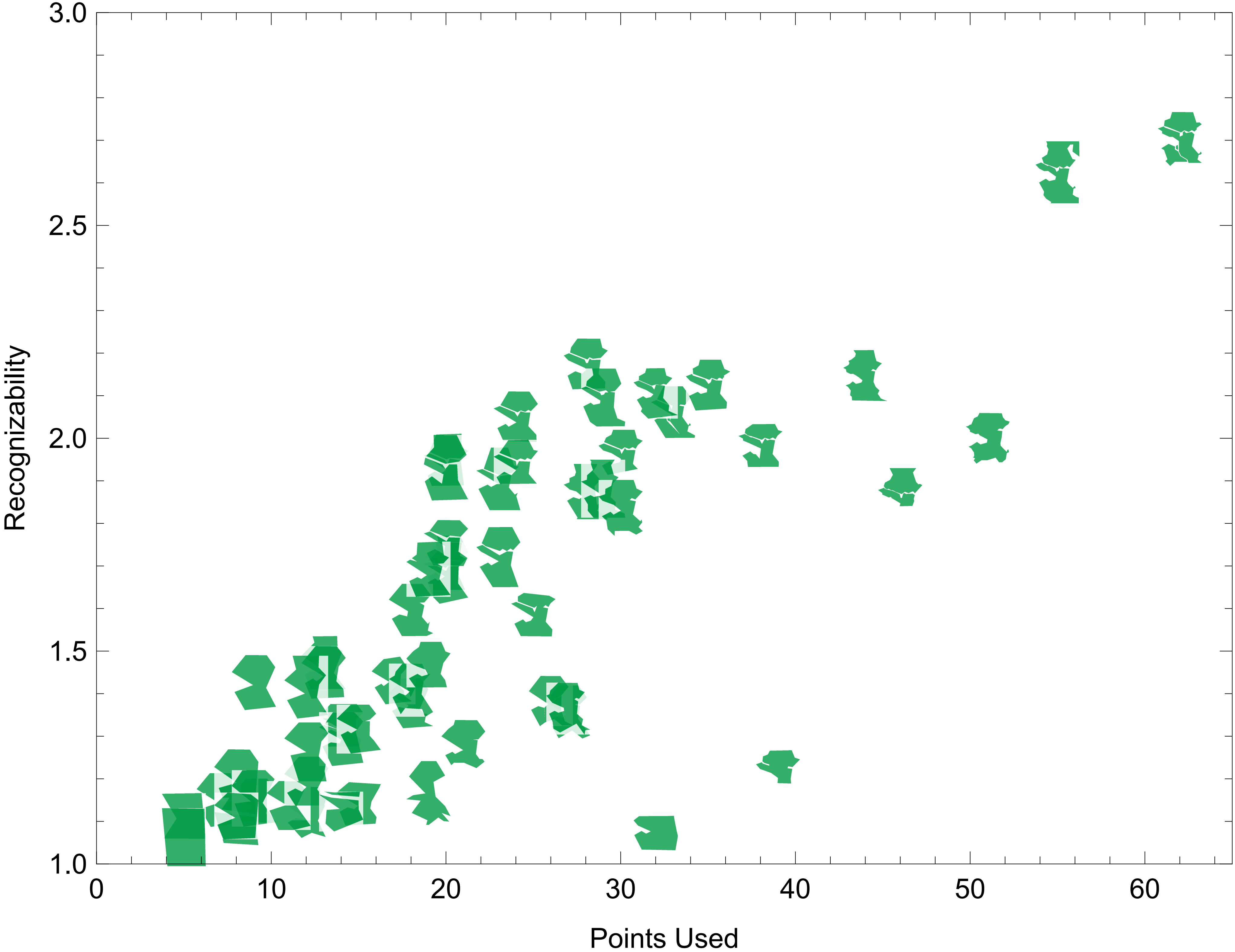


3.0
2.5
2.0
1.5
1.0
0
10
20
30
40
50
60
Recognizability
Points Used

Geography 360
November 3, 2016

GIS data storage and structures

1. *Questions and Announcements*
   - Recognizable Seattles: A Class Perspective (~3800 votes!)
   - Clarifications:
     NO MEETING OR EXAM during Final Exam time.
     Quiz II: December 2nd
     Final project due during Exam week.
   - Near/Far project:
     Part I: ~6 people had the OSM website problem.
     Part I: Not going to peer-review each other's OpenStreetMap contributions
     Part II: Posted. Read it over today. There's an outdoor component.

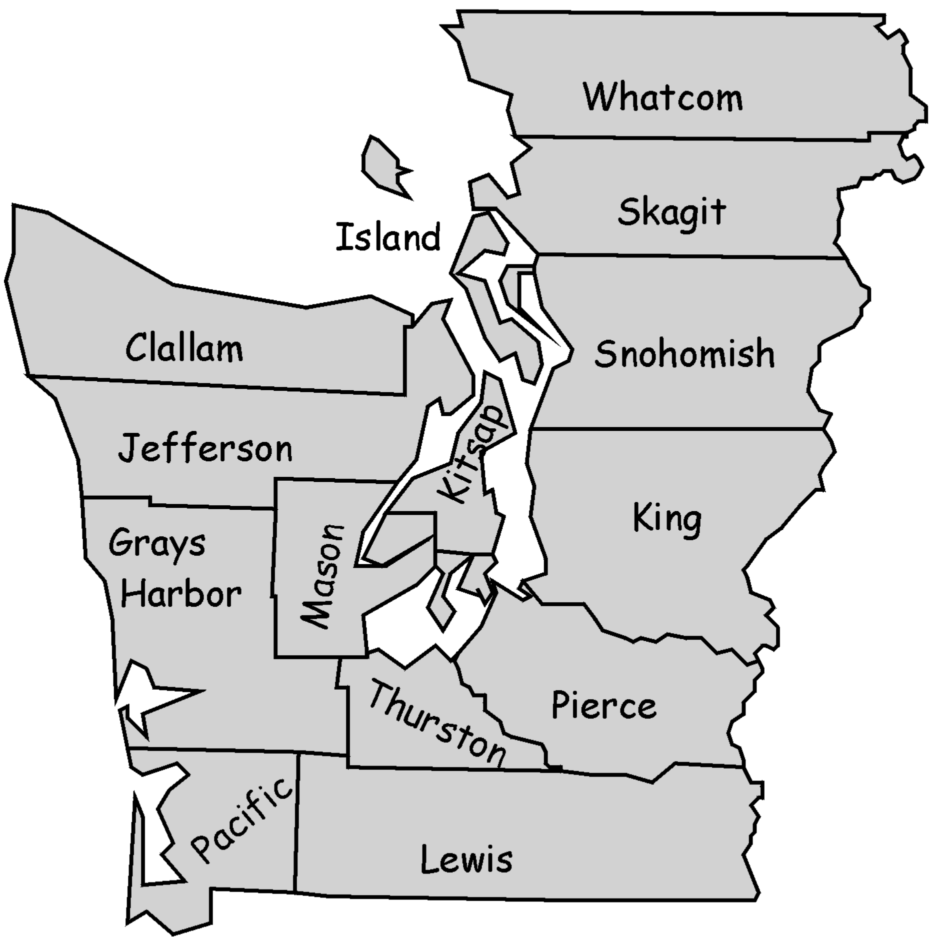
2. *GIS Data: Storage and Structures*
   - Representing attributes
   - Databases and local storage

How do we store information about the characteristics of spatial entities in GIS?

- *Attribute data* is typically **stored in tables** that are linked/associated with the spatial objects in your GIS.

- These tables help us in searching/retrieving objects based on their characteristics and in performing various analyses.

Visualizing Feature Geometries



Storing Associated Attribute Data in a Table

Name	FIPS	Pop90	Area	PopDn
Whatcom	53073	128	2170	59
Skagit	53057	80	1765	45
Clallam	53009	56	1779	32
Snohomish	53061	466	2102	222
Island	53029	60	231	261
Jefferson	53031	20	1773	11
Kitsap	53035	190	391	485
King	53033	1507	2164	696
Mason	53045	38	904	42
Gray Harbor	53027	64	1917	33
Pierce	53053	586	1651	355
Thurston	53067	161	698	231
Pacific	53049	19	945	20
Lewis	53041	59	2479	24

Question: Where is the spatial data associated with these features?

Answer: Usually encoded in the table, too, but stored in an attribute/field that isn't made visible to you as a user.

How do we store information about the characteristics of spatial entities in GIS?

- *Attribute data* is typically **stored in tables** that are linked/associated with the spatial objects in your GIS.
  - Seeming exception: You've seen how GeoJSON and (the user-facing parts of) OpenStreetMap use *key=value* pairs to store attribute data *in lists, not in tables*. So, in OSM, the attributes for Smith Hall are:
    building=university,
    name="Smith Hall",
    operator="University of Washington", etc.
    But behind the scenes, OSM uses tables…

- These tables help us in searching/retrieving objects based on their characteristics and in performing various analyses.

Databases

- A *database* is a collection of data files that is organized to support data storage, manipulation, and retrieval.

- Databases are usually built/managed/used by a 'DBMS', a DataBase Management System.

- A 'spatial' database is usually at the core of a GIS program.

Types of databases

- Different types of databases are used to organize GIS data. Among others:
  - Flat file systems
  - Relational databases
  - NoSQL databases, including graph databases.

- Relational databases presently most widely used in GIS.

Some Terminology for Tables

Attribute
or Item

Record

Name	FIPS	Pop90	Area	PopDn
Whatcom	53073	128	2170	59
Skagit	53057	80	1765	45
Clallam	53009	56	1779	32
Snohomish	53061	466	2102	222
Island	53029	60	231	261
Jefferson	53031	20	1773	11
Kitsap	53035	190	391	485

Row = record = tuple

Column = field = item = attribute

A Flat File (one of the simplest forms of storage for a database system)

	Attribute	Attribute	Attribute
Record	Value	Value	Value
Record	Value	Value	Value
Record	Value	Value	Value

name	surname	address	phone #	order #	item	qty	item	qty	item	qty	item	qty
Leo	Durocher	112 Beal St	5-1307	1	CR7	1						
Rudy	Valentini	1 Hispanola Dr	4-2706	2	F15	1						
Paul	Smith	99 Upstate Ln	0-0000	3	GTO	3	F15	1	B52	1	SR71	1
Adam	Smith	1 Wall St	1-2334	4	626	1						
Atom	Ant	685 Hanbar Rd	4-1222	5	B52	2	CR7	2				
William	Smith	202 Dinkytown	9-9199	6	F111	2						
Alice	Paul	5 Free St.	4-4178	7	SR71	1						
Paul	Smith	99 Upstate Ln	0-0000	8	F15	1						

Relational databases

- Imagine: Dividing up a big flat file into a series of smaller files or ‘tables’.
- Each kind of attribute has its own table.
- Tables can later be **joined** in various combinations by an analyst. The analyst specifies a attribute that both tables have – a **key**.

A simple relational database example….

Property Table

PIN	Sq Ft	Zone
P101	244	resid
P102	5000	comm
P103	6790	mixed

Owner Table

PIN	Owner
P101	Smith
P102	Whoozit
P103	Whatzit

Relational Key

A more complicated example:

Patient File

Patient Record

Key	Check-in	Check Out	Room No.
42	2/1/96	2/4/96	N763
78	2/3/96	2/4/96	N712

Purchases File

Purchase Record

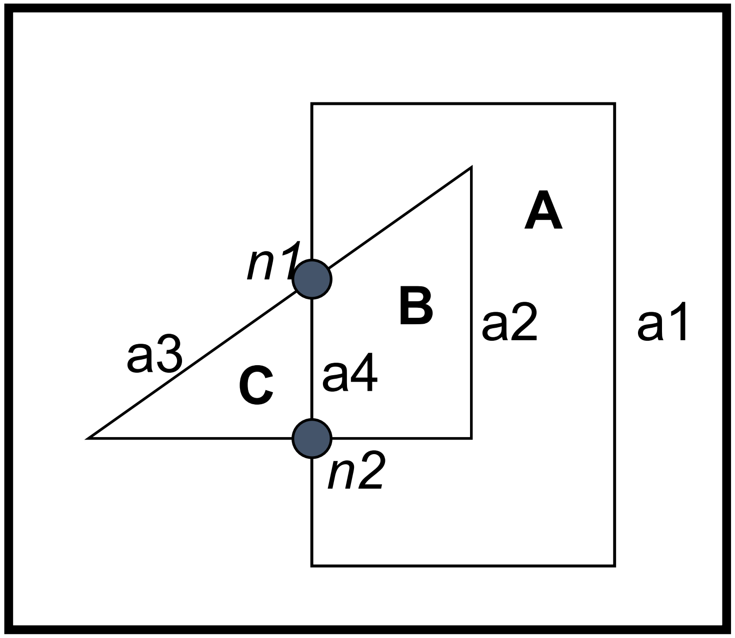
Item	Date	Price	Customer	Key
Skate Board	2/1/96	49.95	John Smith	42
Baseball Bat	2/1/96	17.99	James Brown	978

Accident Report

Date	Injury	Name	Key	Location
2/1/96	Broken Leg	John Smith	42	75 Elm Street
2/2/96	Concussion	Sylvia Jones	654	12 State Street
2/2/96	Cut on Ear	Robert Doe	123	2323 Broad Street

Accident File

Relational database structures can be used to implement the vector data model…



Arc Coordinate Data

Arc	StartXY	IntermediateXY	EndXY
a1	4,5	(4,8), (8,8), (8,1), (4,1)	4,3
a2	4,5	(6,7), (6,3)	4,3
a3	4,5	(1,3)	4,3
a4	4,3		4,5

Arc Topology

Arc	Start	End	Left	Right
a1	n1	n2		A
a2	n1	n2	A	B
a3	n1	n2	C	
a4	n2	n1	C	B

Node Topology

Node	Arcs
n1	a4, a2, a1, a3
n2	a2, a4, a3, a1

Polygon Topology

ID	Arcs
A	a1, a2
B	a2, a4
C	a3, a4

Why relational databases?

- Relatively simple to implement—well understood and ubiquitous in computing today.
- Can handle many data types.
- Encourages (requires) careful thinking in the design stages.
- Yet within the scope of a given design you've implemented, the relational database does <u>not</u> require prior knowledge about how you will *manipulate* data.

Disadvantages/challenges of relational databases

- In an era with “big data,” there may be practical limits to the scale at which data can be stored and queried in typical relational databases.

- Can perform poorly in representing and calculating with heterogeneous and dense relationships whose types may not be well understood ahead of time (compare: graph databases!)

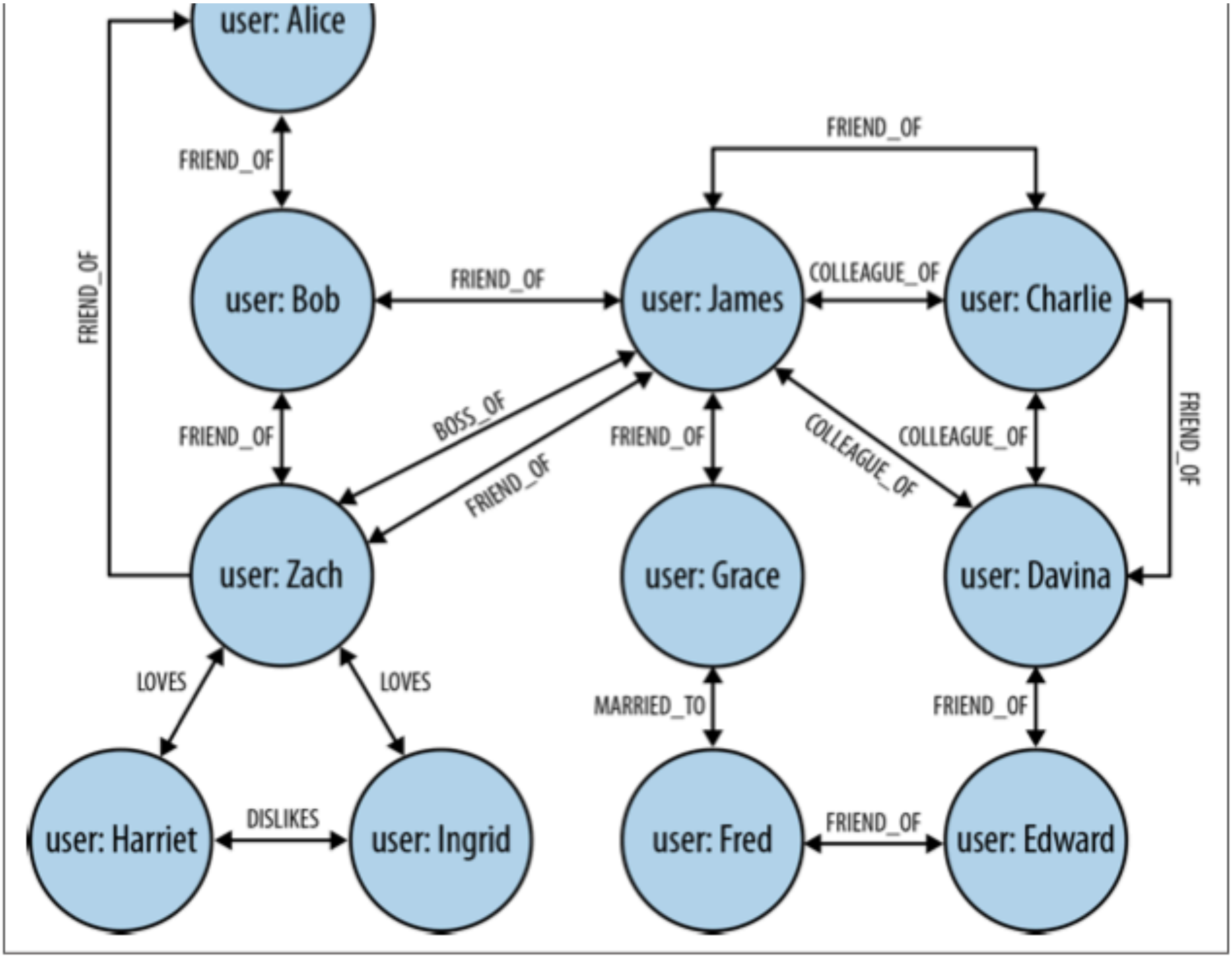
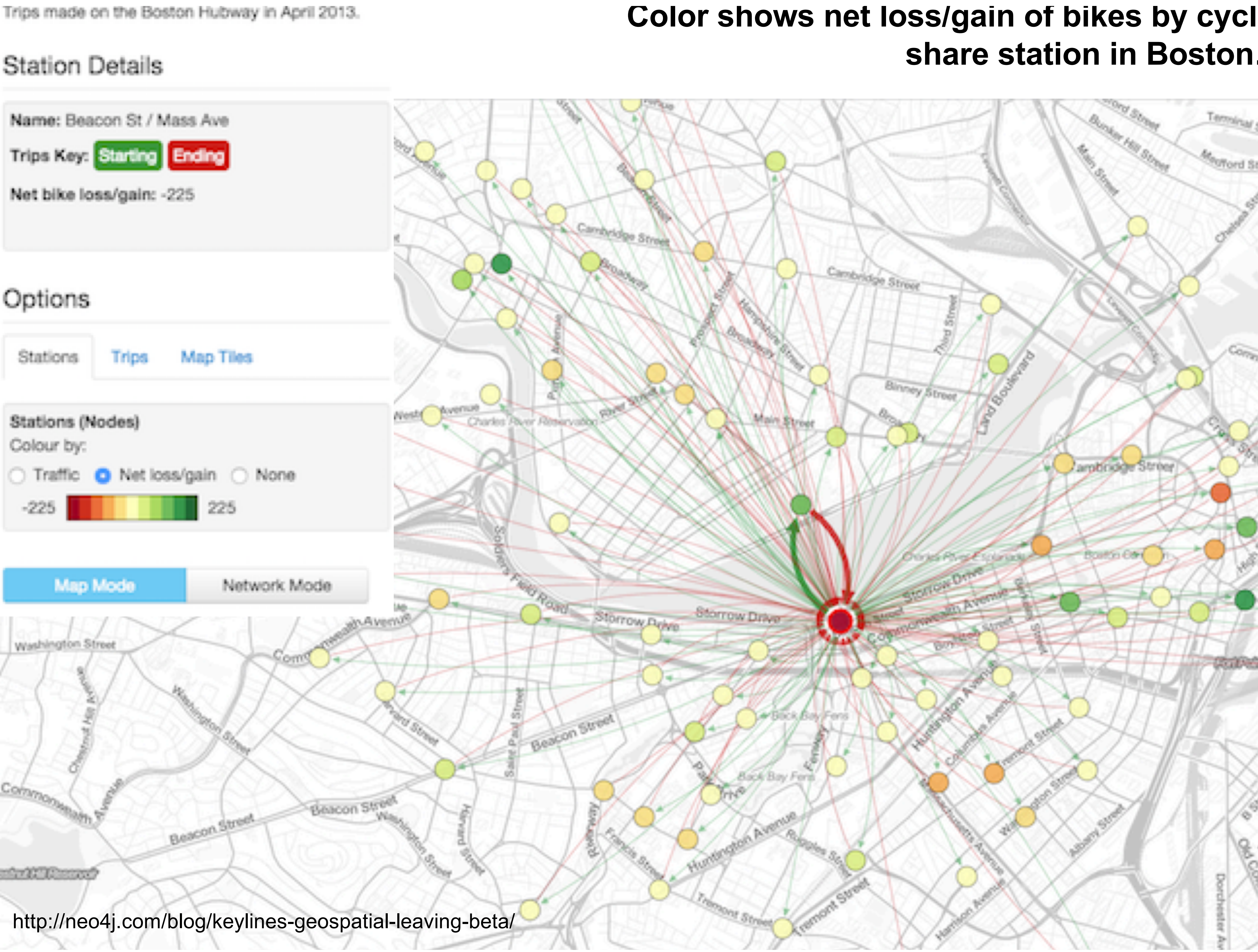


Figure 2-5. Easily modeling friends, colleagues, workers, and (unrequited) lovers in a graph

Robinson et al.'s 2013 *Graph Databases*

Color shows net loss/gain of bikes by cycle share station in Boston.



http://neo4j.com/blog/keylines-geospatial-leaving-beta/

DataBase Management Systems (DBMS)

- A DBMS is a software package that facilitates data storage, manipulation, and retrieval.

- A DBMS usually has:
  - Data definition language
  - Data dictionary [“metadata”]
  - Data entry & update modules
  - Report generator
  - Query language