Nearest-state-county-finder

JINGYI ZHANG, YUZHE XU, CYNTHIA YOUNG
Final report slide Dec 8th, Boston University

Project Abstract

- This project is to create a system, which can efficiently find the nearest state or county of the given input location.
- In this system, users are asked to input the latitude and longitude of the targeted location, and are required to input k (1<= k <=10) as number of reference points they want to find.

    - To implement this, we loaded input data from USGS.
    - We used the Haversine Distance formula to compute distance
    - We used a KD tree to do the nearest neighbor search, and used majority voting to narrow down the state and county.

Files

Data < 2023_Gaz_counties_national.csv> : Downloaded from USGS. It includes the state (USPS) , county(NAME) , LATITUDE(INTPTLAT) and LONGITUDE (INTPTLONG), which are needed for finding nearest K reference points. We converted it from txt to the csv format to better extrapolate.

USPS	GEOID	NSICOD	NAME	ALAND	AWATER	ALAND_SQMI	AWATER_SQMI	INTPTLAT	INTPTLONG
AL	1001	161526	Autauga County	1539631461	25677536	594.455	9.914	32.532237	-86.64644
AL	1003	161527	Baldwin County	4117725048	1132887203	1589.863	437.41	30.659218	-87.746067
AL	1005	161528	Barbour County	2292160151	50523213	885.008	19.507	31.870253	-85.405104
AL	1007	161529	Bibb County	1612188713	9572302	622.47	3.696	33.015893	-87.127148
AL	1009	161530	Blount County	1670259100	14860281	644.891	5.738	33.977358	-86.56644
AL	1011	161531	Bullock County	1613083468	6030667	622.815	2.328	32.101759	-85.717261
AL	1013	161532	Butler County	2012002548	2701199	776.839	1.043	31.751667	-86.681969
AL	1015	161533	Calhoun County	1569248377	16534047	605.89	6.384	33.770516	-85.827909
AL	1017	161534	Chambers County	1545068570	16988729	596.554	6.559	32.915504	-85.394032

load.cpp: source code for finding the nearest state and county.

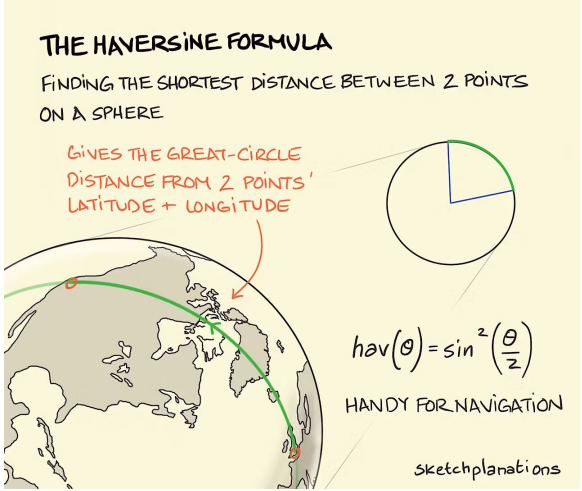
makefile: file used to compile load.cpp (make / make run)

input/output: .txt files for testing and outputting the result

Method of Code

Distance Computation-Haversine Distance:

- The distance between two points (reference point and input point) is calculated using the **haversine distance**, which involves latitude and longitude.
  a = sin²(Δlat/2) + cos(lat1) * cos(lat2) * sin²(Δlon/2)
  c = 2 * atan2(sqrt(a), sqrt(1-a))
  distance = R * c
  R is the radius of the Earth
  Haversine formula takes into account the curvature of the Earth's surface, making it more accurate for longer distances

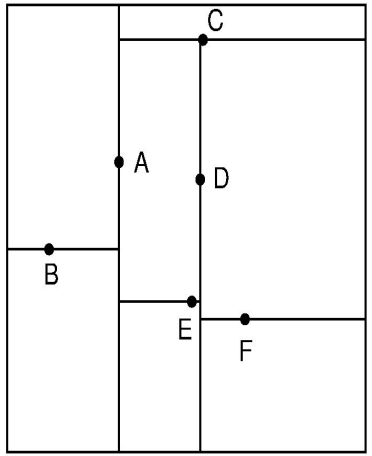


State and County Identification -majority voting :

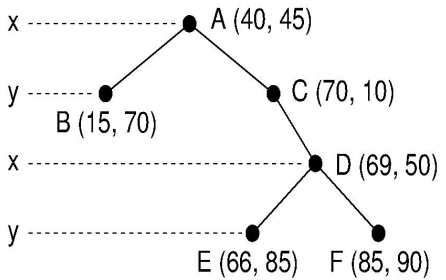
- Once the nearest K reference points are found, the system , perform a **majority voting** among the K nearest points, to determine the state of the searched point.

Method of Code: KD Tree

- KD-Tree is a data structure that is particularly suitable for multi-dimensional space search queries
- Efficient space division
  - When build the tree, insert points, effectively divides the space into smaller regions, each containing similar points
- Fast nearest neighbor search, using BST property
  - Suitable for nearest neighbor search
  - Quickly narrow down the search area



(a)

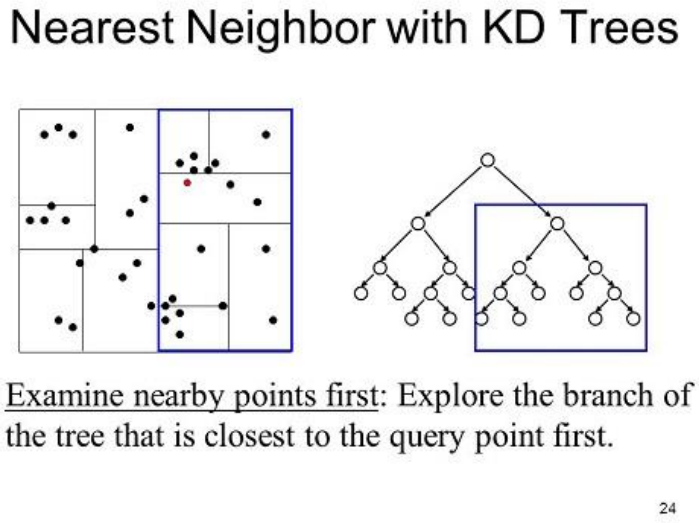


(b)

Method of Code: Nearest Neighbor Search

- Checks if the current node is closer and then selects a subtree to search deeper. If there are closer points in the subtree then search path will be explored further (pruning)

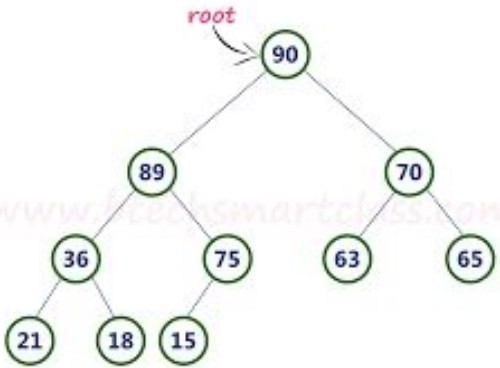
- If the subtree on the other side is likely to contain a closer point, then that subtree is also searched.



Method of Code: Max Heap

- We used Max heap to store the closest points found and their distance from the query point. It allows us to efficiently keep track of the closest points found so far.

1. Compare new node with top of queue
2. If new node distance < top one
3. Top pop up
4. New node push in



Input Output Example vs Real Points

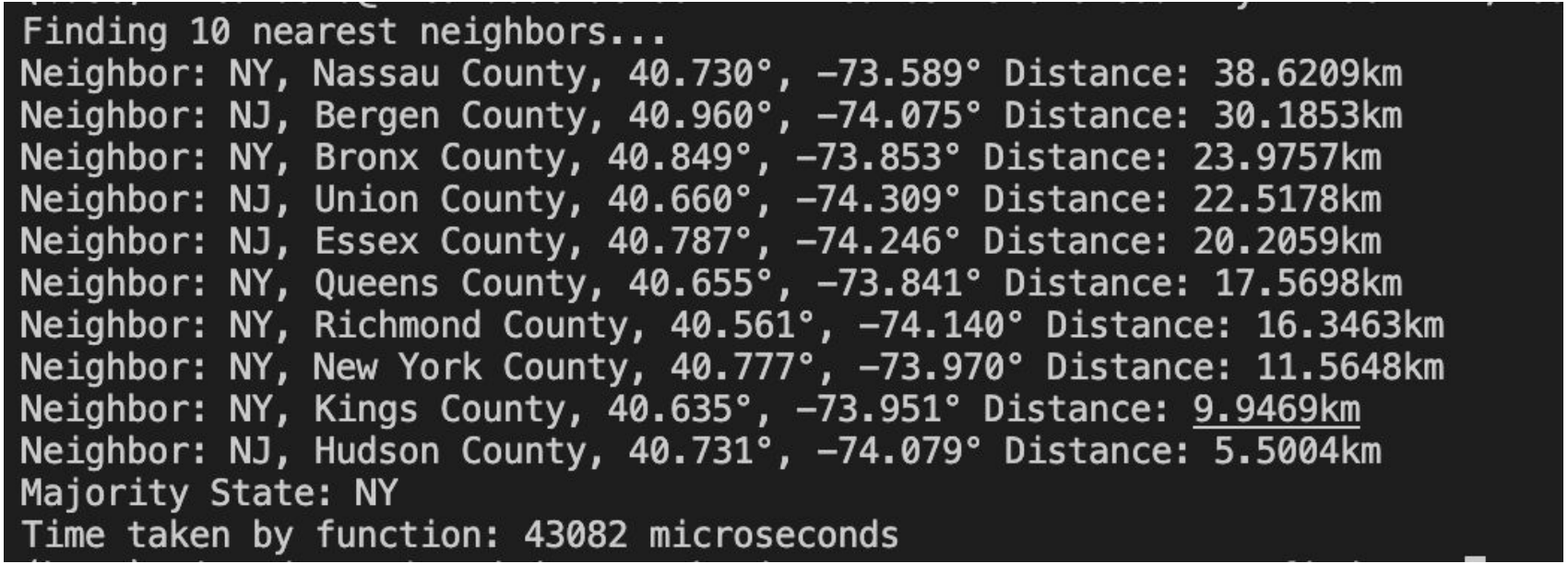
```
Finding 10 nearest neighbors...
Neighbor: RI, Bristol County, 41.707°, -71.287° Distance: 73.0934km
Neighbor: NH, Rockingham County, 42.989°, -71.099° Distance: 71.0281km
Neighbor: MA, Worcester County, 42.312°, -71.940° Distance: 68.9414km
Neighbor: MA, Bristol County, 41.749°, -71.089° Distance: 66.8733km
Neighbor: RI, Providence County, 41.870°, -71.579° Distance: 66.2446km
Neighbor: MA, Plymouth County, 41.987°, -70.742° Distance: 50.1298km
Neighbor: MA, Essex County, 42.643°, -70.865° Distance: 37.9704km
Neighbor: MA, Middlesex County, 42.482°, -71.395° Distance: 28.0771km
Neighbor: MA, Norfolk County, 42.172°, -71.181° Distance: 20.8224km
Neighbor: MA, Suffolk County, 42.339°, -71.018° Distance: 7.0945km
Majority State: MA
Time taken by function: 49504 microseconds
```

BU CDS Building:

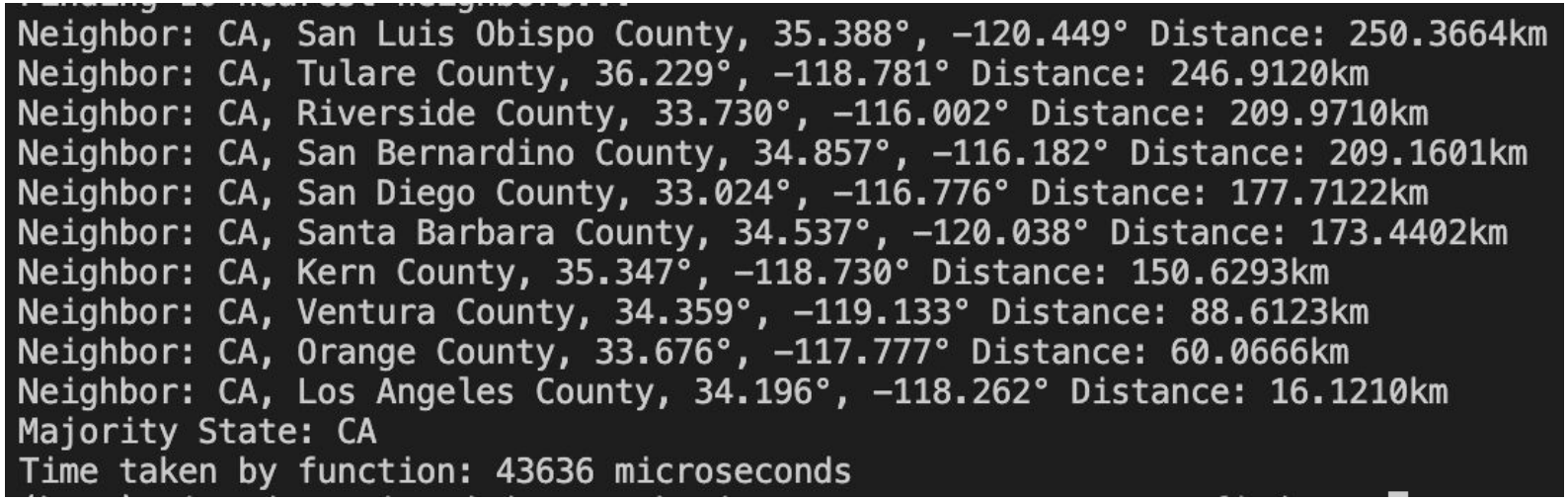
42.3499° N, 71.1032° W (42.3499 / -71.1032 / 10 as input)

State: MA (Massachusetts)

County: Suffolk



Statue of Liberty
coordinate: 40.6892° N, 74.0445° W (40.6892 / -74.0445 / 10 as input)
State: NY (New York)
County: New York County



LA county

coordinate: 34.0522° N, 118.2437° W , K=10 (34.0522 / -118.2437 / 10 as input)

State: CA (California)

County: Los Angeles County

```
Finding 10 nearest neighbors...
Neighbor: CA, Sonoma County, 38.525°, -122.926° Distance: 3843.1936km
Neighbor: CA, Marin County, 38.051°, -122.746° Distance: 3840.7220km
Neighbor: CA, Mendocino County, 39.432°, -123.443° Distance: 3836.5870km
Neighbor: CA, San Francisco County, 37.727°, -123.032° Distance: 3805.4267km
Neighbor: AK, Aleutians East Borough, 55.245°, -161.997° Distance: 3782.6591km
Neighbor: HI, Hawaii County, 19.598°, -155.502° Distance: 322.0059km
Neighbor: HI, Kauai County, 22.012°, -159.706° Distance: 194.9330km
Neighbor: HI, Maui County, 20.856°, -156.602° Distance: 151.1400km
Neighbor: HI, Kalawao County, 21.219°, -156.974° Distance: 102.7453km
Neighbor: HI, Honolulu County, 21.461°, -158.202° Distance: 27.9637km
Majority State: HI
Time taken by function: 42917 microseconds
```

Pearl Harbor in **Hawaii** **(21.36 / -157.95 / 10 as input)**

Latitude: 21.362374050539042

Longitude: -157.95365360883602 , This is an edge test