

Conceptual Problem Set for Spatial Data Science

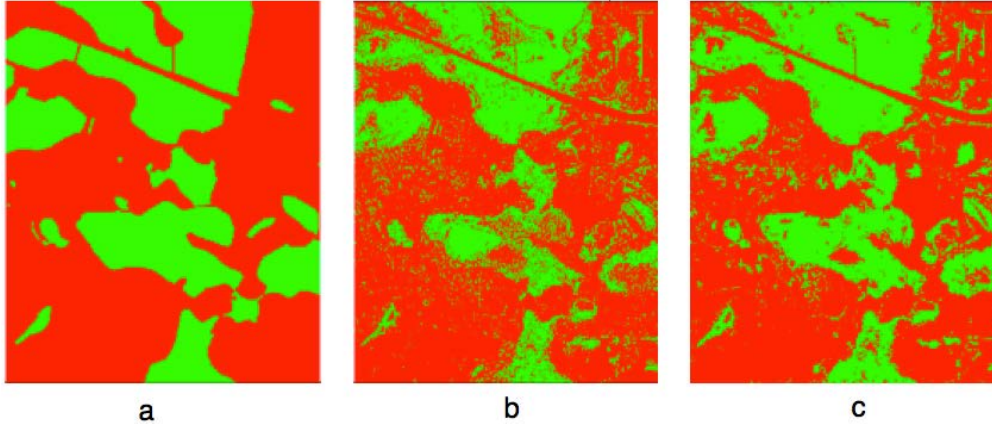
1. Which violate independence assumption?
 - a) "Mohamed Lee" is a rare name, even though "Mohamed" is the most frequent first name and "Lee" is the most frequent last name.
 - b) Near things are more related than distant things.
 - c) Nearby video frames often show common people and objects.
 - d) All of the above
2. Which does not illustrate spatial-hotspot pattern family?
 - a) Roads with unusually high rate of traffic accidents
 - b) Areas with unusually high concentration of museums
 - c) Neighborhood with unusually high rate of an infectious disease (or crime)
 - d) Country with largest geographic area
3. Which does not illustrate spatial-outlier pattern family?
 - a) A dozen pine trees growing among thousands of cypress
 - b) A historical palace (e.g., Forbidden City) in the center of a modern city (e.g., Beijing)
 - c) Waterfalls in the river Mississippi
 - d) Vast plains uninterrupted by hills (e.g., parts of Pampas, Argentina away from hills)
4. Which does not illustrate spatial co-occurrence (e.g., colocation)?
 - a) Loud sound temporally follows bright flash of a lightning.
 - b) Nuclear power plants usually locate near water
 - c) Egyptian plover birds live close to Crocodiles
 - d) College campuses often have bookstores nearby.
5. Which of the following are properties of spatial data?
 - a) Autocorrelation
 - b) Heterogeneity
 - c) Implicit relationships (e.g., neighbor)
 - d) All of the above

Answer Key for Conceptual Quiz

1. (d)
2. (d)
3. (d)
4. (a) It is coincidence in time
5. (d)

Technical Problem Set for Spatial Data Science

1. Which of the following three figures exhibits highest spatial autocorrelation?

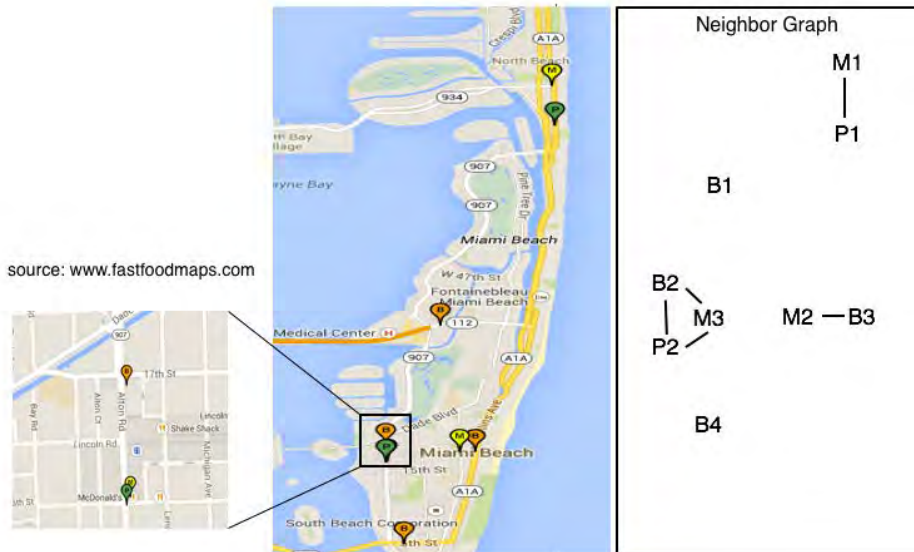


- a) Figure a
- b) Figure b
- c) Figure c
- d) They all have the same level of spatial auto-correlation

2. Which violate the identical distribution assumption underlying traditional methods?

- a) Cancer cell heterogeneity makes treatment of cancer difficult.
- b) No two places on the Earth are exactly alike.
- c) All politics is local.
- d) All of the above

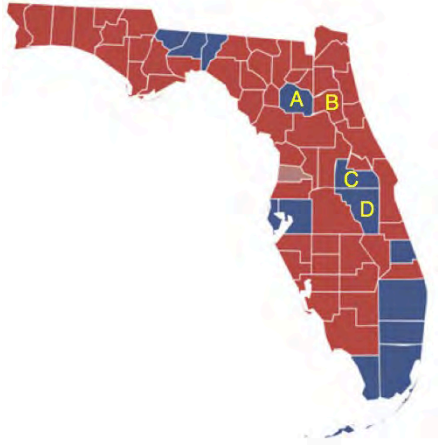
3. Consider following map (left half) and neighbor graph (right half) of fast-food restaurants by Burger King (B), McDonald's (M), and Pizza Hut (P).



Which pair has the highest participation index (an interest measure for colocation)?

- a) (Burger King, Pizza Hut)
- b) (Pizza Hut, McDonald's)
- c) (Burger King, McDonald's)
- d) They all have the same value of participation index

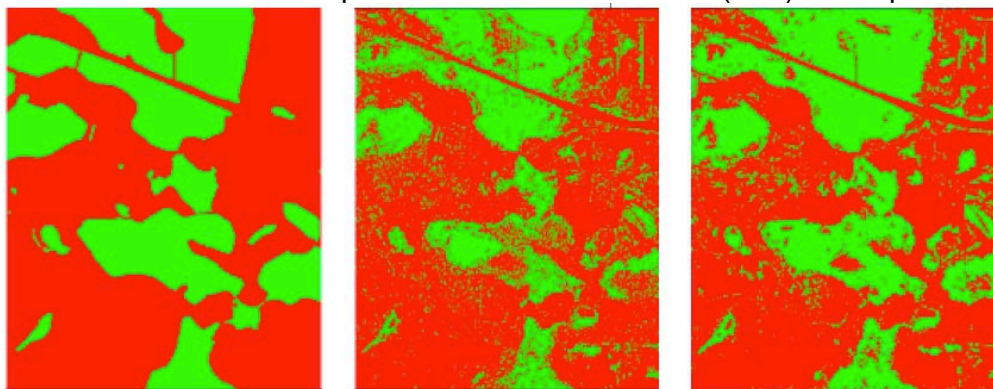
4. The following figure shows the 2012 United States president election results (blue vs. red) for Florida counties, e.g., A, B, C, D, etc.



Which county is a spatial outlier in the above map?

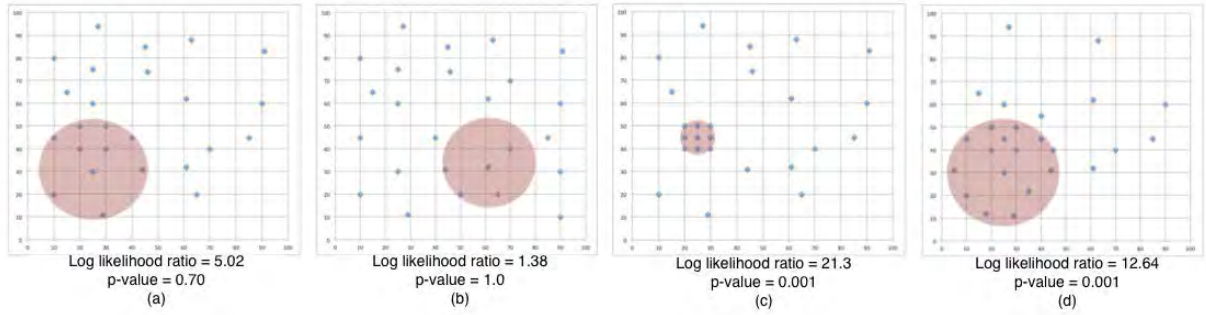
- a) County A
- b) County B
- c) County C
- d) County D

5. Assuming map A represents ground truth, which map is more likely a results from traditional model with independent identical distribution (i.i.d.) assumption?



- a) Figure b
- b) Figure c

6. Consider following four datasets, each with a candidate circles with its log likelihood ratio and p-value.



Which figures best illustrate statistically-significant hotspots?

- a) Figure a and Figure b
- b) Figure b and Figure c
- c) Figure c and Figure d
- d) Figure d and Figure a

Answer Key for Technical Quiz

1. (a) It has lowest level of salt and pepper noise
2. (d)
3. (P,M). Its participation Index is 2/3
4. A
5. (b) It has higher salt and pepper noise
6. (c)