

Applied Spatiotemporal Data Mining

Day 1: Overview

Guofeng Cao

www.spatial.ttu.edu



Department of Geosciences

Texas Tech University

guofeng.cao@ttu.edu

Chinese University of Geosciences, June 2018



Course Description

Content overview

- This course will introduce concepts and commonly used methods in quantitative analysis of (geographic) spatial data
- Contents include:
 - Characteristics of spatial data
 - Pattern recognition methods of spatial data
 - Spatiotemporal statistical analysis methods
- Class webpage: <https://github.com/surfcao/summer2018-cug>



Course Description

Audience

- This class is intended for students (undergraduate and graduate students) from relevant disciplines (e.g., geography, geology, environmental science and social sciences) who are interested in analysis of spatial data
- Students will be encouraged to engage this course with their thesis/dissertation topics and research interests

Please note:

- General knowledge of statistics or quantitative skills will be very helpful and equations will appear significantly in some lectures
- We have a mix of undergraduate and graduate students with diverse background and expectation



Course Description

Course objectives

- After completing this course, undergraduates students are expected to learn how to:
 - formulate real-world problems in the context of geographic information systems and spatial analysis
 - utilize mainstream software tools (commercial or open-source) to solve spatial problems
 - communicate results of spatial analysis in the forms of writing and presentation
- In addition to the above, graduates students are expected to:
 - have a good understanding of spatial analytical methods
 - apply the introduced methods in the dissertation and thesis research
 - evaluate and assess the results of alternative methods



Course Format

Lectures

- Instructor: Guofeng Cao (guofeng.cao@ttu.edu)
- # 3 Building 604
- 2:30-5:30pm
- TAs: Zehao Yuan and Jianfeng Zhou



Lab Assignments

Lab assignments

- One assignment per day
- R



Final Project

Final project

- The project could be used as a setting for your thesis and dissertation topics, other course topics or research interests
- Group collaboration is encouraged, but for each group, no more than two graduates are allowed
- Start to think of the project ideas early and communicate with the instructor and TA for comments
- Project presentation or poster session: *PechaKucha* style
<http://en.wikipedia.org/wiki/PechaKucha>
- Project report: no more than 8 pages with single space and size 12 font



Grading policy

- Four lab assignments: 60%
- Final project: 40%
- Class and lab attendance is mandatory



- Required
 - O'Sullivan, David and David J. Unwin, 2010. Geographic Information Analysis
 - Cressie, N., & Wikle, C. K. (2011). *Statistics for Spatio-temporal Data*. John Wiley & Sons.
 - Bivand Roger S., Pebesma, Edzer J., and Gmez-Rubio, Virgilio (2008), *Applied Spatial Data Analysis with R*, Springer.
- Optional:
 - de Smith, Michael J., Paul A. Longley and Michael F. Goodchild (2013), *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, 4th Edition. Available in both print and web (*free!*) version at <http://www.spatialanalysisonline.com>



Quiz

Please try to answer the following questions



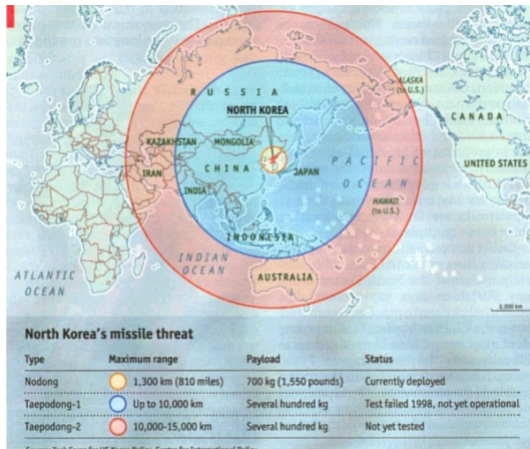
Quiz

1. Which one of the following best approximate the radius of the Earth?
 - (a) 4356 km
 - (b) 6356 km
 - (c) 8356 km
2. Which one of the following best approximate the latitude of Wuhan or this classroom?
 - (a) 30°
 - (b) 40°
 - (c) 45°
3. How far does one degree of latitude approximate on the ground?
 - (a) 10 km
 - (b) 110 km
 - (c) 1100 km
4. GPS device on my smart phone gave me the reading (47.640120461583138, -122.12971039116383). Does it make sense?



Quiz

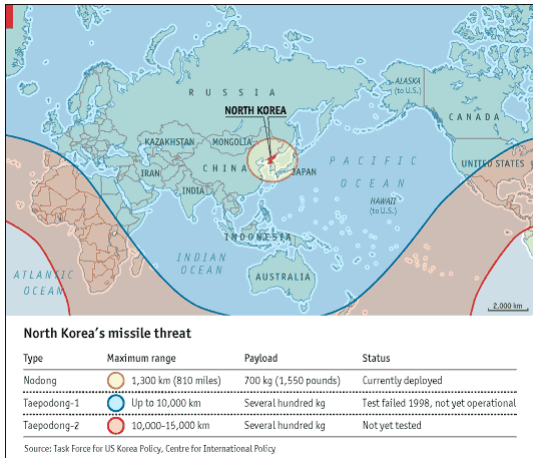
6. What is buffer analysis?
7. How would you find out the total population resided within 5 km of highway in China?
8. The following is a map published in an article of *The Economist* describing the missile threat of North Korea. Is there anything wrong with the following map?





Quiz

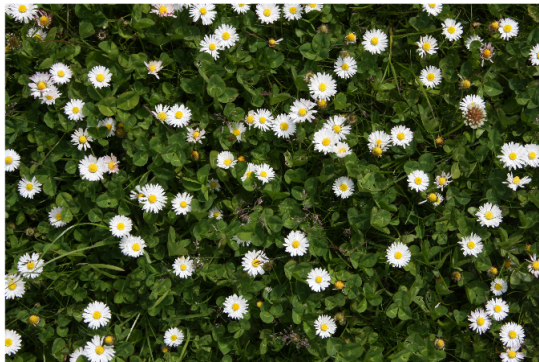
Then *The Economist* realized it, and gave the corrected map as the following.





Quiz

9. Does the distribution of these daisies look randomly for you?



10. Suppose for counties of Texas, we found that the counties that consume more coffee have less cases of heart diseases. Can we conclude that drinking more coffee causes less heart diseases?



Thank you

Questions/comments?